

Enquiries should be addressed to:

Keith R Hayes  
CSIRO Mathematics, Informatics and Statistics  
GPO Box 1538 Hobart, Tasmania, Australia, 7001  
Castray Esplanade, Hobart, Australia, 7000  
Telephone : +61 3 6232 5260  
Fax : +61 3 6232 5485  
Email : [keith.hayes@csiro.au](mailto:keith.hayes@csiro.au)

### **Distribution List**

Client	(1)
Publications Officer	(X)
Stream Leader	(1)
Authors	(1)

### **Copyright and Disclaimer**

© CSIRO To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of CSIRO.

### **Important Notice**

CSIRO advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, CSIRO (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

# Contents

<b>1 Introduction</b>	8
1.1 Project background	8
1.2 Report structure and outline	8
1.3 Fundamental issues	10
1.4 Notation	11
<b>2 Risk and uncertainty</b>	14
2.1 What is risk?	14
2.2 Why and how is risk assessed?	17
2.3 What is uncertainty?	21
2.3.1 Linguistic uncertainty	23
2.3.2 Variability	25
2.3.3 Epistemic uncertainty	26
<b>3 Uncertainty analysis</b>	28
3.1 Linguistic uncertainty	28
3.2 Epistemic uncertainty and variability	29
3.2.1 Methods overview	29
3.2.2 Model uncertainty	32
3.2.3 Completeness	38
3.2.4 Scenario uncertainty	38
3.2.5 Subjective judgement	38
3.2.6 Measurement error	42
3.2.7 Parametric uncertainty and sampling uncertainty	42
3.3 Dependence	48
<b>4 Uncertainty analysis methods</b>	60
4.1 Analytical methods	60
4.2 Probabilistic methods	62
4.2.1 Monte Carlo simulation	62
4.2.2 Probability boxes and probability bounds analysis	66
4.3 Graphical methods	73
4.3.1 Bayesian networks	73
4.3.2 Loop analysis	77
4.3.3 Fuzzy cognitive maps	81
4.4 Non-probabilistic methods	85
4.4.1 Fuzzy set theory	85
4.4.2 Interval analysis	86
4.4.3 Info-gap theory	88
<b>5 Uncertainty analysis in practise</b>	93
5.1 Who's using what?	93
5.2 Methods evaluation	96

5.2.1	Assumptions . . . . .	96
5.2.2	Technical complexity . . . . .	97
5.2.3	Software availability . . . . .	99
5.2.4	Decision utility . . . . .	99
<b>6</b>	<b>Discussion and recommendations . . . . .</b>	<b>102</b>
6.1	Uncertainty and qualitative risk assessment . . . . .	102
6.2	Honest quantitative risk assessment . . . . .	103
6.3	Uncertainty, statistics and quantitative risk assessment . . . . .	104
6.4	A recommended strategy for uncertainty analysis . . . . .	106

## Acknowledgements

In completing this report I have been assisted and encouraged by friends, colleagues and collaborators. Simon Barry (CSIRO) and Gareth Peters (UNSW) have patiently guided me through modern statistical methods, and commented on various aspects of the report. Scott Ferson introduced me to Probability Bounds Analysis and provided me with a copy of his pbox R libraries. The PBA analysis presented in this report would not have been possible without his generosity.

My Hobart team - Jeff Dambacher (CSIRO) and Geoff Hosack (CSIRO) - have been a constant source of support and inspiration. They also introduced me to the concept of qualitative modeling. Jeff contributed to Appendix B of this report and Geoff contributed to the discussion on fuzzy cognitive maps in Section 4.3.

I would like to thank Greg Hood (Bureau of Rural Science) and Petra Kuhnert (CSIRO) for their encouragement, comments along the way and programming advice. Alan Walsh (ANU) and Xunguo Lin (CSIRO) acted as sounding boards for half-baked ideas and helped clarify my thoughts. Brent Henderson (CSIRO), Yakov Ben-Haim (Technion), Scott Ferson (Applied Biomathematics), James Franklin (UNSW) and two anonymous reviewers provided comments that have improved the report. Lynda Watson (CSIRO) kindly assisted with proof reading.

This report is the product of the Australian Center of Excellence for Risk Analysis (ACERA) and CSIRO. In preparing this report, the author acknowledges the financial and other support provided by the Australian Government Department of Agriculture, Fisheries and Forestry (AGDAFF) and the University of Melbourne.

## Executive summary

This report reviews uncertainty and uncertainty analysis methods in risk assessment, with a specific focus on issues related to import risk assessment. The report is motivated by the availability of qualitative and quantitative methods for import risk assessment. It examines how the challenges posed by uncertainty influence choices between these two approaches. The project's terms of reference are to summarise and categorise the different sources of uncertainty in risk assessment problems, and review the practicality and applicability of a range of treatment methods. The report is intended for scientists and managers involved in, or contemplating the use of, qualitative or quantitative risk assessment. Whilst the report focusses on import risk assessment, readers from other application domains will find that much of the information and analysis presented here is relevant to them.

Uncertainty is a term used to encompass many concepts. It has been described, defined and categorised in many different ways, using different names for the same things and occasionally the same name for different things. This report identifies four basic sources of uncertainty: uncertainty that arises through the vagarious nature of language (linguistic uncertainty), the uncertainty created by our limited understanding of natural systems (epistemic uncertainty), the uncertainty created by the irreducible variation in these systems (variability), and finally the uncertainty associated with our value systems and management decisions (decision uncertainty).

The report examines in detail the various sources of linguistic uncertainty, epistemic uncertainty and variability in scientific endeavors and risk-related problems. It summarises probabilistic, non-probabilistic and graphical methods for treating and propagating these sources of uncertainty through risk assessment under the headings of five basic strategies: ignore it, eliminate it, envelope it, average over it or factorise it. It also examines the related problem of dependency that occurs when arithmetic operations are performed with random variables.

The principal impediment to uncertainty analysis within qualitative risk assessment is that variability and epistemic uncertainty are confounded with each other and with linguistic uncertainty. Separating the three sources of uncertainty requires, as a minimum, that linguistic uncertainty is eliminated from the problem as far as possible. Fuzzy sets and possibility theory provide a mechanism that was specifically designed to eliminate two important sources of linguistic uncertainty (vagueness and ambiguity). These sources of uncertainty, however, can also be eliminated with probability theory via carefully implemented elicitation methods and probability bounds analysis. This approach has the additional advantages of: a) being able to minimise other well known heuristics and biases in human perception and judgements of uncertain events; and, b) couching its analysis within the realms of probability theory which is likely to be more familiar to decision makers than evidence or possibility theory.

Furthermore, there are a range of issues with qualitative approaches to risk assessment that relate to the science-quality criteria of transparency, repeatability and falsifiability, and the decision-utility criteria of precision and accuracy, namely:

- qualitative risk assessment predictions cannot be (in)validated with observations, and uncertainty cannot be coherently propagated through risk functions without translating qualitative metrics of likelihood or consequence into numerical metrics;
- arithmetic operations, such as product, performed with linguistic descriptions of likelihood and/or consequence can be biased and non-commutative. These problems only become apparent when vagueness is eliminated from the analysis using numerical defi-

nitions of terms such as “high”, “medium” or “low”. This report uses interval analysis to show that the qualitative risk operations performed by the AGDAFF for import risk assessment lead to biased and non-commutative results; and,

- the effects of dependency between risk-generating events cannot be coherently explored with qualitative descriptions of uncertain events, leading to different interpretations of observed outcomes and potentially paradoxical assumptions about the relatively likelihood of events.

The problem of epistemic uncertainty and variability in quantitative risk assessment starts with the availability of data. This report distinguishes data - observations of a process - understanding and beliefs about a process. The key to uncertainty analysis in the absence of data is elicitation. Elicitation converts beliefs into outcomes, models or parameters that enable regression and forward propagative uncertainty analysis. Again, structured elicitation techniques also provide an opportunity to avoid or minimise heuristic bias and dysfunctional group effects, and are therefore “good practice” in quantitative risk assessment.

Import risk assessment is often performed in the absence of empirical observations. Uncertainty analysis in this context can be performed using elicitation and forward uncertainty propagation methods. The key challenges associated with epistemic uncertainty and variability can be addressed or eliminated by making assumptions about the structure of the risk-generating process (the model), the shape, scale and/or location of the probability distribution(s) that represent variability and/or epistemic uncertainty in model parameters, and the nature of the dependency between the parameters of the model. These types of assumptions have an important bearing on the results of the risk assessment. The principal objectives of propagative uncertainty analysis are to report, and where practical, test the effect of these assumptions on the overall risk estimate, and in doing so achieve an honest assessment. In approaching this problem the analyst can adopt one or more of the strategies identified above, namely:

- simply ignore: this is sometimes defensible for parametric uncertainty and model structure uncertainty, but only in limited circumstances. For example, model structure uncertainty can be ignored where the model or risk function is dictated by legislation or guidelines. In this context the risk assessment results are only defensible as a guide to the relative magnitude of risk but this can be useful as a risk-screening decision aid;
- eliminate: this is possible for variability (and to a limited degree dependence) by either building a more complex risk model that models variability or via choosing a simpler assessment endpoint that enables a simpler risk model with a lower parameter dimension. The first approach may not be attractive in data-limited circumstances because it can increase model structure uncertainty. The latter approach is only tenable if meaningful decision criteria can be stipulated by a decision maker for the simpler endpoint;
- compare and envelope: comparative strategies are akin to sensitivity analysis and seek to highlight the effect of assumptions on risk estimates. Enveloping methodologies place bounds on the best and worst estimates and seek to guarantee that the true result will lie within these bounds. Interval analysis, probability boxes and probability bounds analysis can satisfy this guarantee for variability and dependence with minimal conditions, for example, that the true value of an uncertain quantity lies within an elicited interval. Over-confident expert opinion is clearly a challenge in this regard. Info-gap theory attempts to place an upper bound on the effects of uncertainty on decisions but its recommendations may be sensitive to initial conditions. Moreover, comparison and enveloping cannot

guarantee that the effects of model structure uncertainty have been completely addressed because the set of possible models is infinite. In the absence of data and statistical inference this problem is unconstrained but is best approached by consulting widely and comparing as many plausible models as is possible within the resources available to the study. Techniques such as influence diagrams, loop analysis and fuzzy cognitive maps can be helpful in this context;

- average over: an analyst can average over variability and several sources of epistemic uncertainty, including model structure uncertainty, using techniques such as second-order Monte Carlo Simulation and Bayesian model averaging. Again, however, in the absence of data this problem is unconstrained and computationally more demanding than the compare and envelope strategy. The range of plausible model structures, or alternative probability density functions, that can be addressed within the resources of a single study are therefore likely to be smaller with this strategy, and in the case of model structure uncertainty it can lead to risk estimates that are incompatible with accepted theories; and.
- model and factorise: this strategy is applicable to variability and dependence and in the presence of data also provides a means to identify parsimonious descriptions of cause and effect and thereby treat model structure uncertainty. Copulas and Bayesian networks can be used to treat dependence and partition different sources of variability in a risk assessment problem. These methods can be used in data-poor situations but the lack of data-based constraints can still undermine attempts to provide a systematic analysis. Moreover, the full benefits of Bayesian networks, and statistical graph theory and hierarchical modeling in general, cannot be realised in the absence of data.

Virtually all risk assessment frameworks emphasise the importance of monitoring and review, and this report emphasises that scientifically credible risk assessments should ultimately become a statistical exercise by making predictions about the risk-generating process that are testable and eventually tested against observations. Any historical distinction between uncertainty analysis and statistics is not constructive in a risk assessment context. A more useful distinction is to consider forward uncertainty propagation methods as the initial tools of uncertainty analysis that enable honest assessments to proceed before observations are made and data are collected. Thereafter, uncertainty analysis should increasingly move to an inferential mode that relies on statistics and uncertainty analysis methods to characterise and quantify variability and epistemic uncertainty in data sets relevant to the problem in hand.

A synthesis of these discussions, together with an examination of the pros and cons of different uncertainty analysis methods, suggests the following overall strategy for uncertainty analysis in import risk assessment:

1. use formal elicitation techniques to canvass the opinions, construct conceptual models and parameterize the beliefs of stakeholders and experts. Use either predictive or structural elicitation methods to convert conceptual models into statistical, qualitative and/or mechanistic models and convert beliefs about stochastic variables into numerical intervals with assigned levels of confidence;
2. ensure feedback is embedded within the elicitation procedure (to minimise the potential for misunderstanding) and apply an advocacy-like procedure to ensure that all aspects of the risk assessment are rigorously reviewed;
3. state risk-decision criteria (risk acceptability levels) in a numeric, measurable fashion for



as many of the steps in the risk-generating process as is possible, including steps leading up to the overall assessment endpoint;

4. maintain plausible diverse opinions and in the first instance envelope this diversity using techniques such as loop analysis, comparisons of alternative risk functions, interval analysis, probability boxes and probability bounds analysis. If the upper bound on the subsequent risk estimate is lower than the decision criteria associated with the assessment endpoint, report the result and consider the need for monitoring strategies that enable (in)validation of as many of the steps in the risk-generating process as possible within the resources available to the assessment. If possible, collect data and use statistical inference methods to check that the risk-generating process is operating within the bounds predicted for each step of the process by the risk assessment;
5. if the lower bound on the enveloped risk estimate is higher than the decision criteria associated with the assessment endpoint consider prohibiting, stopping or otherwise mitigating the risk-generating process and if necessary repeat the risk assessment with risk management steps in place, and include within the assessment the impact of management, and the effects of decision uncertainty upon this; and,
6. if the upper and lower bounds of the enveloped risk estimate straddle the decision criteria associated with the assessment endpoint consider first the effects of dependence and the mitigating effects of positive or negative dependence. For example, a potential application of positive quadrant dependence arises in import risk assessment because the probability of detecting organisms at the border should be positively dependent on the number of organisms that arrive at the border - i.e. as the number of infected units rises so should the probability of their detection. Treating these events as independent denies the reality of inspection regimes, inflates uncertainty bounds and can lead to paradoxical simulations where large numbers of infected units are multiplied by a small probability of detection (and vice-versa) in naive simulations.

The strategy outlined above is designed to enable uncertainty analysis with the minimum amount of assumptions. The objective here is for the assessment to be roughly right rather than precisely wrong. If the enveloped predictions continue to cross the assessment endpoint then two avenues are available to the assessor and manager:

- consider prohibiting or otherwise deferring the risk-generating process and the risk assessment, and collect data that enables statistical inference and a more precise empirical estimate of the risk function (statistical or mechanistic model) and/or the variables (risk factors) associated with this model; or,
- use the most plausible assumptions about the model structure and its variables to provide a more precise risk prediction by modeling and/or factorising the uncertainty associated with the problem using techniques such as Bayesian Networks and second-order Monte Carlo Simulation supported by linear or non-linear estimates of dependence.

It is very important with the second option that the assumptions associated with the analysis are clearly communicated together with the effects of alternative plausible assumptions on the risk estimate where possible. It is also important that monitoring strategies are designed and implemented in order to (in)validate as many of the steps in the risk-generating process as possible and thereby enable a gradual departure from data-poor circumstances to data-rich circumstances, and a move towards the inference opportunities of modern statistical methods.

# 1 Introduction

## 1.1 Project background

In October 2007 the Australian Center of Excellence for Risk Assessment (ACERA) entered into a two year collaborative research agreement with the CSIRO Division of Mathematics, Informatics and Statistics (CMIS), together with the Australian Bureau of Rural Sciences (BRS), the Australian National University (ANU) and Applied Biomathematics. The research project (ACERA reference 0705) funded by this agreement is divided into three parts. Part I explores some of the practical issues that arise when Monte Carlo Simulation is applied to quantitative, pathway-based, import risk assessment. Part II investigates a range of approaches to uncertainty analysis in qualitative and quantitative risk assessment, with particular emphasis on their applicability to import risk assessment. Part III explores others issues that are relevant to decisions regarding the choice qualitative or quantitative risk assessment approaches, such as quantitative modeling in the absence of data and the use of historical data in import risk assessment. This is the final report of the second part of the project.

The objectives of Part II of the project are to:

- provide a thorough description of the different types of uncertainty in qualitative and quantitative risk assessment, including a synopsis of the taxonomy and lexicon of uncertainty across disciplines such as the physical and biological sciences;
- review and compare different uncertainty analysis methods, and where possible identify patterns of use within or across disciplines and risk assessment endpoints;
- assess the practicality, assumptions, data requirements, ease of use and software availability of different uncertainty analysis methods, and (where possible) illustrate this assessment with reference to qualitative and quantitative case studies; and,
- identify paradoxes that may arise when using qualitative and quantitative models in the presence of uncertainty. In particular, illustrate the paradoxes that may occur when different uncertainty analysis methods (and their assumptions) are applied to pathway-based import risk assessment

This report is the primary output of Part II of the project. Other outputs from this part of the project include contributions to journal articles (Hosack et al., 2008), two peer reviewed conference papers (Kuhnert and Hayes, 2009; Kuhnert et al., 2009), a workshop on uncertainty analysis techniques presented at the 2009 conference of the Australian and New Zealand Chapter of the Society for Risk Analysis (<http://www.acera.unimelb.edu.au/sra/2009/index.html>), together with analysis and other material contributions to the ACERA “demonstration project” (ACERA project 09/01) (Burgman et al., 2009). Some of the material in these research outputs is collated and further contextualised in this report.

## 1.2 Report structure and outline

Section 1 of the report outlines the project background and summarises the project’s research outputs. It concludes with a discussion of some fundamental issues associated with four different “schools” of uncertainty analysis that measure uncertainty in different ways. It sets the scope for the remainder of the report by introducing the concept of imprecise probability, identifying an important equivalency between two of the four schools, and highlighting how the report addresses the other two schools of thought.

Section 2 discusses risk and uncertainty, and identifies four broad categories of uncertainty: decision uncertainty, linguistic uncertainty, epistemic uncertainty and variability. It identifies the various sources of uncertainty within each of these categories and the often confusing nomenclature associated with each category. This section also highlights the potential for bias and non-commutative results that can occur in qualitative risk assessment but which only become evident when linguistic uncertainty is treated mathematically.

Section 3 examines uncertainty analysis and treatments for uncertainty. It places uncertainty analysis within the broader objectives of prediction, inference and understanding, and clarifies the different approaches to uncertainty analysis adopted by modelers, statisticians and risk analysts. This section identifies two basic modes of uncertainty analysis: a forward propagation mode and an inferential mode. The forward propagation mode is appropriate when there are no, or very few, observations of the process of interest. The inferential mode is appropriate when there are observations of the process. Pathway-based import risk assessments typically lack observations of the assessment endpoint hence uncertainty analysis in this context is usually constrained to the forward propagation mode. Section 3, however, emphasises that observations of the risk-generating process can (and should) be made, enabling this mode of uncertainty analysis to be augmented by a range of powerful inference techniques. Section 3 also outlines a number of basic strategies for treating uncertainty and dependency in quantitative risk assessment, such as factorise it, envelope it or average over it, and introduces methods for treating model structure uncertainty and parametric uncertainty within each of these strategies.

Section 4 addresses uncertainty analysis techniques that are appropriate when the objective is risk prediction in the absence of observations (data) of the risk-generating process. These techniques can be used in the initial stages of a risk management programme - i.e. as part of the risk assessment but before observations of the risk-generating process are made to test its predictions - and they can be applied to pathway-based import risk assessment. Section 4 examines in detail many of the methods introduced in Section 3 but does not address the inferential mode of uncertainty analysis (statistics). Statistics is an enormous discipline and Section 3 highlights a number of excellent textbooks on various aspects of statistical science. Interested readers should seek further guidance in these texts. Section 4 categorises uncertainty propagation techniques into one of four types - analytical, probabilistic, non-probabilistic and graphical - and examines how these can be used to address parametric uncertainty and model structure uncertainty.

Section 5 examines the use of forward uncertainty propagation techniques within the broader risk assessment community. It presents the results of a literature search on the use of these techniques by the risk assessment community, and then evaluates these techniques against a number of criteria that are designed to assess their practicality and utility in a model-based decision-making process.

Section 6 of the report concludes with a discussion of, and recommendations for, uncertainty analysis within qualitative and quantitative import risk assessment. It re-emphasises the role of uncertainty analysis and statistics within risk management programmes and examines paradoxes that can occur in risk assessment if critical issues such as dependency are handled in a naive fashion. The section concludes with a general framework for handling uncertainty in quantitative risk assessment.

Scattered through out the report are text boxes that refer to a hypothetical case study that was completed during the course of the project (Burgman et al., 2009). The objective of these text

boxes is to highlight particular issues associated with qualitative and quantitative approaches to import risk assessment that were addressed during the case study, including *inter alia* the potential bias that can occur with qualitative matrix-based risk assessment methods, and the application of quantitative methods such as Monte Carlo simulation and probability bounds analysis. A complete description of the case study can be found in Burgman et al. (2009).

### 1.3 Fundamental issues

This report is concerned with possibility and probability, and how these concepts are used in risk assessments to characterise, analyse and propagate epistemic uncertainty and variability. O'Hagan and Oakley (2004) claim that probability is an appropriate representation for all forms of uncertainty. Using probability to represent epistemic uncertainty, however, is a slightly controversial issue (Aven, 2010). It is clearly tenable under a Bayesian interpretation of probability, but some of the common practise in Bayesian methods, such as the use of the Uniform distribution to represent epistemic uncertainty, impart more information than is typically warranted by the available evidence (Tucker and Ferson, 2003; Zio, 2008). This facet of precise probability distributions, together with a variety of "puzzles" (Halpern, 2005) has led scholars to consider alternative theories of uncertainty, particularly for that part of uncertainty associated with incomplete knowledge (epistemic uncertainty).

There are at least three other schools or theories, other than probability theory, to represent and propagate uncertainty: evidence theory, possibility theory and plausibility theory (Helton et al., 2004; Halpern, 2005; Dubois, 2010). This report places most of its emphasis on probability theory for three reasons:

- probability theory is by far the most widely used theory of uncertainty, it is well understood, and an enormous amount of predictive and inferential methodology has been designed around it;
- our attention is not restricted to precise probability. The report describes an imprecise probability model (upper and lower probabilities) (Walley, 1991) that mixes interval analysis techniques and probability theory to simultaneously represent epistemic uncertainty and variability; and,
- the belief and plausibility functions of evidence theory, are equivalent to the upper and lower probability measures of imprecise probability (Ferson et al., 2003; Regan et al., 2004; Baudrit et al., 2007a,b) and in certain circumstances (nested random sets) are also equivalent to the necessity and possibility measures of possibility theory. Upper and lower probabilities retain the main advantages of evidence theory and possibility but are easier to interpret and use in practical contexts.

Whilst the focus here is probability theory, the report does address methods that are often employed in risk assessment and were developed or associated with other schools of uncertainty analysis. For example, it briefly examines the use of fuzzy sets and info-gap theory. The former aligns with possibility theory, whilst the later is a form of (or at least similar to) Wald's maximin principle, neither of which involve probability. It is not possible within the resources of this project to provide a comprehensive account of the other schools of uncertainty analysis, and good textbooks exist that already do this (Halpern, 2005). In deciding which methods to cover, and which to exclude, I have been guided by the practise of risk assessment, as evidenced by method citations in the literature, and my own experience.

The report touches upon qualitative descriptions of uncertainty but does not examine non-numeric approaches to uncertainty in any detail. Terms such as “low likelihood” or “medium certainty” confound linguistic uncertainty with variability and epistemic uncertainty. When these terms are undefined they convey a vague, ordinal sense of confidence, but offer no further insight into the sources of uncertainty and how it might affect the outcome of a decision. Linguistic uncertainty can only be separated from epistemic uncertainty and variability when terms such as “low”, “medium” and “high” are quantitatively defined in a manner that is appropriate to the context in hand. Once quantitatively defined, qualitative descriptions of uncertainty may be useful for the purposes of communicating uncertainty, but they are superfluous to its analysis. It is important to note that upper and lower probabilities allow the analyst to address uncertainty about variability in a coherent fashion, removing the necessity of precise distribution functions, and avoiding the need to resort to qualitative descriptions such as “low confidence in a high likelihood”. The distinction between precise and imprecise probability is missed by some authors (see for example Hansson (2004)).

Finally the report restricts its attention to uncertainty in the language, input data, models and parameters of a risk assessment exercise. It does not discuss other “locations” of uncertainty, such as the context of the problem, and only briefly touches on the selection of output variables (Walker et al., 2003). It notes but does not address in detail “decision uncertainty”, that is the uncertainty associated with the values, intentions and behaviour of human beings (Ascough et al., 2009). These are legitimate sources of uncertainty but they are not within the scope of this project. Furthermore, in import risk assessment some of these sources of uncertainty are minimised by international and national regulation and guidelines.

## 1.4 Notation

Inconsistent notation, both within and between references, is a potential source of confusion. I have tried to maintain a consistent notation in this report but this often requires departing from the notation used in the literature that it is cited here. Random variables are denoted with capital Roman letters, such as  $X$  or  $Y$ . Lower case letters such as  $x, y$  denote realisations of random variables. Bold face is used to denote vectors, and non-bold face denotes scalars. Functions are generically denoted as  $f(\cdot)$ ,  $h(\cdot)$  or  $g(\cdot)$ . The use of  $p(\cdot)$  is reserved for the probability density function, and  $F(\cdot)$  denotes the cumulative distribution function.  $E[X]$  denotes the expectation of a random variable and  $E[h(X)]$  the expectation of a function of random variable. Roman capital  $P$  denotes the probability of an event.

### BOX 1: HYPOTHETICAL CASE-STUDY

The hypothetical case-study mirrors a real pathway-based, import risk assessment. The risk assessment calculates the probability that a pest insect will be introduced, establish and spread in Western Australia following a single year of fruit imports from orchards in a foreign country. In this case study, the risk assessment model details nine steps starting with prevalence of the pest insect in the exporting nation's orchards, and culminating with their establishment and spread in Australia (Figure B1.1). The model is consistent with international and Australian standards for import risk assessment.

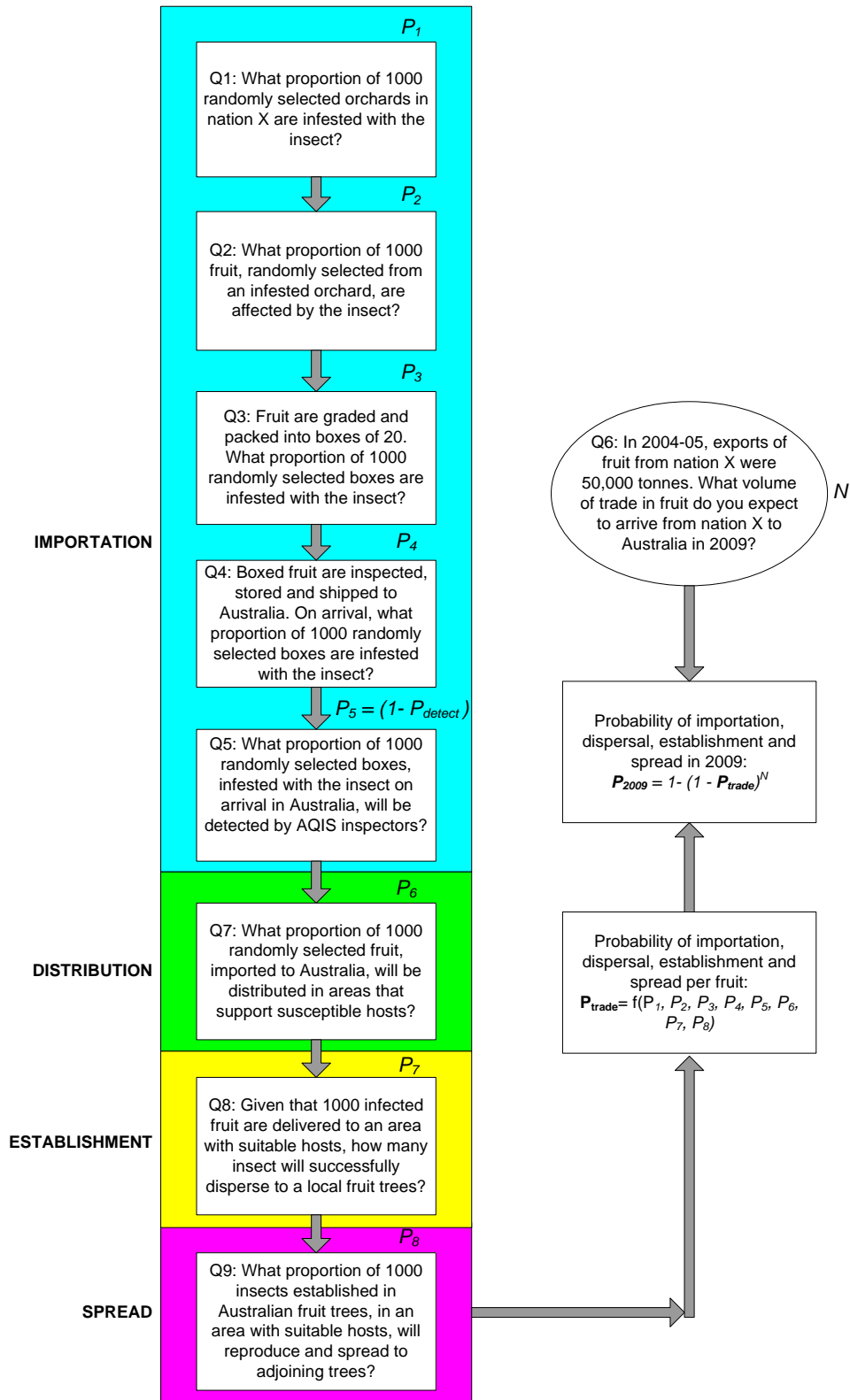
The case study's assessment endpoint is the establishment and spread of a non-native insect in Australia but it does not address the consequences of this event. The insect, however, is a known to be pest species that would reduce the value of Australian fruit production if it were to become established in Australian orchards. The probability of the assessment endpoint following trade with the exporting nation ( $P_{trade}$ ) is a function of the nine steps in the model (Figure 1.1). Typically the probability of each step is assumed to be independent of the others, leading to a simple function that multiplies the probabilities at each step. This report examines the implications of this assumption, and the effects of alternative, often more plausible, assumptions. It is important to note that this report does not address the issue of unit mixing that occurs in real commodity distribution networks, but is not captured in this risk assessment model. This issue is addressed in separate dedicated study completed under Part I of the project.

The case study also examines issues that may arise with Biosecurity Australia's risk-assessment framework and the language-based descriptions of probability used in this framework (Table B1). The case study provides a platform to compare the consistency of risk estimates based on a range of different predictive uncertainty analysis methods and model assumptions.

Likelihood	Frequency	Probability
High	The event would be very likely to occur	$0.7 < P \leq 1$
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	$0 < P \leq 0.000001$

**Table B1: Language based descriptions and numerical definitions of probabilities used in the hypothetical case-study**

Intervals for each of the probabilities  $P_1$  to  $P_8$ , and volume of trade  $N$ , in the hypothetical risk-model were elicited from seven experts using a structured four-step elicitation procedure (Spiers-Bridge et al., 2010). Beta and Normal distributions were fitted to these intervals, depending on whether the quantity in questions was a proportion or a real-valued variate, and then pooled using a linear pooling method with equal weights assigned to each expert (Stone, 1961; Clemen and Winkler, 1999; O'Hagan et al., 2006). The elicited intervals, fitted distributions and pooled distributions for all nine steps in the risk assessment model, together with a complete description of the analysis methods, are fully described in Appendix A and Burgman et al. (2009).



**Figure 1.1:** The risk assessment model for the hypothetical case-study, highlighting the nine step model, the questions for each of the steps that were presented to an expert group, and the four steps (importation, distribution, establishment and spread) in Biosecurity Australia's import risk assessment procedure that each question addresses

## 2 Risk and uncertainty

### 2.1 What is risk?

The concepts of risk and uncertainty are intimately linked. Risk occurs because the past, present and future are uncertain. Willett (1901) defined risk as the “objectified uncertainty regarding the occurrence of an undesirable event”. Knight (1921) defined risk as “measurable uncertainty”, reserving the term uncertainty for those things that could not be measured. More recent definitions emphasise that risk is concerned with the chance of undesired events, usually within a specified time frame, and almost universally identify probability as the key metric in this context (Table 1).

Mathematical definitions of risk vary across disciplines (Table 2). The definition most relevant to import risk assessment describes risk as a function of  $n$  variables or “risk factors” ( $x_i$ )

$$Risk = h(x_1, x_2, \dots, x_n). \quad (2.1)$$

Haimes (2009) suggests that the risk factors in Equation 2.1 can include time, the probability of an undesired event, the probability of consequences given the event, the magnitude of these consequences and the state of the system in question, such as its performance capability, “vulnerability” and “resilience” to undesired events. In other words the factors that contribute to risk can be diverse and complex. Uncertainty enters this equation because the risk factors and/or the risk function vary in time or space, or because they are not completely known (or both). If the risk factors are treated as random variables to reflect uncertainty, and this uncertainty is measured with probability, then risk is a function of the joint distribution of the risk factors (Cox et al., 2005). This approach to risk and uncertainty is advantageous because it enables mathematically coherent strategies that can represent and propagate uncertainty and deal with the important question of dependency between risk factors (Section 3.3).

Import risk assessment assigns probabilities to the chances of pests and diseases successfully negotiating the hurdles between introduction, establishment and spread. Quantitative import risk assessment functions therefore resemble the general risk Equation 2.1 wherein risk is a function of  $n$  variables - the probability of successfully negotiating the hurdles to establishment and spread. In qualitative import risk assessment the calculation of risk is internalised by those conducting the assessment. In quantitative import risk assessment the calculation is performed via an explicit model and formula (such as Figure 1.1).

The chief advantages of quantitative risk assessment are: a) it helps minimise one source of uncertainty (language); and, b) it enables two other sources (epistemic uncertainty and variability) to be analysed in a manner that exposes conclusions that are not consistent with the rules of conditional probability. Burgman (2005) discusses the results of experiments that show how the agreement between experts’ estimates of risk increases if they are able to compare their interpretation of terms such as “high”, “medium” and “low” and the context in which they are used. Moreover, Moskowitz and Sarin (1983) demonstrate that assessors routinely violate the basic rules of probability calculus when assessing the likelihood of interdependent events, unless their conditional beliefs are elicited and presented in the form of a conditional probability table. The effects of linguistic uncertainty and the presence of calculus errors are very difficult to expose with qualitative descriptions of probability and internalised risk calculations. By exposing these issues, quantitative risk assessment encourages two important science-quality criteria: transparency in the assessment and repeatability between assessments and assessors.



Definition	Reference
The objectified uncertainty regarding the occurrence of an undesirable event	Willett (1901)
Measurable uncertainty	Knight (1921)
A measure of the probability and severity of adverse effects	Lowrance (1976)
The probability that a consequence will occur	Rasmussen (1981)
The probability that a particular adverse event occurs during a stated period of time, or results from a particular challenge	The Royal Society (1983)
The probability of harm	Wachbroit (1991)
The likelihood of an undesired event occurring as result of some behaviour or action (including no action)	Hayes (1997)
The magnitude of an adverse event multiplied by the likelihood of its occurrence	Mullin and Bertrand (1998)
The combination of the magnitude of an adverse event and the probability of its occurrence	Environmental Risk Management Authority (1999)
The probability of future loss	Byrd and Cothern (2000)
The combination of the probability of an event and its consequences	ISO/IEC (2002)
The chance that something bad will happen	Brillinger (2002)
The probability of occurrence of an undesired event	van Straalen (2002)
The chance, within a time frame, of an adverse event with specific consequences	Burgman (2005)
The chance, within a prescribed time frame, of an adverse event with specific (usually negative) consequences	Fox and Burgman (2008)
The probability of an unwanted outcome or consequence occurring	Wooldridge (2008)
The effect of uncertainty on objectives	ISO (2009)

**Table 2.1:** Examples of different definitions of risk in the literature

Function	Arguments	Discipline	Reference
$R = h(P_c, C(v))$	Risk ( $R$ ), probability of consequence ( $P_c$ ), value of the consequence to the risk taker ( $C(v)$ )	Engineering	Rowe (1977)
$Y = h(X)$	An unknown observable quantity ( $Y$ ), a vector of unknown observable quantities on a more detailed level ( $X$ )	Engineering	Apeland et al. (2002)
$R = h(x_1, x_2, \dots, x_n)$	Risk ( $R$ ), risk factors ( $x_i$ )	Ecology	Nayak and Kundu (2001)
$R = [s_i, l_i, x_i]_c$	Risk ( $R$ ), scenarios ( $s_i$ ), likelihood ( $l$ ), consequences ( $x_i$ ), complete ( $c$ )	Engineering	Kaplan (1997)
$R = ((L_1, O_1) \dots (L_n, O_n))$	Risk ( $R$ ), likelihood ( $l$ ), outcome( $O$ ), number of possible outcomes ( $n$ )	Engineering	Kumamoto and Henley (1996)
$HQ = \frac{C \cdot I}{EM \cdot RfD}$	Hazard quotient ( $HQ$ ), concentration ( $C$ ), intake rate ( $I$ ), body mass ( $BM$ ), unit-less reference dose ( $RfD$ )	Toxicology	Hammonds et al. (1994)
$HQ = \frac{PEC}{PNEC}$	Hazard quotient ( $HQ$ ), predicted environmental concentration ( $PEC$ ), predicted no effect concentration ( $PNEC$ )	Toxicology	Calow and Forbes (2003)
$\frac{h(R, EEC)}{p(R EEC) p(EEC)} =$	Risk (number of genera affected) ( $R$ ), expected environmental concentration ( $EEC$ )	Toxicology	Warren-Hicks et al. (2002)
$R = -E[\mu(X - \bar{X})]$	Risk ( $R$ ), expected utility ( $E[u(\cdot)]$ ), random variable (lottery) ( $X$ ), the mean of the lottery ( $\bar{X}$ )	Decision theory	Jia et al. (1999)
$\frac{dP}{dt} = h(r, P) - C$	Amount of a valued property ( $P$ ), time ( $t$ ), growth rate of the property ( $r$ ), removal due to anthropogenic factors( $C$ )	Fisheries	Smith et al. (2007)
$R = h(V, U)$	Risk ( $R$ ), uncertainty ( $U$ ), variability ( $V$ )	Biosecurity	Murray (2002)
$P(I) = 1 - (1 - C_1 \cdot C_2)^N$	Probability of agent entry ( $P(I)$ ), country factor ( $C_1$ ), commodity factor ( $C_2$ ), number of animal import units ( $N$ )	Biosecurity	OIE (1996)
$P(I) = 1 - (1 - p)^v$	Probability of incursion ( $P(I)$ ), probability of incursion when importing one unit of volume of product ( $p$ ), total volume of imported products ( $v$ )	Biosecurity	AGDAFF (2001)

**Table 2.2:** Examples of different risk definitions functions associated with different disciplines

## 2.2 Why and how is risk assessed?

We perform risk assessment to help protect human values, as reflected in the assessment end-points, from the processes and events that threaten them. The primary aim of the assessment is to separate the wheat from the chaff - i.e. separate the (high) risk events that we should worry about from the (low) risk events that we should not. Hence, risk assessment is a decision aid that aims to rank or quantify risks to human values in order to prioritise management actions and the allocation of resources. Science-quality criteria require the assessment to be transparent, repeatable and systematic, and its predictions to be theoretically falsifiable. Decision makers would also prefer risk predictions to be precise and accurate. The extent to which a risk assessment can reliably, transparently and accurately separate high risk events from low risk events is an important test of its utility in a decision-making framework. The extent to which the assessment makes predictions that can be measured and falsified is an important test of its scientific credibility. It may be difficult to falsify a stochastic prediction<sup>1</sup>, but it must, in principle, be possible for the prediction-making process to be considered scientific.

Science measures things on four different scales: nominal, ordinal, interval or ratio<sup>2</sup>. A nominal scale classifies or identifies variables without implying any order. For example, categorising insect pests according to those that have wings and those that do not. Nominal variables are qualitative, categories that differ in quality not quantity, and cannot therefore be summed, subtracted, multiplied or divided. Nominal scales express equality and have a restricted set of permissible statistics (Stevens, 1946): categories can be counted and the category with the greatest number of items (the mode) can be identified. The variability of nominal variables can also be measured with a range of simple indices (Wilcox, 1967). Nominal scales, however, do not play an important role in risk assessment because they lack a sense of order - i.e. they cannot separate high risk events from low risk events.

An ordinal scale, such as high, medium or low, measures the rank order of variables but provides no information on the distance between variables. Permissible statistics include those of nominal variables together with percentiles and the median (Stevens, 1946). The position of ordinal variables in the quantitative-qualitative classification is fuzzy (Agresti, 2002). They can express equality (“high” = “high”), they also indicate whether one variable is greater or less than another (“high” > “medium” > “low”) but they cannot be summed, subtracted, multiplied, etc. in a meaningful manner. Ordinal scales are very important in qualitative risk assessment because they produce ordinal risk metrics.

Ordinal scales give the sense of an underlying continuous variable, and it is common practice in qualitative risk assessment to add, multiply or otherwise combine ordinal variables to produce intervening variables or overall risk estimates. The result of these operations, however, can be arbitrary and difficult to interpret. Wooldridge (2008), for example, suggests that “high” × “high” should result in something smaller than “high” if the underlying continuous variable is a probability. Cox et al. (2005), however, point to examples where a “high” probability of contaminated food, and a “high” probability of not cooking it properly, results in a “high” probability of exposure to an adverse health outcome. Neither of these propositions, however, can be falsifiable until the terms “high”, “medium” and “low” are numerically defined in a manner that is appropriate to the context of the problem. Hence, undefined qualitative risk

---

<sup>1</sup>Ferson (1996a) notes that validating a probabilistic prediction can be difficult, particularly for rare events, and offers a range of additional quality assurance techniques.

<sup>2</sup>But see Velleman and Wilkinson (1993) for a critique of this classification

terms do not meet one of the criteria that define good-quality science, namely the ability to measure and falsify (at least theoretically) the predictions of the risk assessment. Furthermore, the potential bias and non-commutative conclusions of qualitative risk assessment schemes that involve<sup>3</sup> ordinal variables cannot be exposed unless these terms are defined (Box 2). Again, quantitative definitions of qualitative risk terms help maintain another important requirement of good-quality science: transparency in the risk assessment process.

Interval and ratio scales are quantitative scales. Interval scales measure the distance between two values in a way that is invariant to their location on the scale. For example the difference between “9” and “10” is the same as that between “1” and “2”. Zero on the interval scale is a matter of convention or convenience and does not mean the absence of the quantity being measured. Ratio scales are the same as interval scales except “0” on a ratio does mean the absence of the quantity being measured, hence ratios are meaningful. For example “4” on a ratio scale means twice as much of the quantity being measured as “2”, but on an interval scale it does not because the “0” of an interval scale is arbitrary. Stevens (1946) argues that permissible statistics of interval scale data include all those of nominal and ordinal data, and the mean, standard deviation, rank order correlation and product moment correlation, but not the coefficient of variation, whereas all types of statistics are permissible with ratio scale data.

Ratio-scale data are very important to quantitative risk assessment because it produces ratio-scale estimates where “0” means the absence of risk. Cox (2008) suggests one way to check the quality of a qualitative risk assessment is to ensure that the rank order of its estimates do not changes when its operations are converted to a ratio scale. These checks require that the ordinal metrics of qualitative risk assessment, such as “low likelihood” are converted to a ratio scale. Converting ordinal risk metrics in this fashion helps to improve the quality of qualitative risk assessment by exposing potential bias (Box 2). Accurate rank-order risk estimates, however, are not the same as accurate absolute risk estimates. Converting qualitative risk metrics to check for a consistent rank-order does not guarantee accurate risk estimates because the arithmetic operations in a qualitative risk assessment almost always include unsupported assumptions, most commonly that the variables in a risk assessment model are independent. Violations of this assumption can easily lead to inaccurate assessments, particularly for tail risks (Section 3.3).

Burgman (2005) argues that good quality risk assessments are “complete” and “honest”. Complete risk assessments are defined as those that undertake all stages of the risk assessment cycle including monitoring and validation of predictions. Honest risk assessments are those that are faithful to the assumptions about the kinds of uncertainty embedded in the assessment, carry these uncertainties through the analysis, and represent and communicate them reliably and transparently. Two of the three science-quality criteria identified above - transparency and falsifiability - feature prominently in Burgman’s definition of a complete and honest risk assessment. These criteria, together with a full account of the sources of uncertainty in the assessment, and its effects on risk-based decisions, are the primary scientific motivation for quantitative risk assessment. The minimum requirements for qualitative risk assessment to meet these criteria are: a) the assumptions and calculation steps are clearly documented; b) all sources of uncertainty are highlighted and their potential effect on the analysis identified; and c), ordinal risk estimates are defined on a ratio scale.

---

<sup>3</sup>Convolution is the mathematical operation which finds the distribution of the sum of independent random variables from the distribution of its addends. The term is generalised here to include operations such as difference, product and quotients, as well operands other than distributions (e.g. intervals, probability boxes).

## BOX 2: BIAS AND NON-ASSOCIATIVITY IN QUALITATIVE RISK ASSESSMENT

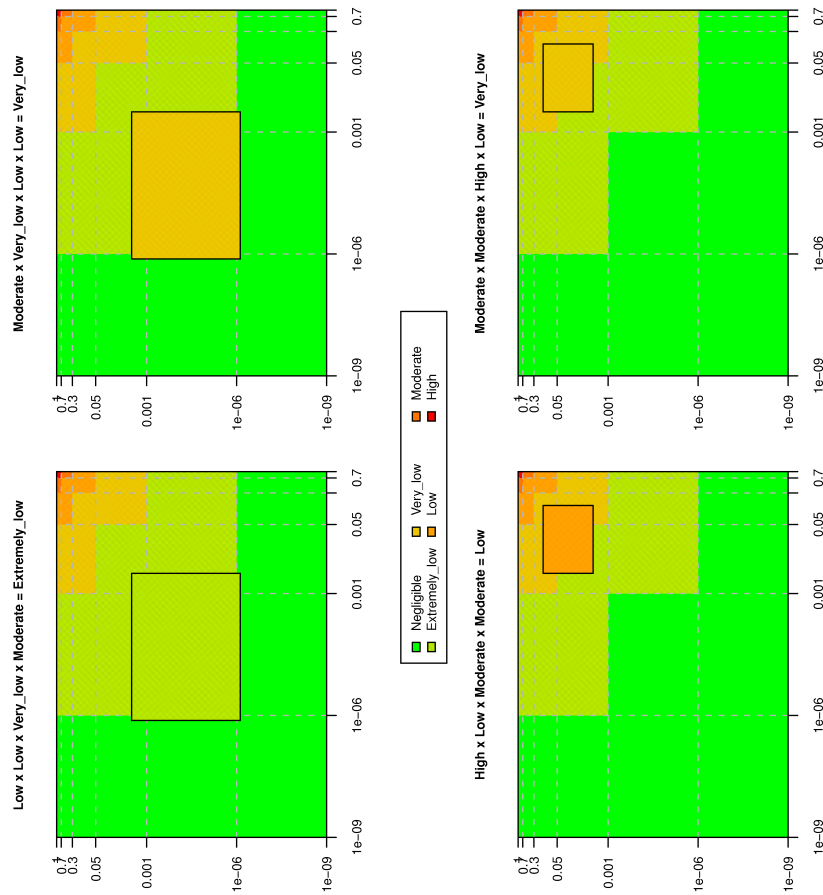
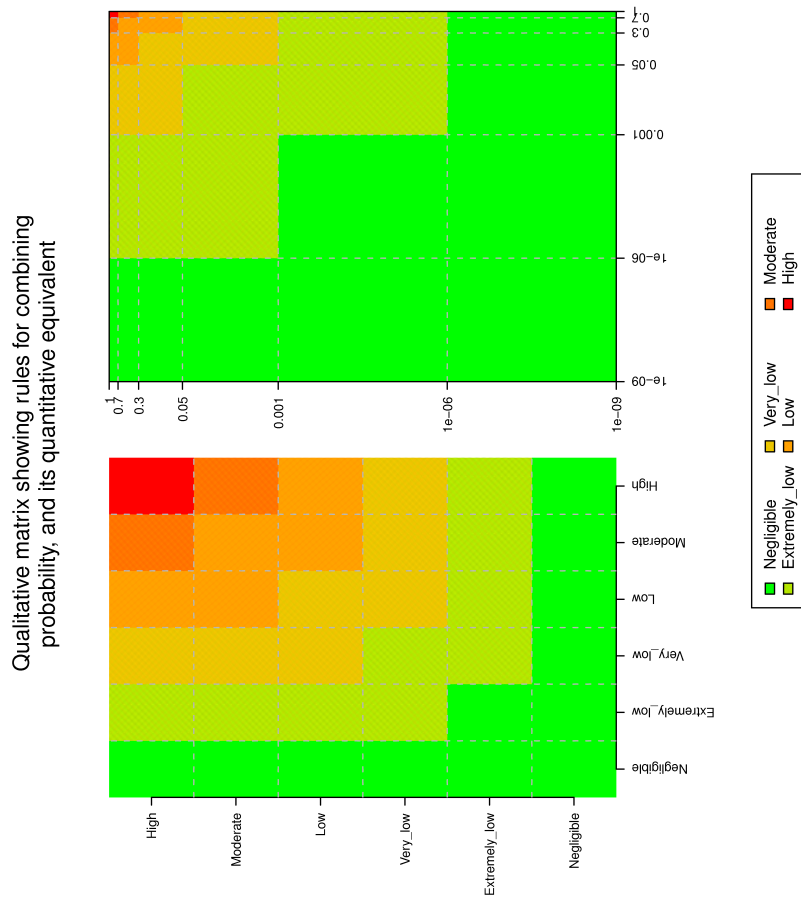
Australia's qualitative import risk assessment schema calculates the probability of import, establishment and spread using qualitative convolution rules illustrated in the matrix on the left hand side of Figure 2.1. The matrix determines the outcome of the product of (for example) a "moderate" probability of entry, with a "low" probability of distribution, resulting in a "low" probability of entry. The matrix on the right hand side of Figure 2.1 shows the same combination rules but with ratio-scale definitions of the terms "high", "moderate" and "low" etc. (Table B1) substituted. This matrix is shown on a log-scale for the purposes of clarity only.

Substituting qualitative terms with their numerical definitions allows the analyst to test for bias and non-commutative results that can arise because the construction of the convolution matrix is arbitrary. McCarthy et al. (2007) highlighted the potential for non-associative results with language based convolutions, noting that in the Australian schema ("low" x "low") x "very low" gives an "extremely low" result. Whereas "low" x ("low" x "very low") gives a "very low" result. Hence the result of the convolution is sensitive to the way in which the product operation is bracketed.

The matrices in Figure 2.2 illustrate an effect similar to that identified by McCarthy et al. (2007). Here the matrices show the result of the product of the four steps (importation, distribution, establishment and spread) in Australia's import risk assessment procedure. The colour of the square box within the matrix shows the result of the convolution identified in the title of each matrix. The top left matrix shows the "extremely low" result that occurs when the qualitative convolution matrix is used to calculate the product "low" x "low" x "very low" x "moderate". The top right matrix shows the (non-commutative) result "very low" when the order of this product is changed. The matrices in the bottom of Figure 2.2 show the same effect for a different set of probabilities.

It is important to recognise that qualitative convolutions schemes can also lead to bias. The quantitative definitions of the terms "high", "moderate" and "low" adopted in this context permit the application of interval analysis (Moore, 1966). The location of the coloured boxes in Figure 2.2 show the interval that results from the convolution in the title of each of the matrices. The colour of the box corresponds to the result of the qualitative convolution. The difference between the colour of the box and the colours in the matrix that it spans illustrates the bias. The greater the colour difference the greater the bias. For example, the boxes plotted on the matrices in the bottom row show that the product "high" x "low" x "moderate" x "moderate" results in the interval [0.0032, 0.147] irrespective of the order or bracketing of the operation. This interval spans the categories "extremely low" and "low" because its lower bound is > 0.001 and its upper bound is < 0.3. The qualitative convolution, however, returns the result "low" or "very low" (depending on the order of the operation) suggesting that the interval spans [0.05, 0.3] or [0.001, 0.05]. The difference between these intervals is bias.

The bias, non-associative and non-commutative effects demonstrated here are peculiar, but not limited, to Biosecurity Australia's qualitative matrix combination rules - a similar effect can occur in any qualitative risk matrix. The severity of the effect is determined by *inter alia* the number of operations (four in this case) and the particular structure of the matrix. These considerations are arbitrary and at the analyst's discretion. Current guidelines for constructing qualitative risk matrices (Cox, 2008) do not address this issue, and it is probably best avoided by simply replacing qualitative combination rules with interval arithmetic (Section 4.4.2).



## 2.3 What is uncertainty?

Uncertainty is a term used to encompass many concepts (Morgan and Henrion, 1990). It has been defined as a degree of ignorance (Beven, 2009), a state of incomplete knowledge (Cullen and Frey, 1999), insufficient information (Murray, 2002), or a departure from the unattainable state of complete determinism (Walker et al., 2003). In truth, the scientific literature contains many definitions, descriptions and typologies of uncertainty. This picture is complicated further by different lexicons that use different names for the same thing, and occasionally the same name for different things.

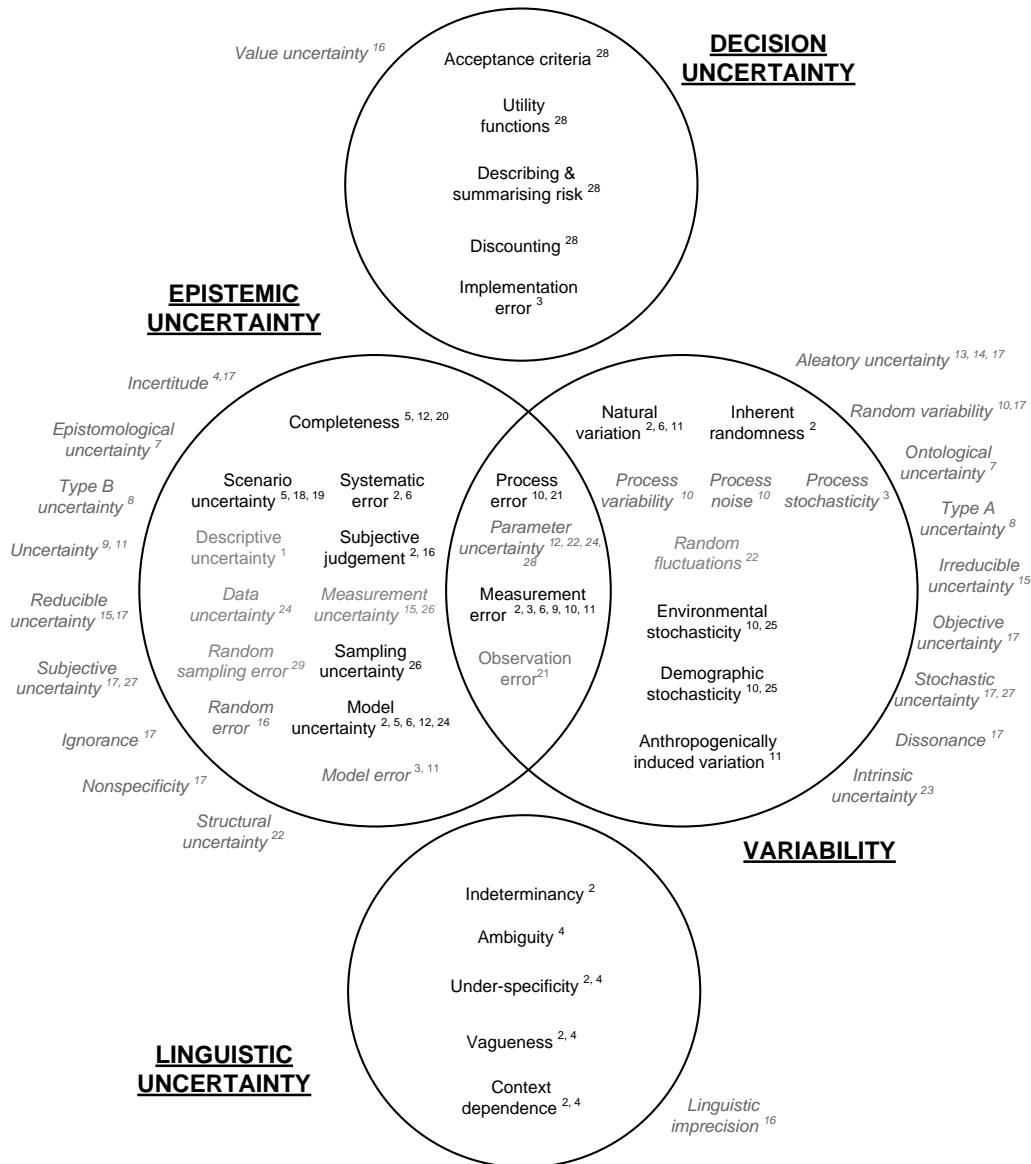
Regan et al. (2002a, 2003) identify two main sources of uncertainty: epistemic uncertainty and linguistic uncertainty. Epistemic uncertainty is the uncertainty associated with knowledge. It is the uncertainty created by imperfect knowledge about something that is in principle knowable, and therefore in principle reducible with additional research and observation. It is synonymous with terms such as incertitude (Carey and Burgman, 2008) and epistemological uncertainty (Gillund et al., 2008), and is sometimes simply referred to as "uncertainty" (Frey and Burmaster, 1999; McCann et al., 2006). Statisticians and modelers often refer to it as model uncertainty or model error. Other authors, however, emphasise that model uncertainty represents only one of many sub-categories of epistemic uncertainty (Regan et al., 2002a, 2003; Ryan, 2008) (Figure 2.3).

Many authors, including myself, separate variability from epistemic uncertainty, and identify it as a third main source of uncertainty (O'Hagan et al., 2006; Hayes et al., 2007a; Carey and Burgman, 2008; Ascough et al., 2009). Variability is the uncertainty associated with diversity or heterogeneity. It is the natural or anthropogenically induced variation in a population over space and time (Anderson and Hattis, 1999; McCann et al., 2006). Its key distinguishing feature is that it cannot be eliminated with additional research or observation. In some circumstances it can be reduced (in the sense of a lower variance) but it cannot be eliminated - it is an inescapable reality of the real world. It is described in the literature as irreducible uncertainty (Tucker and Ferson, 2003), random variability (Bolker, 2008), ontological uncertainty (Gillund et al., 2008), or more cryptically still as "Type A" uncertainty (Hammonds et al., 1994).

Finkel (1990) identifies a fourth source of uncertainty - decision uncertainty - that enters policy analysis after risks have been estimated. Finkel defines decision uncertainty as the uncertainty that arises where there is ambiguity or controversy about how to quantify or compare social objectives. It is synonymous with Morgan and Henrion's (1990) "value" uncertainty. Finkel goes on to identify various examples of decision-making uncertainty including measures used to describe and summarise risk, defining acceptance criteria, choosing utility functions and ways to aggregate individual utilities, and the (discounting) problem that occurs when comparing immediate consequences with delayed ones.

Another source of decision uncertainty that arises in managed systems, after the risk estimation process, is implementation error, defined by Harwood and Stokes (2002) as the uncertainty associated with incorrect, irrational or illegal behaviour of human beings that leads to imperfect policy implementation. The uncertainty surrounding preferences, values and behaviour is not reducible, and not solely, or even most appropriately, in the domain of scientific inquiry. Decision uncertainty is therefore outside the scope of this report. It is, however, important and can influence how risk assessment results are interpreted and communicated (Ascough et al., 2009), and should be recognised as a separate source of uncertainty.

**Figure 2.3:** Venn diagram illustrating four broad categories of uncertainty, (epistemic uncertainty, variability linguistic uncertainty and decision uncertainty), together with more specific types of uncertainty identified by various authors within these broad categories (black font), and the different, and sometimes confusing, nomenclature used to describe these different types of uncertainty (grey italic font)



References: 1. Rowe (1977); 2. Regan et al. (2002); 3. Harwood and Stokes (2002); 4. Carey and Burgman (2008); 5. Ryan (2008); 6. Regan et al. (2003); 7. Gillund et al. (2008); 8. Hammonds et al. (1994); 9. Frey and Burmaster (1999); 10. Bolker (2008); 11. McCann et al. (2006); 12. Vesely and Rasmuson (1984); 13. O'Hagan et al. (2006); 14. Pate-Cornell (1996); 15. Tucker and Ferson (2003); 16. Morgan and Henrion (1990); 17. Karanki et al. (2009); 18. Draper et al. (2000); 19. Hayes et al. (2008); 20. Baybutt (1989); 21. Clark (2007); 22. Charles (1998); 23. OGTR (2005); 24. Cressie et al., (2007); 25. Morris and Doak (2002); 26. Ferson et al., (2003); 27. Helton (1997); 28. Finkel (1990); 29. Frey and Rhodes (1998)



On the whole, the various sources and categories of uncertainty identified in the literature can be classified into one of these four categories: epistemic uncertainty, variability, linguistic uncertainty and decision uncertainty. It is important to note, however, that the distinction between these different categories is not always clear. The term process error for example is often used to mean all sources of variability in a data set that cannot be explained by the variables of a statistical or process-based model (Clark, 2007). Process error is therefore a mixture of the uncertainty generated by the simplified reality of the model, that is potentially reducible (via a better model), and the irreducible variability of the response variable.

Bolker (2008) categorises process error as variability that affects the future dynamics of the ecological system, as distinguished from measurement error that does not affect the system dynamics. This distinction is important because error in process model creates prediction errors that can increase with time, whereas measurement error does not unless the measurements are subject to systematic bias such as drift<sup>4</sup>. Modern (hierarchical) statistical methods can separate these two sources of variation and provide estimates of process error that are not confounded with measurement (observation) error (Section 3.3).

Ryan (2008) suggests that measurement error could fall under the category of random sampling error. Other authors, however, consider measurement error as reducible and therefore a type of epistemic uncertainty (Harwood and Stokes, 2002; Tucker and Ferson, 2003; Ferson et al., 2003). Measurement error is also synonymous with Rowe's (1977) definition of measurement uncertainty (inability to assign values to variables in a system). Hence the terms process error and measurement error are difficult to neatly categorise. The term parameter uncertainty is also difficult to categories because it is used by some authors to describe *inter alia* measurement error and random error (Finkel, 1990).

In short there is no single classification that neatly captures the different categorisations given to the various sources of uncertainty that one encounters in the literature. Existing classifications and their sometimes confusing nomenclature, reflect differences between disciplines driven by different domains, different objectives and the availability of data. The nomenclature and classification summarised in Figure 1 captures many, but not all, of the idiosyncrasies in the literature. In any particular application, however, the best that one can hope for are carefully defined terms, and context specific descriptions, of the types and sources of uncertainty associated with the problem in hand.

### **2.3.1 Linguistic uncertainty**

Linguistic uncertainty is an important but often overlooked source of uncertainty in risk assessment. Linguistic uncertainty is important because it is pervasive in the workshops, committees and other face-to-face, language-based, methods that qualitative risk assessment relies on to assess and communicate risk (Carey and Burgman, 2008). It is overlooked because in many taxonomies of uncertainty (Beck, 1987; Smith and Shugart, 1994; Pate-Cornell, 1996; Charles, 1998; Walker et al., 2003), risk assessment standards (Food and Agriculture Organisation, 1996; Australia, 2004) and other risk-guidance publications (Barnthouse et al., 1986; USEPA, 1992; Murray, 2002) there is no mention of this source of uncertainty.

---

<sup>4</sup>Drift is the continuous degradation in the accuracy of measurements with time

Linguistic uncertainty arises because language, and our use of it, is not precise. Regan et al. (2002a) identify five types of linguistic uncertainty:

- vagueness occurs because words allow border-line cases. The term “endangered” for example is vague because some species are neither endangered nor not endangered - there are borderline cases (Regan et al., 2002a). The terms used to distinguish likelihood and consequences categories in qualitative risk assessment are typically vague - words such as low, negligible, moderate, or frequent, large, etc. all permit borderline cases;
- context dependence is the uncertainty created by failing to fully specify the context in which a proposition is to be understood. For example describing an oil spill as “small” creates uncertainty - small for what? An oil container, a dinghy, a port, an ocean? (Burgman, 2005). It is important to note that the term small is also vague and that this source of uncertainty will remain even if the context is fully specified;
- ambiguity is uncertainty associated with the meaning of words. Words can have more than one meaning and it is often not clear what the meaning is. A good example from the bio-invasion literature is the term “invasive”. This word is variously used by different authors to refer to species that are considered to be weeds, species that are spreading, species that are harmful, and so forth. Inconsistent use of the same word can confound attempts to compare the results of apparently similar studies (Hayes and Barry, 2008);
- indeterminacy is a subtle and insidious form of linguistic uncertainty. It arises because the future use of a term may not be completely fixed by its current use. Pertinent examples (for import risk assessment) of this type of linguistic uncertainty occur through taxonomic revisions, or the discovery of ecomorphs, strains or other practically important differences between groups of organisms within the same species (see for example Hayes et al. (2009)). Prior to a taxonomic revision or the discovery of sub-species groups, analysts are unaware of the ambiguity associated with the species name;
- under-specificity creates uncertainty through unwanted generality. Unwanted generality can occur in many ways, leading to many ways to interpret (for example) the statement, “there is a 70% chance of rain” - will it rain for 70% of the day, will it rain over 70% of an area, or do you mean a 70% chance of at least some rain, at a particular point? (Burgman, 2005). Unwanted generality can also occur through imprecise descriptions of locations (e.g. inland Australia), processes (e.g. fishing), stressors (e.g. debris), etc. (Carey and Burgman, 2008; Regan et al., 2002a).

Linguistic uncertainty may be deliberate or inadvertent. People may use vague terms deliberately to avoid giving an impression of precision, or they may use them because they are poor communicators. In either case vagueness and the other sources of linguistic uncertainty create problems for risk assessment, particularly when analysts attempt to quantify certainty with terms such as “highly certain” or “medium confidence”. O’Hagan et al. (2006) and Morgan and Henrion (1990) emphasise that verbal expressions of uncertainty mean different things to different people, and sometimes mean different things to the same person in different contexts. Verbal expressions of certainty do not therefore provide a consistent (e.g. between assessments or even between assessors) basis for uncertainty analysis. It is interesting to note that qualitative descriptions of uncertainty figure prominently in the reports produced by Working Group III (WG III) of the Intergovernmental Panel on Climate Change (IPCC) when describing climate-change impacts. WG III describe these terms as self explanatory, which is a testament to how often the importance of linguistic uncertainty is underestimated in scientific endeavors.

### 2.3.2 Variability

Variability is caused by fluctuations or differences in a quantity or process. It is manifest as a multiplicity of “true values” (Finkel, 1990), or as the diversity among members of a population (Frey and Rhodes, 1998). The categorisation developed here (Figure 2.3) recognises the following sources of variability:

- natural variation is the variability that occurs naturally in populations across time and space. Some authors distinguish two sub-categories of temporal variability: environmental stochasticity and demographic stochasticity. Environmental stochasticity is the temporal variation in the mean vital rates (birth, death, growth, etc.) of organisms, caused by the abiotic and biotic forces that effect a population. Demographic stochasticity is temporal variation in population growth rates driven by chance variation in the actual fate of different individuals. Morris and Doak (2002) suggests that this is essentially the same as the inherent randomness that causes variation in the number of heads and tails when repeatedly flipping a coin (see below);
- inherent randomness describes repeated processes that have no discernible deterministic pattern and are unpredictable in detail. Regan et al. (2002a) argue that genuine examples of inherently random processes are difficult to find, and that classic random processes (such as tossing a coin) are only unpredictable because we do not have enough information about the dynamics of the process and its initial conditions to reliably predict its outcome. Even with this information, however, the outcomes of a coin toss would still be patternless and therefore inherently random (Franklin, 2009). The extent to which complex natural systems, such as ecosystems and weather systems, are inherently random is a mute point. For practical purposes the variation associated with these systems cannot be reduced or described by deterministic cause and effect relationships because we are unlikely to ever hold enough information to completely characterise their dynamics, and very slight differences in their initial conditions can quickly propagate to large differences in later conditions.

As noted earlier, variability cannot be treated in the same sense as epistemic uncertainty or linguistic uncertainty - i.e. it cannot be minimised. It can, however, be characterised and propagated through a risk assessment using a variety of techniques that are discussed here. An important practical consideration in this context, however, is the concept of a population. Variability is the irreducible diversity or heterogeneity of a population. A clear definition and understanding of variability therefore requires an unambiguous description of the relevant population (Anderson and Hattis, 1999). In practise this may not be a simple task, particularly where the risk bearing units are separated or aggregated in, for example, a distribution network.

Furthermore, data used to capture and represent variability in a risk assessment may reflect only a sub-set of the population of interest, or conversely may reflect a much larger group that are exposed or otherwise predisposed to hazards to different degrees. Care must therefore be taken when characterising the variability of a population to ensure that the population is appropriate to the issue in hand. Thompson (1999) provides general guidance in this regard. Nauta (2000) provides an example within the context of propagating variability and uncertainty through a quantitative microbial risk assessment. Modern statistical techniques, such as random effects models (McCulloch and Searle, 2001), provide a means to test for the presence of sub-groups within a data set that may indicate important sub-population level processes.

### 2.3.3 Epistemic uncertainty

Epistemic uncertainty stems from a lack of data, understanding and knowledge about the world. In a risk assessment there are many ways in which our knowledge may be incomplete. The categorisation developed here (Figure 2.3) recognises the following:

- model uncertainty is the uncertainty in our conception or description of a system. Hence, Rowe's (1977) classification of "descriptive uncertainty". All risk assessments are predicated on a conceptual model of the system in question. The extent to which this model is transparent varies between assessments. Quantitative assessments are usually explicit, whereas qualitative assessment often leave the assessor(s) conceptual model unspecified. Transparent risk assessment models can range in quality from simple cartoons (Port of Melbourne Corporation, 2006) to mathematical descriptions of ecosystem processes using difference or differential equations (Kot, 2001). They can be idealised representations of a statistical process that generated a data set (e.g. a Poisson process for counts), or more complex hierarchical representations of both the ecological process and observational process that generates data (Clark and Bjornstad, 2004). Uncertainty occurs in the choice of variables that are included or excluded from the model, the relationships between these variables in the model and the scale of the representation (Vesely and Rasmuson, 1984; Emlen, 1989; Regan et al., 2002a; Melbourne and Chesson, 2005). A very problematic outcome of model uncertainty is being faced with a choice of two or more models that describe a data set equally well but give very different predictions (Chatfield, 1995; Pascual et al., 1997);
- completeness meaning have all relevant risk factors and phenomena been considered (Ryan, 2008). More generally, completeness refers to uncertainty when enumerating all possible states of the world or elementary outcomes (Halpern, 2005). Completeness uncertainty is similar to model uncertainty but it is distinguished here because in risk assessment this uncertainty occurs very early in the initial hazard identification stages (Vesely and Rasmuson, 1984). An example in the context of import risk assessment is whether or not all potential pests and pathways associated with a new commodity have been identified and included within the risk assessment, prior to the development of a model of introduction, establishment and spread. The importance of completeness in risk assessment is sometimes identified in risk-defining functions. Kaplan (1997) and Kumamoto and Henley (1996), for example, define risk with functions that emphasise the need to completely enumerate the number of ways things can go wrong (Table 2.1);
- scenario uncertainty is the epistemic uncertainty associated with predictions of future situations. Draper et al. (2000) define it as different sets of conditions in locations and time (the future), resulting from both anthropogenic and natural processes. Ryan (2008) describes it as the uncertainty generated if a model is applied to situations outside the one under study, particularly where the model is used to make predictions about future events. Hayes et al. (2007a) identify "pressure scenarios" as combinations of uncertain ecosystem models and uncertain future perturbations to these systems. Similarly Dambacher et al. (2009) identify "perturbation scenarios" as possible future influences on fisheries and harvest practises. Scenario uncertainty is similar to completeness but it is distinguished here because it specifically addresses the uncertainty associated with the future. An example in an import risk assessment context may be changing trade patterns, or the effect of climate change on the distribution of pests in an exporting nation;

- subjective judgement occurs as a result of interpretation of data, for example where there is insufficient data to make a reliable judgement about parameter values (Regan et al., 2002a). It often manifests as differences in opinion between credible experts and is usually reducible by obtaining more data. Vesely and Rasmuson (1984) refer to this source of uncertainty as parameter uncertainty, noting that it occurs when experts make judgements about constants for which there is little or no data. Subjective judgement, however, is not exclusively restricted to parameter values (it is also important in model construction) and parameter uncertainty does not neatly fall into either of the categories of epistemic uncertainty or variability (Figure 2.3). Vesely and Rasmuson (1984) identify two other sources of parameter uncertainty: inapplicability and vagueness created by extrapolating general values from a population of interest (e.g. all water-cooled nuclear power plants in the USA) to a particular case (e.g. a particular water-cooled nuclear power plant). In statistics this source of uncertainty is often categorised and treated as a random effect, falling in the category of variability because it represents irreducible variation within a population across space or time (McCulloch and Searle, 2001; Cressie et al., 2007);
- systematic error is bias. Regan et al. (2002a) define it as the difference between the true value of the quantity of interest and the value to which the mean of the measurements converges as sample size increase. Systematic error can arise through accidental or deliberate exclusion of certain data (censoring), poorly calibrated measuring devices, non-random or poorly stratified field surveys, extrapolating between laboratory settings or surrogate organisms to other sites, subjects and species, or indeed any form of consistently incorrect process. The mechanisms that can generate consistently incorrect records are unfortunately numerous and often subtle. Burgman (2005) provides examples that highlight the diverse mechanisms that can generate bias;
- measurement uncertainty occurs because measuring devices and observers are imperfect. Measurement uncertainty, also known as measurement error, observation error or random error, manifests as (apparently) random variation in the measured value of a quantity (Regan et al., 2002a). The magnitude of measurement uncertainty depends on the number of measurements taken, the variation between measurements, the accuracy of the measuring device and the skill and training of the observer. The sources and characteristics of measurement uncertainty appear to be straightforward and well understood. Various authors, however, do not unanimously categorise it (Figure 2.3), probably because the term subsumes irreducible sources of uncertainty (finite measurement accuracy) with other sources of error that are clearly reducible, such as the skill of the observer;
- sampling uncertainty is related to measurement uncertainty. It is defined by Ferson et al. (2003) to be the epistemic uncertainty about the density function of a variable (i.e. its variability) that arises because only a portion of the individuals in a population have actually been measured. Frey and Rhodes (1998) refer to this source of uncertainty as the “random sampling error” that is introduced by estimating a statistic, such as the sample mean or standard deviation, from a limited number of samples<sup>5</sup>. In a risk assessment sampling uncertainty is reflected as uncertainty in the variability of the input parameters (risk factors) of a risk function. It is therefore useful to consider it as a type of parametric uncertainty.

---

<sup>5</sup>This definition should not be confused with Morgan and Henrion’s (1990) definition of random error caused by imperfect measuring devices (i.e. measurement uncertainty).

### 3 Uncertainty analysis

Cox and Baybutt (1981) define uncertainty analysis as a process that quantifies the uncertainty in a risk estimate, and partitions this uncertainty among the variables or risk factors that contribute to the risk estimate. Helton and Davis (2002) define it similarly as the answer to the question: what is the uncertainty in  $f(X)$  given the uncertainty in  $X$ ? This report takes a slightly more general view. Here we define uncertainty analysis as a three-step process that:

1. recognises, identifies and minimises linguistic uncertainty;
2. recognises, identifies, and where ever possible characterises, variability and epistemic uncertainty in the risk factors  $X$  and risk function  $f(X)$ ; and,
3. estimates the effect of epistemic uncertainty and variability on the outcomes of a risk assessment and reports this effect in an open and clear fashion.

This definition emphasises the different types and sources of uncertainty in risk assessment, and the importance of propagating epistemic uncertainty and variability through the assessment in an honest fashion.

#### 3.1 Linguistic uncertainty

Regan et al. (2002a) list a variety of treatments for linguistic uncertainty including providing precise numerical definitions for vague terms, and carefully specifying the context of terms and their meaning when these terms are potentially ambiguous. Providing precise numerical definitions for terms such as low, medium and high is a popular treatment for vagueness when assessing and communicating uncertainty. This approach, however, may impose a level of precision that analysts find difficult to work with (particularly in data-poor situations), does not guarantee that language-based operations will be commutative and unbiased (Box 2), and it can create a conundrum known as “Sorites Paradox”. The practical importance of the Sorites paradox occurs where small changes close to the boundary of a term, result in large changes in risk estimates. Consider, for example, a qualitative assessment that defines “low” likelihood as an event with an annual frequency of  $\leq 0.3$ , and “medium” likelihood as an event with an annual frequency  $\leq 0.7$ . An event with an annual frequency of 0.31 is therefore classified as “medium”. Analysts and stakeholders will probably be uncomfortable with this classification and it may dramatically change a subsequent risk estimate.

It is difficult to systematically minimise or eliminate all the various sources of linguistic uncertainty in a risk assessment. Evidence for this, together with further guidance on treatments, can be found in a variety of studies that have recently examined the use and interpretation of language-based descriptions of uncertainty in IPCC reports. Patt and Dessai (2005) and Budescu et al. (2009), for example, independently demonstrate that scientists and policy makers interpret terms such as “likely” or “very unlikely” in very different ways, and continue to do so even when they have read a set of numerical definitions for these terms. These studies demonstrate that the heuristics people use when judging uncertain events (Section 3.2.5) are not easily replaced with more consistent and mathematically coherent methodologies.

The study conducted by Patt and Dessai (2005) demonstrates that people’s interpretation of language-based descriptions of uncertainty is influenced by the context of the statement, particularly the perceived severity of the uncertain event that is being assessed. Budescu et al. (2009) also point to the (potentially complex) interplay of different sources of linguistic un-

certainty, wherein people assign different quantitative interpretations to a term such as “very unlikely” because they also assign different meaning to under-specified or context dependent terms such as “frequent”, “large” or “abrupt” that accompany the description of the uncertain event in question.

One suspects that in risk assessment problems the effects of linguistic uncertainty will persist unless people’s beliefs in uncertain events are carefully elicited with techniques that are specifically designed to expose misunderstandings, reduce heuristic biases, and ultimately replace language-based expressions of uncertainty with numerical expressions. For example, the treatments for linguistic uncertainty offered by the two studies cited above include identifying the source and type of uncertainty being described and the requirement that every probability term be accompanied by a numerical interval, but this interval need not be consistent for the same terms but allowed to vary in different contexts.

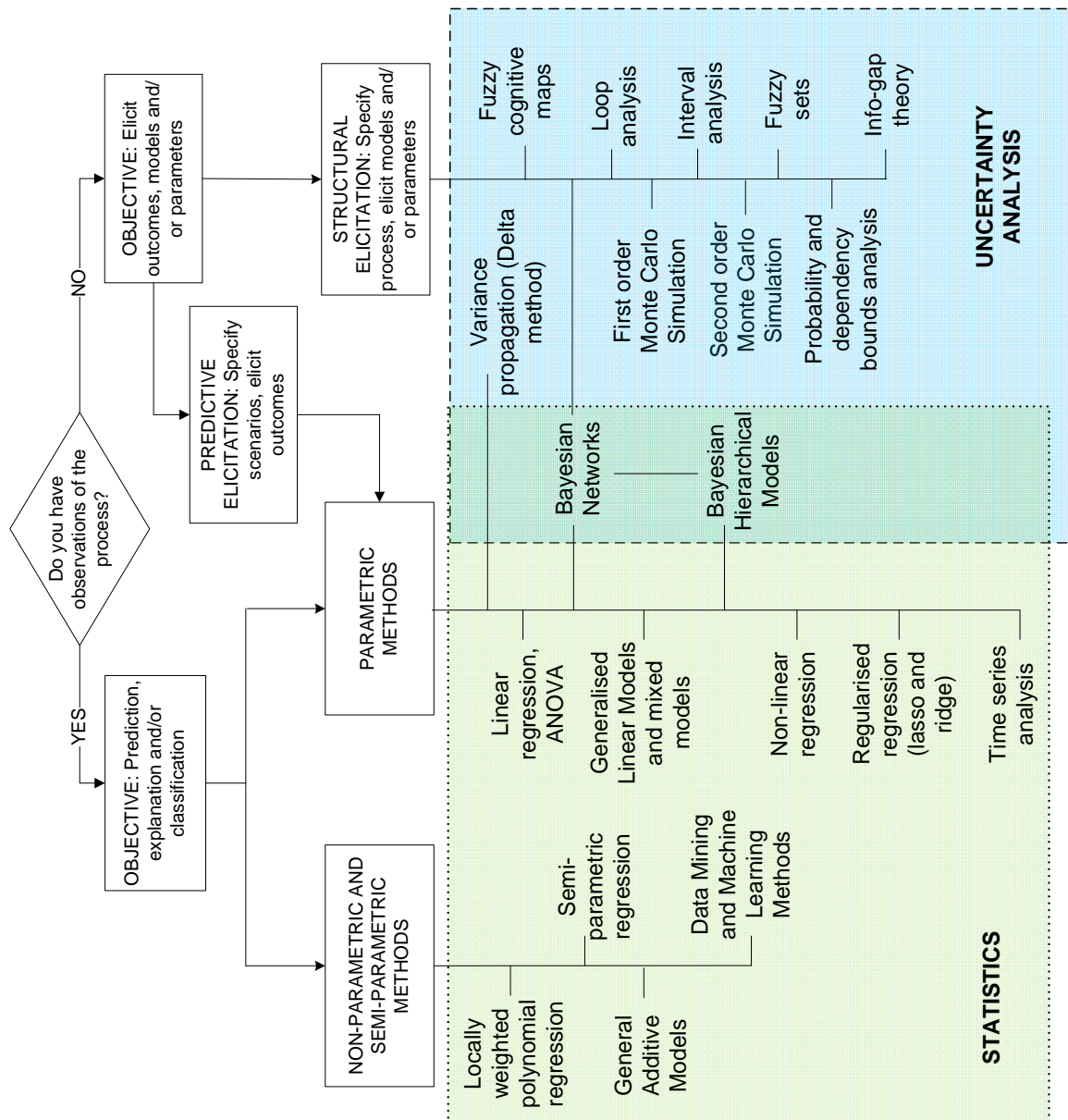
The methods recommended by O’Hagan et al. (2006) and Spiers-Bridge et al. (2010) are capable of eliciting a context-specific numerical interval but it is important to recognise that numerical intervals are only one of many numerical methods for modeling the partial information and imprecision associated with language (Walley, 1996). Fuzzy sets for example provide a calculus for vagueness, and hence a way to avoid the practical implications of Sorites paradox (Goguen, 1969; Klir and Folger, 1988), whilst Walley and Cooman (2001) argue that upper (and lower) probabilities provide a more general way to model simple forms of vagueness such as “the risk is low”. Fuzzy sets and imprecise probability, however, have not been widely adopted in risk assessment (Section 4), and Fuzzy sets have also been criticised on the grounds that they are not good representations of how humans actually process vague terms (Morgan and Henrion, 1990).

## **3.2 Epistemic uncertainty and variability**

### **3.2.1 Methods overview**

There are many methods to characterise, treat and propagate epistemic uncertainty and variability through an analysis but comparatively little guidance on the use of these methods in quantitative risk assessment. Beven (2009) is a notable exception in this regard providing comprehensive guidance on how to choose an uncertainty analysis method. Beven’s approach distinguishes “forward uncertainty propagation methods” that can be used in the absence of data, and methods that are conditioned upon or assimilate data. Here we maintain this distinction and recognise two different modes of uncertainty analysis: a forward, propagative mode that is informed by expert knowledge and experience, but lacks data, and a conditioning, inference, mode supported by knowledge, experience and conditioned by data (Figure 3.1).

Data is the starting point for any consideration of epistemic uncertainty and variability. It is important at this point to distinguish between data (observations of a process) and understanding and beliefs about a process. The latter is generated via experience and previous observations of the process in question (or similar processes), and is often the basis for quantitative and qualitative risk assessment, particularly where relevant data are absent. The key to uncertainty analysis in the absence of data is elicitation. Elicitation converts beliefs into outcomes, models or parameters that enable forward propagative uncertainty analysis. Structured elicitation techniques also provide an opportunity to avoid or minimise linguistic uncertainty, heuristic bias and dysfunctional group effects (Table 3), and are therefore good practise in quantitative risk assessment (Burmaster and Anderson, 1994).



**Figure 3.1:** Overview of probabilistic and some non-probabilistic methods for characterising, analysing and propagating variability and epistemic uncertainty in quantitative risk assessment, highlighting statistical methods, methods that risk analysts typically refer to as uncertainty analysis, and the elicitation routes to these methods



Elicitation can proceed in one of two ways: structural or predictive (Figure 3.1). Structural, or direct, elicitation specifies the process of interest and elicits from experts and stakeholders, models or model parameters. The elicitation typically focuses on the probability of an event within the process or on the probability of a large number of possible values for uncertain events in the process (Hampton et al., 1973). There are many methods for eliciting point estimates and distributions. These methods are discussed further in Section 3.2.5.

Data provided by structural elicitation (or actual observation) provides the basis for a large number of probabilistic and some non-probabilistic methods of forward uncertainty propagation. Probability-based methods include Bayesian Belief Networks, Monte Carlo simulation and probability bounds analysis, and are often the methods that risk analysts' refer to when they talk about uncertainty analysis techniques (Figure 3.1). Import risk assessment is typically performed in the absence of empirical observations. Hence, uncertainty analysis in this context is often performed using probabilistic forward propagation methods. Relevant non-probabilistic techniques include fuzzy sets and interval analysis, and graphical techniques such as fuzzy cognitive maps and qualitative modeling (aka loop analysis). These methods are discussed further in Section 4.

Predictive elicitation specifies scenarios that describe the process of interest and elicit from the expert and/or stakeholder the probability or value of the outcome. Embedded within the scenarios are explanatory variables or "keys", that condition the expert's response. In a risk assessment context these keys are risk factors. The elicited outcome is then treated as the dependent variable in regression-like problem (Kaldane and Wolfson, 1998; Barry and Lin, 2010). The basic idea behind predictive elicitation can be traced back to Egon Brunswick's theories of representative experimental design and the "lens model" that identifies the probabilistic relationship between cues (information on which a forecast or prediction is based), the event that is subsequently observed and the prediction (Brunswick, 1955, 1956). The primary advantage of predictive elicitation is that it enables the powerful machinery of modern regression techniques. The analysis can include, for example, the effect of experts in the regression model (as a fixed or random effect) and distinguish process error (attributable to the model, the cues) and judgement error (attributable to individual experts).

If the assessor has observations of the process then elicitation is no longer strictly necessary, but it can serve as way to test the predictive accuracy of experts (Stewart, 1990). Observations enable prediction, explanation and classification using data-mining and/or model-based statistical methods. In Figure 3.1 these techniques are loosely termed statistics, however, the distinction between uncertainty analysis and statistics is not constructive in a risk assessment context. Virtually all risk assessment frameworks emphasise the importance of monitoring and review, and here we emphasise that scientifically credible risk assessments should ultimately become a statistical exercise by making predictions about the risk-generating process that are testable and eventually tested against observations. Hence a more useful distinction is to consider forward uncertainty propagation methods as the initial tools of uncertainty analysis, that enable honest assessments (*sensu* Burgman) to proceed before observations are made and data are collected. Thereafter, uncertainty analysis should increasingly move to an inferential mode that relies on statistics and uncertainty analysis methods to characterise and quantify variability and epistemic uncertainty in data sets relevant to the problem in hand.

### 3.2.2 Model uncertainty

The structure of a model that describes a risk-generating process is an important determinant of the risk assessment result. The results of all risk assessments, qualitative or quantitative, are conditional on a model, at the very least a conceptual model. Hence model uncertainty is as relevant to qualitative risk assessment as it is to quantitative risk assessment. The only difference in this context is the transparency of the model and the analysis of uncertainty associated with it.

Models serve two roles: as heuristic devices they can provide insight into a system's behaviour and provide greater understanding of complex processes. As predictive devices they form the center piece of model-based decision support systems such as risk assessment and management strategy evaluation (Sainsbury et al., 2000; Walker et al., 2003). It is in the predictive role that model structure uncertainty is particularly important because ignoring it creates overly precise predictions. In practise, however, this source of uncertainty is often overlooked (Draper, 1995; Chatfield, 1995; Laskey, 1996; Breiman, 2001; Arhonditsis et al., 2007; Hill et al., 2007). The models used in pathway-based import risk assessment are predictive and typically highly abstract, decomposing complex bio-invasion events into a simple series of steps to which probabilities are assigned in order to estimate the overall risk (Figure 1.1).

Levins (1993) lists three desirable properties of a model: it should be precise, generalisable and realistic. Similar desiderata, such as relevance, flexibility and realism or "physicality" are listed in the literature (Reichart and Omlin, 1997; Pastorok et al., 2002). It is important to note, however, that it is impossible to maximise all three properties because the model will be no less complex than the real world. All models must sacrifice at least one property in order to maximise the other two, and hence model uncertainty is unavoidable. Model structure uncertainty can be handled in one of four ways:

- ignore it;
- treat it by comparing the predictions of alternative plausible models;
- treat it by conducting a second order uncertainty analysis that averages over alternative plausible models; or,
- envelope it by bounding the predictions of alternative models.

#### **Ignore it**

Model structure uncertainty is almost always ignored in qualitative risk assessment (because the conceptual model is internalised) and is often ignored in quantitative risk assessment. In some circumstances this is a defensible approach. For example, the structure of a risk model is sometimes dictated by normative industry standards - such as the ecotoxicity hazard quotient - or is mandated under national or international standards - such as the OIE animal product import risk assessment guidelines and models (Table 2.2, Morley (1993)). Mandated risk functions or models provide a comparable basis for risk ranking because differences between assessments cannot be attributed to different models. If the risk function or model structure is not mandated, and uncertainty surrounding its structure is ignored, then the analyst is simply assuming that the model used in the assessment is the best one for the purpose at hand. This assumption should be clearly stated in the assessment and where-ever possible addressed using the approaches identified in this report.

If the structure of a model is not mandated then the analyst must make a number of decisions

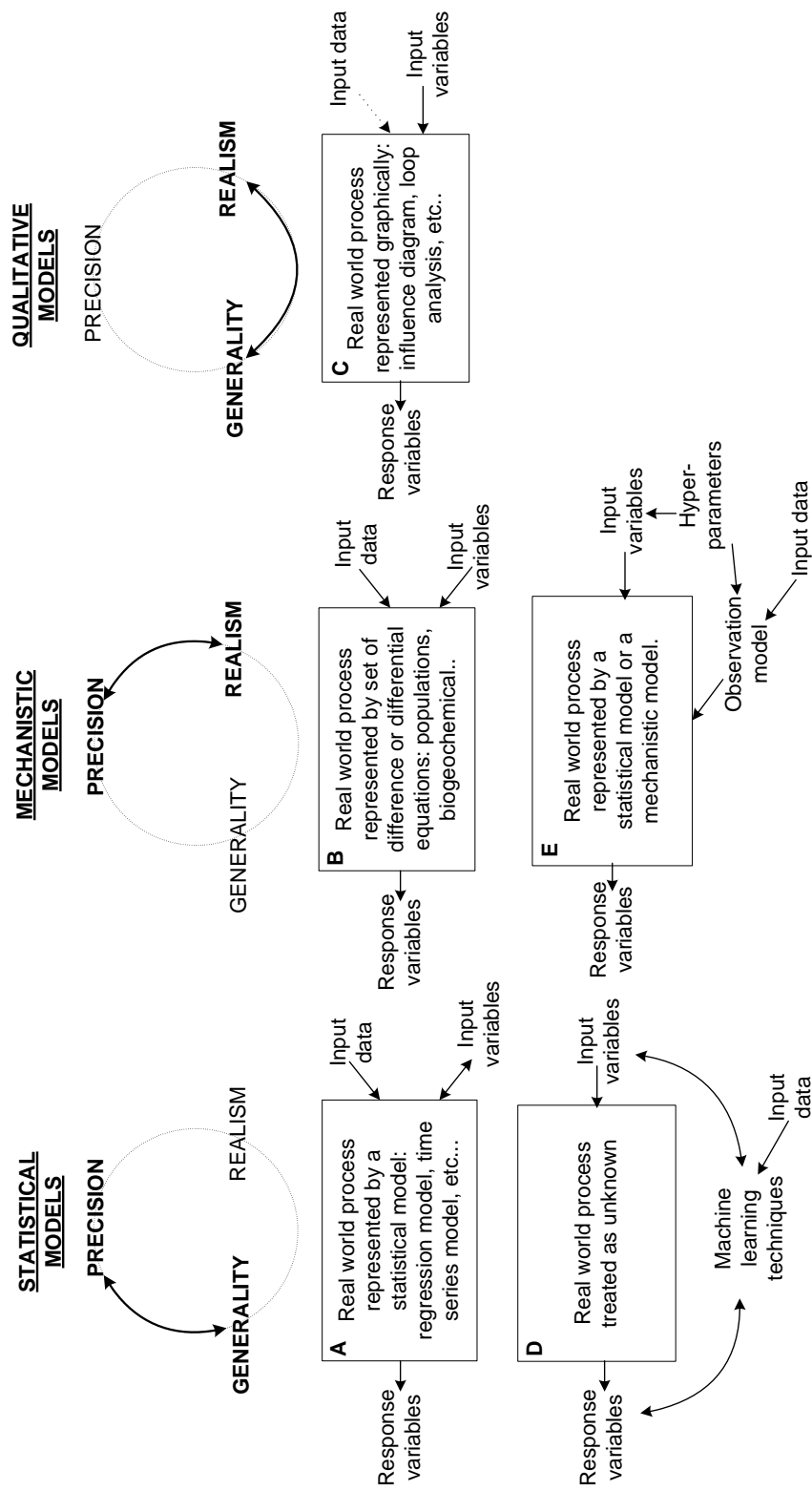
about which variables to include or exclude from the model, which components should be aggregated or disaggregated, the nature of the relationship between disaggregated components (e.g. linear or non-linear), what stochastic terms (process error, observation error, environmental variability, etc.) are to be included, and if they are included, how they are represented (normally distributed for example). These decisions can have an important effect on the results of the model. Investigating the effects of a range of plausible model structures is a recommended alternative to simply ignoring uncertainty about model structure (Hilborn and Mangel, 1997; Pascual et al., 1997). The application of this approach, however, varies between disciplines. This variation reflects, at least in part, three different modeling paradigms - statistical, mechanistic and qualitative - that can be loosely interpreted as maximising different pairs of Levin's desirable model properties (Figure 3.2).

### **Compare alternatives: Statistical models**

According to Breiman (2001) and Bolker (2008) the vast majority of statistics is based on a data-modeling culture that portrays real-world processes as a stochastic function of explanatory variables (caricature A of Figure 3.2). In statistics, however, the term “model” is ambiguous. It can refer to a single parameter, as in the use of a Poisson distribution as a “model” of the variability between random events in space or time, where there is no structure per se, to regression models of the form  $y = \beta\mathbf{X} + \epsilon$  where there is clear scope for structural uncertainty because  $\beta\mathbf{X}$  can vary from a simple linear combination of  $n$  explanatory variables,  $\beta\mathbf{X} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n$ , to much more complicated functions incorporating transformations of, and interactions between, the explanatory variables  $\mathbf{X}$  (Hastie et al., 2001). The discussion in this section refers to the latter interpretation.

The more recent machine learning or “algorithmic modeling culture” (Breiman, 2001; Hastie et al., 2001) (caricature D in Figure 3.2) eschews models in favour of algorithms that accurately predict patterns in data sets. Machine learning techniques rely on non-parametric techniques such as regression trees to select variables, and metrics of predictive accuracy to determine their credentials. They usually divide a data set into training and testing sets, but do not a priori specify a model for the data, and hence are able, at least in theory, to avoid the problem of model structure uncertainty. The choice or availability of explanatory variables, however, may have been based on a conceptual model, and hence these techniques are not completely freed from the uncertainty in the underlying conceptual model. Moreover, the distinction between these two cultures is not distinct as modern statistical techniques often mix tenets of parametric and non-parametric statistics (Ruppert et al., 2003; Friedman and Popescu, 2008).

The statistical data modeling culture typically treats alternative models by testing how well they fit a data-set using step-wise variable selection methods (Hocking, 1976; Faraway, 2002), together with metrics such as the Likelihood Ratio or Bayes Factor (Kass and Raftery, 1995), and/or information metrics such as Aikake's Information Criteria, the Deviance Information Criteria or the Bayesian Information Criteria (Burnham and Anderson, 1994; Ellison, 2004; Link and Barker, 2006). In multiple regression problems, however, the set of possible models can be very large ( $2^p$  where  $p$  is the number of potential explanatory variables) and parameter estimates will be biased if a model is formulated from the same data set that the parameter estimates are subsequently inferred from (Chatfield, 1995). Modern regression techniques tackle this problem by applying a penalty term to the loss function (the difference between observations and predictions) that reduces the magnitude of very large regression coefficients, and eliminates very small regression coefficients, thereby effectively estimating the model and its parameters simultaneously (Hastie et al., 2001).



**Figure 3.2:** Summary of three modeling paradigms (statistical, mechanistic and qualitative) highlighting two of the three desirable properties of a model that each emphasises, together with five caricatures of the modeling approach associated with each paradigm (reading left to right) and emergent variations within each paradigm (reading top to bottom).

### Average over alternatives: Statistical models

A second order analysis of model structure uncertainty can be completed in an inferential mode using Bayesian Model Averaging (BMA) (Draper, 1995; Hoeting et al., 1999). Hoeting *et al* describe the essential elements of BMA as follows: the posterior probability of the quantity of interest  $\Delta$  is given by:

$$P(\Delta|D) = \sum_{k=1}^K P(\Delta|M_k, D)P(M_k|D). \quad (3.1)$$

where the posterior probability of model  $M_k$  conditional on the data  $D$  is given by

$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum_{i=1}^K P(D|M_i)P(M_i)}, \quad (3.2)$$

The (marginal) likelihood of the data  $p(D|M_k)$  under model  $k$  is found by integrating across the model's parameters  $\theta_k$

$$P(D|M_k) = \int P(D|\theta_k, M_k)P(\theta_k|M_k)d\theta_k, \quad (3.3)$$

where  $P(D|\theta_k, M_k)$  is the likelihood of the data and  $P(\theta_k|M_k)$  is the prior probability of  $\theta_k$  for model  $M_k$ .

BMA emphasises that inference on a quantity of interest is conditional on the model and provides estimates that are weighted averages of the predictive distribution of each model, where the weights are the posterior probabilities of each model. The approach is very similar to the treatment of model uncertainty using second-order Monte Carlo Simulation, except that in this case solutions are conditioned on, and constrained by, observations. Models that are poorly supported by the data contribute little to the overall posterior estimate of the quantity of interest so long as they are assigned relatively uninformative prior probability.

The main difficulties with BMA are: a) there is limited guidance on how to specify the prior probabilities  $P(M_k)$  for each model; b) the integrals in Equation 3.3 can be difficult to compute, although they have been solved for certain classes of statistical model, and they can also be replaced with a BIC approximation under certain circumstances (Wintle et al., 2003); and c), the number of plausible models in the summation in Equation 3.1 can be very large.

This last issue can be tackled by using model selection methods to reduce the set of possible models down to a smaller set of plausible models using a method such as “Occam’s window”. Occam’s window calculates the ratio of the posterior probability of a small model and a larger model, and discards the larger model unless it is strongly supported by the data. This process proceeds in a pair-wise fashion, eliminating models that are not conclusively supported by the data. The selection criteria, however, is somewhat arbitrary. Hoeting et al. (1999), for example, suggest strong evidence for the large model as a posterior model probability that is 20 times greater than that of the smaller model. An alternative solution to the problem of a large number of candidate models is trans-dimensional Markov methods (Sisson, 2005), the most well known of which is reversible jump MCMC (Green, 1995). Reversible jump MCMC enables averaging across a large range of models and avoids any arbitrary distinction between strong and weak evidence. This approach, however, is technically and computationally demanding. Designing efficient between-model proposals is a difficult problem, particularly for models of different dimensions. Solving this problem is still an active area of statistical research (Fan et al., 2009).

### **Compare alternatives: Mechanistic models**

Mechanistic models emphasise precision and realism over generality, using difference or differential equations to portray real world processes (caricature B, Figure 3.2). These models are not easily generalised because each new context requires new initial conditions and a new set of parameter values. Hence, analysing the effects of alternative models structures within the mechanistic model paradigm can be a time-consuming process because of the overheads associated with parameterising each alternative structure.

Model uncertainty in this paradigm is again addressed by either considering a range of alternative structures (Pascual et al., 1997; Fulton et al., 2004) or by treating the model structure as an uncertain parameter and averaging over this parameter in a second-order uncertainty analysis. In the former approach, prudent models are typically chosen on the grounds of parsimony and ideally validated with independent data sets and clear falsification criteria (Fedra et al., 1981; Reckhow and Chapra, 1983; Oreskes et al., 1994; Rykiel, 1996; Turley and Ford, 2009).

A critical issue in this context is that in absence of data there are no strong constraints on the number of plausible models that should be examined. Theoretically the number of plausible models is infinite. Pascual et al. (1997) note this problem but suggest that in practise thoughtful ecologists should be able to limit their construed scenarios to a (presumably small) set of reasonable models. Models that exhibit ecologically impossible behaviour are clearly not plausible, but without observations the analyst may still face an inconveniently large number of plausible alternative structures. The number of “reasonable” mechanistic models may be constrained during uncertainty analysis if observations of the processes of interest are available. Unreasonable models will find relatively little support - i.e. have a relatively low likelihood - under the data. Hence, reasonable models can be identified in the mechanistic modeling paradigm by either specifying reasonable ranges on the predictive behaviour of the model’s parameters, dictated by observation and/or ecological theory (Fedra et al., 1981) or, where data are available, using methods such as Generalised Likelihood Uncertainty Estimation (Beven, 1993; Romanowicz and Beven, 2006), goodness of fit cost functions (Villars et al., 1998), target diagram methods (Los and Blaas, 2010) or formal Bayesian inference methods (Arhonditsis et al., 2007; Peters et al., 2010).

### **Average over alternatives: Mechanistic models**

Second order uncertainty analysis for mechanistic models can be achieved in a forward propagative mode via simulation. The simulation assigns a distribution function to a class of plausible models and treats this as a “model hyper-parameter” in a second-order Monte Carlo Simulation. For example, with two competing models the distribution would be a Bernoulli distribution taking two values with probability determined by the relative likelihood that either model is the right one (Ferson et al., 2003). This distribution is sampled on each pass of a Monte Carlo simulation, a model is chosen according to its relative likelihood, and the simulation completed in the usual manner by sampling from the distributions that represent variability in the model’s parameters. This technique results in predictions that represent an average of all models in the plausible class, weighted by the relative probability of each model. In this approach, however, the analyst is again faced with the unconstrained problem of specifying a number of plausible or possible models, and their relative likelihood. In the absence of other prior information, an analyst may simply assign equal likelihood to all plausible models.

Second order MCS and BMA provide a means to avoid statistical and mechanistic predictions that are overly precise, by averaging predictions across plausible model structures. Second

order MCS is a relatively simple approach whereas BMA requires considerable expertise and experience to successfully implement. Ferson et al. (2003), however, suggest that there is a far greater problem with these approaches: by representing model uncertainty via a mixture of possible models, these approaches may effectively average incompatible theories. They can result in posterior predictive distributions that are not supported by any of the theories that underlie the individual models.

#### **Envelope alternatives: Mechanistic models**

In light of potential difficulties with average over alternative models, Ferson et al. (2003) suggest an alternative strategy for handling model uncertainty is to envelope or bound the results of alternative plausible models. Large numbers of plausible models represent a practical challenge to this approach, and may lead to risk results that are too broad to be of practical value, but it is nonetheless sensible and computationally simpler than averaging.

Perhaps the best known example of the enveloping approach for mechanistic models is the multi-model global averages of surface warming produced by Working Group I of the Intergovernmental Panel on Climate Change (IPPC) (Figure 3.3 (IPPC, 2007)). This example shows the application of two different uncertainty treatment strategies: model uncertainty treated by enveloping the predictions of a range of alternative mechanistic models, and scenario uncertainty (discussed in Section 3.2.4) treated by considering 6 alternative future emission scenarios.

#### **Compare alternatives: Qualitative models**

Qualitative models emphasise generality and realism at the expense of precision (caricature C, Figure 3.2). Qualitative modeling techniques divert attention to the relationships and general trends between variables and away from the estimation of precise parameter values. Current applications range from simple influence diagrams (Cox et al., 2003; Niemeijer and de Groot, 2008; Negus et al., 2009), fuzzy cognitive maps<sup>6</sup> (Ozesmi and Ozesmi, 2003; Ramsey and Veltman, 2005; Ramsey and Norbury, 2009) to loop analysis and the use of signed digraphs that are mathematically similar representations of the differential equations that underlie most mechanistic process models (Dambacher et al., 2002, 2003b; Hosack et al., 2008). All of these methods are discussed further in Section 4.

An important facet of the qualitative modeling paradigm is that it does not strive to accurately parameterise cause and effect relationships and is therefore less reliant on data. It relies instead on the experience and beliefs that underlie the conceptual models of cause and effect held by all scientists and stakeholders. Model structure uncertainty in this paradigm is addressed by using a technique that enables the opinions of many different stakeholders to be canvassed and quickly converted into a range of plausible models that reflects the diversity of their opinions.

Qualitative models are constructed graphically, and are therefore well suited to situations that involve diverse groups of stakeholders or collaborators who may hold different views of a system but are unable to describe them mathematically. Simple (less than three or four variables) signed digraphs can be analysed by hand within a few minutes. More complex graphs and qualitative models can be analysed just as quickly if the assessor is familiar with certain software (refer to Section 4). Qualitative models can therefore treat model structure uncertainty by quickly exploring the implications of a large range of plausible models, prior to settling on a subset for subsequent (quantitative or qualitative) analysis. They are therefore a good way to elicit prior beliefs on model structure.

---

<sup>6</sup>Some readers may wonder why I have classified fuzzy cognitive maps as a qualitative modeling technique. The reasons for this are made apparent in Section 4

Again an important challenge in this context is the potentially very large number of plausible models. In the qualitative modeling paradigm it can be difficult to identify reasonable models from the range of plausible models because their predictions refer only to the direction of change, not its magnitude. Reasonable models can be identified by comparing the actual direction (positive, negative, unchanged) of change of a model's variables with the predicted direction of change using a simulation derived metric of sign determinacy (Hosack et al., 2008) but the issue of statistical (in)validation of qualitative model predictions remains an open research question.

### **3.2.3 Completeness**

Completeness uncertainty is treated in the hazard identification stage of a risk assessment. The purpose of the hazard identification is to identify what can go wrong and how things go wrong - i.e. enumerate and characterise all elementary outcomes. Hazard analysis tools range from simple unstructured brainstorming, to more systematic top-down (e.g. fault tree analysis, failure modes and effects analysis) or bottom up (event tree analysis, Hazard and Operability analysis, Hierarchical Holographic Modeling) approaches (Kletz, 1999; Hayes, 2002a,b; Hayes et al., 2004). Hazard analysis techniques such as fault tree and event tree analysis, can also be used to help construct "complete" descriptions of exposure pathways, and hence in the context of pathway-based import risk assessment, they can be viewed as a means to address model structure uncertainty as well (Burgman, 2005).

### **3.2.4 Scenario uncertainty**

Scenario uncertainty is typically addressed by postulating a range of "what if" scenarios that are designed to examine the effects of future conditions that are fundamentally different from current conditions. This can be achieved under any of the three modeling paradigms identified above, although some authors recommend that analysts shift towards increasingly simple, order of magnitude, models when making predictions far beyond the original domain of the model - e.g. when making predictions many decades into the future (Casman et al., 1999). Draper et al. (2000) use a combination of macro- and micro-scenarios to postulate possible situations millions of years into the future. Macro-scenarios are defined as high-level statements of future conditions (such as climate change), whereas micro-scenarios are more detailed descriptions of these scenarios within the modeling context adopted for the assessment. These authors state that the best results are obtained when future scenarios are as exhaustive as possible, emphasising the similarity between scenario uncertainty and completeness.

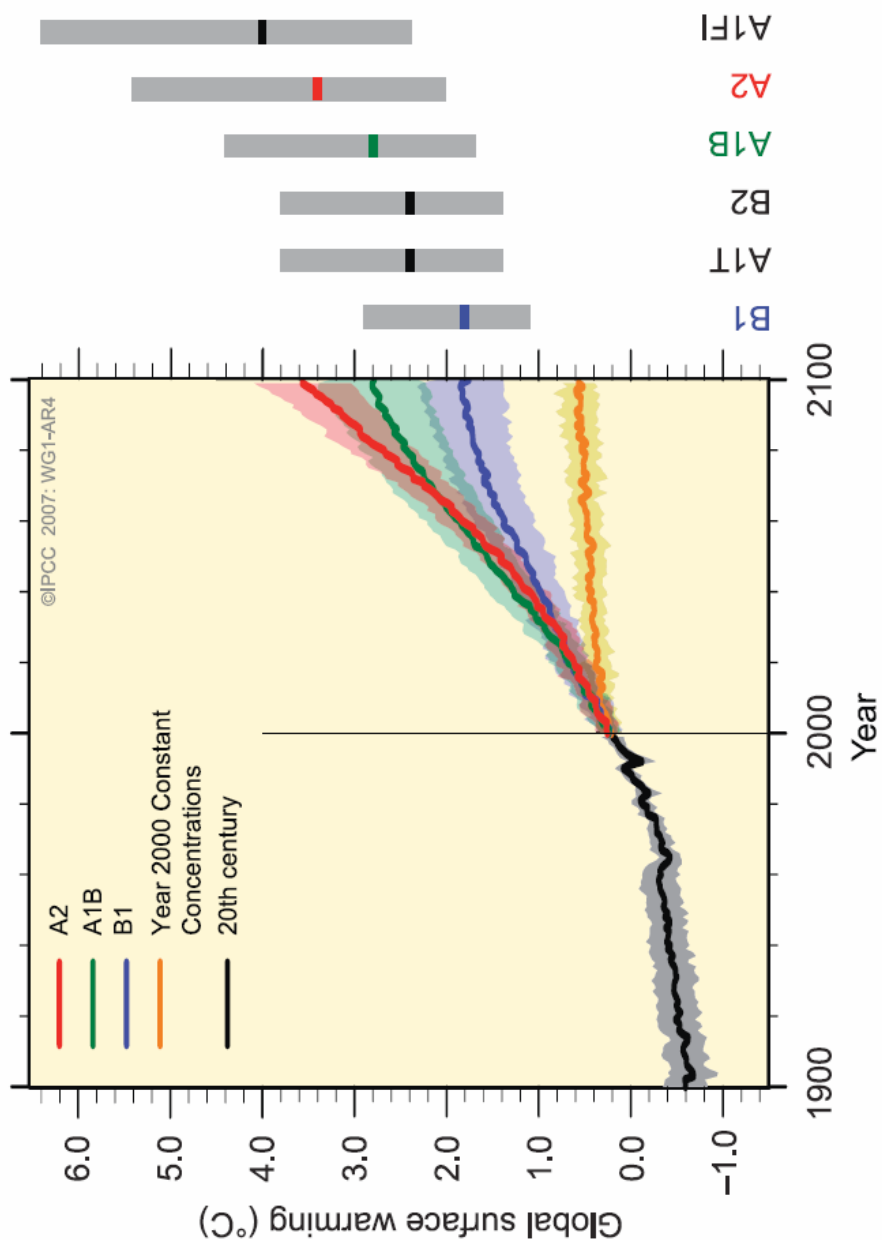
### **3.2.5 Subjective judgement**

Subjective judgement is an important source of uncertainty in all forms of risk assessment, but is particularly important in qualitative risk assessment because it relies exclusively on this form of information. Subjective judgement contributes to both model structure uncertainty and parameter uncertainty. Treatments for model structure uncertainty are discussed above and in Section 2.5.1.

Subjective risk estimates are most reliable (predictions and outcomes are well calibrated) when relevant data is plentiful, feedback on the accuracy of the predictions is immediate and the experts have strong incentives to improve their performance (Franklin et al., 2008). At least two of these criteria, however, are not met for import risk assessment (as currently practised in Australia). Hence, in data-scarce situations, subjective judgement is a necessary but potentially inaccurate source of information.



### MULTI-MODEL AVERAGES AND ASSESSED RANGES FOR SURFACE WARMING



**Figure 3.3:** IPCC estimates of global average surface warming (relative to 1980–1999) showing examples of two treatment strategies for different types of uncertainty: model uncertainty treated by enveloping the predictions of alternative mechanistic models (grey bars on right hand side), in this instance a simple climate model, several earth system models of intermediate complexity and a large number of Atmosphere–Ocean General Circulation Models; and, scenario uncertainty treated by postulating 6 future emission scenarios (solid coloured lines). (Source: IPCC (2007))

Subjective judgements about uncertain parameters are usually obtained by asking experts for a (qualitative or quantitative) estimate of the parameter (Hampton et al., 1973; Regan et al., 2002a). Naive, unstructured approaches to this problem should be avoided because:

- terms such as “low likelihood” confound linguistic uncertainty with other forms of uncertainty, including subjective judgement. The different sources of uncertainty that are compounded within these terms cannot be separated without (at least) numerically defining the meaning of terms such as “low”;
- humans use particular cognitive structures, and well-documented heuristics when making judgements about uncertain problems (Anderson, 1998; Hubbard, 2009). These cognitive structures can fail, and the heuristics can lead to outright errors unless information is elicited and presented to assessors in particular formats (Table 3.1); and,
- unstructured group elicitation is prone to “dysfunctions of interactive group processes”, such as group pressure for conformity and the influence of dominant personalities, that may exacerbate individual heuristic effects, particularly overconfidence (Bier, 2004).

Fortunately there are many ways to elicit models and parameter values from experts in a manner that helps to avoid or minimise heuristic bias and dysfunctional group effects. The four-step elicitation procedure adopted in the case study (Box 1) was specifically designed to address many of these issues, and the literature contains ample guidance and additional methodology. Renooij (2001) and O’Hagan et al. (2006) for example, provide very readable introductions to this topic. Jenkinson (2005) and Burgman (2005) provide comprehensive accounts of methods for eliciting probabilities, distributions and risk assessment models. Kynn (2008) provides an excellent summary of expert heuristics and offers ten guidelines for eliciting probability estimates from experts. Kaplan and Burmaster (1999) provide similar guidance.

Franklin et al. (2008) argues that some sources of heuristic bias can be attenuated by accountability, and recommends that in data-poor situations subjective judgement and quantitative methods are mixed in an “advocacy” framework similar to that used by banks and financial regulators to analyse operational risk. The advocacy framework replaces the immediate feedback of data with expert review and potentially hostile stakeholder reaction, mediated by an independent panel. This situation is not too dissimilar to the current practice of import risk assessment and may therefore be relatively easy to implement once the necessary quantitative elements were incorporated into the current practice.

Ferson et al. (2003) summarise five different approaches that can simultaneously capture epistemic uncertainty and variability from *inter alia* expert opinion, and discusses ways to aggregate multiple estimates. Some of these techniques are used in subsequent sections of this report to analyse the information elicited for the project case study. It is important to recognise that numerical definitions of terms such as “low”, coupled with structured elicitation techniques, provides a mechanism to minimise linguistic uncertainty and a means of transition from qualitative to quantitative risk assessment. Qualitative risk assessment is unnecessary once an expert’s opinion has been carefully elicited and numerically defined. The main impediments to quantitative risk assessment are the overheads associated with formal elicitation techniques (but in practise these may be no greater than those of unstructured workshops) and the computational hurdles of performing arithmetic with random variables (Section 2.5).

Subjective judgement is also important in the treatment of systematic error, particularly in decisions to ignore it by assuming, for example, the carcinogenic potency of a large dose of a

**Table 3.1:** A summary of some of the key issues and heuristics that can result in bias when parameter values are naively elicited from assessors<sup>7</sup>

Issue	Description and implications
Overconfidence	Assessors tend to overestimate the accuracy of their beliefs or alternatively underestimating the uncertainty in a process. This leads to, for example, difficulties in estimating the tails of a distribution that represents the variability of a parameter, or equivalently confidence intervals that are consistently too narrow and do not span the true value of a parameter (where this is known)
Availability	Assessors link their probability estimates to the frequency with which they can recall an event. Unusual, catastrophic, newsworthy and recent events are more easily remembered than distant or routine events that are not newsworthy. Hence assessors underestimate the frequency of the latter (e.g. annual number of deaths by stroke) and overestimate the frequency of the former (e.g. annual number of deaths caused by tornadoes)
Representativeness	Assessors judge the probability that A belongs to B by how representative or similar A is to B, leading to a range of phenomena such as base rate neglect, the law of small numbers and insensitivity to sample size. For example the string of coin tosses HTHHTH is deemed to be more likely than HHHTTT or HTHHTT (when in fact it is not) because the process of tossing a coin is known to be random and the first string looks more random than the second or third
Anchoring and adjustment	An assessor (or group of assessors) tends to anchor around any initial estimate and adjust their final estimate from this value irrespective of the accuracy of the initial estimate. The subsequent adjustment is usually insufficient to correct the initial inaccuracy due to the effect of overconfidence
Scale format and translation	An assessor's response to the same problem tends to vary depending on the scale and manner in which information is presented. Transforming a variable to another scale sometimes creates confusion and mathematically incoherent estimates
Motivational bias and affect	Assessors can provide inaccurate or unreliable estimates because it is beneficial for them to do so. Proponents of a project may overestimate benefits and underestimate risks or costs because of enthusiasm for the project and/or ambition within a company. This can be particularly problematic when assessing the risks of a new technology because those with substantive technological knowledge often stand to gain from its adoption
Non-adherence to probability axioms	Assessors fail to provide probability estimates that adhere to the axioms of probability (e.g. conditional probability axioms) unless the information is presented in the form of a frequency, frequency tree or with the assistance of a conditional probability table

Source references: Morgan and Henrion (1990), Kaldane and Wolfson (1998), Burgman (2005), Kynn (2008) and Kuhnert and Hayes (2009)

chemical in mice is the same as its low dose potency in humans (Morgan and Henrion, 1990). Regan et al. (2002a) note that theoretical grounds for systematic error may be identified but in practise it is very difficult to recognise it when it occurs. Burgman (2005) recommends generic treatment strategies such as diligent inspection of experimental theory, comparing estimates with scientific theory, independent studies, replication and careful attention to detail. Other generic advice stresses caution when extrapolating laboratory studies to real world situations, or making predictions based on regression analysis of historical data (Morgan and Henrion, 1990).

### **3.2.6 Measurement error**

Measurement error can occur in conjunction with systematic error leading to observations that are both imprecise and inaccurate. Measurement error is treated by reporting observations within the limits of the measuring device, and applying standard statistical techniques that measure variability around the mean, such as the range or variance of a variable, or the standard error of the mean (Sokal and Rohlf, 1995).

### **3.2.7 Parametric uncertainty and sampling uncertainty**

The various forms of variability encountered in risk assessment problems are usually characterised and propagated by assigning a probability distribution to the factors or variables (parameters) that are inputs to the risk function. When implementing this approach the analyst must choose the type of distribution (Normal, Triangular, Gamma, etc.) and the parameters that define its location, shape and spread. The first problem is the large number of distributions that might be used to describe the variability in a model parameter. Patil et al. (1984a,b), for example, describe 136 discrete univariate and multivariate distributions, and 151 continuous univariate distributions.

The second problem is that the analyst may also be uncertain about the parameters of any particular distribution due primarily to the effects of measurement error, subjective judgement or sampling uncertainty. Sampling uncertainty arises because we rarely measure all of a population. It is evident in the difference between the empirical distribution functions derived from two independent samples taken from the same population, or more simply the different summary statistics associated with these samples. Sampling uncertainty is an important source of parametric uncertainty in risk assessment because limited observations are often the basis for characterising the variability of the factors (input parameters) in risk functions.

There are five ways to treat parametric uncertainty in a risk assessment:

- ignore it;
- eliminate one or more of the parameters from the model via a simpler or more complex risk function;
- treat the variability by conducting a sensitivity analysis that compares the effects on the risk estimate of a variety of different density functions;
- treat the variability by enveloping it so that the results of the risk assessment are guaranteed to lie between upper and lower bounds; or,
- treat it by averaging over the uncertainty in the parameters of a distribution function using a second-order uncertainty analysis method.

### Ignore it

Ignoring uncertainty about the type of density function can sometimes be justified on theoretical grounds. Gardner and O'Neill (1983) suggest that a probability distribution should be chosen on the basis of: a) the probabilistic properties of the process being modelled; b) the empirical distribution of the data; and, c) any information on the expected distribution of the system's behaviour. Thompson (1999), however, argues that the single-most important criterion for selecting the type of distribution is knowledge about the nature of the process that generated the random variable.

Many parametric distributions have been developed to describe the scientific findings that similar processes lead to similar patterns of variability (Thompson, 1999). If the mechanistic process that generates variability is well understood then the analyst might simply choose an appropriate density function. For example, the choice of a Normal (Gaussian) distribution might be justified on the grounds that the uncertain input results from a large number of small independent errors. Equivalent theoretical justifications exist for many distributions commonly employed in risk assessment (Stephens et al., 1993; Ferson et al., 2003) (Tables 3.2 and 3.3).

An alternative rationale for choosing one probability density function over all others can be found in the Maximum Entropy principle. The entropy of a discrete random variable  $X$  is given by the differential entropy function

$$H(X) = - \int p(x) \ln p(x) dx. \quad (3.4)$$

The maximum entropy principle argues that a density function should be chosen that maximises this entropy function, subject to the constraints of the total law of probability - i.e. the density function must integrate to 1, and other constraints provided by the available information - i.e. any observations of the process to hand. The rationale for this argument is that the maximum entropy function subject to these constraints represents the minimum amount of information that is consistent with these constraints. For example, the density function that maximises the differential entropy, subject to the constraints that the first and second moments are equal to a sample mean and variance, is the Normal distribution with these moments.

The mathematics associated with the entropy theory are quite complex. The constrained maximisation is performed using Lagrange multipliers and the calculus of variations. It does, however, provide an arguably objective rationale for selecting a probability density function and leads to a number of results that support "uninformative" priors used in Bayesian inference (Vose, 2000). Lee and Wright (1994) and Ferson (1996a) discuss the application of this approach to risk assessment problems. A more detailed introduction to the genesis and mathematics of the maximum entropy principle can be found in Bishop (2006).

Ignoring uncertainty about the parameters of a given distribution function might also be justified on the grounds that a single distribution provides an excellent fit to the available data. Murray (2002), Thompson (1999) and Cullen and Frey (1999) provide good guidance on the parametric and non-parametric methods that can be used to gauge the concordance between a distribution and a data set. Large data sets (hundreds of observations), however, may be required to accurately distinguish the tails of distributions if the ratio of the standard deviation to the mean is greater than one (Haas, 1997). In ecological systems, however, data sets of this size are relatively rare. In more typical situations analysts have very little data to guide the choice of a distribution, or may find that two or more distributions fit (small) data sets equally well but lead to very different predictions, particularly in the tails.

**Table 3.2:** Common parametric distributions and the generating process used to justify their selection in quantitative risk assessment (amended from Thompson (1999))

Name	Range	Mechanistic basis for selection
Beta ( $a, b$ )	$0 \leq x \leq 1$	Fractional quantities including the probability of observing an event given data of $n$ trials and $r$ successes. Conjugate prior for a Bernoulli likelihood. Uncertainty in the probability of an event occurring. A modified four parameter version is sometimes used for subjective distributions
Binomial ( $n, p$ )	$0 \leq x \leq n$	Number of success in $n$ attempts where each attempt has a probability $p$ of succeeding and probability $1 - p$ of failing. A Bernoulli distribution is a special case with $n = 1$ . Used to model dose-response data (e.g. number of animals with disease out of the number exposed)
Cauchy ( $a, b$ )	$-\infty < x < +\infty$	Cauchy (0, 1) arises from the ratio of two independent Normal variates
Chi-squared ( $\nu$ )	$0 \leq x < +\infty$	Sum of the squares of $\nu$ Normal variates. Describes the sampling distribution for a small sample size variance of a Normal distribution. Special case of the Gamma $\alpha = \frac{\nu}{2}, \beta = 2$
Error function ( $h$ )	$-\infty < x < \infty$	Distribution of normal random errors
Erlang ( $\alpha, \beta$ )	$0 \leq x < +\infty$	Describes the time between events. Equivalent to a Gamma distribution with discrete value of $\alpha$
Exponential ( $\beta$ )	$0 \leq x < +\infty$	Describes the inter-arrival times between random, successive, independent events that occur at a constant average rate. $\beta$ represents the mean time between events. Special case of the Weibull with $\alpha = 1$
Gamma ( $\alpha, \beta$ )	$0 \leq x < +\infty$	Describes the time required for a specified number of events ( $\alpha$ ) to occur, given a random Poisson process with a constant average time interval of $\beta$ between events. The sum of independent exponential random variables. Equals exponential when $\alpha = 1$ and equals chi-squared when Gamma $\alpha = \frac{\nu}{2}, \beta = 2$
Geometric ( $p$ )	$0 \leq x < +\infty$	Results from an exponential decline. Describes the number of failures that occur between successes given a probability $p$ of success. Equals the negative binomial with $n = 1$ . Discrete version of the exponential distribution
Gumbel ( $a, b$ )	$-\infty < x < +\infty$	One of the Extreme Value distributions. Used to estimate a maximum observable value. Represents the limit of the maximum value of $n$ identically distributed continuous random variables as $n$ approaches infinity
Hypergeometric ( $S, n, N$ )	$0 \leq x \leq \min(n, S)$	Describes the number of items of one type randomly chosen without replacement in a sample of size $n$ from a population of size $N$ containing $S$ such items (and $N - S$ other types)

**Table 3.3:** Continuation of parametric distributions commonly used in risk assessment and the generating process used to justify their selection

Name	Range	Mechanistic basis for selection
Logistic ( $\alpha, \beta$ )	$-\infty < x < +\infty$	Approximately describes the middle of the range of a set of identically distributed variables - e.g. the average of the highest and lowest values observed for identical items. Commonly used to describe population growth
Lognormal ( $\mu, \sigma$ )	$0 \leq x < +\infty$	Arises from multiplicative processes. Describe quantities that are the product of a large number of other quantities. Used to describe quantities with occasionally large values (right skewed quantities) such as the concentration of a substance in the environment, the number of species in a family of community, resistance to chemicals and bacteria on plants (Ott, 1995; Limpert et al., 2001)
Negative Binomial ( $n, p$ )	$0 \leq x < +\infty$	Describes the number of attempts to succeed $n$ times given a constant probability of success $p$ on each attempt. Related to the Geometric distribution
Normal ( $\mu, \sigma$ )	$-\infty < x < +\infty$	Also known as the Gaussian distribution. Arises from the addition of a large number of random variables (as long as none of them dominate the variance). Results from the Central Limit Theorem of probability. Used for a large number of physical and biological processes such as dispersion
Pareto ( $a, b$ )	$a \leq x < +\infty$	Also known as the Zeta distribution. Describes variables with their highest probability at a minimum value $a > 0$ , and for which probability density declines geometrically above that value
Poisson ( $\lambda$ )	$0 \leq x < +\infty$	Describes the number of relatively rare independent events occurring within a fixed time interval or unit of space, such as counts of microbes and radiation in an interval of time
Raleigh ( $b$ )	$0 \leq x < +\infty$	Describes the time to the occurrence of an event. Maximum entropy distribution when arithmetic mean and geometric mean are known. Equivalent to a chi-squared (2) distribution when $b = 1$ , and a Weibull ( $2, \sqrt{2}b$ ) distribution
Student's t ( $\nu$ )	$-\infty < x < +\infty$	Describes the sampling distribution of the mean of a Normal random variable for a small sample size
Triangular ( $a, b, c$ )	$a \leq x \leq c$	Often used when the minimum, maximum and mode of a variable assumed to be perfectly known.
Uniform ( $a, b$ )	$a \leq x \leq b$	Assumes all intervals between precisely known upper and lower limits have constant equal probability. Maximum entropy distribution when only the minimum and maximum values are known
Weibull ( $\alpha, \beta$ )	$0 \leq x < +\infty$	Describes the lifetime of an item, and the time between occurrence of events (e.g. success or failure) when the probability of occurrence changes with time

### **Model or eliminate it**

An alternative treatment for parametric uncertainty is to try and emulate the process that creates the variability reflected in the parameter via a more complex risk function, or eliminate the parameter altogether via a simpler risk function. Modeling variability effectively transfers the problem of parametric uncertainty into one of model structure uncertainty by creating a more complex model. This will rarely be an attractive option. A much simpler and potentially more attractive option is to re-state the problem in simpler terms. Hayes (1997) emphasises that the complexity of risk assessment models are largely driven by the assessment endpoints and recommends choosing endpoints that are as simple as possible whilst still being relevant for policy and management purposes. Assessment endpoints at the end of risk event chains, such as the spread of an established pest, inevitably entail more complex models than endpoints selected at intermediate points in an event chain, such as the number of organisms that successfully arrive at a nation's borders.

Reducing model complexity by choosing endpoints that are as simple as possible whilst still remaining relevant for management purposes can be an effective treatment for parametric uncertainty. The trick here is to offer risk managers a range of endpoints within the risk-generating chain of events and ask them, "What is the simplest assessment endpoint that allows you to make decisions on the acceptability of this risk?" before embarking on the assessment. This strategy was successfully applied to the risks associated with ballast water introductions in Australia (Hayes, 2003; Hayes et al., 2009). This is clearly not an option, however, where the risk assessment endpoint is mandated by national or international legislation.

### **Compare alternatives**

If the analyst cannot ignore, simplify or model an uncertain parameter, and has trouble justifying a distribution on the grounds of theory, and/or relevant data sets are simply too small to accurately distinguish between distributions, then the next alternative is to explore the implications of parametric uncertainty by comparing the effects that different plausible distributions have on the results of the risk assessment. This is better than simply choosing one distribution on the grounds of, for example, mathematical convenience. This approach, however, is incomplete because the number of plausible distributions may be large and there is no a priori guidance on how many distributions are "enough" to fully capture the uncertainty in the input parameter. A complete solution, however, can be achieved by enveloping the uncertainty surrounding the input parameter using probability boxes.

### **Envelope alternatives**

Ferson et al. (2003) provides an excellent summary of ways to envelope the distribution of an uncertain model parameter using probability boxes built on the basis of whatever information is available. Probability boxes are a class of imprecise probabilities that can be used to describe a distribution whose shape is precisely known (or assumed) but whose parameters are uncertain, a distribution whose parameters are precisely known but whose type is uncertain, and a distribution of uncertain type and parameters. These methods form the foundation of probability bounds analysis and are discussed further in Section 4. In this context the strength and distinguishing hallmark of probability bounds analysis is that it frees the analyst from having to make any assumptions about the type of the distribution function or its uncertain parameters. Probability boxes can be created that are "best-possible" meaning that they are guaranteed to envelope the true quantity in the smallest box possible given the available data (Ferson and Hajagos, 2004).



### Average over alternatives

The final strategy for dealing with parametric uncertainty is to average over it<sup>8</sup> using a second-order uncertainty analysis method. In this context “second-order” refers to methods that propagate uncertainty about the parameters (mean, standard deviation, etc.) of a given distribution function through the risk assessment. This approach therefore assumes that the analyst is able to select one or more probability distribution functions to represent the variability in a risk factor, and then accurately characterise the uncertainty in the parameters of this distribution. This characterisation can be achieved in five different ways, four of which require at least two observations (samples) of the parameter in question:

- using expert opinion; Hoffman and Hammonds (1994) suggest that expert opinion can be used to characterise uncertainty in the variability associated with input parameters in quantitative risk assessment. This is not, however, a compelling approach given the overheads associated with expert elicitation methods and the large range of more objective numerical methods;
- via standard parametric results; there are a number of standard statistical results that describe the reliability (standard error) of statistics derived from random samples, and more importantly provide intervals (confidence limits) that for the purposes of a risk assessment can be interpreted as bounds on the true value of the unknown population parameter (Freund, 1992; Sokal and Rohlf, 1995). These approaches, however, require some assumption about the nature of the variability in the underlying population, typically that the variability in the population is normally distributed;
- numerically via bootstrapping; the bootstrap is a data-based simulation method that can be used to quantify the sampling uncertainty associated with a statistic that is estimated from a limited number of samples (Efron and Tibshirani, 1994; Frey and Burmaster, 1999). Bootstraps can be parametric or non-parametric. Non-parametric bootstrap involves resampling a data set with replacement many times, and generating an estimate of the statistic on each occasion in order to characterise its variability. The main shortcoming of this approach is that the values of each bootstrap sample will always lie between the maximum and minimum values of the original data set. This can create substantial bias when characterising the population-level variability in a sample statistic, particularly with small data sets. Parametric bootstraps avoid this problem but at the cost of having to make assumptions about the distribution of the underlying population variability. The application of the parametric bootstrap to risk assessment is discussed extensively in Frey and Rhodes (1998), Cullen and Frey (1999) and Vose (2000);
- via Maximum Likelihood; the likelihood function actually provides two ways to construct confidence intervals around sample-based statistics, through the likelihood profile and via the Fisher Information matrix (Clark, 2007). Both approaches require the analyst to identify the probability density function that describes the variability in the population from which sample observations were drawn. The observations are used to determine the maximum likelihood estimate (MLE) of the parameters of the probability density function that is known or assumed to describe variability in the population. The likelihood profile uses a likelihood ratio constructed around the MLE, and a test statistic called the deviance that has a  $\chi^2$  distribution, to generate confidence intervals for the sample statistic centered on the MLE. The Fisher Information method uses the curvature of the log-likelihood

---

<sup>8</sup>It is important to recognise that averaging reduces variability and therefore in a sense erases uncertainty.

function in the neighborhood of the MLE to estimate the variance of the sample statistic (Burmaster and Thompson, 1998; Frey and Burmaster, 1999). The curvature, however, is a quadratic approximation - a Taylor series expansion around the MLE up to the second order terms - and can give poor results with non-Normal likelihood functions and small data sets (Clark, 2007).

- via Bayesian predictive intervals; the Bayesian equivalent of confidence intervals - credible intervals - are simply percentiles of the posterior distribution and are therefore a natural byproduct of Bayesian inference. The Bayesian posterior distribution is proportional to the product of the likelihood of the population parameters given the observations, and the prior probability of the parameters. This approach therefore requires additional assumptions regarding the prior distribution of the population parameters. The additional information these assumptions bring to the problem can be considered advantageous or simply another source of unwanted subjective input. See Hayes (1998) and ? for further discussion, from a risk assessment perspective, on some of the issues associated with Bayesian inference.

For large sample sizes each of the methods discussed above yield similar results. When sample sizes are small, however, they can give substantially different results. The choice of one method over another is dictated by computational considerations (e.g. is the likelihood function differentiable?) and the contextual information associated with the problem in hand (e.g. are assumptions about the nature of the population variability tenable?). Clark (2007) provides a very readable explanation of each of these methods, together with additional guidance on their use.

### 3.3 Dependence

Risk assessment entails arithmetic with uncertain and variable quantities. This uncertainty and variability imposes a number of theoretical and practical challenges, one of which is the effect of dependence. Dependence between the variables of a risk assessment model implies that the outcomes of the random variables are somehow related to each other. For example, if  $X$  and  $Y$  are two continuous real-valued random variables, then positive linear dependence implies that when the value of  $X$  is large relative to its full range, then so is the value of  $Y$ . Likewise when the value of  $X$  is small relative to its full range, then so is the value of  $Y$  (Regan et al., 2004). It is important to recognise that the dependence between random variables can be complex and not necessarily linear and, despite contrary suggestions in the literature (Smith et al., 1992; Slob, 1994; Bukowski et al., 1995), its effect on risk estimates cannot be accurately characterised by a sensitivity analysis that varies a correlation coefficient between + 1 and -1. Furthermore, low or zero correlation between two random variables does not generally mean that they are independent (Ferson and Hajagos, 2006, 2004; Ferson, 1996a).

The potential sources of dependence are many and varied (Table 3.4). Two sources of dependence, functional dependence and variable repetition, occur via construction of the risk function or model. They are perhaps trivial sources of dependence but nonetheless have important implications. Functional dependence prevents (except in special cases) analysts from rearranging equations with uncertain operands in the usual (high school) way to solve for an unknown (Ferson, 1996a). Variable repetition inflates probability bounds beyond those that are “best possible” (Section 4). Statistical dependence occurs in the parameters of some distribution functions also by virtue of construction. The shape parameters of the Gamma and Beta distributions

are important cases. Scatter plots of these parameters, constructed by fitting the distributions to bootstrapped samples, display patterns of non-linear dependency that cannot be adequately described by a linear correlation coefficient (Frey and Rhodes, 1998).

The other types of dependence identified in Table 3.4 are a facet of the real-world. They are diverse and include the spatial and temporal dependence commonly observed in time series and spatially-explicit processes (such as the spread of an invading organism), the dependence created by biological "laws" such as the body weight and surface area of organisms, and the dependence between the opinions of experts created by common schooling or exposure to the same (limited) body of evidence. Many of these sources are directly relevant to import risk assessment problems.

The types and sources of dependence identified here are not comprehensive. Hickman (1983), for example, identifies many sources of dependency in complex industrial systems (nuclear power plants) that are not listed here. Li (2000) identifies "macroeconomic forces" as an important source of dependence in financial instruments, such as Collateralized Loan Obligations (CLOs). These forces cause loan defaults to increase during times of recession and decrease during times of economic growth, creating positive dependence in CLOs. This positive dependence was partially responsible for the Global Financial Crisis of 2009. There are doubtless other potential sources of dependency in the risk assessment problems. These examples listed here and in Table 3.4 illustrate the potentially diverse and sometimes subtle sources of dependence that may be encountered in risk assessment.

Dependence between experts is difficult to avoid but it can be minimised by canvassing the opinions of a wide range of experts and stakeholders, and by using the structured elicitation techniques discussed in Section 2.3. Functional dependence cannot be avoided but it only presents a practical problem in limited circumstances (back calculation with MCS). Variable repetition can sometimes be avoided by restructuring the risk function or model so that variables are not repeated (Ferson, 1996a).

The other sources of dependence identified in Table 3.4 can be treated in one of four ways:

- ignore it;
- envelope it such that the results of the risk assessment are guaranteed to lie between upper and lower bounds irrespective of the dependence;
- model or infer the joint distribution that describes the dependence using copulas; or,
- factorise the joint distribution that describes the dependence into a series of conditionally independent distributions and then model or infer the parameters of these distributions using Bayesian statistical methods.

### **Ignore it**

Smith et al. (1992) suggest that dependency between random variables can be safely ignored when the correlations between the variables are weak, when the variables have only a small influence on the result of the risk assessment, and when the variables are "relatively well known" - i.e. have a small<sup>9</sup> standard deviation. Ignoring dependence between risk factors and assuming they are independent considerably simplifies risk assessment arithmetic. The expectations of binary operations on real random variables are simple function, and the joint distribution of two or more random variables is simply the product of their marginal distributions (Table 3.5).

---

<sup>9</sup>The meaning of "small" here is context dependent, and in this instance under specified.

**Table 3.4:** Some sources of dependence that may be encountered in risk assessment

Type	Description and examples	Reference
Functional	In a simple risk function such as $R = \frac{AB}{C}$ the risk ( $R$ ) is a function of $A$ and cannot therefore be independent of $A$ . The dependence is implied by the structure of the model	Ferson (1996a)
Repetition	Variables that occur repeatedly in a risk function are (by definition) perfectly dependent and should not be independently sampled	
Common cause	In abnormal operating environments the operating characteristics of components may all be simultaneously altered. For example, the failure rates of safety components in industrial systems may not be independent during a fire or a flood if the common cause of the fire or flood destroys the redundancy built into the safety system	Ferson and Hajas (2004)
Temporal	Serial autocorrelation between observations $(x_1, x_2, \dots, x_n)$ of a random variable $X$ . Commonly encountered in financial and ecological data sets. For example, the vital rates of an organism are commonly correlated within and between years: good years for survival also tend to be good years for reproduction and growth	Morris and Doak (2002)
Spatial	Spatial autocorrelation between observations $(x_1, x_2, \dots, x_n)$ of a random variable $X$ . Many ecological phenomena are structured by forces that have a spatial component. Hence, the similarity (variance) of features at different sites is dependent on the distance between the sites	Legendre and Legendre (1998)
Same source	Observations and measurements errors of different variables taken from the same source may be correlated. For example, measurement and measurement errors of different dioxin congeners taken from the same individual are likely to be positively correlated	Hart et al. (2003)
Biological	Many biological variables are dependent. For example, the body size of an organism at sexual maturity and its fecundity, the body weight of an individual and its food intake rate, the weight of an individual and their skin surface area	Tucker and Ferson (2003)
Elicitation	Information elicited from experts may be dependent for many reasons. For example, they all reflect a single school of thought, they all have seen the same (partial) set of data, they are influenced by a single dominant individual	Bier (2004)
Statistical	Dependence between sampling distributions of parameters of parametric distributions. For example, non-linear dependence between samples of the parameters $\alpha$ and $\beta$ of a Gamma distribution	Frey and Burmaster (1999)

Assuming risk factors are independent, however, can lead to bias and overly precise predictions (Regan et al., 2002b). Moreover, low or even zero correlation coefficients do not imply independence, and dependency can influence tail risks even if correlation is zero (Ferson and Hajagos, 2004).

### Envelope alternatives

The problem of computing the distribution of the sum of two random variables  $X$  and  $Y$ , when nothing is known about the dependency between them has a long history. Frechet (1935) solved the problem for conjunction and disjunction of events  $A$  and  $B$  with probability  $P(A)$  and  $P(B)$ , but it took another 52 years before the more general case was solved for random variables with marginal distributions  $F(x)$  and  $F(y)$ , and for a larger class of operations (Frank et al., 1987). Frank's theoretical solutions, however, are not well suited to calculation via a computer because they involve finding the largest and smallest values over an infinite number of cases<sup>10</sup>.

Williamson and Downs (1990) subsequently operationalised Frank's solution by providing algorithms that compute the bounds in a way that accounts for the discretisation error introduced by encoding continuous distributions with a finite computer representation. Their algorithms allow the convolution of two random variables with unknown dependency for all binary operations such as product, ratio, sum and difference (Table 3.5). Moreover, the solutions are guaranteed to enclose all distributions that could arise as a result of the convolution no matter what dependence there may be between the variables, and for individual binary operations they are point-wise best possible, that is, they cannot be any narrower without excluding distributions that could arise under some dependence between  $X$  and  $Y$  (Ferson and Hajagos, 2006). The algorithms of Williamson and Downs also have a number of other attractive properties:

- they require less computer time and memory than convolutions obtained via Monte Carlo simulation, and are much more convenient than sensitivity studies that explore the effects of different dependence;
- they can be combined in risk assessment calculations with assumptions of independence so that some variables are assumed to be independent, whilst no assumptions are made about other variables; and,
- their “outward bounded” discretisation of a continuous distribution function allows probability boxes to be constructed from precise and imprecise distributions with an arbitrarily small discretisation error. This provides a platform that extends Frank's solutions for precise marginal distributions to imprecise marginal distributions, thereby enabling probability bounds analysis (Section 4).

Applying the Williamson and Downs algorithms in a pair-wise fashion to the marginal distributions of all the random variables in a risk function can result in broad upper and lower bounds for the overall risk estimate (see Section 3.2.2). These bounds can be tightened if the analyst has some knowledge about the sign of the dependency between the variables using the “partially specified dependence” methods described by Ferson and Hajagos (2006). For example, if two variables are known to be positively dependent, then the convoluted distribution function of the sum can be found using algorithms that are similar to Williamson and Downs (Table 3.5).

---

<sup>10</sup>Frank's solutions are for infinite sets and entail infimum (*inf*) and supremum (*sup*) functions. For finite sets the infimum equals the minimum and the supremum equals the maximum

Strategy	Example algorithms	Description and constraints	References
Ignore (marginals)	$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)$	Joint distribution of real random variables $X$ and $Y$ with marginal distribution functions $F_X(x)$ and $F_Y(y)$	Ferson and Hajagos (2004)
Ignore (moments)	$\mu_{z=x+y} = \mu_x \cdot \mu_y, \quad \mu_{z=\frac{x}{y}} = \mu_x \cdot \mu_{\frac{1}{y}}$ $\sigma_{z=x+y}^2 = \sigma_x^2 + \sigma_y^2, \quad \sigma_{z=x \cdot y}^2 = \sigma_x^2 \cdot (\mu_y)^2 + \sigma_y^2 \cdot (\mu_x)^2$	Binary operations on real random variables $X$ and $Y$ with mean $(\mu_x, \mu_y)$ and variance $(\sigma_x^2, \sigma_y^2)$	Welsh et al. (1988) Smith et al. (1992)
Envelope (marginals)	$\begin{aligned} \underline{F}_{X+Y}(z) &= \max_{z=x+y} \{ \max [F_X(x) + F_Y(y) - 1, 0] \} \\ \overline{F}_{X+Y}(z) &= \min_{z=x+y} \{ \min [F_X(x) + F_Y(y), 1] \} \\ \underline{F}_{X-Y}(z) &= \max_{z=x+y} \{ \max [F_X(x) - F_Y(-y), 0] \} \\ \overline{F}_{X-Y}(z) &= 1 + \min_{z=x+y} \{ \min [F_X(x) - F_Y(-y), 0] \} \end{aligned}$	Binary operations on real random variables $X$ and $Y$ with marginal distribution functions $F_X(x), F_Y(y)$ resulting in upper $(\overline{F})$ and lower $(\underline{F})$ bounds	Frank et al. (1987) Williamson and Downs (1990) Ferson and Hajagos (2004) Regan et al. (2004)
Envelope (marginals)	$\begin{aligned} \underline{F}_{X \cdot Y}(z) &= \max_{z=x \cdot y} \{ \max [F_X(x) + F_Y(y) - 1, 0] \} \\ \overline{F}_{X \cdot Y}(z) &= \min_{z=x \cdot y} \{ \min [F_X(x) + F_Y(y), 1] \} \\ \underline{F}_{X \div Y}(z) &= \max_{z=x \cdot y} \{ \max [F_X(x) - F_Y(1/y), 0] \} \\ \overline{F}_{X \div Y}(z) &= 1 + \min_{z=x \cdot y} \{ \min [F_X(x) - F_Y(1/y), 0] \} \end{aligned}$	Binary operations on real positive random variables $X$ and $Y$ with marginal distribution functions $F_X(x), F_Y(y)$ resulting in upper $(\overline{F})$ and lower $(\underline{F})$ bounds	Frank et al. (1987) Williamson and Downs (1990) Ferson and Hajagos (2004) Regan et al. (2004)
Envelope (marginals)	$\begin{aligned} \underline{F}_{X+Y}(z) &= \max_{z=x+y} \{ \max [F_X(x) \cdot F_Y(y)] \} \\ \overline{F}_{X+Y}(z) &= \min_{z=x+y} \{ \min [1 - (1 - F_X(x))(1 - F_Y(y))] \} \end{aligned}$	Sum of two positive quadrant dependent variables with marginal distribution functions $F_X(x), F_Y(y)$ resulting in upper $(\overline{F})$ and lower $(\underline{F})$ bounds	Ferson and Hajagos (2006)

**Table 3.5:** Summary of strategies for dealing with dependency in risk assessment, together with example algorithms and constraints

Strategy	Example algorithms	Description and constraints	References
Model it (moments)	$\sigma_{z=x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y$ $\sigma_{z=x,y}^2 = \sigma_x^2(\mu_y)^2 + \sigma_y^2(\mu_x)^2 + 2\rho\sigma_x\sigma_y\mu_x\mu_y$	Binary operations on real positive random variables $X$ and $Y$ with mean $(\mu_x, \mu_y)$ and variance $(\sigma_x^2, \sigma_y^2)$ and Pearson product moment correlation coefficient $\rho$	Welsh et al. (1988) Smith et al. (1992)
Model it (Normal copula)	$X_i = F_i^{-1}[\Phi(Z_i)], \quad Z_i = \sum_{j=1}^i c_{ij} Y_j,$ $Y_j \sim N(0, 1), \quad CC^T = \Sigma = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,k} \\ \sigma_{k,1} & \sigma_{k,k} \end{bmatrix}$	Generate $k$ correlated random variates $X_{i,k}$ with marginal distribution functions $F_i$ and $\Sigma = k \times k$ positive semi-definite Pearson correlation matrix (via the elements $c_{ij}$ of the lower triangular matrix $C$ obtained numerically by the Cholesky decomposition)	Ferson et al. (2004) Clemen and Reilly (1999)
Model it (Frank copula)	$C[F(x), F(y)] = -\frac{1}{\alpha} \ln \left[ 1 + \frac{(\exp(-\alpha x) - 1)(\exp(-\alpha y) - 1)}{\exp(-\alpha) - 1} \right]$	Bivariate distribution $H(x, y) = C[F(x), F(y)]$ with marginal distribution $F(x)$ and $F(y)$ where $\alpha$ determines the level of dependence between $x$ and $y$ . As $\alpha \rightarrow 0$ , $C(F(x), F(y))$ approaches $F(x)F(y)$ implying independence	Clemen and Reilly (1999) Haas (1999)
Factor it (BN)	$p(\mathbf{X}) = \prod_{k=1}^K p(x_k   pa_k)$	Factorise the joint distribution of a directed acyclic graph with $K$ nodes, where $pa_k$ denotes the set of parents of $x_k$	Lauritzen and Spiegelhalter (1988) Spiegelhalter et al. (1993)
Factor it (BHM)	$p(\mathbf{y}; \mu, \tau) = \prod_{i=1}^n p(y_i   \mu, \tau) p(\mu) p(\tau)$	Two parameter BHM: observations $y_i$ follow a Normal distribution with mean $\mu$ and precision $\sigma^2 = \tau^{-1}$ . Mean assumed to have standard Normal distribution $\mu \sim N(0, 1)$ and precision a Gamma distribution $\tau \sim Ga(2, 1)$ . Observations conditionally independent given $\mu$ and $\tau$ , and $\mu$ and $\tau$ are independent	Ades and Lu (2003) Gilks et al. (1996)

**Table 3.6:** Continuation of the summary of strategies for dealing with dependency in risk assessment

### Model the joint distribution

The dependency between random variates can be modelled and simulated in a number of ways, using either the parameters of their joint distribution functions or via samples drawn from their marginal distribution functions in a fashion that is faithful to the known or assumed dependency between them. Welsh et al. (1988) and Smith et al. (1992) provide algorithms for computing the moments of correlated variables under various convolutions (Table 3.6). Scheuer and Stoller (1962) provide a numerical method for generating correlated normal random variates, whilst Iman and Conover (1982) provide a more general distribution-free method for inducing correlation between variables during a simulation. This method is popular and it has been adopted in various risk assessment software packages such as @Risk (Haas, 1999).

Methods such as Iman and Conover's are accurate, numerically inexpensive and fairly simple to implement, but it is important to recognise that they cannot represent all forms of dependence, and in many cases are limited to (potentially unrealistic) linear or monotonic dependency. For this reason, statisticians often cross-check the value of a correlation coefficient against a scatter plot of the data used to calculate the coefficient in order to protect against the potentially misleading impression that high (or low) correlation coefficients can give (see for example Hutchinson and Lai (1990) or Ferson and Hajagos (2004)). Scatter plots, however, are typically unavailable to analysts performing data-free simulation with risk assessment software. In these circumstances analysts must either accept that they are simulating correlation, not dependency, or use other techniques that are capable of representing more complex forms of dependence that can occur in reality. One such technique entails dependence functions known as copulas.

Copulas are multivariate functions that couple marginal distributions together in ways that can characterise complex non-linear dependencies either via inference or in forward propagative analysis of uncertainty. Nelsen (1999) provides the following definition of a bivariate copula: Consider a pair of random variables  $X$  and  $Y$ , with distribution functions  $F(x) = P(X \leq x)$  and  $F(y) = P(Y \leq y)$  respectively and a joint distribution function  $H(x, y) = P(X \leq x, Y \leq y)$ . To each pair of real numbers  $(x, y)$  we can associate three numbers that all lie on the interval  $[0, 1]$ :  $F(x)$ ,  $F(y)$  and  $H(x, y)$ . Each pair of real numbers leads to a point  $(F(x), F(y))$  in the unit square  $[0, 1] \times [0, 1]$  and this ordered pair in turn corresponds to a value of  $H(x, y)$  in  $[0, 1]$ . The function that assigns the value of the joint distribution to each ordered pair of values from the individual (marginal) distributions is called a copula, typically denoted  $C(F(x), F(y))$ . More generally, for a  $n$  dimensional random vector  $U$  on the unit cube, a copula  $C$  is

$$C(U_1, U_2, \dots, U_n) = P(U_1 \leq u_1, U_2 \leq u_2, \dots, U_n \leq u_n). \quad (3.5)$$

One important result from the theory of copulas is Sklar's Theorem (1959). The theorem states that if  $H$  is an  $n$ -dimensional distribution function with continuous marginal distributions functions  $F_1, F_2, \dots, F_n$ , then there exists a unique  $n$ -copula  $C$  such that for all  $x$  in  $\mathcal{R}^n$

$$H(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)). \quad (3.6)$$

If each  $F_i$  is discrete then  $C$  is unique on  $\text{Range}(F_1) \times \text{Range}(F_2) \times \dots \times \text{Range}(F_n)$ . Sklar's theorem is important because it proves that any  $n$ -dimensional probability distribution can be expressed in terms of its marginals via a copula. Hence there are no limitations to the types of dependency that can be represented with copulas.

Theoretically there are an infinite number of ways to link the ordered values of a multivariate distribution function to its marginal distributions, and hence in theory an infinite number of



copulas. In practise, however, analysts tend to stick to a handful of copulas that have useful mathematical and statistical properties. Common examples in the risk assessment literature include the normal copula, Frank copula (Table 3.6), Gumbel copula, Joe copula and Clayton copula (Jouini and Clemen, 1996; Clemen and Reilly, 1999; Haas, 1999; Li, 2000; Embrechts et al., 2001; Yan, 2007). Figure 3.4 shows contour plots of the density functions of four popular bivariate copulas, with identical marginal distributions and the same dependence measure (in this case Kendall's tau = 0.5). The different patterns of dependence induced by each of the copulas are clearly evident.

One very important property of copulas is that they provide multivariate dependence structures that are separate from their marginal distributions, and since any continuous random variable can be transformed to be uniform over the range  $[0, 1]$  by its probability integral transformation, copulas can be constructed with any continuous marginal distribution. For example, Figure 3.5 shows the contour plots for the same set of copulas in Figure 3.4 but this time with Beta (10, 2) and Lognormal (5, 1) marginals. This example demonstrates the complex patterns of dependency that copulas are capable of representing. Note that this pattern of dependency cannot be mimicked with a linear correlation coefficient.

Two other important results from the theory of copulas are the Frechet-Hoeffding limits for joint distribution functions of random variables and the product copula  $\Pi(u, v) = u \cdot v$ . Frechet (1951) and Hoeffding (1940) showed that for any bivariate copula  $C$  and for all  $u, v$  on the unit square:

$$W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) = M(u, v). \quad (3.7)$$

Hence by Sklar's theorem if  $X$  and  $Y$  are random variates with a joint distribution  $H(x, y)$  and marginal distributions  $F(x)$  and  $F(y)$ , then for all  $(x, y)$

$$\max(F(x) + F(y) - 1, 0) \leq H(x, y) \leq \min(F(x), F(y)). \quad (3.8)$$

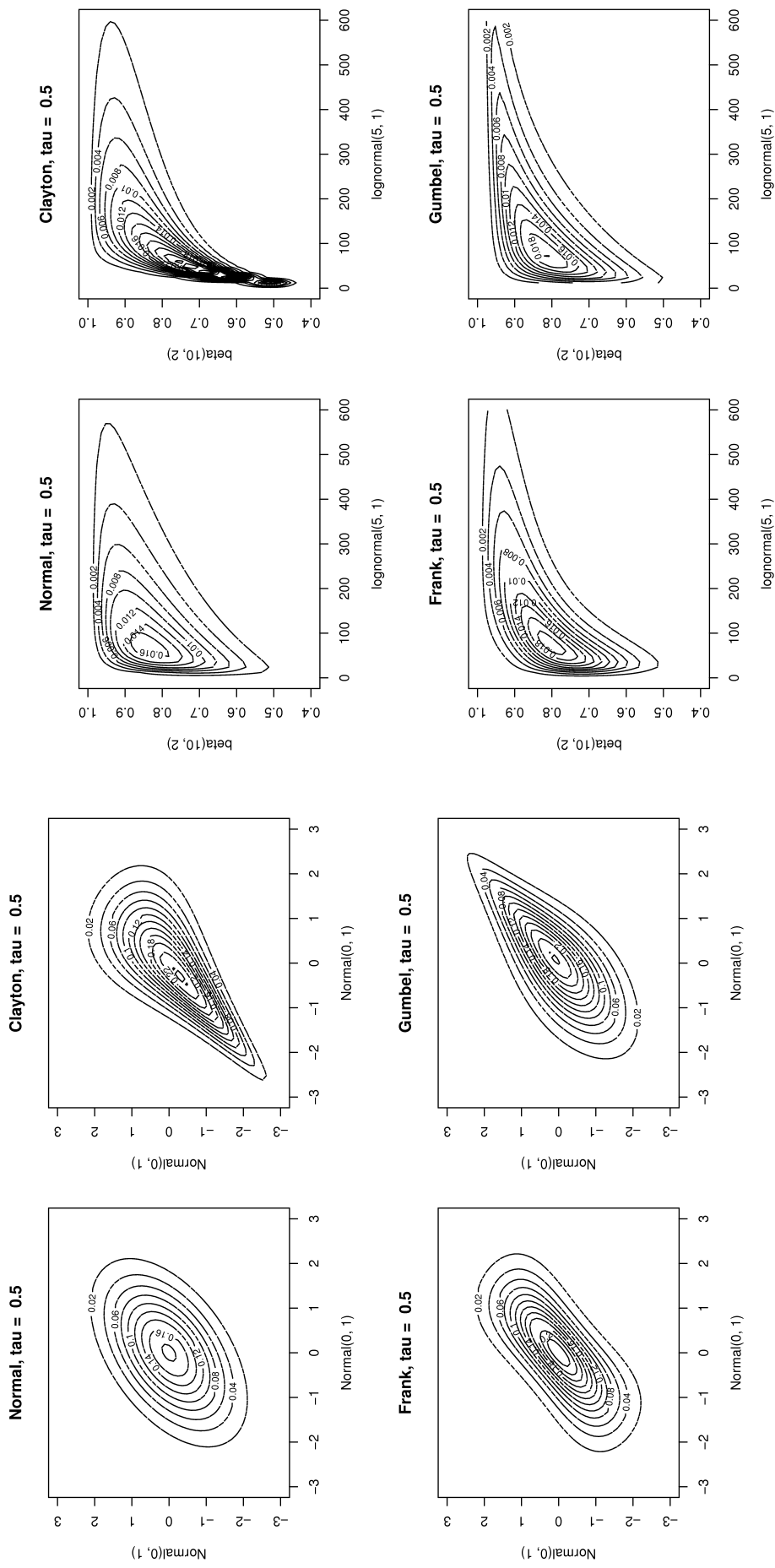
The Frechet-Hoeffding limits are important because the lower bound  $W^n$  is smaller than every  $n$ -copula, and the upper bound  $M^n$  is larger than every  $n$ -copula, and hence via Sklar's theorem the value of all distribution functions must lie within these bounds. The Frechet-Hoeffding limits therefore bound the joint distribution function of  $n$  random variables irrespective of the dependency between the variables. Williamson and Down's algorithms make use of these limits to calculate bounds on convolutions of random variables without having to make any assumptions about the nature of the dependency between them (Table 3.5).

The product copula is important because, again via Sklar's theorem, two random variables  $X$  and  $Y$  are independent if and only if their joint distribution has the product copula  $C(F(x), F(y)) = F(x) \cdot F(y)$ .

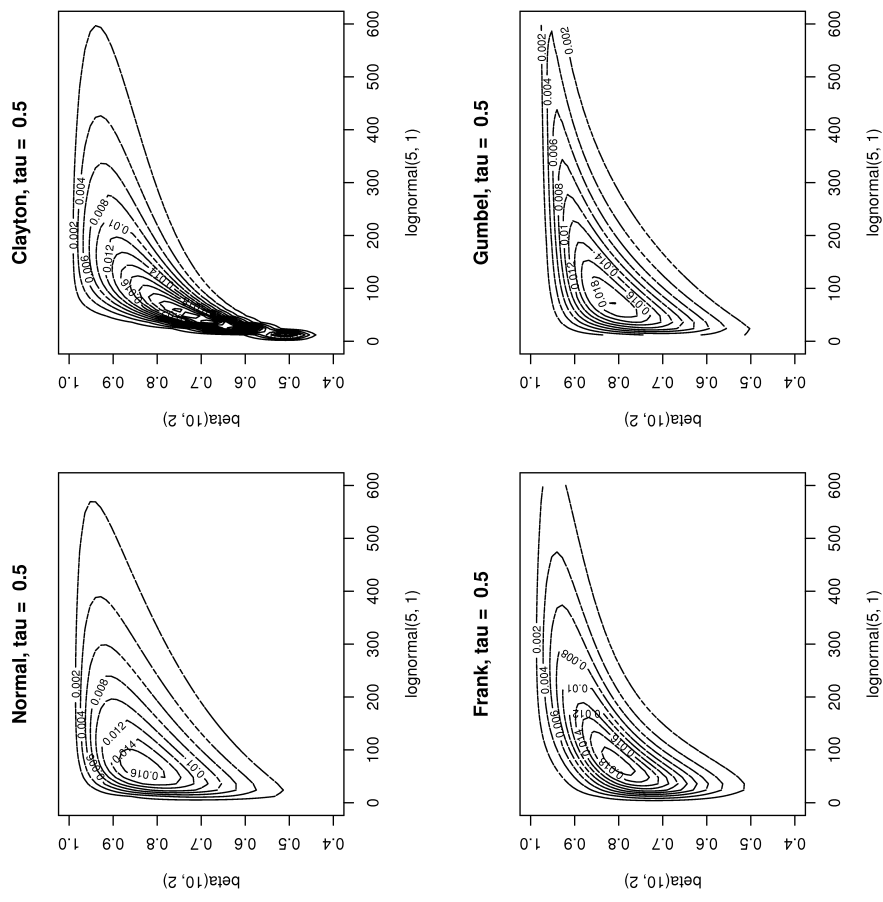
Using copulas to model dependency requires an appropriate metric of dependence, typically transformed versions of Spearman's rank correlation coefficient or Kendall's tau,<sup>11</sup> and an appropriate copula. Choosing an appropriate copula, however, invokes model structural uncertainty. The true data generating mechanism for a bivariate or multivariate data set is usually unknown and several candidate copulas may fit the data equally well.

---

<sup>11</sup>Most copulas are parameterised by a dependency parameter that in many cases is a simple transformation of a rank-order correlation coefficient such as Spearman's rank correlation coefficient or Kendall's tau



**Figure 3.4:** Probability density surfaces of bivariate distributions constructed via four copulas. All copulas are shown with the same standard normal marginals and the same Kendall's Tau correlation coefficient to highlight the effect of the different type of copula on the bivariate probability surface



Empirical copulas and log-likelihood methods for fitting parametric copulas are available and statisticians typically recommend fitting a number of candidate copulas to the data and choosing among ones with the highest likelihood (Yan, 2007; Nelsen, 1999). The list of appropriate copulas may be constrained by their inherent properties,<sup>12</sup> but in many cases the analyst will still be faced with a potentially large number of candidates. Like all model selection problems, the particular choice of copula can have important implications for the results of the risk assessment, particularly because different copulas assign different levels of dependency to the upper and lower tails of bivariate distributions - an effect known as “tail dependency” (Haas, 1999; Embrechts et al., 2001). This effect is clearly evident in the Clayton and Gumbel copulas illustrated in Figure 3.5.

The main advantage of copulas is that they provide a direct and relatively simple way to model joint dependency. A risk function of  $n$  random variables can be directly modelled by assigning marginal distributions to each of the variables, estimating a rank correlation coefficient between each variable and choosing an appropriate copula  $C^n$ . Copulas therefore provide an alternative to the potentially complex calculations that are needed to express a joint distribution as the product of conditional independent distributions (see below).

Another important advantage of copulas is that they allow the analyst to propagate the effects of tail dependence through a risk assessment. This can be important if the parameters of a problem are independent under “normal” conditions (e.g the process is operating around the mean or the mode of the parameters concerned) but become dependent at the extreme ranges of a parameter. It is important to note, however, that this type of behaviour may not be evident in data sets collected under “normal” conditions and may therefore be a potential source of completeness or scenario uncertainty, treatments for which are discussed in Section 3.

Copulas can also be used with probability boxes and Dempster-Shafer structures to convolve imprecisely known random variables using the “diagonal difference” method (Ferson and Hajagos, 2004). In propagative uncertainty analysis, however, implementing a copula-based approach requires a reliable estimate of the rank correlation matrix. This is an inherently difficult task. Clemen and Reilly (1999) identify methods to elicit rank correlations from experts, but these methods are rarely used in ecological risk assessment, and in practice copulas are mainly used in the inferential mode of uncertainty analysis.

### Factor the joint distribution

If risk factors are treated as random variables, then risk is a function of their joint density. An alternative way to model, simulate and perform inference with a multivariate density is to factor it into a series of simpler (usually univariate), conditionally independent densities. At the heart of this approach are the two fundamental rules of probability (Bishop, 2006):

$$p(X) = \sum_Y p(X, Y) \quad (\text{Sum rule})$$

$$p(X, Y) = p(Y|X)p(X), \quad (\text{Product rule})$$

where  $p(X, Y)$  is the joint distribution of the random variables  $X$  and  $Y$ , and  $p(X|Y)$  is the conditional distribution of  $X$  given  $Y$ .

---

<sup>12</sup>Some types of copula can only model positive dependency (Archimedean copulas of dimension 3 or higher), and some are non-comprehensive - i.e. cannot represent the full  $[-1, +1]$  range of Kendall's tau or Spearman's rank correlation coefficient

From the product rule, together with the symmetry property  $p(X, Y) = p(Y, X)$  the following relationship known as Bayes theorem immediately follows

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}, \quad (\text{Bayes rule})$$

where the denominator can be expressed, via the sum rule as

$$p(X) = \sum_Y p(X|Y)p(Y). \quad (3.9)$$

The product rule holds for cases where  $X$  and  $Y$  are not independent. Repeated application of this rule allows high-dimension multivariate probability density functions to be factored into products of simpler, conditionally-independent, density functions that can be sampled one (or more) at a time. For example, an arbitrary joint density function  $p(X, Y, Z)$  can, by application of the product rule, be written as

$$p(X, Y, Z) = p(X|Y, Z)p(Y, Z). \quad (3.10)$$

A second application of the product rule to the second term on the right hand side of 3.10 gives

$$p(X, Y, Z) = p(X|Y, Z)p(Y|Z)p(Z). \quad (3.11)$$

Hence the potentially complex joint density has been factored into a product of simpler independent density functions. This factorisation holds for any choice of joint density.

Some of the simplest examples of these conditional independence relationships are found in the conditional probability tables that underlie Bayesian Networks. In this context the joint probability density of the network is factored into the product of a set of conditional densities, one for each node of the network, but conditioned only on the parents of that node (Section 4). Bayesian networks allow variability to be treated in a forward propagative mode that (typically) relies on experts to complete the conditional probability tables associated with each node. Factorisation of joint probability density functions, however, is a much more general strategy and is widely applied in an inferential model of uncertainty analysis. Clark (2003) provides a relatively simple example that illustrates the process well. The model in this case is a statistical model of the fecundity of an endangered species, where the data  $y_i$  are observations of the number of offspring produced by  $i = 1, \dots, n$  breeding pairs. The fecundity of the species is described by a Poisson distribution with rate parameter  $\lambda_i$ . The likelihood for a data set of  $n$  independent identically distributed observations is

$$p(y|\lambda) = \prod_{i=1}^n \text{Poisson}(y_i|\lambda_i). \quad (3.12)$$

In this model the fecundity of the species is allowed to vary between breeding pairs - i.e. not all breeding pairs have precisely the same rate parameter, some are more, or less, successful than others. The variability between breeding pairs is observed to be greater than the variance of the Poisson distribution, hence fecundity it is described by a Gamma distribution

$$p(\lambda|\alpha, \beta) = \text{Gamma}(\lambda_i|\alpha, \beta), \quad (3.13)$$

with parameters  $\alpha$  and  $\beta$  that may be considered constant or assigned distributions themselves. Note here that the probability of  $\lambda_i$  is written as a conditional probability to emphasise that

it depends on the parameters  $\alpha$  and  $\beta$ . The joint (posterior) density of the model parameters conditioned on the observations is factored into a set of univariate, conditionally independent, density using Bayes' rule (omitting the normalising term in the denominator) and the product rule

$$p(\lambda, \alpha, \beta | y_{1:n}) \propto \overbrace{\prod_{i=1}^n \text{Poisson}(y_i | \lambda_i) \prod_{i=1}^n \text{Gamma}(\lambda_i | \alpha, \beta)}^{\text{Conjugate pair}} p(\beta) p(\alpha). \quad (3.14)$$

Each of the density functions on the right hand side of 3.14 can be sampled either one at a time or as a pair. The first two densities on the right hand side of the equation form a conjugate-pair, meaning that their product produces (after some algebra) another standard univariate distribution which is easily sampled. Typically in models of this type the normalising constant omitted from 3.14 has no analytical solution. In this case the posterior density is numerically derived using Markov Chain Monte Carlo (MCMC) methods. These methods include Gibbs sampling where conjugacy allows direct sampling from standard distributions or the Metropolis-Hastings algorithm where it is not possible to directly sample for one of the parameters in the model because its probability density is the product of two densities that are not conjugate.

The chief advantage of models such as 3.14 is that they allow inference over many sources of real-world variability, including variability between individuals in time and space (Link et al., 2002; Clark and Bjornstad, 2004). These types of models also allow the analyst to incorporate observation models between the data and the process model, (caricature E, Figure 3.2). This is an important and exciting development in statistical science. These hierarchical models provide a potentially more accurate distinction between the adequacy of alternative models by separating the variability in a data set that is not explained by a model (process noise) from the observation error in the data set and the population-level variability of input variables (Clark, 2003, 2007).

There are a range of frequentist and Bayesian inferential methods for models that distinguish process noise and observation error. Some of the earliest, the Kalman Filter and Error in Variables method (Schnute, 1994), assume linear process models and normally distributed error variables. These approaches have been subsequently extended to non-linear models and non-Gaussian error distributions using spline approximations (Kitagawa, 1987) or advanced Sequential Monte Carlo methods and adaptive MCMC methods (Peters et al., 2010). Again the methodology here relies on MCMC methods to approximate the analytically intractable integrals associated with these distributions.

MCMC calculations are sometimes difficult and computationally intensive (Cressie et al., 2007). The potential for more precise predictions of process error, however, has encouraged statisticians and modelers to use hierarchical models, and examples relevant to biosecurity, including disease surveillance and the spread of invasive species, are now available in the literature (Ranta et al., 2005; Hooten and Wikle, 2008). Interestingly, many of applications of these methods cross the traditional divide between statistical and mechanistic models by transforming differential equations to difference equations and treating the problems of modeling fitting in the presence of observation error as a regression problem (Clark, 2007).

## 4 Uncertainty analysis methods

Methods for forward propagative uncertainty analysis have been mentioned on various occasions throughout the previous sections of the report. This section provides a systematic summary of these methods. In each case the summary is necessarily brief. The objective here is to provide the reader with a basic understanding of the method, and their advantages and disadvantages.

### 4.1 Analytical methods

#### The Delta method

The Delta method is a technique for finding approximations to the moments (particularly the variance) of functions of random variables when these moments cannot be directly evaluated (Oehlert, 1992). The technique also goes by a variety of other names such as the method of moments or the mean-value first-order second-moment method (Morgan and Henrion, 1990; Kuo et al., 2007). The Delta method works by replacing a function  $h$  of  $n$  random variables  $h(\mathbf{x}) = h(x_1, x_2, \dots, x_n)$  by its Taylor series in powers  $(x_i - \bar{x}_i)$

$$\begin{aligned} h(\mathbf{x}) = h(\bar{\mathbf{x}}) &+ \sum_{i=1}^n (x_i - \bar{x}_i) \left[ \frac{\partial h}{\partial x_i} \right]_{x=\bar{x}} \\ &+ \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}_i)(x_j - \bar{x}_j) \left[ \frac{\partial^2 h}{\partial x_i \partial x_j} \right]_{x=\bar{x}} + \text{HOT}, \end{aligned} \quad (4.1)$$

where  $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$  are the means of the random variables  $x_i$  and HOT represents the higher order terms in the Taylor series expansion. For many practical applications the higher order terms and cross product terms in Equation 4.1 are ignored leaving what is known as the first order approximation of the function  $h(\mathbf{x})$

$$h(\mathbf{x}) \approx h(\bar{\mathbf{x}}) + \sum_{i=1}^n (x_i - \bar{x}_i) \left[ \frac{\partial h}{\partial x_i} \right]_{x=\bar{x}}. \quad (4.2)$$

Taking the moments about the origin and about the mean of Equation 4.2 provides a first-order approximation for the mean and variance

$$E[h(\mathbf{x})] \approx h(\bar{\mathbf{x}}) \quad (4.3)$$

$$\begin{aligned} \text{Var}[h(\mathbf{x})] &\approx \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(x_i, x_j) \left[ \frac{\partial h}{\partial x_i} \right]_{x=\bar{x}} \left[ \frac{\partial h}{\partial x_j} \right]_{x=\bar{x}} \\ &\approx \sum_{i=1}^n \text{Var}(x_i) \left[ \frac{\partial h}{\partial x_i} \right]_{x=\bar{x}}^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(x_i, x_j) \left[ \frac{\partial h}{\partial x_i} \right]_{x=\bar{x}} \left[ \frac{\partial h}{\partial x_j} \right]_{x=\bar{x}}. \end{aligned} \quad (4.4)$$

If we know or assume that the inputs  $(x_i, x_j)$  are independent, then the second term in Equation 4.4 is zero and we recover the predictive method-of-moments formula for the variance of functions of random variables in Table 3.5. Alternatively, if we assume or know the covariance between the inputs then Equation 4.4 provides the basis for the variance formulas in the first row of Table 3.6. This should be apparent by recalling the definition of the Pearson product moment correlation coefficient:

$$\rho_{x,y} = \frac{\text{Cov}(x,y)}{\sigma_x \cdot \sigma_y}.$$

Equations 4.3 and 4.4 are exact in some simple cases such as the sum or difference of independent random variables, or more generally where the function  $h$  is linear in its input variables. These exact cases can also be extended to the products and ratios of powers of uncertain variables  $h(\mathbf{x}) = \prod_{i=1}^n x_i^{a_i}$ . Taking a log transformation reduces this function to a simple (linear) weighted sum of the log of the uncertain inputs  $\ln[h(\mathbf{x})] = \sum_{i=1}^n a_i \ln(x_i)$ , resulting in the following exact<sup>13</sup> formulas for the mean and variance

$$E[\ln(h(\mathbf{x}))] = \sum_{i=1}^n a_i E[\ln(x_i)] \quad (4.5)$$

$$\text{Var}[\ln(h(\mathbf{x}))] = \sum_{i=1}^n a_i^2 \text{Var}[\ln(x_i)] + 2 \sum_{i=1}^n \sum_{j=i+1}^n a_i a_j \text{Cov}[\ln(x_i), \ln(x_j)]. \quad (4.6)$$

Risk functions that involve products and ratios of powers are quite common, particularly in ecotoxicology (Table 2.2) and import risk assessment. Indeed the case study developed during this project is a good example (see Figure 1.1). Slob (1994) discusses uncertainty analysis in this context further, highlighting the additional results that follow if the risk factors  $\mathbf{X}$  are independent and identically (log-normally) distributed.

Nowadays, the Delta method is not widely used in ecological risk assessment (Section 5) although it was historically more popular in engineering contexts. The main advantages of this method are that the numerical calculations are simple and easy to implement once the algebraic analysis has been completed. Some authors also argue that the method is clear and intuitive, generally decomposing the variance of the output into the sum of the contributions from each input (Morgan and Henrion, 1990). There are, however, several disadvantages to this approach:

- the complexity of the algebra increases rapidly with more complex models, due to the large number of interactions in the covariance terms. This problem is exacerbated if the cross-product term or any of the higher order terms in Equation 4.1 are retained;
- the propagation of uncertainty is restricted to the parameters of the distribution, most often the mean and variance. The method says nothing about the tails of the resulting distribution. This issue can only be addressed in very limited cases (through the central limit theorem) or by making very strong assumptions about the distributions of the uncertain risk factors, such as independent and identically normally distributed (Slob, 1994); and,
- the method is not accurate if the risk function is not smooth (not continuously differentiable) or if important covariance terms are omitted.

#### Other analytical methods

Kuo et al. (2007) and Yu et al. (2001) cite two other (approximate) analytical uncertainty analysis methods: Rosenblueth's Point Estimation Method (RPEM) and Harr's Point Estimation Method (HPEM). These methods, however, do not appear to be widely known or applied. A Web of Science search returned only 5 and 4 hits respectively. Yu et al. (2001) also reports markedly different results between Monte Carlo Simulations and RPEM and HPEM, and poor performance with non-linear risk functions.

<sup>13</sup>The formulas are exact so long as the geometric means and geometric covariance matrix for  $\mathbf{X}$  are finite

## 4.2 Probabilistic methods

### 4.2.1 Monte Carlo simulation

#### First order Monte Carlo simulation

Monte Carlo Simulation (MCS) belongs to a class of computational algorithms known generally as Monte Carlo methods. The class of Monte Carlo methods is large and varied but all of its algorithms rely on randomly generating samples from a defined input domain, such as a set of probability density functions, to solve problems that do not have an analytical solution. Monte Carlo methods have a long and successful history, that dates back to at least the turn of the 20<sup>th</sup> Century, although the term “Monte Carlo” was only coined in the 1940’s by John von Neumann as a code word for secret work on the diffusion of neutrons (Burgman, 2005).

In risk assessment circles, Monte Carlo Simulation is used to propagate the effects of variable risk factors through  $n$ -dimensional risk functions in cases where the complexity of the risk function and/or the distribution of the individual risk factors, preclude an analytical solution. In this context MCS can be considered the mainstay of probabilistic quantitative risk assessment (Rasmussen, 1981; Vose, 2000). Monte Carlo methods, however, have a much broader application than quantitative risk assessment and are used widely in a range of statistical problems to solve analytically intractable optimisation and integration problems, such as finding the maximum of a multi-modal likelihood function, or estimating the normalising constant (an integral) in a Bayesian Hierarchical Model such as Equation 3.14 (Robert and Casella, 1999)

Monte Carlo methods work by using simulation methods to approximate integrals of functions of random variables

$$\int_{\chi} h(x)p(x)dx, \quad (4.7)$$

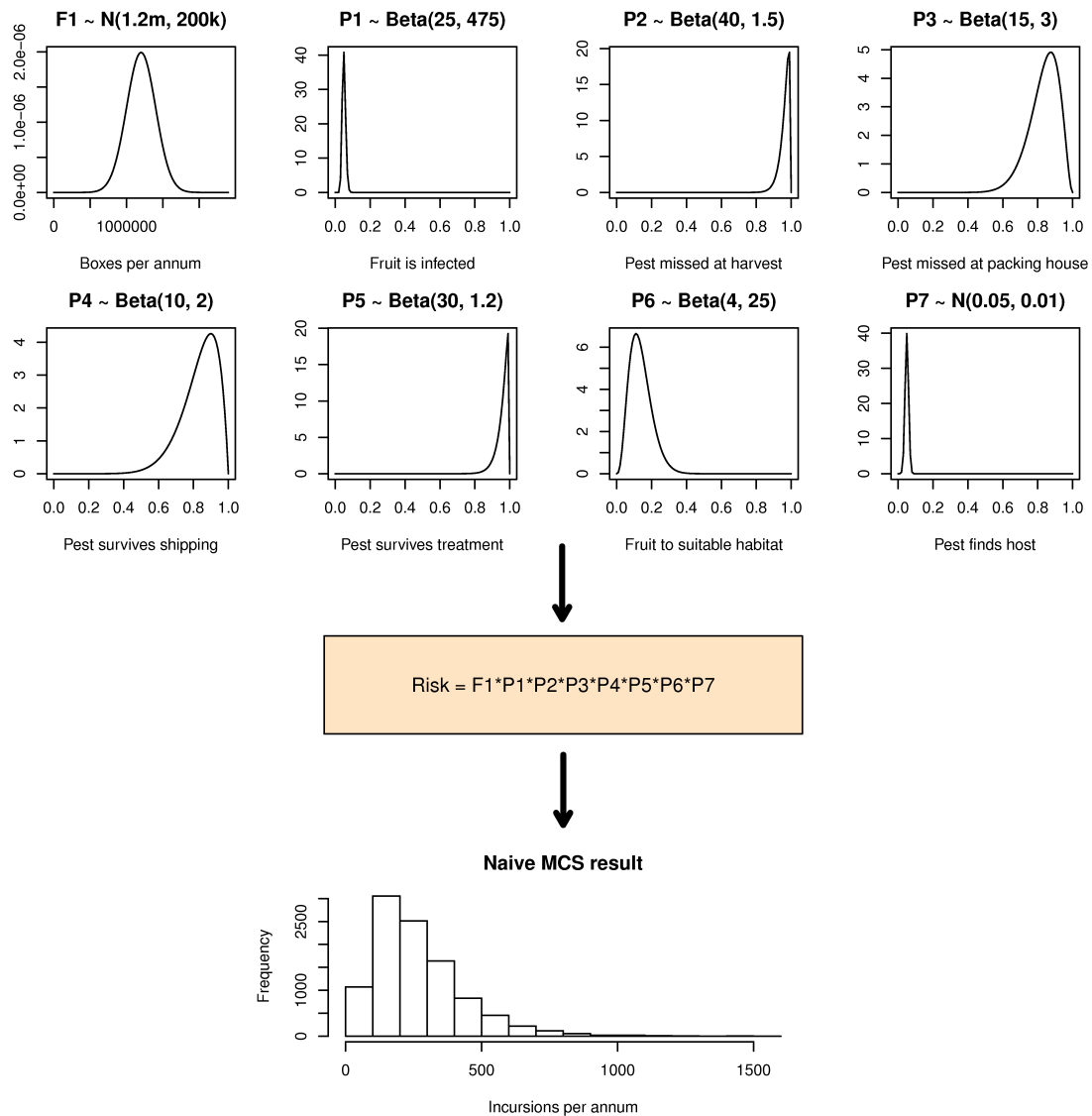
where  $p(x)$  is a probability density,  $h(x)$  is a function (in our context a risk function) of the random variable  $X$  and  $\chi$  denotes the set where this random variable takes its values (usually equal to the support of the density  $p()$ ) (Robert and Casella, 2010). Monte carlo integration approximates Equation 4.7 by generating a sample  $(X_1, \dots, X_n)$  from the density  $p(x)$  with the empirical average

$$\bar{h}_n = \frac{1}{n} \sum_{j=1}^n h(x_j). \quad (4.8)$$

Importantly, the error associated with this approximation decreases by a factor of  $1/\sqrt{n}$  irrespective of the dimensions of the function  $h(x)$ .

Monte carlo simulation treats the values  $h(x_j)$  as possible outcomes of the (risk) function  $h()$  and collates these into an empirical distribution function. Figure 4.1 provides a schematic representation of a first order Monte Carlo Simulation in a typical import risk assessment context. In this example the risk function  $h(\mathbf{X})$  is a simple product of 8 input parameters, whose variability is represented by standard density functions - i.e density functions that can be sampled using the standard inverse distribution method (Vose, 2000). The MCS approximates the probability density of the risk function by taking (usually independent) samples from each of the variable inputs, applies the risk function deterministically to each of the  $N$  sets of 8 samples, and collates the result in a histogram or cumulative distribution function.





**Figure 4.1:** Simple schematic of a first order Monte Carlo Simulation for an import risk assessment model. The model is run many times. On each occasion a variate is randomly selected from the parametric distributions representing variability in the model parameters (here assumed to be independent). The results of the all model runs are typically collated in a histogram showing the variability of the overall risk estimate.

In practical applications the probability density functions associated with each of the risk factors may represent several sources of uncertainty, including variability and some forms of epistemic uncertainty such as subjective judgement, measurement error and sampling uncertainty. They may therefore be derived from repeated measurements of a process, theoretical arguments and/or elicitation of an expert's beliefs. The various ways to choose a distribution from among the large number of standard parametric probability density functions are discussed in Section 3.2.8. Alternatively empirical distribution functions may be used, but again these will generally require hundreds of observations in order to accurately characterise the full range of variability.

The schematic in Figure 4.1 is an example of a *first-order* Monte Carlo Simulation - i.e. it ignores uncertainty in the parameters of the distribution functions. Many authors emphasise the need to separate parametric variability - that is irreducible variability in the input parameters of risk functions - from theoretically reducible epistemic uncertainty about this variability (Hoffman and Hammonds, 1994; Hattis and Burmaster, 1994; Clark, 2007). As discussed in Section 3.2, uncertainty about the variability of an input parameter (risk factors) can occur for a number of reasons, principally through sampling uncertainty but it can also reflect differences of opinion between experts, scenario uncertainty and model structure uncertainty. The effect of this uncertainty is usually propagated through Monte Carlo simulation by averaging over it in a second order simulation.

### **Second order Monte Carlo simulation**

Second order MCS recognises that the parameters of the probability distribution functions that describe the variability of risk factors may not be precisely known. In second order MCS this uncertainty is propagated through the simulation by creating a probability density or interval to describe the epistemic uncertainty in the parameters of the input distribution. This density or interval is sampled in an "outer" simulation to obtain an estimate of the parameters that are subsequently used to create density functions for the input parameters that are sampled in the "inner" simulation in the usual fashion. This procedure is sometimes referred to as "two-dimensional" MCS to reflect the two simulations (the inner and outer) that occur. The various parametric and non-parametric methods for constructing confidence intervals around sample statistics, that can be subsequently sampled in the outer simulation of a second order MCS, are discussed in Section 3.2.7. Modern examples of this approach to uncertainty analysis can be found in Wu and Tang (2004), Pouillot et al. (2007) and Vicari et al. (2007).

The main advantage of Monte Carlo simulation is that the approximation error is not influenced by the dimensions of the function  $h(x)$ . The accuracy of the simulation depends on the number of samples  $N$  and this is not influenced by the number of input variables in the risk assessment problem. Monte Carlo integration is therefore resilient to the "curse of dimensionality" which is why it is widely used in complex statistical inference problems. MCS is also immune to the problem of repeated uncertain parameters<sup>14</sup> so long as the simulation is correctly programmed. Moreover, Monte Carlo simulation is supported by a large number of user-friendly software packages such as @Risk (Vose, 2000) that can accommodate the influence of linear correlation between uncertain input parameters where this is known or assumed.

There are also a number of important problems with Monte Carlo Simulation (Ferson and Long, 1995; Ferson, 1996b; Ferson and Ginzburg, 1996) chief amongst which are: a) its high information requirements that must be met with either data or assumptions; b) the potential for non-

<sup>14</sup>Repeated uncertain parameters artificially inflate uncertainty estimates with interval-based methods of uncertainty propagation. It is discussed further in Section 4.4.1

linear dependency between input variables and the parameters of their density functions; and, c) the unwarranted precision that occurs when precise probability density functions are used to propagate epistemic uncertainty through risk functions. Ferson and Ginzburg (1996) summarises this last problem well. They demonstrate how a Monte Carlo simulation of the product of two random variables with marginal uniform distributions, chosen to represent two uncertain parameters on the grounds of Laplace's principle of insufficient reason, results in a distribution with a marked central tendency that is not justified by the original information regarding the input parameters. Moreover they show that this result is not caused by (and is therefore additional to) the problems of linear or non-linear correlation between the parameters.

The concerns and criticisms directed at MCS by Ferson's publications are well warranted, particularly where precise density functions are used in an uncritical fashion to characterise the many forms of variability and epistemic uncertainty in the input variables of a risk function. Baudrit et al. (2007a) identify three types of variables in risk assessment problems: random variables observed with total precision, imprecisely observed random variables and deterministic parameters whose value is imprecisely known. These different quantities reflect at least four potential sources of uncertainty: variability, sampling uncertainty, subjective judgement and measurement error. First order MCS simply amalgamates all of these quantities and their various sources of uncertainty into a single density function. This creates a source of potential confusion and error (Nauta, 2000). Cullen and Frey (1999), for example, list seven different ways to interpret the probability density functions that are used in a MCS:

- they represent subjective judgement due to inapplicable, inappropriate or otherwise unavailable data sets. In this case uncertainty about the appropriateness of a data set exists that may not be characterised by a statistical analysis;
- they represent epistemic uncertainty only. This is the most usual approach with no regard to the distinction between epistemic uncertainty and variability;
- they only represent variability. This assumes that probability densities for variability are precisely known;
- they represent variability and parametric uncertainty. This approach recognises that for small sample sizes the parameters of a density function are uncertain, and characterises this uncertainty using (for example) the sampling distribution of the observed statistic. Uncertainty propagation requires second-order simulation and recognition of the potential for statistical dependency between the parameters of the sampling distribution (Table 3.4);
- they correct for the observed bias (overconfidence) in variability projections derived from measurement error and subjective judgement<sup>15</sup> (Shlyakhter, 1992, 1994)
- they represent systematic error, measurement error and variability. Measurements made with imperfect measuring devices will include variation due to the underlying population variability and the random and/or systematic errors introduced by the instrument; and,
- they represent a generalised case of variability and epistemic uncertainty. This is arguably the most realistic case. Data usually represents a limited sample of observations made with an imprecise and sometimes biased instrument. This creates uncertainty about the possible bias and variance of the measuring device (observation error) and parametric uncertainty about the probability density functions used to represent population variability.

---

<sup>15</sup>Interestingly I am unaware of any actual application of this potentially very powerful result

ity. This generalised case can be complicated by (non-linear) dependency between the various sources of error. For example, measurement errors may be proportional to the quantity of the value being measured (a source of heteroscedasticity).

The potential for complex non-linear dependency between the sources of variability in real data sets is another important weakness in Monte Carlo simulation. Techniques for propagating linear dependency through an analysis when drawing samples in a first- or second-order MCS are available (Section 3.3). Non-linear dependency can be modelled using copulas but examples in ecological risk assessment are very rare. Rather the potential for dependency (linear and non-linear) is routinely ignored in MCS application (Section 5.1). Complex non-linear dependency can be addressed using the hierarchical Bayesian inference techniques discussed in Section 3.3, particularly approaches that explicitly incorporate an observation model between the data and the real-world process of interest (caricature E in Figure 3.2). The complexity of Cullen and Frey's general case is the primary motivation for these types of models (Clark, 2007).

Dependency between the different sources of variability will tend to have the most prominent effect on the tails of the risk estimates - i.e. in the low probability, high consequence events. The effects of linear dependence can be easily investigated using an approach known as "dispersive MCS" (Ferson, 1996b; Ferson and Hajagos, 2004). This approach imposes maximum linear correlation between the samples drawn from the input parameters of the risk function in MCS. Ferson and Hajagos (2004) provide a simple Excel-based example of this method.

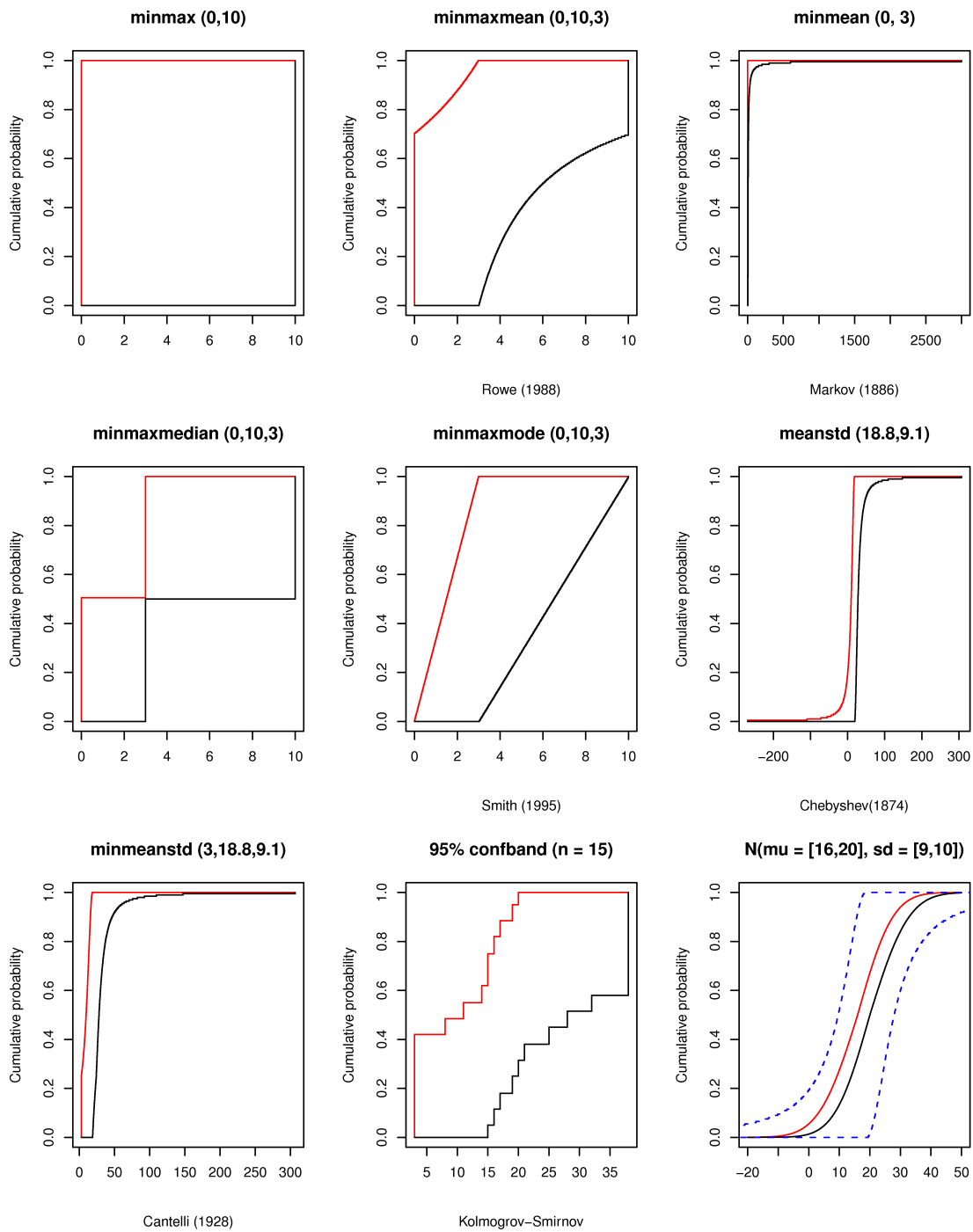
#### 4.2.2 Probability boxes and probability bounds analysis

The term probability bounds analysis was coined by Ferson (2002) to describe a collection of methods and algorithms that allow risk assessment calculations to be performed with probability boxes, or p-boxes. These methods include the important case of dependency bounds analysis, developed by Williamson (1989) and Williamson and Downs (1990) to calculate the upper and lower bounds on the distribution of a function of random variables when only the marginal distributions of the variables are known (Table 3.5). 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) represents the class of distribution functions  $F(x)$  specified by a pair of left (upper) and right (lower) distribution functions  $\bar{F}(x)$  and  $\underline{F}(x)$  such that  $\underline{F}(x) \leq F(x) \leq \bar{F}(x)$  for all  $x$ , and, optionally, additionally specified by bounds on the parameters or type of distribution (Ferson and Hajagos, 2004). 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 parameters or type of a variable's distribution function. If these 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 location, scale or shape parameters are assumed to be precisely correct. Ferson (2002) and Ferson et al. (2003) provide comprehensive descriptions of the different ways to construct p-boxes based on the various forms of (usually limited) data that are available to risk analysts. Figure 4.2, for example, shows probability boxes constructed using nine different methods:

- **{min, max}**: this is the one of the simplest p-boxes, suitable for situations where all that is known is that a random variable cannot be smaller than  $a$ , and cannot be larger than  $b$ . In the example shown here the limits are  $[0, 10]$ ;

- **{min, max, mean}**: if the limits and mean of the variable are known then the p-box can be tightened by a mathematical constraint known as Rowe’s range-mean inequality. In the example shown in Figure 4.2 the limits are  $[0, 10]$  and mean is 3. The upper and lower limits of the p-box are reduced to satisfy the requirement that the probability mass to the left of the mean is balanced by the mass to its right;
- **{min, mean}**: knowing only one endpoint of the variables range (i.e. its upper or lower limit) and its mean leads to a very wide p-box constructed via the Markov inequality. In the example shown in Figure 4.2, the lower limit is known to be 0, and the mean is known to be 3, but the upper limit is unknown;
- **{min, max, median}**: knowledge of the median, as well as the limits, of a variable, “pinches” the p-box to the definite point at the 50<sup>th</sup> quantile. If the value of the median is uncertain then the pinching is less severe. In the example shown in Figure 4.2 the limits are again  $[0, 10]$  and the median is 3;
- **{min, max, mode}**: if the shape of the variable’s distribution function is known to be unimodal, and its range and mode are known, the upper and lower limits of the p-box can also be substantially narrowed over the {min,max} case. The {min,max,mode} example in Figure 4.2 is for a variable on the range  $[0, 10]$  with a mode at 3;
- **{mean, std}**: the Chebyshev inequality is used to construct a p-box where the mean and standard deviation (or variance) of a variable is known but the shape of its distribution function is not. The example in Figure 4.2 shows the p-box for a sample of observations whose mean is 18.8 and whose standard deviation is 9.1. No other assumptions about the endpoints or shape of the distribution function are made;
- **{min, mean, std}**: if one of the endpoints of the variables range is known, together with the mean and standard deviation, then the p-box can be substantially constrained via the Cantelli inequality (Benzi et al., 2007). The example shown in Figure 4.2 shows the effect of knowing the mean (18.8), standard deviation (9.1) and the minimum value (3) of a variable. This approach can be generalised to the situation where both endpoints are known;
- **confidence intervals**: there are a number of ways to construct p-boxes from the confidence intervals of a sample statistic or empirical distribution function. These p-boxes account for sampling uncertainty caused by small sample sizes, and include the extreme case where the sample consists of a single observation (Wall et al., 2001). Figure 4.2 shows an imprecise empirical distribution function constructed by using the Kolmogorov-Smirnov statistic to create 95% confidence limits for all values of the hypothetical sample  $\mathbf{X} = (17, 11, 14, 38, 15, 3, 15, 16, 20, 25, 21, 28, 8, 32, 19)$  (Sokal and Rohlf, 1995);
- **imprecise parametric distribution function**: the final example in Figure 4.2 shows a p-box that represents an imprecise parametric distribution function. Here the shape of the distribution function is assumed to be known (e.g. normally distributed) but the analyst is uncertain about the parameters of the distribution - the mean is thought to lie on the interval  $[16, 20]$  and the standard deviation on the interval  $[9, 10]$ . The p-box constructed via the Chebyshev inequality for the same (imprecise) mean and standard deviation is shown as the dotted blue lines. The difference between the two bounds shows the strength of the assumption that the variable has a Normal distribution.



**Figure 4.2:** Probability boxes constructed with nine different methods reflecting the different amount and quality of information typically available in risk assessment problems

With the exception of the confidence-limit methods, the p-boxes described above have two very important properties: they are “rigor-preserving” and “best-possible” (Ferson et al., 2003). Rigor-preserving means that the p-box is guaranteed to enclose the underlying distribution or quantity so long as the information that it is based on is correct - e.g. the range of the variable is truly contained within the stated limits. Best-possible means that the p-box cannot be made any tighter without further information or assumptions. These properties are very important because they guarantee that p-boxes provide an honest account of the uncertainty in a variable (sensu Burgman, 2005) and also provide a way to measure the strength of some of the assumptions that are often introduced into a risk assessment. For example, if we assume nothing about the shape of the distribution function of a random variable, the Chebyshev inequality allows us to calculate the upper bound on the probability of a selecting a random value that is five standard deviations or more larger than its mean. The answer is 1 in 25 or 4%. If we assume, however, that this random variable has a Normal distribution then the chance of selecting the same value is approximately 1 in 3.5 million. Hence the assumption that a random variable has a Normal distribution is a strong assumption because it says an awful lot about the probability of observing values that are far from the mean.

Assumptions about the shape of a distribution function may be justified on the grounds of the mechanism that is thought or known to be generating variability in a parameter (Tables 3.2 and 3.3) but in many practical situations assumptions about the shape of a distribution are often made for the purposes of computational convenience. Assumptions about the dependence or independence between risk parameters are also typically made on the grounds of computational convenience rather than any empirical information that demonstrates that they are, for example, independent. In these circumstances, p-boxes provide an “honest” uncertainty propagation alternative to gauge the strength of these assumptions and their effects on the output of a risk assessment. Probability boxes can also be used to envelope the different opinions of experts whose estimates of uncertain parameters are elicited, thereby faithfully capturing the diversity of their opinions.

Probability bounds analysis proceeds by applying the algorithms for enveloping dependency (Table 3.5) to probability boxes. For example, if the p-box for a finite quantity  $X$  is  $[\overline{F}_X, \underline{F}_X]$ , and the p-box for a finite quantity  $Y$  is  $[\overline{F}_Y, \underline{F}_Y]$  then the Frechet-Hoeffding limits (Equation 3.8) for the quantity  $X + Y$  computed without dependence assumption is (Ferson and Hajagos, 2004)

$$\begin{aligned}\overline{F}_{X+Y}(z) &= \min_{z=x+y} \min(\overline{F}_X(x) + \overline{F}_Y(y), 1) \\ \underline{F}_{X+Y}(z) &= \max_{z=x+y} \max(\underline{F}_X(x) + \underline{F}_Y(y) - 1, 0).\end{aligned}$$

The Hypothetical case study compares and contrasts the results of pooling versus enveloping expert opinion in a probability bounds analysis that also looks at the effect of the independence assumption. The probability bounds estimates are subsequently compared to the results of a Monte Carlo simulation based on the pooled (average) opinions of a group of experts, and the assumption that the probability of each step in the Hypothetical risk model is independent of the other steps (Box 3).

### BOX 3: PBA AND MCS APPLIED TO THE HYPOTHETICAL CASE STUDY

Probability boxes were created for the Hypothetical case study by enveloping the individual (precise) distribution functions elicited from each of the experts. Enveloping is an aggregation procedure that preserves all of the variability between experts and their individual uncertainty. These envelopes, shown in Figure 4.3, were created using Scott Ferson's S4Pbox.R functions. Histograms of the MCS samples taken from each of the pooled distributions for the nine steps of the Hypothetical risk model are shown in Figure 4.4. Red vertical lines show the 10th and 90th percentiles of the data (dashed lines) around the mean (solid line). The potentially complex effects of pooling are evident in the multiple modes of some of these histograms, particularly that for the proportion of infested orchards.

Method	Opinions	Dependence	Q(95)
Qualitative	NA	NA	[0.05, 0.3]
MCS	Pool	Independent	0.0024
PBA	Pool	Independent	[0.0024, 0.003]
PBA	Pool	No assumption	[0, 1]
PBA	Envelope	Independent	[0, 0.25]
PBA	Envelope	No assumption made	[0, 1]

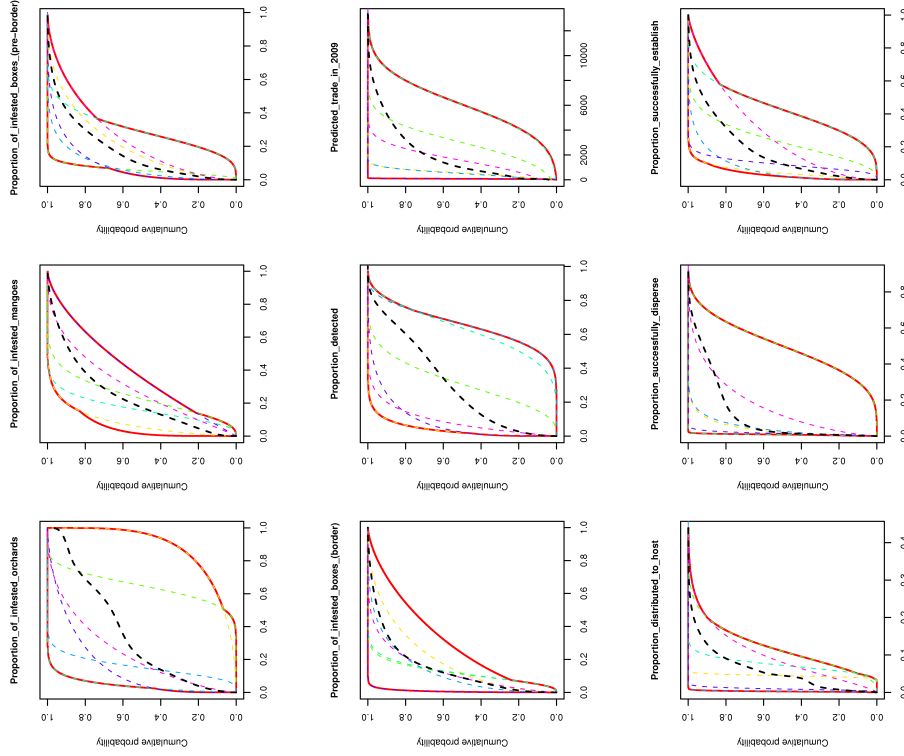
**Table B2: 95th percentile of the annual probability of introduction, establishment and spread of Hypothetical following a qualitative and quantitative risk assessment, with Monte Carlo Simulation and Probability Bounds Analysis results shown**

Figures 4.5, 4.6 and Table B2 summarise the results of the Hypothetical risk assessment using PBA and MCS. Figure 4.5 shows the annual probability of entry, establishment and spread per fruit, whilst Figure 4.6 shows the annual probability for  $N$  fruit based on expert estimates of the volume of trade in a year. The shaded purple boxes shows the qualitative risk assessment result "Low" and the interval definition of this term (Table B1).

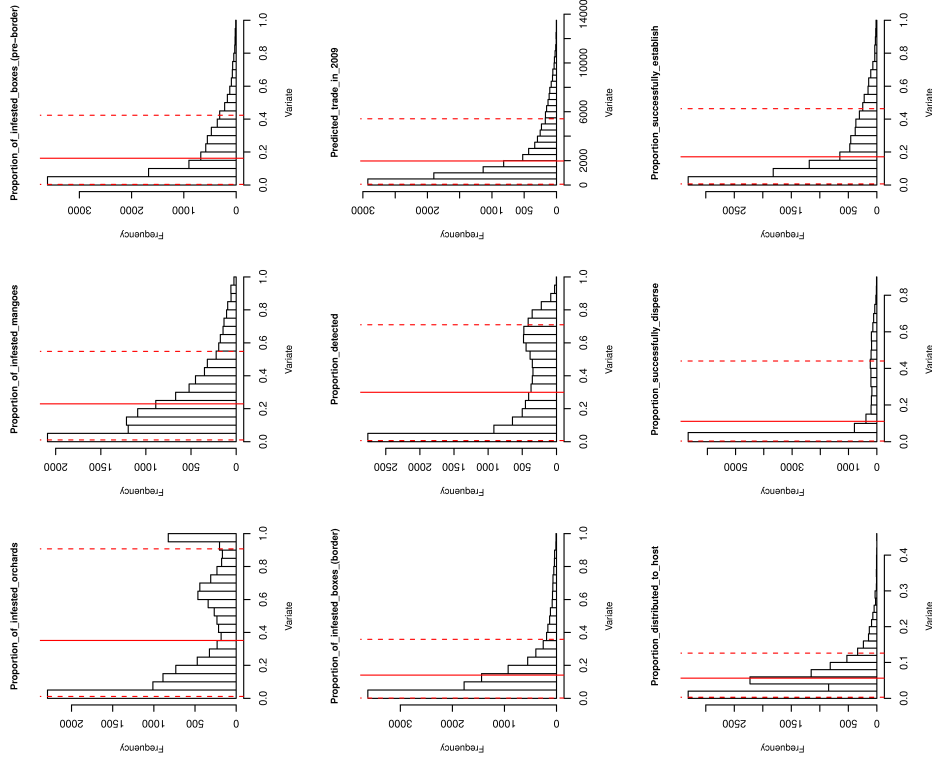
PBA results (coloured lines) are guaranteed to encompass the MCS result (black line). The top left quadrant of each figure has been enlarged to show this. The other important result is that, in this instance, the effect of pooling is swamped by the dependence assumptions. The upper bounds on the 95th percentile of the enveloped PBA result (blue lines) approaches the upper bound of the qualitative result if we assume independence in the model (Table B3). This however seems *a priori* unlikely because it assumes, for example, that the probability of detection at the border is independent of the number of infected fruit that arrive at the border. If we relax this assumption the upper bound on the individual per fruit risk is close to the qualitative prediction, but the upper bounds on the 95th percentile of the risk following one year of trade reaches 1 for both pooled and enveloped predictions, much higher than the qualitative estimate and the MCS estimate.

These results show the strength of the dependence assumption in qualitative and quantitative risk assessment, and also highlight the potential for risk estimates to be very broad if these assumptions are removed. Very broad risk estimates are likely to span decision criteria and may therefore be criticised as unhelpful in a decision-making process. They can serve to highlight, however, the limitations of current information and the effects of (sometimes) unjustified assumptions. They should not therefore be seen as a limitation of honest risk assessment, but rather a strength.

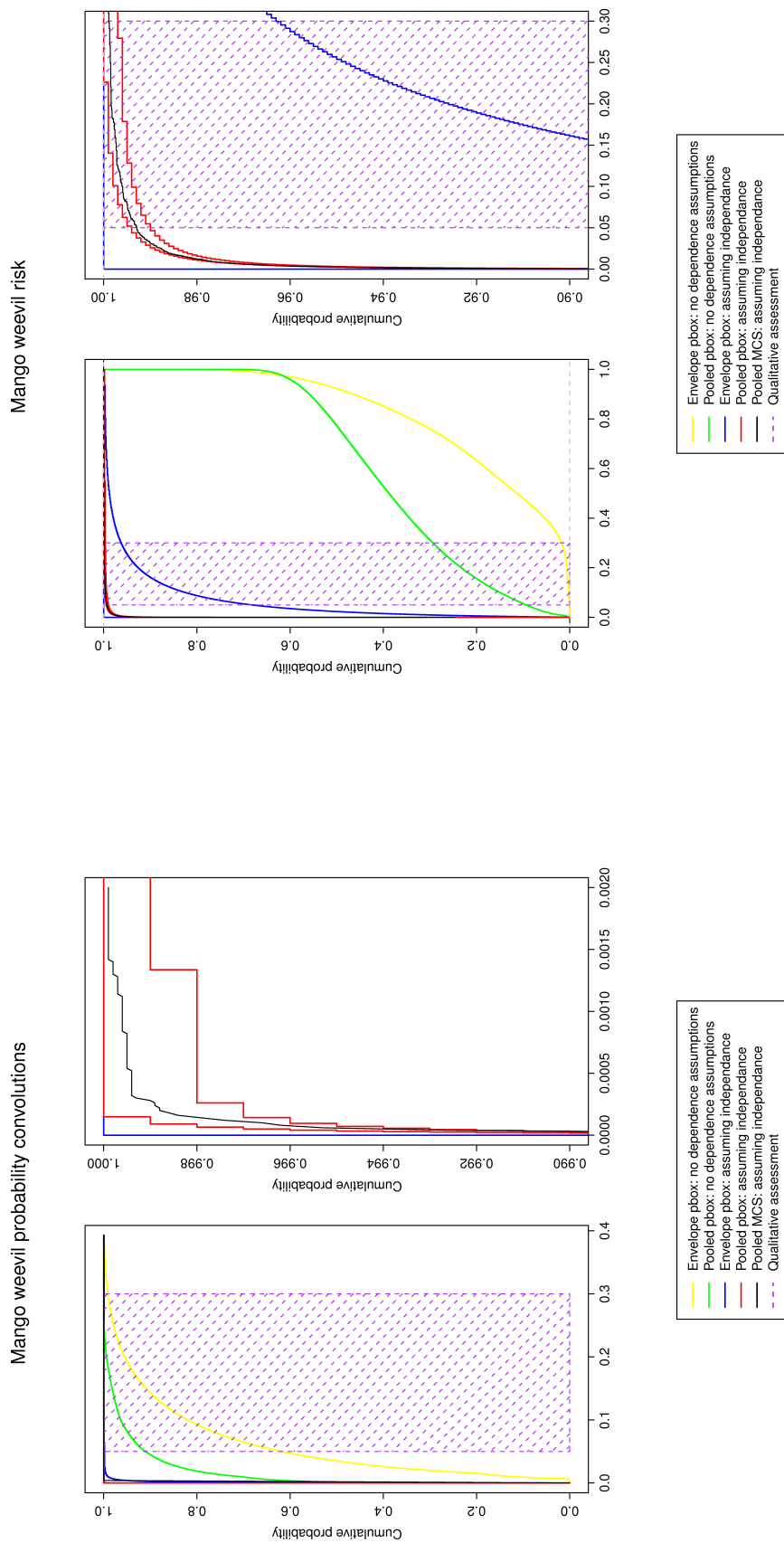




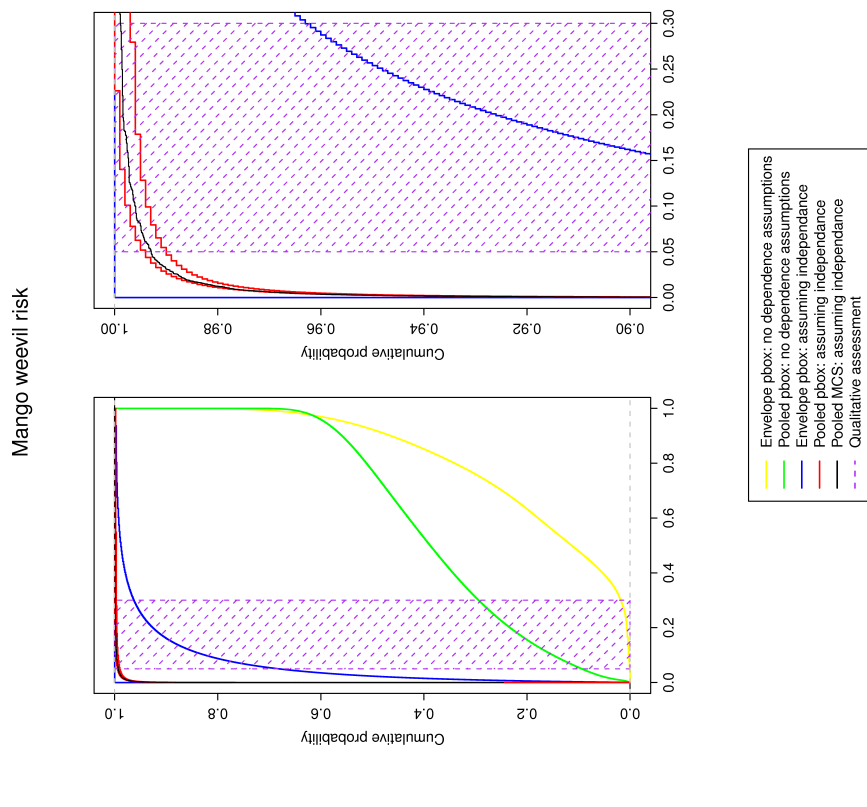
**Figure 4.3:** Probability boxes that envelope (red lines) each of the expert's distribution functions (coloured dashed lines), together with the pooled distribution function (black dashed line) for each of the steps in the hypothetical risk assessment model



**Figure 4.4:** Monte Carlo approximations of the pooled probability densities functions for each of the steps in the hypothetical risk assessment model showing in some cases the complex multimodal density of the pooled opinions



**Figure 4.5:** Annual probability of entry, establishment and spread of hypothetical per fruit estimated quantitatively using Monte Carlo Simulation (black line) and Probability Bounds Analysis (yellow, green, blue and red lines) with and without the assumption of independence. Purple dashed box is overall qualitative risk assessment result



**Figure 4.6:** Results of the hypothetical risk assessment showing the qualitative risk estimate (purple dashed box), the results of the Monte Carlo Simulation assuming independence between each step of the risk model (black line) and the results of the Probability Bounds Analysis with and without the assumption of independence (yellow, green, blue and red lines)

## 4.3 Graphical methods

### 4.3.1 Bayesian networks

Bayesian Networks (BNs)<sup>16</sup> cross the divide between qualitative models, mechanistic models and statistical models (Section 3), and are one of the few methods that can perform forward uncertainty propagation with little or no data, and statistical inference when data is available. BNs provide a transparent, mathematically coherent way to express one's belief in a conceptual model, and the conditional probability of events, in a manner that can be updated as data are gathered during the monitoring and validation stages of an assessment. BNs have much to offer as a risk assessment tool and have been identified as a pragmatic and scientific approach to modeling complex systems in the presence of “high uncertainty” (Hart and Pollino, 2008).

BNs are a relatively new tool. They emerged during the late 1980s and early 1990s as a synthesis of developments in statistical graph theory (Wright, 1934; Shipley, 2000) and Artificial Intelligence, specifically as solutions for conditional probability distributions within complex causal networks (Pearl, 1986; Lauritzen and Spiegelhalter, 1988). The initial development and uptake of BNs focussed largely on medical applications, and examples relevant to the diagnosis of medical conditions predominate early statistical literature. The advantages of a probabilistic description of the relationships in a complex system, however, were quickly recognised by ecologists and by the late 1990's Bayesian networks were being applied to prediction and diagnosis in ecological systems (see reviews in McCann et al. (2006) and Uusitalo (2007)). Today they are a relatively popular method of uncertainty propagation (and inference) with some biosecurity examples in the literature (Henrion, 1989; McMahon, 2005; Peterson et al., 2008; Hood et al., 2009).

The term “Bayesian Network” was coined by Pearl (1986) to describe the “dependency-graph” representation of any joint distribution  $P(x_1, \dots, x_n)$ . The graphical representation of the joint distribution function is achieved via a Directed Acyclic Graph (DAG), that consists of a set of nodes linked by directed (one-way) arrows that indicate the conditional relationship between nodes. Nodes are comprised of states that are independent, mutually exclusive and exhaustive propositions about the values that the variable represented by the node can take. The arrows between nodes describes the particular product-rule decomposition of the joint distribution that in turn reflects the presumed or inferred cause and effect relationship in the system being studied. For example, this factorisation of a three-variable joint distribution

$$P(x_1, x_2, x_3) = P(x_3|x_1, x_2)P(x_2|x_1)p(x_1), \quad (4.9)$$

represents a unique DAG with two arrows linking the node  $x_3$  to its “parents” ( $x_1, x_2$ ) to represent the factor  $P(x_3|x_1, x_2)$ , and one arrow linking the node  $x_2$  to its parent  $x_1$  to represent the factor  $P(x_2|x_1)$  (Bishop, 2006). More generally, the unique decomposition of any joint distribution of a set of risk factors (nodes) represented by a Bayesian Network can be written

$$p(\mathbf{X}) = \prod_{k=1}^K p(x_k|pa_k), \quad (4.10)$$

where  $K$  is number of nodes in the DAG,  $pa_k$  denotes the set of parents of each node and  $x_k$  the values of the variable at the node conditional on the values of its parents.

---

<sup>16</sup>Bayesian Networks are variously described in the literature as probability networks, influence networks, belief networks, Bayesian belief networks, and so forth

Equation 4.10 highlights the causal nature of the directed acyclic graph: nodes within the graph are dependent only on their parents, and are conditionally independent of each other once their parents are given. Having constructed the DAG, the Bayes Network is completed by specifying the conditional probability associated with each variable (node). This can be achieved in several ways (Friedman, 2004)

- via deterministic functions (mechanistic models) that determine the value of  $x_k$  given the input parameters  $pa_k$  of the function;
- via any statistical regression model that relates the response variable  $x_k$  to the explanatory variables  $pa_k$ ; and/or,
- via decision trees or conditional probability tables (CPTs) that describe the probability of discrete values of  $x_k$  for all combinations of the discrete values of its parents.

The structure of a BN represents an assumption about the joint distribution of the risk factors (represented by the nodes of the directed acyclic graph) that are deemed relevant to the problem in hand. In other words the DAG represents a qualitative conceptual model of cause and effect. The conditional probability models associated with each node are a quantitative, statistical or mechanistic model of these causal relations.

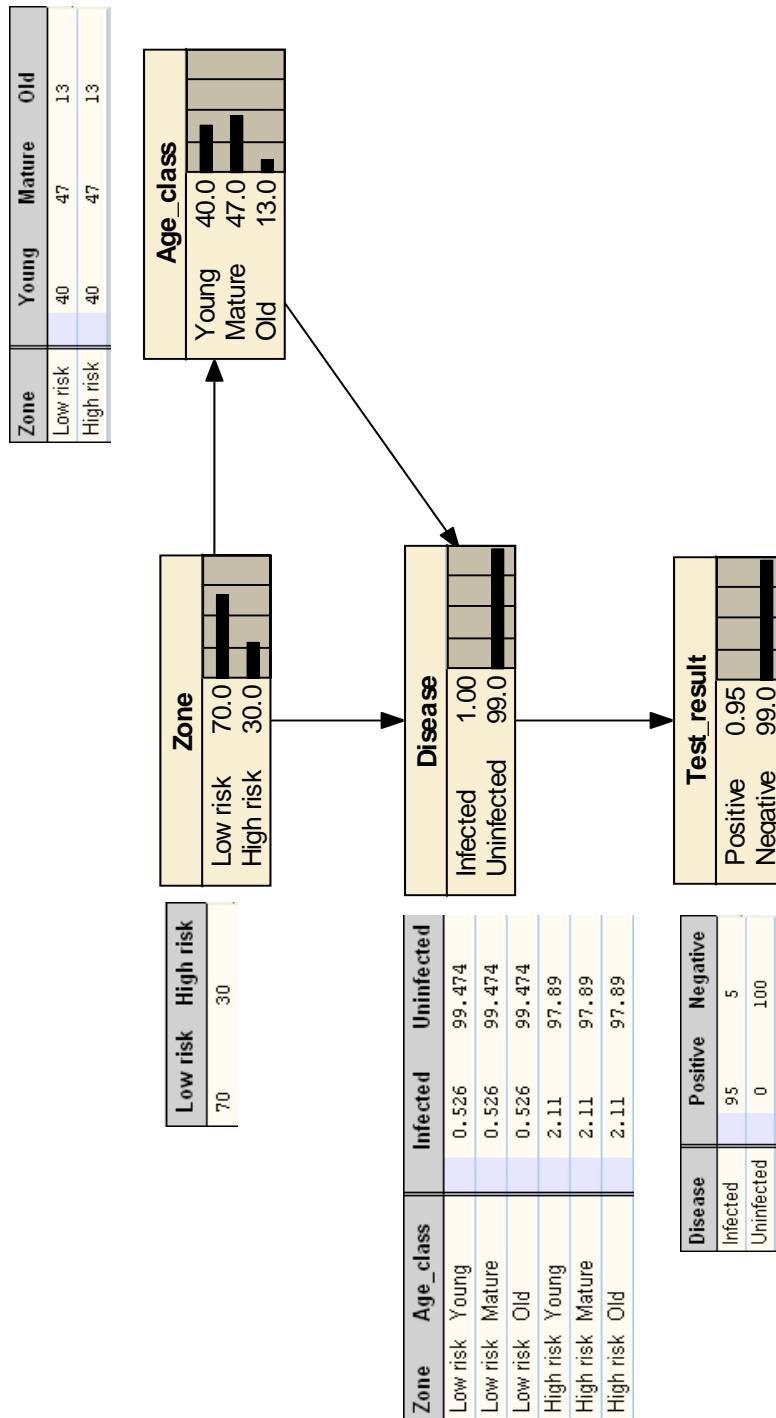
The DAG in Figure 4.7 shows the factorisation of the joint distribution of a four variable bio-surveillance test problem into a set of conditional relationships between the test result, the disease status of the test animal, its age class and the disease zone in which it was raised

$$P(\text{test}, \text{disease}, \text{age}, \text{zone}) = P(\text{test}|\text{disease})P(\text{disease}|\text{age}, \text{zone})P(\text{age}|\text{zone})P(\text{zone}).$$

Figure 4.7 also shows the conditional probability tables that quantify the relationship between the variables within the structure dictated by the DAG. In this example the nodes “zone” and “age-class” are prior variables - i.e. have no parents within the DAG. Notice also how the joint distribution of “disease” depends on the number of states of each of its parents. Here construction of the CPT for the “disease” node requires 12 entries. More generally a binary node with  $n$  binary parents requires  $2^n$  entries to complete its CPT which can represent a considerable analytical burden.

Having constructed the DAG and completed the quantitative component of the network, the BN provides for prediction and diagnosis. In the predictive mode the BN provides the probability of the states of all nodes in the network conditional on the original values specified for the prior nodes. Diagnosis conditional on observations, however, is achieved through the use of Bayes’ theorem. Most BN software packages, for example, allow the user to specify the observed values of a node within the network, which will automatically update the values of all other nodes in the network, via Bayes’ theorem and the fundamental rules of probability (Section 3.3), to be consistent with this observation.

Bayesian Networks are an attractive uncertainty analysis tool for many reasons. They are well suited to problems with small or incomplete data sets and when parameterised manually they are not restricted by a minimum sample size (Uusitalo, 2007). BNs are very flexible - they can be constructed using empirical data, expert opinion or a mixture of both (Wooldridge and Done, 2004). Bayesian networks can also incorporate prior information from a diverse range of disciplines in participatory settings, this facet together with their graphical representation of cause and effect, make them well suited to cross-disciplinary collaboration (Pollino et al., 2007).



**Figure 4.7:** Simple Bayesian Network showing the Directed Acyclic Graph that dictates the conditional cause and effect relationships in a four variable bio-surveillance problem, together with the conditional probability tables that quantify the relationship between these variables (source: Hood et al. (2009))

Perhaps the biggest advantage of BNs, however, is that the process of building the DAG, and quantifying the conditional relationships between the nodes of the network, forces the analyst(s) to think very carefully about the mechanisms, processes and context of their problem. The BN allows the analyst to express their beliefs about things that are and are not casually connected, and the graphical presentation of this information facilitates the participation of, and communication to, stakeholders and other interested parties. Furthermore, the BN approach acknowledges that dependencies between nodes may be uncertain and/or variable and the explicit use of the conditional decomposition of the joint probability distribution forces experts to express probabilistic dependency in a mathematically coherent manner. The use of conditional probability tables, for example, quickly exposes any inconsistency between an expert's belief in an event  $A$  and his or her conditional belief in event  $A$  given  $B$  (Moskowitz and Sarin, 1983).

The current application of Bayesian networks to ecological risk assessment problems also suffers from a number of drawbacks, notably:

- the predictions of a BN are very sensitive to the structure of the DAG (Druzdzel and van der Gaag, 2000) but it is rare to see the issue of model structure uncertainty addressed in ecological applications. This is particularly true for large networks, presumably because of the overheads associated with constructing and quantifying a single network, let alone plausible alternatives. Figure 4.8, for example, highlights the importance of the structural assumptions that underlie any particular DAG. The DAG in the top half of the figure was designed to emulate the European and Mediterranean Plant Protection Organisation (EPPO) import risk assessment scheme (Kuhnert et al., 2009). The independence between pest concentration, survival and growth, and between volume of pest arriving and detection, however, are questionable. The DAG in the bottom half of the figure presents an alternative, arguably more plausible structure, that explicitly allows for negative (or positive) density-dependent growth effects during transit, and conditions the probability of detection upon the number of arriving individuals. Note the reversal of the probability of entry. This effect is partly due to the negative density-dependence relationships (that might plausibly be positive), but overwhelmingly due to the positive dependence between probability of detection and number of pests arriving with a commodity (which is very unlikely to be anything other than positive);
- in data-rich situations it is possible to automate the construction and parameterisation of the network (Friedman, 2004). It is very rare, however, to see this in ecological practice<sup>17</sup>. The majority of BNs (at least in ecological applications) are constructed and quantified using expert opinion and in large networks this requires estimates of hundreds or even thousands of conditional probabilities. As noted above, BNs protect against conditionally incoherent probability estimates, but they provide no protection against the other heuristics and cognitive biases discussed in Section 3.5. Henrion and Breese (1991) and Renooij (2001) discuss this issue with specific reference to Bayesian Networks;
- BN predictions are sensitive to a range of computational issues, notably the way in which continuous prior information is discretised and the way in which data is scaled prior to implementing the network (Kuhnert et al., 2009); and,
- by definition a Directed Acyclic Graph cannot capture the effects of feedback in a system - i.e. a path traced through the structure of DAG cannot pass through a variable (node)

---

<sup>17</sup>I am aware of only one example: Milns, I., Beale, C. M. and Smith, V. A. (2010) Revealing ecological networks using Bayesian network inference algorithms, Ecology Online Preprint, 23 Feb 2010

more than once. This is a potentially important restriction in ecological systems because the dynamics of these systems are typically mediated by positive and negative feedback between components of the system (see Section 4.3.2). Another implication of this restriction is that DAGs cannot be used to capture the dynamics of time-variant processes without building a separate network for each time step. The overheads associated with constructing and quantifying separate networks appears to preclude manual parameterisation, restricting this application to data rich situations.

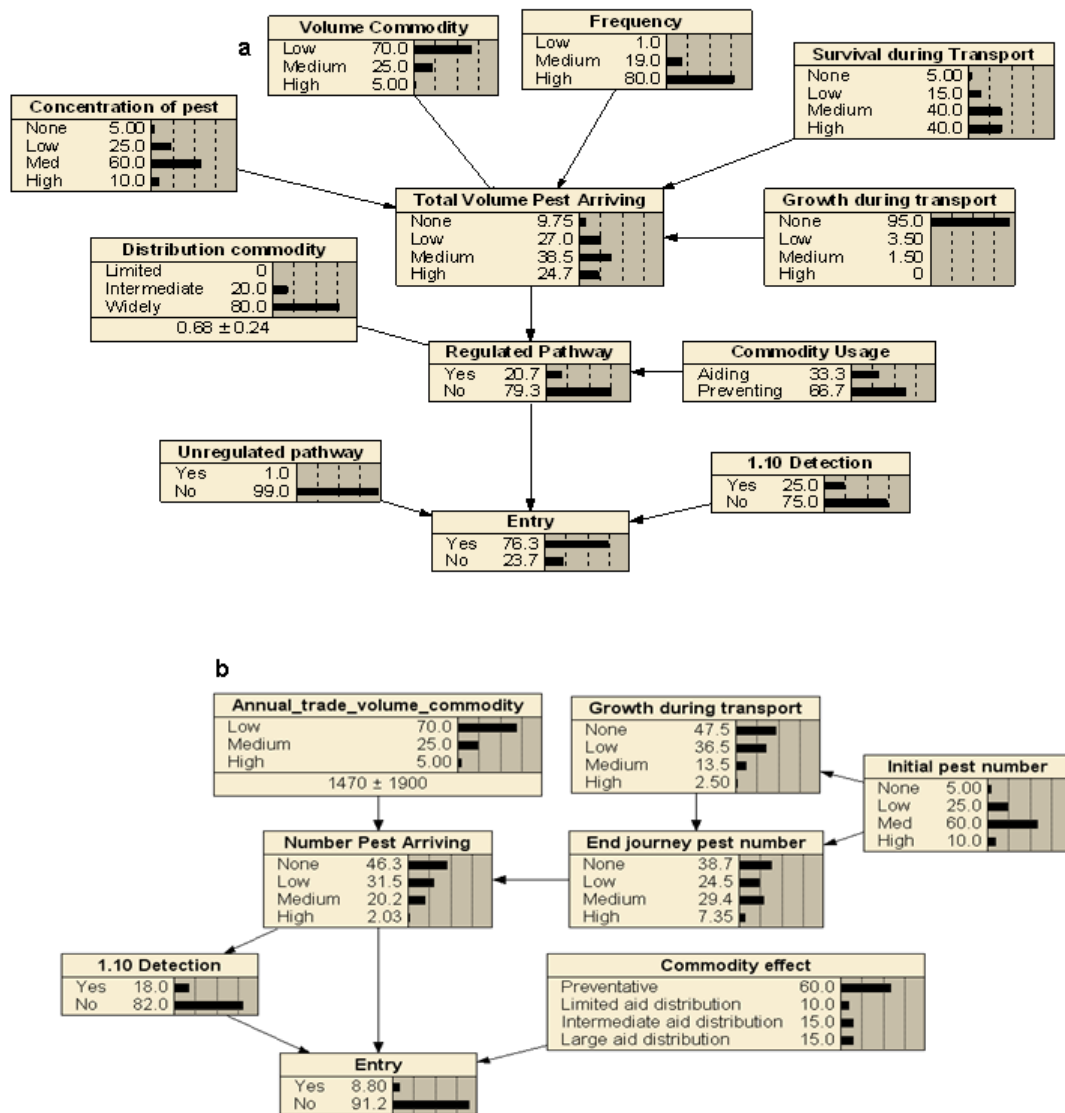
In data-poor situations the advantages of Bayesian networks can be realised, and their disadvantages most readily minimised, when they are applied to relatively simple systems - i.e. when the overheads associated with constructing and quantifying alternative model structures are not overly onerous, and when feedback is not important to the process at hand. The most appropriate applications in these situations are those such as the bio-surveillance problem exemplified in Hood et al. (2009). In data-rich situations BNs can be used to represent and analyse uncertainty in large complex systems, whose dynamics are likely to be determined by time varying processes and feedback cycles. The extension to these systems relies on the more general and more recent developments of statistical graph theory, and to date these development do not appear to have been applied to ecological risk assessment problems.

### 4.3.2 Loop analysis

All risk assessments, qualitative or quantitative, are constructed around a conceptual model of the system in question. Loop analysis (also known as qualitative modeling) provides a quick, rigorous and transparent method that enables certain predictions to be made about the behaviour of this model and to explore the effects of model uncertainty on these predictions. Qualitative modeling is best suited to the early “problem formulation” stage of a risk assessment (USEPA, 1992), prior to the identification of hazards and estimation of risk (Hayes et al., 2007b).

Qualitative modeling proceeds by determining the system’s structure, which is defined by the variables of the system and the relationships by which they are linked. In biological systems, variables are typically interacting populations of different species, and their dynamics can be accounted for by generalized Lotka-Volterra equations, wherein each contributes towards the birth or death of another. Similarly, the dynamics of human social and economic systems can be described by the interactions of different sectors and entities of society (such as governing bodies, social customs, and markets) that control flows of resources, goods, and services that are either measurable, such as money, or immeasurable, such as status and world view.

Variables and relationships in loop analysis are portrayed by Sign-Directed Graphs (SDGs) (or signed digraphs), where a link from one variable to another ending in an arrow ( $\rightarrow$ ) represents a positive direct effect, such as births produced by consumption of prey, and a link ending in a filled circle ( $\rightarrow\bullet$ ) represents a negative direct effect, such as death from predation. All possible ecological relationships can be described in this manner: predator-prey or parasitism ( $\rightarrow\bullet$ ), mutualism ( $\leftrightarrow$ ), commensalism ( $\rightarrow$ ), interference competition ( $\bullet\rightarrow\bullet$ ), and amensalism ( $\rightarrow\bullet$ ). Self-effects are shown by links originating and ending in the same variable, and are typically negative ( $\odot$ ), as in self-regulated variables, but can also be positive ( $\cup$ ) where variables are self-enhancing. Importantly, loop analysis ignores the strength of the pairwise relationships in the SDG by assigning one of two unit signs  $-1$ , or  $+1$  to each interaction. Furthermore, the interactions in the SDG are typically considered to be fixed and independent of population size. However, there can be interactions that are modified by the abundance of a third variable, which creates additional direct effects in the system (Dambacher and Ramos-Jiliberto, 2007).



**Figure 4.8:** Two BNs designed to represent the EPPO risk assessment schema, highlighting the influence of model structure uncertainty on the network predictions. In the DAG shown in a) pest concentration, survival during transport and growth during transport are independent of each other. Similarly the total volume of pest arriving and the probability of detection at the border are also independent. Both of these propositions are unlikely to be true. The DAG in b) includes a density dependent relationship between initial pest concentration, survival and growth during transport, and makes detection at the border dependent on the number of pests arriving



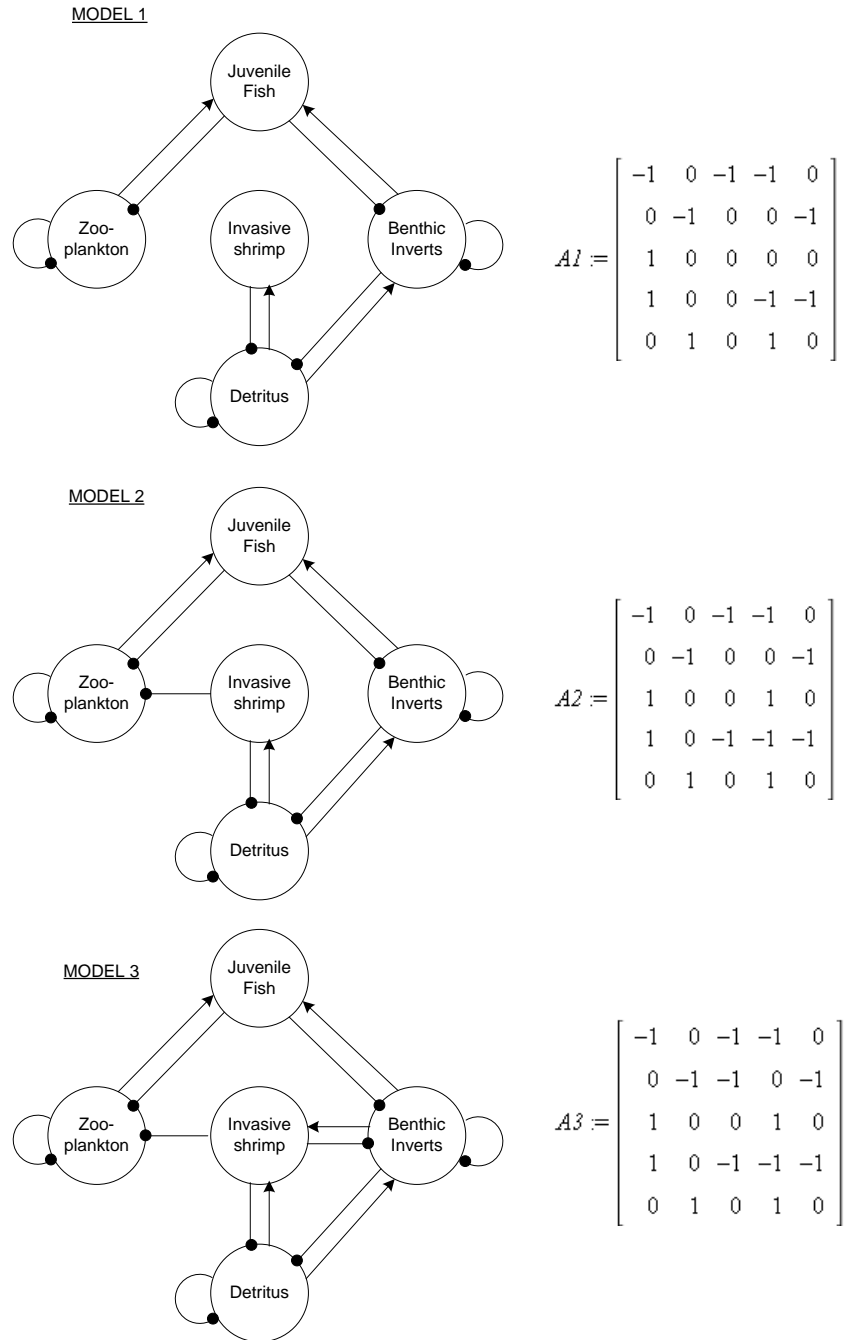
Once the structure of a system is defined then it is possible to: a) analyse the system's feedback which determines the qualitative conditions for system stability; and b) examine its response to sustained (press) perturbations. System feedback is governed by the products of the interactions in the SDG. Negative feedback returns the opposite effect to an initial change in a variable, and acts to maintain a system's equilibrium. The overall stability of a system can be judged and understood according to two criteria that depend on the relative sign and balance of the system's feedback cycles (Puccia and Levins, 1985; Dambacher et al., 2003a). In general, stability requires that the net feedback in a system is negative, and that feedback at lower levels of the system is stronger than feedback at higher levels. Negative feedback ensures that a system's dynamics are self damped, and stronger feedback at lower levels ensures that a system will not overcorrect and exhibit unrestrained oscillations. As system size and complexity increases, the symbolic contingencies underlying the conditions for stability in any one model become too complex to interpret through the Signed Digraph. To address this problem Dambacher et al. (2003a) developed a set of stability metrics that can be used to judge the potential for stability in large complex models.

A press perturbation is defined as a sustained change to a rate of birth, death or migration of a species (Dambacher et al., 2002), or the equivalent increase or decrease in mass, value or flow of other non-biological variables in the SDG. The response of simple systems subject to press perturbations can be predicted by examining the SDG and calculating the product of the sign of the direct effects from the impacted node to all other nodes, multiplied by the sign of the "complementary subsystem" (the feedback of the variables not on the path from the input to the response variable). Again, however, in complex systems this quickly becomes impossible due to the large number of paths and complementary subsystems. Predictions for these systems is achieved via an equivalent algebraic analysis of the system's community matrix (refer to Appendix B).

The utility of loop analysis in a risk assessment context is as a method of forward uncertainty propagation for model structure uncertainty and scenario uncertainty. The Signed Directed Graphs can be quickly constructed with a range of different stakeholders to capture different conceptual models and thereby investigate the potential effects of model structure uncertainty, and/or different perturbation scenarios (Figure 4.9). For example, Hayes et al. (2008) coined the term "pressure scenarios" to describe the combination of uncertain model structure and uncertain future stresses on systems, and used loop analysis to identify system responses that were either consistent across, or idiosyncratic of, these scenarios. Dambacher et al. (2009) adopt a similar approach to scenario uncertainty in commercial fisheries.

The advantages of qualitative analysis of conceptual models early in the risk assessment process are numerous: loop analysis can represent conceptual models in a transparent fashion and help minimise the effects of linguistic uncertainty. Like Bayesian networks, the graphical structure of the SDG is attractive to stakeholders without mathematical training and can be used to elicit conceptual models from a diverse range of different disciplines. Loop analysis also has a rigorous mathematical foundation that can identify unstable (and therefore potentially implausible) conceptual models, the direction of the response of variables subject to multiple, simultaneous, pressures, and the probability of sign determinacy - i.e. the probability that the direction of response will be correct irrespective of the magnitude of the interaction strengths (parametric uncertainty) that it ignores.

By ignoring the magnitude of the interaction coefficients in the community matrix, Qualita-



**Figure 4.9:** Three hypothetical qualitative models (left) and their equivalent community matrix (right), showing possible interactions between a non-native invasive shrimp and four components of the invaded ecosystem: detritus, zooplankton, benthic invertebrates and juvenile fish. Lines with arrows indicate positive effects, and those with filled circles denote negative effects. Self-effects are shown by lines that start and end at the same node. The different models are used to explore three different hypotheses: the shrimp feeds only on detritus (Model 1); the shrimp feeds on detritus and competitively interferes with zooplankton (Model 2); and, the shrimp feeds on detritus, benthic invertebrates and competitively interferes with zooplankton (Model 3). By analysing the sign of the interaction terms in the community matrices the analyst can predict the (in this case indirect) effect of the shrimp on juvenile fish. (Source: Hayes et al. (2007b))

tive modeling (caricature C, Figure 3.2) achieves generality and realism, but at the expense of precision. The lack of precision, and the other equilibrium assumptions associated with loop analysis, entail a number of important drawbacks

- the technique assumes that all interaction within the SDG are equally “strong”. The implications of this assumption can only be examined in a limited sense by considering models with and without interactions that are deemed “weak” or otherwise unimportant to the overall dynamics of the system;
- the technique cannot address questions such as “how much should we spend on  $x$  to get more of  $y$ ?”. The predictions of loop analysis are restricted to the direction (increase, decrease or ambiguous) of change of each variable in the SDG, it says nothing about the magnitude of change;
- qualitative modeling describes system dynamics through a set of linear differential equations (Appendix B). It therefore assumes that the system’s equilibrium, whether it is described by fixed points or sustained bounded fluctuations, will exhibit familiar levels or trajectories of abundance. Furthermore, qualitative predictions of a variable’s response to press perturbations describe a linear shift from one equilibrium to another, and do not address transient behaviour between equilibria, and cannot be used to make predictions for systems that are always held away from an equilibrium by constant external forcing.

In practise, the assumptions associated with qualitative modeling require that: a) there is some level of resolution (in space and time) and some level of aggregation of the system’s variables, at which the system displays familiar dynamics that can be adequately described by linear differential equations; b) that these dynamics are relevant to the problem at hand; and, c) that the model is built at this level of resolution. Hence, when building qualitative models the modeler must choose from among a (ideally systems-based) hierarchy of possible model structures at, for example, increasing spatial resolution such that the variables within the model are relevant to the question being answered, whilst the constant, non-linear or random variations that are omitted from it are not (Levins, 2006). In complex systems, however, this process is difficult to codify and requires an experience modeler.

### 4.3.3 Fuzzy cognitive maps

The term cognitive map is used to describe a variety of conceptual constructs but it is most commonly associated with an influence diagram (or causal map) that shows the variables (variously termed states, nodes, concepts, etc.) deemed to be important to a problem and the direct effects (variously termed as arcs, edges, links, interactions, etc.) between these variables (Siau and Tan, 2005). The term cognitive map was first used in the 1970’s by the political scientist Robert Axelrod to represent graphical portrayals of social scientific knowledge (Kosko, 1986). These maps are precisely the Sign Directed Graphs that support loop analysis - i.e. qualitative, graphical models allowing two-way, positive and negative, causal effects between variables of a system.

Kosko (1986) coined the term Fuzzy Cognitive Map (FCM) to describe a cognitive map in which “causal weights” - numbers on the interval  $[-1, 1]$  - are added to the direct links of the Signed Digraph. It is instructive to note that Kosko (1988) refers to Signed Digraphs as “simple FCMs” with causal edge weights in the set  $\{-1, 0, 1\}$ , hence causality occurs to a “maximal degree”, whereas FCMs allow “degrees of causality” to be represented. This helps illustrate the similarities and differences between loop analysis and FCM. The use of term “fuzzy” in FCM,

however, seems to be open to interpretation. In some applications of FCM the causal weights placed upon the cognitive map are precise numbers (Hobbs et al., 2002; Ozesmi and Ozesmi, 2004) whereas in other applications they are fuzzy sets describing, for example, linguistic measures of relative abundance (Ramsey and Norbury, 2009). Biosecurity relevant examples of FCM in the literature include Ramsey and Veltman (2005) and Ramsey and Norbury (2009)

Graph theory enables a number of similarity statistics to be derived from Fuzzy Cognitive Maps describing, for example, the connectivity of a map. The utility of FCMs in risk assessment, however, again lies in exploring the implications of model structure and scenario uncertainty. The equivalent of perturbation analysis in loop analysis is achieved by solving different maps (with and without the press perturbation) via an iterated matrix operation that finds the roots of the linear differential equations represented by the “adjacency matrix” - the FCM equivalent of the community matrix in loop analysis.

The adjacency matrix  $\mathbf{W} = [w_{ij}]$  of dimension  $n$  contains the causal edge weights of the cognitive map (Figure 4.10). Variables that have no influence on one another are assigned a causal edge weight of zero. The value of each node  $S_i \in [0, 1]$  of the map is given by a function of the weighted sum of its parents and the causal edge weights that link it to its parents

$$S_i = f_i(\mathbf{W}\mathbf{S}_j), \quad (4.11)$$

where  $f_i(\cdot)$  is any monotonic function that returns a value on the interval  $[0, 1]$ . Any suitably bounded function can be used and examples in the ecological literature include the logistic function (Ozesmi and Ozesmi, 2004) and step functions with a lower “activation” level below which  $S_i$  is zero (Hobbs et al., 2002). The resulting values of each node, however, have various interpretations depending on the analysis approach. For example, they can represent a classic fuzzy set membership number (Section 4.4.1) that describes the extent to which the node meets a logical proposition (e.g. habitat is suitable for this species), they can describe the normalised value of a state variable (e.g. the biomass of a species), or the probability of a random event (Hobbs et al., 2002), and so forth.

Nodes within a FCM are often divided into endogenous nodes ( $S_E$ ), whose values are calculated by solving the FCM, and exogenous nodes ( $S_F$ ) whose values are fixed by the user to represent press perturbations such as a management intervention. The steady-state (equilibrium) values of endogenous nodes are derived by solving the graph. The solution is found by multiplying the adjacency matrix by an initial states vector  $S_{E,0} = [S_{1,0}, S_{2,0}, \dots, S_{E,0}]$  resulting in an amended states vector  $S_{E,1}$  whose elements are then transformed on to the interval  $[0, 1]$  using the function  $f_i(\cdot)$ . The transformed state vector is again multiplied with the adjacency matrix and the process repeated until convergence. For example, with a logistic transformation function we have

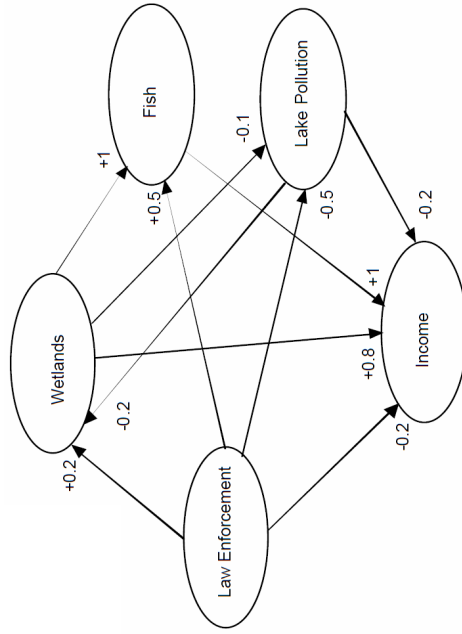
$$\begin{aligned} S_{E,1} &= \frac{1}{1 + \exp(-[\mathbf{W}(S_{E,0} + S_F)])} \\ S_{E,2} &= \frac{1}{1 + \exp(-[\mathbf{W}(S_{E,1} + S_F)])} \\ &\vdots \\ S_{E,t} &= \frac{1}{1 + \exp(-[\mathbf{W}(S_{E,t-1} + S_F)])}, \end{aligned} \quad (4.12)$$

until there is no further change in the value of  $S_{E,t}$  (Figure 4.10). Gauss-Seidel variations that help accelerate convergence have also been used in the literature (Hobbs et al., 2002).

FCM's share many of the advantages of loop analysis: relatively quick, transparent, graphically based, and therefore a good way to elicit conceptual models from a diverse range of stakeholders. They can examine the implications of diverse opinions about plausible model structures and can be used to examine the implications of scenario uncertainty, through for example, the different management regimes associated with different pressure scenarios. The sign and magnitude of the steady-state values also provide additional information on the direction and relative magnitude of change in each of the maps's nodes. The steady-state solution can also incorporate the effects of linguistic uncertainty via the use of fuzzy sets for the causal weights and it seems possible to generalise this to parametric uncertainty via the use of an interval (although I am unaware of any examples of this). There are, however, some drawbacks with FCM's:

- simple signed FCM's (Signed Directed Graphs) are easier to construct with experts and are more reliable than real-valued FCM's because experts are more likely to agree on the causal sign of a direct effect than on its magnitude; (Kosko, 1988);
- FCM analysis is based on the same set of assumptions (e.g. assumes linear relationships between variables and a new stable equilibrium) as loop analysis and therefore suffers from the same limitations in this regard;
- the units of causality in a FCM can be vague and this can create problems when interpreting the results. Ramsey and Veltman (2005) and Ramsey and Norbury (2009), for example, suggest that in the context of community ecology the predictions from a FCM can be "loosely" defined as relative abundance scaled to the interval [0,1]. This definition, however, is problematic for FCM's that incorporate biological and non-biological variables. Moreover, the non-linear transformation function in Equation 4.3.3 is arbitrary (any transformation onto the unit interval can be used) and this will corrupt whatever units the expert had in mind when assigning magnitude to the interactions in the map. Comparing different steady-state FCM's allows predictions of change following a press perturbation to the system, but the arbitrariness of units in the steady-state solutions will again confound interpretation. Importantly, the magnitude of the change contains no information because the absolute value of the increase or decrease is a function of the arbitrary non-linear transformation, and it is not immediately clear if the iterated transformation used to find the roots will maintain the rank order of predicted change under different transformations; and,
- there does not appear to be any explicit stability analysis applied to FCM's and in many applications there are no self-effects applied to the variables in the maps - the diagonal elements of the adjacency matrix are zero. As a result there is no guarantee that the iterative calculations in Equation 4.3.3 will converge to a single value. If there are no negative entries along the main diagonal of the adjacency matrix - i.e. no negative self-effects, then the system cannot be asymptotically stable (May, 1974). Moreover, if all the main diagonal entries are zero, then the system may be neutrally stable and exhibit oscillatory solutions with magnitude depending on initial state. However, this oscillatory solution is known to be unstable: small changes to the underlying equations will disrupt the oscillations (Edelstein-Keshet, 1988).

Fuzzy cognitive maps provide an intermediate step between loop analysis and a fully quantitative, linear, ecosystem model. I am unaware, however, of any studies that have compared the extent to which the predictions of each approach are interpretable and/or consistent with each other.



	Amount of wetland	Fish population	Pollution	Livelihood	Enforcement of laws
Amount of wetland	0	1	-0.1	0.8	0
Fish population	0	0	0	1	0
Pollution	-0.2	-1	0	-0.2	0
Livelihood	0	0	0	0	0
Enforcement of laws	0.2	0.5	-0.5	-0.2	0
FCM simulation					
Initial state vector	1	1	1	1	1
	0.5	0.52497919	0.47003595	0.569546224	0.5
	0.50029964	0.51399455	0.48500450	0.536483643	0.5
	0.500149955	0.51326165	0.48500300	0.535800301	0.5
	0.50014997	0.51325424	0.48500375	0.535757901	0.5
	0.500149963	0.51325421	0.48500375	0.535757526	0.5
Steady state					
	0.500149963	0.51325421	0.48500375	0.535757524	0.5
	0.500149963	0.51325421	0.48500375	0.535757524	0.5
	0.500149963	0.51325421	0.48500375	0.535757524	0.5

**Figure 4.10:** Fuzzy cognitive map (top) and adjacency matrix (bottom) together with its steady-state solution for Uluabat Lake, Turkey. The steady state (equilibrium) solution for this model suggests that pollution units will be lower than fish population or livelihood. The interpretation is that pollution is not severely harming fish populations or livelihoods (source: Ozesmi and Ozesmi (2003) and Ozesmi and Ozesmi (2004))

## 4.4 Non-probabilistic methods

### 4.4.1 Fuzzy set theory

Fuzzy set theory generalizes several classic notions of concise sets that underlie, for example, the axioms of probability theory. Two important concepts generalised by fuzzy set theory are the notions of membership and relation, that describe the presence or absence of association.

The concept of membership is generalised in two subtly different ways. The first defines a *fuzzy set* by generalising the characteristic function<sup>18</sup> of a crisp set with a membership function ( $\mu_A$ ) on the range  $[0, 1]$  that describes the extent to which members ( $x$ ) of a universal set<sup>19</sup> ( $X$ ) are members of the set  $A$ . The closer the output of the membership function is to 1, the greater the degree of membership (Klir and Folger, 1988). For example, a membership function for the fuzzy set of real numbers that are “close to 0” could be

$$\mu_A(x) = \frac{1}{1 + 10x^2}. \quad (4.13)$$

Put more formally: if  $X$  is the universe of discourse, with elements  $x$ , the fuzzy set  $A$  is defined as a set of ordered pairs  $A = \{x, \mu_A(x) | x \in X\}$  where  $\mu_A(x)$  is the membership function of  $x$  in  $A$ .

Fuzzy sets are the constructs of possibility theory because the proposition “ $X$  is  $A$ ” that associates the variable  $X$  with the concept  $A$ , is deemed to restrict the possible values that  $X$  may take. The possibility distribution that this proposition induces is numerically equal to the membership function of  $A$  (Goh et al., 2007).

A relationship between fuzzy sets and imprecise probabilities occurs via the notion of the  $\alpha$ -cut of a fuzzy set. The  $\alpha$ -cut of a fuzzy set  $A$  is defined as the crisp set  $A_\alpha$  that contain all the elements of the universal set  $X$  that have a membership function greater than or equal to  $\alpha$ :  $A_\alpha = \{x \in X | \mu_A(x) \geq \alpha\}$ . Since  $\alpha$  lies on the range  $[0, 1]$  the upper and lower distribution functions associated with a fuzzy set  $A$  are defined by the (crisp) minimum and maximum values of the  $\alpha$ -cut of  $A$ . These constructs are often termed “possibility-probability” distributions (Feng et al., 2010).

The second generalisation of membership is the *fuzzy measure*. A fuzzy measure is defined by a function that assigns a number on the interval  $[0, 1]$  to the power set of  $A$ .<sup>20</sup> The concept of a fuzzy measure, as distinct from a fuzzy set, can be illustrated by the problem of trying to guess the age of a person we are told is “middle-aged”. The power set in this example might be the ten crisp subsets  $\{40, 41, \dots, 49\}$ . A fuzzy measure could be used to represent our guess of the person’s age. The subset that is assigned the highest value by the fuzzy measure represents our best guess of the person’s age. Contrast this problem with one formulated in terms of fuzzy sets in which we know the person’s age but must determine the extent to which he or she is considered to be “old” or “young”.

Fuzzy sets and fuzzy measures are used to treat different types of linguistic uncertainty. Fuzzy sets provide a way to treat vagueness. Fuzzy measures provide a way to treat ambiguity and under-specificity. In the first example above the phrase “close to 0” is vague. This source of

---

<sup>18</sup>The characteristic function of a set  $A$  assigns the value 1 to each element of the universal set if it is a member of  $A$  and 0 otherwise

<sup>19</sup>The universal set is the set of all possible elementary outcomes or elements of concern in any particular problem. In a risk assessment context it could represent all the possible events that contribute to an undesired outcome. It is therefore subject to completeness uncertainty

<sup>20</sup>The power set of  $A$  is the family of subsets that belong to  $A$ .

linguistic uncertainty was treated with a membership function that assigns a precise number on the interval  $[0, 1]$  to an element (in this case a real number  $x$ ) that unequivocally expresses membership to a set whose boundaries are not clearly defined (in this case the set of real numbers close to zero). In the second example the evidence for the age of a person was ambiguous. Fuzzy measures treat ambiguous or under-specified statements by assigning a precise number on the interval  $[0, 1]$  to subsets whose membership of a set  $A$  is unclear.

The practical application of fuzzy sets to risk assessment appears to be largely restricted to the treatment of variability. The simplest application involves fuzzy set operations for union (OR), intersection (AND), and additive complement (NOT), defined for two fuzzy sets  $A$  and  $B$  as

$$\begin{aligned} (\text{AND}) \quad & \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) \\ (\text{OR}) \quad & \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \\ (\text{NOT}) \quad & \mu_{\bar{A}}(x) = 1 - \mu_A(x). \end{aligned} \tag{4.14}$$

These rules have been used to propagate uncertainty through fault trees and to aggregate verbal (fuzzy logic) descriptions of cause and effect relationships (Duque-Ocampo et al., 2006). More general arithmetic and algebraic operations can also be extended to fuzzy sets via the “extension principle” (Zadeh, 1975). The extension principle describes the mapping of fuzzy sets  $A_1, \dots, A_n$  in  $X$  to a Fuzzy set  $B$  in  $Y$  through a function  $f(\cdot)$ , where  $B = f(A_1, \dots, A_n)$ . The membership function of  $B$  is given by

$$\mu_B(y) = \sup_{y=f(x_1, \dots, x_n)} \left\{ \min \left[ \mu_{A_1}(x_1), \dots, \mu_{A_n}(x_n) \right] \right\}. \tag{4.15}$$

Fuzzy sets are not a very popular method to treat vague terms in risk assessment (Section 5.1) and there do not appear to be any biosecurity related examples in the literature. Fuzzy sets do, however, appear to be used more often than probability bounds analysis (based on a very limited sample). This is puzzling since the latter provides the same functionality as the former but without the unfamiliar vernacular of possibility theory. Probability bounds analysis also provides clear and unambiguous methods to deal with dependency between variables during pair-wise convolutions, whereas the arithmetic and algebraic operations of fuzzy set theory do not appear to make allowance for dependency between sets. Theoretically this seems possible via some dependency function on  $\alpha$  but I am unaware of any published examples.

#### 4.4.2 Interval analysis

Interval analysis (Moore, 1966) is one of the simplest ways to propagate all forms of variability and epistemic uncertainty through a risk function. The methods of interval analysis, specifically interval arithmetic, are also related to deterministic “conservative” methods of risk assessment that propagate upper or lower bounds (but not both) through risk functions to find the worst case. Worst-case analysis, however, can lead to hyper-conservative risk estimates and arbitrary levels of protection (Cogliano, 1997; Ferson, 2002; Burgman, 2005). Interval analysis propagates the upper and lower bounds through a risk assessment and therefore shows simultaneously the best and worst case. This offers a view of the other side of the coin but the resultant bounds may still be hyper-optimistic or -conservative.

Interval analysis replaces all point estimates in a risk function with intervals that are presumed or known to contain the true value of an uncertain quantity. These intervals can therefore subsume many different sources of uncertainty, and may accordingly be derived in many different ways. For example, they may subsume measurement error and represent a classical confidence interval



for a mean. They may subsume variability and represent the observed or theoretical quantities of an empirical distribution function, theoretical probability density function or Bayesian posterior distribution. Burgman (2005) provides a useful summary of the different ways to construct and choose an interval. Finally, they may subsume subjective judgement, parametric uncertainty, scenario uncertainty or a combination of all these, and represent the best guess or carefully elicited beliefs of a group of experts. In the later case, it is often useful for the interval to be accompanied by some indication of the expert's confidence that the true value lies within it.

Interval arithmetic propagates interval-valued variables through risk functions. If  $X$  and  $Y$  are uncertain variables represented by intervals on the range  $[x_1, x_2]$  and  $[y_1, y_2]$  then the following arithmetic operations hold<sup>21</sup>

$$\begin{aligned} X + Y &= [x_1 + y_1, x_2 + y_2] \\ X - Y &= [x_1 - y_2, x_2 - y_1] \\ X \times Y &= [\min(x_1y_1, x_1y_2, x_2y_1, x_2y_2), \max(x_1y_1, x_1y_2, x_2y_1, x_2y_2)] \\ X \div Y &= [\min(x_1/y_1, x_1/y_2, x_2/y_1, x_2/y_2), \max(x_1/y_1, x_1/y_2, x_2/y_1, x_2/y_2)] \quad 0 \notin Y. \end{aligned} \tag{4.16}$$

The primary advantages of interval analysis are: a) it is simple and does not require any assumptions about the shape or parameters of a density function or about the potential dependency between uncertain inputs; and, b) it is rigorous - i.e. so long as the uncertain inputs truly lie within their respective intervals then interval analysis result will truly bound the result of an explicit, monotonic, risk function (Ferson, 2002). Although conceptually very simple, interval analysis has some important practical limitations

- division is only defined when the divisor does not contain zero. Hence for two uncertain variables  $x = [2, 19]$  and  $y = [-5, 15]$  the function  $f(x, y) = x/y$  is not defined. A related problem occurs in probability bounds analysis - the Frechet-Hoeffding limits for multiplication and division are undefined for distributions that span zero. Solutions to this problem exist for standard (precise) distributions but not currently for probability boxes (*pers comm* Scott Ferson);
- interval analysis becomes much more difficult if the risk function is not explicitly known (e.g. a computer model whose details are unknown). In this situation, the simple arithmetic operations in Equation 4.16 may not return the true bounds of the function. For example, if the range of an interval input happens to span the maximum or minimum value of function, then the maximum or minimum value of the function will lie on the interior on the interval input (Ferson, 2002). If the (unknown) risk function is, or can be assumed to be, approximately linear, then alternative methods exist to propagate interval-valued inputs (Kreinovich and Nguyen, 2009); and,
- the range of an interval analysis result can be very large with risk functions that involve large numbers of uncertain inputs and arithmetic operations. Moreover, interval analysis provides no measure of the plausibility of values within the resulting range. Thus, if the resulting interval spans a resulting criteria it provides no information about the likelihood that the criteria will be exceeded, only that in a worst or best case situation it could be exceeded. In some circumstances, this could undermine the utility of the assessment.

---

<sup>21</sup>The operations for multiplication and division are simpler if the intervals represent uncertain probabilities on the range  $[0, 1]$

### 4.4.3 Info-gap theory

Ben-Haim (2006) describes Info-gap theory (IGT) as a method for evaluating the effects of severe uncertainty that is radically different from all current theories of decision-making under severe uncertainty. IGT is different because it offers a non-probabilistic approach to decision-making under uncertainty. It provides a quantitative representation of Knight's concept of "true uncertainty" for which "there is no objective measure of probability" (Ben-Haim, 2004). It has been identified as a method that is suitable when the information base is so depauperate that the analyst cannot parameterize a probability distribution, decide on an appropriate distribution or even identify lower or upper bounds on possible parameter values. It is therefore identified by some practitioners as the method of choice when your uncertainty is "unbounded" (Halpern et al., 2006).

Since the first comprehensive description of the theory in 2001, IGT has attracted many proponents, principally in the fields of ecology, engineering and economics (see <http://info-gap.com/>) and there currently at least eight biosecurity examples in the published literature (Yemshanov et al., 2009; Burgman et al., 2010; Carrasco et al., 2010; Yemshanov et al., 2010; Davidovitch et al., 2009; Rout et al., 2009; Moffitt et al., 2008, 2007). There is, however, an on-going debate surrounding IGT that revolves around two claims: a) IGT is not a radically new theory but rather a reformulation of minimax analysis that has been known in the mathematical research literature for over 60 years; and b) its results are sensitive to initial estimates and are not therefore robust to "severe uncertainty" (Sniedovich, 2007, 2008, 2010). It is also worth noting that Ferson and Tucker (2008) highlight similarities between IGT, fuzzy arithmetic and probability boxes

Info-gap theory seeks to find decisions that are robust to uncertainty. It measures the robustness of a decision by showing how wrong an initial model, or an initial estimate of its parameters, can be before a critical point is reached. It also allows the analyst to compare alternative decisions and identify points along a "horizon of uncertainty", at which decisions should change if critical requirements must be met.

The Info-Gap solution for decision-making under uncertainty entails three elements:

- a decision space  $Q$  that includes a number of alternative decisions, actions or choices ( $q \in Q$ ) available to a decision maker. These choices may be alternative surveillance strategies, risk management strategies, decisions about the size of nature reserves, etc.
- a reward function  $R$  that measures how successful the decision is. Reward functions are typically simple process models, risk functions or utility functions, that measure the performance of a decision. Associated with the reward function is a critical reward value  $r_c$  that must be met; and,
- a non-probabilistic model  $U$  for the uncertain quantities  $u$  in the reward function, parameterised by the term  $\alpha$  that measures the amount of uncertainty.

A generic Info-Gap model based on these three constructs is described by the "robustness function"  $\hat{\alpha}(q, r_c)$

$$\hat{\alpha}(q, r_c) = \max \{ \alpha : r_c \leq \min_{u \in U(\alpha, \tilde{u})} R(q, u) \}. \quad (4.17)$$

On first appearance the notation surrounding IGT can appear daunting. It is worth therefore reiterating in words what the robustness function means. Equation 4.17 states that the robustness of a decision  $q$  with uncertain elements  $u$  is the maximum amount of uncertainty  $\alpha$  such that the

minimum reward associated with the decision  $R(q, u)$  is greater than the critical reward value  $r_c$ . The aim of IGT is to help guide the decision-maker in choosing the “best decision” where best is defined as the decision with the greatest robustness function conditional on the reward function and the initial estimates  $\tilde{u}$  of the uncertain elements in this function.

The uncertain quantities in the reward function are almost always the parameters of the process model or risk function, but in certain circumstances can include the model itself (see below). Furthermore, ecological applications of Info-gap theory almost always use a “fractional-error” uncertainty model that creates an expanding interval ( $\alpha \geq 0$ ) around an initial or best estimate of each uncertain parameter ( $\tilde{u}_i$ )

$$\left| \frac{u_i - \tilde{u}_i}{\tilde{u}_i} \right| \leq w_i \alpha. \quad (4.18)$$

This model creates a family of nested intervals that become broader as the value of  $\alpha$  increases

$$(1 - w_i \alpha) \tilde{u}_i \leq u_i \leq (1 + w_i \alpha) \tilde{u}_i. \quad (4.19)$$

The weight parameter  $w_i$  allows the analyst to moderate the influence of individual parameters on the horizon of uncertainty, for example setting low values ( $w_i \ll 1$ ) for parameters which are known to be less variable or less uncertain than others. In ecological practice, however, the weight parameter is usually set to 1 in which case the fractional error model will apply the same amount of uncertainty uniformly to all uncertain parameters. In some applications this may be unrealistic, as some parts of a problem may be much better known or characterised than others (Ferson and Tucker, 2008). Other non-probabilistic uncertainty models include envelope-bound models ( $|\tilde{u}_i - u_i| \leq w_i \alpha$ ) and ellipsoid-bound models ( $(u_i - \tilde{u}_i / \tilde{u}_i)^2 \leq (w_i \alpha)^2$ ) (Moilanen et al., 2006). Models that allow for dependencies between uncertain parameters are also available (Ben-Haim, 2006).

Info-Gap’s most important property is its definition of the best decision as that which is most immune to the uncertainty in the decision maker’s model of the world, represented by the reward function. This is an important counterpoint to optimality-based approaches that identify the best decision as that which maximises the reward function. The word “optimal” implies that the decision could be no better but this is conditional on the reward function (model) and may not be true for variations of this function. Info-gap theory places the uncertainty in the reward function at the center of the decision-making process and encourages decision makers to maximise immunity to uncertainty rather than maximising reward. This is a very sensible approach, and one that recognises that there is no such thing as an “optimal” decision in a non-deterministic decision-making process.

Info-gap theory is ideally suited to the situation wherein the analyst is comfortable specifying one or more models for the problem at hand (the reward function), and an initial estimate for its uncertain parameters, but is uncomfortable or unwilling to specify a probabilistic model, for example a precise parametric distribution, to represent their uncertainty in values that these parameters may take. IGT uncertainty models impose less restrictions and less structure (than precise parametric distributions) on the possible or likely values that an uncertain parameter may take, and may therefore offer a more comfortable alternative to analysts in situations of “severe uncertainty”.

The conclusions of an Info-gap analysis, however, can depend on the analyst’s initial estimate  $\tilde{u}$  and their belief in the plausibility of deviations from this estimate, as the following example demonstrates. Figure 4.11 replicates a published IGT solution to a bio-surveillance problem.

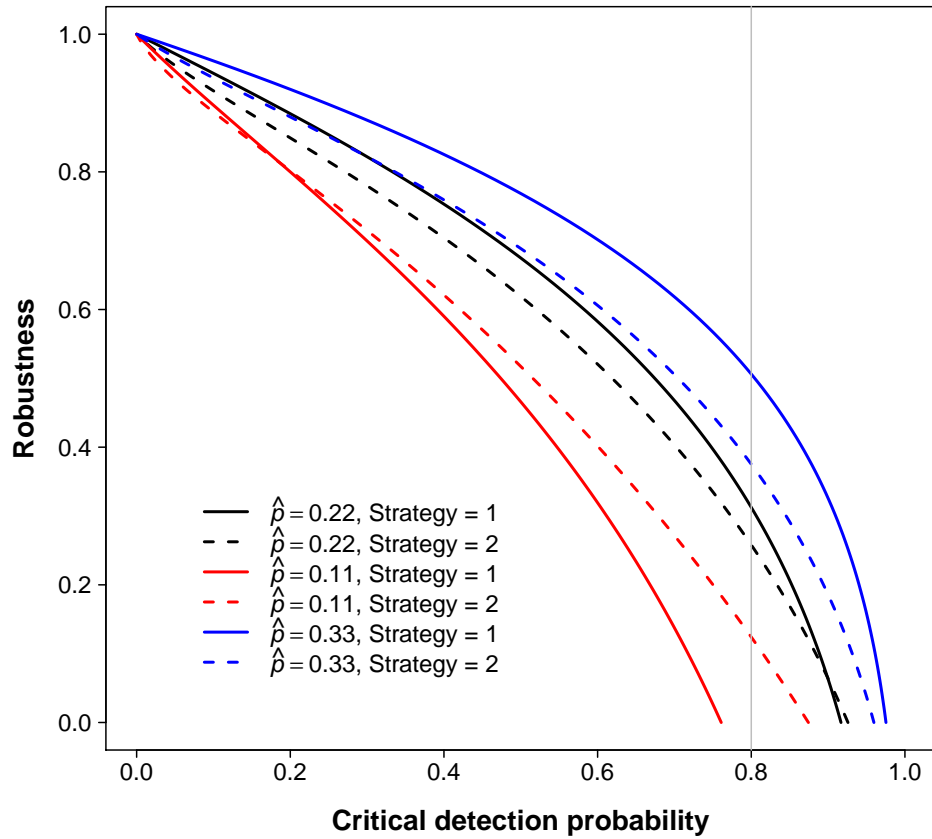
In this example the decision maker is faced with two bio-surveillance strategies. The first strategy places more emphasis on standard surveillance techniques, such as traps, sampling lures and other well structured sampling methods. The second strategy places more emphasis on unstructured observations by trained field personnel and casual observations by relatively untrained workers. The uncertain parameter in this example is the probability  $p$  of detecting a pest species in a single sample.

The authors' original estimate of the detection probability  $\tilde{p}$  is 0.22. In Figure 4.11 we explore the implications of varying this initial estimate by  $\pm 50\%$  - i.e. starting the analysis with  $\tilde{p} = 0.33$  and  $\tilde{p} = 0.11$ . In all cases we want to ensure that the probability of detecting a new pest is at least 80% - the grey vertical line in Figure 4.11. Three conclusions can be drawn from this comparative analysis:

- if  $\tilde{p} = 0.33$  then strategy 1 has a robustness at the critical reward level of approximately 0.5. With a fractional-error model this means that the initial estimate can lie on the range  $0.33 \pm 0.17$  and still guarantee a detection probability of at least 0.8 with this strategy. The robustness of strategy 2 is lower for this critical detection probability and indeed for all levels of robustness  $\leq 1$ . Strategy 1 is therefore unambiguously the best;
- if  $\tilde{p} = 0.22$  then strategy 1 has a robustness at the critical reward level of approximately 0.3, meaning that the initial estimate can lie on the range  $0.22 \pm 0.07$  and still ensure an 80% chance of detecting a new pest. The robustness of the second strategy, however, is much closer to the first. Notice that the first estimate supports  $\tilde{p} = 0.22$  as a "safe possibility" under strategy 1 because 0.22 lies within the interval  $[0.16, 0.5]$  but the second conclusion does not identify  $\tilde{p} = 0.33$  as a safe possibility under strategy 1 because it lies outside the interval  $[0.15, 0.29]$ ; and,
- if  $\tilde{p} = 0.11$  then strategy 1 cannot satisfy the critical reward requirement ( $p \geq 0.8$ ) and strategy 2 becomes the preferred strategy with a robustness approximately equal to 0.1. Notice that IGT highlights the potential for strategy 1 to be unsatisfactory with such a low initial estimate because  $\tilde{p} = 0.11$  lies outside both of the previous intervals, but the reversal of the strategy preference, albeit with low robustness, is not revealed unless the analyst tests the effects of alternative initial starting values.

It is important to recognise that the effect of the initial estimate on IGT conclusions will vary on a case-by-case basis. This example simply serves to highlight the possibility of a better management strategy under different initial conditions. This might seem obvious but this facet of the theory is not widely recognised in the ecological literature. Rout et al. (2009) is a rare exception in this regard, warning that Info-gap theory should not be mindlessly applied, and should not be applied where the uncertainty is so severe that a "reasonable" initial estimate cannot be selected. The term reasonable, however, is very vague in this context.

The possibility of a better management strategy under an alternative initial condition raises another important issue. If, for a given critical reward, the preferred management strategy varies with the initial estimate of an uncertain parameter, the analyst will be forced to consider the plausibility of different parameter values in order to identify a preferred strategy, particularly if the robustness of the alternative strategies are similar. IGT, however, does not provide any estimate of the plausibility of different values, this issue is left with the analyst. Analysts who were attracted to IGT because they are very uncertain, and hence reluctant to specify a probability distribution for a model's parameters, may be disappointed to find that they need to specify the plausibility of possible parameter values in order to identify a robust management strategy.



**Figure 4.11:** Robustness functions for two surveillance strategies highlighting the possibility for the preferred order of different management strategies to vary under different initial estimates of uncertain parameters (in this case the probability of detecting a pest species in a single sample). Black lines show the original result published by Davidovitch et al. (2009). Red and blue lines show alternative plausible values for the uncertain parameter. The vertical grey line is the critical reward requirement. In this case the requirement is that the probability of detecting a new pest incursion must be  $\geq 0.8$

In biosecurity risk assessment one of the most severe forms of uncertainty is our limited understanding of complex ecological processes that manifests as model structure uncertainty. IGT does not provide a ready-made solution to this problem and, as with many other applications of uncertainty analysis, this form of uncertainty is typically not addressed in ecological applications. IGT provides an alternative non-probabilistic way to express uncertainty, but in most ecological applications it is applied to uncertain parameters of probabilistic models, such as the rate of a Poisson process, or the probability of detecting a pest in a trap. Its greatest strength is that it places uncertainty at the forefront of the decision selection problem.

An important point is that its recommendations could be sensitive to the initial estimates of the uncertain parameters. As a method of uncertainty analysis it is not unique in this regard, but, as Figure 4.11 demonstrates, small departures from an initial estimate can still lead to different conclusions.

This is important because IGT does not distinguish between the likelihood of different initial estimates. Hence, if recommendations based on an Info-gap analysis change with different initial estimates, and these estimates are highly uncertain (for example two equally credible experts have different views on the “best” initial estimate) then the theory may not be able to unambiguously identify the best course of action. If the robustness is low at the point where the preference order of the two decisions change (conditional on the required reward) then the theory highlights that the current level of understanding and information is insufficient for reliable decision-making. This insight, of course, presumes that analysts test for the effect of different initial conditions when using IGT.

## 5 Uncertainty analysis in practise

### 5.1 Who's using what?

Risk assessment and uncertainty analysis is practised across a wide range of disciplines. A Web of Science literature search was designed and implemented at the outset of the project to gauge the popularity of different forward uncertainty propagation techniques across various risk assessment application domains. Table 5.1 shows the boolean search terms applied to the title (TS) field of the Web of Science records. The search terms were modified following initial feedback at the project workshop held on the 28th July 2008.

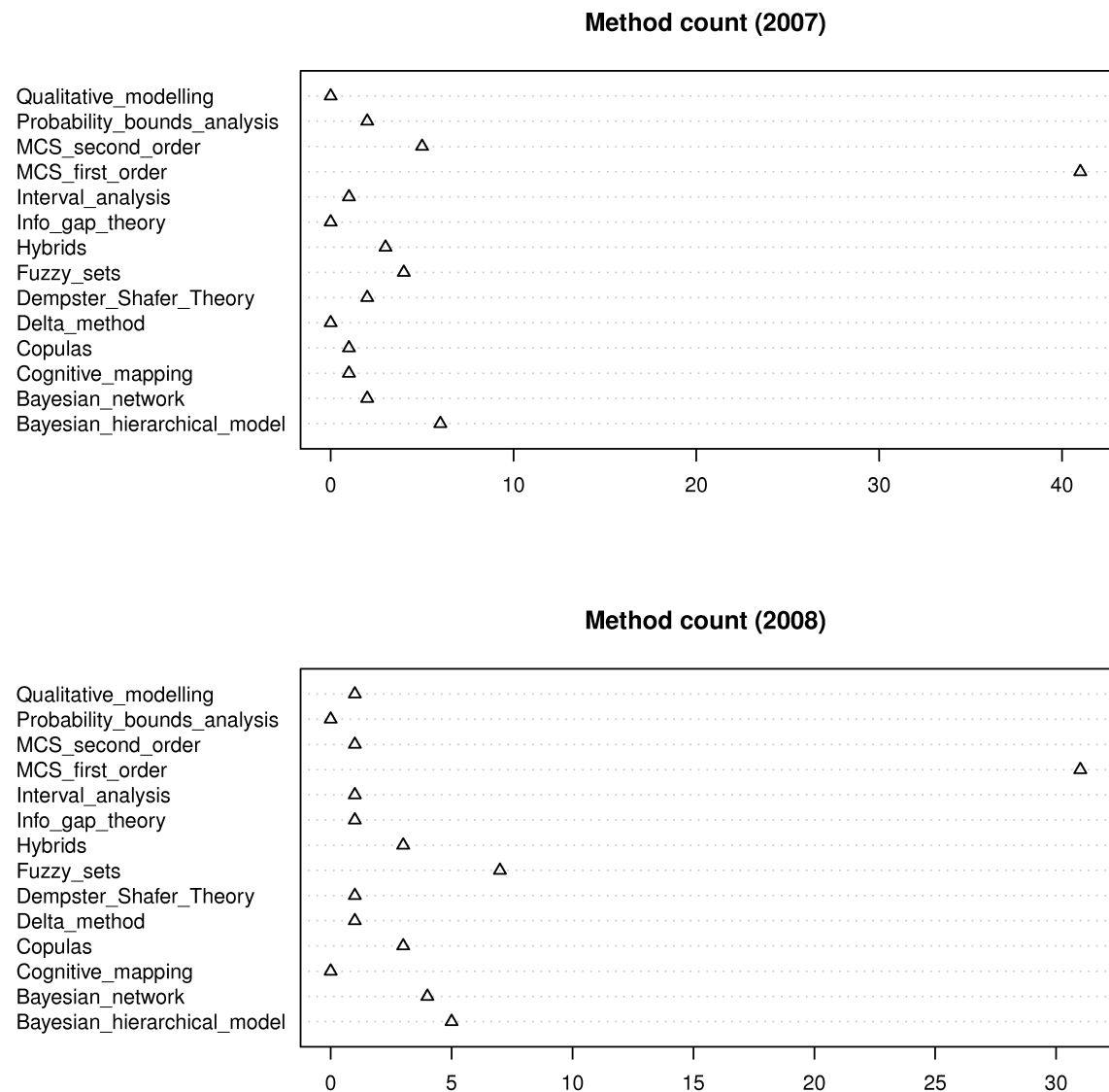
Identifier	Search terms
#3	#2 AND #1
#2	TS=(“fuzzy set*” OR “delta method” OR “interval analys?s” OR “monte carlo simulation*” OR “copula*” OR “probability bound* analys?s” OR “dependency bound* analys?s” OR “cauchy deviate” OR “info gap” OR “dempster shafer” or (hierarchical SAME bayes*) OR “qualitative model?ing” OR “loop analys?s” OR “bayes* belief network*” OR “bayes* analys?s of computer code output*” OR “cognitive map?”)
#1	(TS=“risk assessment” OR TS=“risk analys?s” OR TS=“invasive species” OR TS=“alien invasive species” OR TS=“non indigenous species” OR TS=“non native species” OR TS=“alien species” OR TS=“foreign species” OR TS=“introduced species” OR TS=“biosecurity” OR TS=“bio security” OR TS=“pest”)

**Table 5.1:** Boolean search run on the Web of Science to identify risk assessment studies that use forward propagative uncertainty analysis techniques

The initial application of the literature search recovered 349 references between the years 2000 to 2008. References for the year 2007 and 2008 (139 in total) were subsequently collated and categorised according to uncertainty analysis method, application domain and risk assessment endpoint (Figures 5.1 and 5.2).

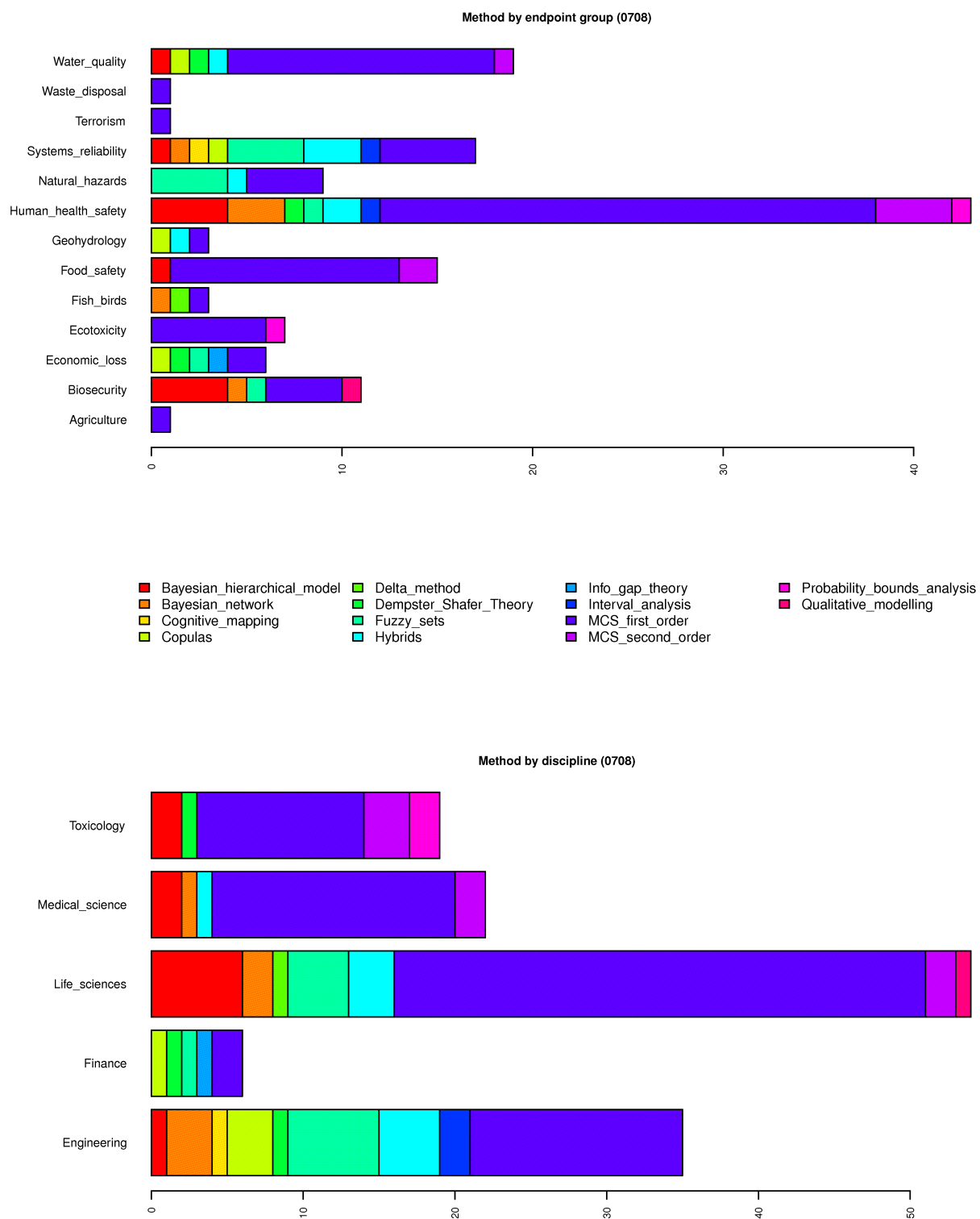
The outstanding result of this literature search is the clear popularity of first order Monte Carlo Simulation. Over the two-year period more than 70 quantitative risk assessment studies used this method to propagate uncertainty through a risk function (Figure 5.1). The analysis of methods by discipline and by endpoint (Figure 5.2) did not identify any clear patterns of usage in this, or indeed the other, uncertainty analysis methods. Bayesian methods - Bayesian networks and Bayesian hierarchical models - figure prominently in the remaining methods, together with fuzzy sets and hybrids - i.e. studies that use two or more of the methods listed here. None of these methods, however, are anywhere near as popular as first order Monte Carlo Simulation.

The analysis by discipline and endpoint (Figure 5.2) suggests that human health and safety, followed by water quality, systems reliability and food safety, are the most common endpoints in quantitative risk assessment that use the uncertainty analysis methods identified here. MCS methods appear to dominate in fields such as toxicology, medicine and life sciences (covering endpoints such as biosecurity, food safety, water quality, etc.) with a somewhat more even mix of methods in engineering and financial applications.



**Figure 5.1:** Dot plots illustrating the frequency of fourteen forward propagative uncertainty analysis techniques in risk assessment studies published in 2007 and 2008





**Figure 5.2:** Bar plots showing the frequency with which different forward uncertainty propagation methods were used by different disciplines and for different risk assessment endpoints in the years 2007 and 2008.

## 5.2 Methods evaluation

Section 4 of the report summarises ten different ways to propagate variability and epistemic uncertainty through risk functions. Whilst the advantages and disadvantages of each approach are discussed, it is perhaps not immediately apparent which method might be best suited to any particular situation. This section of the report attempts to evaluate each of the methods against a set of evaluation criteria in order to shed further light on this issue. The evaluation criteria are:

- Assumptions: What assumptions must be made by the analyst in order to implement the method? How strong are these assumptions? How much data is required to support these assumptions?
- Technical complexity: Is the method well understood? What level of technical expertise is required to implement the method? Is this level of expertise widely available?
- Software availability: What types of software are available to implement the method? What level of training is required to implement the software? What are the computational overheads associated with the method?
- Decision utility: How useful is the approach to decision makers? Will it consistently produce reliable answers when implemented by different competent analysts? How wide are the upper and lower bounds of the answer likely to be?

These criteria were chosen because they address issues of practical implementation, and may therefore help practitioners identify which method(s) might best serve their particular needs. Ultimately this comparison, together with the comparison of their individual advantages, will help to identify an overall strategy for uncertainty analysis that is outlined in Section 6 of the report.

### 5.2.1 Assumptions

The three most important assumptions in propagative uncertainty analysis are: a) the structure of the model or risk function; b) the probabilistic or non-probabilistic function(s) used to represent the different sources of variability and epistemic uncertainty in the model variables; and c) the nature of the dependency between the variables and (where appropriate) the parameters of their density functions where uncertainty is represented probabilistically.

On the whole, model structure uncertainty is rarely addressed in quantitative or qualitative risk assessments. Loop analysis and fuzzy cognitive maps provide a relatively rapid way to address this issue requiring, in the first instance, no more than a simple influence diagram. Both rely on Signed Directed Graphs, thereby allowing for the presence of feedback that is disallowed in the Directed Acyclic Graphs. Both methods, however, assume that the dynamics of the system in question can be adequately portrayed with linear differential equations, and loop analysis provides no information on the transient dynamics of the system as it shifts between equilibria.

The extent to which Fuzzy Cognitive Maps can accurately portray transient dynamics is unclear because the interaction weights are normalised (in a potentially non-linear manner) as the graph is solved. Loop analysis assumes that interactions between variables in the model are in some undefined sense equally “strong”. Vague predictions are the price of this very loose assumption. Fuzzy cognitive maps can accommodate subjective uncertainty about the interaction weights through fuzzy sets and therefore presumably through simple intervals as well.

The predictions of all of the other propagative uncertainty techniques identified here are also

conditional on the assumed model, and all require the variables (risk factors) of this model to be parameterised in some fashion. In the absence of observations, first and second order Monte Carlo simulation require the most onerous assumptions in this context, not only in relation to the probability density function used to represent variability and/or epistemic uncertainty, but also in the nature of the dependency between these variables. In practise the default assumption is one of independence but this is not a conservative approach.

Probability bounds analysis performs much better in this regard. Not only does it free the analyst from having to make assumptions about the shape, scale or location parameters of the probability density function, it also enables him or her to separate the effects of variability from epistemic uncertainty, without the need of additional (hyper) parameterisation, and explore the bounds created by positive or negative non-linear dependency. The price of this assumption-free flexibility is potentially large bounds. Interval analysis offers the same advantages at the same price but with the additional cost of providing no information on the plausibility of the values that lie within the endpoints of the interval answer.

Info-gap theory rates reasonably well against the assumption criteria. In most applications its predictions are conditional on a single model, and ecological examples usually assume that the horizon of uncertainty increases at an equal rate across all of the models parameters. It is possible to info-gap model uncertainty but this is rare in ecological practice (see however Fox et al. (2007) and McDonald-Madden et al. (2008) for good examples). Info-gap solutions do not rely on extensive parameterisation, although they do require the analyst to choose among a number of alternative models of uncertainty. IGT also makes no assumptions about dependency since it does not rely on a probabilistic interpretation of uncertainty.

Bayes Networks make no assumptions about dependency. They explicitly acknowledge the conditional dependence between nodes of the network, and force the analyst to address this in a mathematically coherent manner. The predictions of Bayes Networks, however, are sensitive to the model structure, and Directed Acyclic networks impose additional assumptions on the nature of the cause and effect relationships that they represent, specifically excluding the effects of feedback in the system. Bayesian Networks are methodologically restricted in this regard, however, “loopy Belief Networks” and dynamical networks are an active area of statistical research. There are, however, very few published applications of dynamic networks applied to ecological problems or biosecurity related problems (see Sebastini et al. (2006) and Johnson et al. (2010) for examples in this context).

### **5.2.2 Technical complexity**

The technical complexity of the techniques described above varies considerably between methods. Info-gap theory is the most technically demanding whilst interval analysis is the least demanding in this regard. Info-gap solutions are often not analytical and can involve technically-advanced, constrained optimisation. Probability bounds analysis, Bayesian Networks and fuzzy set theory are the next most technically challenging methods. The language used to describe Probability Bounds analysis and Bayesian Networks is usually very clear (exceptionally so in the case of PBA), and because these methodologies are probabilistic, the language is likely to be familiar to secondary or tertiary educated practitioners.

The basic concept behind Probability Bounds Analysis - an upper and lower bound on a distribution function - is not difficult to grasp. The algorithms and limits that propagate these bounds through arithmetic operations, however, are not trivial and are not widely known. Moreover,

these algorithms are currently implemented in a pair-wise fashion, thus requiring careful programming and occasionally restructuring of a problem to avoid the occurrence of repeated variables, and some arithmetic operations are not currently available for probability boxes that span zero. Extensions of the limits used by PBA to the more general case of three or more variables are complex and still an open research question.

The basic concepts, methods and algorithms of fuzzy set theory are not overly complex. The language associated with this theory, however, is unlikely to be familiar to anyone without specialist training in set theory and alternative schools of uncertainty analysis. This is likely to present a reasonable barrier to risk practitioners who are presented with this material for the first time. The theory of fuzzy sets is obfuscated further because it lacks an unambiguous axiomatic framework. The links between evidence theory, possibility theory and imprecise probability theory are often not clearly articulated in the literature. There is some overlap between the methods and algorithms of these theories but this interpretation often does not match the original motivation of, for example, belief functions or possibility theory (Baudrit et al., 2007a). This situation may reflect the fact that evidence theory and possibility theory are young theories when compared to probability theory, and that their techniques and methods are still evolving.

The basic theory underlining the conditional structure of a Bayesian Network (the sum and product rules) is not difficult to grasp and does not require a high level of statistical training. Moreover, the Bayesian Networks developed for ecological risk assessment applications are usually restricted to relatively simple systems with discrete conditional probability tables at each of the network nodes. These structures can be built with minimal training. The formal methods of inference and diagnosis in Bayes Networks, however, require a greater level of statistical training to understand, although this complexity is typically hidden behind a software application. The statistical theory of graphical models is also much more extensive than that suggested by the usual applications of Bayes Nets in risk assessment, and requires a high (post-tertiary) degree of specialised training to understand and implement. This may be one reason why the structure of a Bayesian Network is rarely inferred from data in risk assessment problems, and why the full scientific utility of statistical graph theory is under-utilised in ecological risk assessment problems.

The Delta method is simple to understand and easy to implement for simple models. It quickly becomes more challenging, however, as the number of parameters in the model increases, particularly if the cross-product terms and/or second-order terms are retained, requiring of the analyst a good grasp of calculus. Loop analysis and fuzzy cognitive maps are also similarly easy to implement and solve with very simple systems of two or three variables. Loop analysis is particularly appealing in this context as symbolic solutions are available for these types of systems. Solutions for more complex systems, however, require a good tertiary-level understanding of matrix algebra and calculus.

The theory behind first- and second-order Monte Carlo simulations is relatively easy to understand (particularly given the ready availability of software and instruction documentation). Understanding the effects of linear and non-linear dependency requires a greater level of expertise, and the algorithms used to generate correlated and/or dependent random variables require a tertiary level of statistical training. Again, however, these algorithms are usually hidden from the user by software. Unfortunately the difference between linear correlation and non-linear dependency is also sometimes missing from the manuals that accompany this software.

### 5.2.3 Software availability

The popularity of Monte Carlo simulation may be due in part to large number of software packages that are available to the user. MCS simulations can be performed with a large variety of commercial software packages such as Crystal Ball, @RISK, MATLAB, Maple, Mathematica, etc. and freeware packages such as R and PopTools. R libraries for second-order MCS have also recently been developed and are freely available (Pouillot and Delignette-Muller, 2010). @RISK and PopTools run in Microsoft's Excel programme and can therefore be implemented in simple spreadsheet formats without the need for any programming skills.

Bayesian Networks are also supported by a fairly large class of commercial software programmes, including Netica, Hugin Expert and MATLAB, together with some freeware programmes such as MSBNx, SMILE and GeNIe. Some of the commercial platforms can also be downloaded for free and run in a "limited mode". These packages typically come with extensive documentation and user-friendly GUI's and therefore require little, if any, programming skills. More general applications of statistical graph theory can be implemented with statistical packages such as R but these require substantial programming skill.

Probability bounds analysis can be performed with one commercial software package RAMAS Risk Calc (Ferson, 2002) and a freeware package known as Statool (Berleant et al., 2003). Scott Ferson has also developed a series of functions in R that can be used to perform PBA within the R environment. These functions are not currently available as an R library on CRAN but can be accessed by contacting Scott directly (scott@ramas.com) and sourced in the usual manner. Statool and Risk Calc are supported by detailed instruction manuals. Risk Calc allows immediate-mode calculation (much like a calculator) and is easy to run once the user becomes familiar with its syntax and functions. Risk Calc also enables the user to implement interval analysis techniques. A software add-in for Microsoft Excel that supports arithmetic with intervals, probability distributions and p-boxes, within Excel's spreadsheet format, is also currently being developed (*pers comm* Scott Ferson).

Loop analysis calculations require specialised commercial mathematical software such as Maple or MATLAB. Symbolic representation of these calculations is currently only available through Maple. The calculations require a high level of programming skills. Fuzzy set theory, Info-gap theory, interval analysis, the Delta method and fuzzy cognitive maps can also be implemented and solved using these programmes and a wide range of statistical software applications including R. Again implementing these methods with these software packages requires a detailed understanding of the algorithms associated with each method, as well as training and familiarity with the software programming language. With the exception of interval analysis, I am unaware of any commercial or freeware software packages that are dedicated to these methods.

### 5.2.4 Decision utility

The decision-utility criteria address the extent to which each of the methods provides consistently precise and accurate answers. Clearly this will be partly determined by the nature of the problem in hand and the amount of data available to the analysis, but a few general comments in this regard are possible. Monte Carlo simulation provides a mechanism for consistent and relatively precise predictions. The precision of the prediction is driven by the variance of the probability density functions used to describe uncertain input parameters, the tail-shape of these density functions, the dependency between them and the risk function that describes the problem at hand. Distributions with low variance, low tail mass (e.g. the Normal distribution),

that are assumed to be independent and propagated through simple risk functions, with small parameter dimension, will lead to more precise predictions.

The extent to which MCS predictions might be accurate, however, depends on how well the assumptions of the analysis are supported by data and/or theory. In data-rich situations - i.e. situations supported by hundreds of observations - it is possible to identify parametric density functions that accurately represent the data. Empirical information on the covariance structure of the data should also highlight the presence of dependency, and if linear, this can also be incorporated into the simulation. The predictions are still conditional on the model, but alternative plausible models can be easily built and simulated with existing software packages. Conditional on the risk function (the model), and correct programming, this situation provides the best opportunity for accurate risk prediction so long as the data are representative of the risk-generating conditions associated with the problem at hand.

In data-poor situations the assumptions that must be made to run a Monte Carlo Simulation are unlikely to be well supported. Furthermore the large number of factors that contribute to a Monte Carlo answer, namely: the risk function, the type of density function of each risk factor, the parameterisation of these density functions and the nature of the dependency between the input parameters, are likely to preclude any systematic evaluation of the effect of plausible alternative assumptions on the risk estimate. Probability bounds analysis provides a rigorous and robust alternative in these situations, that can be tailored to reflect whatever data is available and alleviate the requirements for assumptions regarding the type and/or parameters of the density functions, and the dependence between them. In data-poor situations, however, probability bounds analysis is likely to result in very wide bounds that span the decision maker's management criteria. Probability bounds estimates are also still conditional on the risk function. Interval analysis is also likely to suffer from the same problems, but it suffers further in this regard because it does not offer the opportunity for identifying where the bulk of the probability mass lies within its final bounds.

Info-gap theory provides an alternative decision-support mechanism in these circumstances. Info-gap results are still conditional on the risk function (although in theory this can also be "info-gapped") but it makes no assumptions about the probabilistic characteristics of uncertain input parameters or the dependency between them. In ecological practice, however, uncertain input parameters are typically treated as independent. Info-Gap solutions can be sensitive to the initial parameter estimates, and the decision utility of its solutions may be undermined if these estimates are poorly known. This situation can be improved by evaluating plausible combinations of initial values but in practice this requires *a priori* information on the plausible range of each parameter, and could entail a large number of computations. Moreover, Info-Gap solutions are most useful when the robustness curves for alternative management strategies do not cross. In these situations the methodology will unambiguously identify the best management strategy (again conditional on the model). If the robustness curves cross, however, the utility of the methodology is again diminished because it provides no information on the likelihood of the horizon of uncertainty associated with the change in best management action.

Bayesian Networks, like Monte Carlo simulation, offer the potential for accurate and precise predictions. In this instance, however, it is more difficult to explore the implications of model structure uncertainty (if the conditional probability tables are completed manually) and the predictions of a Directed Acyclic Graph are unlikely to be accurate if the dynamics of the system in question are sensitive to, or dictated by, feedback. The primary advantage of Bayes networks,

and indeed statistical models more generally, is that they offer the analyst the opportunity to test the likelihood of the data under different models, and hence test for agreement between the model and its parameterisation using observations of the risk generating process. In data-rich situations it is possible to automatically populate and structure a Bayes Net so that it provides the most parsimonious description of the conditional relationship in the data. This capability offers substantial decision utility but it requires the decision maker to place a certain amount of trust in the mathematics that underlie it.

It is difficult to build a compelling case for the decision utility of fuzzy set theory because: a) extensive methods exist to convert experts' beliefs and interpretations of linguistic terms into numerical intervals, removing the necessity of methods designed to treat vagueness and ambiguity; and b) the advantages of the upper and lower bounds of belief and plausibility measures can be achieved with a method (imprecise probability) that is highly likely to be more familiar to decision makers. The decision utility of the Delta method also ranks low because of its inability to describe the tails of risk problems - i.e. the low probability, high consequence events, that often have an important bearing on risk management.

Loop analysis deliberately makes vague predictions. In some circumstances knowing the broad direction of change may be sufficient for management purposes, and knowing the extent to which this prediction is robust across a number of plausible models, may enhance its utility considerably. Loop analysis, however, cannot be used to assess the unit benefit for unit cost of management and this limits its utility as a management decision aid. Fuzzy cognitive mapping may be able to provide additional utility by providing some indication of the relative magnitude of change but the corruption of the absolute magnitude of the interaction weights means that it also cannot measure the absolute "bang for buck". The primary utility of both of these methods is a means to explore the implications of model structure uncertainty prior to embarking on more precise methods of uncertainty analysis or inference.

## 6 Discussion and recommendations

### 6.1 Uncertainty and qualitative risk assessment

Risk assessment entails judgements and arithmetic operations with stochastic variables. Uncertainty enters the process via the language used to describe and contextualize the assessment, via our limited knowledge about the most appropriate structure of a risk function and via the inherent variability of the variables in these functions. There are important theoretical and practical reasons to keep these different sources of uncertainty separated throughout the risk assessment that are primarily motivated by the need to separate uncertainty that in theory can be reduced with additional resources (linguistic uncertainty and epistemic uncertainty) from that which cannot (variability). Moreover experience shows that risk assessments that do not explicitly attempt to separate linguistic uncertainty and epistemic uncertainty from variability can provide ambiguous and/or overconfident predictions.

The principal impediment to uncertainty analysis within qualitative risk assessment is that variability and epistemic uncertainty are confounded with each other and with linguistic uncertainty. Separation of the three sources of uncertainty requires, as a minimum, that linguistic uncertainty is eliminated from the problem as far as possible. Fuzzy sets and possibility theory provide a mechanism that was specifically designed to eliminate two important sources of linguistic uncertainty (vagueness and ambiguity). These sources of uncertainty, however, can also be eliminated with probability theory via formal elicitation methods and probability bounds analysis. This approach has the additional advantage of: a) being able to minimise other well known heuristics and biases in human perception and judgements of uncertain events; and, b) couching its analysis within the realms of probability theory which is likely to be more familiar to decision makers than evidence or possibility theory.

Some of the problems associated with qualitative risk assessment, particularly those that perform arithmetic on qualitative expressions of likelihood (or indeed consequence) via matrix-based methods, are only apparent when one source of linguistic uncertainty (vagueness) is eliminated via numerical definition of terms such as “high”, “medium” or “low”. In these circumstances interval analysis provides a means to compare the predictions of the qualitative assessment with the equivalent numerical result. The analysis presented here demonstrates that these types of qualitative assessment are prone to bias and non-commutative results. This occurs because the convolution rules dictated by the matrix-based operations are arbitrary and in general will not conform to mathematical rules. The simple example shown in Figure 2.2 illustrates this problem in relation to Biosecurity Australia’s risk assessment schema. It is, however, a much more general problem that accompanies the other issues with qualitative risk matrices already identified in the literature (Cox et al., 2005; Cox, 2008).

These issues aside, there are a number of other fundamental issues with qualitative approaches to risk assessment that are related to the science-quality criteria of transparency, repeatability and falsifiability, and the decision-utility criteria of precision and accuracy, namely:

- qualitative risk assessment predictions cannot be (in)validated with observations, and uncertainty cannot be coherently propagated through risk functions without translating qualitative metrics of likelihood or consequence into numerical metrics; and,
- the effects of dependency between risk-generating events cannot be coherently explored allowing for different interpretations of the factors responsible for the same observed outcomes, and potentially paradoxical assumptions about the relative likelihood of events.



Overall it is very difficult to see how science- and decision-quality criteria can be maintained with qualitative risk estimates. Language is vague and ambiguous thereby reducing transparency and precision. It can be interpreted differently by different analysts thereby challenging repeatability. Most importantly, it cannot be falsified and cannot therefore in principle be shown to be inaccurate. All of these quality criteria require linguistic descriptions of likelihood and consequence to be numerically defined. Once this step is carefully taken, language may be useful for the purposes of communicating risk and uncertainty but it need not play any further role in risk assessment.

## 6.2 Honest quantitative risk assessment

Once linguistic uncertainty is minimised, there are three important issues that should be addressed as part of the uncertainty analysis stage of a risk assessment:

- what variables (risk factors) are included or excluded from the risk model, and how is uncertainty in the model structure captured and propagated through the assessment?
- how is variability and epistemic uncertainty represented in the variables of the model, and how is this propagated through the assessment?
- are the variables in the risk model independent, and if not, how is actual or potential dependency captured and propagated through the assessment?

In the absence of data - i.e. in a forward propagative mode of uncertainty analysis - the challenges associated with each of these three issues are typically reduced or eliminated from an assessment by making assumptions about the model structure, the shape, scale or location of the distribution that represents variability and/or epistemic uncertainty in the model's variables, and the nature of the dependency between these variables. These types of assumptions have an important bearing on the results of the risk assessment. The principal objectives of propagative uncertainty methods are to report, and where practical test, the effect of these assumptions on the overall risk estimate, and in doing so achieve an honest assessment.

In approaching this problem the analyst can adopt one or more of a number of general strategies to epistemic uncertainty, variability and dependency:

- simply ignore: this is sometimes defensible for parametric uncertainty and model structure uncertainty, but only in limited circumstances. For example, model structure uncertainty can be ignored where the model or risk function is dictated by legislation or guidelines. In this context the risk assessment results are only defensible as a guide to the relative magnitude of risk but this can be useful as a risk-screening decision aid;
- eliminate: this is possible for variability (and to a limited degree dependence) by either building a more complex risk model (to capture the cause and effect processes that create variability) or via choosing a simpler assessment endpoint that enables a simpler risk model with a lower parameter dimension. The first approach may not be attractive in data-limited circumstances because it can increase model structure uncertainty. The latter approach is only tenable if meaningful decision criteria can be stipulated by a decision maker for the simpler endpoint;
- compare and envelope: comparative strategies are akin to sensitivity analysis and seek to highlight the effect of assumptions on risk estimates. Enveloping methodologies place bounds on the best and worst estimates and seek to guarantee that the true result will lie

within these bounds. Interval analysis, probability boxes and probability bounds analysis can satisfy this guarantee for variability and dependence with minimal conditions, for example, that the true value of an uncertain quantity lies within an elicited interval. Overconfident expert opinion is clearly a challenge in this regard. Info-gap theory attempts to place an upper bound on the effects of uncertainty on decisions but its recommendations may be sensitive to initial conditions. Moreover, comparison and enveloping cannot in principal provide guarantees for model structure uncertainty. In the absence of data and statistical inference this problem is unconstrained but is best approached by consulting widely and comparing as many plausible models as is possible within the resources available to the study. Techniques such as influence diagrams, loop analysis and fuzzy cognitive maps can be helpful in this context;

- **average over:** an analyst can average over variability and several sources of epistemic uncertainty, including model structure uncertainty, using techniques such as second-order Monte Carlo Simulation and Bayesian model averaging. Again, however, in the absence of data this problem is unconstrained and computationally more demanding than the compare and envelope strategy. The range of plausible model structures, or alternative probability density functions, that can be addressed within the resources of a single study are therefore likely to be smaller with this strategy, and in the case of model structure uncertainty it can lead to risk estimates that are incompatible with accepted theories; and.
- **model and factorise:** this strategy is applicable to variability and dependence and in the presence of data also provides a means to identify parsimonious descriptions of cause and effect and thereby treat model structure uncertainty. Copulas and Bayesian networks can be used to treat dependence and partition different sources of variability in a risk assessment problem. These methods can be used in data-poor situations but the lack of data-based constraints can still undermine attempts to provide a systematic analysis. Moreover, the full benefits of Bayesian networks, and statistical graph theory and hierarchical modeling in general, cannot be realised in the absence of data.

When considering an honest appraisal of model structure uncertainty it also important to bear in mind that model complexity is not synonymous with accuracy. Complex mathematical models are sometimes seen as more realistic and relevant than simple models (Bartell et al., 2003). There is, however, no evidence to support the contention that complex ecological models are more accurate than their simpler counterparts (Reckhow, 1994; Arhonditsis and Brett, 2004; Fulton et al., 2004). The aim of an honest quantitative risk assessment is therefore to use the most parsimonious representation of the risk-generating process as possible. Parsimony, however, can only be tested when risk predictions are compared to observations.

### **6.3 Uncertainty, statistics and quantitative risk assessment**

Smith (2002) suggests that there is a major difference in the way statisticians and risk analysts view uncertainty because the models used in risk assessment are not simple empirical (data-based) models but mechanistic models. Mechanistic models can be divided into two types: deterministic and stochastic. Deterministic models have no stochastic components and typically ignore most sources of real-world variability. They therefore have limited utility in a risk assessment context. Stochastic mechanistic models include at least one random component to represent real world variability, and at this point the boundary between mechanistic model and statistical model becomes fuzzy.

Virtually all quantitative risk assessments describe risk as a function of one or more variables or risk factors:  $\text{Risk} = f(\mathbf{X})$  (Table 2.2). The function  $f(\cdot)$  can take many different forms but it usually represents some mechanistic description of the risk-generating process. The function captures variability by treating one or more of the input variables  $\mathbf{X}$  as a random variable - i.e.  $f(\mathbf{X})$  is a stochastic mechanistic model. More sophisticated analysis can separate variability and epistemic uncertainty via hierarchical models or by treating input variables as imprecise distribution functions, as described in Section 3. By contrast, the explanatory variables in a statistical model do not have to represent theoretically important components or processes of the risk-generating mechanism. In practise they are usually a set of available observed covariates whose collection may or may not have been informed by domain-based understanding and theory.

Statistical models are not, however, defined by, or restricted to, non-mechanistic descriptions of physical or biological processes. Indeed some of the most exciting developments in statistical science in the last two decades has been in the area of data assimilating models (State Space Models, Hidden Markov Models, etc.). These approaches can convert a set of differential equations describing, for example, the growth and spread of an invasive species (a mechanistic model) into a regression problem (a statistical model) by including a process error term to represent all the real-world variability that is not captured by the model and/or by treating the outcomes of the mechanistic model as a latent stochastic process that is observed imperfectly (caricature E, Figure 3.2). Statistical inference on the model parameters, the latent state(s), process error and/or observation error terms typically proceeds via Bayesian Monte Carlo methods such as Gibbs sampling, Metropolis-Hastings within Gibbs or more advanced adaptive MCMC methods and Sequential Monte Carlo methods, depending on *inter alia* the complexity of the mechanistic model (Carlin and Louis, 2009; Koller and Friedman, 2009). A growing number of examples of ecological applications of these methods are now available in the literature (Clark, 2003; Clark and Bjornstad, 2004; Clark, 2007; Arhonditsis et al., 2007; Peters et al., 2010).

In the cases highlighted above, the traditional distinction between “statistical model” and “mechanistic model” (*sensu* Smith (2002)) is not useful or instructive. The only important difference between quantitative risk functions (based on stochastic mechanistic models) and statistical models is that the latter are specifically designed to explain observed variation. The complexity of a statistical model is therefore constrained by the need to estimate the model and/or its parameters from the available data set. Quantitative risk functions on the other hand can be specified prior to, or in the absence of, any observations. Hence there is no a priori restriction on the complexity of a quantitative risk model other than domain-based knowledge and the computing power needed to run it.

It is bad practice, and indeed unscientific, to make but not test risk assessment predictions. Furthermore virtually all risk assessment frameworks emphasise the importance of monitoring (collecting data), reviewing risk assessment predictions and where necessary, repeating the assessment cycle if predictions are not supported by observations. Hence all risk assessment exercises, and the analysis of variability and epistemic uncertainty, should be viewed as a statistical inference problem informed by observations of the risk-generating process and its outcome. The forward uncertainty propagation techniques identified under the broad heading of “uncertainty analysis” in Figure 3.1 are largely a means to start the risk assessment process prior to the availability of observations. Thereafter they should be supported and ultimately replaced by the inference methods identified under the broad heading of “statistics”. The distinction between the two groups of methods should not be seen as an enduring distinction between two separate

disciplines.

Validating or invalidating risk predictions with observations can be a challenging problem, particularly for rare events (Ferson, 1996a). Validation should, however, be targeted at the entire risk-generating process and not simply at the endpoint of the process. Virtually all risk functions and conceptual models entail a series of events that contribute to the overall assessment endpoint. Furthermore, these events are sometimes not all extremely rare. Hence, while it may not be possible or indeed desirable to gather observations on the risk endpoint, it will almost always be possible to expose inputs, assumptions and intermediate variables in the risk function to actual observations.

In import risk assessment, for example, the survival and initial establishment phase is often the most improbable because it is the most stringent and difficult phase for an invading organism to successfully negotiate, and is it also one of the most difficult steps to physically monitor. Earlier steps in the event chain, however, such as the incidence of infection at collection stations in the exporting nation or the proportion of infected produce that reaches the border are easier to monitor and there are strong theoretical grounds to suspect higher rates of occurrence. Steps in the risk-generating process with these characteristics are potentially easier to monitor and thereby allow inference and (in)validation of quantitative risk assessment predictions and uncertainty analysis assumptions.

## **6.4 A recommended strategy for uncertainty analysis**

The discussion of the preceding sections has highlighted a number of general strategies for linguistic uncertainty, epistemic uncertainty, variability and dependence in risk assessment. They have highlighted science-quality and decision-utility criteria, the difficulties with qualitative risk assessment in this context, and the importance of observations in constraining epistemic uncertainty and (in)validating risk assessment predictions and assumptions. A synthesis of these discussions, together with an examination of the pros and cons of different uncertainty analysis methods, suggests the following overall strategy for uncertainty analysis in data-poor situations:

1. use formal elicitation techniques to canvass the opinions, construct conceptual models and parameterize the beliefs of stakeholders and experts. Use either predictive or structural elicitation methods to convert conceptual models into statistical, qualitative and/or mechanistic models and convert beliefs about stochastic variables into numerical intervals with assigned levels of confidence;
2. ensure feedback is embedded within the elicitation procedure (to minimise the potential for misunderstanding) and apply an advocacy-like procedure to ensure that all aspects of the risk assessment are rigorously reviewed;
3. state risk-decision criteria (risk acceptability levels) in a numeric, measurable fashion for as many of the steps in the risk-generating process as is possible, including steps leading up to the overall assessment endpoint;
4. maintain plausible diverse opinions and in the first instance envelope this diversity using techniques such as loop analysis, comparisons of alternative risk functions, interval analysis, probability boxes and probability bounds analysis. If the upper bound on the subsequent risk estimate is lower than the decision criteria associated with the assessment endpoint, report the result and consider the need for monitoring strategies that enable (in)validation of as many of the steps in the risk-generating process as possible within

the resources available to the assessment. If possible, collect data and use statistical inference methods to check that the risk-generating process is operating within the bounds predicted for each step of the process by the risk assessment;

5. if the lower bound on the enveloped risk estimate is higher than the decision criteria associated with the assessment endpoint consider prohibiting, stopping or otherwise mitigating the risk-generating process and if necessary repeat the risk assessment with risk management steps in place, and include within the assessment the impact of management, and the effects of decision uncertainty upon this; and,
6. if the upper and lower bounds of the enveloped risk estimate straddle the decision criteria associated with the assessment endpoint consider first the effects of dependence and the mitigating effects of positive or negative dependence. For example, a potential application of positive quadrant dependence arises in import risk assessment because the probability of detecting organisms at the border should be positively dependent on the number of organisms that arrive at the border - i.e. as the number of infected units rises so should the probability of their detection. Treating these events as independent denies the reality of inspection regimes, inflates uncertainty bounds and can lead to paradoxical simulations where large numbers of infected units are multiplied by a small probability of detection (and vice-versa) in naive simulations.

The strategy outlined above is designed to enable uncertainty analysis with the minimum amount of assumptions. The objective here is for the assessment to be roughly right rather than precisely wrong. If the enveloped predictions continue to cross the assessment endpoint then two avenues are available to the assessor and manager:

- consider prohibiting or otherwise deferring the risk-generating process and the risk assessment, and collect data that enables statistical inference and a more precise empirical estimate of the risk function (statistical or mechanistic model) and/or the variables (risk factors) associated with this model; or,
- use the most plausible assumptions about the model structure and its variables to provide a more precise risk prediction by modeling and/or factorising the uncertainty associated with the problem using techniques such as Bayesian Networks and second-order Monte Carlo Simulation supported by linear or non-linear estimates of dependence.

It is very important with the second option that the assumptions associated with the analysis are clearly communicated together with the effects of alternative plausible assumptions on the risk estimate where possible. It is also important that monitoring strategies are designed and implemented in order to (in)validate as many of the steps in the risk-generating process as possible and thereby enable a gradual departure from data-poor circumstances to data-rich circumstances, and a move towards the inference opportunities of modern statistical methods.

The decision to proceed with, defer, prevent or mitigate a risk-generating process is informed by risk assessment but ultimately made on socio-economic grounds. The objective of the strategy outlined above is to keep the risk assessment component of this process as useful, scientific and as honest as possible, within the limits of currently available methodology. The extent to which this strategy is applicable to a given situation will be in part dictated by the specific circumstances of the assessment. It is not, for example, intended as a strategy for risk screening although some of its initial elements may still prove useful in this context. The different approaches to elicitation (predictive versus structural) may also provide an opportunity to ei-

ther complement or skip propagative uncertainty analysis methods. The core components of the strategy are probability-based methods supported by observations and a gradual move towards statistical inference. I recommended that these components are retained in any particular application.

## References

- Ades, A. E. and Lu, G. (2003). Correlations between parameters in risk models: Estimation and propagation of uncertainty by markov chain monte carlo. *Risk Analysis*, 23:1165–1172.
- AGDAFF (2001). Guidelines for import risk analysis. Technical report, Government Department of Agriculture, Fisheries and Forestry, Canberra, Australia.
- Agresti, A. (2002). *Categorical Data Analysis*. John Wiley and Sons, Inc.
- Anderson, E. L. and Hattis, D. (1999). Foundations: Uncertainty and variability. *Risk Analysis*, 19:47–49.
- Anderson, J. (1998). Embracing uncertainty: The interface of bayesian statistics and cognitive psychology. *Conservation Ecology*, 2:1–27.
- Apeland, S., Aven, T., and Nilsen, T. (2002). Quantifying uncertainty under a predictive epistemic approach to risk analysis. *Reliability Engineering and Systems Safety*, 75:93–102.
- Arhonditsis, G., Qian, S., Stow, C., Lamon, E., and Reckhow, K. (2007). Eutrophication risk assessment using bayesian calibration of process-based models: Application to a mesotrophic lake. *Ecological Modelling*, 208:215–229.
- Arhonditsis, G. B. and Brett, M. T. (2004). Evaluation of the current state of mechanistic aquatic biogeochemical modeling. *Marine Ecology Progress Series*, 271:13–26.
- Ascough, J. C., Maier, H. R., Ravalico, J. K., and Strudley, M. W. (2009). Future research challenges for incorporation of uncertainty in environmental and ecological decision making. *Ecological Modelling*, 219:383–399.
- Australia, S. (2004). Risk management, australia and new zealand standard 4360:2004. Technical report, Standards Australia, Strathfield, Australia.
- Aven, T. (2010). On the need for restricting the probabilistic analysis in risk assessments to variability. *Risk Analysis*, 30:354–360.
- Barnthouse, L. W., Suter, G. W., Bartell, S. M., Beauchamp, J. J., Gardner, R. H., Linder, E., O'Neill, R. V., and Rosen, A. E. (1986). User's manual for ecological risk assessment. Technical report, Environmental Sciences Division Publication No. 2679, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.
- Barry, S. C. and Lin, Q. (2010). Point of truth calibration: Putting science into scoring systems. *Risk Analysis*, in press.
- Bartell, S. M., Pastorok, R. A., Akcakaya, H. R., Regan, H., Ferson, S., and MacKay, C. (2003). Realism and relevance of ecological models used in chemical risk assessment. *Human and Ecological Risk Assessment*, 9:907–938.
- Baudrit, C., Couso, I., and Dubois, D. (2007a). Joint propagation of probability and possibility in risk analysis: Towards a formal framework. *International Journal of Approximate Reasoning*, 45:82–105.
- Baudrit, C., Guyonnet, D., and Dubois, D. (2007b). Joint propagation of variability and imprecision in assessing the risk of groundwater contamination. *Journal of Contaminant Hydrology*, 93:72–84.

- Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Water Resources Research*, 23:1393–1442.
- Ben-Haim, Y. (2004). Uncertainty, probability and information-gaps. *Reliability Engineering and System Safety*, 85:249–266.
- Ben-Haim, Y. (2006). *Info-Gap Decision Theory*. Elsevier, London, England.
- Benzi, M., Bnezi, M., and Seneta, E. (2007). Francesco paolo cantelli. *International Statistical Review*, 75:127–130.
- Berleant, D., Xie, L., and Zhang, J. (2003). Statool: A tool for distribution envelope determination (denv), an interval-based algorithm for arithmetic on random variables. *Reliable Computing*, 9:91–108.
- Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16:41–51.
- Beven, K. (2009). *Environmental Modelling: An Uncertain Future*. Routledge, London, England.
- Bier, V. (2004). Implications of the research on expert overconfidence and dependence. *Reliability Engineering and Systems Safety*, 85:321–329.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer, New York, USA.
- Bolker, B. M. (2008). *Ecological Models and Data in R*. Princeton University Press, Princeton, New Jersey, USA.
- Breiman, L. (2001). Statistical modelling: the two cultures. *Statistical Science*, 16:199–215.
- Brillinger, D. R. (2002). Some examples of risk analysis in environmental problems. pages 1–22. Key note talk at the University of California, Berkeley, USA. Available online: <http://www.stat.berkeley.edu/brill/Papers/interface2002.pdf> (accessed Dec. 2008).
- Brunswick, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62:193–217.
- Brunswick, E. (1956). *Perception and the representative design of psychological experiment*. University of California Press, Berkley, USA.
- Budescu, D. V., Broomell, S., and Por, H. H. (2009). Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. *Psychological Science*, 20:299–308.
- Bukowski, J., Korn, L., and Wartenburg, D. (1995). Correlated inputs in quantitative risk assessment: The effects of distributional shape. *Risk Analysis*, 15:215–219.
- Burgman, M. A. (2005). *Risks and Decisions for Conservation and Environmental Management*. Cambridge University Press, Cambridge, England.
- Burgman, M. A., Colyvan, M., Christian, R., Dobiecki, B., Elith, J., Fidler, F., Grant, N., Hauser, C., Hayes, K. R., Jones, S., Korb, K., Kuhnert, P., McBride, M., McCarthy, M., Nicholson, A., Robinson, A., Rout, T., Speirs-Bridge, A., and Walshe, T. (2009). Demon-



- strating risk analysis capabilities. Technical report, Australian Centre of Excellence for Risk Analysis, Melbourne University, Melbourne, Australia.
- Burgman, M. A., Wintle, B. A., Thompson, C. A., Moilanen, A., Runge, M. C., and Ben-Haim, Y. (2010). Reconciling uncertain costs and benefits in bayes nets for invasive species management. *Risk Analysis*, 30:277–284.
- Burmester, D. E. and Anderson, P. D. (1994). Principles of good practise for the use of monte carlo techniques in human health and ecological risk assessment. *Risk Analysis*, 14:477–481.
- Burmester, D. E. and Thompson, K. M. (1998). Fitting second-order parameteric distributions to data using maximum likelihood estimation. *Human and Ecological Risk Assessment*, 4:319–339.
- Burnham, K. P. and Anderson, D. R. (1994). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York, USA.
- Byrd, D. M. and Cothorn, R. C. (2000). *Introduction to Risk Analysis*. Government Institutes, Rockville, Maryland, USA.
- Calow, P. and Forbes, V. E. (2003). Does ecotoxicology inform ecological risk assessment? *Environmental Science and Technology*, 37:146–151.
- Carey, J. M. and Burgman, M. A. (2008). Linguistic uncertainty in qualitative risk analysis and how to minimise it. *Annals of the New York Academy of Science*, 1128:13–17.
- Carlin, B. P. and Louis, T. A. (2009). *Bayesian Methods for Data Analysis*. Chapman and Hall, London, England.
- Carrasco, L. R., Baker, R., MacLeod, A., Knight, J. D., and Mumford, J. D. (2010). Optimal and robust control of invasive alien species spreading in homogenous landscapes. *Journal of the Royal Society Interface*, 7:529–540.
- Casman, E. A., Morgan, M. G., and Dowlatabadi, H. (1999). Mixed levels of uncertainty in complex policy models. *Risk Analysis*, 19:33–42.
- Charles, A. T. (1998). Living with uncertainty in fisheries: analytical methods, management priorities and the canadian ground fishery experience. *Fisheries Research*, 37:37–50.
- Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 158:419–466.
- Clark, J. S. (2003). Uncertainty and variability in demography and population growth: A hierarchical approach. *Ecology*, 84:1370–1381.
- Clark, J. S. (2007). *Models for Ecological Data: An Introduction*. Princeton University Press, Princeton, New Jersey, USA.
- Clark, J. S. and Bjornstad, O. N. (2004). Population time series: process variability, observation errors, missing values, lags and hidden states. *Ecology*, 85:3140–3150.
- Clemen, R. T. and Reilly, T. (1999). Correlations and copulas for decision and risk analysis. *Management Science*, 45:208–224.

- Clemen, R. T. and Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19:187–203.
- Cogliano, V. J. (1997). Plausible upper bounds: Are the sums plausible? *Risk Analysis*, 17:77–84.
- Cox, D. C. and Baybutt, P. (1981). Methods for uncertainty analysis: A comparative survey. *Risk Analysis*, 1:251–258.
- Cox, L. A. (2008). What's wrong with risk matrices? *Risk Analysis*, 28:497–512.
- Cox, L. A., Babayev, D., and Huber, W. (2005). Some limitations of qualitative risk rating systems. *Risk Analysis*, 25:651–662.
- Cox, P., Niewohner, J., Pidgeon, N., Gerrard, S., Fischhoff, B., and Riley, D. (2003). The use of mental models in chemical risk protection. developing a generic workplace methodology. *Risk Analysis*, 23:311–324.
- Cressie, N., Calder, C. A., Clark, J. S., ver Hoef, J. M., and Wikle, C. K. (2007). Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling. Technical report, Technical Report No. 798, Ohio State University, Columbus, Ohio, USA.
- Cullen, A. C. and Frey, H. C. (1999). *Probabilistic techniques in exposure assessment: A handbook for dealing with uncertainty in models and inputs*. Plenum Press, New York, USA.
- Dambacher, J., Luh, H., Li, H., and Rossignol, P. (2003a). Qualitative stability and ambiguity in model ecosystems. *American Naturalist*, 161:876–888.
- Dambacher, J. M., Gaughan, D. J., Rochet, M. J., Rossignol, P. A., and Trenkel, V. M. (2009). Qualitative modelling and indicators of exploited ecosystems. *Fish and Fisheries*, 10:305–322.
- Dambacher, J. M., Li, H. W., and Rossignol, P. A. (2002). Relevance of community structure in assessing indeterminacy of ecological predictions. *Ecology*, 83:1372–1385.
- Dambacher, J. M., Li, H. W., and Rossignol, P. A. (2003b). Qualitative predictions in model ecosystems. *Ecological Modelling*, 161:79–93.
- Dambacher, J. M. and Ramos-Jiliberto, R. (2007). Understanding and predicting effects of modified interactions through a qualitative analysis of community structure. *The Quarterly Review of Biology*, 82:227–250.
- Davidovitch, L., Stoklosa, R., Majer, J., Nietrzeba, A., Whittle, P., Mengersen, K., and Ben-Haim, Y. (2009). Info-gap theory and robust design of surveillance for invasive species: The case study of barrow island. *Journal of Environmental Management*, 90:2785–2793.
- Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57:45–97.
- Draper, D., Saltelli, A., Tarantola, S., and Prado, P. (2000). *Sensitivity Analysis*, chapter Scenario and parametric sensitivity and uncertainty analyses in nuclear waste disposal risk assessment: The case of GESAMAC, pages 275–292. John Wiley and Sons, Ltd., New York, USA.

- Druzdzel, M. J. and van der Gaag, L. C. (2000). Building probabilistic networks: Where do the numbers come from? *IEEE Transactions on Knowledge and Data Engineering*, 12:481–486.
- Dubois, D. (2010). Representation, propagation and decision issues in risk analysis under incomplete probabilistic information. *Risk Analysis*, 30:361–368.
- Duque-Ocampo, W., Huguet-Ferre, N., Domingo, J. L., and Schuhmacher, M. (2006). Assessing water quality in rivers with fuzzy inference systems: A case study. *Environment International*, 32:733–742.
- Edelstein-Keshet, L. (1988). *Mathematical Models in Biology*. Random House, New York, USA.
- Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman and Hall, New York, USA.
- Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology Letters*, 7:509–520.
- Embrechts, P., Lindskog, F., and McNeil, A. (2001). Modelling dependence with copulas and applications to risk. Technical report, Swiss Federal Institute of Technology, Zurich, Switzerland.
- Emlen, J. M. (1989). Terrestrial population models for ecological risk assessment: A state-of-the-art review. *Environmental Toxicology and Chemistry*, 8:831–842.
- Environmental Risk Management Authority, E. (1999). Identifying risks. Technical report, ERMA report reference ER-TG-01-1 9/99, Environmental Risk Management Authority, New Zealand.
- Fan, Y., Peters, G. W., and Sisson, S. A. (2009). Automating and evaluating reversible jump mcmc proposal distributions. *Statistical Computing*, 19:409–421.
- Faraway, J. J. (2002). *Practical Regression and Anova using R*. <http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>.
- Fedra, K., Van Straten, G., and Beck, M. B. (1981). Uncertainty and arbitrariness in ecosystems modelling: A lake modelling example. *Ecological Modelling*, 13:87–110.
- Feng, L. H., Hing, W. H., and Wan, Z. (2010). The application of fuzzy risk in researching flood disasters. *Natural Hazards*, 53:413–423.
- Ferson, S. (1996a). Automated quality assurance for ecological risk models. *Human and Ecological Risk Assessment*, 2:558–569.
- Ferson, S. (1996b). What monte carlo methods cannot do. *Human and Ecological Risk Assessment*, 2:990–1007.
- Ferson, S. (2002). *RAMAS Risk Calc 4.0 Software: Risk Assessment with Uncertain Numbers*. Lewis Publishers, Boca Raton, Florida, USA.
- Ferson, S. and Ginzburg, L. R. (1996). Different methods are needed to propagate ignorance and variability. *Reliability Engineering and System Safety*, 54:133–144.
- Ferson, S. and Hajagos, J. G. (2004). Arithmetic with uncertain numbers: rigorous and (often) best possible answers. *Reliability Engineering and Systems Safety*, 85:135–152.

- Ferson, S. and Hajagos, J. G. (2006). Varying correlation coefficients can underestimate uncertainty in probabilistic models. *Reliability Engineering and Systems Safety*, 91:1461–1467.
- Ferson, S., Kreinovich, V., Ginzberg, L., Myers, D. S., and Sentz, K. (2003). Constructing probability boxes and Dempster-Shafer structures. Technical report, SAND2002-4015, Sandia National Laboratories, Albuquerque, New Mexico, USA.
- Ferson, S. and Long, T. F. (1995). *Environmental Toxicology and Risk Assessment, Third Volume*, chapter Conservative uncertainty propagation in environmental risk assessment, pages 97–110. American Society for Testing and Materials, Philadelphia, USA.
- Ferson, S., Nelsen, R. B., Hajagos, J., Berleant, D. J., Zhang, J., Tucker, W. T., Ginzburg, L. R., and Oberkampf, W. L. (2004). Dependence in probabilistic modeling, Dempster-Shafer theory and probability bounds analysis. Technical report, SAND2004-3072, Sandia National Laboratories, Albuquerque, New Mexico, USA.
- Ferson, S. and Tucker, W. T. (2008). Probability boxes as info-gap models. *IEEE Transactions*.
- Finkel, A. M. (1990). Confronting uncertainty in risk management: A guide for decision makers. Technical report, Center for Risk Management Resources for the Future, Washington, USA.
- Food and Agriculture Organisation, F. (1996). International standards for phytosanitary measures, part 1 - import regulations: Guidelines for pest risk analysis. Technical report, Food and Agriculture Organisation of the United Nations, Rome, Italy.
- Fox, D., Ben-Haim, Y., Hayes, K. R., McCarthy, M., Wintle, B., and Dunstan, P. (2007). An info-gap approach to power and sample size calculations. *Environmetrics*, 18:189–203.
- Fox, D. R. and Burgman, M. (2008). *Encyclopaedia of Quantitative Risk Analysis and Assessment*, chapter Ecological risk assessment, pages 1600–1603. John Wiley and Sons Ltd., Chichester, England.
- Frank, M. J., Nelson, R. B., and Schweizer, B. (1987). Best possible bounds for the distribution of a sum - a problem of Kolmogorov. *Probability Theory and Related Fields*, 74:199–211.
- Franklin, J. (2009). *What science knows and how it knows it*. Encounter Books, New York, USA.
- Franklin, J., Sisson, S. A., Burgman, M. A., and Martin, J. K. (2008). Evaluating extreme risks in invasion ecology: learning from banking compliance. *Diversity and Distributions*, 14:581–591.
- Frechet, M. (1935). Generalisations du theoreme des probabilites totales. *Fundamenta Mathematica*, 25:379–387.
- Frechet, M. (1951). Sur les tableaux de correlation dont les marges sont donnees. *Annals of the University of Lyon, Section A*, 9:53–77.
- Freidman, J. H. and Popescu, B. E. (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics*, 2:916–954.
- Freund, J. E. (1992). *Mathematical Statistics*. Prentice Hall, Upper Saddle River, New Jersey, USA.

- Frey, H. C. and Burmaster, D. E. (1999). Methods for characterising variability and uncertainty: Comparison of bootstrap simulation and likelihood-based approaches. *Risk Analysis*, 19:109–130.
- Frey, H. C. and Rhodes, D. S. (1998). Characterisation and simulation of uncertain frequency distributions: Effects of distribution choice, variability, uncertainty and parameter dependence. *Human and Ecological Risk Assessment*, 4:423–468.
- Friedman, N. (2004). Inferring cellular networks using probabilistic graphical models. *Science*, 303:799–805.
- Fulton, E. A., Parslow, J. S., Smith, A. D. M., and Johnson, C. R. (2004). Biogeochemical marine ecosystem models ii: the effect of physiological detail on model performance. *Ecological Modelling*, 173:371–406.
- Gardner, R. H. and O'Neill, R. V. (1983). *Uncertainty and Forecasting of Water Quality*, chapter Parameter uncertainty and model predictions: A review of Monte Carlo results, pages 245–257. Springer-Verlag, Berlin, Germany.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practise*. Chapman and Hall, London, England.
- Gillund, F., Kjolberg, K. A., von Krauss, M. K., and Myhr, A. I. (2008). Do uncertainty analyses reveal uncertainties? using the introduction of dna vaccines to aquaculture as a case. *Science of the Total Environment*, 407:185–196.
- Goguen, J. A. (1969). The logic of inexact concepts. *Synthese*, 19:325–373.
- Goh, Y. M., McMahon, C. A., and Booker, J. D. (2007). Development and characterisation of error functions in design. *Research in Engineering Design*, 18:129–149.
- Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732.
- Haas, C. N. (1997). Importance of distributional form in characterising inputs to monte carlo risk assessments. *Risk Analysis*, 17:107–113.
- Haas, C. N. (1999). On modelling correlated random variables in risk assessment. *Risk Analysis*, 19:1205–1214.
- Haimes, Y. Y. (2009). On the complex definition of risk: A systems-based approach. *Risk Analysis*, 29:1647–1654.
- Halpern, B. S., Regan, H. M., Possingham, H. P., and McCarthy, M. A. (2006). Accounting for uncertainty in marine reserve design. *Ecology Letters*, 9:2–11.
- Halpern, J. Y. (2005). *Reasoning About Uncertainty*. MIT Press, Cambridge, England.
- Hammonds, J. S., Hoffman, F. O., and Bartell, S. M. (1994). An introductory guide to uncertainty analysis in environmental and health risk assessment. Technical report, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.
- Hampton, J. M., Moore, P. G., and Thomas, H. (1973). Subjective probability and its measurement. *Journal of the Royal Statistical Society, Series A*, 136:21–42.

- Hansson, S. O. (2004). Philosophical perspectives on risks. *Techné*, 8:10–34.
- Hart, A., Smith, G. C., Macarthur, R., and Rose, M. (2003). Application of uncertainty analysis in assessing dietary exposure. *Toxicology Letters*, 140-141:437–442.
- Hart, B. and Pollino, C. A. (2008). Increased use of bayesian network models will improve ecological risk assessments. *Human and Ecological Risk Assessment*, 14:851–853.
- Harwood, J. and Stokes, K. (2002). Coping with uncertainty in ecological advice: lessons from fisheries. *Trends in Ecology and Evolution*, 18:617–622.
- Hastie, T., Tibshirani, R., and Freidman, J. (2001). *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, Berlin, Germany.
- Hattis, D. and Burmaster, D. E. (1994). Assessment of variability and uncertainty distributions for practical risk analyses. *Risk Analysis*, 14:713–730.
- Hayes, K. R. (1997). A review of ecological risk assessment methodologies. Technical report, CRIMP Technical Report 13, CSIRO Division of Marine Research, Hobart, Australia.
- Hayes, K. R. (1998). Bayesian statistical inference in ecological risk assessment. Technical report, CSIRO Division of Marine Research, Hobart, Australia.
- Hayes, K. R. (2002a). Identifying hazards in complex ecological systems, part 1: Fault tree analysis for biological invasions. *Biological Invasions*, 4:235–249.
- Hayes, K. R. (2002b). Identifying hazards in complex ecological systems, part 2: Infections modes and effects analysis for biological invasions. *Biological Invasions*, 4:251–261.
- Hayes, K. R. (2003). *Bioinvasions: Pathways, Vectors, and Management Strategies*, chapter Biosecurity and the role of risk-assessment, pages 382–414. Island Press, Washington, D.C., USA.
- Hayes, K. R. and Barry, S. C. (2008). Are there any consistent predictors of invasion success? *Biological Invasions*, 10:483–506.
- Hayes, K. R., Gregg, P. C., Gupta, V. S. R., Jessop, R., Lonsdale, M., Sindel, B., Stanley, J., and Williams, C. K. (2004). Identifying hazards in complex ecological systems, part 3: Hierarchical holographic model for herbicide tolerant oilseed rape. *Environmental BioSafety Research*, 3:1–20.
- Hayes, K. R., Kapuscinski, A. R., Dana, G., Li, S., and Devlin, R. H. (2007a). *Environmental Risk Assessment of Genetically Modified Organisms, Volume 3: Methodologies for Transgenic Fish*, chapter Introduction to environmental risk assessment for transgenic fish, pages 1–28. CABI Publishing, Oxfordshire, England.
- Hayes, K. R., Leriche-Guichard, A., McEnnulty, F. M., Patil, J., Barney, M., and Cooper, S. (2009). Ballast water service level agreement (sla) iv. Technical report, Final report for the Australian Government Department of Agriculture, Fisheries and Forestry, CSIRO Division of Mathematics and Information Sciences, Hobart, Australia.
- Hayes, K. R., Lynne, V., Dambacher, J. M., Sharples, R., and Smith, R. (2008). Ecological indicators for the exclusive economic zone waters of the south west marine region. Techni-

- cal report, Final report for the Australian Government Department of the Environment and Heritage, CSIRO Division of Mathematics and Information Sciences, Hobart, Australia.
- Hayes, K. R., Regan, H. M., and Burgman, M. A. (2007b). *Environmental Risk Assessment of Genetically Modified Organisms, Volume 3: Methodologies for Transgenic Fish*, chapter Introduction to the Concepts and Methods of Uncertainty Analysis, pages 188–208. CABI Publishing, Oxfordshire, England.
- Helton, J. C. and Davis, F. J. (2002). Illustration of sampling-based methods for uncertainty and sensitivity analysis. *Risk Analysis*, 22:591–622.
- Helton, J. C., Johnson, J. D., and Oberkampf, W. L. (2004). An exploration of alternative approaches to the representation of uncertainty in model predictions. *Reliability Engineering and System Safety*, 85:39–71.
- Henrion, M. (1989). *Uncertainty in Artificial Intelligence 3*, chapter Some practical issues in constructing belief networks, pages 161–173. Elsevier Science, Amsterdam, Holland.
- Henrion, M. and Breese, J.S. and Horvitz, E. (1991). Decision analysis and expert systems. *AI Magazine*, Winter:64–91.
- Hickman, J. W. e. a. (1983). A guide to the performance of probabilistic risk assessments for nuclear power plants: Volume i. Technical report, NUREG-CR-2300, United States Nuclear Regulatory Commission, Washington, USA.
- Hilborn, R. and Mangel, M. (1997). *The Ecological Detective: Confronting Models with Data*. Princeton University Press, Princeton, New Jersey, USA.
- Hill, S. L., Walters, G. M., Punt, A. E., McAllister, M. K., Le Quere, C., and Turner, J. (2007). Model uncertainty in the ecosystem approach to fisheries. *Fish and Fisheries*, 8:315–336.
- Hobbs, B. F., Ludsin, S. A., Knight, R. L., Ryan, P. A., Biberhofer, J., and Ciborowski, J. J. H. (2002). Fuzzy cognitive mapping as a tool to define management objectives for complex ecosystems. *Ecological Applications*, 12:1548–1565.
- Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, 32:1–49.
- Hoeffding, W. (1940). *Collected Works of Wassily Hoeffding*, chapter Scale-invariant correlation theory, pages 57–108. Springer-Verlag, New York, USA.
- Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14:382–417.
- Hoffman, F. O. and Hammonds, J. S. (1994). Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Analysis*, 14:707–712.
- Hood, G. M., Barry, S. C., and Martin, P. A. J. (2009). Alternative methods for computing the sensitivity of complex surveillance systems. *Risk Analysis*, 29:1–13.
- Hooten, M. B. and Wikle, C. K. (2008). A hierarchical bayesian non-linear spatio-temporal model for the spread of invasive species with application to the eurasian collared-dove. *Environmental Ecology and Statistics*, 15:59–70.

- Hosack, G. R., Hayes, K. R., and Dambacher, J. M. (2008). Assessing model structure uncertainty through an analysis of system feedback and bayesian networks. *Ecological Applications*, 18:1070–1082.
- Hubbard, D. W. (2009). *The Failure of Risk Management: Why its Broken and How to Fix it*. John Wiley & Sons, New Jersy, USA.
- Hutchinson, T. P. and Lai, C. D. (1990). *Continuous Bivariate Distributions Emphasising Applications*. Rumsby Scientific Publishing, Adelaide, Australia.
- Iman, R. L. and Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics*, A9:1749–1842.
- IPPC (2007). *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter Summary of Policy Makers. Cambridge University Press, Cambridge, England.
- ISO (2009). Iso 31000:2009 risk management – principles and guidelines. Technical report, International Organisation for Standardisation, Geneva, Switzerland.
- ISO/IEC (2002). Guide 73:2002 risk management: Vocabulary guidelines for use in standards. Technical report, International Organisation for Standardisation, Geneva, Switzerland.
- Jenkinson, . (2005). The elicitation of probabilities - a review of the statistical literature. Technical report, Sheffield University, Sheffield, England.
- Jia, J., Dyer, J., and Butler, J. (1999). Measures of perceived risk. *Management Science*, 45:519–532.
- Johnson, S., Fielding, F., Hamilton, G., and Mengersen, K. (2010). An integrated bayesian network approach to lyngbya majuscula bloom initiation. *Marine Environmental Research*, 69:27–37.
- Jouini, M. N. and Clemen, R. T. (1996). Copula models for aggregating expert opinions. *Operations Research*, 44:444–457.
- Kaldane, J. B. and Wolfson, L. J. (1998). Experiences in elicitation. *Journal of the Royal Statistical Society, Series D (The Statistician)*, 47:3–19.
- Kaplan, S. (1997). The words of risk analysis. *Risk Analysis*, 17:407–417.
- Kaplan, S. and Burmaster, D. (1999). Foundations: How, when, why to use all of the evidence. *Risk Analysis*, 19:55–62.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90:773–795.
- Kitagawa, G. (1987). Non-gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Society*, 82:1032–1041.
- Kletz, T. (1999). *HAZOP and HAZAN: Identifying and Assessing Process Industry Hazards*. Taylor and Francis, London, England.
- Klir, G. J. and Folger, T. A. (1988). *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, Englewood Cliffs, USA.



- Knight, F. (1921). *Risk, Uncertainty and Profit*. Houghton Mifflin, Boston, USA.
- Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT Press, Massachusetts, USA.
- Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man-Machine Studies*, 24:65–75.
- Kosko, B. (1988). Hidden patterns in combined and adaptive knowledge networks. *International Journal of Approximate Reasoning*, 2:377–393.
- Kot, M. (2001). *Elements of Mathematical Ecology*. Cambridge University Press, Cambridge, England.
- Kreinovich, V. and Nguyen, H. T. (2009). Towards neural-based understanding of the cauchy deviate method for processing interval and fuzzy uncertainty. Technical report, Departmental Technical Reports (CS), Paper 34, University of Texas at El Paso, [http://digitalcommons.utep.edu/cs\\_techrep/34](http://digitalcommons.utep.edu/cs_techrep/34).
- Kuhnert, P. M. and Hayes, K. R. (2009). *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, chapter How believable is your BBN?, pages 4319–4325. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation.
- Kuhnert, P. M., Hayes, K. R., Martin, T. G., and McBride, M. F. (2009). *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, chapter Expert opinion in statistical models, pages 4263–4268. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation.
- Kumamoto, H. and Henley, E. J. (1996). *Probabilistic Risk Assessment and Management for Engineers and Scientists*. IEEE Press, Piscataway, USA.
- Kuo, J. T., Yen, B. C., Hsu, Y. C., and Lin, H. F. (2007). Risk analysis for dam overtopping: Feitsui reservoir as a case study. *Journal of Hydraulic Engineering*, 133:955–963.
- Kynn, M. (2008). The *Š*heuristics and biases $\checkmark$  bias in expert elicitation. *Journal of the Royal Statistical Society Series A*, 171:239–264.
- Laskey, K. B. (1996). Model uncertainty: Theory and practical implications. *IEEE Transactions on Systems Man and Cybernetics, Part A Systems and Humans*, 26:340–348.
- Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their applications to expert systems. *Journal of the Royal Statistical Society Series B (Methodological)*, 50:157–224.
- Lee, R. C. and Wright, W. E. (1994). Development of human exposure-factor distributions using maximum-entropy inference. *Journal of exposure analysis and environmental epidemiology*, 4:329–341.
- Legendre, P. and Legendre, L. (1998). *Numerical Ecology*. Elsevier, Amsterdam, Netherlands.
- Levins, R. (1968). *Evolution in Changing Environments: Some Theoretical Explorations*. Princeton University Press, Princeton, New Jersey, USA.

- Levins, R. (1993). A response to orzack and sober: Formal analysis and the fluidity of science. *Quarterly Review of Biology*, 68:547–555.
- Levins, R. (2006). Strategies of abstraction. *Biological Philosophy*, 21:741–755.
- Li, D. X. (2000). On default correlation: A copula function approach. *The Journal of Fixed Income*, 9:43–54.
- Limpert, E., Stahel, W. A., and Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *Bioscience*, 51:341–352.
- Link, W. A. and Barker, R. J. (2006). Model weights and the foundations of multimodel inference. *Ecology*, 87:2626–2635.
- Link, W. A., Cam, E., Nichols, J. D., and Cooch, E. G. (2002). Of bugs and birds. markov chain monte carlo for hierarchical modelling in wildlife research. *Journal of Wildlife Management*, 66:277–291.
- Los, F. J. and Blaas, M. (2010). Complexity, accuracy and practical applicability of different biogeochemical model versions. *Journal of Marine Systems*, 81:44–74.
- Lowrance, W. W. (1976). *Of Acceptable Risk*. William Kaufmann, Los Altos, USA.
- May, R. M. (1974). *Stability and Complexity in Model Ecosystems*. Princeton University Press, Princeton, New Jersey, USA.
- McCann, R. K., Marcot, B. G., and Ellis, R. (2006). Bayesian belief networks: applications in ecology and natural resource management. *Canadian Journal of Forest Research*, 36:3053–3062.
- McCarthy, M. A., Burgman, M. A., and Gordon, I. (2007). Review of the use of period of trade and trade volume in import risk analysis. Technical report, Final report, ACERA project 0702, Australian Centre of Excellence for Risk Analysis, Melbourne University, Melbourne, Australia.
- McCulloch, C. E. and Searle, S. R. (2001). ), *Generalised, Linear and Mixed Models*. John Wiley and Sons, Inc. New York, USA.
- McDonald-Madden, E., Baxter, P. W. J., and Possingham, H. P. (2008). Making robust decisions for conservation with restricted money and knowledge. *Journal of Applied Ecology*, 45:1630–1638.
- McMahon, S. M. (2005). Quantifying the community: Using bayesian learning networks to find structure and conduct inference in invasions biology. *Biological Invasions*, 7:833–844.
- Melbourne, B. A. and Chesson, P. (2005). Scaling up population dynamics: integrating theory and data. *Oecologia*, 145:179–187.
- Moffitt, J. L., Stranlund, J. K., Field, B. C., and Osteen, C. D. (2007). *Robust Inspection for Invasive Species with a Limited Budget*, chapter 2, pages 7–22. Springer, New York.
- Moffitt, J. L., Stranlund, J. K., and Osteen, C. D. (2008). Robust detection protocols for uncertain introductions of invasive species. *Journal of Environmental Management*, 89:293–299.

- Moilanen, A., Runge, M. C., Elith, J., Tyre, A., Y., C., Fegraus, E., Wintle, B. A., Burgman, M., and Y., B.-H. (2006). Planning for robust reserve networks using uncertainty analysis. *Ecological Modelling*, 199:115–124.
- Moore, R. E. (1966). *Interval Analysis*. Prentice-Hall, Eaglewood Cliffs, New Jersey, USA.
- Morgan, M. G. and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, England.
- Morley, R. S. (1993). A model for the assessment of the animal disease risks associated with the importation of animals and animal products. *Revue Scientifique et Technique de l'Office International des Epizooties*, 12:1055–1092.
- Morris, W. F. and Doak, D. F. (2002). *Quantitative Conservation Biology: Theory and Practice of Population Viability Analysis*. Sinauer Associates Inc., Sunderland, USA.
- Moskowitz, H. and Sarin, R. K. (1983). Improving the consistency of conditional probability assessments for forecasting and decision making. *Management Science*, 29:735–749.
- Mullin, T. J. and Bertrand, S. (1998). Environmental release of transgenic trees in canada: Potential benefits and assessment of biosafety. *The Forestry Chronicle*, 74:203–219.
- Murray, N. (2002). *Import Risk Analysis: Animals and Animal Products*. Ministry of Agriculture and Forestry, Wellington, New Zealand.
- Nauta, M. J. (2000). Separation of uncertainty and variability in quantitative microbial risk assessment models. *International Journal of Food Microbiology*, 57:9–18.
- Nayak, T. K. and Kundu, S. (2001). Calculating and describing uncertainty in risk assessment: The bayesian approach. *Human and Ecological Risk Assessment*, 7:307–328.
- Negus, P., Moller, G., Blessing, J., Davis, L., Marshall, J., and Dobbie, M. (2009). Stream and estuary assessment program (seap): An assessment framework for riverine ecosystems. Technical report, Queensland Government Department of Natural Resources and Water. Brisbane, Australia.
- Nelsen, R. B. (1999). *An Introduction to Copulas*. Springer-Verlag, New York, USA.
- Niemeijer, D. and de Groot, R. S. (2008). A conceptual framework for selecting environmental indicator sets. *Ecological Indicators*, 8:14–25.
- Oehlert, G. W. (1992). A note on the delta method. *The American Statistician*, 46:27–29.
- O'Hagan, A., Buck, C. E., Daneshmand, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Expert's Probabilities*. John Wiley and Sons, Ltd., Chichester, England.
- O'Hagan, A. and Oakley, J. E. (2004). Probability is perfect but we can't elicit it perfectly. *Reliability Engineering and System Safety*, 85:239–248.
- OIE, O. I. d. E. (1996). International animal health code: Mammals, birds and bees. Technical report, Office Internationale des Epizooties, Paris, France.
- Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994). Verification, validation and confirmation of numerical models in the earth sciences. *Science*, 263:641–646.

- Ott, W. R. (1995). *Environmental Statistics and Data Analysis*. Lewis Publishers, Boca Raton, USA.
- Ozesmi, U. and Ozesmi, S. (2003). A participatory approach to ecosystem conservation: Fuzzy cognitive maps and stakeholder group analysis in uluabat lake, turkey. *Environmental Management*, 31:518–531.
- Ozesmi, U. and Ozesmi, S. L. (2004). Ecological models based on people's knowledge: a multi-step fuzzy cognitive mapping approach. *Ecological Modelling*, 176:43–64.
- Pascual, M. A., Kareiva, P., and Hilborn, R. (1997). The influence of model structure on conclusions about the viability and harvesting of serengti wildebeest. *Conservation Biology*, 11:966–976.
- Pastorok, R. A., Bartell, S. M., Ferson, S., and Ginzburg, L. R. (2002). *Ecological Modeling in Risk Assessment. Chemical Effects on Populations, Ecosystems and Landscapes*. Lewis Publishers, Boca Raton, USA.
- Pate-Cornell, M. E. (1996). Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering and System Safety*, 54:95–111.
- Patil, G. P., Boswell, M. T., Joshi, S. W., and Ratnaparkhi, M. V. (1984a). *Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work, Volume I: Discrete Models*. International Co-operative Publishing House, Maryland, USA.
- Patil, G. P., Boswell, M. T., Joshi, S. W., and Ratnaparkhi, M. V. (1984b). *Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work, Volume II: Continuous Univariate Models*. International Co-operative Publishing House, Maryland, USA.
- Patt, A. and Dessai, S. (2005). Communicating uncertainty: lessons learned and suggestions for climate change assessment. *Comptes Rendus Geosciences*, 337:425–441.
- Pearl, J. (1986). Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 29:241–288.
- Peters, G., Hosack, G., and Hayes, K. R. (2010). Non-linear state space model selection via adaptive particle markov chain monte carlo (adpmcmc). *Annals of Applied Statistics*, submitted.
- Peterson, D. P., Rieman, B. E., Dunham, J. B., Fausch, K. D., and Young, M. K. (2008). Analysis of trade-offs between threats of invasion by nonnative brook trout (*salvelinus fontinalis*) and intentional isolation for native westslope cutthroat trout (*oncorhynchus clarkii lewisi*). *Canadian Journal of Fisheries and Aquatic Sciences*, 65:557–573.
- Pollino, C. A., Woodberry, O., and Nicholson, A. E. (2007). Parameterisation of a bayesian network for use in ecological risk assessment. *Environmental Modelling and Software*, 22:1140–1152.
- Port of Melbourne Corporation, P. (2006). Supplementary environmental effects statement channel deepening project summary brochure. Technical report, Port of Melbourne Corporation, Melbourne, Australia.
- Pouillot, R. and Delignette-Muller, M. L. (2010). Evaluating variability and uncertainty seper-

- ately in microbial quantitative risk assessment using two r packages. *International Journal of Food Microbiology*, 142:330–340.
- Pouillot, R., Miconnet, N., Afchain, A. L., Delignette-Muller, M. L., Beaufort, A., Rosso, L., Denis, J. P., and Cornu, M. (2007). Quantitative risk assessment of listeria monocytogenes in french cold-smoked salmon: I. quantitative exposure assessment. *Risk Analysis*, 27:683–700.
- Puccia, C. and Levins, R. (1985). *Qualitative Modeling of Complex Systems: An Introduction to Loop Analysis and Time Averaging*. Harvard University Press, Cambridge, Massachusetts, USA.
- Ramsey, D. S. L. and Norbury, G. L. (2009). Predicting the unexpected: using a qualitative model of a new zealand dryland ecosystem to anticipate pest management outcomes. *Austral Ecology*, 34:409–421.
- Ramsey, D. S. L. and Veltman, C. (2005). Predicting the effects of perturbations on ecological communities: What can qualitative models offer? *Journal of Animal Ecology*, 74:905–916.
- Ranta, J., Tuominen, P., and Maijala, R. (2005). Estimation of true salmonella prevalence jointly in cattle herd and animal populations using bayesian hierarchical modeling. *Risk Analysis*, 25:23–37.
- Rasmussen, N. C. (1981). The application of probabilistic risk assessment techniques to energy technologies. *Annual Review of Energy*, 6:122–138.
- Reckhow, K. H. (1994). Water quality simulation modelling and uncertainty analysis for risk assessment and decision making. *Ecological Modelling*, 72:1–20.
- Reckhow, K. H. and Chapra, S. C. (1983). Confirmation of water quality models. *Ecological Modelling*, 20:113–133.
- Regan, H. M., Akcakaya, H. R., Ferson, S., Root, K. V., Carroll, S., and Ginzburg, L. R. (2003). Treatments of uncertainty and variability in ecological risk assessment of single-species populations. *Human and Ecological Risk Assessment*, 9:889–906.
- Regan, H. M., Colyvan, M., and Burgman, M. A. (2002a). A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications*, 12:618–628.
- Regan, H. M., Ferson, S., and Berleant, D. (2004). Equivalence of methods for uncertainty propagation of real-valued random variables. *International Journal of Approximate Reasoning*, 36:1–30.
- Regan, H. M., Hope, B. K., and Ferson, S. (2002b). Analysis and portrayal of uncertainty in a food-web exposure model. *Human and Ecological Risk Assessment*, 8:1757–1777.
- Reichart, P. and Omlin, M. (1997). On the usefulness of over parameterised models. *Ecological Modelling*, 95:289–299.
- Renooij, S. (2001). Probability elicitation for belief networks: issues to consider. *The Knowledge Engineering Review*, 16:255–269.
- Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer, New York, USA.

- Robert, C. P. and Casella, G. (2010). *Introducing Monte Carlo Methods with R*. Springer, New York.
- Romanowicz, R. J. and Beven, K. J. (2006). Comments on generalised likelihood uncertainty estimation. *Reliability Engineering and System Safety*, 91:1315–1321.
- Rout, T. M., Thompson, C. J., and McCarthy, M. J. (2009). Robust decisions for declaring eradication of invasive species. *Journal of Applied Ecology*, 46:782–786.
- Rowe, W. D. (1977). *An Anatomy of Risk*. John Wiley and Sons, New York, USA.
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Cambridge University Press, Cambridge, England.
- Ryan, L. (2008). Combining data from multiple sources, with applications to environmental risk assessment. *Statistics in Medicine*, 27:698–710.
- Rykiel, E. J. (1996). Testing ecological models: The meaning of validation. *Ecological Modelling*, 90:229–244.
- Sainsbury, K. J., Punt, A. E., and Smith, A. D. M. (2000). Design of operational management strategies for achieving fishery ecosystem objectives. *ICES Journal of Marine Science*, 57:731–741.
- Scheuer, E. M. and Stoller, D. S. (1962). On the generation of normal random vectors. *Technometrics*, 4:278–281.
- Schnute, J. T. (1994). A general framework for developing sequential fisheries models. *Canadian Journal of Fisheries and Aquatic Sciences*, 51:1676–1688.
- Sebastini, P., Mandl, K. D., Szolovits, P., Kohane, I. S., and Ramoni, M. F. (2006). A bayesian dynamic model for influenza surveillance. *Statistics in Medicine*, 25:1803–1816.
- Shipley, B. (2000). *Cause and correlation in biology*. Cambridge University Press, Cambridge, England.
- Shlyakhter, A. I. (1992). Sea-level rise or fall. *Nature*, 357:25.
- Shlyakhter, A. I. (1994). An improved framework for uncertainty analysis: Accounting for unsuspected errors. *Risk Analysis*, 14:441–447.
- Siau, K. and Tan, X. (2005). Improving the quality of conceptual modeling using cognitive mapping techniques. *Data and Knowledge Engineering*, 55:343–365.
- Sisson, S. A. (2005). Transdimensional markov chains: A decade of progress and future perspectives. *Journal of the American Statistical Association*, 100:1077–1089.
- Slob, W. (1994). Uncertainty analysis in multiplicative models. *Risk Analysis*, 14:571–576.
- Smith, A. D. M., Hobday, A. J., Webb, H., Daley, R., Wayte, S., Bulman, C., Dowdney, J., Williams, A., Sporcic, M., Dambacher, J., Fuller, M., Furlani, S., Griffiths, S., Kenyon, R., and Walker, T. (2007). Ecological risk assessment for the effects of fishing. Technical report, Final report R04/1072 for the Australian Fisheries Management Authority, CSIRO Marine and Atmospheric Research, Hobart, Australia.

- Smith, A. E., Ryan, P. B., and Evans, J. S. (1992). The effect of neglecting correlations when propagating uncertainty and estimating the population distribution of risk. *Risk Analysis*, 12:467–474.
- Smith, E. (2002). *Encyclopaedia of Environmetrics*, chapter Uncertainty analysis, pages 2283–2297. John Wiley and Sons Ltd., Chichester, England.
- Smith, E. P. and Shugart, H. H. (1994). *Ecological Risk Assessment Issue Papers*, chapter Issue paper on uncertainty in ecological risk assessment, pages 8.1 – 8.53. United States Environmental Protection Agency, Office of Research and Development, Washington DC, USA.
- Sniedovich, M. (2007). The art and science of modeling decision-making under severe uncertainty. *Decision Making in Manufacturing and Services*, 1:111–136.
- Sniedovich, M. (2008). From shakespere to wald: Modeling worst-case analysis in the face of severe uncertainty. *Decision Point*, 22:8–9.
- Sniedovich, M. (2010). A bird's eye view of info-gap decision theory. *The Journal of Risk Finance*, 11:268–283.
- Society, T. R. (1983). Risk assessment: Report of a royal society study group. Technical report, The Royal Society, London, England.
- Sokal, R. R. and Rohlf, J. (1995). *Biometry*. WH Freeman and Company, New York, USA.
- Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., and Cowell, R. G. (1993). Bayesian analysis in expert systems. *Statistical Science*, 8:219–283.
- Spiers-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., and Burgman, M. A. (2010). Reducing overconfidence in the interval judgements of experts. *Risk Analysis*, 30:512–523.
- Stephens, M. E., Goodwin, B. W., and Andres, T. H. (1993). Derviving parameter probability density functions. *Reliability Engineering and System Safety*, 42:271–291.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103:677–680.
- Stewart, T. R. (1990). A decomposition of the correlation coefficient and its use in analysing forecasting skill. *Weather and forecasting*, 5:661–666.
- Stone, M. (1961). The opinion pool. *Annals of Mathematical Statistics*, 32:1339 – 1342.
- Thompson, K. M. (1999). Developing univariate distributions from data for risk analysis. *Human and Ecological Risk Assessment*, 5:755–783.
- Tucker, W. T. and Ferson, S. (2003). Probability bounds analysis in environmental risk assessment. Technical report, Applied Biomathematics, New York, USA.
- Turley, M. C. and Ford, D. E. (2009). Definition and calculation of uncertainty in ecological process models. *Ecological Modelling*, 220:1968–1983.
- USEPA, U. S. E. P. A. (1992). Framework for ecological risk assessment. Technical report, United States Environmental Protection Agency, Office of Research and Development, Washington DC, USA.

- Uusitalo, L. (2007). Advantages and challenges of bayesian networks in environmental modelling. *Ecological Modelling*, 203:312–318.
- van Straalen, N. M. (2002). *Species Sensitivity Distributions in Ecotoxicology*, chapter The theory of ecological risk assessment based on species sensitivity distributions, pages 37–102. Lewis Publishers, London, England.
- Velleman, P. F. and Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician*, 47:65–72.
- Vesely, W. E. and Rasmuson, D. M. (1984). Uncertainties in nuclear probabilistic risk analyses. *Risk Analysis*, 4:313–322.
- Vicari, A. S., Mokhtari, A., Morales, R. A., Jaykus, L. A., Frey, H. C., Slenning, B. D., and Cowen, P. (2007). Second-order modeling of variability and uncertainty in microbial hazard characterization. *Journal of Food Protection*, 70:363–372.
- Villars, M., de Vries, I., Bokhorst, M., Ferreira, J., Gellers-Barkman, S., Kelly-Gerreyn, B., Lancelot, C., Menesguen, A., Moll, A., Patsch, J., Radach, G., Skogen, M., Soiland, H., Svendsen, E., and Vested, H. (1998). Report of the asmo modelling workshop on eutrophication issues, 5<sup>th</sup> november 1996. Technical report, Netherlands Institute for Coastal and Marine Management, RIKZ, The Hague, The Netherlands.
- Vose, J. (2000). *Risk Analysis: A Quantitative Guide*. John Wiley and Sons, Chichester, England.
- Wachbroit, R. (1991). *Risk Assessment in Genetic Engineering*, chapter Describing risk, pages 368 – 377. McGraw-Hill, New York, USA.
- Walker, W. E., Harremoes, P., Rotmans, J., Van der Sluijs, J. P., Van Asselt, M. B. A., Janssen, P., and Kreyer von Krauss, M. P. (2003). Defining uncertainty. a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4:5–17.
- Wall, M. M., Boen, J., and R., T. (2001). An effective confidence interval for the mean with samples of size one and two. *The American Statistician*, 55:102–105.
- Walley, P. (1991). *Statistical reasoning with imprecise probability*. Chapman and Hall, London, England.
- Walley, P. (1996). Measure of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58.
- Walley, P. and Cooman, G. (2001). A behavioral model for linguistic uncertainty. *Information Sciences*, 134:1–37.
- Warren-Hicks, W. J., Parkhurst, B. R., and Butcher, J. B. (2002). *Species Sensitivity Distributions in Ecotoxicology*, chapter Methodology for aquatic ecological risk assessment, pages 345–382. Lewis Publishers, Boca Raton, USA.
- Welsh, A. H., Peterson, A. T., and Altmann, S. A. (1988). The fallacy of averages. *The American Naturalist*, 132:277–288.
- Wilcox, A. R. (1967). Indices of qualitative variation. Technical report, ORNL-TM-1919, Oak Ridge National Laboratory, Oak Ridge, USA.



- Willett, A. (1901). *The Economic Theory of Risk and Insurance*. University of Pennsylvania Press, Philadelphia, USA.
- Williamson, R. C. (1989). *Probabilistic Arithmetic*. PhD thesis, Thesis submitted for the degree of Doctor of Philosophy, Department of Electrical Engineering, University of Queensland.
- Williamson, R. C. and Downs, T. (1990). Probabilistic arithmetic i: Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, 4:89–158.
- Wintle, B. A., McCarthy, M. A., Volinsky, C. T., and Kavanagh, R. P. (2003). The use of bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, 17:1579–1590.
- Wooldridge, M. (2008). *Microbial Risk Analysis of Foods*, chapter Qualitative risk assessment, pages 1–28. ASM Press, Washington DC, USA.
- Wooldridge, S. and Done, T. (2004). Learning to predict large scale coral bleaching from past events: A bayesian approach using remotely sensed data, in situ data, and environmental proxies. *Coral Reefs*, 23:96–108.
- Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5:161–215.
- Wu, F. C. and Tang, Y. P. (2004). Second order monte carlo uncertainty/variability analysis using correlated model parameters: application to salmonid embryo survival in risk assessment. *Ecological Modelling*, 177:393–414.
- Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21:1–21.
- Yemshanov, D., Koch, F. H., Ben-Haim, Y., and Smith, W. D. (2010). Robustness of risk maps and survey networks to knowledge gaps about a new invasive pest. *Risk Analysis*, 30:261–276.
- Yemshanov, D., Koch, F. H., Ben-Haim, Y., and Smith, W. D. (in press). Detection capacity, information gaps and the design of surveillance programs for invasive forest pests. *Journal of Environmental Management*, pages 1–12.
- Yu, P. S., Yang, T. C., and Chen, S. J. (2001). Comparison of uncertainty analysis methods for a distributed rainfall-runoff model. *Journal of Hydrology*, 244:43–59.
- Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning - i. *Information Sciences*, 8:199–249.
- Zio, E. (2008). Risk informed regulation: Handling uncertainty for a rational management of safety. *Nuclear Engineering and Technology*, 40:327–348.

## Appendix A: Hypothetical elicitation and analysis

For variates that take values on the interval  $[0, 1]$  we fitted a Beta distribution function to each of the expert's second intervals by using a standard optimisation routine (the `optim` function in R). The optimisation routine finds the parameters  $(\alpha, \beta)$  of a Beta distribution function 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:

$$SS = [tq_1 - pbeta(p_1, \alpha, \beta)]^2 + [tq_2 - pbeta(p_1, \alpha, \beta)]^2$$

where  $tq$  is a vector of target quantiles (10th and 90th for an 80% confidence interval),  $p$  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.

For variates that take values on the real line (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) a standard distribution is assumed for the variate (e.g. beta or normal); and, (4) the distribution is symmetrical around the lower and upper bounds.

The distribution fitting procedure described above results 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:

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

where the subscript  $i$  represents each expert,  $n$  is the total number of experts,  $F_i(\theta)$  is the expert's distribution function with parameter vector  $\theta$ , and  $w_i$  is the weight 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 simple, intuitive and will faithfully capture the range of opinions expressed by each expert. The pooling procedure, however, results in a non-parametric distribution. This is not, however, a serious impediment to any subsequent calculations.

Having fitted and pooled the distributions for each question in the risk assessment model the analysis is in a position to compare the results of a Monte Carlo Simulation (MCS) with a Probability Bounds Analysis (PBA). For the MCS, the pooled distributions are sampled using a standard inverse distribution method (also known as the inverse transform sampling method). In this application 2000 samples were randomly selected from each of the pooled distribution as inputs into the risk assessment model. The subsequent risk estimates were collected, sorted and portrayed in an empirical distribution function. The Probability Bounds Analysis was performed using the S4 library developed by Scott Ferson, Applied Biomathematics, using the methods described in Ferson and Hajagos (2004), Tucker and Ferson (2003) and Ferson (2002).

## Appendix B: Press perturbation predictions

Section 4.3.2 presented an overview of loop analysis (qualitative modeling) and introduced the concept of a Signed Directed Graph (SDG). This appendix presents the mathematical details of a press perturbation analysis based on the system's community matrix, derived from the system of equations that are graphically represented by the SDG. Computer programs for these analyses can be found in the most recent revision of Supplement 1 of Dambacher et al. (2002) in Ecological Archives E083-022-S1 at <http://www.esapubs.org/archive/>.

Consider the growth rates of  $n$  interacting populations of species, described by the following system of linear differential equations

$$\frac{dN_i}{N_i dt} = \sum_{j=1}^n \alpha_{ij} N_j + \beta_i - \delta_i + \iota_i - \varepsilon_i \quad (i = 1, \dots, n) \quad (6.1)$$

where the per capita rate of change in population abundance of  $N_i$  is controlled by density-independent rates of birth ( $\beta_i$ ), death ( $\delta_i$ ), immigration ( $\iota_i$ ), and emigration ( $\varepsilon_i$ ), and  $\alpha_{ij}$  density-dependent interactions. Equation 6.1 can be generally stated as

$$\frac{dN_i}{N_i dt} = g_i(N_1, N_2, \dots, N_n; p_1, p_2, \dots, p_m), \quad (6.2)$$

where the growth function of each population ( $g_i$ ) is determined by the system's variables ( $N_i$ ) and the growth rate parameters  $\mathbf{p} = (\alpha_{ij}, \beta_i, \delta_i, \iota_i, \varepsilon_i)$

At equilibrium, population abundances ( $N_i$ ) are constant and defined by  $\frac{dN_i}{N_i dt}|_{N^*} = 0$ . In the neighborhood of an equilibrium point, density-dependent interactions, represented by the community matrix  $\mathbf{A}$  (Levins, 1968) (see below) determine the balance between the equilibrium population abundances, and the density-independent rates of growth, via the linear matrix equation  $\mathbf{A}\mathbf{N}^* = -\mathbf{k}$ , where  $\mathbf{k} = \beta_i - \delta_i + \iota_i - \varepsilon_i$ .

The elements of the community matrix  $\mathbf{A}$  are calculated as

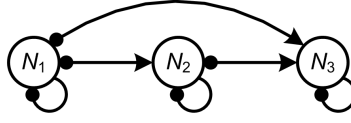
$$\alpha_{ij} = \frac{\partial \left( \frac{dN_i}{N_i dt} \right)}{\partial N_j} \Big|_{N^*} = \frac{\partial g_i}{\partial N_j} \Big|_{N^*}. \quad (6.3)$$

These elements define the relationships or direct effects between system variables. An alternative approach is to consider the system's Jacobian matrix ( $\mathbf{A}'$ ), which is derived from the  $dN_i/dt$  form of Equation 6.3

$$\alpha'_{ij} = \frac{\partial \left( \frac{dN_i}{dt} \right)}{\partial N_j} \Big|_{N^*} = \frac{\partial N_i g_i}{\partial N_j} \Big|_{N^*} \quad (6.4)$$

The sign structure of the community and Jacobian matrices are identical, and thus for the purpose of qualitative modeling, either one can be used. Quantitative stability analysis, however, requires the use of the Jacobian matrix, although calculations of perturbation response can proceed with either the community or Jacobian matrix.

Prior to stability and perturbation analysis, the relationships between variables are portrayed by sign digraphs. For example a simple three variable omnivory system can be portrayed in the following Sign Directed Graph The corresponding community matrix



$$\mathbf{A} = \alpha_{ij} = \begin{bmatrix} -a_{1,1} & -a_{1,2} & -a_{1,3} \\ +a_{2,1} & -a_{2,2} & -a_{2,3} \\ +a_{3,1} & +a_{3,2} & -a_{3,3} \end{bmatrix} \quad (6.5)$$

contains the equivalent information.

Any long-term impact on an ecosystem can be interpreted and evaluated as a sustained change in one of the system's  $p_m$  growth parameters. This is the case whether the change comes from within the system via a density-dependent parameter, as in Mendelian selection, or externally by way of a density-independent parameter, as in change coming from the environment or change from management, development, or experimental purpose. The predicted change in the equilibrium level of each variable is found (via the implicit function theorem) by differentiation of Equation 6.2 with respect to  $p_m$

$$\frac{\partial N^*}{\partial p_m} = -\mathbf{A}^{-1} \frac{\partial \mathbf{g}}{\partial p_m}. \quad (6.6)$$

Given the matrix equality

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \text{adj}(\mathbf{A}), \quad (6.7)$$

where “det” is the matrix determinant and “adj” is the adjoint matrix (or classic adjoint matrix), Equation 6.6 can be more conveniently expressed as

$$dN^* = \underbrace{\frac{1}{\det(-\mathbf{A})}}_{\text{overall feedback}} \underbrace{[\text{adj}(-\mathbf{A})]}_{\text{complementary feedback}} \underbrace{\left( \frac{\partial \mathbf{g}}{\partial p_m} \right) \partial p_m}_{\text{strength of perturbation}}. \quad (6.8)$$

Here we have, via Cramer's Rule, the solution for  $\partial N^*$  - the difference between the old and new equilibrium abundance for each population), and  $(\partial \mathbf{g} / \partial p_m) \partial p_m$  - the strength or magnitude of a given input or perturbation.

For our example system, the solution for the effects of a parameter change will be

$$\partial N^* = \frac{\begin{bmatrix} a_{1,1}a_{3,3} + a_{2,3}a_{3,2} & -a_{1,2}a_{3,3} - a_{1,3}a_{3,2} & a_{1,2}a_{2,3} - a_{1,3}a_{2,2} \\ a_{2,1}a_{3,3} - a_{2,3}a_{3,1} & a_{1,1}a_{3,3} + a_{3,1}a_{1,3} & -a_{1,1}a_{2,3} - a_{1,3}a_{2,1} \\ a_{2,1}a_{3,2} + a_{2,2}a_{3,1} & a_{1,1}a_{3,2} - a_{1,2}a_{3,1} & a_{1,1}a_{2,2} + a_{2,1}a_{1,2} \end{bmatrix}}{a_{1,1}a_{2,2}a_{3,3} + a_{1,1}a_{2,3}a_{3,2} + a_{2,1}a_{1,2}a_{3,3} + a_{2,1}a_{1,3}a_{3,2} + a_{3,1}a_{1,3}a_{2,2} + a_{3,1}a_{1,2}a_{2,3}} \left( \frac{\partial \mathbf{g}}{\partial p_m} \right) \partial p_m \quad (6.9)$$

By comparing Equations 6.8 and 6.9 it is evident that the matrix  $\mathbf{A}$  has two separate functions in determining a population's response to a perturbation. Through the adjoint of  $\mathbf{A}$ , all direct and indirect effects in the system are combined in complementary feedback cycles (Dambacher et al., 2002), which mediate the relative variation in the response of each population, whilst the determinant of  $\mathbf{A}$  (which constitutes the overall feedback of the system) scales the magnitude of each variable's response to a perturbation. For example, for a given input, if the overall feedback is relatively weak, then the effect of the complementary feedback cycles on population abundance will be relatively large. Stability analysis requirements (Dambacher et al., 2003a) show that the determinant of  $-\mathbf{A}$  must be positive and this requires that  $a_{3,1}a_{1,2}a_{2,3}$  to be relatively weak. Note that the use of the determinant of  $-\mathbf{A}$  maintains a sign convention in the adjoint matrix for even- and odd-sized systems (Dambacher et al., 2002).

The sign direction (increase, decrease or ambiguous response) of each of the individual populations to a system perturbation can be discerned from the adjoint matrix. A positive input to a variable-through an increase in birth rate or decrease in death rate-is read down a column, and response predictions for each variable are read along the rows. For example for a positive input to  $N_2$ , the prediction of decreased abundance of  $N_1$  is determined by two feedback cycles with negative sign - i.e.  $\text{adj}(-\mathbf{A})_{1,2} = -a_{1,2}a_{3,3} - a_{1,3}a_{3,2}$ , while the response of  $N_3$  is ambiguous because it is determined by two feedback cycles of opposing sign - i.e.  $\text{adj}(-\mathbf{A})_{3,2} = a_{1,1}a_{3,2} - a_{1,2}a_{3,1}$ . Note: where inputs to a variable are negative, through a decrease in birth rate or increase in death rate, then the signs of the adjoint matrix elements are simply reversed.

Ambiguous predictions from the adjoint matrix are interpreted via a technique of weighting the net number of feedback cycles to the absolute number in a response  $\bar{U}$  i.e. the weighted prediction for a response prediction is equal to the net number of feedback cycles divided by the total number of cycles (Dambacher et al., 2003b). For instance, the predicted response of  $N_3$  for an input to  $N_2$  is ambiguous because there are the same number of positive and negative feedback cycles. If, however, there were a total of four feedback cycles in a perturbation response, three of which were positive and one negative, then the net number of cycles would be two and the weighted prediction of the response would be  $2/4 = 0.5$ . The sign determinacy of responses with weighted predictions 0.5 has been shown to generally be  $> 90\%$  through simulations using random parameter space (Dambacher et al., 2003b). Below this threshold the sign determinacy of responses declines to zero for weighted predictions equal to zero. Hosack et al. (2008) show that this relationship is insensitive to different probability density functions assigned to the elements of the adjoint matrix.

In the analysis of linear systems, multiple inputs have an additive effect on the equilibrium of a variable through the superposition principle. Thus if there was simultaneous inputs that increased the birth rate of  $N_1$  and increased the death rate of  $N_2$  (for example by culling), then the predicted response of  $N_3$  is given by

$$dN_3^* = \frac{1}{\det(-\mathbf{A})} \left[ (a_{2,1}a_{3,2} + a_{2,2}a_{3,1}) \frac{\partial g_1}{\partial \beta_1} d\beta_1 - (a_{1,1}a_{3,2} - a_{1,2}a_{3,1}) \frac{\partial g_2}{\partial \delta_2} d\delta_2 \right] \quad (6.10)$$







### Contact Us

Phone: 1300 363 400  
+61 3 9545 2176

Email: [enquiries@csiro.au](mailto:enquiries@csiro.au)

Web: [www.csiro.au](http://www.csiro.au)

### Your CSIRO

Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills for building prosperity, growth, health and sustainability. It serves governments, industries, business and communities across the nation.