NOVEL APPROACHES TO PLANT PEST RISK ASSESSMENT

LIHONG ZHU

Thesis submitted in partial fulfilment of the requirements of the University of Greenwich for the Degree of Doctor of Philosophy

June 2009

AKNOWLEDGEMENTS

I must say, from the bottom of my heart, I could not have asked for better supervisors. I would like to express my genuine gratitude to my supervisors Dr John Holt and Dr Rob Black. It has been an honour to work with them. It has been a truly unique combination of supervisions I have received: Rob, with a brilliant insight into the regulatory and legal perspective of this work; John, with excellent knowledge in the quantitative aspect of this study. I am deeply grateful to John and Rob for their continuous encouragement and invaluable suggestions and support, their wisdom and knowledge. I also greatly appreciate the opportunities Rob and John provided throughout the course of this work, which enabled me to go to several international conferences and a number of beneficial visits to some world class organisations. They possess an exceptional ability of inspiration that make me feel positive even under the most difficult situation. Over the years, they and their families have been a significant presence in my life, to whom I know I could turn whenever I need.

Furthermore, I wish to express my cordial appreciation to Dr Richard Baker, of Central Science Laboratory (CSL), UK, for the constructive discussion of my work; Dr David McNamara and Mr Andrea Orlinski, of European Plant Protection Organisation (EPPO), for providing useful data and valuable discussion. My sincere thanks also go to Risk Analysts in CSL, EPPO and United States Department of Agriculture, Plant Quarantine officials in Africa, who provided answers to my questionnaire.

This work was funded by the Higher Education Funding Council for England, through the University of Greenwich.

My sincere thanks are extended to my colleagues and friends at the Natural Resources Institute, for their support during my time there. Special gratitude goes to Mrs Chanban Black and Dr Susan Seal, for their long-lasting friendship, encouragement and understanding.

I owe everything to my mother and late father, without whom, I won't be the person I have become today. I thank them for their love and their high expectation of me. My parents taught me high principles and moral standards, which I benefit forever. My thanks also go to my sister and brother, for their love and unwavering support. I also owe my late parents-in-law much for their love and support, who helped so much in raising my son Beichen when my husband and I were both working on our PhDs.

My children Beichen and Helen award me such joy and motivate me to complete this thesis. I am deeply sorry for the compromises at times.

I am forever indebted to my husband Qing Gong, who has been my intellectual and emotional sounding board, who is always there to listen to me, inspires me and offers valuable ideas. I could not have done this without his endless love, encouragement and patience.

No word can express my deep love for my beloved family. To them I dedicate this thesis.

NOVEL APPROACHES TO PLANT PEST RISK ASSESSMENT

LIHONG ZHU

ABSTRACT

Pest risk assessment is an essential yet problematic stage in pest risk analysis (PRA) that concerns the likelihood and consequences of pest introduction. The aim of this study was to develop methodologies for risk assessment and to explore different approaches that could lead to the development of new methods for practical PRA in line with the requirement of "scientific justification" by World Trade Organisation and Food and Agriculture Organisation of the United Nations.

Current international practices were discussed and research reviewed on qualitative and quantitative approaches to risk assessment.

It was proposed that risk assessment be divided into two steps: Pest risk identification (PRI) and pest risk evaluation (PRE). Mind Mapping was a valuable tool for PRI that reduced ambiguity and increased transparency.

Approaches to PRE were proposed that facilitated the scoring and weighting of risk factors, and the subsequent combining of risk scores.

Several methods were developed to incorporate weighting into PRA, which included subjectively assigned weighting and Delphi technique-derived weighting.

Metrics for combining risk scores into an overall risk value were also explored, compared and evaluated.

Correlation and interaction between risk factors were analysed, which revealed that some risk factors were highly correlated and some were relatively independent, which meant there was some information redundancy, and therefore simplification of risk assessment was possible.

Cluster analysis was applied to risk factor scores and different clusters of risk factors were identified: some more appropriate for preliminary assessment; some for determining the level of risk; and some could be eliminated.

A method to apply Principal Components Analysis (PCA) to derive weighting for individual risk factors was developed. PCA could be applied to historical data of pest introductions, previous PRA cases, or expert opinion.

Genetic algorithms implemented in the software BEAGLE, were applied to PRA data. The rules obtained could distinguish high-risk situations with high accuracy, which was useful in predicting the risk of an organism by using a simplified set of conditions.

The results showed that weightings and rules differed for different taxonomic groups. Therefore it was implausible to develop a generic scheme in this way. However, it may be possible to develop patterns based on taxonomy.

The results of applying several different techniques all suggested that by grouping risk factors for different purposes, risk assessment could be simplified without compromising rigor, because a) some factors were redundant; b) some factors are more important than others; and c) high risk situation could be predicted with a few key factors.

Key words: PRA, risk assessment, plant health, mind mapping, weighting, correlation, Delphi study, correlation, cluster analysis, principal components analysis, machine leaning, risk assessment simplification.

ABBREVIATIONS/ACRONYMS

ALB	Asian longhorned beetle
ALR	Acceptable level of risk
APHIS	The Animal and Plant Health Inspection Service, USA
APPPC	Asia and Pacific Plant Protection Commission
BEAGLE	Evolutionary Algorithm Generating Logical Expression
CEPM	The Committee of Experts on Phytosanitary Measures
CFIA	The Canadian Food Inspection Agency. Canada
CPC	Crop protection compendium
CSL	Central Science Laboratory. UK
DEFRA	Department for Environment, Food and Rural Affairs. UK
EPPO	European and Mediterranean Plant Protection Organisation
EU	The European Union
FAO	Food and Agriculture Organisation of the United Nations
GA	Genetic Algorithm
GATT	General Agreement on Tariffs and Trade
ICPM	Interim Commission on Phytosanitary Measures
IPM	Integrated pest management
IPPC	International Plant Protection Convention
ISPM	International Standards for Phytosanitary Measures
IUCN	The International Union for Conservation of Nature
LMO	Living modified organisms
NAPPO	North American Plant Protection Organisation

NPPO	National Plant Protection Organisation
OIE	The International Office of Epizootics
PRA	Pest Risk Analysis
PC	Principal component
PCA	Principal components analysis
PHRAU	Plant health risk assessment unit, Canada.
PHSI	The Plant Health and Seeds Inspectorate, UK
PRE	Pest risk evaluation
PRI	Pest risk identification
QP	Quarantine pest
RPPO	Regional Plant Protection Organisations
SPS Agreement	Agreement on the Application of Sanitary and Phytosanitary Measures
USDA	The United States Department of Agriculture
WTO	World Trade Organisation

SYNOPSIS

The last two decades have been an era of globalisation. Growing international trade and tourism have greatly increased the movements of goods and people around the world, and have been associated with more introductions of the plant pests into new areas, and have resulted in substantial economic, environmental, and ecosystem damage (Pimentelo *et al.*, 2000, 2001; Orwig, 2002). For example, the invasion of *Bemisia tabaci* worldwide has had profound consequences. *B. tabaci* was described over 100 years ago and has since become one of the most important pests worldwide in subtropical and tropical agriculture as well as in greenhouse production systems. It adapts easily to new host plants and geographical regions and has now been reported from all global continents except Antarctica. In the last decade, international transport of plant material and people has contributed to its geographical spread. *B. tabaci* has been recorded from more than 600 plant species and there may be many additional hosts not yet formally documented (Henneberry *et al.*, 2001).

Two conflicting problems in the international trade in plant and plant products are (a) how to ensure that the spread and introduction of pests of plants and animals are being prevented? (b) how to ensure that consequently strict health and safety regulations are not being used as an excuse for protecting domestic producers from competition?

There are two international treaties regarding international trade versus plant protection that approach the problem from opposite directions. The Agreement on the Application of Sanitary and Phytosanitary Measures (SPS Agreement) (WTO, 1995) aims at free trade whilst recognising the need " to protect ... plant life or health" (WTO 1995). The International Plant Protection Convention (IPPC) has "the purpose ... to prevent the spread and introduction of pests of plants"[1] (IPPC, 1997).

In response to the demands of the international trade rules, pest risk analysis (PRA) was introduced to bridge these two conflicting issues. PRA is a structured decision making process of "evaluating biological or other scientific and economic evidence to determine whether a pest should be regulated and the strength of any phytosanitary measures to be taken against it" (IPPC, 1995a). One objective of PRA is to assess the potential risk and impact arising from pest introduction and spread.

It is now widely accepted that pest risk analysis and the subsequent plant quarantine decision-making procedures should be both scientific based and transparent (WTO 1995). In short, phytosanitary measures should be based on PRA to reduce a) the pest introduction to a new area, b) unnecessary barriers to international trade.

Pest risk analysis comprises pest risk initiation, pest risk assessment and pest risk management (IPPC, 1995a).

---

[1] Originally this was so (IPPC 1951) but the 1997 revised text of IPPC re-aligns things to ensure that plant quarantine measures are applied consistent with free trade

Pest risk assessment is a technique for identifying, characterising, quantifying and evaluating hazards. Irrespective of the application, risk assessment seeks to answer the following questions: a) what can go wrong? b) how likely is it to happen? and c) if it happens, what consequences are expected? (Oryang, 2002). In the wider context of risk analysis, a further question should be resolved: how to manage (eliminate or reduce) the hazard to an acceptable level? A good risk analysis should be convincing, scientifically justified and transparent, and document any areas of uncertainty for further review.

In common with risk assessment in some other disciplines, a number of problems must be overcome, e.g. subjectivity, uncertainty, non-quantifiable variables, and the need to integrate information into a simple statement of risk.

The international framework for PRA (ISPM No. 11) lacks methodology for (a) handling economic and social criteria and to integrate these with biological criteria; (b) combining the assessments under different criteria into a meaningful overall risk score; and (c) coping with uncertainty throughout the whole assessment. The risk criteria recognised in the International Standards for Sanitary and Phytosanitary Measures (ISPMs) (IPPC, 1995a) include geographical and regulatory criteria, introduction and/or spread potential and economic consequences. The lack of specific guidance hampers detailed examination of risk criteria with the result that some important risk elements may be overlooked.

Considering the problems involved in PRA, the general aims of this thesis are to:

- Explore the rationale of risk assessment structure in plant quarantine decision-making.
- Explore the development of new methods for practical PRA in line with the requirement of "scientific justification" by WTO and FAO[2].

Within these broad aims, certain specific objectives are addressed in the following chapters:

- To develop unambiguous and consistent criteria for risk assessment, i.e. the identification of pest risks;
- to explore the relationships among risk factors;
- to explore the possibility of incorporating weighting into pest risk assessment;
- to explore the possibility of simplification of risk assessment;
- to examine the degree of generality in risk assessment criteria for PRA.

To set the scene for the thesis, an outline is now provided of the work described in the succeeding chapters.

Chapter 1 provides a general context for pest risk analysis and its legislative and economic background. The concepts of pest, risk, risk analysis, plant

---

[2] ISPM No11 (IPPC 2004) now concerns invasive species and living modified organisms as well as quarantine pests in the strict sense, but the latter will be the main focus of this Thesis.

quarantine, international trade, are discussed in this chapter; the requirements of the SPS agreement and IPPC are also presented here which sets PRA in the full context of plant protection and international trade.

Chapter 2 reviews the current international standards and methods for PRA; regional PRA schemes and national PRA practices are discussed as examples, and points out the problems involved in PRA and sets out the aim and objectives of this study.

Chapter 3 sets out in a general way the issues that may arise in any situation where PRA maybe appropriate. The author looks comprehensively at the components of PRA and those issues arising that need to be considered in PRA practice. A structure for pest risk assessment process is proposed as follows. *Risk identification*: Mind Mapping is used to identify the risk factors involved in a risk assessment which take into account the geographical and regulatory criteria; introduction potential which concerns entry, establishment and spread probability; and the consequences on economic, environment and society. *Risk evaluation*: discusses possible ways to obtain an overall risk estimation, which consists of scoring, weighting of the risk factors and combining the risk scores. Some case studies are used to illustrate the process.

Chapter 4 elaborates some of the issues identified in Chapter 3, in particular how to derive weightings, the relationships among risk factors and how to define high risk situations. A variety of statistical techniques are used. These issues concern: a) the derivation of weightings from historical PRA data; b) the possibility of generalisation of risk assessment by looking at data redundancy and evolving key rules for high risk situation; c) discriminating risk factors for risk assessment, and d) subjectivity ('expert judgment') in risk assessment.

Chapter 5 concludes this study. General discussion and future directions of PRA research are considered in this chapter.

Table of Contents

CHAPTER 1 PEST RISK AND PEST RISK ANALYSIS

Section 1.1 explains the concepts of plant pests and alien species and then explores the threat of plant pest introduction posed by the growth in international trade, identifies the pathways for the plant pest introduction, and discusses the economic and environment consequences caused by the introduced plant pests. Section 1.2 defines the concept of plant quarantine. Section 1.3 introduces the concept of risk analysis and risk factors, and defines the (plant) pest risk analysis (PRA) process as a component of decision-making. Section 1.4 explores the importance of pest risk analysis in international trade as required in various trade-related international agreements. Section 1.5 discusses the components of PRA and Section 1.6 concludes the chapter.

1.1 PLANT PEST, ALIEN SPECIES, QUARANTINE PEST AND PEST INTRODUCTION

1.1.1 Plant pest, alien species and quarantine pest

A plant pest is an organism, which is injurious to plants or plant products; it could be any species, strain, or biotype of plant, animal or pathogenic agent (IPPC, 1995b). The US Plant Protection Act has a similar definition, defining plant pest as "any living stage of any of the following that can directly or indirectly injure, cause damage to, or cause disease in any plant or plant product: (A) A protozoan. (B) A nonhuman animal. (C) A parasitic plant. (D) A bacterium. (E) A fungus. (F) A virus or viroid. (G) An infectious agent or other pathogen. (H) Any article similar to or allied with any of the articles specified in the preceding subparagraphs" (Anon, 2000).

Plant pests can cause enormous damage to agriculture and the environment (Pimentel *et al.*, 2000, 2001, 2005). The damage can be economic through lost output, income, investment, and the cost of control as well as psychological shock and panic, unemployment and other social ills. As the presence of the pests in one location poses a threat to an adjacent location or even a distant location, the damage can be social as well as environmental (Pimentel *et al.*, 2000, 2001, 2005; Orwig, 2002).

Plant pests pose the greatest immediate threat when they move as plagues or when they are introduced for the first time into ecologically favourable conditions where there are few natural factors to limit their spread and people do not have experience in managing them. Such occurrences often have the most significant impacts.

Alien species is a synonym for exotic species, foreign species, non-indigenous species, and non-native species. According to the USA Executive Order 13112, "Alien species" is, "... with respect to a particular ecosystem, any species, including its seeds, eggs, spores, or other biological material capable of propagating that species, that is not native to that ecosystem"[1] (Anon, 1999). IUCN (2000) defines "alien species" as "a species, subspecies,

---

[1] "Ecosystem" means the complex of a community of organisms and its environment (USA Executive Order 13112 on Invasive Species).

or lower taxon occurring outside of its natural range (past or present) and dispersal potential (*i.e.* outside the range it occupies naturally or could not occupy without direct or indirect introduction or care by humans) and includes any part, gametes or propagule of such species that might survive and subsequently reproduce" (IUCN, 2000).

Many alien species have disastrous effects on native species (Porter and Savignano, 1990; Kizlinski *et al.*, 2002). Their populations may grow out of control, and they are then called invasive alien species[2], therefore "Invasive species" means an alien species whose introduction does or is likely to cause economic or environmental harm or harm to human health (Anon, 1999).

A quarantine pest is a pest of potential economic importance to the area endangered thereby and not yet present there, or present but not widely distributed and being officially controlled (IPPC, 1995b).

1.1.2 Global threats of pest introduction

The movement of plants and plant products from one area to another always involves the risk of introducing plant pests and diseases into new areas especially in the current era of trade liberalisation. The unnatural introduction of species into new environments is the major reason of ecosystem destruction and species extinction. The plants at risk may be crops, forest

---

[2] In the context of conservation, invasive alien species can also be defined as "A species outside of its native range that threatens the survival or reproduction of native plants or animals or threatens to reduce biological diversity" (North Carolina Division of Parks and Recreation Department of Environment and Natural Resources 1998).

trees or ornamentals, not necessarily related to the original import. Introduced organisms have also had disastrous effects on the diversity of ecological systems (Black and Sweetmore, 1995a, b, c).

One example is the introduction and the subsequent spread of chestnut blight (*Cryphonectria parasitica*) in North America in 1904. This fungal disease effectively eliminated overstory chestnut (*Castanea dentate*) from forests and has dramatically altered the species composition of eastern North American forests (Liebhold *et al.*, 1995).

More and more cases of new pest introductions, which have caused severe damages to crop production and environment, are seen all the time despite the efforts of strict phytosanitary regulations (Pimentel *et al.*, 2000, 2001, 2005).

International trade is the main reason of pest introduction (Simberloff, 1996; Zhu *et al.*, 2000; Smith *et al.*, 2007). The growing trade in products of agriculture, forestry, horticulture and aquaculture, e.g. fresh fruits and vegetables, has significantly increased the chance of exotic pests being introduced to a new area.

The global traffic network has further enabled new pests to spread to the new areas through hitchhiking[3], in that non-native insects travel to places where they've never existed before by moving with cargo, in baggage, or at large in

---

[3] A pest that is carried by a commodity and, in the case of plants and plant products, does not infest those plants or plant products [CEPM, 1996; revised CEPM, 1999]

carriers instead of their host material. Gypsy moth has been frequently intercepted in New Zealand on used vehicles imported from Japan. Another example is the outbreaks of Asian longhorned beetle in North America; it is believed that it arrived there in the wooden packing material used in cargo shipments from China.

The second major reason for pest establishment is climate change due to global warming and the increase in protected crops. Climate change may enable some pests from warmer climates to survive and become established in some areas that used to be less favourable (Simberloff, 2000b). The same applies to protected or sheltered crops, greenhouses providing suitable mini-environments to some pests that historically are unlikely to survive and establish in such regions.

Coakley *et al.* (1999) reviewed the research on climate change and plant disease management. Results indicated that climate change could alter stages and rates of development of the pathogen, modify host resistance, and result in changes in the physiology of host-pathogen interactions. The most likely consequences are shifts in the geographical distribution of host and pathogen and altered crop losses, caused in part by changes in the efficacy of control strategies (Coakley *et al.*, 1999).

An example of new pest establishment due to climate change is the new report of *Diplodia pinea* (Desmaz.) in Estonia. *D. pinea* was found in Estonia for the first time in the autumn of 2007. It is believed that hard droughts registered in Estonia in 2002 and 2006, following the general trend of climatic change, is the cause of this pathogen from southern Europe to the north. In Central Europe hard drought encouraged *D. pinea* to become epidemic in 2003 (Hanso and Drenkhan, 2009).

Another cause of increase in pest introduction is the growth of tourism. It is not only the intensive global traffic and trade, which provides a major means of pest entry; passengers may smuggle fresh plant produce and spread pest infestation elsewhere. Containers, packaging materials, airline food, etc., could also harbour pests and transport them to any destination.

1.1.3 Pathways for plant pest introduction

Pathways for plant pest introduction are the means by which a species is moved from one location to another. There are many such pathways. Alien species have sometimes entered into a new area naturally. For example, a large proportion of the insect species in Southern Florida arrived by flight, assisted by winds, from the West Indies, the Bahamas, and the Yucatan peninsula of Mexico (Frank and McCoy, 1995).

Other alien species introductions were either created or enhanced by human activities, which are basically of two types. The first type is intentional, which is the result of a deliberate action to move an organism from one location to another. Examples of intentional introductions include the intended movement of living seeds, whole plants, or pets, including the introduction of biological control agents, and intended trade of some species.

The second type is unintentional, which is the result of unintended human activities. In a study on the origins and pathways of 325 non-native invertebrate plant pests established in Great Britain between 1787 and 2004, Smith *et al.* (2007) found that 67.8% of 101 non-native invertebrate plant pests, for which sufficient information exists, were introduced unintentionally. Of the post-1970 species posing a significant pest risk to cultivated hosts, 43.6% were introduced unintentionally by human being, compared to 5.5% that established naturally. In the US, some of the most serious introduced plant pests, such as chestnut blight and white pine blister rust were introduced to North America with imported nursery stock. Examples of unintentional pathways are soil associated with the trade of nursery stock, importation of fruits and vegetables (e.g. plant pests), and the international movement of people (e.g. pathogens). In these and countless other unintentional pathways, the movement of species is an indirect result of human activities.

Human activity-facilitated plant pest introduction is increasingly seen as a worldwide problem. In particular, the trade in agriculture, forestry and aquaculture etc., which has increased immensely in the last few decades, is considered a major pathway of pest introduction. This greater than ever international trade has dramatically increased the chance of new species being introduced into new areas. These introductions have been caused by movement with commodities, vehicles, passengers, mail, etc.

Air transportation, water transportation, and land transportation, in particular air transportation, provide efficient ways of travel for alien species. For example, mosquitoes have survived flights from Africa to Britain in passenger cabins (Gratz *et al.*, 2000), and snakes have travelled in cargo bays from Guam to Hawaii (Claiborne, 1997).

The use of containers, the huge metal boxes that are stacked up on ships and off-loaded directly on to trains or trucks, has provided a "quantum leap" in the efficiency of transportation, both for trade goods and for exotic animals and plants. Previously, seaports were the routes of entry for many exotics, but with container transport, the biological invaders are picked up and delivered directly to inland destinations all over the world. Containers provide a sheltered environment; they sit for weeks waiting to be loaded or unloaded, giving hitchhikers[4] plenty of time to embark or disembark; and they are difficult for customs and quarantine inspectors to search thoroughly (Bryant, 2002). It is believed that container shipments of used tires from Japan brought the Asian tiger mosquito (*Aedes albopictus*) to the U.S., South Africa, New Zealand, Australia and Southern Europe (Novak, 1992).

On some occasions, the pathway of a new pest introduction cannot be easily identified. An example is a fruit fly outbreak in Mauritius in 1996. It was thought that the species was *Bactrocera dorsalis* from India, carried in airline

---

[4] Hitchhiker is also called a contaminating pest, which is carried by a commodity and, in the case of plants and plant products, does not infest those plants or plant products [CEPM, 1996; revised CEPM, 1999]

passenger food from Delhi. However, several years later, after eradication, the species was re-identified as *B. invadens*, originally from Sri Lanka but likely to have been introduced from the African mainland. The pathway has not been identified but is more likely to be commercial fruit imports (Black 2009 pers. comm.).

1.1.4 Impacts of alien species

Alien species affect indigenous species in a number of ways: direct predation (including herbivory), competition for resources (food and territories), habitat alteration or degradation (e.g. shading out of native plant species by species such as *Rhododendron*; toxicity of plant breakdown products inhibiting growth of native species; destruction); spread of disease (e.g. crayfish plague spread by American signal crayfish (*Pacifastacus lenisculus*), Dutch elm disease; and genetic pollution, e.g. hybridisation between introduced Sika deer (*Cervus Nippon*) and native red deer (*Cervus elaphus*) (DEFRA 2003).

According to Pegg (2003), more than 50,000 non-indigenous invasive species are estimated to cost the United States some $138 billion annually in damage, losses, and pest control, and are forever changing a variety of the nation's ecosystems (Pimentel *et al.*, 2000). Globally, the cost of introduced non-indigenous invasive species is estimated to be around US$1.4 trillion per annum (Pimentel *et al.*, 2001).

Besides the easily recognised direct damage to agriculture, economic and environmental forest species, more difficult to calculate is the habitat destruction and biodiversity impact, especially that by invasive alien species. Aquaculture, ornamentals and amenity species, natural vegetation and animals can also be affected by introduced species. For example, in the United States, approximately 400 of the 958 species that are listed under the Endangered Species Act are considered at risk from competition from non-indigenous invasive species (Wilcove *et al.*, 1998).

1.2 PLANT QUARANTINE AND SUSTANABLE DEVELOPMENT

FAO defines "plant quarantine" as "all activities designed to prevent the introduction and/or spread of quarantine pests or to ensure their official control" (IPPCD, 1995b). The purpose of plant quarantine is to reduce or prevent the risk of pest introduction and to control the spread of pests and diseases when outbreaks do occur. Plant quarantine measures were first put in place in reaction to catastrophes caused by the introduction of pests from other parts of the world. This was classically the case for the measures established after the introduction of *Viteus vitifoliae* (grapevine phylloxera) from North America into Europe at the end of the 19th century (Smith, 2000). Another good example of measures established is the measures against Colorado beetle introduced into the UK, which enabled the UK to protect against a pest which is common elsewhere in the EU; the measure is still effective (Black, 2003).

The basis of plant quarantine is the phytosanitary or plant protection legislations of individual countries – it is therefore regarded as regulatory control of plant pests. Plant quarantine has an important role in the protection of natural resources and environment. It consists of preventative actions to prevent the undesirable consequences of pest spread, pest establishment, and the associated significant economic damage. It can also avoid the use of pesticides to control or eradicate the introduced pests. It maybe cost effective even just by delaying an inevitable pest introduction[5]. It benefits the whole society rather than the affected area or farmers only, as it reduces the chances of the introduced pests to spread from the one area to another and reduce the costs for the whole society to control or eradicate the introduced pests.

Plant quarantine may therefore be regarded as a component of the Integrated Pest Management (IPM) strategy and has been shown to be very cost-effective in developed countries for reducing the risk of pest introduction, and is likely to be even more so in the often less-developed countries of the tropics and sub-tropics (Black and Sweetmore, 1995b).

---

[5] Something is a quarantine pest only if it poses a threat, by definition. As long as the threat is there, the only consideration is the cost-effectiveness of measures. The general impression is that without plant quarantine, the pest threat would almost certainly materialise, given the way a quarantine pest is defined – can be introduced into a new area by human activities, and most importantly, cause significant economic damage.

1.3 RISK ANALYSIS AS A DECISION-MAKING TOOL

1.3.1 The concept of risk and risk factors

The word *Risk* has been widely used for different purposes by different researchers. In its simplest form, risk is the likelihood of an adverse effect. For example, the Concise Oxford Dictionary of Current English (1999) defines "risk" as "the chance or possibility of loss or bad consequence". A more general concept of risk is the likelihood and the degree of an adverse effect. For example, Kaplan and Garrick (1981) and Haimes (2004) described *risk* as the probability of loss or injury and the degree of probability of such loss. Risk is therefore a combination of the probability, or frequency, of occurrence of a defined adverse effect and the magnitude of the consequences of the occurrence (Warner, 1992). Another fundamental concept used in risk analysis is *hazard*, which means source of danger or risk according to the Concise Oxford Dictionary of Current English (1999). For example, *hazard* is used to refer to an organism that potentially may cause unwanted consequences in New Zealand biosecurity risk analysis procedure.

Risk factor is another important concept in risk analysis. The term *risk factor* was first coined by heart researcher Dr. Thomas R. Dawber and his collaborators in a series of papers in 1961, where they attributed heart disease to specific conditions (blood pressure, cholesterol, smoking), which he defined as *risk factors* (Dawber and Kannel, 1961; Kannel *et al.*, 1961). A *risk factor* is a variable associated with an increased risk of adverse effect. Risk factors are correlated with, but not necessarily causal to the risk.

1.3.2 The understanding of the term *risk analysis*

Risk analysis was proposed originally by David Hertz (Hertz, 1964), who regarded it as a natural and logical extension of the sensitivity analysis approach (Rappaport, 1967) in investment risk management. Sensitivity analysis is used in many areas of problem solving, to understand the effects of uncertainty, by illustrating the range of outcomes that can occur. By using risk analysis, managers can contemplate and confront the future uncertain environment in which they operate.

The term *risk analysis* denotes methods that aim to develop a comprehensive understanding and awareness of the risk associated with a particular event of interest. Decision and risk analysis can be viewed as having two major roles. Firstly, they offer a broad perspective for structuring the process of decision-making, and secondly, they provide a set of techniques for evaluating the worth of alternative decision options. Both analytic approaches involve decomposing and structuring the problem, assessing the uncertainty and values of the possible outcomes, and determining the preferred strategy in terms of some specified choice criteria (Hertz and Thomas, 1983).

Risk analysis provides a tool to decision-making. The contribution which risk analysis can make is to help decision-makers' thinking processes, and this is done in the first instance by forcing them to confront the structure of the decision problem in a relatively unemotional and objective manner. It enables the decision-maker to examine, discuss, and eventually understand why one course of action might be more desirable than other alternatives.

1.3.3 Pest risk and pest risk factors

Pest risk is often defined as the economic and environmental harms to agriculture, environment, and citizens that would be caused by pest introduction in an area where the pest does not already occur.

The term *Pest risk factor* is used throughout this thesis to refer to any event or pest characteristic that can result in an introduction and spread of a plant pest.

Contributing pest risk factors can be conveniently grouped into seven broad categories (Black and Abdallah, 1997; IPPC, 2001):

- *Pathway*: Any means that allows the entry or spread of a plant pest
- *trading partner*: The country/region where the pest occurs
- *host range*: A host is an organism that harbours a virus, parasite, insects, or other fauna. A host typically provides nourishment and/or shelter. A host range is a collection of hosts a pest can utilize
- *dispersal potential:* The ability to spread naturally
- *climate suitability*: The climate suitability for a potential plant pest to establish and spread in the PRA area

- *economic impact*: The potential economic damages that may arise from the pest introduction; and
- *environmental impact*: The potential environmental and ecological damages that may arise from the pest introduction.

1.3.4 Pest risk analysis and plant quarantine decision-making

Historically, there was no unified definition of pest risk analysis. The SPS Agreement does not refer to "risk analysis", but uses the term "risk assessment" in a general way. The secretariat of the IPPC uses "risk assessment" to describe a component of risk analysis (Stage 2).

The term "pest risk analysis" has been used to refer to the evaluation of the biological factors affecting importation decisions (Khan, 1979). Although risk analysis has a long history in other disciplines, its application for phytosanitary decision-making only emerged in the late 1980s (Griffin, 2002). While the quarantine policies of most countries have historically been based on an assessment of pest or disease risks, PRA has only become prominent as a discrete scientific discipline since the formation of the WTO in 1995 (Stynes, 2002).

The establishment of SPS Agreement motivated the Secretariat of the International Plant Protection Convention to develop international standards for phytosanitary measures and even made a revision of the convention itself necessary (FAO 1997). The 1997 revision took on board the basic principles of the SPS Agreement by requiring that phytosanitary measures be technically justified on the basis of PRA. The risk must be identified and assessed and any measures taken must be commensurate with the risk (Pemberton, 2000).

The development of ISPMs can be dated back to the early 1990s. The International Workshop on the Identification, Assessment, and Management of Risks due to Exotic Agricultural Pests, in Virginia, October 1991, was deemed a milestone in PRA harmonisation. It was in that workshop, an international standard on PRA process of three stages was proposed: initiation of PRA, pest risk assessment and pest risk management (Yang *et al.*, 1991).

A decade later, several international and regional standards for PRA have been established. ISPM No. 2 "Guidelines for pest risk analysis" (IPPC, 1995a) was the first, and has been widely recognised and used by the National Plant Protection Organisations (NPPO) of the member countries, even though it is only a conceptual standard for PRA.

In 1995, *pest risk assessment* was formally defined as part of PRA, which "evaluate the probability of the entry, establishment and spread of a pest and of the associated potential economic consequences" (IPPC 1995a). The goal of risk assessment is to assess the likelihood and consequences arising from pest introduction and spread.

Pest risk analysis is defined as a structured decision making process of "evaluating biological or other scientific and economic evidence to determine whether a pest should be regulated and the strength of any phytosanitary measures to be taken against it" (IPPC, 1995a).

There is now a widely accepted requirement that the pest risk analysis and the subsequent plant quarantine decision-making procedures should be both scientific and transparent (WTO, 1995, see 1.4.1). In response to the demands of the international trade rules, pest risk analysis (PRA) approaches have been developed, including the following components (IPPC, 1995a): Initiation (of pest risk analysis), Pest Risk Assessment, and Pest Risk Management.

Some recent PRAs have also included a fourth component: Pest risk communication and documentation. This is not a discrete stage of PRA; it is continuous throughout the PRA process, the purpose is to reconcile the views of scientists, stakeholders, politicians, etc., in order to achieve a common understanding of the pest risks and to develop credible pest risk management options.

Details of PRA stages are reviewed in Chapter 2.

1.4 PEST RISK ANALYSIS AND INTERNATIONAL TRADE

There are two international treaties regarding international trade vs. plant protection: the Agreement on the Application of Sanitary and Phytosanitary Measures (SPS Agreement) and the International Plant Protection Convention (IPPC).

PRA has a role to balance the two conflicting issues of promoting international trade and preventing the spread and introduction of pests. It is recognised that the 1997 version of IPPC does incorporate risk assessment as the fundamental basis for phytosanitary decision-making. Its ultimate objective is to aid rational decision-making by highlighting all probable factors that an introduction of a new pest may result in, and providing recommendations for risk mitigation.

1.4.1 The requirements of the World Trade Organisation

The Uruguay Round of the GATT (General Agreement on Tariffs and Trade) brought plant health under the umbrella of the World Trade Organisation (WTO) for the first time through the "Agreement on the Application of Sanitary and Phytosanitary Measures" (Pemberton, 2000).

As of April 2003, International trade between the current 146 members of the WTO was guided by the 1994 Uruguay Round Agreements. These agreements provided obligatory rules intended to ensure that governments extend free market access to each other's products and services.

The Agreement on the Application of Sanitary and Phytosanitary Measures, which entered into force with the establishment of the World Trade Organisation on 1 January 1995, sets out the basic rules for food safety, and animal and plant health regulations. The essential aim of the SPS Agreement, which builds on the previous General Agreement on Tariffs and Trade (GATT), is to maintain the sovereign right of any government to provide the level of health protection it deems appropriate, but also to ensure that these sovereign rights are not misused for protectionism and do not result in unnecessary barriers to international trade.

Scientific justification and risk assessment

In accordance with the central doctrine of the SPS Agreement, phytosanitary measures that may affect international trade shall be based either on international standards or risk assessment supported by scientific principles and evidence (WTO, 1995).

Where PRA is required to justify the phytosanitary decisions made on imported items, it is expected that:

- some criteria should be taken into account e.g. economic factors, entry, establishment, spread and cost-effectiveness etc.;
- the appropriate level of phytosanitary measures with minimal restrictions for trade should be determined;

Transparency

The SPS Agreement has the intention to make the sanitary and phytosanitary measures more transparent. Governments are required to notify other countries of any new or changed phytosanitary requirements that affect trade, and to set up offices ("Enquiry Points") to supply more information on new or existing measures on request. They are also required to be open to scrutiny on how they apply their plant health regulations. The increased transparency also protects the consumer and trading partners' interests, from hidden protectionism through unnecessary technical requirements.

Particularly, it is stated that if requested, countries must make known what factors they took into consideration, the procedures they used and the level of risk they determined to be acceptable.

Although many governments had already used risk analysis in their SPS measures (without necessarily following formal risk analysis procedures), the SPS Agreement encourages (and effectively requires) the wider use of systematic risk analysis among all WTO member countries and for all relevant products (WTO, 1995).

In the area of phytosanitary decision-making, it is often the case that a degree of lack of information exists; a risk assessment can provide a transparent framework for decision-making based on the best available information, taking into account expert judgment and uncertainty.

Other principles

The SPS agreement also lays down requirements concerning consistency, equivalency, non-discrimination and minimal impact of phytosanitary measures.

The WTO member countries are encouraged to use international standards, guidelines and recommendations where they exist. However, SPS allows countries to set out their own higher standards, which must be based on science. They must be consistent, not arbitrary or unjustifiable, between countries that have identical or similar conditions.

An *Acceptable Level of Risk* (ALR) can be achieved in alternative ways - providing they are technically and economically feasible and also provide the same level of plant health - governments should select those less trade restrictive measures to meet their health objectives. Furthermore, if another country can show that the measures it applies provide the same level of health protection, these should be accepted as equivalent (see SPS Article 4, 5) (WTO, 1995).

1.4.2 The requirements of the International Plant Protection Convention

In relation to the Uruguay Round Agreements of the WTO, particularly the SPS Agreement, the International Plant Protection Convention (the IPPC) plays the vital role of providing international standards for phytosanitary measures affecting trade implemented by governments. The IPPC has been the responsibility of the Director-General of the Food and Agriculture Organisation of the United Nations (FAO) since first adopted by the FAO conference in 1951 (coming into force in1952). It has been amended in 1979 and again in November 1997[6]. The purpose of the IPPC is to secure common and effective action to prevent the spread and introduction of pests of plants and plant products, and to promote measures for their control.

The IPPC was identified in the SPS Agreement as the reference for phytosanitary standards. A series of conceptual standards has been established under the IPPC, International Standards for Phytosanitary Measures (ISPMs), to assist in harmonising phytosanitary decision-making procedures. However, as there is no specific pest-related international phytosanitary standard equivalent to animal health standard under the International Office of Epizootics (OIE), WTO member governments must base their phytosanitary measures on risk assessment (Black, 2003).

Regional Plant Protection Organisations

The 1997 new revised IPPC encourages member countries to cooperate with each other in establishing regional plant protection organisations (RPPOs) in appropriate areas (IPPC, 1997, Article IX). RPPOs participate in various activities to achieve the objectives of the IPPC.

---

[6] The 1997 amendment is now in force and binding on all signatory countries to the IPPC whether or not they ratified the amendment.

There are several influential RPPOs established, such as the European and Mediterranean Plant Protection Organisation (EPPO), the designated RPPO within Europe, which has developed a serial of regional standards on PRA for use in Europe. EPPO's counterparts in North America and Asia are North American Plant Protection Organisation (NAPPO) and the Asia and Pacific Plant Protection Commission (APPPC). RPPOs are playing a more and more important role in:

- Developing an international strategy against the introduction and spread of pests;
- Encouraging the harmonisation of phytosanitary regulations and all other areas of official plant protection action;
- Promoting the use of modern, safe and effective pest control methods;
- Providing a documentation service on plant protection.

(http://www.fao.org/ag/AGP/AGPP/PQ)

Each member countries of the IPPC is obliged to make provision for an official National Plant Protection Organisation (NPPO) (IPPC, 1997, Article IV). Among other responsibilities, the following are directly related to PRA and international trade:

- The issuance of phytosanitary certificates;
- The surveillance of growing plants, and of plants and plant products, particularly of the occurrence, outbreak and spread of pests, and of controlling those pests;

- The inspection of consignments of plants and plant products, and other regulated articles moving in international traffic, particularly with the object of preventing the introduction and/or spread of pests;
- The disinfestation or disinfection of consignments of plants, plant products and other regulated articles moving in international traffic, to meet phytosanitary requirements;
- The protection of endangered areas and the designation, maintenance and surveillance of pest free areas and areas of low pest prevalence;
- The conduct of PRA.

(http://www.fao.org/ag/AGP/AGPP/PQ)

1.4.3 Free trade and international disputes

Many developing countries point out that 'free' trade has in fact not been so free for them. It is the richer countries that have benefited the most from the international liberalisation of trade. In 1997, Canada, the EU countries, Japan, and the USA accounted for almost two thirds of world exports. The least developed countries, with 10 per cent of the world's population, had only 0.3 per cent of the world trade - half the share they had 20 years before.

One important aspect of the WTO is its "dispute settlement procedure", which makes it necessary to align the IPPC with the SPS. To close the potential loophole that measures ostensibly supported by SPS might be used for protectionist purposes, the SPS Agreement gives the right to a government to challenge another country's SPS measures on the grounds that they are not justified by scientific evidence (WTO, 1995). However, the dispute settlement process is ideally for disputes between equally powerful partners. In order to avoid costly disputes, WTO members are encouraged to report trade concerns over food safety, veterinary and phytosanitary measures to the SPS Committee of WTO in the first instance. One of the problems even with this lower-level forum for 'complaints' is that the developing countries lack the information resources, manpower, technical assistance, and legal resources, and ultimately to push their interests to the highly technical WTO panel discussions. That is why there is an urgent need for improved and practical PRA methodologies for developing countries, particularly in handling uncertainty and assessment of subjective criteria like the economic and social impact of pest introduction.

1.5 CONCLUSIONS

This chapter has provided a general context for pest risk analysis and its legislative and economic background. Some of the concepts and terms that will be used throughout this thesis, e.g. pest, risk, risk analysis, risk assessment, plant quarantine, international trade, and their relationship were discussed in this chapter. This chapter also discussed the great threats posed by introduced pests to the global ecosystem and economy, explained that plant quarantine is mostly a response to human activities rather than to the natural pest movement, and that it is a component of sustainable development. The requirements of the SPS agreement and IPPC were briefly explored here with the intention to set PRA in the full context of plant protection and international trade. Required by WTO and FAO, phytosanitary decision-making should be based on rational PRA to reduce a) the pest introduction to a new area, b) unnecessary barriers to international trade. This chapter further reviewed that PRA, a process to determine whether a pest should be regulated and if so, the strength of phytosanitary measures, had become increasingly important in relation to the growing international trade, from economic, legal and political perspectives.

CHAPER 2 REVIEW OF GUIDELINES, PRACTICE AND RESEARCH IN PEST RISK ANALYSIS

2.1 OVERVIEW OF THE INTERNATIONAL STANDARDS ON PRA

Under the IPPC, three international standards for phytosanitary measures (ISPMs) on pest risk analysis (PRA) have been developed and adopted:

- ISPM No. 2 (2007): Framework for pest risk analysis
- ISPM No. 11 (2004): Pest risk analysis for quarantine pests including analysis of environmental risks and living modified organisms
- ISPM No. 21 (2004): Pest risk analysis for regulated non-quarantine pests[1].

2.1.1 SPM No. 2: Framework for pest risk analysis (2007)/ Guidelines for pest risk analysis (IPPC, 1995)

The original ISPM Publication No. 2: Guidelines for Pest Risk Analysis was the first international standard for PRA, which was endorsed in November 1995 and published in February 1996 by ICPM (IPPC, 1995).

Initiating the process involves identification of pests or pathways for which the PRA is needed. Pest risk assessment determines whether each pest identified as such, or associated with a pathway, is a quarantine pest, characterized in

---

[1] PRA for regulated non-quarantine pests is out of the scope of this thesis and is not discussed here.

terms of likelihood of entry, establishment, spread and economic importance. Pest risk management involves developing, evaluating, comparing and selecting options for reducing the risk. Details of the three stages are discussed in 2.2.

ISPM No 2 was revised in 2007 and developed into "Framework for pest risk analysis" (IPPC, 2007).

With the publication of ISPM2, PRA and the relevant concepts such as pest risk assessment, pest risk management were formally defined, and this standard was since widely accepted and followed by most of the NPPOs. However, this standard is more a conceptual guidance on PRA; it sets out the general framework for PRA but does not specify how PRA is to be done - leaving the interpretation and implementation of the guidelines to the PRA practitioners.

2.1.2 ISPM No.11: Pest risk analysis for quarantine pests including analysis of environmental risks and living modified organisms (2004)/ Pest risk analysis for quarantine pests (IPPC, 2001)

The original ISPM No. 11: Pest Risk Analysis for Quarantine Pests, was drafted in 1999, and was endorsed by ICPM in April 2001.

ISPM11 is an expansion of ISPM2. It adds significant details to many aspects of the PRA process. For example, under ISPM11, pest risk assessment includes not only the characterisation of pest risk in terms of likelihood of entry, establishment, spread and economic consequences, but also the documentation of the aspects and degree of uncertainty.

To reflect the increasing requisite of the consideration of environmental risks, a supplement to ISPM11 on analysis of environmental risks was endorsed in April 2003 by ICPM. It was agreed that it should be integrated into ISPM No. 11.

Subsequently it became the first version of the revised ISPM11: ISPM No 11 rev1: Pest Risk Analysis for quarantine pests including analysis of environmental risks (IPPC, 2003). It provided additional information to address the full range of pests covered by IPPC and included details regarding the analysis of risks of plant pests to the environment and biological diversity, including those risks affecting uncultivated/unmanaged plants, wild flora, habitats and ecosystems contained in the PRA area.

This revised ISPM extended the use of some terms to attempt to cover the analysis of environmental risks. For instance, it made it clear that "pests" not only refer to those that directly affect plants, but also to organisms that indirectly affect plants and environment such as weeds, invasive plants and plant species or cultivars that are imported for planting.

With respect to a plant being assessed as a pest with indirect effects, wherever a reference is made to a host or a host range, it is understood to refer instead to a suitable habitat (that is a place where the plant can grow) in the PRA area.

In the case of organisms that affect plants indirectly, through effects on other organisms, the terms host/habitat will extend also to those other organisms.

In the case of plants to be imported, the concepts of entry, establishment and spread have to be considered differently; the intended habitat is the place where the plants are intended to grow and the unintended habitat is the place where the plants are not intended to grow.

Another added aspect to the original ISPM No.11 was to provide more detailed guidance on PRA for living modified organisms (LMO). This supplement was based on ISPM11 Rev. 1. It provided guidance on the criteria for evaluating potential risks to plants and plant health posed by LMOs. It did not alter the scope of ISPM No. 11 but intended to clarify issues related to PRA for LMOs. It was approved by ICPM subsequently in 2004.

The original ISPM No 11 and its two supplements was combined together and became the revised ISPM No 11: Pest risk analysis for quarantine pests including analysis of environmental risks and living modified organisms (IPPC, 2004)

The problem with the guidelines remains that no matter however detailed explanation of PRA it provided, ISPM11 did not recommend any specific methods to conduct a PRA. Nor does it provide guidance on how detailed a PRA should be under different circumstances. Moreover, little or no guidance is given as to how the available data is used to estimate the risk from each criterion (biological, economic, environmental and social), and even more significantly, how an overall prognosis should be made from the assessment of individual factors is left unspecified.

2.2 THE THREE STAGES OF PEST RISK ANALYSIS

According to ISPM2 and ISPM11, there are three stages in a PRA process:

- Stage I: PRA initiation;
- Stage II: Pest risk assessment; and
- Stage III: Pest risk management

2.2.1 Stage 1. Initiating the PRA process (IPPC, 2003)

This stage includes all the steps to ensure that a PRA is indeed required. PRA could be initiated either by the identification of a pest that could present a potential hazard or by the identification of an imported commodity, a new plant species, or other new pathways that may allow the introduction and/or the spread of pests. This process normally includes:

- Identifying a pest that could present a potential hazard or identifying a pathway that quarantine pests may associate

- Determining whether an appropriate PRA process may have already been carried out.
- And whether the pest concerned is a potential quarantine pest or whether potential quarantine pests are identified. Meanwhile, geographic and biological characteristics of the pests, and the areas at risk (PRA areas) should be identified (IPPC, 2001).

As part of the initiation process, previous risk assessments for the same or close relative species should be identified and cited. Information on the pests of concern, pathways, conditions, and risk management options should be collected. Also important is co-operation between the stakeholders, e.g. trading partners, regulatory bodies and PRA practitioners in the importing and exporting countries.

At the end of Stage 1, pests, pathways and PRA area are identified; information is collected; previous PRAs are reviewed and quarantine pest candidates are listed. Figure 2.2 shows the process of PRA initiation.

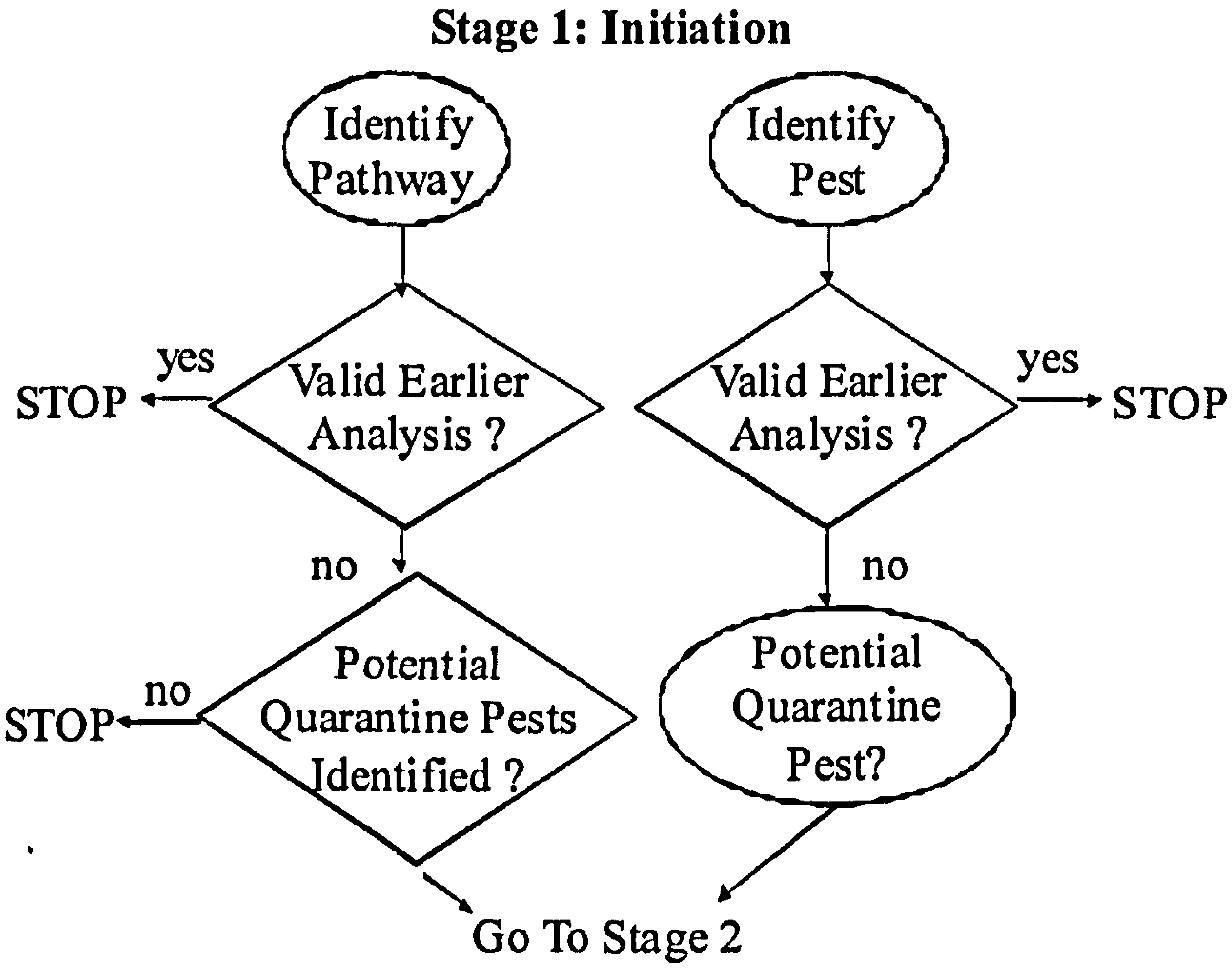


Figure 2.2 PRA stage 1: PRA initiation (IPPC, 1996)

2.2.2 Stage 2. Pest risk assessment (IPPC, 2003)

The introduction and spread potential and consequences of the potential quarantine pests identified in stage 1 will be evaluated at stage 2. This assessment can be done qualitatively and/or quantitatively.

There is usually wide agreement that if the introduction of a pest presents unacceptable consequences, risk management measures will be warranted. It is only when the strength of the risk measures, the cost of control, or the ranking of pests for significance is in question, that it is necessary to examine economic factors in greater detail (IPPC, 2001).

At the end of stage 2, the following work should have been done:

- Pests considered for risk management identified
- Endangered area identified
- Probability of introduction assessed
- Economic consequences (including environmental risks) assessed
- Degree and area of uncertainty documented

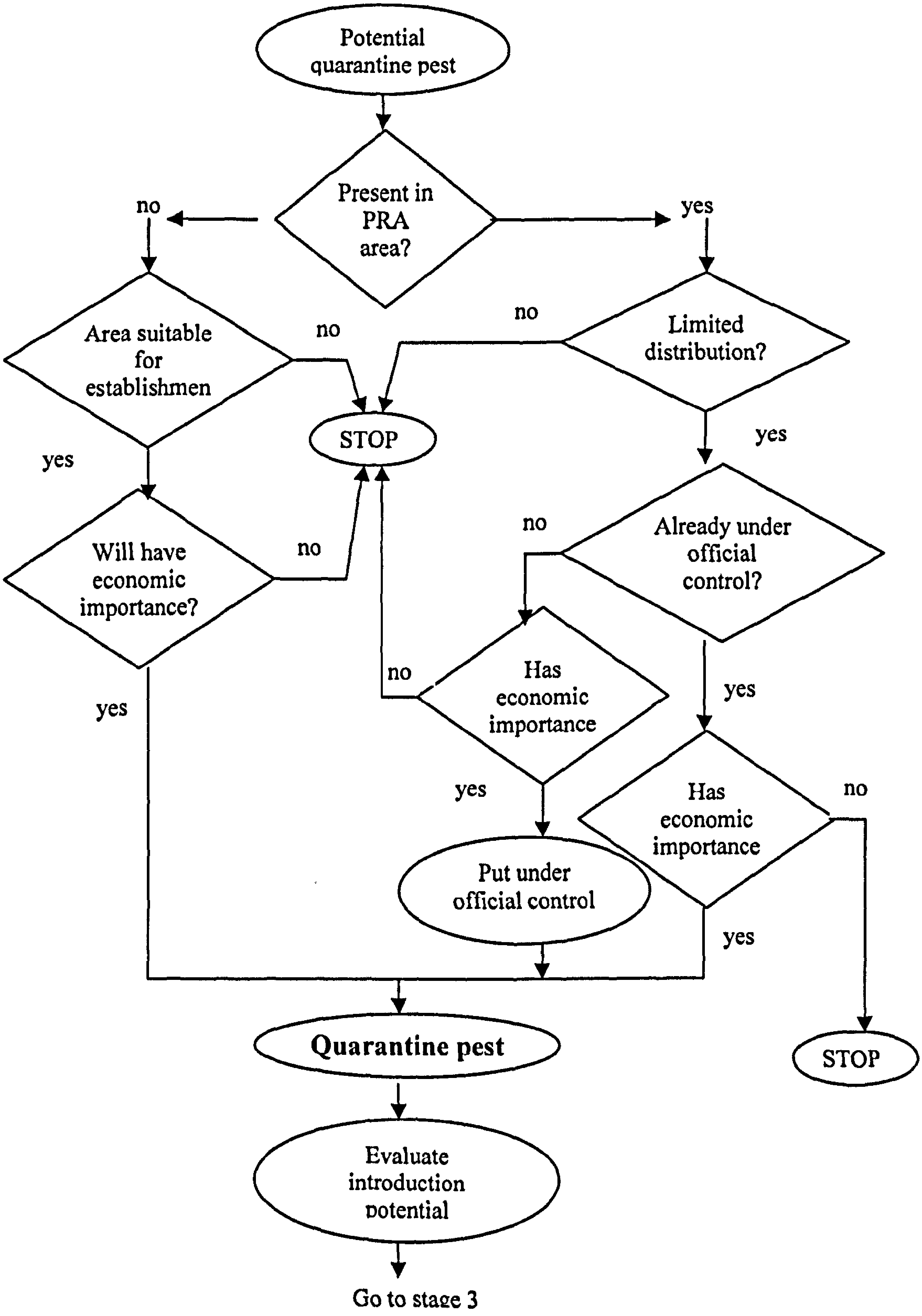


Figure 2.3 PRA stage 2: pest risk assessment

2.2.3 Stage 3. Pest risk management (IPPC, 2003)

The third stage of pest risk analysis is to make a judgment, on whether risk management is required and the strength of risk management measures to be used based on information obtained from stage 2. If the risk assessed is found to exceed the ALR, possible risk management measures should be identified to reduce the risk to, or below the ALR. The principle for pest risk management is to manage risk to achieve the required degree of safety that can be justified and is feasible within the limits of available options. In the event that several equivalent options are available, the least trade restrictive approaches should be recommended (WTO, 1995).

The decision-making process should be based on the information as follows:

- Reasons for initiating the process
- Finding of the pest categorisation phase
- Estimation of the probability of introduction to the PRA area
- Evaluation of potential economic and environment consequences in the PRA area

According to the principle of modification[2] in ISPM1, the implementation of a particular pest risk management measure should be monitored and reviewed

---

[2] Modification: As conditions change, and as new facts become available, phytosanitary measures shall be modified promptly, either by inclusion of prohibitions, restrictions or requirements necessary for their success, or by removal of those found to be unnecessary (FAO, 1995: Principles of Plant Quarantine as Related to International Trade).

to ensure that the measure be modified promptly should any new information become available.

Figure 2.4 shows the process of pest risk management.

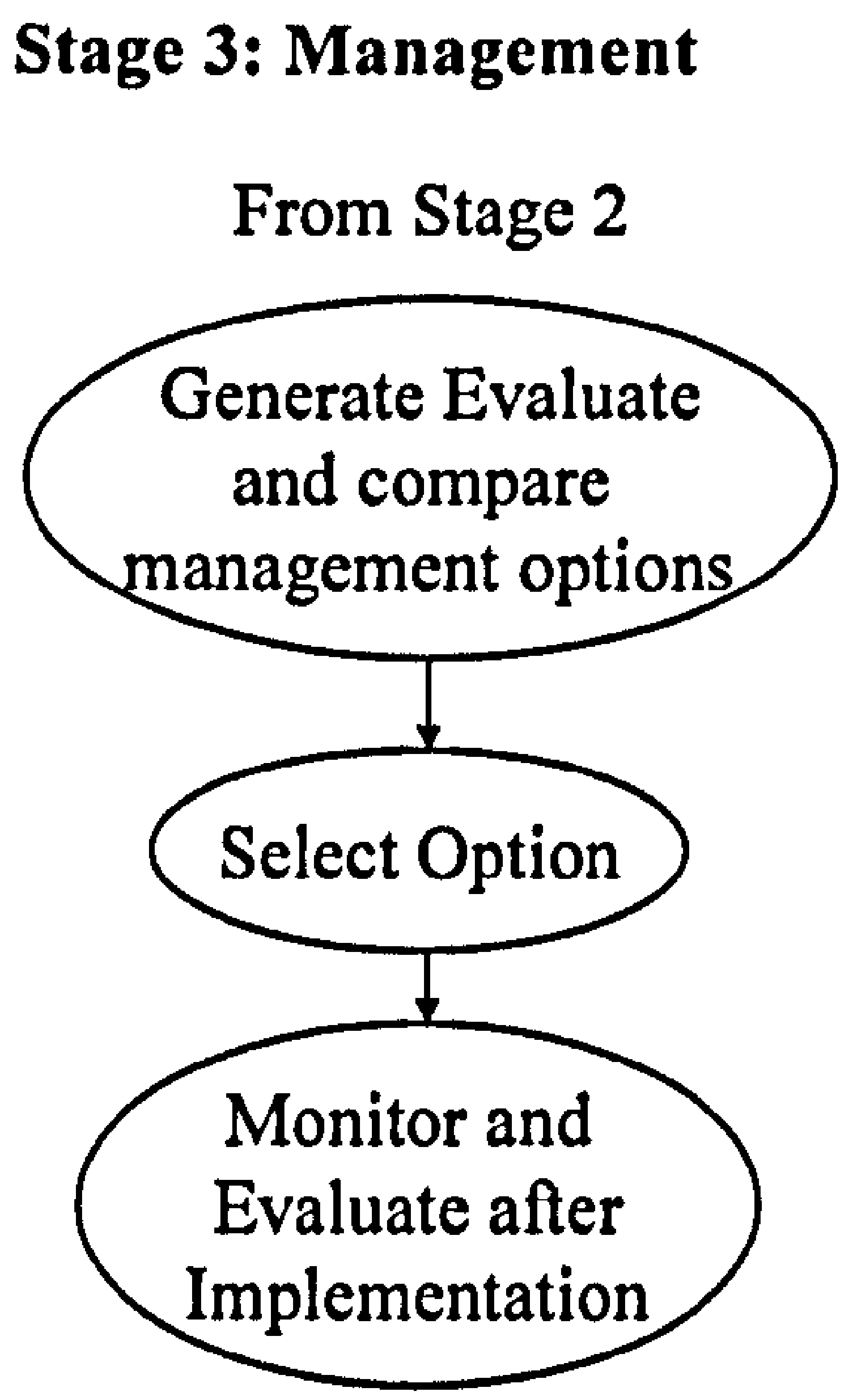


Figure 2.4 PRA stage 3: Pest risk management

2.3 REGIONAL AND NATIONAL APPROACHES TO IMPLEMENTING THE ISPMs ON PRA

Some Regional Plant Protection Organisations such as EPPO and NAPPO have also established PRA guidelines or schemes, which followed the general principles of the ISPMs but are more sophisticated and operable.

2.3.1 The EPPO standards on PRA

Based on many years of experience of their experts on pest risk analysis and phytosanitary regulations, harmonised with the ISPMs, the EPPO established three standards in 1998: Guidelines on Pest Risk Analysis, which includes three parts:

- PM 5/1 (1) Check-list of information required for pest risk analysis
- PM 5/2 (1) Pest risk analysis to decide immediate action to be taken on interception of a pest in an EPPO country
- PM 5/3 (1) Pest risk assessment scheme

Subsequently, PM 5/4 Pest risk management scheme was approved in September 2000 to complete this series of EPPO Phytosanitary Measures on PRA; PM 5/2 (1) was reviewed, the revised PM 5/2 (2) Pest risk analysis on detection of a pest in an imported consignment replaced the previous PM 5/2 (1).

This EPPO series of guidelines on PRA (PM 5/1-4) provides detailed guidance on the analysis of risk from individual pests in relation to their potential status as quarantine pests or regulated non-quarantine pests, concerning different elements of pest risk analysis, and also the different purposes for which PRA is performed.

As part of this research was based on the EPPO schemes, they will be elaborated in more detail in the following sections.

2.3.1.1 PM 5/1 (1) Check-list of information required for pest risk analysis (PRA) (EPPO, 1998)

It specifies the information that should be considered before deciding that a given pest qualifies as having the characteristics of a quarantine pest, such as the organism, biological characteristics, geographical distribution, host plants, potential for establishment in PRA area, control, transport, and economic impacts. This is the first comprehensive information checklist for PRA.

2.3.1.2 PM 5/2 (2) Pest risk analysis on detection of a pest in an imported consignment (EPPO, 2002)

The first version of PM 5/2 Pest risk analysis to decide immediate action to be taken on interception of a pest in an EPPO country was a simplified questionnaire style PRA scheme (answered by 'yes' or 'no') that was used when an unfamiliar pest was intercepted within EPPO region for the first time. It allowed the importing country to decide whether action should be taken to the particular consignment in a relatively short time.

The second version of PM 5/2 approved in September 2001. PM 5/2 (2) Pest risk analysis on detection of a pest in an imported consignment is a simplified PRA scheme that is intended to be used when an unfamiliar pest is detected in an imported consignment. It allows a quick decision as to what phytosanitary action to take with regard to the particular consignment.

This scheme recognises the fact that most of the time when an unfamiliar pest is detected in a consignment there is little available information about some or all of the important elements necessary for PRA. Consequently, throughout this simplified scheme, it is necessary to make assumptions based on expert opinion, and often with consideration given to possible worst-case scenarios. Hence, this procedure is not intended to be a substitute for a full PRA procedure that is used to decide whether a pest can be categorized as a quarantine pest and the phytosanitary measures to be taken.

2.3.1.3 PM 5/3 (1) Pest risk assessment scheme (EPPO, 1998)

This scheme presents detailed instructions for the first two stages of PRA (Figure 2.5): initiation of a PRA process and pest risk assessment. It helps the PRA practitioners to decide whether a pest risk exists and to conduct the so-called quantitative assessment[3] of that risk, based on questions to which replies are given on a 1-9 scale. However, expert judgment has to be used in interpreting the replies.

---

[3] The author has reservation about the terminology used in some PRA schemes on qualitative and quantitative assessment. In the author's view, it is inappropriate to call an assessment "quantitative or semi-quantitative" simply by introducing score/number into an assessment. See "Research on pest risk assessment" in later section.

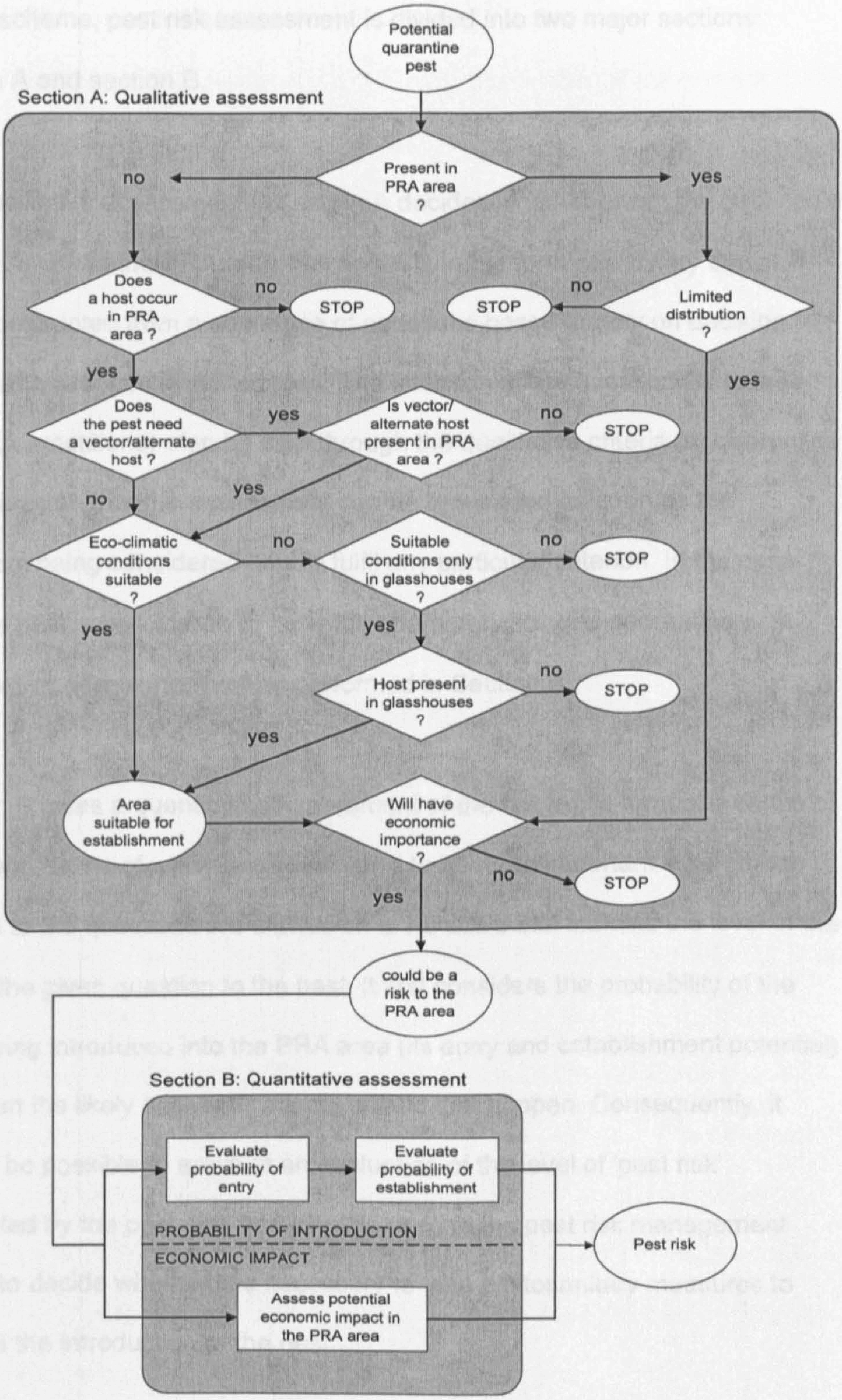


Figure 2.5 Pest risk analysis, Stage 2: Assessment (EPPO, 1998)

In this scheme, pest risk assessment is divided into two major sections: section A and section B.

The qualitative assessment in section A decides whether or not the pest could present a risk to the PRA area. Section A is in the form of a binary decision tree, constructed from a sequence of questions based largely on decision points with two "yes or no" options. The intention of the questions is to lead the PRA practitioner step by step through the qualitative criteria for quarantine pest status so that the assessment can be terminated as soon as the organism being considered fails to fulfil any particular criterion. In the case that the pest is considered to have the characteristics of a quarantine pest, quantitative assessment will be performed in Section B.

Section B gives a quantitative assessment of the risk in the form of a series of questions, some of which are considered to be more important than others. Scores to the questions are elicited on a 1-9 scale that indicate the level of the risk of the given question to the pest. It first considers the probability of the pest being introduced into the PRA area (its entry and establishment potential) and then the likely economic impact should that happen. Consequently, it should be possible to arrive at an evaluation of the level of 'pest risk' presented by the pest; this can then be used in the pest risk management phase to decide whether it is necessary to take phytosanitary measures to prevent the introduction of the pest.

The EPPO Pest Risk Assessment Scheme was the first scheme to indicate that some risk factors/questions are more important than others and suggests that risk scores can be weighted, prior to being combined in an appropriate way.

One of the differences between this scheme and the ISPM 11 is that the EPPO scheme places *spread potential* of the pest within the scope of *economic impact* because the speed and extent of the spread is regarded as more related to the economic loss than introduction.

2.3.1.4 PM 5/4 (1) Pest risk management scheme (EPPO, 2000)

The purpose of the EPPO pest risk management is to decide: (1) whether phytosanitary measures are required to reduce the risk from a certain pest to an acceptable level; and (2) which measure or measures can or should be applied. This scheme is composed of a sequence of numbered steps, most of which are questions. The steps are followed successively for each of the major pathways likely to carry the pest or (for a commodity-initiated analysis) for each of the pests likely to be associated with the pathway. The decision-making scheme identifies the measures appropriate for an individual pest (in the case of a pest-initiated analysis); but it can also be used for each of the several pests that could be carried on a pathway (in the case of a commodity-initiated analysis).

2.3.2 Examples of NPPO PRA practice

Member countries of WTO and IPPC are encouraged to apply the ISPMs. However, recognising the diversities of resources, circumstances, target pests or commodities between different countries, it is understandable that different countries may employ different approaches to PRA, provided they are in line with ISPMs. Various methods e.g. qualitative or quantitative approaches are used to fit the different purposes. Some examples of national PRA approaches are presented below.

2.3.2.1 The UK PRA Approaches

In the UK, Department for Environment, Food and Rural Affairs (DEFRA) is responsible for the implementation of plant health regulations within England and Wales. The Plant Health Division of DEFRA is responsible for the plant health aspects of imports of plants, plant produce, soil and other growing media. The Plant Health and Seeds Inspectorate (PHSI) carries out import and export inspections, issues phytosanitary certificates, oversees eradication campaigns and the operation of the Plant Passport scheme. The Central Science Laboratory (CSL)[4], an executive agency of DEFRA, provides scientific support on plant health measures. CSL identifies pests on samples submitted by the PHSI and advises on interceptions and outbreaks of pests

---

[4] On 1 April 2009, CSL became part of The Food and Environment Research Agency (Fera). However, CSL is still used throughout this thesis.

and diseases. PRA for pests and commodities is conducted by a team of Pest Risk Analysts within Plant Health Group at CSL.

CSL has developed a condensed version of the recognised schemes for rapid risk assessment: a short, summary qualitative scheme that contains the major factors in the ISPM2 and 11. It can be completed very quickly to decide action against pest interceptions and whether a detailed analysis is required before committing extra resources. The EPPO PRA scheme can then be used for detailed analysis (Baker *et al.*, 1999).

One PRA example done by CSL concerns Asian long-horned beetle (ALB) (*Anoplophora glabripennis*), a wood-boring pest found in New York in 1996 that was introduced with solid wood packaging material (SWPM) from China (APHIS 1996). This event initiated a summary PRA, followed by a detailed EPPO pest risk assessment by CSL and Forest Research Commission (Evans *et al.*, 1998; MacLeod *et al.*, 2003; MacLeod *et al.*, 2002). The assessment concluded that there was a risk of entry and establishment of this pest to the EPPO region and the species had the potential to cause serious damage to trees including willows and poplar species, both of forestry importance in the UK. It recommended that the insect should be considered a quarantine pest and that it would be prudent to introduce phytosanitary measures to protect the EPPO region (McLeod *et al.*, 2002, 2003).

Another example is that CSL in 1997 conducted a summary PRA for watermelon silver mottle virus (WSMV) and subsequently a detailed assessment following the EPPO scheme. It showed that the mean risk of introduction was 7.3 (on a 1-9 scale) and the mean risk of economic impact was 6.5. As a result of this PRA, MSWV was listed as an A1 quarantine pest[5] for both the EU and EPPO (Sansford, 2002).

A more recent example is the PRA for *Aceria tulipae* (Keifer) conducted by CSL (MacLeod, 2007). The PRA was initiated by the finding of *Aceria tulipae* on onion (*Allium cepa*) sets from the Netherlands, and UK was proposed to be the PRA area. The PRA concluded that there was a risk of entry and establishment of *Aceria tulipae* to the UK. However, the potential economic and environmental impacts, which *Aceria tulipae* is likely to cause, is minor even without official control. Furthermore, outbreak of this pest is likely to be eradicated. As a result, no phytosanitary measures are recommended (McLeod, 2007).

These examples demonstrate that PRA is a useful tool for phytosanitary decision-making, and in particular:

- Risk analysis can be done in a different way for a different purpose. For example, the summary scheme was used to enable a quick decision, and the EPPO scheme was used for a detailed analysis;
- The same scheme produced different outcomes i.e. different levels of risk for different organisms; in contrast, if a scheme always produced the same outputs for different organisms under different circumstances, for example, high risk for every organism, it clearly would have little value.

---

[5] A1 list of pests are pests recognized not to be present in any part of the EPPO region and that present a risk to most or all parts of the region; and the A2 list of pests are pests with a limited distribution in EPPO region, presenting a risk of further spread.

2.3.2.2 The USDA guidelines on weed-initiated pest risk assessment

In the USA, the Animal and Plant Health Inspection Service (APHIS) of the United States Department of Agriculture (USDA) is responsible for the implementation of plant health regulations. In August 1998, it developed the Weed-initiated Pest Risk Assessment: Guidelines for Qualitative Assessment, following the weed policy developed by APHIS in 1994. The guidelines propose that risk assessment should be used as a basis for weed exclusion decisions, and pest-initiated, qualitative pest risk assessments are conducted to determine whether a weed species should be regulated.

Under this scheme, the pest risk is assessed using five Risk Elements (RE):
RE #1: Habitat Suitability;
RE #2: Spread potential after establishment, Dispersal Potential;
RE #3: Economic Impact;
RE #4: Environmental Impact; and
RE #5: Likelihood of Introduction/Spread.

RE #1-4 focus on the consequences of introduction and RE # 5 focuses on the likelihood of introduction. Each risk element is rated from the highest (3) to the lowest (1), and an estimate pest risk potential (PRP) is produced by simply summing the RE scores.

2.3.2.3 The Canadian PRA approach

In Canada, The Plant Health Division of the Canadian Food Inspection Agency (CFIA) is responsible for the implementation of plant health regulations and pest risk management (the stage three of the pest risk analysis), translating pest risk assessment data into policies implementing phytosanitary measures. The Plant Health Risk Assessment Unit (PHRAU) of the CFIA is responsible for pest risk assessment (stage 1 and stage 2 of the pest risk analysis). The PHRAU provides about 50 risk assessments a year to the Plant Health and Production Division (Watler, 2002).

In Canada, PRA is carried out to provide scientific support to make policy changes; in response to a request to import a new commodity; new discoveries within Canada; surveys or other information indicating that a situation has changed. Canada adopts a qualitative risk assessment scheme, which considers the risk factors covered by the current ISPMs. For each risk factor, a qualitative rating (guidelines are given) is made as low, medium or high, which is then converted to a numerical value as 1, 2 or 3, respectively. Subsequently, an overall rating of risk is derived by summing up the score of each factor.

A problem related to risk rating in this scheme was to achieve a high overall rating to reflect the perceived high risks for the species with high economic impact but with low environmental impact (Watler, 2002, personal communication). This in fact demonstrates the need to introduce weighting into the scheme that recognises that some important factors are more influential than others (Zhu *et al.*, 2000, 2002, see 3.5 and 4.4 for further discussion).

2.4 RESEARCH ON PEST RISK ASSESSMENT METHODOLOGIES

Risk assessment methods can be broadly characterized as qualitative or quantitative. Qualitative assessments usually rely on binary or ordinal scoring of risk, whilst quantitative assessments usually employ stochastic and probabilistic approaches. Subjectivity ('Expert judgment') is the weakest point with qualitative assessments, whereas lack of data (experimental or heuristic) limits the application of quantitative approaches[6] (Zhu *et al.*, 2002).

2.4.1 Characterising the risk

Whichever approach is employed, the first step has usually been to characterise the risk factors in some systematic way; this equates to identifying "what can go wrong?". In the EPPO risk assessment scheme, for example, there are about 45 risk factors, each of which takes the form of a question. Identifying and structuring risk factors has not usually involved any particular methodologies.

---

[6] The term quantitative approach used here is referred to the process of measurement that is based on empirical observation and mathematical expression of quantitative relationships. However, the author acknowledges the common usage of this term in the PRA world, where most people regard ordinal scoring as quantitative and indeed EPPO stage B is called quantitative but is based on scores.

However, Zhu *et al.* (2000) used mind mapping[7] to facilitate risk identification by disaggregating pest risk into a series of nested risk factors, from general to specific. This highlights the dependencies among risk factors and helps to distinguish factors that are manageable ("control points") as an aid to the selection of risk management measures.

2.4.2 Quantifying specific risk factors

For those risk factors that are amenable to quantification, it is possible to provide detailed predictions. Some computer-based approaches have been used to assess risks associated with specific factors such as establishment potential.

Spatial analysis using geographic information systems was employed in the USA to monitor pest outbreaks, in order to assess the hosts at risk and the risk of spread. It has recently been applied to med fly, karnal bunt, and citrus canker (Sequeira, 2002b).

An automated software, CLIMEX decision-support system developed by Sutherst and Maywald in 1985, is applicable to any pest species, provided there are data on its current geographical distribution. The software provides insights into the species' performance in new environments (Sutherst *et al.*, 1991).

---

[7] Mind mapping is a technique to use diagrams to represent words, ideas, tasks or other items linked to and arranged radially around a central key word or idea. It aims to generate, visualize, structure and classify ideas, and as an aid in study, organisation, problem solving, decision making, and writing. An introduction to mind mapping can be found in Appendix 2.

CLIMEX was used in Australia, the UK and New Zealand to evaluate the risk of establishment of exotic species in relation to climate in a new environment. Baker *et al.* (1998, 2003) applied CLIMEX to predict the potential distribution of Colorado beetle (*Leptinotarsa decemlineata*) and western corn rootworm *(Diabrotica virgifera virgifera)* in the UK under current climate conditions and under global climate change (Baker *et al.*, 1998, 2003). A study of the establishment potential of Asian longhorned beetle (*Anoplophora glabripennis*) in Europe using CLIMEX showed that almost 40% of Europe has a climate suitable for *A. glabripennis* establishment, although biotic factors must also be taken into consideration. This risk analysis contributed to the decision to add *A. glabripennis* to the list of quarantine pests whose introduction and spread within all EC Member States is banned (MacLeod *et al.*, 2002, 2003).

CLIMEX was applied using analogous climates and global insect pest distribution data to identify potential sources of new invasive insect pests for New Zealand (Peacock and Worner, 2006).

Dobesberger (2002) described some multivariate analysis techniques used to predict establishment potential. Examples included a multiple linear regression model for soybean rust, discriminant analysis for bacterial leaf blight of rice, and logistic regression for pink bollworm[8].

---

[8] The pink bollworm (*Pectinophora gossypiella*), *lagarta rosada* in Spanish, is an insect known for being a pest in cotton farming. The adult is a small thin gray moth with fringed wings. The larva is a dull white eight-legged caterpillar with conspicuous pink banding along its dorsum. The larva reaches one half inch in length. The pink bollworm is native to Asia but has become an invasive species in most of

Peacock *et al.* (2006) investigated the influence of climate variables on insect establishment patterns by using discriminant analysis to classify the climatic preferences of two groups of polyphagous insect species intercepted at New Zealand's border. This study showed that multivariate statistical techniques such as discriminant analysis can help distinguish the climatic limits of insect distributions over large geographical scales.

Probabilistic scenario analysis (PSA) has been used since the 1940s to assess the risks associated with nuclear technology, other engineering applications, financial analyses, and general economic evaluations. A PSA example implemented in PRA was described by Oryang (2002) striving to present PSA as a structured and practical approach. Scenario type risk analyses were also used in Australia in import risk analysis (Stynes, 2002).

Other quantitative techniques may have application in PRA as well. For example, probabilistic risk analysis based on systems analysis and Bayesian probability has long been used in disciplines such as astronautics and nuclear safety, when there are seldom enough data for a classical statistical analysis (Pate-Cornell and Dillon, 2001).

McDowell (2002) discussed various data analysis techniques and predictive models that may be useful for PRA.

---

the world's cotton-growing regions. It reached the cotton belt in the southern United States by the 1920s. It is a major pest in the cotton fields of the southern California deserts.

Holt *et al.* (2006) discussed using likelihood ratios based on Bayesian theory in PRA. Assigning likelihood ratios instead of scores allows a more rigorous probabilistic treatment of the data, which can offer more effective discrimination between organisms. For each component of the assessment, the likelihood ratio expresses how many times more likely is the evidence that the organism poses a risk as a quarantine pest than if it doesn't. According to Bayesian theory, the product of the likelihood ratios for all components and the prior odds gives the posterior odds that the organism poses a risk as a quarantine pest, given the evidence available. If no prior information was available, the prior odds were regarded as neutral. To express this as a probability the posterior odds are divided by itself plus one.

Application of artificial neural networks has showed potential in plant protection and biosecurity. Artificial neural networks are predictive tools that have the ability to detect and approximate non-linear relationships from the data. In a study by Peacock *et al.* (2007), artificial neural networks were used to predict the geographic distribution of groups of polyphagous plant pests. Using climate variables as predictors, artificial neural network models were compared with binary logistic models for predicting insect distribution. Using bootstrapping, artificial neural networks were shown to predict insect presence and absence significantly better than the binary logistic regression models (Peacock *et al.*, 2007).

2.4.3 Simplifying risk assessment

Pate-Cornell and Dillon (2001) made the point that "It is generally impossible to include all components and all event scenarios in a PRA [referring to probabilistic risk analysis], and an adapted screening procedure is necessary. This screening procedure is meant to filter out the scenarios that are low contributors to the overall risk while retaining the important ones". With similar issues in mind, Zhu *et al.* (2002) investigated decision-making in risk assessment in an attempt to identify the more important risk factors and assess their consistency between different cases. Zhu *et al.* (2002) suggested using multivariate statistics (Principal Components Analysis, PCA) and genetic algorithms to simplify the risk assessment without losing important information.

2.4.4 Uncertainty and expert judgment in PRA

Difficulties in assessing risk under uncertainty are obvious and the use of expert opinion with its associated subjectivity is inevitable. Major uncertainties in PRA concern the behaviour and pest status of non-indigenous organisms in new environments (McDowell, 2002). Inputs based on expert judgment are also essential to developing probabilistic models of pest risk (Dobesberger, 2002). As is often the case, the problem with the precautionary approach is that conservative estimates of the pest risk are used as a way to accommodate uncertainty, with alarming and discouraging results for trade. Bias often exists in expert judgment. Zhu *et al.* (2000) suggested using Delphi techniques to reduce the individual biases in expert judgment. Such an approach requires a pool of PRA practitioners to give their opinions independently and adjust the outcome collectively. EPPO, through their PRA panel, uses a similar approach for some PRA cases, although not necessarily by means of a formal Delphi study. Zhu *et al.* (2001) also described various sources of uncertainty and suggested several methods such as using fuzzy logic, sensitive analysis and Monte Carlo simulation to handle these.

2.5 CONCLUSIONS

In this chapter, the international guidelines and the regional standards on PRA are reviewed, as well as some examples of various approaches by a number of countries.

Zhu *et al.* (2002) also reviewed the existing literature and research on PRA. In common with risk assessment in some other disciplines, a number of problems must be overcome, e.g. subjectivity, uncertainty, non-quantifiable variables, and the need to integrate information into a simple statement of risk. It is hard to define the risks of pest introduction, which are not totally characterised and related with many human activities, as well as involving many uncertainties.

Various bodies have made efforts in developing PRA methodologies. Yet it is found that the current practices and methods exhibited a number of characteristics that left scope for improvement or further development, and it is necessary to examine a number of problem areas in relation to PRA:

- There is a lack of methods to assign a score to a risk factor[9];
- There is a lack of method to combine the risk scores to give a meaningful overall risk level;
- There is a lack of methods to handle economic and social criteria and to integrate these with biological criteria;
- The current approach to risk scoring does not allow for a statement of uncertainty;
- There is no clear methodology for a summation or combination of individual risk scores and assessments under different criteria;
- There is no expression for the degree of importance for risk criteria and individual risk factors;
- There is a lack of case study data as examples[10];
- There is a lack of acknowledgement of relationships and duplications among risk factors;
- There are different types of ambiguity in PRA[11];
- Even when the risk factors are identified, it is still not possible to make an evaluation on the probability of the risk due to a lack of data.

---

[9] MacLeod and Baker published a paper on assigning descriptions to scores for the questions on entry and establishment for the EPPO pest risk assessment scheme (MacLeod and Baker, 2003).

[10] This review was originally done in 1999 - 2000, when there was hardly any publication on PRA for a specific organism following international/regional guidelines, it was subsequently presented at the BCPC conference in 2002. Since then, there have been a number of publications regarding specific PRA cases, e.g. those published by the CSL staff. However, the other statements still hole true.

[11] See Appendix 1 for discussion on ambiguity in PRA.

It is impossible to address all the issues identified above in this study, therefore the focuses are on the following issues, which aim to investigate ways of improvement on the existing approaches:

a) how to structure a risk assessment;
b) how to ascribe a score;
c) how to combine the individual scores into a final statement of risk level;
d) how to derive weightings and then incorporate these into the risk assessment;
e) what are the relationships among the risk factors;
f) is implication of risk assessment possible?
g) what factors/rules are most important in determining the level of pest risk.

In Chapter 3, methods are proposed to address issues a – d; in Chapter 4, some statistical techniques are applied to investigated issues d – f.

CHAPTER 3 STRUCTURING THE COMPONENTS OF PEST RISK ASSESSMENT

3.1 INTRODUCTION

Pest risk assessment is the evaluation of the probability of the introduction[1] and spread of a pest and of the associated potential environmental and economic consequences (IPPC, 1999); it is arguably the most problematic stage in PRA.

Risk assessment in general is a structured science-based process to estimate the likelihood and severity of risk with attendant uncertainty (Coleman and Marks, 1999). It involves techniques for identifying, characterising, quantifying and evaluating hazards.

In this thesis, case studies are used to assist in identifying ambiguities with current pest risk assessment guidelines and methods, and in gaining a thorough understanding of pest risk analysis (see Appendix 1 for discussion on ambiguity). PRA case studies of some lepidopteran, dipteran and coleopteran pests were used to examine the assessment of introduction potential. The factual information was taken from various reference materials, including the Crop Protection Compendium (CAB International, 2000), EPPO's Plant Quarantine Retrieval System (EPPO, 1999), Quarantine Pest for Europe (Smith *et al.*, 1997), Internet resources, and literature reviews.

---

[1] Introduction includes the entry and establishment of an organism into a new area.

This chapter is organised as follows: Section 3.2 focuses on the identification of the factors involved in pest risk; section 3.3 explores the methods employed to evaluate the pest risk; section 3.4 concludes the Chapter.

3.2 PEST RISK IDENTIFICATION (PRI)

Risk assessment in the wider disciplines is sometimes regarded as having two stages. The first in which the various sources of the risk (hazards) are identified and structured, and the second in which some kind of evaluation takes place to decide the level of the risks that have been identified (Warner, 1992).

Taking this idea from risk assessment from another field therefore, it is proposed in this thesis that the pest risk assessment be divided into two steps: pest risk identification and pest risk evaluation. The rest of this section describes the approach to PRI, followed by PRE in Section 3.2.

3.2.1 Introduction - the concept of Pest Risk Identification (PRI)

Identifying and structuring risk factors has not usually involved any particular methodologies. Recognising the ambiguities (see appendix 1) and problems inherent in the international guidelines, in this chapter attempts were made to understand the source of pest risk, and to specify the risk factors.

Whichever approach is employed to conduct a risk analysis, the first step has usually been to characterise the risk factors in some systematic way; a step that identifies the risk factors i.e. all the possible events or characteristics that may result in a pest introduction in the context of PRA. This corresponds to identifying what can go wrong. Hence, pest risk identification (PRI) is defined hereafter as "the process of identification and characterisation of the potential sources of the pest risk".

The objective of PRI was to capture all of the risk factors, specify the nature of the risks and uncertainties faced, and structure the problem by classifying the risk factors involved. It is argued here that identifying the potential pest risk and developing an understanding of its structure are essential prerequisites to the subsequent steps. In the process of PRI, the source, cause and effect of the potential pest risk were reviewed and described in detail.

Risk factors[2] have been explored in the ISPMs and various regional PRA guidelines. The ISPM2 suggested some general risk criteria; the subsequent ISPM11 gave some further explanation of these criteria by providing a series of general factors being considered in a risk assessment. The various regional and national schemes also recommend such factors. In the EPPO risk assessment scheme, for example, there are about 45 risk factors, each of which takes the form of a question.

3.2.2 The application of mind mapping in PRI

In this study, the relationships between risk factors was visualised by using mind mapping (MindManager, 2002) (see Appendix 2 for a detailed description of mind

---

[2] In this thesis, *risk factor* is used to refer to all the possible events or characteristics that may result in a pest introduction or spread in a new area.

mapping and MindManager). Mind mapping helps facilitate risk identification by disaggregating pest risk into a hierarchy of risk factors, from general to specific, and subsequently to visualise the relationships among those factors.

The FAO PRA guidelines (IPPC, 2001) have clearly specified the factors associated with geographical and regulatory criteria; consequently the focus of this study was on the exploration of risk factors for introduction potential and economic consequences. Risk criteria[3] were sub-divided into risk factors from general to specific. Risk factors were disaggregated until sufficiently specific to be presented by a series of direct, tangible questions (See Appendix 7 for a case study of PRA on the introduction potential of Fall Webworm). The risk concerned could then be estimated with less subjectivity. Based on appropriate items of information or assumptions, a risk analyst can then assess this factor qualitatively or quantitatively.

3.2.3 Results – PRI, the structure of risk factors

In the following case studies, identifying pest risk factors for introduction potential and economic consequences, structured by mind mapping, were explored in detail. The risk criteria were broken down into a hierarchy of increasing specificity. The risk factors identified were mostly based on the FAO guidelines and the EPPO PRA scheme. However, these risk factors have been regrouped, some inappropriate or impractical factors were excluded, some factors were disaggregated further, and some new factors were added.

---

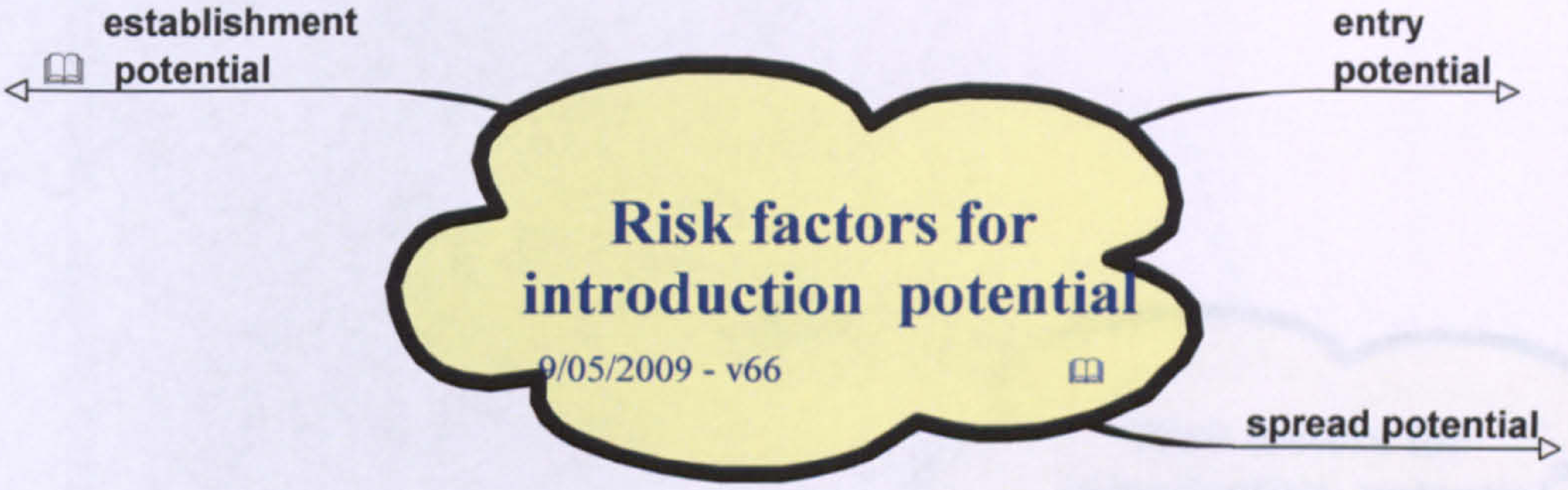
[3] In this thesis, *risk criteria* refer to the broad headings of the characters that may result in a pest being defined as a quarantine pest or regulated non quarantine pest

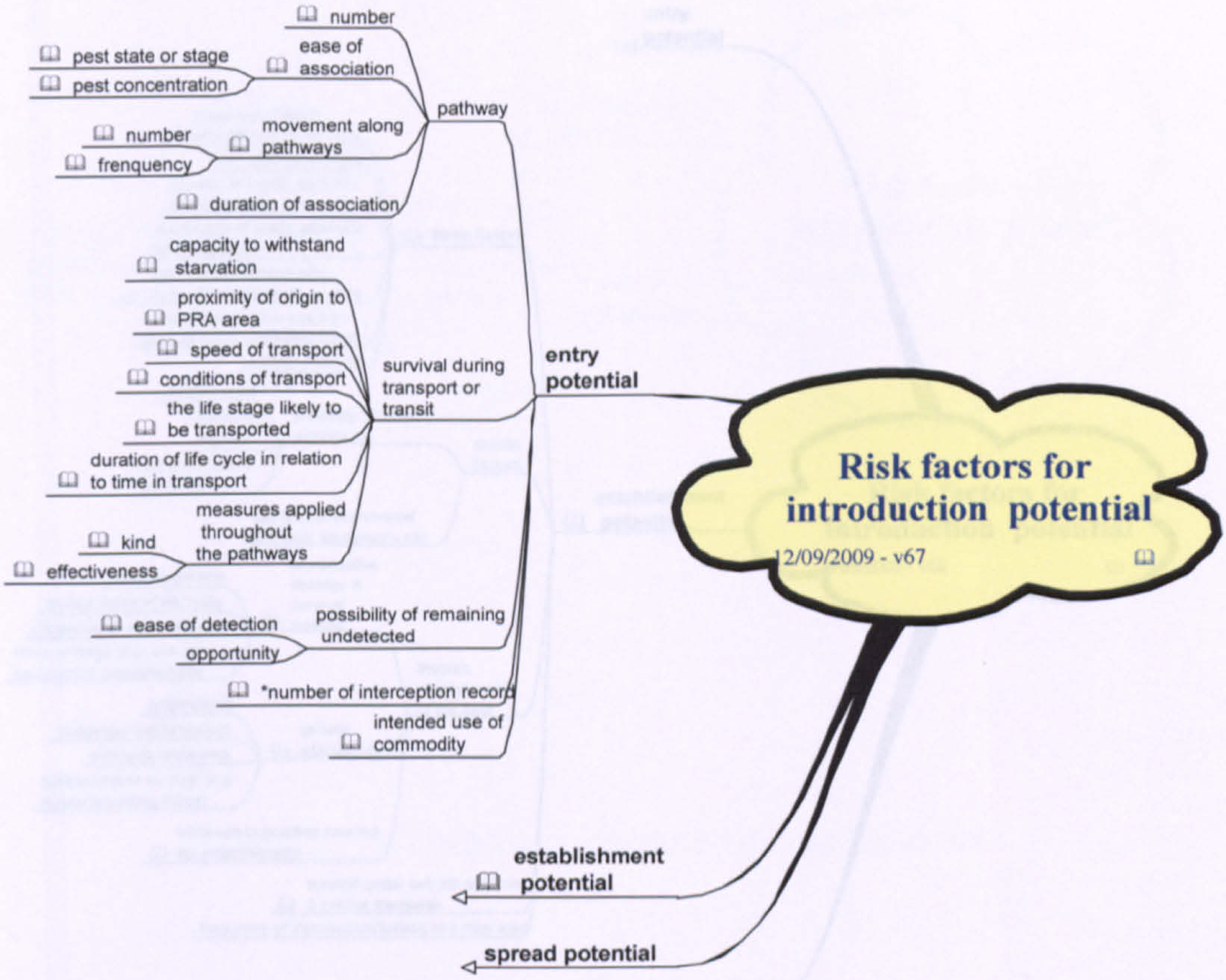
3.2.3.1 Risk factors for introduction/spread potential

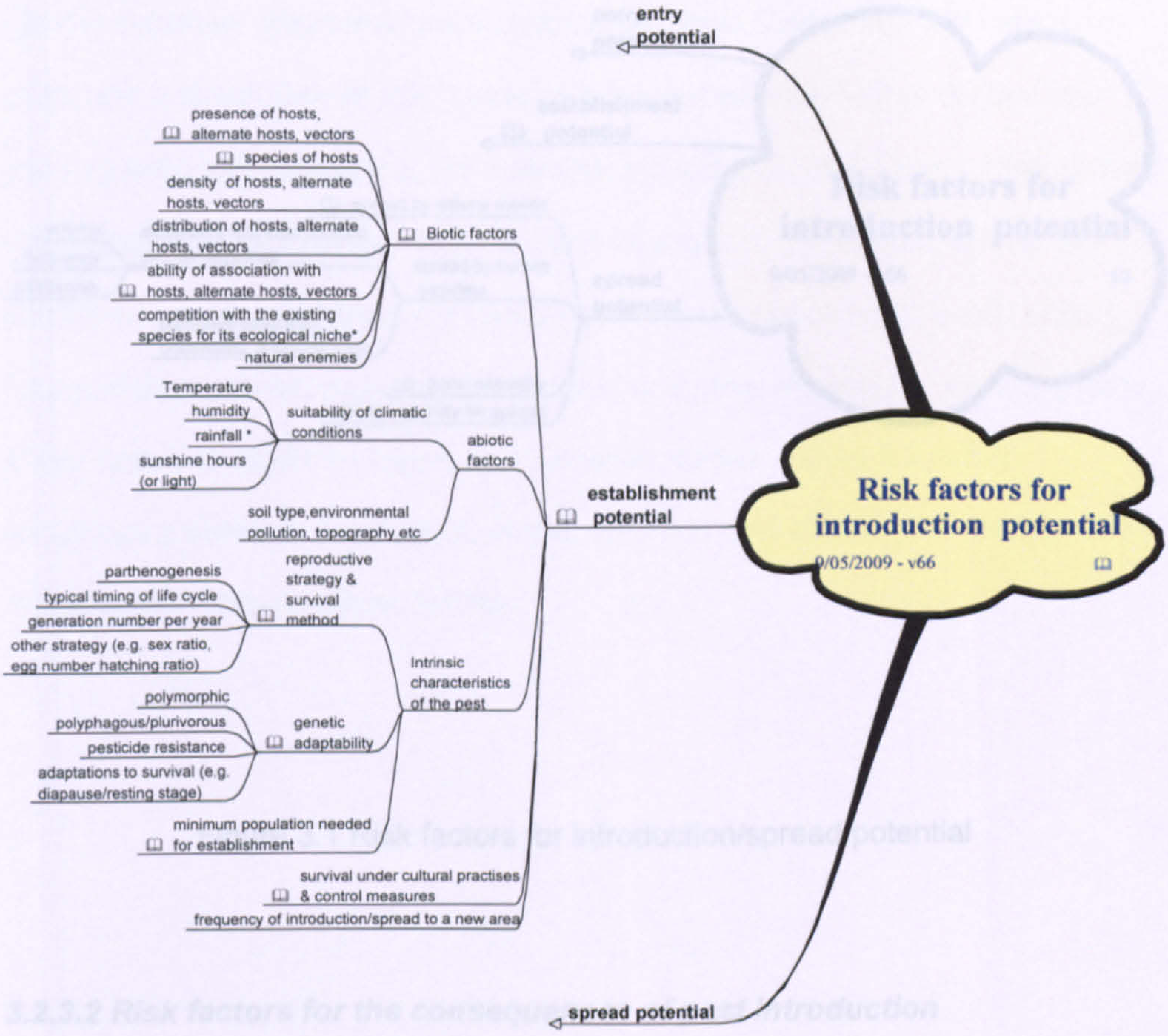
The case studies initially concentrated on insect pests. Following FAO guidelines, approximately 40-45 risk factors were identified for the criterion of introduction potential. These risk factors were further divided into three main categories: entry potential, establishment potential and spread potential. Each of the three main categories was further disaggregated into increasingly specific factors in subsequent branches. For example, entry potential was further divided into five sub-categories: entry pathway, survival during transport or travel, possibility of remaining undetected, number of interception, and intended use of commodity. Where applicable, sub-categories were divided further. For example, entry pathway was further divided into four factors: the number of pathways; the ease of association, which was further divided into pest state or stage and pest concentration; movements along pathways, including number and frequency of movements; and duration of association. Figure 3.1, which comprises a series of mind maps, shows the general mind map for the risk structure of the introduction/spread potential. Most of the risk factors had an underlying question, marked as a "bookmark"[4]. For example, for the risk factor duration of association, the question "Do the time of the stage or state of the pest, which would be associated with pathways, last long?" was asked and the following note, "Introduction at many different times of the year will increase the likelihood that the pest will be at a stage or state to expose to a suitable host for establishment" was recorded.

---

[4] Bookmarks can be found in the mindmaps throughout the thesis. For example, the book symbol beside "establishment" in the mindmap on next page. A bookmark means there is extra information attached to that branch, it may be a question, or a note.


establishment potential
entry potential
Risk factors for introduction potential
9/05/2009 - v66
spread potential


Risk factors for introduction potential
12/09/2009 - v67
entry potential
pathway
number
ease of association
pest state or stage
pest concentration
movement along pathways
number
frenquency
duration of association
survival during transport or transit
capacity to withstand starvation
proximity of origin to PRA area
speed of transport
conditions of transport
the life stage likely to be transported
duration of life cycle in relation to time in transport
measures applied throughout the pathways
kind
effectiveness
possibility of remaining undetected
ease of detection
opportunity
*number of interception record
intended use of commodity
establishment potential
spread potential



3.1 Risk factors for introduction/spread potential

3.2.3.2 *Risk factors for the consequences of pest introduction*

The assessment of the consequances of pest introduction is another challenging feature in the context of PRA because of the difficulty in the estimation of the environmental and social, as well as economic factors, which may be important in some cases. There was not much guidance on how to assess the economic impact caused by pest introductions in the existing international standards and regional guidelines. Based on the FAO guidelines and case studies, this study identified approximate 50 factors for assessing economic consequences. These factors account for both direct and indirect pest impacts. Similar to the identification of risk factors for introduction potential, the mind mapping process started from the more general objective of assessing economic consequence of

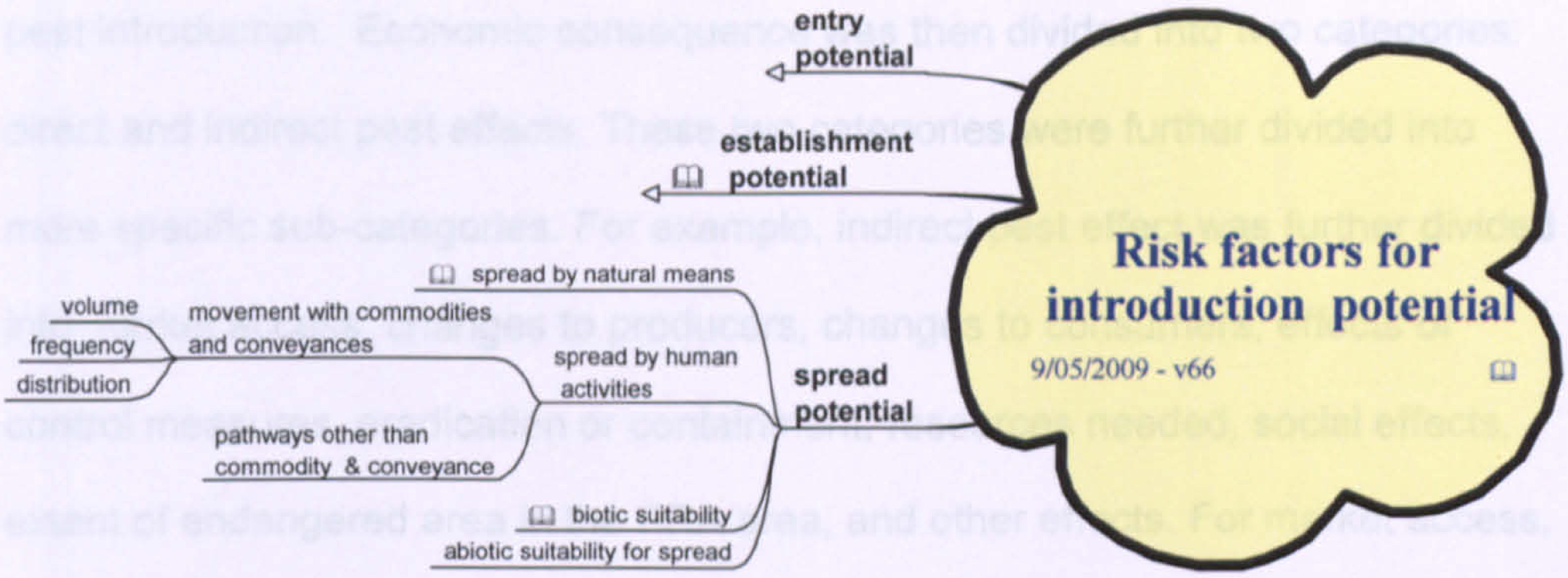
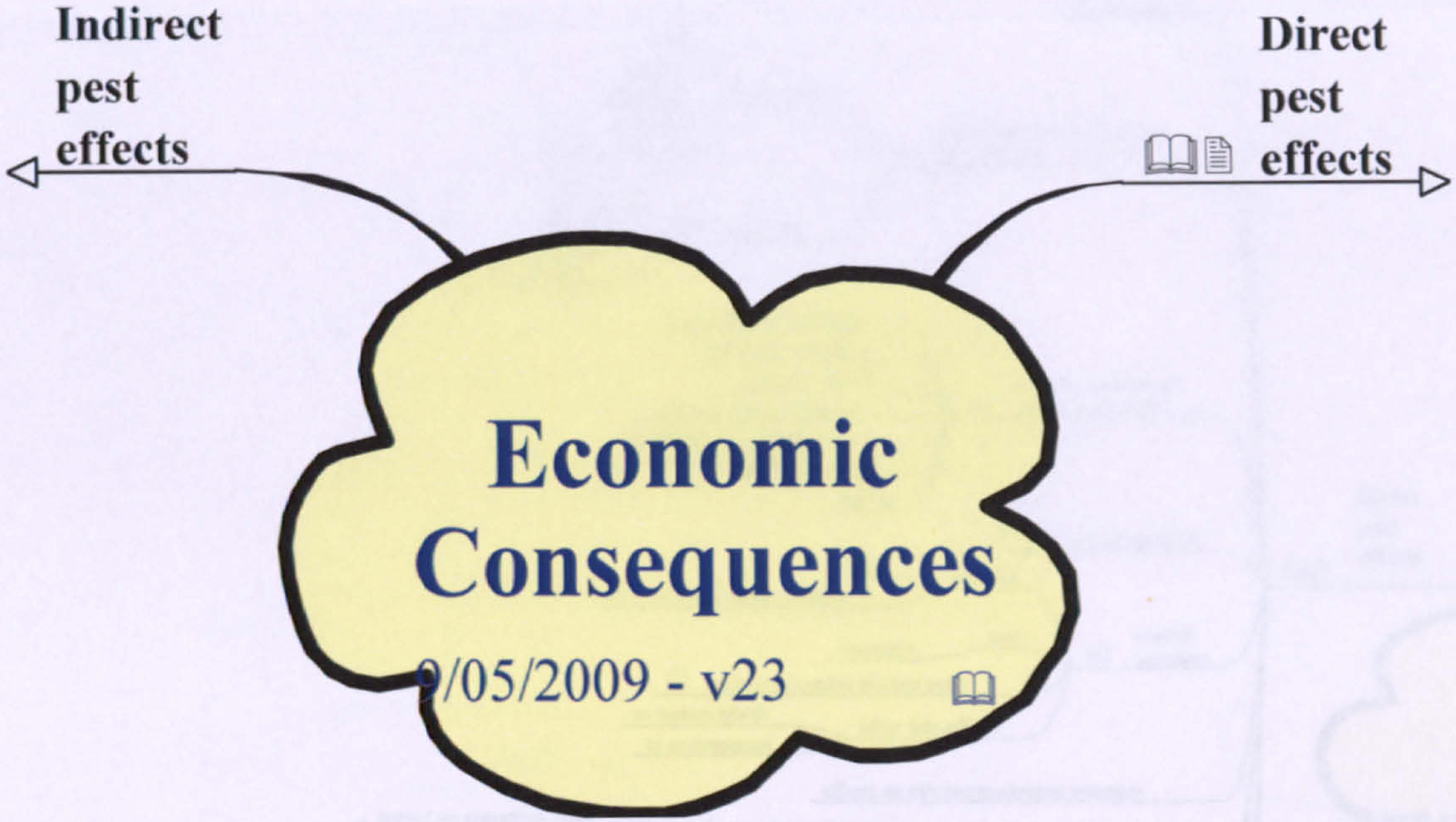


Figure 3.1 Risk factors for introduction/spread potential

3.2.3.2 Risk factors for the consequences of pest introduction

The assessment of the consequences of pest introduction is another challenging feature in the context of PRA because of the difficulty in the estimation of the environmental and social, as well as economic factors, which may be important in some cases. There was not much guidance on how to assess the economic impact caused by pest introductions in the existing international standards and regional guidelines. Based on the FAO guidelines and case studies, this study identified approximate 60 factors for assessing economic consequences. These factors account for both direct and indirect pest impacts. Similar to the identification of risk factors for introduction potential, the mind mapping process started from the more general objective of assessing economic consequence of pest introduction. Economic consequence was then divided into two categories: direct and indirect pest effects. These two categories were further divided into more specific sub-categories. For example, indirect pest effect was further divided into market access, changes to producers, changes to consumers, effects of control measures, eradication or containment, resources needed, social effects, extent of endangered area in the PRA area, and other effects. For market access, it was further divided into export and domestic market. Figure 3.2, which comprises a series of mind maps, shows the mind map structure of considerations for economic consequences in PRA.

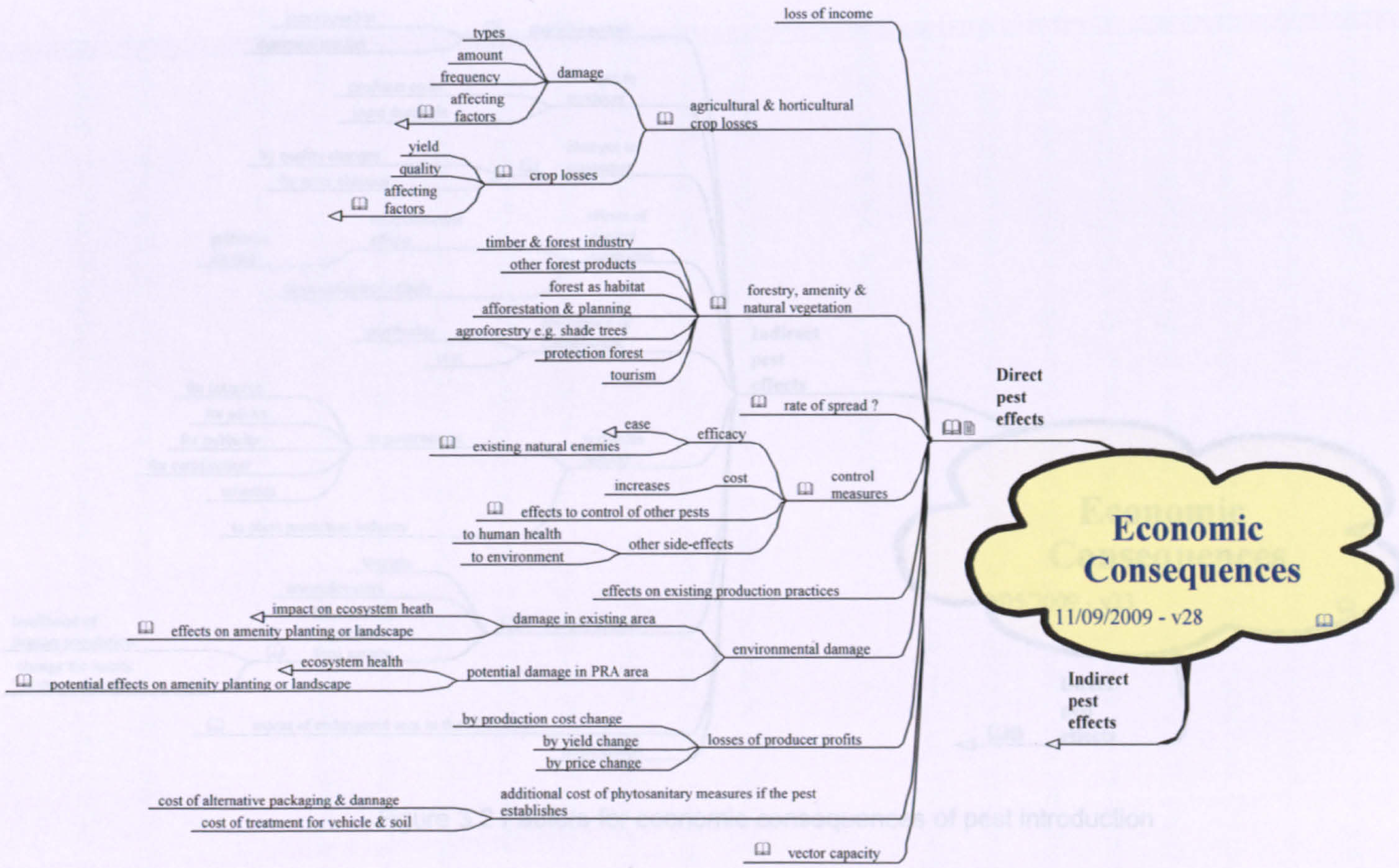


Economic Consequences

9/05/2009 - v23

Indirect pest effects

Direct pest effects

Economic Consequences
11/09/2009 - v28
Direct pest effects
Indirect pest effects
loss of income
agricultural & horticultural crop losses
damage
types
amount
frequency
affecting factors
crop losses
yield
quality
affecting factors
forestry, amenity & natural vegetation
timber & forest industry
other forest products
forest as habitat
afforestation & planning
agroforestry e.g. shade trees
protection forest
tourism
rate of spread ?
control measures
efficacy
ease
existing natural enemies
cost
increases
effects to control of other pests
other side-effects
to human health
to environment
effects on existing production practices
environmental damage
damage in existing area
impact on ecosystem heath
effects on amenity planting or landscape
potential damage in PRA area
ecosystem health
potential effects on amenity planting or landscape
losses of producer profits
by production cost change
by yield change
by price change
additional cost of phytosanitary measures if the pest establishes
cost of alternative packaging & dannage
cost of treatment for vehicle & soil
vector capacity

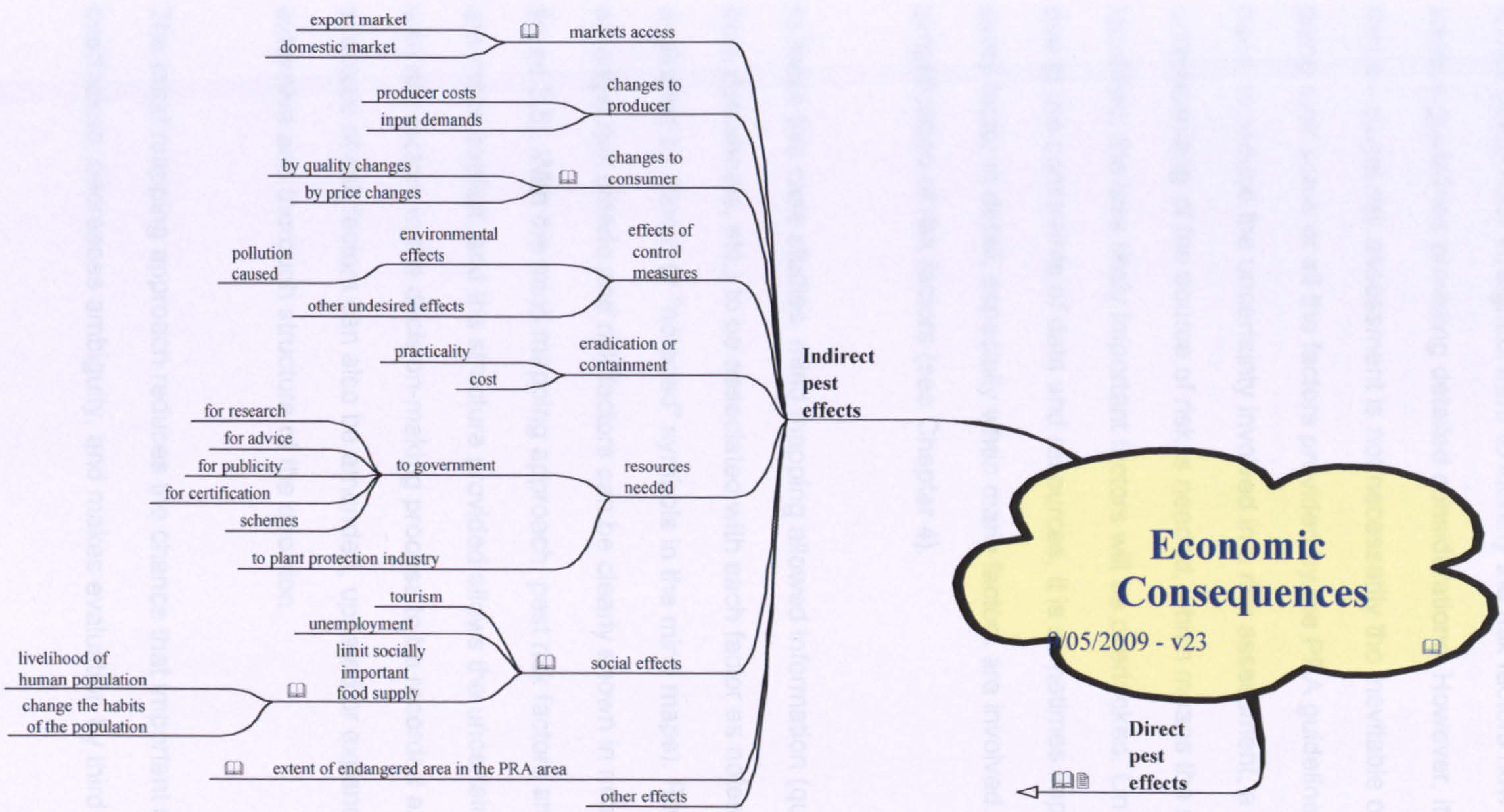


Figure 3.2 Factors for economic consequences of pest introduction

3.2.4 Conclusions

It may seem fairly straightforward to identify the risk factors now that there are various guidelines providing detailed considerations. However, it is argued here that a rational risk assessment is not necessarily the inevitable outcome of simply going over some or all the factors provided by the PRA guidelines. On the one hand, to reduce the uncertainty involved in a risk assessment, a thorough understanding of the source of risk is needed, which means the more factors identified, the less likely important factors will be overlooked. On the other hand, due to the constraints of data and resources, it is sometimes impossible to assess every factor in detail, especially when many factors are involved. This requires the simplification of risk factors (see Chapter 4).

In these two case studies, mind mapping allowed information (questions, facts from datasheets, etc.) to be associated with each factor as notes or hyperlinks (indicated by "book" or "notepad" symbols in the mind maps). Relationships amongst risk criteria and risk factors can be clearly shown in mind maps (see figure 3.5). With the mind mapping approach, pest risk factors and assumptions are made explicit, and the structure provided allows the uncertainties associated with risk factors in the decision-making process to be recorded and reviewed. The structure of risk factors can also be amended, updated or expanded to a more extensive and thorough structure of the decision.

The mind mapping approach reduces the chance that important risk factors are overlooked, decreases ambiguity, and makes evaluation by third parties easier. It also encourages debate and allows others to see where important differences exist in risk factors, assumptions, weightings, scores, and uncertainties. This enables critical comments and reviews to be obtained, and facilitates re-analysis or re-examination. Thus, the quality and depth of PRA can be improved. Mind mapping also highlights dependencies among risk factors and helps to distinguish factors that are manageable ("control points") as an aid to the selection of risk management measures (Zhu *et al.*, 2000, 2001).

3.3 PEST RISK EVALUATION (PRE)

3.3.1 Introduction

Following the identification of risk factors and sometimes the elimination of some factors by simplification of the assessment (see Chapter 4), the next logical step is the evaluation (quantification) of the level of risk. In some fields, risk quantification is defined as "the estimation of a given risk by a statistical and/or analytical modelling process" (Warner, 1992). Risk evaluation or quantification is the detailed assessment of the potential impact of the risks identified, and often involves sophisticated statistical calculations in an attempt to accurately predict the potential impact of the risks (Warner, 1992).

In this thesis, it is proposed that "pest risk evaluation" is defined as "the process of estimating the degree of risk of the introduction of a pest and its subsequent consequences". In this context "risk" usually refers to the introduction (entry and establishment) and spread of an economically damaging pest, as well as its consequences. It is intended that at the end of PRE, an overall level or statement of risk would be achieved.

In this thesis, it is proposed that PRE is subdivided into three phases: scoring risk factors, weighting risk factors, and combining risk scores.

The EPPO scheme indicated that the numerical scores may be combined, weighted and averaged in appropriate ways, but no particular method of calculation was specially recommended (EPPO, 1997). Therefore, one objective of this chapter was to review and develop methods to achieve an overall risk assessment in the form of a numerical result that was logically and biologically meaningful.

3.3.2 Scoring risk factors

3.3.2.1 The purpose of scoring the risk

It is popular to use a risk score to convey the level of risk in current pest risk assessment practice (e.g. EPPO, USDA, Canadian guidelines); this is usually achieved by assigning a number to a specific risk factor according to the available information. The number is a point on an ordinal scale, a high score meaning a greater risk. Figure 3.3 shows the role of risk scoring in a risk assessment.

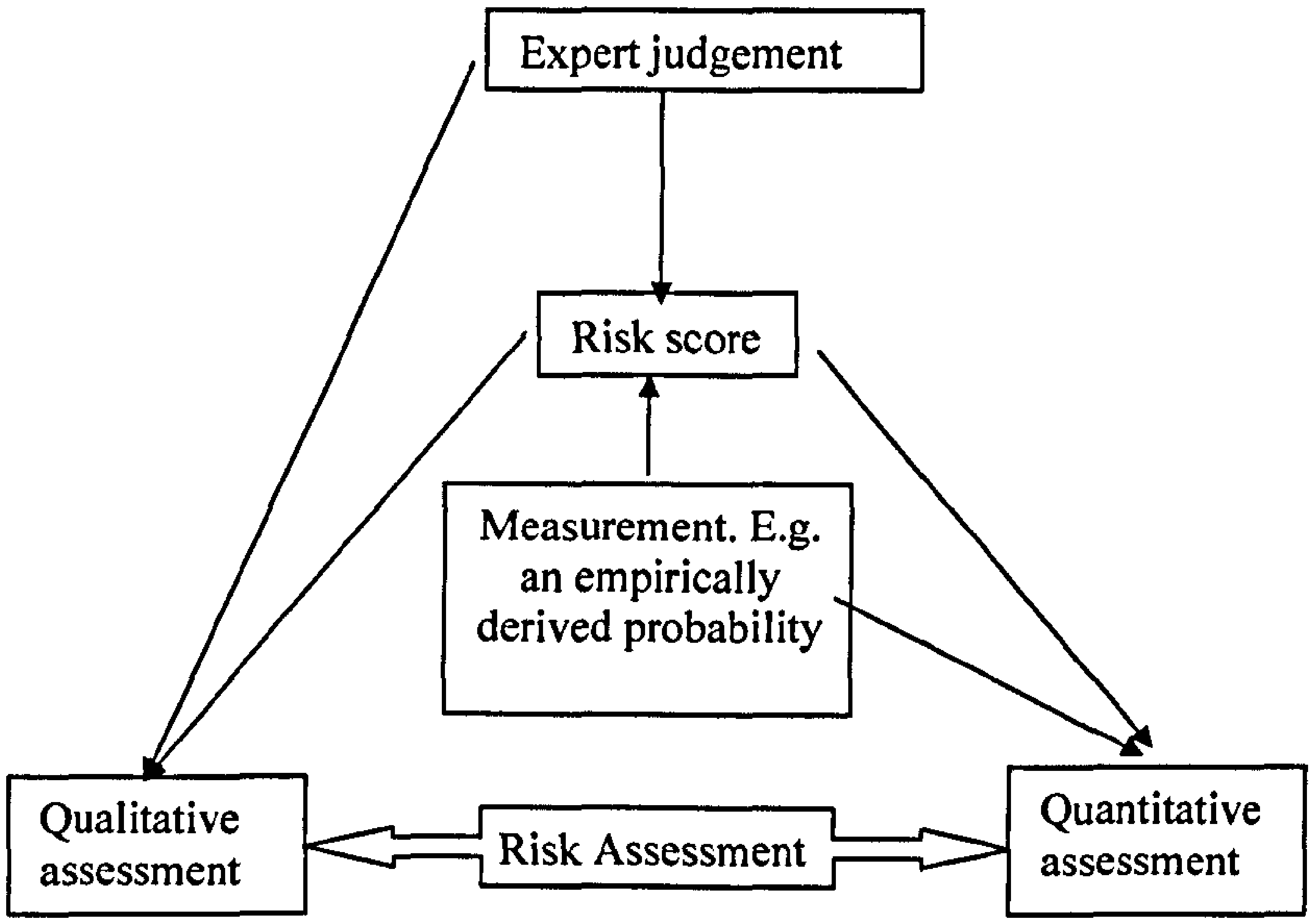


Figure. 3.3 The role of risk scoring in risk assessment

As shown in Figure 3.5, a risk score can express both subjective and objective information. It can derive from both expert judgment and empirical measurement. It can then feed into both qualitative and quantitative pest risk assessment. Therefore the advantage of using a scoring system is that it gives a general idea of the degree of risk under a particular circumstance and allows the use of simple statistics like the sum or the mean. The limitation is that the score is generally subjective and can't be easily validated.

3.3.2.2 Scoring system in current PRA practice

Currently, there are several scoring systems used in PRA practice: 1-9, 0-3, 1-3 and 0-4 scoring systems. A greater number of intervals allow greater precision or at least the appearance of greater precision. It is pertinent to ask, however, what degree of gradation is appropriate for the precision of the assessor's judgment.

1 - 9 scoring system. The advantage of a 1-9 scoring system is that it allows a user a relatively high degree of resolution to express the perceived different levels of risk. However, there is a danger that the resolution of the scale exceeds that of the judgment concerned. Since a score is subjectively assigned, often from a mixture of information sources (objective plus subjective), it is possible that the same risk analyst may give a different score to the same factor based on the same information on different occasions.

A 1-9 scoring system was initially adopted in the EPPO's PRA scheme Section B: Quantitative Evaluation. In this section, a series of questions are presented (EPPO, 1997). Replies to these questions are expressed as scores on a 1-9 scale, i.e. the questions require an evaluation from minimum probability or impact (1) to maximum probability or impact (9). For example, one question related to entry potential is "1.1 How many pathways could the pest be carried on? A rough guidance for scoring was given: few = 1; many =9, this question can be scored from 1 to 9 depending on the number of the pathways identified.

MacLeod and Baker (2003) provided descriptions for each number on the 1-9 scale to the questions in the EPPO scheme. Such description would not only simplify the task faced by risk assessors, but also serve to standardise responses, thus enhancing the ability to compare the risks posed by different pests. It was concluded that it was possible to provide descriptions for each score on the 1-9 scale for questions where assessments can be based on a numerical value. For questions that can only be answered with an expert opinion in the form of words, such as low, medium or high, only general guidance can be given (MacLeod and Baker, 2003).

Following the EPPO scheme the 1-9 scoring system was used in this thesis for all the case studies. In the case studies, risk scores were given to each risk factor[5]. Scores for each risk factor were personal judgments based on available information and experience.

1 - 3 scoring system. The advantage of 1- 3 system is that it is simple in use, i.e. a user only needs to characterise the risk as low, medium and high, which correspond to 1, 2 and 3, respectively. However, it is felt that it doesn't give sufficient scope to express more subtle differences. For instance, one may argue that although risk may be characterised as medium, one factor may has a low/medium risk and the other may be better expressed as a medium/high risk.

Black and Abdallah (1997) applied the 1-3 scoring system in a preliminary PRAs for Tanzania, in which a three-point 1-3 scoring system was adopted, where 1 = low risk, 2 = medium risk and 3 = high risk; other characters were used for uncertainties. Table 3.1 shows an example of the keys for scoring.

---

[5] The factors on the most disaggregated level of branches in the Mind Map

Table 3.1 The key for scoring geographic and regulatory status (after Black and Abdallah1997).

Score	Key
1	Fully distributed in Tanzania and not subject to control
2	Still spreading/subject to control
3	Not present at all
X	Status uncertain
Y	Further information required

An extension of the 1-3 scoring system is 0-3 scoring system. Based on the 1-3 scoring system, zero is assigned to a risk factor if the associated risk is negligible. 0-3 scoring system has been adopted in the Canadian PRA scheme. Table 3.2 shows example guidelines for rating establishment potential in the Canadian PRA Scheme.

Table 3.2 Guidelines for rating establishment potential – based on the combination of all relevant factors (after Watler 2002)

Risk rating	Criteria
Negligible (0)	Will not survive in Canada e.g. Stewart's wilt of corn
Low (1)	One third or less of range of host(s) in Canada e.g. Oriental fruit moth
Medium (2)	One to two thirds of range of host(s) in Canada e.g. Blueberry maggot
High (3)	Most or all of range of host(s) e.g. Soybean cyst nematode

A 0-4 scoring system was adopted in the UK non-native risk assessment scheme (DEFRA 2005). In this scheme, the assessor is required to choose one of five levels of responses, i.e. very low - 0, low - 1, medium - 2, high - 4, very high –5, justifying these with a written, referenced comment.

3.3.2.3 Subjectivity and reliability of risk scores

One problem with the notion of scoring in general is the subjectivity of linking the notion of the risk to the risk score. Hence, the reliability of a risk score depends on the person who interprets the available information, who brings to bear his/her skill, knowledge and view. As a result it is possible or even likely that different people will give a different score for the same situation. It is therefore important to define as far as possible rules of how a score is given, which link the notion of risk to the score. For example, in 1 –9 scoring system, risk scores 7, 8 and 9 are all regarded as a high risk, a user may not tell the difference between scoring 7 and 8 or 8 and 9.

To deal with the problem, one solution is to develop a transparent and consistent standard for ratings, in that "using actual examples for comparison wherever possible, which means that two assessors evaluating the same information should reach the same rating, and that people reading the assessment should be able to follow the process" (Watler 2002)[6].

---

[6] Watler (2002) on Canadian PRA Schemes.

The second possible solution is to use Delphi study: a group of PRA staff assess the risk for the same pest or commodity individually; present their results and discuss; amend their assessments again according to the discussion and reach a consensus. In practice EPPO uses a Delphi system, i.e. PRAs are done by a panel of experts working independently followed by meetings/discussions, until an agreed assessment is achieved (McNamara 2002, pers. comm.). The author has done this with UK PRA staff on Asian longhorned beetle and with EPPO staff on several forestry pests. In each case a consensus was reached.

3.3.2.4 Scoring risk factor: a case study of Asian longhorned beetle from China to Europe

A case study on PRA for Asian longhorned beetle (ALB) from China to Europe is shown here to illustrate how to score risk factors. ALB is a major pest in China, Korea and Japan, where it kills many species of broadleaved trees, such as maples (including sycamore), poplars, alders, willows, cherries, apples, horse chestnut, elm, mulberry, boxelder, etc. The larval stages of ALB are well protected within untreated wood and, therefore, it is possible for the beetle to be carried via international trade and to emerge at the final destination.

It became established in New York (discovered in 1996) and Chicago (discovered in 1998). The beetle was believed to have arrived in North America in the wooden packing material used in cargo shipments from China. In both New York and Chicago, damage to street trees was high. Cutting down the infested trees, sanitation and quarantine were being exercised as the only viable management option. By 2001, the US federal government had already mapped out a $365 million plan using a tree vaccine, armies of bug spotters and even acoustic tools in the hopes the ALB would be eradicated in New York and Illinois by 2009. A study by the USDA Forest Service determined that if the ALB became established across the US, it would probably kill 30% of all urban trees – at a compensatory value of $669 billion.

In the UK, specimens of ALB have been intercepted at several locations. As yet, there was no evidence to indicate that it had successfully attacked trees in Britain. Analysis of climate data by scientists at the Central Science Laboratory suggested that most of England and Wales and some warmer coastal areas of Scotland were suitable for beetle establishment and breeding (MacLeod *et al.*, 2002). The greatest risks came from the presence of the beetle in packaging material associated with a very wide range of commodities from China. Extensive damage to both urban and woodland/forest trees was expected if the beetle established in Britain.

A detailed PRA of ALB from China to Europe following the EPPO scheme is given in Appendix 3. This PRA was done in 2000 by the author based on the biological and phytosanitary characteristics of the ALB, and discussion with members of the CSL staff[7]. In this case study, of the 12 risk factors being considered for assessing entry potential, the likelihood of survival in transit, and spread through commodity

---

[7] Baker and Macleod eventually published paper son ALB PRA.

distribution were given the highest scores, while the likelihood of multiplication during transit were given the lowest risk score.

Of the 13 risk factors being considered for assessing establishment potential, wild plants aid dispersal and maintenance, the lack of natural enemies were assigned the highest scores, similarity in climatic conditions and other abiotic factors, difficulty in controlling, and lack of competition from existing species were given the second highest scores, while introduction to a new area was given the lowest score as there were only two cities in the US.

Of the 19 risk factors considered for assessing economic impacts, lack of natural enemies was given the highest score, environmental damage in existing area and difficulty in control were given the second highest scores, whereas disruption of control measures to other pests and pest resistance were given the lowest scores.

Certain issues were raised from this case study:

- The ambiguities in the main risk criteria needed to be clarified (see Appendix 1);
- Weightings needed to be introduced to reflect the different importance of each risk factors (see 3.3.3 and Chapter 4);
- Methods were needed for deriving a meaningful overall risk score other than just simply summing the risk scores of each factors (see 3.3.4);
- Once a single value of overall risk was obtained, it needed a meaningful interpretation (see Section 3.3.5).

3.3.3 Weighting risk factors

3.3.3.1 Incorporating weighting into pest risk assessment

After identifying and scoring risk factors, it is possible to make an overall assessment from the evaluation of individual factors. What makes this difficult is that the criteria are probably not all equally important.

Given that risk criteria have been disaggregated into a considerable number of specific risk factors during PRI, it is likely that some factors are more important and some less important. If simply combined by an average into an overall risk, the consequence might be that no overall risk assessment will be particularly high or low. However, the reality is that not all the risk factors contribute the same to the overall risk: some risk factors are more important than others and it is appropriate that they contribute more to the assessment. This is illustrated by a somewhat simplified example in Table 3.3.

Table 3.3. A hypothetical example illustrates the impact of simple average of risk factor scores

Risk factor	Factor 1	Factor 2
Risk score	1	9
Weighting	0.2	0.8
Simple average	5	
Weighted average	7.4	

A possible solution is to introduce weighting into risk assessment. A weighting is a value, which is given to something according to how important or significant it is. (Collins Cobuild English Dictionary 2001).

The use of weighting has hardly been explored in PRA. The EPPO PRA scheme suggested that:
" The numerical scores may be combined, weighted and averaged in appropriate ways that may enable the assessor who uses them consistently to make useful comparisons between pests, pathways and hosts. No particular mode of calculation is specifically recommended by EPPO."(EPPO, 1997).

The possibility of incorporating weighting into pest risk assessment practice was explored here to reflect the impact of perceived importance of each risk factor, such that a risk assessment is more biologically and economically meaningful.

3.3.3.2 Methods developed to obtain weightings - subjectively assigned weightings

In the following sections, different approaches to the derivation of weighting using case studies are proposed:
Weighting subjectively assigned by a single assessor;
Weighting obtained by consensus of expert opinions from a pool of specialists.

To investigate the use of subjective weighting systems, a weighting was assigned to each risk factor to reflect its perceived importance according to the author's experience and knowledge. Two subjective weighting systems were investigated: 0-1 weighting system and 0-3 weighting system. Two case studies were used to present this idea. In these case studies, risk factors were presented in the format of mind maps, structured as a hierarchy from general to specific. Risk scores were given in the range of 1- 9.

a. 0 – 1 weighting system.

Weightings were assigned to each risk factor at the top level (see mind maps) in the range of 0 - 1 such that the sum of weightings for the factors at that level equalled 1. The same approach was repeated for successive levels down to the risk factors at the bottom level. Therefore for each risk factor, its overall weighting will be the product of a hierarchy of weightings (from top to bottom levels). In allocating the weightings, risk factors were considered or compared only with those factors in the same level of the hierarchy and independent of other levels.

Fig 3.4 shows a case study on the *introduction and spread potential* of Natal fruit fly (*Ceratitis rosa*) from East Africa to Europe. The first level of introduction/spread potential comprised three factors: *entry, establishment and spread*, the assigned weightings were 0.4, 0.4 and 0.2, respectively.

Each factor was disaggregated in stages and weightings (0-1) assigned at each stage of the hierarchy such that sum of the weightings at each branch-point was equal to 1. For example, *establishment* is partly decided by *abiotic factors*, which is again partly decided by *host range*. The overall weighting for *host range* in Figure 3.6 was calculated as: (weighting for host range 0.2) * (weighting for biotic factors 0.3) * (weighting for establishment potential 0.4) = 0.024.

Each factor was calculated in the same way, such that the sum of the weightings for all risk factors at the same level equalled 1, no matter how many risk factors were considered. Thus the overall risk score was normalized and independent of the number of risk factors.

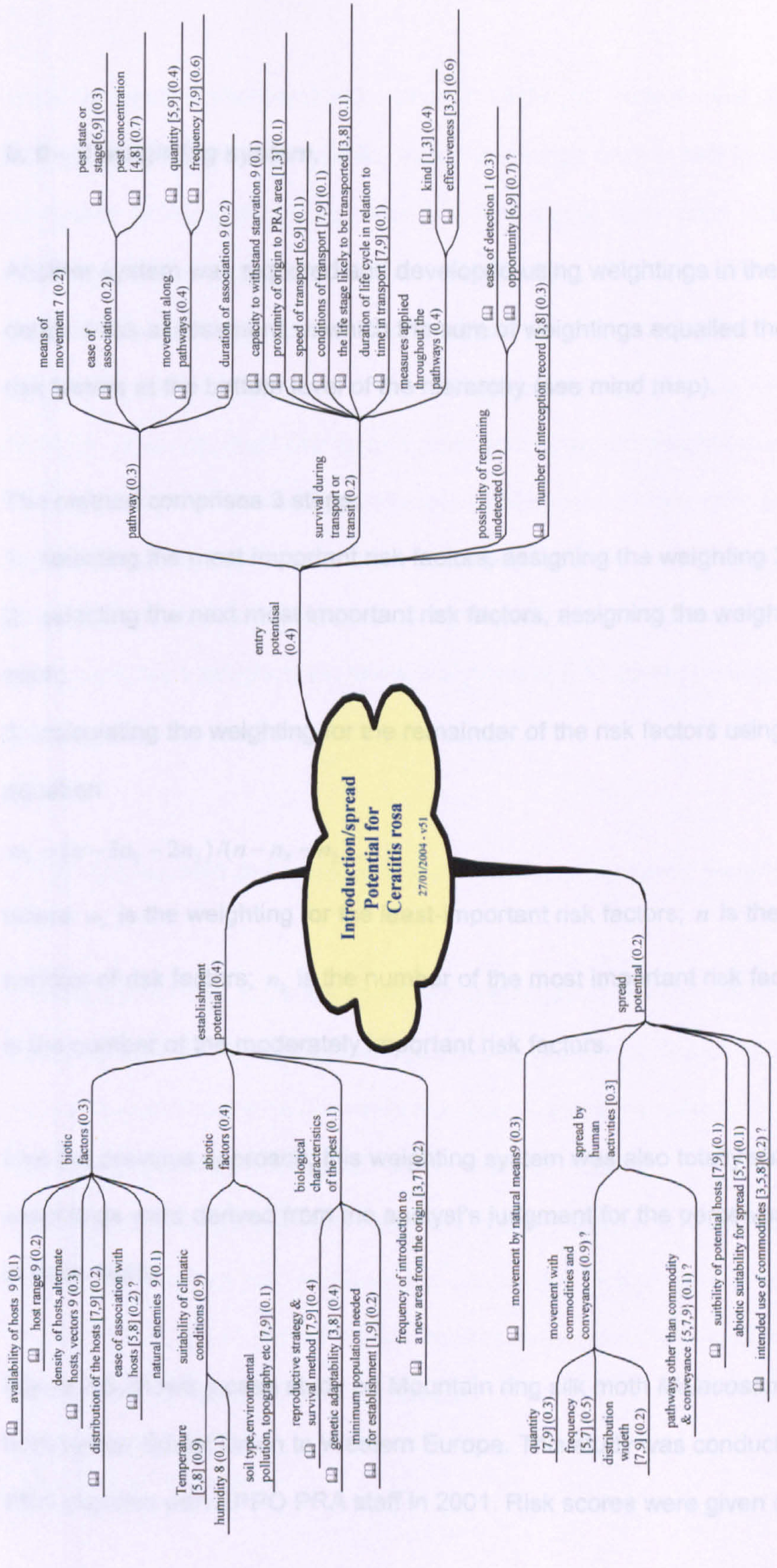


Figure 3.4 Introduction/spread potential for *Ceratitis rosa* from East Africa to Europe.

b. 0 – 3 weighting system.

Another system was explored and developed using weightings in the range 0-3 for detailed risk assessment, in which the sum of weightings equalled the number of risk factors at the bottom level of the hierarchy (see mind map).

The method comprises 3 steps:

1. selecting the most important risk factors, assigning the weighting 3 to each;
2. selecting the next most important risk factors, assigning the weighting 2 to each;
3. calculating the weighting for the remainder of the risk factors using the equation:

$w_r = (n - 3n_3 - 2n_2)/(n - n_3 - n_2)$,

where w_r is the weighting for the least-important risk factors; n is the total number of risk factors; n_3 is the number of the most important risk factors and n_2 is the number of the moderately important risk factors.

Like the previous approach, this weighting system was also totally subjective and weightings were derived from the analyst's judgment for the perceived importance of each factor.

Figure 3.5 shows a case study on Mountain ring silk moth *Molacosoma parallela* from former Soviet Union to Western Europe. This study was conducted during a PRA exercise with EPPO PRA staff in 2001. Risk scores were given on a 1 – 9 scale (numbers inside brackets). Although all the risk factors were structured as a hierarchy from general to specific, factor weightings were in fact assigned / calculated following the above process. For example, host range, ability of association with hosts, temperature, economic loss, effect on yield or quality, quarantine treatment in source area, capacity to survive during transport, spread by natural means, distribution, and extension of endangered area were perceived[8] to be the most important risk factors, and were assigned weightings of 3. Density of hosts, means of movement, and movement along pathway were perceived as the next most important risk factors, and were assigned weightings of 2. The perceived important risk factors were highlighted in the mind map. All the other risk factors were assigned the same weighting of 0.17 based on the calculation of weightings for the least important risk factors.

3.3.3.3 Issues with the approach of subjectively assigned weighting

It was identified that there are some disadvantages of this approach:

The assessment of weightings was subjective, and the choice of values for the important and moderately important risk factors was constrained;

the hierarchy structure of the risk factors was not utilised;

less important factors were all assigned the same weighting, which depended on how many important and moderately important risk factors there were;

less important factors did not usually contribute much to the overall risk assessment, yet still had to be assessed.

---

[8] Based on the author's judgment.

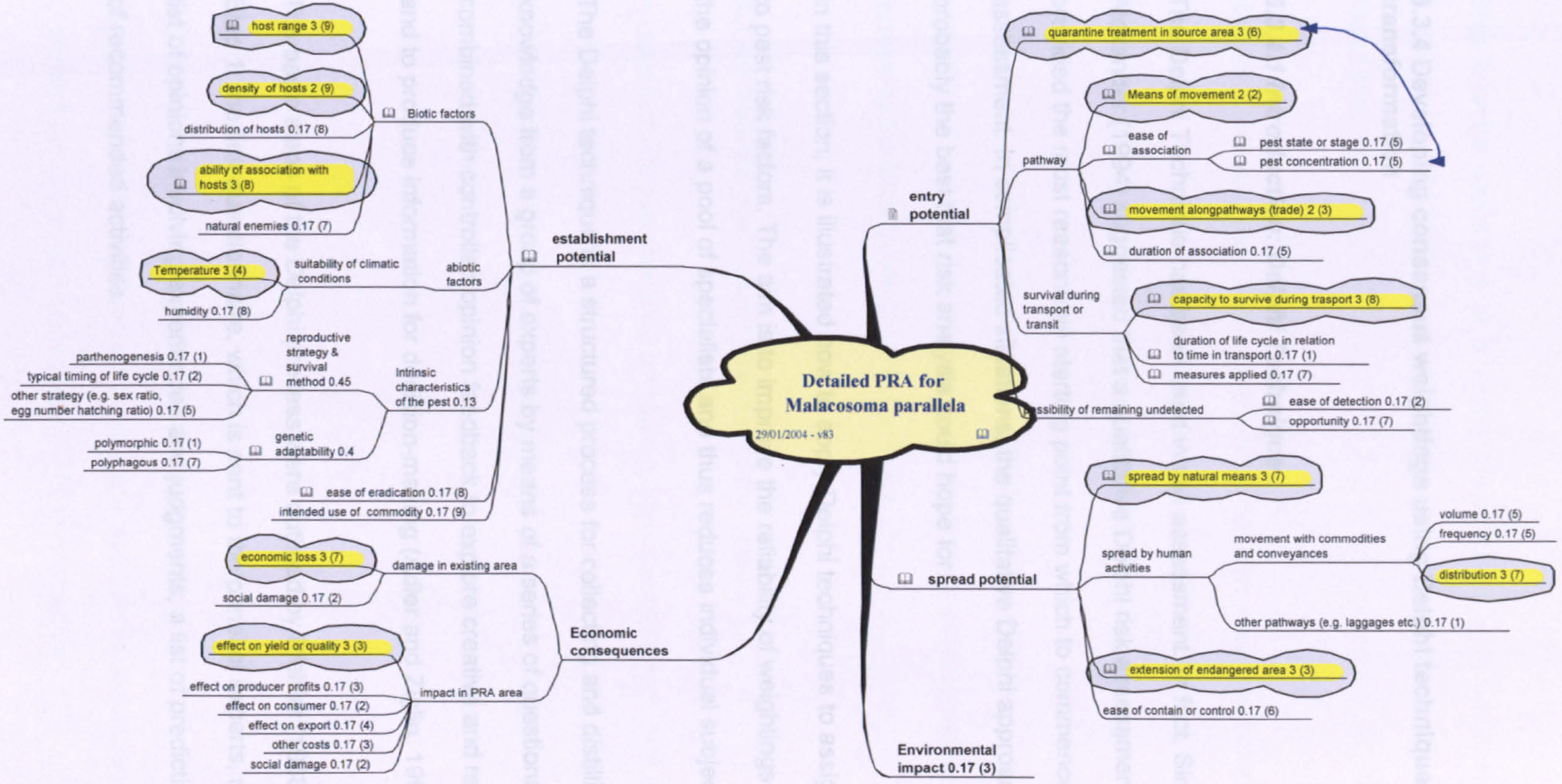


Figure 3.5 Detailed risk assessment for *Malacosoma parallela*

3.3.4 Developing consensus weightings using Delphi technique and data transformation

3.3.4.1 Introduction: Delphi Technique

The Delphi Technique has been used in risk assessment. In fact, Simberloff and Alexander (1994) suggested that a qualitative Delphi risk assessment procedure provided the most reasonable starting point from which to commence risk assessment. In complicated situations, the qualitative Delphi approach was probably the best that risk analysis could hope for.

In this section, it is illustrated how to apply Delphi techniques to assign weightings to pest risk factors. The aim is to improve the reliability of weightings by surveying the opinion of a pool of specialists and thus reduces individual subjectivity.

The Delphi technique is a structured process for collecting and distilling knowledge from a group of experts by means of a series of questionnaires combined with controlled opinion feedback to explore creative and reliable ideas and to produce information for decision-making (Adler and Ziglio, 1996).

The basic steps of the Delphi process were outlined by Pfeiffer (1968):
Step 1: the first questionnaire, which is sent to the panel of experts, may ask for a list of opinions involving experiences and judgments, a list of predictions, and a list of recommended activities.

Step 2: on the second round, a copy of the collective list is sent to each expert and the expert is asked to rate or evaluate each item by some criterion of importance.

Step 3: the third questionnaire includes the list, the ratings indicated, and the consensus, if any. The experts are asked to either revise their opinions or discuss their reasons for not coming to consensus with the group.

In this thesis, the first and second steps of the Delphi study were conducted, which aimed to develop a methodology for collecting expert judgment and assessing the importance of each risk factor.

3.3.4.2 Method: the Questionnaire and the survey

A questionnaire was designed and sent to plant quarantine officials and risk analysts in Africa, EPPO region and USDA (Appendix 4). Each was asked to rank the risk factors. After receiving the responses, data analysis was conducted, method for obtaining weightings from the expert survey was developed and, finally, weightings for factors for introduction/spread were derived.

In this study, step 1 was to compile a list of risk factors for the questionnaire. Risk factors were drawn from ISPM No 11 and the EPPO RA scheme, as ISPM No 11 and EPPO RA scheme were enforced international/regional standards, which had been consulted with the member countries, it was safe to assume that they presented the opinion of the community of the risk analysts.

A questionnaire was subsequently designed[9], which comprised four parts, concerning the main risk criteria, risk factors, rules for scoring the risks, and risk management options. This questionnaire focused on pest risk analysis in general rather than on pest risk analysis for a specific pest[10].

It was expected that the experts would rank the risk factors, comparing with the other risk factors at the same hierarchical level. Risk factors at the same hierarchical level were to be ranked from the most important (1) to the least important (the ranking of which depends on the number of risk factors at the same hierarchical level). Risk factors can be ranked as equally important. New factors could also be added. (See Appendix 4: Delphi study – a survey for pest risk analysis: weighting risk elements according to importance, for the details of the questionnaire).

The questionnaire was then circulated to plant quarantine officials in Africa, EPPO region, and USDA/PHIS, in order to obtain their opinion on the importance of risk factor.

Seventeen responses were received. The expert responses are shown in Table 3.4, which has been slightly modified. For example, in evaluating the entry potential, the fifth expert gave the same rank of 1 to *the number* and *the ease of*

---

[9] Questions were formulated with the help of Dr Richard Baker from CSL (Central Science Laboratory, UK), and were based on the FAO PRA guidelines and the EPPO pest risk assessment scheme.

[10] In the questionnaire in this chapter, the author didn't give any guidance as to what group or groups of pests they should have in mind when answering the questionnaire, as at that time none of the existing PRA schemes suggested taxon-difference. It only became clear after the studies done in Chapter 4, which suggested that some risk factors were very taxon-specific.

pathway. In the table, the ranks for *the number* and *ease of pathway* were modified to 1.5 to take into account of the equal rankings. This modification maintained consistency in the sum of the rankings. Table 3.5 shows the frequency of the risk factor rankings from the survey response.

Table 3.4 Expert surveys: rankings of the risk factors

Risk factor/Ranking/Response			1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th
ENTRY POTENTIAL			2.5	1	1	1.5	2	1	2.5	3	2	1	1.5	1	1	3	3	2	2
Pathway			1	1	1	2	1.5	1	3	1	1	1.5	1.5	4	4	1	2.5	1.5	1.5
	number						1.5			1	3	2	2	2					1.5
	ease						1.5			3	1	2	2	1				1	1.5
	movement						3.5			2	2	2	2	3				2.5	3.5
	duration						3.5			4	4	4	4	4				2.5	3.5
Survival during transport			3	2	2	2	3.5	3	1	3	4.5	1.5	3.5	1	1	2	2.5	4	3.5
	withstand starvation						2.5			6	3.5	7	6.5	7				4	2.5
	proximity						2.5			2	6.5	6	6.5	6				4	2.5
		500km		1		1	1	1	1	1			1		1	1	1		
		1000km																	
		2000km																	
		depends on pest epidemic			1						1	1		1				1	1
	speed						2.5			4	6.5	1.5	4.5	4				4	2.5
	condition						6			1	3.5	1.5	4.5	5				4	6
	life stage						2.5			3	1.5	3	2.5	3				4	2.5
	life cycle duration						6			5	1.5	5	2.5	2				4	6
	measures						6			7	5	4	1	1				4	6
Undetected possibility				4	3	4	1.5	2	2	5	4.5	3	3.5	2.5	2	4	4	4	1.5
Interception number			2	3	4	2	3.5	4		4	2	4	1.5	2.5	3	3	1	4	3.5
	1-2 times				1	1	1				1	1							1
	3-5 times			1				1		1			1		1	1	1	1	
	6-10 times								1										
	more than 10 times													1					
Intended use of commodity			4	5	5	5	5	5		2	3	5	5	5	5	5	5	1.5	5

ESTABLISHMENT POTENTIAL	1	2.5	1	3	2	2	2.5	1	1	2.5	1.5	2	3	1	1.5	2	2
Biotic habitat	2	4	2	1.5	2	2		2	1.5	2	5	5	1	1	2	3	2
host availability					2.5			1.5	1	1	1	1				1	2.5
host species					2.5			4	5	2.5	3	4				4.5	2.5
host amount					6			6	2.5	2.5	3	3				4.5	6
host distribution					6			3	2.5	4	3	2				4.5	6
association ease					2.5			5	4	5	5	5				4.5	2.5
competition					6			7	7	6	6.5	7				4.5	6
natural enemies		1	6	6	2.5	5		1.5	6	7	6.5	6			2.5	4.5	2.5
Abiotic habitat	1	5			5	3		3	1.5	2	1	1	5	2	1	3	5
Intrinsic characteristics	5	2.5	3	3.5	2	1		4	4	2	4	2	2	3	3.5	3	2
reproduction strategy																	
parthenogenesis	1	3	1	2.5	4	4		4	1		2.5	1	3	1		2.5	4
life cycle time	3	3	2	4	2.5	4	1.5	3	4		2.5	4	3		3	2.5	2.5
generations/year	3	1		1	1	4		1	2.5		4	2	3		1	2.5	1
reproductive strategy	3	3		2.5	2.5	4	1.5	2	2.5		1	3	1	2	2	2.5	2.5
adaptation																	
polymorphic			1		2			4	4		4	3.5	1	2	3	2.5	2
polyphagous/plurivorous	1	1		1	2	1		1	1	1	1	1.5	3	1	1	2.5	2
pesticide resistance	2.5	2.5			4		1	3	3	3	2.5	1.5	3	3	3	2.5	4
survival adaptation	2.5	2.5	2	2	2	2		2	2	2	2.5	3.5	3	4	3	2.5	2
Survival under current measures	3	2.5	1	1.5	2	4	1.5	1	5	5	3	4	4	4	5	3	2
Frequency of introduction	4	1	4	3.5	4	5	1.5	5	3	4	2	3	3	5	3.5	3	4
1-2 times			1	1	1				1	1						1	1
3-5 times		1				1		1			1	1	1	1	1		
6-10 times							1										
more than 10 times	1																
SPREAD POTENTIAL	2.5	2.5	3	1.5	2	3	1	2	3	2.5	3	3	2	2	1.5	2	2
By natural means	1	1	2	2	3.5	1	2	2	1	2.5	2	3	1	1	3	2.5	3.5
By human activities	2	3	2	2	1.5	2	1	1	2	2.5	1	2	2	4	4	2.5	1.5
Biotic suitability	3	2	2	2	1.5	3		3	3.5	2.5	4	4	3	2	2	2.5	1.5
Abiotic suitability	4	4	4	4	3.5	4		4	3.5	2.5	3	1	4	3	1	2.5	3.5

DIRECT PEST EFFECT	1	1	1	1	1.5	1.5	1.5	1	1	1	1	1	1.5	1.5	1.5	1.5	1.5
Host value	2	2		3	3.5	1	2.5	5.5	2	5	2	2	5	3	3.5	4	3.5
Host damage	3	5		3	3.5	2	2.5	3	1	1.5	2	3	3	2	1	4	3.5
Crop losses	4	2	4.5	3	3.5	3	2.5	1	3	1.5	2	1	1	1	2	4	3.5
Population to reach damaging densities					7			7	5	4	5	4				4	7
Spread rate (incl.host number & distribution)	1	5	1.5	3	3.5	4	6	4	6	6	5	5	4	5	5	4	3.5
Effect on production	5	7	3	6.5	3.5	6	6	5.5	7	3	5	6	6	6	6.5	4	3.5
Cost of control measures	6	2	1.5	3	3.5	5	2.5	2	4				2	4	3.5	4	3.5
Direct environmental and ecosystem impact	7	5	4.5	6.5		7	6	8					7	7	6.5		
INDIRECT PEST EFFECT	2	2	2	2	1.5	1.5	1.5	2	2	2	2	2	1.5	1.5	1.5	1.5	1.5
PHYTOSANITARY MEASURES																	
Certification	3	1.5	2.5	2.5	1.5	1.5	2	1	2.5	5	5	4	3.5	3	4	5	1.5
Pest-free area	1	5.5		2.5	1.5	1.5		2.5	1	1	1	2.5	3.5	1	3	2.5	1.5
Pre-clearance	5	5.5	1	5.5	3	4		2.5	4	4	3	4	3.5	2	6	4	3
Treatment	2	1.5	4	2.5	4.5	6	2	6	5	2	3	2.5	3.5	5	2	1	4.5
Post entry quarantine	4	3.5	5	2.5	4.5	4		4	6	3	3	1	3.5	4	1	6	4.5
Road/border inspection		3.5	2.5	5.5	6	4	2	5	2.5	6	6	5	3.5		5	7	6
Systems approach																2.5	

Table 3.5 Expert surveys: frequencies of rankings of risk factors

Risk factor/Ranking/Frequency			1/1.5	2/2.5	3/3.5	4/4.5	5/5.5	6/6.5	7/7.5	8
ENTRY POTENTIAL				8	6	3				
Pathway				12	2	1	2			
	number			3	3	1				
	ease			5	2	1				
	movement				5	3				
	duration				1	2	5			
Survival during transport				4	5	6	2			
	withstand starvation				2	1	1		2	2
	proximity				3		1		4	
		500km		10						
		1000km								
		2000km								
		depends on pest epidemic		7						
	speed			1	2		4		1	
	condition			2		1	2	1	2	
	life stage			1	3	3	1			
	life cycle duration			1	2		1	2	2	
	measures			2			2	1	2	1
Undetected possibility				2	4	3	6	1		
Interception number										
	1-2 times			6						
	3-5 times			8						
	6-10 times			1						
	more than 10 times			2						
Intended use of commodity				1	1	1	1	12	1	

ESTABLISHMENT POTENTIAL	7	8	2				
Biotic habitat	4	8	1	1	2		
host availability	6	2					
host species		3	1	3	1		
host amount		2	2	1		3	
host distribution		2	2	2		2	
association ease		2		2	4		
competition				1		4	3
natural enemies	2	3		1	1	5	1
Abiotic habitat	5	2	3		4		
Intrinsic characteristics	1	6	5	3	1		
reproduction strategy							
parthenogenesis	5	3	2	4			
life cycle time	1	5	5	4			
generations/year	6	3	2	2			
reproductive strategy	3	8	3	1			
adaptation							
polymorphic	2	4	2	3			
polyphagous/plurivorous	11	3	1				
pesticide resistance	2	4	6	2			
survival adaptation		12	3	1			
Survival under current measures	4	3	3	4	3		
Frequency of introduction	2	1	6	5	3		
1-2 times	7						
3-5 times	9						
6-10 times	1						
more than 10 times	1						
SPREAD POTENTIAL	3	8	5				
By natural means	6	7	4				
By human activities	5	9	1	2			
Biotic suitability	2	7	5	2			
Abiotic suitability	2	2	5	7			

DIRECT PEST EFFECT	9	8						
Host value	1	6	5	1	3			
Host damage	3	4	7	1	1			
Crop losses	5	4	5	3				
Population to reach damaging densities				3	2		3	
Spread rate (incl.host number & distribution)	2		3	4	5	3		
Effect on production			4	1	3	7	2	
Cost of control measures	1	4	4	3	1	1		
Direct environmental and ecosystem impact				1	1	3	4	1
INDIRECT PEST EFFECT	8	9						
PHYTOSANITARY MEASURES								
Certification	5	4	3	2	3			
Pest-free area	8	4	2		1			
Pre-clearance	1	2	4	5	3	1		
Treatment	2	6	2	3	2	2		
Post entry quarantine	2	1	4	6	1	2		
Road/border inspection		3	2	1	4	4	1	
Systems approach		1						

3.3.4.3 Data analysis of Delphi survey: How to derive risk factor weightings

After the collation of experts' responses, two steps of data analysis were conducted in order to derive pest risk factor weights. The first step was to convert each expert's ranking (expert's assessment of the importance) of each risk factor into a normalized weight (Fig. 3.6). It was explored using the following formulae to calculate the weightings of each risk factor:

$$Wr_i = (n+1-rk_i) \qquad (1)$$

$$Wr_i = (n+1-rk_i)^2 \text{ or} \qquad (2)$$

$$Wr_i = {}^{1}\!/_{rk_i} \text{ or} \qquad (3)$$

$$Wr_i = \exp^{\left(a\left(n+1-rk_i\right)\right)} \qquad (4)$$

where Wr_i is the weighted ranking of risk factor i; rk_i is the original ranking, n is the number of risk factors being considered. a is a parameter, the value of which determines the rate at which the importance of risk factors decrease with the risk score. In this study, $a = 0.5$ was set as an example.

The reason underlying the use of these formulae is that as the rankings increase, the importance of a risk factor or a risk element decreases compared with its counterparts. While the weighted rankings in formula (1) decrease linearly with the actual ranking, the weighted rankings in formula (2) decreases quadratically with the actual rankings, the weighted rankings in (3) is the inverse of the actual rankings, and the weighted rankings in (4) decrease exponentially with the actual rankings.

Based on the above rules for generating weighted ranking, Table 3.6 shows the weighted rankings for 8 risk factors of a representative risk factor with the original rankings between1 – 8.

Table 3.6 Weighted rankings for original rankings 1-8

	Rankings (r_k)															
Weighted ranking	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	Sum
1/rk	1	0.67	0.5	0.4	0.33	0.29	0.25	0.22	0.2	0.18	0.17	0.15	0.14	0.13	0.13	4.76
(9-rk)	8	7.5	7	6.5	6	5.5	5	4.5	4	3.5	3	2.5	2	1.5	1	67.5
$(9\text{-}rk)^2$	64	56.3	49	42.3	36	30.3	25	20.3	16	12.3	9	6.3	4	2.3	1	374
exp(0.5*(9-rk))	54.6	42.5	33.1	25.8	20.1	15.6	12.2	9.5	7.4	5.8	4.5	3.5	2.7	2.1	1.7	241

Subsequently, the weighted rankings were normalised using the following formula:

$$Nr_i = Wr_i / \sum_{i=1}^{i=8} Wr_i$$

where Nr_i is the normalised weighted ranking i, such that the sum of normalised weighted rankings equals 1. Table 3.7 shows the normalized weighted rankings for the four weighting methods.

Table 3.7 Normalised rankings

Normalised ranking Nr	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	Sum
1/rk	0.21	0.14	0.11	0.08	0.07	0.06	0.05	0.047	0.042	0.038	0.035	0.032	0.03	0.028	0.026	1
(9-rk)	0.12	0.11	0.1	0.1	0.09	0.08	0.07	0.067	0.059	0.052	0.044	0.037	0.03	0.022	0.015	1
$(9\text{-}rk)^2$	0.17	0.15	0.13	0.11	0.1	0.08	0.07	0.054	0.043	0.033	0.024	0.017	0.011	0.006	0.003	1
exp(0.5*(9-rk))	0.23	0.18	0.14	0.11	0.08	0.06	0.05	0.039	0.031	0.024	0.019	0.014	0.011	0.009	0.007	1

Different weighting transformations place different relative importance on high-ranked and low-ranked risk factors.

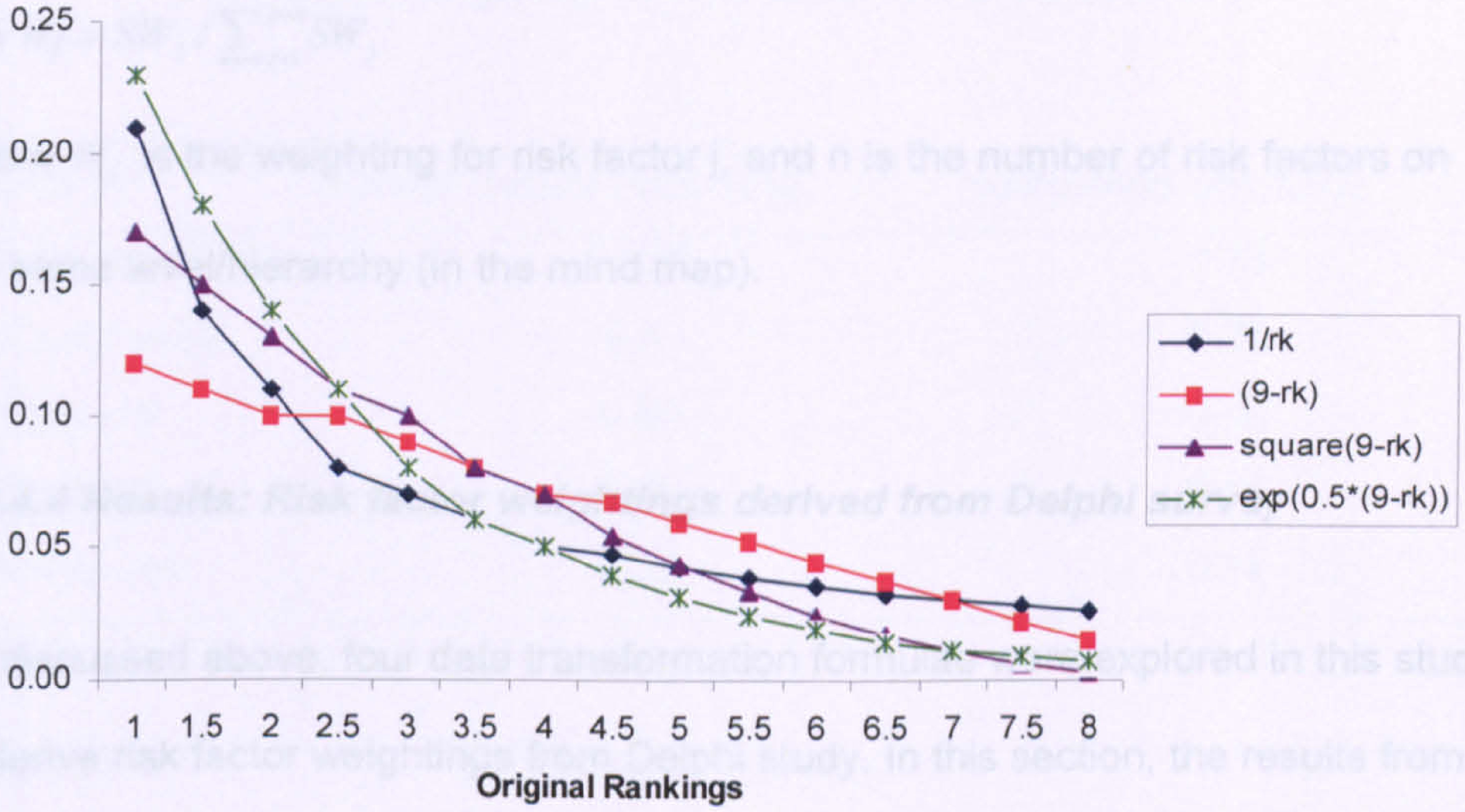


Figure 3.6 Normalised weighted rankings[11]

The second step is to aggregate the normalized rankings by all the experts for a particular risk factor to derive an overall weighting for the risk factor. The basic rule for assigning weighting applied: "All risk factors were considered or compared only with those factors in the same level and independent of others" (see 3.4.3.3.2).

Weightings for each risk factor were derived from the following equations. The sum of the factor weighting is calculated first:

(1) $SW_j = \sum_{i=1}^{i=8} Wr_{ij} * fq_{ij}$

Where SW_j is the sum of the weighting for risk factor j, it is derived by summing the product of the weighted ranking i and its frequency fq_{ij}; the factor weighting is calculated subsequently:

(2) $Wj = SW_j / \sum_{j=1}^{j=n} SW_j$

where W_j is the weighting for risk factor j, and n is the number of risk factors on the same level/hierarchy (in the mind map).

3.3.4.4 Results: Risk factor weightings derived from Delphi survey

As discussed above, four data transformation formulae were explored in this study to derive risk factor weightings from Delphi study. In this section, the results from the reciprocal transformation $Wr_i = \frac{1}{rk_i}$ are presented.

Table 3.8 shows the weightings for the three factors under introduction probability (first level in the mind map) derived from this approach.

Table 3.8 Weightings for risk factors for introduction probability

Risk factor	Sum of occurrence* weighted ranking	Normalised weighting
Entry	11.13	0.37
Establishment	10.7	0.36
Spread	8.2	0.27

---

[11] Normalised ranking (9-rk) is not a completely straight line because the rankings were normalised and rounded.

Similarly, weightings for risk factors at the next level of the hierarchy were calculated and shown in Table 3.9 – 3.11.

Table 3.9 Weightings for risk factors for entry potential

Risk factor	Sum of frequency/ranking	Normalised weighting
Pathway	12.07	0.33
Survival transit	8.4	0.23
Undetected	5.86	0.16
Interception	6.39	0.17
Intended use	4.32	0.12

Table 3.10 Weightings for risk factors for establishment potential

Risk factor	Sum of frequency/ranking	Normalised weighting
Biotic factor	8.32	0.23
Abiotic factor	7.47	0.21
Intrinsic characteristics	6.42	0.18
Survival measures	7.33	0.21
Introduction frequency	5.92	0.17

Table 3.11 Weightings for risk factors for spread potential

Risk factor	Sum of frequency/ranking	Normalised weighting
Natural means	10.54	0.32
Human activities	9.47	0.29
Biotic suitability	6.75	0.21
Abiotic suitability	6.12	0.19

Weightings for individual risk factors for introduction/spread potential derived from reciprocal transformation are shown in Table 3. 12 and Figure 3.7. Weightings derived from other transformations can be found in Appendix 5.

Table 3.12 Factor weightings derived from Delphi Survey for introduction/spread

Risk factor				Weighting
Entry				0.37
	Pathway			0.33
			number	0.27
			ease	0.37
			movement	0.22
			duration	0.14
	Survival during transport			0.23
			withstand starvation	0.1
			proximity	0.12
			speed	0.14
			condition	0.16
			life stage	0.17
			life cycle duration	0.14
			measures	0.17
	Undetected possibility			0.16
	Interception number			0.17
	Intended use of commodity			0.12
Establishment				0.36
	Biotic habitat			0.23
			host availability	0.31
			host species	0.12
			host amount	0.1
			host distribution	0.11
			association ease	0.1
			competition	0.06
			natural enemies	0.2
	Abiotic habitat			0.21
	Intrinsic characteristics			0.18
		reproduction strategy		
			parthenogenesis	0.27
			life cycle time	0.19
			generations/year	0.29
			reproductive strategy	0.25
		adaptation		
			polymorphic	0.17
			polyphagous/plurivorous	0.41
			pesticide resistance	0.19
			survival adaptation	0.22
	Survival under current measures			0.21
	Frequency of introduction			0.17
Spread				0.27
	By natural means			0.32
	By human activities			0.29
	Biotic suitability			0.21
	Abiotic suitability			0.19

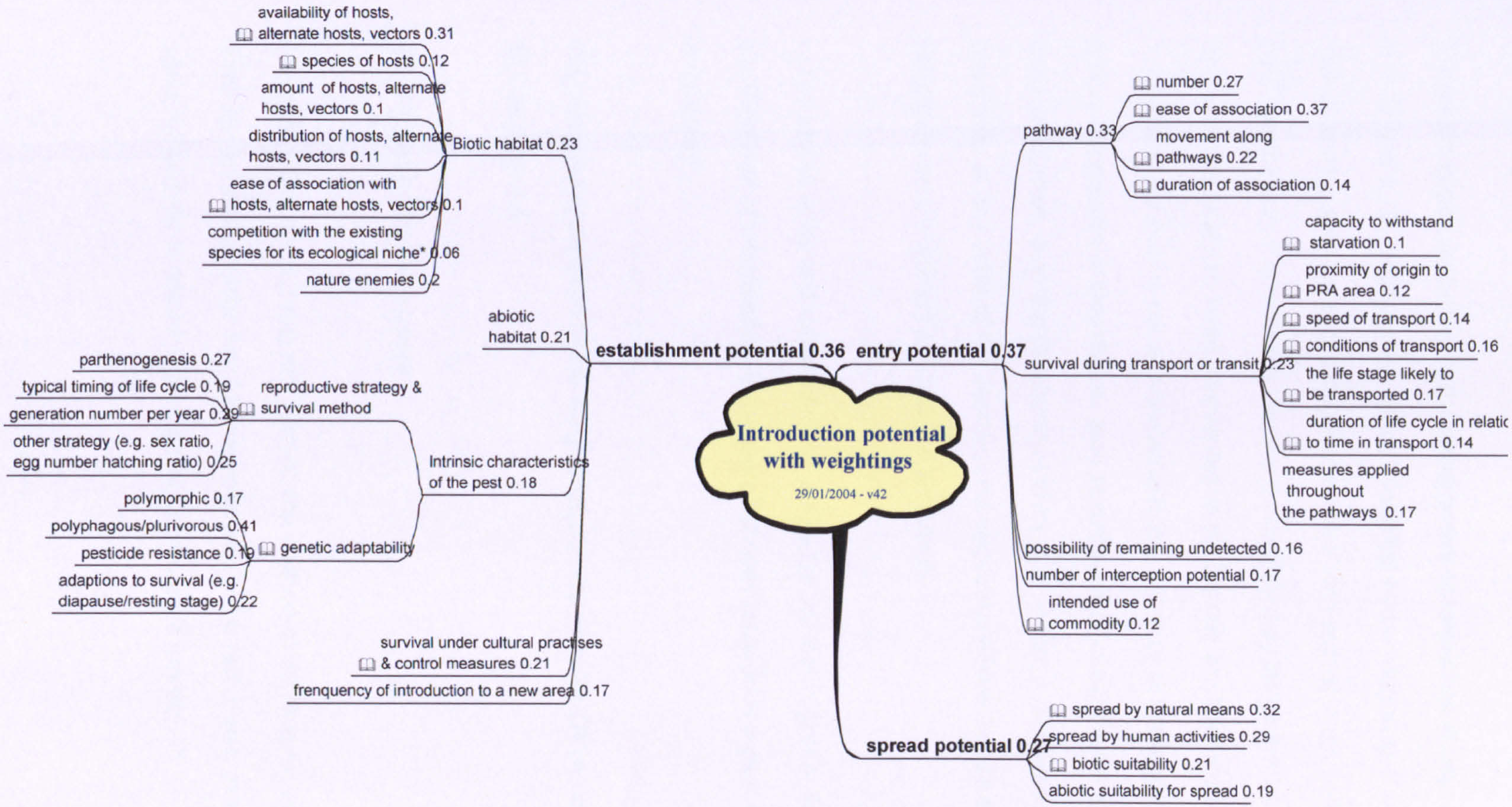


Figure 3.7 Weightings for introduction/spread potential (from expert survey)

The result shows that *entry* and *establishment* potential had similar weightings of 0.37 and 0.36, respectively, *spread* potential had a weighting of 0.27. This makes sense, as *entry* and *establishment* are both critical in the context of a pest introduction, it is possible that the probability may range from 0-1. For example, if a pest has been intercepted at the border, the likelihood of entry is 1. Whereas if a pest is not associated with the pathway at the origin, e.g. the pest is not present in that region, also assuming that all possible means of entry are known, then the likelihood of entry is 0. In fact, a risk assessment could stop at any point of the assessment sequence where risk becomes 0 (negligible as it is termed in some RA schemes).

In contrast to *entry* and *establishment*, *spread* is not that critical in deciding the likelihood of introduction. Spread is often inevitable once a pest becomes established.

The potential application of the weightings derived from the Delphi study is discussed in 3.4.

3.3.5 Combining risk scores

After scoring and weighting risk factors, the method of deriving an overall risk score based on the weighting and score for each risk factor was considered. Various methods of combining weighted scores were developed.

3.3.5.1 Development of appropriate combination metrics

In this study, three methods were proposed and compared: simple averaging, weighted averaging[12], and biased weighted averaging.

Simple averaging

The simplest way to combine risk factor scores to achieve an overall score was by simple averaging, the equation was

$$\text{average risk } r_a = \left[\sum_{i=1}^{i=n} a_i\right] / n,$$

where a_i was the score of risk factor i, n was the total number of risk factors.

Weighted averaging

If weighting were to be incorporated, a way to combine risk scores was by weighted averaging. The formula to calculate the weighted average risk score was as follows:

$$\text{weighted averaged risk } r_w = \sum_{i=1}^{i=n} a_i w_i$$

where w_i was the weighting of risk factor i, which is derived as explained in section 3.2.4, and $\sum_{i=1}^{i=n} w_i = 1$.

---

[12] As suggested in the EPPO PRA scheme (EPPO, 1997).

High and low biased weighted averaging

From the author's experience combined with discussion with other risk analysts, extreme scores (high or low) often had an overriding effect on the perceived level of risk.

The following example demonstrates the overriding power of an extreme score. If the prevalence of an organism at the origin of a commodity is extremely low, which may due to the existing pest control and quarantine measures, the likelihood of the organism being associated with the pathway may be extremely low (but not negligible). However, if the organism can very easily survive in the transit and transport and can also easily go undetected then these factors would warrant a high risk score. If simple averaging were used, the likelihood of entry may be medium or even high, as the extremely low score was balanced by two high scores. However, based on experience and intuition the likelihood of entry could be very low under such situation.

Such a situation could not be reflected using simple averaging, which does not acknowledge the over-riding effect of extreme scores in the result. In general, linear transformations do not give an emphasis to extreme scores.

In order to stress the impact of overriding low or high scores as discussed above, two equations for biased weighted averaging were developed that took into account the impacts of both weighting and extreme risk scores. By incorporating two complementary transformations, one of which gave more importance to high scores and the other to low scores.

Equations for biased weighted averaging were:

(1) high-score bias weighted risk $r_h = 5 + \ln\left\lfloor \sum_{i=1}^{i=n} \exp(a_i - 5) w_i \right\rfloor$, and a

(2) low-score bias weighted risk $r_l = 5 - \ln\left\lfloor \sum_{i=1}^{i=n} \exp(5 - a_i) w_i \right\rfloor$.

Thus, the biases were achieved by summing the scores as exponents; the scores 1,2...9 transformed to e^{-4}, e^{-3}...e^{4} and e^{4},e^{3}...e^{-4}, for the high- and low-score biases, respectively (Zhu *et al.*, 2000).

3.3.5.2 Comparison and discussion of the three metrics

Although averaging is the simplest way to combine risk scores, in practice, however, averaging may not accurately reflect the decision-making processes, because either a very high or a very low risk may have an over-riding impact on the actual decision. Here an extreme hypothetical example is taken, in which there is little chance for a pest to enter into a new area, but the pest has a high potential to establish, to spread and cause severe economic damage. In practice, the overall risk cannot be deemed high as suggested by the averaged risk (Table 3.13), it is likely that the risk is regarded as low. In fact, in this simplified example, the entry potential acts as a constraint to the risk.

Table 3.13 A hypothetical example of the impact of simple average

Risk factor	Risk score
Entry potential	1
Establishment potential	9
Spread potential	9
Economic impact	9
Averaged risk score	7

Another drawback of simple average was that it was very difficult to determine separation points to distinguish low, medium and high risk, if required, or to decide the quarantine status, especially for an inexperienced PRA practitioner.

The three combination methods were compared performing with 14 official PRA data sets (scores for individual risk factors and quarantine status) (McNamara, 2002, pers. comm.), which were based on EPPO PRA scheme (EPPO 1997), where a 1 – 9 scoring system was adopted.

Nevertheless, the quarantine status was not decided solely upon the risk scores: "*All of these PRAs were done during several meetings, and, at the end of the PRAs, the* [PRA]*Panel had to decide whether each of the pests should be a quarantine pest. However, the decision on quarantine status was not based on the scores in the PRA; it was a consensus decision of the* [PRA]

Panel based on their own expert opinions." (McNamara 2002, pers. comm.). (see Appendix 6 for datasheet).

The weightings used to obtain weighted averaged risks and biased weighted risks were taken directly from the EPPO PRA scheme (EPPO 1997), which suggested that questions with asterisks were considered more important than others. A weighting of 3 was assigned to each of these questions (8 altogether), weightings for other questions were calculated from the equation in 3.4.3. 2. The weightings were subjectively assigned by a single assessor.

Table 3.14 and Figure 3.8 show the overall risk scores for 14 forestry pests using different combination methods, as well as their quarantine status.

Table 3.14 Risk scores for 14 forestry pests and their quarantine status (PRA data provided by Dr. McNamara, EPPO, 2002)

Species	Average risk score	Weighted average risk score	High biased weighted risk score	Low biased weighted risk score	Quarantine status
Dendrolimus sibiricus	5.88	5.61	6.08	5.36	Q
Aeolesthes sarta	5.36	4.84	5.27	4.36	Q
Xylotrechus altaicus	5.4	5.22	5.68	5.05	Q
Scolytus morawitzi	5.98	5.79	6.17	5.16	Q
Ips subelongatus	5.2	4.9	5.37	4.37	Q
Dendroctonus ponderosae	5.62	5.38	6.28	5.4	Q
Tetropium gracilicorne	5.25	5.47	5.86	4.91	Q
Ips hauseri	5.3	5.18	5.6	4.63	Q
Dasychira albodentata	5.3	5.12	5.4	4.78	NQ
Erannis jacobsoni	5.09	4.67	5.22	4.09	NQ
Malacosoma parallela	5.21	4.66	5.27	4.28	NQ
Melanophila guttulata	5.15	4.8	5.47	4.51	NQ
Sphinx morio	4.63	4.44	5.19	4.84	NQ

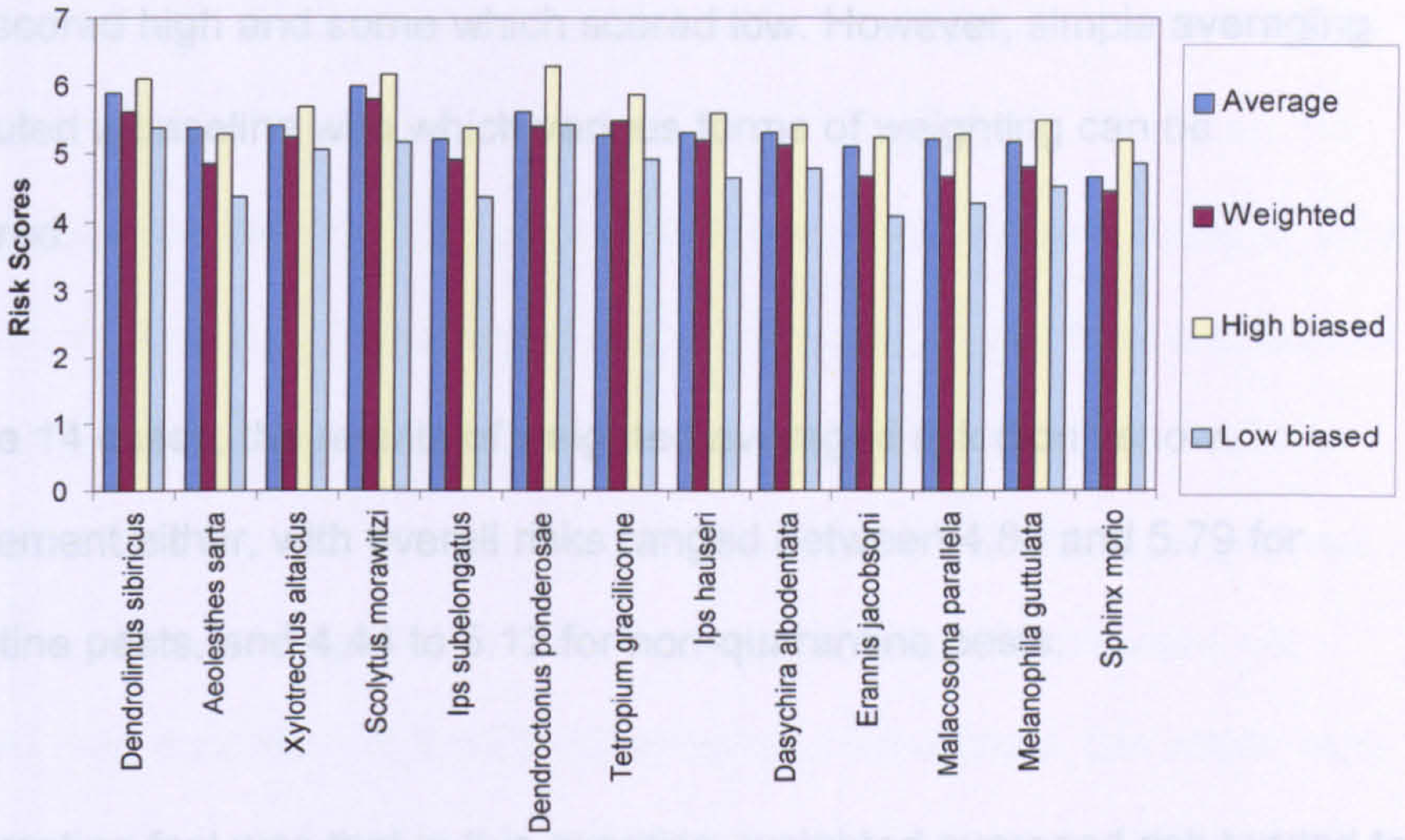


Figure 3.8 Risk scores for 14 forestry pests

The results of simple averaged risk revealed that: the simple averaged risk scores for quarantine pests (Q) ranged from 5.2 (Ips subelongatus) to 5.98 (Scolytus morawitzi); the simple averaged risk scores for non-quarantine pests (NQ) ranged from 4.63 (Sphinx morio) to 5.3 (Dasychira albodentata). There was thus some overlap between the two groups.

Conveniently assuming equal intervals, with a 1 – 9 scoring system, 1 – 3.33 usually indicates a low risk, 3.34 – 6.33 a medium risk and 6.34 – 9 a high risk. In fact, all the combined risk score in the above cases fall into the interval of medium risk i.e. 3.34 – 6.33. This was possibly because there were many risk factors (40 - 44) being considered and it is inherent that the overall risk score tended to be in the middle given that there were often some factors which scored high and some which scored low. However, simple averaging constituted a baseline with which various forms of weighting can be compared.

In these 14 cases, the results of weighted averaged risk didn't show improvement either, with overall risks ranged between 4.84 and 5.79 for quarantine pests, and 4.44 to 5.12 for non-quarantine pests.

An interesting fact was that in this exercise, weighted averaged risk tended to be lower than averaged risk. In an attempt to explain this somewhat unexpected result, the author looked at the way that weightings were given as well as the dataset, it was found that:
(a) no higher weighting is given to factors effecting *entry potential*, and
(b) some of the factors with higher weighting (1.26 *reproduction strategy*, 1.30 *introduction to a new area* and 1.24 *environment aiding establishment*) were scored relatively low or no data were available.

This unexpected result was not a feature of how the weightings were assigned, rather, it was because that those high weighting risk factors were directly taken from that suggested in the EPPO scheme without any adjustment, so do not necessarily identify all important risk factors.

The example of *Aeolesthes sarta* was used to explain how the result could be interpreted. For *Aeolesthes sarta*, simple averaged risk was 5.36, weighted averaged risk was 4.84, high and low biased weighted risks were 5.27 and 4.32 respectively. Given that *Aeolesthes sarta* was regarded as quarantine pest by the EPPO PRA panel, the result did not seem ideal. However, the scoring data revealed that the factors related to *survival during transportation* and *control measures, frequency of commodity arrivals and their distribution, host plants, natural enemies, control measures* and *eradication possibility* were all scored high, i.e. 7 or above, and this may be the reason why it was deemed to present a high risk. However, most of these factors were not weighted high according to the EPPO scheme. Furthermore, the seven high-weighting factors, as suggested in the EPPO scheme, were scored 3, 3, 3, 4, 6, 3, 3 and 6, respectively.

This explained why the weighted average and high and low biased weighted average were lower than the simple average. Again, this also demonstrated that incorporating weighting could influence the overall risk, be it enlarging the constraining effect of a low risk score or the reverse. It was a matter of how accurate the weighting reflecting the perceived importance. The impact of the weightings was also reflected in the results of the biased weighted risks.

The three metrics were also compared (Zhu *et al.*, 2000) for four pests in Table 3.15.

Table 3.15 Comparison of three metrics for four Pests

	Fall webworm (*Hyphantria cunea*)	Codling moth (*Cydia pomonella)*	Fijian fruit fly (*Bactrocera passiflorae*)	Natal fruit fly (*Ceratitis rosa*)
Averaged Risk				
r_a	6.03	5.28	5.95	6.79
Weighted Risk				
r_w	6.34	5.25	5.68	6.75
High Biased Weighted Risk				
r_h	7.54	6.77	7.51	7.68
Low Biased Weighted Risk				
r_l	4.25	3.34	3.44	4.65

To illustrate the differences between the metrics, suppose that an assessment of whether the risk was high, medium or low was required. The 1-9 scale was divided into three equal parts: 1-3.66, 3.67 -6.33 and 6.34-9 corresponding to low medium and high risks, respectively. A change from medium to high risk occurred for the fall webworm, if a weighted average rather than a simple average was used. The mind map for this species (see Appendix 7) revealed that certain higher-scoring risk factors were thought to be more important and therefore given high weights.

The high- and low-biased averages were interpreted in the following way. A high value (≥6.34) of the high-biased score indicated that one or more of the risk factors from which it was derived had a very high score. Such a situation might be regarded as high-risk even if other risk factors have moderate scores. However, a low value (≤3.66) of the low-biased score indicated that a constraint existed for one or more of the component risk factors, thereby possibly negating the impact of high scores for other factors.

Comparing the outcomes for the weighted averages and the low/high-biased averages, there was no difference for both fall webworm and Natal fruit fly: high risk for both metrics. With Fijian fruit fly and the codling moth the outcomes differed: in both cases being medium for the weighted average and low for the low/high-biased average. The mind map for the Fijian fruit fly revealed the existence of many importation *pathways* and a high chance of *pest survival during transport*. However, the *climate* and *host range* within the PRA area were largely unsuitable – these constraints meant that the overall risk was correctly regarded as low. For the codling moth, marginal *climate suitability* in the PRA area and a low volume of fruit *trade* between the source area and the PRA area also led to a low-biased score in the low-risk range (i.e.≤3.66). Again, therefore, given the existence of factors that constrains the risk, the prognosis was correctly judged to be low-risk even though some high-risk factors were also present.

A case study of *introduction/spread potential* for fall webworm (*Hyphantria cunea* Drury) can be found in Appendix 7. This case study demonstrated the whole risk assessment process proposed in this chapter. It used mind mapping to identify the risk factors. A score and weighting were assigned to each risk factor, and overall risk scores were calculated with the proposed combination methods and result was discussed.

3.4 CONCLUSIONS AND DISCUSSION: STRUCTURING THE COMPONENTS OF PEST RISK

An exploration of new methodologies to assist in PRA has been described and these methodologies subjected to some evaluation with case studies.

It is proposed to divide pest risk assessment stage into two steps: Pest Risk Identification (PRI) and Pest Risk Evaluation (PRE).

The application of mind mapping provided a means to identify all possible risk factors that may lead to a pest introduction and avoid important factors being overlooked, and potentially, to reduce ambiguity and increase transparency.

The sources of pest risk could be identified by disaggregating the main considerations for pest risk (e.g. introduction potential and consequences) into a series of more specific risk factors, such that the ambiguities are reduced and the quality and depth of PRA are improved. Risk factors for introduction potential and economic consequences were put into a structured format visualised by mind mapping.

In all the case studies, it proved possible to structure the pest risk analysis using mind mapping. Mind mapping allowed the pest risk be disaggregated into a series of more specific risk factors, and to which a score and weighting could be attributed.

Nevertheless, a limitation of mind mapping exists: mind mapping only allows a hierarchical approach. Applying this in PRA, the risk criteria/factors can be disaggregated as far as appropriate, usually into different number of levels (see figure 3.1, where the number of hierarchies for the risk factors differ). This might be sometimes misleading as it might be thought that if there were more factors under a heading, then this heading might be regarded as more important. However, this is not true, if the weighting results from the Delphi study were taken for example, *abiotic habitat* was not disaggregated further, but it has a weighting of 0.21; whereas at the same hierarchical level, *intrinsic characteristics* of the pest was further disaggregated into 8 risk factors, but it only has a weighting of 0.18.

Some approaches to the evaluation of risk (PRE) were proposed which facilitate the scoring of risk factors, and the subsequent weighting and combining of risk scores.

The purpose of scoring the risks and various scoring systems were discussed.

The methods for introducing weighting to risk assessment are new in PRA. Incorporating weighting into PRA was proposed and methodologies for obtaining weightings were developed.

Weightings can be subjectively assigned to risk factors and methods were proposed and discussed.

The Delphi technique can be applied to obtain a pool of experts' opinion. Methods to derive weightings from expert opinion were developed. The weightings obtained from this approach can help better understand the different importance of each risk factor. The result of weighting also provides a more objective opinion because they were derived from a wide range of risk analysts and quarantine officials, who brought in different experience, knowledge and background. If there were bias present in the expert opinion, the result might well reflect the collective bias. Furthermore the Delphi technique provides a means that different biases from individual experts might balance each other.

The weighting results from this study could provide a starting point for a risk analyst to commence his/her own analysis, should he/she wish to incorporate weighting into the assessment. The weightings might be adjusted to reflect his/her own perception of the importance of each risk factor, as long as the sum of the weightings remains equal to 1.

Different data transformation formulae were proposed in this chapter to derive weightings from the Delphi survey. As the aim of this study was to find out the perceived importance of each risk factor and to rank them, it did not matter which transformation formula to choose, as long as it could provide a good distinction between the rankings.

The author used a weighting based on the reciprocal formula, 1/rank to illustrate the weightings obtained from the expert survey throughout the

Chapter. The reasons for choosing this formula were that it gave a strong bias towards higher rankings; it provided a sharp distinction between rankings, and it is easy to understand and calculate.

The usual third step of a Delphi study was not performed in this study, in that the analyses of the expert response were not returned to the PRA experts for review. Hence, the application of the Delphi approach was limited, partly due to the constraint of time and difficulty of obtaining expert responses.

Overall, incorporating weighting was a novel approach, which accounted for the perception of different importance of the risk factors; therefore any overall risk value derived was more meaningful and accurate.

Apart from the methods discussed in this chapter, the possibility of obtaining objective weighting from historical PRA data with statistical techniques was investigated and will be discussed in detail in Chapter 4.

Various methods for combining individual risk scores into an overall risk value were developed and discussed; weighted average and high and low biased weighted average were proven to be better reflecting the perceived quarantine status in a number of case studies.

CHAPTER 4. EMPIRICAL APPROACHES FOR PEST RISK ASSESSMENTS

There are a number of problems associated with the current risk assessment process. One of these is that with the numerous risk factors considered, some risk factors are likely to be highly related to each other (multicollinearity[1]). Hence, it might not be necessary to include all the risk factors in a quantitative risk assessment in order to arrive at a reliable and meaningful overall assessment.

Another problem is that the risk factors are not all equally important in the process of risk assessment. The more important risk factors should contribute more to the final risk assessment than those less important. However, determining which factors are important is a difficult task and this problem becomes particularly acute in quantifying the importance of risk factors.

The third problem is that even if a synthesis of overall risk score were derived, it is still subject to human judgment to decide whether the risk is acceptable.

Using two different datasets of pest risk assessments, based on different risk assessment schemes, this chapter tries to use various empirical approaches to investigate the problems mentioned above. Section 4.1 describes the two sets of data used in the empirical analysis. Section 4.2 uses correlation

---

[1] Multicollinearity refers to a situation in which two or more explanatory variables in a regression model are highly correlated. It is perfect multicollinearity if the correlation between two independent variables is equal to 1 or -1.

analysis to show that there are significant correlations among some risk factors, whilst some risk factors are independent from the others. Risk factors are correlated with the overall risk to differing extents. Section 4.3 investigates the relative contributions of the risk factors. The idea of discriminating factors is considered, which might be used as indicators to identify high-risk pests. Principal Component Analysis (PCA) is applied to pest risk assessment data for 252 potential quarantine pests in Tanzania. Section 4.4 explores how empirically to reduce the number of risk factors to be considered in PRA and also uses PCA as a means to derive risk factor weightings. Section 4.5 investigates how to find patterns and influential risk factors in the data by using an automatic machine learning technique (Genetic Algorithms), which aims to derive generalised rules that help identify high risk situations and support the decision making process.

4.1 DATA

The data used in these studies includes two datasets of pest risk assessments, based on different risk assessment schemes: IPPC ISPMs No 2, and EPPO P/M 5(2).

The first dataset used in this study comprised 252 risk assessment cases for potential quarantine pests for Tanzania (Tanzanian data hereafter) prepared by Black and Abdallah (1997), including fungi, bacteria, viruses and virus-like diseases, insects, phytoplasma, nematodes, and mites. Of these potential quarantine pests, there are 78 fungal species, 18 bacteria, 60 viruses and virus-like organisms, 62 insects, 11 phytoplasma, 19 nematodes and 4 mites. The risk assessment was based on the IPPC ISPM No 2: Pest risk analysis for quarantine pest (IPPC 1995a).

Seven risk factors were considered in this assessment:

Climate suitability

Host range

Dispersal potential

Economic impact

Environmental impact

Pathway

Trading partners

For each risk factor, scores of 1, 2 or 3 were given according to the level of risk for that factor for that pest, where 1 represents low risk, 2 medium risk, and 3 high risk.

All scores were based on the pest distribution and Tanzania's economy at the time. Pests have since spread and Tanzania's agriculture has diversified and intensified. For example, Tanzania now exports Irish potato and the tubers are a pathway for *Meloidogyne chitwoodi* so this pest would now have an economic impact of 3, giving a score total of 16 (Holt *et al.*, 2006). *M. chitwoodi* was known to be present in South Africa since 1999 (EPPO, 1999).

In Black and Abdallah (1997), the overall risk rating for each species was calculated as the sum of the risk scores for the seven risk factors. An overall risk was originally obtained by summing the scores. Because the scoring system has only three scale points and seven components, there are only 15 possible values for the sum of the scores (7–21 inclusive). Black and Abdallah (1997) categorised the results into low, medium and high risk (corresponding to score bands 7–10, 11–14 and 15–21, respectively).

A summary of the results of the risk assessments, grouped by taxa are shown in Table 4.1 (after Black and Abdallah (1997))

Table 4.1Risk assessment results of quarantine pests for Tanzania

	Low Risk	Medium Risk	High Risk	Number of species
Bacteria	8	7	3	18
Fungi	34	39	5	78
Insects	14	25	23	62
Mites	3	1	0	4
Nematodes	9	7	3	19
Phytoplasma	9	2	0	11
Viruses or virus-like organisms	45	15	0	60
Total	122	96	34	252

The conclusions originally derived from these data were of considerable importance as they formed the basis for official plant quarantine harmonization activities in the East African community (Holt *et al.*, 2006). Detailed examination of the individual cases suggested good correspondence between this categorisation and the assessors' perceptions of high, medium or low risk (Holt *et al.*, 2006).

The second data set used in these analyses includes 15 risk assessment cases for insect species, based on the EPPO risk assessment scheme. These were prepared by McNamara, Orlinski, Baker (pers. comm. 2000-2001[2]) and the author. EPPO risk assessment scheme is based on the IPPC guidelines, but with more detailed considerations for each risk factor (EPPO data hereafter). Hence although the numbers of risk factors are different for the two datasets, they covered the same biological, economic, environmental and social criteria.

The 15 insect species comprised *Malacosoma parallela* (mountain ring silk moth), *Aeolesthes sarta* (city longhorn beetle), *Dendrolimus superans sibiricus* (Siberian silk moth), *Dasychira albodentata* (coniferous orgiid), *Scolytus morawitzi* (scolytid of Morawitz), *Anoplophora glabripennis* (Asian longhorned beetle), *Hyphantria cunea* (fall webworm), *Cydia pomonella* (codling moth), *Ceratitis rosa* (Natal fruitfly), *Bactrocera passiflorae* (Fijian fruit fly), *Xylotrechus altaicus* (Altay longhorned beetle), *Choristoneura fumiferana* (spruce budworm), *Erannis jacobsoni* (geometrid of Yacobson), *Pissodes strobi* (white pine weevil) *and Dendroctonus ponderosae* (black hills beetle).

Under the EPPO scheme employed in 2001, there were 44 risk factors to be considered in each pest risk assessment. Risk factors were scored on a 1-9

---

[2] Since then, EFSA Plant Health panel has produced EU guidelines for PRA. However, the risk score data used here were still as when they were discussed back in 2001.

scale[3] covering the likelihood of introduction (*entry* and *establishment*) and the consequences of introduction (including potential for spread). Missing scores for some risk factors (e.g. n/a) were handled by assigning a score of 5 to keep their neutral position.

Table4.2 shows the risk factors considered under the EPPO scheme and the results of some basic statistics of the factor scores for 15 insect species.

---

[3] To rank the perceived level of risk by assigning ordinal numbers (e.g. 1 to 3 or 1 to 9) to each risk factor is a common approach in pest risk assessment. (Zhu *et al.*, 2000, Black *et al.*, 1997, EPPO 1997, Holt *et al.*, 2006)

Table 4. 2. Descriptive statistics of the scores of 44 risk factors for 15 insect species

	Mean	Median	Standard Deviation	Sample Variance	Min.	Max.
Number of pathways	4.27	5	1.75	3.07	2	8
Association with pathway	5.67	6	1.45	2.1	2	8
Pest concentration	5.27	5	1.44	2.07	3	8
Surviving existing practises	5.93	6	1.75	3.07	3	9
Remaining undetected	5.4	5	1.35	1.83	4	8
Surviving in transit	7.4	8	0.91	0.83	6	9
Multiplying in transit	2.47	1	1.77	3.12	1	5
Movement along pathway	4.07	4	1.79	3.21	1	8
Commodity distribution	6.53	7	1.85	3.41	1	8
Duration of consignment arrival	6.4	7	1.99	3.97	2	9
Transfer to a suitable host	5.93	6	1.67	2.78	2	8
Intended use of commodity	5.87	6	2	3.98	2	9
Host species in PRA area	6.13	6	2.13	4.55	2	9
Host extension	6.6	7	1.99	3.97	1	9
Wild plants aiding dispersal	6.13	6	2.23	4.98	3	9
Climate similarity	5.87	6	2.1	4.41	2	9
Similarity of other abiotic factors	7.67	8	1.18	1.38	5	9
Competition with existing species	6.73	7	1.16	1.35	5	8
Natural enemies	6.93	7	1.44	2.07	5	9
Existing control measures	6.07	6	1.39	1.92	4	8
Reproductive strategy	4.8	5	2.08	4.31	2	8
Low populations being established	5.8	5	1.57	2.46	4	9
Ease of being eradicated	7.53	8	0.83	0.7	6	9
Genetically adaptability	5.8	6	1.78	3.17	3	9
Frequency of introduction	4.13	4	1.55	2.41	2	7
Economic loss in existing area	7.8	8	1.01	1.03	6	9
Environmental damage in existing area	5.93	7	2.05	4.21	3	8

Social damage in existing area	4.47	5	1.88	3.55	2	8
Extension of endangered area	5.27	6	1.79	3.21	3	8
Spread by natural means in PRA area	4.87	5	2.13	4.55	1	7
Spread by human assistance in PRA area	6.6	6	0.91	0.83	5	8
Ease of containing of spread	6.4	7	0.74	0.54	5	7
Direct effect on yield or quality in PRA area	6.13	7	1.46	2.12	3	8
Effect on producer profits	5.47	6	1.46	2.12	3	7
Effect on consumer demand	4	4	1.77	3.14	2	7
Effect on export markets	5.93	6	1.1	1.21	4	7
Other costs due to the introduction	4	5	1.13	1.29	2	5
Environmental damage in PRA area	4.07	4	1.28	1.64	2	7
Social damage in PRA area	3.07	3	1.1	1.21	1	5
Natural enemies	6.4	7	1.4	1.97	4	8
Ease of the pest being controlled	7.27	8	1.1	1.21	5	8
Impact on control of other pests	2.6	2	1.24	1.54	2	5
Other side-effects	4.6	5	1.06	1.11	1	5
Resistance to plant protection products	3.8	4	1.47	2.17	1	5

4.2 ANALYSIS OF THE RELATIONSHIPS OF RISK FACTORS: INTER-DEPENDENCE AND INTERACTIONS

While a sound pest risk analysis should cover in an objective manner all the biological, economic, environmental and social impacts, it should also take account of the interaction among these impacts and their combined effects.

In order to understand the relationships amongst risk factors, information redundancy and to explore the potential for simplification of the risk assessment, in this section the risk assessment process was examined and the following questions were asked:

- What are the relationships among the risk factors? This is to examine whether they are dependent on each other;
- If the risk factors were not independent of each other, how do they interact with each other?
- Are the correlations between risk factors consistent for different taxonomic groups of the pests? This is important in the context of the potential for general simplification of risk assessment schemes.

The data analysis performed in this study was correlation analysis of risk factors for the two datasets of pest risk assessments described in section 4.1. Correlations were calculated separately for the two sets of risk factors, as well as the subsets of these data where applicable.

4.2.1 Relationships between the risk factors based on the ISPM No 2

Correlations of the seven risk factors in Tanzanian data were calculated using SPSS (v.11) (SPSS Inc 2001).

Table 4.3 shows the result of the correlations between each risk factor and their *p* values, where significant correlations were flagged and highlighted at the 0.01 and 0.05 levels.

The overall risk[4] is partly correlated with each individual risk factor. It is inevitable as the overall risk scores are numerically based on the components, i.e. the summation. Nevertheless the correlations between the overall risk score and each individual risk factor scores indicate which relationships are stronger than others, and therefore how important each individual risk factor is for the overall risk score. However, because the overall score and components scores are not independent, the *P* values are not particularly meaningful and must be taken only as a relative indicator of importance. This reservation/limitation is held throughout this chapter where significance level is applicable.

See Appendix 8 for a detailed explanation of the correlations between risk factors.

---

[4] The overall risk score is the summation of the scores for the 7 risk factors.

Table 4.3 Correlation matrix for seven risk factors in Tanzanian risk assessment

Risk factor	Climate Suitability	Host Range	Dispersal Potential	Economic Impact	Environmental Impact	Pathway	Trading Partner
Climate Suitability	1						
Host Range	0.22(***)	1					
Dispersal Potential	0.27(***)	0.34(***)	1				
Economic Impact	0.27(***)	0.40(***)	0.42(***)	1			
Environmental Impact	-0.108	0.071	0.079	0.072	1		
Pathway	0.16(**)	.019(***)	0.20(***)	0.41(***)	-0.06	1	
Trading Partner	0.13	0.23(***)	0.21(***)	0.33(***)	-0.05	0.62(***)	1
Overall Risk	0.48(***)	0.68(***)	0.63(***)	0.74(***)	0.27(***)	0.59(***)	0.59(***)

Number of species = 252

*** Correlation is significant at the 1% level

** Correlation is significant at the 5% level

Interactions between correlated risk factors were subsequently examined, i.e. which factors might have direct causal effects on others and which might be affected by others[5]. Table 4.4 shows the result.

---

[5] This was based on the author's own judgment of what were likely to be the cause and effect.

Table 4.4 Interaction matrix of risk factors in Tanzanian data

Risk factor/ Interaction	Climate Suitability	Host Range	Dispersal Potential	Economic Impact	Environmental Impact	Pathway	Trading Partner
Climate Suitability							
Host Range	+						
Dispersal Potential	+	+					
Economic Impact	+	+	+				
Environmental Impact							
Pathway		+	-	-			
Trading Partner		+	-	-		-	
Overall Risk	+	+	+	+	+	+	+

Number of species = 252

+ Indicates column has a direct effect on row

- Indicates row has a direct effect on column

The interpretations of the results above were fairly self-evident, e.g.:

If there were many *trading partners*, there would be more *pathways*. A high *dispersal potential* may be linked to a high *economic impact*; if *host range* of a pest was wide, its *dispersal potential* may be high because of the improved colonisation ability. If there were more *trading partners*, there might be higher *economic impact*.

The correlations revealed in this study seemed biologically meaningful. The author has yet to come cross an organism with no economic significance when had high likelihood for other factors, e.g. *host range, dispersal potential*.

Usually if an organism had a higher *dispersal potential* and a wider *host range*, it is more likely for it to find a host and became established. In terms of economic significance, it is a general practice not only to consider the direct damage to the crop, but also consider the indirect impact, e.g. impact on exports and increased costs for control measures, as well as social impact. Environmental impact is also to be considered, e.g. whether it can affect the indigenous flora or fauna. Exceptions might be some latent viruses, which show no symptoms and cause no direct damage; however those viruses usually were very host-specific and rarely had high dispersal potential or wide host range.

Examining the correlations between overall risk and the risk factors, it was interesting to note that *economic impact* (0.74) had the biggest effect on the overall risk, followed by *host range* (0.68), *dispersal potential* (0.63), *pathway* (0.59), *trading partners* (0.59), *climate suitability* (0.48), *environmental impact* (0.27) had the least effect on the overall risk. This strongly reflected the definition of *quarantine pest*, which emphasized a pest's economic importance.

It is interesting to note that the risk factors are logically related to each other – there is overlap between them – one of the reasons why the scheme have so much overlap in their content might be that it tries to capture different aspect of the risk more fully.

4.2.2 Risk factor correlations for different taxonomic groups

Correlations were analysed[6] between risk factors for the different taxonomic groups of potential quarantine pests to see whether they were consistent or differed between groups (tables 4.5 – 4.8). Some interesting relationships are discussed for individual groups. Interactions between significantly correlated factors were examined and are shown in the correlation matrix tables.

---

[6] Details of correlation for individual taxonomic groups are given in the thesis text because the correlations are discussed in the context of their different biology.

4.2.2.1 Fungi

Table 4.5 Correlation and interaction of risk factors for fungi

Risk factor	Climate Suitability	Host Range	Dispersal Potential	Economic Impact	Environmental Impact	Pathway	Trading Partners
Climate Suitability	1						
Host Range	0.06	1					
Dispersal Potential	0.12	-0.02	1				
Economic Impact	0.06	**0.24(**)+**	**0.24(**)+**	1			
Environmental Impact	**-0.38(***)+**	-0.04	0.07	-0.04	1		
Pathway	0.09	0.12	0.14	**0.30(***)-**	**-0.30(***)-**	1	
Trading Partners	0.15	**0.39(***)**	0.13	**0.33(***)-**	**-0.32(***)-**	**0.45(***)-**	1
Overall Risk	**0.33(***)+**	**0.58(***)***	**0.50(***)+**	**0.66 (***)+**	**-0.01**	**0.55(***)+**	**0.64(***)+**

Number of species = 73

*** Correlation is significant at the 1%level

** Correlation is significant at the 5% level

\+ Indicates column has a direct effect on row

\- Indicates row has a direct effect on column

The correlations and the inferences about cause and effect suggest that:

- *Economic impact* was affected by *host range, dispersal potential, pathway*, and *trading partners*;
- *Environmental impact* were affected by *climate suitability, pathway*, and *trading partners*;

- *Dispersal potential* had a slight effect on *economic impact*, and it was relatively independent to the others;
- The overall risk was affected by *economic impact* (0.68), followed by *trading partners* (0.63), *host range* (0.58), *pathway* (0.52), *dispersal potential* (0.48), and *climate suitability* (0.31). *Environmental impact* had little effect on the overall risk.

However, as discussed above, *economic impact* was affected by *host range, dispersal potential, pathway*, and *trading partners*, therefore there was a degree of autocorrelation here, but it still gave a comparative measure of how an individual factor reflected the overall risk.

These correlations made sense by examining the biology of fungi. For example, *dispersal potential* was relatively independent; this may be because many fungal species can reproduce via vegetative spores or through mycelial fragmentation and this allows more rapid dispersal; both asexual and sexual spores are actively dispersed by forcible ejection from their reproductive structures; this may enable a fungus travel long distance without the help of a host or human activity.

4.2.2.2 Viruses and Virus-Like Diseases

Table 4.6 Correlation and interaction of risk factors for virus and virus-like diseases

Risk factor	Climate suitability	Host range	Dispersal potential	Economic impact	Environmental impact	Pathway	Trading partners
Climate suitability	1						
Host range	0.06	1					
Dispersal potential	0.29(**)+	-0.034	1				
Economic impact	-0.04	0.2	-0.04	1			
Environmental impact	0	-0.02	-0.01	0.46(***)-	1		
Pathway	0.03	-0.15	-0.12	0.18	-0.06	1	
Trading partners	-0.07	0.05	-0.05	0.27(**)-	0.18	0.61(***)-	1
Overall Risk	0.46(***)+	0.42(***)+	0.33(***)+	0.58(***)+	0.45(***)+	0.39(***)+	0.58(***)+

Number of species = 60

*** Correlation is significant at the 1%level

** Correlation is significant at the 5% level

\+ Indicates column has a direct effect on row

\- Indicates row has a direct effect on column

The results suggested that:

- *Economic impact* were significantly affected by *environmental impact* and *trading partners*;
- *Host range* was independent of any other risk factors;
- *Climate suitability* significantly affected *dispersal potential;*

- The overall risk was greatly affected by *economic impact* (0.58) and *trading partners* (0.58), followed by *climate suitability* (0.46), *environmental impact* (0.45), *host range* (0.42), *pathway* (0.39), and *dispersal potential* (0.33).

It was interesting to note that *host range* seemed independent to others; this might because most of the viruses have a very specific range of host comparing with other taxonomic groups.

It was worthy to note that *climate suitability* had a direct effect on *dispersal potential.* If we consider the epidemiology of plant virus, we'll see that plant viruses are often transmitted from plant to plant by vectors. These vectors are normally insects such as aphids, whiteflies etc. It then became apparent that *climate suitability* often had direct effect on the dispersal of viruses through its effect on the vectors.

It was also worthy to remark that environmental impact and economic impact were significantly correlated. This might due to the fact that some viruses need alternate hosts such as weeds to complete their life cycle, and weeds played a big part in assessing environmental impact.

4.2.2.3 Insects

Table 4.7 Correlation and interaction of risk factors for insects

Risk factor	Climate suitability	Host range	Dispersal potential	Economic impact	Environmental impact	Pathway	Trading partners
Climate suitability	1						
Host range	**0.35(***)+**	1					
Dispersal potential	**0.26(**)+**	**0.24+**	1				
Economic impact	**0.53(***)+**	**0.44(***)+**	**0.39(***)+**	1			
Environmental impact	0.11	0.08	-0.08	-0.02	1		
Pathway	0.13	**0.34(***)+**	0.22-	**0.49(***)-**	0.14	1	
Trading partners	0.09	0.15	0.14-	**0.31(***)-**	0.15	**0.70(***)-**	1
Overall risk	**0.59(**)+**	**0.69(**)+**	**0.45(**)+**	**0.77(**)+**	**0.34(**)+**	**0.70(**)+**	**0.57(**)+**

Number of species = 68

*** Correlation is significant at the 1% level

** Correlation is significant at the 5% level

+ Indicates column has a direct effect on row

- Indicates row has a direct effect on column

The results suggested:

- *Economic impact* was significantly affected by all other risk factors except *environmental impact*;
- *Environmental impact* was irrelevant to any other risk factors hence independent;

- The over risk was greatly effected by *economic impact* (0.77), followed by *pathway* (0.70), *host range* (0.69), *climate suitability* (0.59), *trading partners* (0.57), *dispersal potential* (0.45), and *environmental impact* (0.34).

It was noted that *climate suitability* and *host range, dispersal potential* were significantly correlated; *host range* and *pathway* were correlated; *pathway* and *trading partners* were correlated. These correlations were obvious by examining the biology of insects. For example, if an insect was polyphagous, i.e. it had a wider *host range*; it would have more chance to be associated with more *pathways*.

4.2.2.4 Nematodes

Table 4.8 Correlation and interaction of risk factors for nematodes

Risk factor	Climate suitability	Host range	Dispersal potential	Economic impact	Environmental impact	Pathway	Trading partners
Climate suitability	1						
Host range	0.31	1					
Dispersal potential	0.14	0.80(***)+	1				
Economic impact	0.43	0.68(***)+	0.73(***)+	1			
Environmental impact	-0.3	-0.05	-0.21	-0.15	1		
Pathway	0.17	0.40+	0.59(***)-	0.67(***)-	0.05	1	
Trading partners	0.19	0.34	0.53(**)	0.63(***)-	0.02	0.93(***)-	1
Overall risk	0.38	0.79(***)+	0.79(***)+	0.84(***)+	0.13	0.83(***)+	0.78(***)+

Number of species = 19

*** Correlation is significant at the 1% level

** Correlation is significant at the 5% level

\+ Indicates column has a direct effect on row

\- Indicates row has a direct effect on column

The results suggested:

- *Economic impact* was significantly affected by all other risk factors except *climate suitability* and *environmental impact*;
- *Climate suitability* and *environmental impact* were irrelevant to any other risk factors hence independent;

- The over risk was greatly effected by *economic impact* (0.84), followed by *pathway* (0.83), *host range* (0.79), *dispersal potential* (0.79), and *trading partners* (0.78). *Climate suitability* and *environmental impact* had no significant effect on the overall risk.

The correlations showed that *climate suitability* and *environmental impact* made little contribution to the overall risk. They were also independent to the other factors. This may be explained by the characteristics of nematodes – that many plant-attacking nematodes are parasitic; many species cause histological damages to roots, therefore the *climate suitability* and *environmental impact* were not that relevant.

4.2.2.5 Discussions of correlations for various taxonomic groups

Correlations of risk factors for bacteria (17 species), phytoplasma (11 species) and mites (four species) are not shown, as some of the correlations could not be computed because at least one of the variables was constant in each group.

It is interesting that some variables were constant across the whole group. If the variable was constant, the correlation would be zero and thus the variable would make no contribution to the overall score.

The correlation analysis for these taxa suggested that *economic impact* consistently contributed the most to the overall risk score across all taxa groups, which strongly reflected the definition of quarantine pest.

In all taxonomic groups discussed above, *economic impact* was significantly affected by other factors. The implication of this would be that *economic impact* could be regarded as a secondary factor, i.e. it is determined by other factors.

4.2.3 Relationships between the risk factors of the EPPO risk assessment scheme

Correlations of the forty-four risk factors used in the EPPO data were calculated using SPSS (v.11), and the interaction between risk factors was investigated.

Correlations and interactions at significant level of 0.01[7] are shown in table 4.9. A detailed explanation of the results is given in Appendix 9.

It is interesting to note that it seems some correlated factors can be grouped together, which could focus on a particular issue. For example, under entry potential, *remaining undetected* is correlated with *surviving existing control practices* and *effect of natural enemies on spread*. Those factors could all be linked back to the biology of a pest, e.g. an insect, if it is internal feeding and

---

[7] Reservation/limitation about *P* value discussed in section 4.2 still holds.

symptoms are not visible, then it is easy to skip detection/inspection, avoid the effect of control measure and natural enemy attack.

There are also similar examples in the results. The next section endeavors to group them and discuss the implication of the result for risk assessment.

Table 4.9 Significant correlation and interaction of risk factors (based on EPPO scheme)

Risk factor/ Pearson corr/ interaction	A1: Number of pathways	A3: Pest concentration	A4: Surviving existing practices	A5: Remaining undetected	A7: Multiplying in transit	A8: Movement along pathway	A10: Duration of consignment arrival	A11: Transfer to a suitable host	B1: Host species in PRA area	B2: Host extension	B3: Wild plants aiding dispersal	B4: Climate similarity	B5: Similarity of other abiotic factors	B7: Natural enemies (est.)	B8: Existing control measures	B9: Reproductive strategy	B10: Low populations being established	B12: Genetic adaptability	B13: Frequency of introduction	C1: Economic loss in existing areas	C2: Environmental damage in existing area	C3: Social damage in existing areas	C4: Extension of endangered area	C5: Spread by natural means in PRA area	C6: Spread by human assistance in PRA area	C7: Ease of containing of spread	C8: Direct effect on yield or quality in PRA area	C9: Effect on producer profit	C10: Effect on consumer demands	C11: Effect on export markets	C12: Other costs due to introduction	C14: Social damage in PRA area	C15: Natural enemies (spr.)	C16: Ease of being controlled
A1: Number of pathways	1																			.68-					.65-									
A3: Pest concentration		1									-0.68																							
A4: Surviving existing practices			1	0.77																														
A5: Remaining undetected			0.77	1																													0.66	
A7: Multiplying in transit					1	0.69										.79+	.68+	.64+	.71-						.66-								-0.66	

A8: Movement along pathway					0.69	1																			.68-									
A10: Duration of consignment arrival							1							.76-				-0.72																
A11: Transfer to a suitable host								1	.67+	.68+																								
B1: Host species in PRA area								.67-	1	.90-																								
B2: Host extension								.68-	.90+	1																								
B3: Wild plants aiding dispersal		-0.68									1																-0.91	-0.7						
B4: Climate similarity												1		-.71-									.79-											
B5: Similarity of other abiotic factors													1													.74-								
B7: Natural enemies (est.)							.76+					-.71+		1				-.65+																
B8: Existing control measures															1			-.72+						-.67-	-.66-								.68+	

B9: Reproductive strategy					.79-											1		0.72							71-								-.73-	
B10: Low populations being established					.68-												1																	
B12: Genetic adaptability					.64-		-0.72							-.65-	-.72-	0.72		1						.71-	.65-								-.71-	
B13: Frequency of introduction					.71+														1															
C1: Economic loss in existing areas	.68+																			1					0.68		.65-	.74-			.68-			
C2: Environmental damage in existing area																					1								-0.67					
C3: Social damage in existing areas																						1										.74-		
C4: Extension of endangered area												.79+											1											

C5: Spread by natural means in PRA area														-.67+			.71+						1										
C6: Spread by human assistance in PRA area	.65+				.66+	.68+								-.66+	.71+		.65+		0.68					1				0.71					
C7: Ease of containing of spread												.74+													1								
C8: Direct effect on yield or quality in PRA area										-0.91									.65+							1	.84-						
C9: Effect on producer profit										-0.7									.74+							.84+	1	.75+					
C10: Effect on consumer demands																				-0.67				0.71			.75-	1			.70-		
C11: Effect on export markets																													1	.75+			.84+
C12: Other costs due to introduction																			.68+										.75-	1			
C14: Social damage in PRA area																					.74+							.70+			1		

C15: Natural enemies (spr.)				0.66	-0.66										.68-	-.73+		-.71+															1	
C16: Ease of being controlled																														.84-				1

Number of species = 15

Only correlations of significance at 0.01 level (2-tailed) are shown.

+ Indicates column has a direct effect on row

- Indicates row has a direct effect on column

The correlation analysis also suggested that some risk factors either had no significant correlation with others or correlated to others at a less significant level ($p = 0.05$)[8] and were therefore relatively independent. These factors (listed as heading row in Table 4.10) and their correlation with all the other risk factors are listed in Table 4.10.

---

[8] Reservation/limitation about P value discussed in section 4.2 still holds.

Table 4.10 Less significant correlation of risk factors (based on EPPO scheme)

Risk factor/ Pearson correlation	A2: Association with pathway	A6: Surviving in transit	A9: Commodity distribution	A12: Intended use of commodity	B6: Competition with existing species	B11: Ease of being eradicated	C13: Environmental damage in PRA area	C17: Impact on control of other pests	C18: Other side-effects	C19: Resistance to plant protection products
A1: Number of pathways							.63-			
A3: Pest concentration	.60-									
A8: Movement along pathway			.59							
A9: Commodity distribution			1					-.62		
A10: Duration of consignment arrival			.62					-.54		
A12: Intended use of commodity				1						
B1: Host species in PRA area						.52-				
B3: Wild plants aiding dispersal	-.60-									
B4: Climate similarity										-.54
B5: Similarity of other abiotic factors									.58	
B6: Competition with existing species					1	.53-				
B7: Natural enemies (est.)		.62								
B11: Ease of being eradicated					.53+	1				
C1: Economic loss in existing areas									.59-	
C2: Environmental damage in existing area							.63-			
C3: Social damage in existing areas					.58	.60				
C6: Spread by human assistance in PRA area	.54+							-.53-		
C8: Direct effect on yield or quality in PRA area									.55-	
C9: Effect on producer profit									.59-	

C12: Other costs due to introduction		.55							.60-	
C14: Social damage in PRA area					.63+					
C17: Impact on control of other pests			-.62					1		

Number of species = 15

Only correlations of significance at 0.05 level are shown.

+ Indicates column has a direct effect on row

- Indicates row has a direct effect on column

Results of the correlations and interactions between the risk factors of the EPPO scheme suggested that most of the risk factors were correlated with each other. This made sense as there were so many factors considered in the scheme and there ought to be some interrelationships among them.

However, those risk factors were correlated to each other at different significance level, some were independent of others. There could be some implications of the results. It was suggested that risk factors be divided into two types:

- Primary risk factor, which has direct effect on but is not affected by others, or independent of others. Such factors include *pest concentration, reproduction strategy, genetic adaptability, number of pathways, host species in PRA area, climate similarity,* and *intended use of commodities* etc.
- Secondary risk factor, which is affected by or derived from other factors. Such factors include *economic loss, social damage, extension of endangered area, spread by human assistance, effect on yield on quality*

in PRA area, effect on producer's profit, and *transfer to a suitable host* etc. Some secondary factors might be modelled by others.

4.2.4 Conclusions and discussions

The analyses of the correlations and interaction between risk factors revealed that:

- Some risk factors are highly correlated to others; some are relatively independent.
- Risk factors of different taxonomic group of pests correlate and interact differently.
- Within the same taxonomic group, each risk factor has different effect on the overall risk; for different taxonomic group, the same risk factor affects the overall risk differently. This maybe due to the characteristics of the pests. This result suggests that it is not plausible to develop a generalised scheme for different taxonomic groups.
- Climate suitability and environmental impact are relatively independent.
- Pathway and trading partner are highly correlated, probably because more trading usually means more pathways; it is possible to combine the two together.
- Risk factors could be classified into two different types: primary factor and secondary factor (see 4.4.2). This seems to have further applications. One possibility would be to develop a summary scheme with all primary factors, which may provide an approximate and quick estimation of the risk level. However, it is still important to find out what role the secondary factor play,

i.e. how do they modify the result. On the other hand, a cluster of inter-correlated factors might as a group form a primary factor. Due to the scope of this thesis, this is not studied further, but may warrant future study.

- Another implication of the result of correlation analysis is that it would be a useful first step for the application of Bayesian nets in PRA. In a recent study of Bayesian application by Holt, some nodes didn't depend on others whilst others were conditional on other nodes. Establishing the conditional factors would be a useful first step for this (Holt 2009, pers. comm.).
- Economic impact and dispersal potential can be interpreted as secondary risk factors as they were directly affected by other risk factors, and there were no direct effect on others.
- When large numbers of risk factors are involved, some very strong correlations are likely to exist, which means there is redundancy involved; those highly correlated factors might be combined or redefined or modelled by others.

4.3 IMPLICATIONS OF THE MEAN AND VARIANCE OF RISK SCORES

Section 4.2 suggested that there were significant correlations among risk factors, especially risk factors under EPPO schemes. In this section, using EPPO data, it was considered whether all the factors were indeed contributing equally to the outcome of a risk assessment, and how they contributed to the overall risk rating. Risk factors were further examined whether they could be grouped.

For example, if a risk factor has a low average risk score in the risk assessment of many different quarantine pests and does not vary much from pest to pest, then the risk factors tend not to contribute much to the overall risk assessment (see 4.2.2.5).

On the other hand, if a risk factor has a high score variation among different quarantine pests, the risk factor may be more likely to provide useful information to determine the level of the risk.

It is therefore necessary to consider the possibility of simplifying the risk assessment process by reducing the number of risk factors. In doing so, it would be easier to distinguish the comparative importance of each factor, thus it would be easier to assign weightings to risk factors, and risk assessment may be simplified.

In this section, using 15 cases of risk assessments carried out under the EPPO regime, the following questions were considered:

1. Are there discriminating risk factors, i.e. one or a combination of risk factors, which could be used to identify high-risk pests?
2. Could some factors be eliminated from the process? If so, what are they?

4.3.1 Method of analysis

With data of only 15 risk assessment cases and 44 risk factors to be considered, the options for analysis were limited. Due to the small number of observations and the relatively large number of variables, the common statistical factor reduction methods, such as factor analysis, discriminate analysis, cannot look directly into the relationship between overall risk assessment and individual risk factor scores.

This study however tried to answer the questions posed above by investigating the patterns of risk factor scores among different pest risk assessment cases under the EPPO scheme rather than looking directly into the relationship between overall risk assessment and individual risk factor score.

Two simple statistics, mean and standard deviation, were first examined (Table 4.2). The mean of a risk factor is the average score of the risk factor for all the species considered for risk assessment, whereas the standard deviation is a measure of how widely the scores for each risk factor are dispersed from the mean across different species.

After looking at the means and standard deviations of the risk factors, a cluster analysis was conducted on the 44 risk factors to determine the natural groupings of the risk factors based on the mean and standard deviation of the risk factor scores.

Cluster analysis, also called segmentation analysis or taxonomy analysis, attempts to identify the natural groupings of observations. That is, cluster analysis seeks to identify a set of groups, which both minimize within-group variation and maximize between-group variation. Cluster analyses were performed in this study using statistical software Stata (version 7.0) (StataCorp, 2001).

4.3.2 Empirical results

4.3.2.1 Clustering the risk factors

The mean of a risk factor represents the average perceived level of risk associated with that risk factor in all the species considered for risk assessment. If the mean of the scores of a risk factor was very high, it is reasonable to assume that the risk assessor has concern about that factor and no matter what the standard deviation is; it should not be eliminated from further consideration in determining whether an organism is a quarantine pest.

The standard deviation indicates to what extent a risk factor can be used as a discriminating factor. If a risk factor had a higher variation, it indicates that risk factor shows relatively big differences across different cases, thus it could help to distinguish (together with others) the level of the overall risk rating; on the contrary, if a risk factor had a low score variation, that means the perceived risk is similar for every species, hence such risk factors do not help to distinguish the level of the risk between the species concerned, it is therefore not considered a discriminating risk factor.

Table 4.2 in section 4.1 showed the risk factors considered under the EPPO scheme and the means and standard deviations of the scores of risk factors for 15 insect species.

Figures 4.1 and 4. 2 show the histograms of the mean and variance of the risk factor scores.

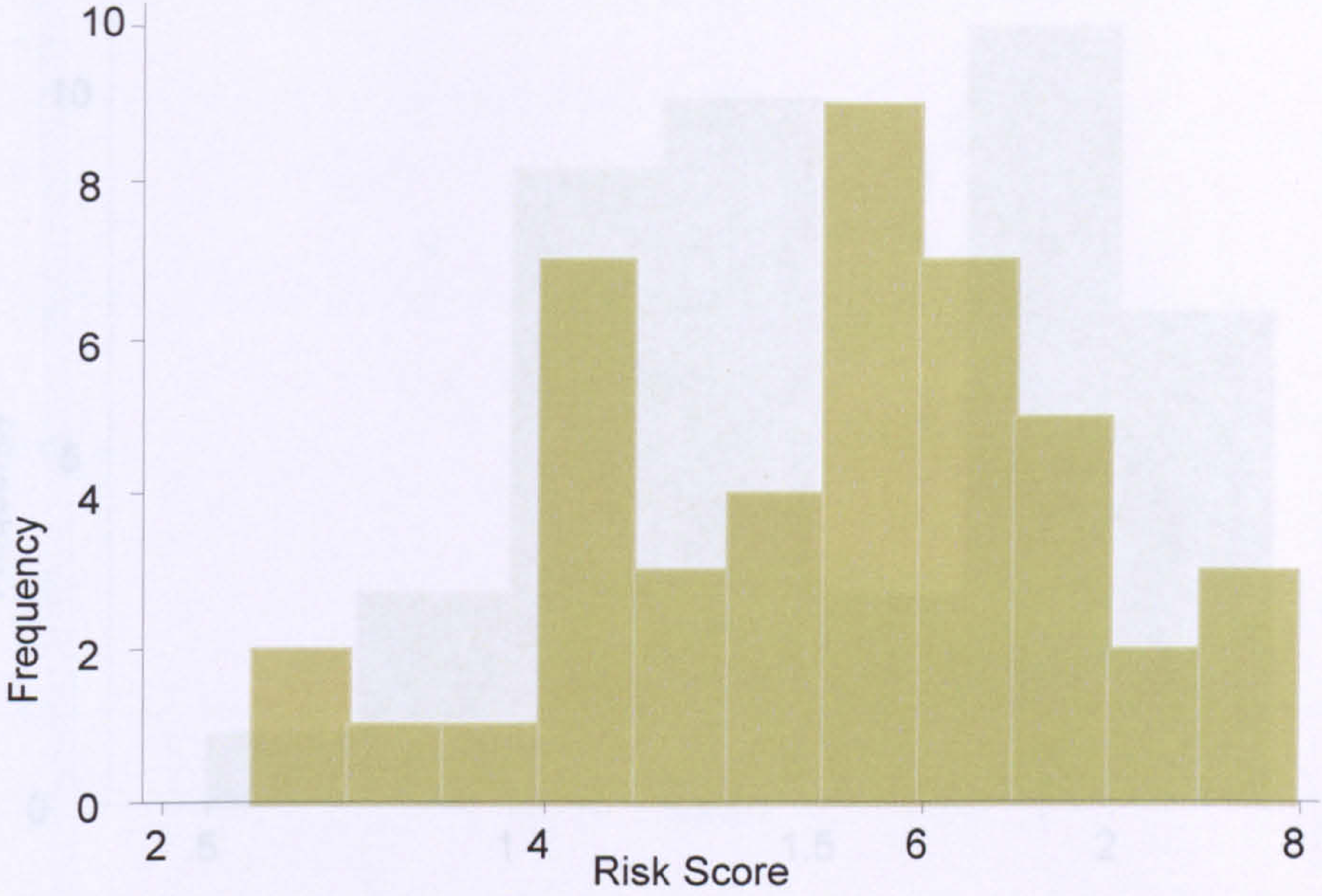


Figure 4.1 Histogram of the mean of the risk factor[9] scores

Figure 4.1 showed that there was a wide distribution of mean values. For example, the average risk score for risk factor *economic loss in existing area* was 7.80[10], and the average risk score for risk factor *similarity of other abiotic factors* was 7.67. However, the average risk score for risk factors *multiplying in transit* and *impact on control of other pests* were only 2.47 and 2.60, respectively.

---

[9] This may indicates a hint of bimodality with peaks at both 4 and 6. However, it is difficult to say. This may be due to the design of the scheme as well, because it is difficult for the assessors to assign risk scores precisely under a 1-9 system.

[10] Which can be explained by the reason of the risk assessments, as most of the pests here were of concern by EPPO, and economic loss in existing area was one of the reasons that brought that pest to EPPO's attention.

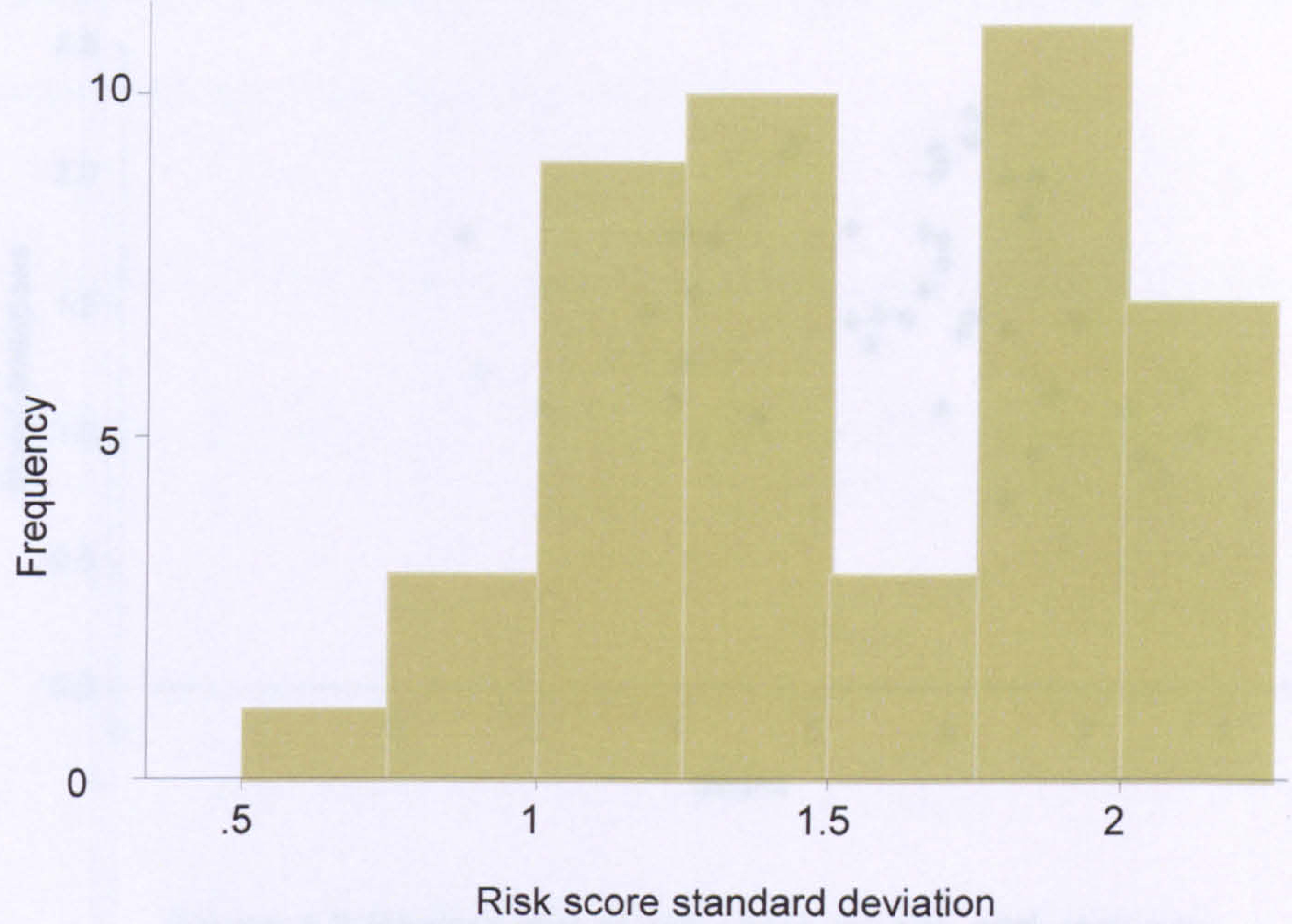


Figure 4.2 Histogram of the standard deviation of the risk scores

Figure 4. 2 suggested that the scores for some risk factors showed large variation between the species concerned, while the risk scores for others were rather similar for all species. For example, the standard deviation for risk factor *wild plants aiding dispersal* was 2.23, while the standard deviation for risk factors *ease of containing of spread* was only 0.74.

Figure 4. 3 shows the scatter plot of risk score means and variance.

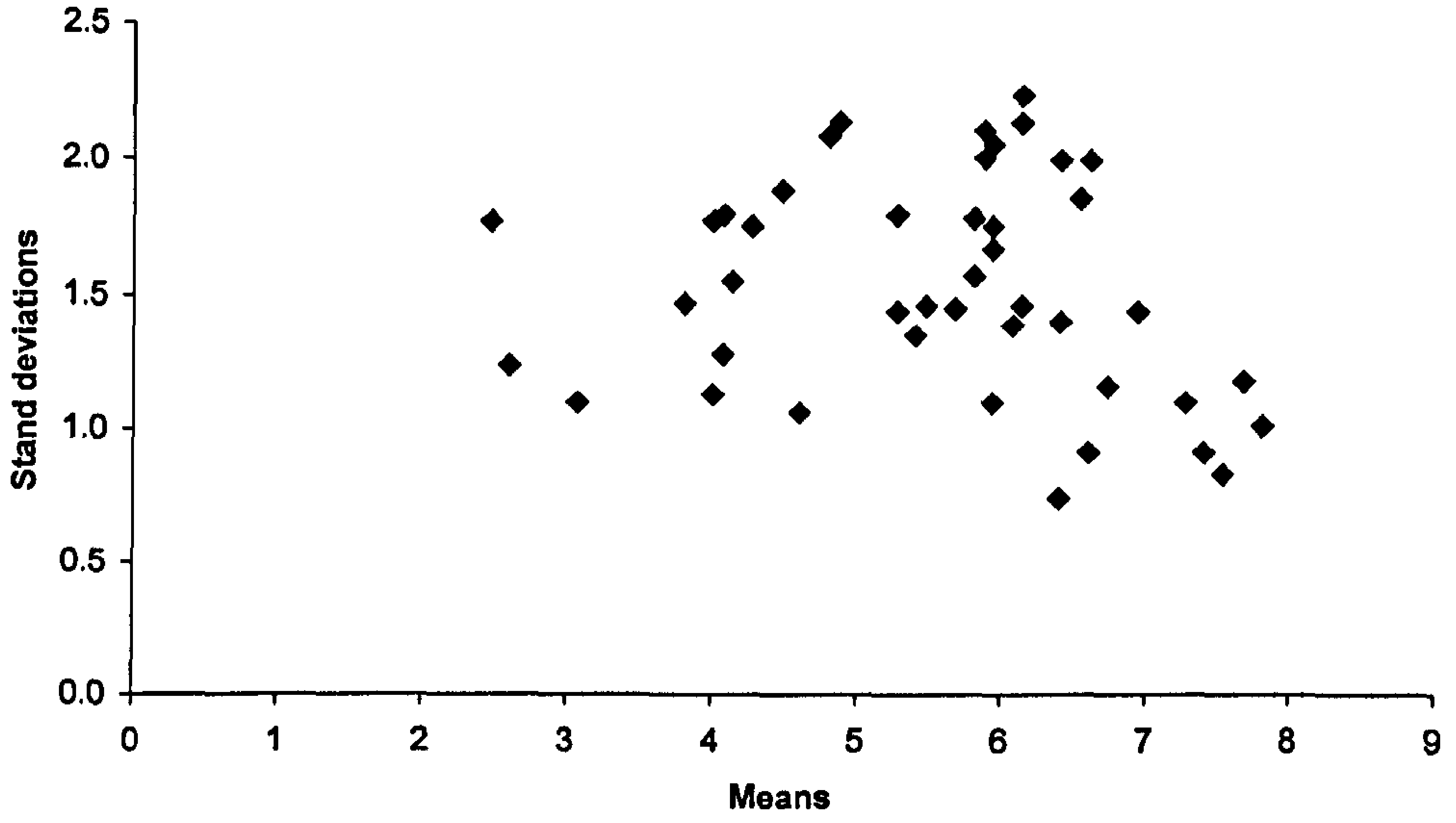


Figure 4.3 Scatter plot of risk score means and variance

In order to investigate any patterns in the risk factors, cluster analyses were performed for risk score means and standard deviations using Stata (StatCorp, 2001). Initially, it was attempted to classify the data into 3, 4, 5, 6, 7, 8, and 9 clusters, in an attempt to obtain the natural groupings of the risk factors.

The Calinski and Harabasz (1974) pseudo-F index were calculated in regards to different clustering in order to determine the number of clusters. Large values of the Calinski and Harabasz (1974) pseudo-F index indicate distinct clustering.

The Calinski and Harabasz pseudo-F index for the different groupings were shown in Table 4.11. The 7-cluster grouping had the largest

Calinski/Harabasz Pseudo-F index, which suggested that the grouping with 7 clusters was the most distinctive clustering.

Table 4.11 Calinski/Harabasz Pseudo-F index of different clustering

Number of Clusters	Calinski/Harabasz Pseudo-F Index
3	82.56
4	68.98
5	89.7
6	50.8
7	**98.59**
8	74
9	69.48

The results of cluster analyses with the grouping of 7 clusters were shown in Table 4.12, which showed the cluster centers. Based on the distribution of the risk score means, it was decided that the means of risk score under 4 were low, the means between 4 and 6 were medium, and the means between 6 and 8 were high. For risk score variance, it was determined that the risk score variance under 1.3 (inclusive) were low, between 1.3 and 1.8 were medium, and over 1.8 were high.

Table 4.12 Cluster centres (7 Clusters)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Mean	5.72	4.12	6.18	2.71	6.61	4.41	7.53
	M	M	H	L	H	M	H
Variance	1.52	1.3	2.04	1.37	1.13	1.9	1.01
	M	L	H	M	L	H	L

Figure 4.4 showed the scatter plot of risk score mean and standard deviation by clusters. It suggested that the 7 clusters identified had indeed reasonably grouped the risk factors into distinctive groups.

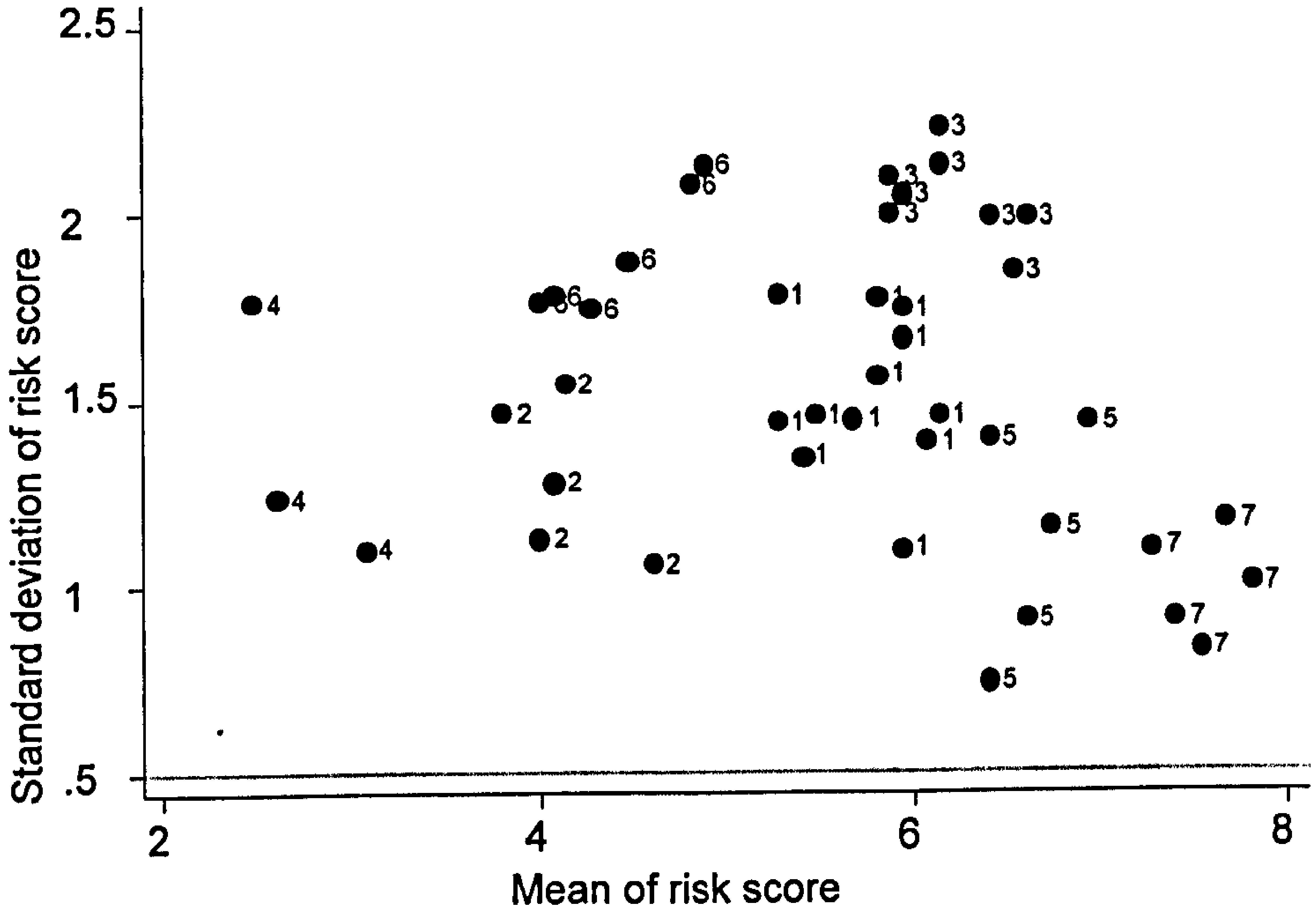


Figure 4.4 Scatter plot of risk score mean and variance by clusters (7 clusters)

Table 4.13 below showed the cluster membership of each risk factor and the meaning of each cluster.

Table 4.13 Cluster memberships of risk factors

Cluster 1 **Medium Mean** **And** **Medium Variance**	**Cluster 2** **Medium Mean** **And** **Low Variance**	**Cluster 3** **High Mean** **And** **High Variance**	**Cluster 4** **Low Mean** **And** **Medium Variance**	**Cluster 5** **High Mean** **And** **Low Variance**	**Cluster 6** **Medium Mean** **And** **High Variance**	**Cluster 7** **High Mean** **And** **Low Variance**
Transfer to a suitable host	Resistance to plant protection products	Environmental damage in existing area	Impact on control of other pests	Competition with existing species	Number of pathways	Economic loss in existing area
Effect on export markets	Environmental damage in PRA area	Host species in PRA area	Multiplying in transit	Ease of containing of spread	Effect on consumer demand	Ease of the pest being controlled
Association with pathway	Other costs due to the introduction	Commodity distribution	Social damage in PRA area	Natural enemies	Reproductive strategy	Surviving in transit
Surviving existing practices	Other side-effects	Duration of consignment arrival		Spread by human assistance in PRA area	Movement along pathway	Similarity of other abiotic factors
Direct effect on yield or quality in PRA area	Frequency of introduction	Intended use of commodity		Natural enemies	Spread by natural means in PRA area	Ease of being eradicated
Remaining undetected		Climate similarity			Social damage in existing area	
Extension of endangered area		Wild plants aiding dispersal				
Effect on producer profits		Host extension				
Pest concentration						
Existing control measures						
Genetic adaptability						
Low populations being established						

4.3.2.1 Risk factors classified in different groups

Based on the 7-cluster result (see Table 4.12 and Figure 4.4), also taking into account the risk factors falling into each clusters (see Table 4.13), the 44 Risk factors were classified into 4 groups:

- Group 1: Risk factors in clusters 5 and 7, which have low score variance with high score mean;
- Group 2: Risk factors in clusters 3 and 6, which have high score variance with high-medium score mean;
- Group 3: Risk factors in cluster 1, which have medium score variance with medium score mean.
- Group 4: Risk factors in Clusters 2 and 4, which have low variance/medium mean or medium variance/low mean

Group 1 (L variance/H mean) includes the following factors:

- *Competition with existing species*
- *Ease of being eradicated*
- *Ease of containing of spread*
- *Ease of the pest being controlled*
- *Economic loss in existing area*
- *Natural enemies (establishment)*
- *Natural enemies (spread)*
- *Similarity of other abiotic factors*
- *Spread by human assistance in PRA area*
- *Surviving in transit*

Group 2 (H variance/H-M mean) includes the following factors:

- *Climate similarity*
- *Commodity distribution*
- *Duration of consignment arrival*
- *Effect on consumer demand*
- *Environmental damage in existing area*
- *Host extension*
- *Host species in PRA area*
- *Intended use of commodity*
- *Movement along pathway*
- *Number of pathways*
- *Reproductive strategy*
- *Social damage in existing area*
- *Spread by natural means in PRA area*
- *Wild plants aiding dispersal*

Group 3 (M variance/M mean) includes the following factors:

- *Association with pathway*
- *Direct effect on yield or quality in PRA area*
- *Effect on export markets*
- *Effect on producer profits*
- *Existing control measures*
- *Extension of endangered area*
- *Genetic adaptability*
- *Low populations being established*

- *Pest concentration*
- *Remaining undetected*
- *Surviving existing practices*
- *Transfer to a suitable host*

Group 4 (L variance/M mean or M variance/L mean) includes the following factors:

- *Environmental damage in PRA area*
- *Frequency of introduction*
- *Impact on control of other pests*
- *Multiplying in transit*
- *Other costs due to the introduction*
- *Other side-effects*
- *Resistance to plant protection products*
- *Social damage in PRA area*

4.3.3 Discussions and conclusions

The results from cluster analysis suggested that the means and variance of risk factor scores could be effectively used to determine the importance of risk factors in terms of how much they contribute to the overall risk assessment and distinguish the level of overall risk.

The following is proposed for the application of the grouping of risk factors.

4.3.3.1 Risk factors for preliminary assessment

Risk factors in Group 1, with low variance/high mean risk factors (clusters 5 and7), contribute little to distinguishing the level of risk between species. However, as revealed by examining the risk factors, they are the most likely definitional factors of a quarantine pest, i.e. implicit in characterising a quarantine pest. Therefore factors in this group could be used for preliminary assessment, to decide whether an organism has the characteristics of a quarantine pest.

4.3.3.2 Risk factors for determining the level of risk

Risk factors in Group 2, with high variance/high-medium mean (clusters 3 and 6), could distinguish the level of risk between species more efficiently. Therefore factors in this group could be used as the key risk factors to determine whether a potential quarantine pest risk is high or low.

Risk factors in Group 3, with medium variance/medium mean (cluster 1), are also of some importance, they have some contribution to distinguishing the level of risk, but not as much as those in Group 2; they also contribute, to a certain extent, to determining whether an organism could be a quarantine pest. By examining the factors in this group, some of them are intrinsic characteristics of an organism, such as reproductive strategy, which are not always contributing much in a risk assessment, especially for a group of similar organisms, but they can become very important in distinguishing organisms from different taxa. Therefore it is proposed that this group of risk factors be kept as factors to determine the level of risk.

4.3.3.3 Risk factors that could be eliminated from the scheme

Risk factors in Group 4, with low variance/low-medium mean (clusters 2 and 4), either contribute little to the assessment of the level of risk or are dependent on or can be expressed by other risk factors. Therefore they could be eliminated from the scheme.

By grouping risk factors for different purposes, risk assessment could be potentially simplified, but rigor still could be retained.

There are some interesting similarities between the results from the cluster analysis and that of the Delphi study (see 3.3.4.4), which gives the author more confidence in these ideas, especially as the results were derived from different data: one was based on real RA cases; the other was based on expert opinion on risk factors without specifying a pest.

For example, some risk factors in Group 4, which were suggested to be eliminated, also obtained low weightings in the Delphi study, e.g. *frequency of introduction* and *resistance to plant protection products.* This aligned well in the two studies: in cluster analysis, those two factors had low variance/low-medium mean, which meant that they were not considered important or contribute much to the overall risk score.

Some factors in Group 2 and 3, which were suggested as useful for determining the level of risk, obtained high weightings in the Delphi study, e.g. *pathway, host availability* and *spread by natural means.* These factors were also aligned well between the two studies.

It should be noted however, that the cluster analyses were based on only 15 species, and most of them were forestry insects. The result may not accurately represent taxonomic groups other than insecta (see discussion in 4.2.4).

Also, the 15 risk assessments used in this study were conducted prior to 2002 so the perceived importance of some factors may have changed along with better understanding of some factors and the changing focus of PRA. For example, RA used to focus on economic importance and crop pests. Over the years, it is realised that environmental and social consequences are also important and they are gaining more attention. Another notable change was the inclusion of invasive plants and living modified organisms in the PRA scope. All these changes would affect the perceived importance of risk factors. Hence if more recent data were available, the result may be different.

Further study is needed to evaluate the applications proposed above. The evaluation would need to answer the following questions:

- Whether an organism would be characterised as a quarantine pest by only using risk factors in Group 1; and

- Whether it would achieve similar results of overall risk by only using factors in Groups 2 and 3.

However, the author is confident about the method proposed in this study, which provides a good insight to the risk factors based on real PRA cases: this method helps to discover what factors are essential and what are redundant.

4.4 INCORPORATING WEIGHTINGS INTO PEST RISK ASSESSMENT: AN APPLICATION OF PRINCIPAL COMPONENTS ANALYSIS

4.4.1 Introduction

As discussed in the previous sections, multicollinearity[11] exists in the risk factors. To determine which factors are important is difficult and this problem becomes particularly acute in measure the importance of risk factors.

A possible solution to the problem of multicollinearity is to reduce the number of factors by removing some factors or combining two or more correlated factors, while accounting for factor importance can be achieved by incorporating weighting into pest risk assessment.

A weighting is a value given to a risk factor according to how important it is perceived to be, or how significant its contribution to the overall risk rating. The larger the value, the more important the factor is to the overall risk assessment.

By giving different weightings to different risk factors, factors that are low contributors to the overall assessment may be filtered out while the important

---

[11] Multicollinearity refers to a situation in which two or more explanatory variables in a regression model are highly correlated. It is perfect multicollinearity if the correlation between two independent variables is equal to 1 or -1.

ones will be retained. Consequently, the process of risk assessment could possibly be simplified to provide a more concise pest risk assessment[12].

In this section, it is proposed to apply a statistical factor reduction technique, Principal Components Analysis, to reduce factor multicollinearity and to derive risk factor weightings.

This part of the chapter is arranged as follows: Section 4.4.2 describes the application of Principal Components Analysis (PCA), and the way to derives risk factors weightings; Section 4.4.3 reports the results of applying PCA technique to Tanzanian data; Section 4.4.4 concludes this study.

4.4.2 Methodology: Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a multivariate analysis technique that was first introduced by Pearson in 1901 (Pearson 1901) and developed independently by Hotelling in 1933 (Hotelling 1933). It is commonly used to eliminate collinearity and reduce the dimensions of a data set with a large number of interdependent variables. As has been shown in 4.2, pest risk assessment tends to lead to data sets of this type.

PCA involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called *principal components* (PCs), which are linear combinations of the variables that explain the maximum amount of variance in the original

---

[12] The benefit of incorporating weighting in risk assessment was discussed in Chapter 3 (3.5.4)

variables (see Appendix 11 for details of the mathematics of PCA).

The first component accounts for the most variance in the variables. Then the second component accounts for the largest share of the remaining variance, and so on.

4.4.3 Empirical results

In this section, PCA was first applied to a data set comprising 252 quarantine pests in Tanzania (Black and Abdallah 1997), and then applied to different taxonomic groups of pests[13].

4.4.3.1 Principal Components Analysis of Tanzania quarantine pests

The initial results are presented in Table 4.14. There were seven PCs identified initially. The eigenvalues of each PC are listed in Column 2 from the largest to smallest in Table 4.14. The third column of Table 4.14 lists the differences between each eigenvalue and its next smaller eigenvalue. Column 4 lists the variance in the seven risk factors explained by each PC, e.g. the first PC explained 36.1% of the total variance, the second explained 16.8%, and so on. Column 5 lists the cumulative variance explained by the PCs. It shows that the first 4 PCs explained 78% of the total variance.

---

[13] *See 4.1 for data description*

The result suggested that the seven original risk factors are correlated with each other and could be reduced, however more than one PC is needed to accommodate all the risk factor variance.

Table 4.14 Principal component analysis of Tanzanian quarantine pest risks

PCs	Eigenvalue	Difference	Proportion	Cumulative
Component1	2.5276	1.34919	0.3611	0.3611
Component2	1.1784	0.13646	0.1683	0.5294
Component3	1.0419	0.32528	0.1489	0.6783
Component4	0.7167	0.05511	0.1024	0.7807
Component5	0.6616	0.12547	0.0945	0.8752
Component6	0.5361	0.19855	0.0766	0.9518
Component7	0.3376	.	0.0482	1

Determining the number of useful PCs of the data and eliminating noisy components is always ambiguous. In regard to how many PCs to retain and extract, two criteria have been applied (see Appendix 11 for details). The first criterion is the Scree Plot test, in which the eigenvalues are plotted in the sequence of the principal factors. The number of factors is chosen where the plot starts levelling off to a linear decreasing pattern. Figure 4.5 suggests a 3 or 4-PC solution, since the eigenvalues begin a linear decline commencing with the fourth or fifth PC.

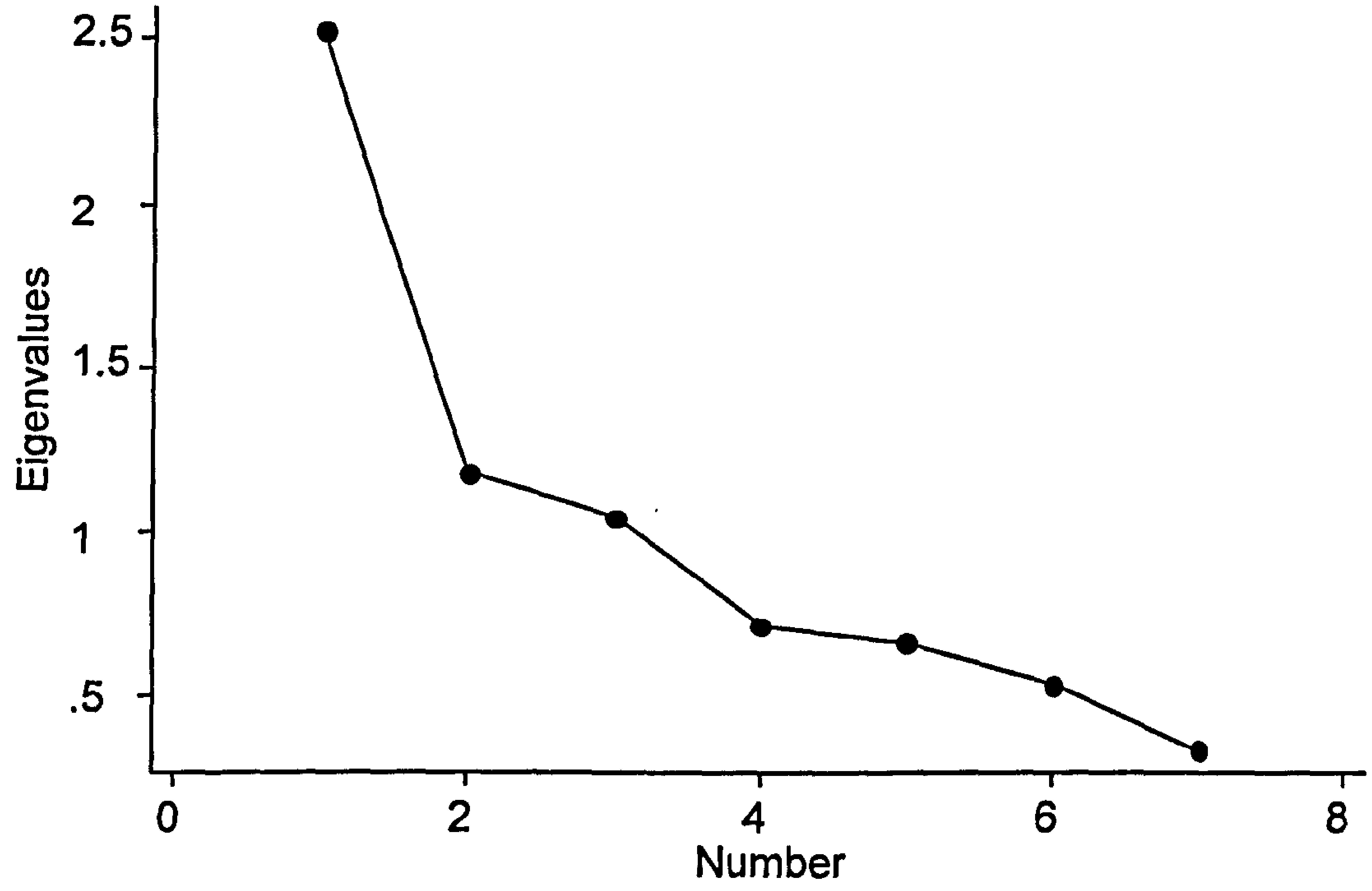


Figure 4.5 Scree plot of eigenvalues after PCA

Another criterion is proposed by Everitt and Dunn (1992) suggests discard all components accounting for less than $(70/n)\%$ of the overall variance, where n is the number of PCs. In this study, the 5th, 6th, and 7th components each accounted for less than 10% of the total variance. This criterion suggested a four-factor solution as the fourth PC explained around 10% of total variance.

Based on these two selection criteria, the first four PCs were retained and extracted, which explained 78% of the total variance.

Principal component is a linear combination of various risk factors, the coefficient of the risk factors being the scoring coefficient on the principal components.

Each scoring coefficient indicates the weighting of a particular risk factor in that principal component axis. The scoring coefficient of the original risk factor indicates its influence. Larger scoring coefficient indicates a greater explanatory power of these risk factors. Table 4.15 shows the scoring coefficients of the four retained PCs. Scoring coefficients with values larger than 0.3 are in bold.

Table 4.15 Scoring coefficients of principal components

Risk Factors	**PC1**	**PC2**	**PC3**	**PC4**
Climate suitability	-0.0665	-0.0213	**0.9316**	-0.0317
Host Range	**0.7953**	-0.0831	-0.152	-0.0754
Dispersal potential	**0.4177**	0.0454	0.2966	0.0954
Economic impact	**0.4298**	0.2299	0.1322	0.0956
Environmental impact	-0.0247	-0.0103	-0.024	**0.9872**
Pathway	-0.0507	**0.6902**	0.0003	-0.0176
Trading Partner	-0.027	**0.6791**	-0.0546	-0.0161
Eigenvalue	2.5276	1.1784	1.0419	0.7167
Variance Explained	36.11%	16.83%	14.89%	10.24%

Each Principal Component (PC) is a linear combination of the scores of various risk factors[14]. The four PCs can be interpreted as follows:

---

[14] The scores for all the various risk factors are standardised with mean 0 and standard deviation of 1.

$$PC1 = -0.0665 * Climate + 0.7953 * Host + 0.4177 * Dispersal + 0.4298 * Economic \\ - 0.0247 * Environmental - 0.0507 * Pathway - 0.027 * Trading$$

$$PC2 = -0.0213 * Climate - 0.0831 * Host + 0.0454 * Dispersal + 0.2299 * Economic \\ - 0.0103 * Environmental + 0.6902 * Pathway + 0.6791 * Trading$$

$$PC3 = 0.9316 * Climate - 0.152 * Host + 0.2966 * Dispersal + 0.1322 * Economic \\ - 0.024 * Environmental + 0.0003 * Pathway - 0.0546 * Trading$$

$$PC4 = -0.0317 * Climate - 0.0754 * Host + 0.0954 * Dispersal + 0.0956 * Economic \\ + 0.9852 * Environmental - 0.0176 * Pathway - 0.0161 * Trading$$

The first PC has large positive coefficients for *host range* (0.7953), *dispersal potential* (0.4177), and *economic impacts* (0.4298). It represents therefore mainly the importance of *host range, dispersal potential*, and the related *economic impacts*. It explains 36.1% of the total variance.

The second PC has large positive coefficients for *pathway* (0.6902) and *trading partner* (0.6791). It represents mainly the importance of *pathway* and *trading partners*. It explains around 17% of the total variance.

The third PC has a large positive coefficient for *climate suitability* (0.9316), which means it represents mainly the importance of *climate suitability*. The fourth PC has a large positive coefficient for *environmental impacts* (0.9872). It represents the importance of *environmental impacts*. The third and fourth

PCs explain around 15% and 10% of the total variance, respectively, which indicates the relatively lower explanatory power of these two risk factors.

It is interesting to note that PC3 and 4 are largely influenced by a single original risk factor. This might suggest that *climate suitability* and *environmental impact* are largely either uncorrelated with or unaffected by other risk factors.

This aligns well with the result of risk factor correlation (see 4.2). Both studies come to a conclusion that *environmental impact* is independent of other factors, and *climate suitability* is not influenced by others.

4.4.3.2 Pest risks and principal components for all pest species

Using the aggregated risk score, Black and Abdallah (1997) classified the risks of potential quarantine pests into high, medium, and low risk. In this section, the relationship was examined between the risk classification by Black and Abdallah (1997) and the PCs extracted as above, an attempt was made to identify whether the pest risk classification can be characterised by one or more PCs.

Figure 4.6 shows the median[15] scores of the extracted PCs within each pest risk classification.

---

[15] The reason to use median score rather than average score is that median score is more likely to capture the property of a pest risk classification. Because the assessments were qualitative, by using

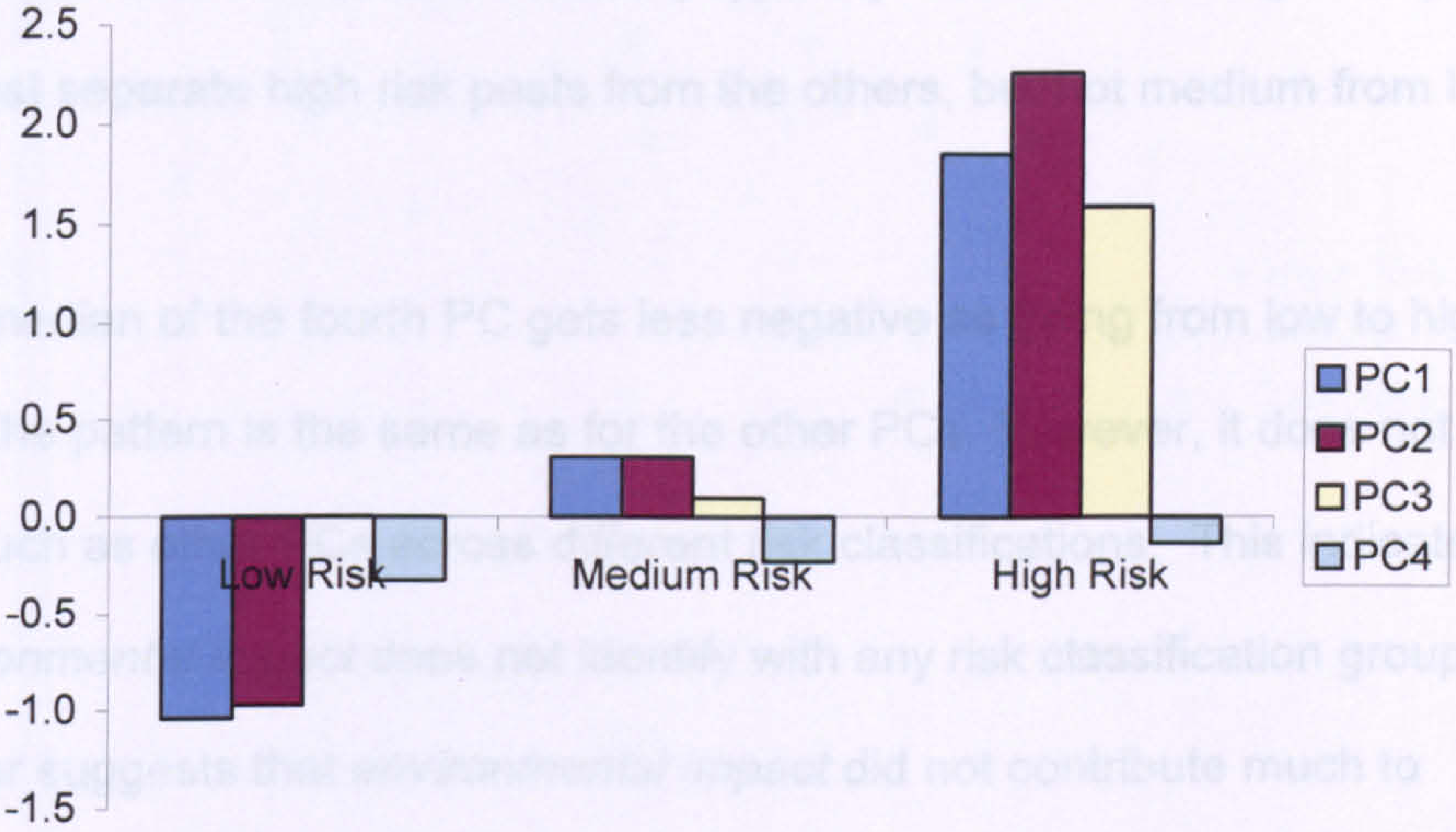


Figure 4.6 Medians of PC scores by pest risk classification

It is noticed that for the high-risk quarantine pests, the first, second, and the third PC scores tend to have higher median values, indicating that high-risk pests tend to have higher risk scores on *host range*, *dispersal potential*, *economic impact*, *pathway*, *trading partner*, and *climate suitability*.

For low risk quarantine pests, the median of the first and second PCs are both negative. This indicates that low risk pests tend to have lower risk scores on *host range*, *dispersal potential*, *economic impact*, *pathway* and *trading partner*.

---

median score can leave out the extreme or marginal members within the classification, which might be incorrectly classified.

The median of the third PC is only non-trivial for high-risk pests, indicating the higher risk score of *climate suitability* typically identifies with high risk pests, or at least separate high risk pests from the others, but not medium from low.

The median of the fourth PC gets less negative as going from low to high risk, thus the pattern is the same as for the other PCs. However, it does not vary as much as other PCs across different risk classifications. This indicates that *environmental impact* does not identify with any risk classification groups. This further suggests that *environmental impact* did not contribute much to distinguish the level of risk when the assessments were originally conducted.

It should be noted that high, medium, and low risks are defined by score aggregation as in Black and Abdallah (1997). There is no absolute measure of risk here. In fact what is being done here is to compare two ways of classifying the level of risk. One is based on the real PRA cases; the other is based on the model developed here, which is a prediction based on a subset of the available information.

The scatter plots of PC2 against PC1 (Figure 4.7), PC3 against PC1 (Figure 4.8), and PC4 against PC1 (Figure 4.9) are plotted.

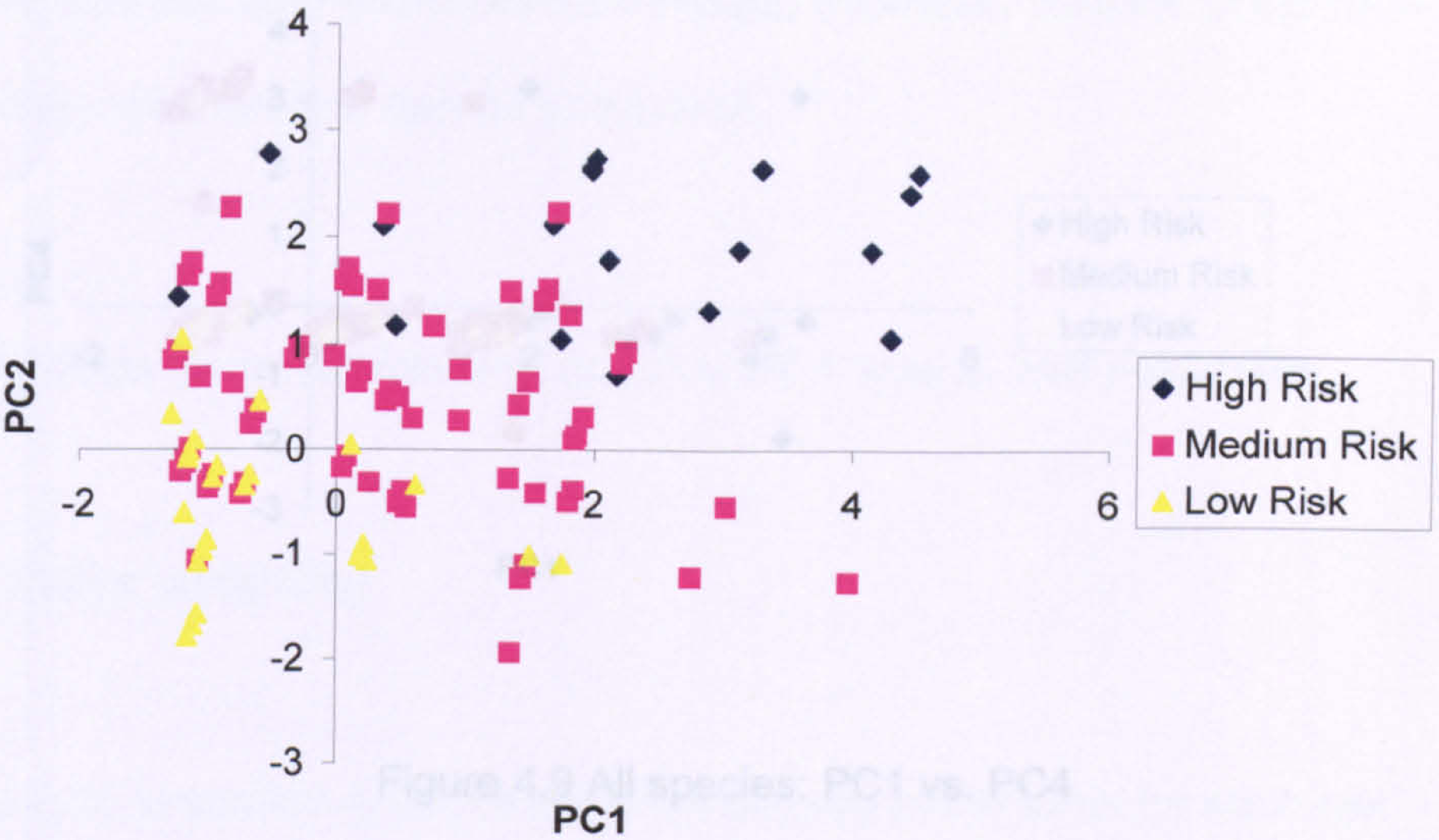


Figure 4.7 All species: PC1 vs. PC2

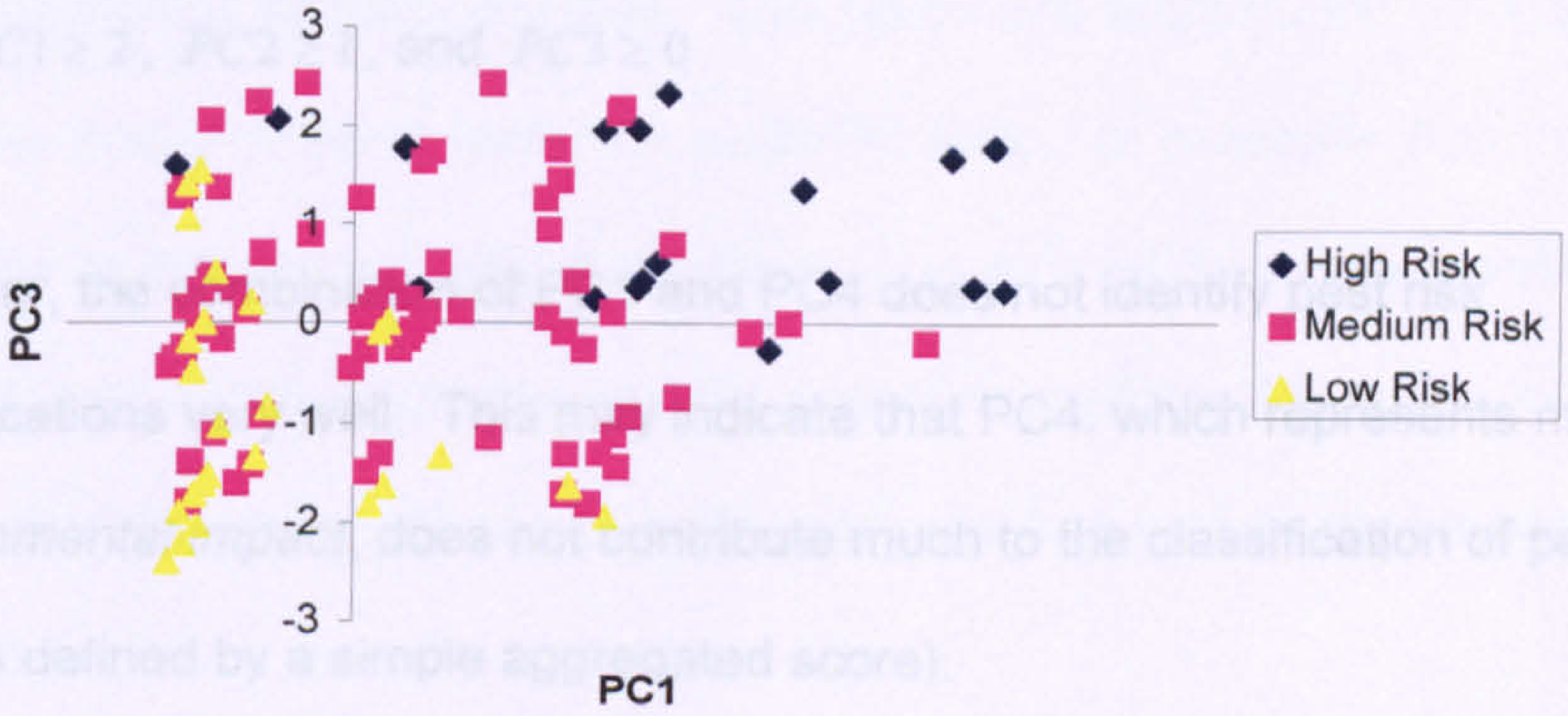


Figure 4.8 All species: PC1 vs. PC3

Figure 4.9 All species: PC1 vs. PC4

These figures suggested that the combination of PC1 and PC2, and the combination of PC1 and PC3, could reasonably indicate pest risk classifications. For example, the majority of the high-risk pests are identified with $PC1 \geq 2$, $PC2 \geq 1$, and $PC3 \geq 0$.

However, the combination of PC1 and PC4 does not identify pest risk classifications very well. This may indicate that PC4, which represents mostly *environmental impact*, does not contribute much to the classification of pest risk (as defined by a simple aggregated score).

This is true by examining the original scores of *environmental impact*, which was given a score of 1 for a large percentage of the pests considered.

It also reconfirmed the criterion for the number of PCs to retain (4.4.2), that PCs explaining less than 10% of the total variance could be eliminated. PC4 (mainly representing *environmental impact*) explained 10.24% of the total variance and therefore can be eliminated.

4.4.3.3 Risk factor weighting, species PCA score, and pest risks

Risk factor weighting

Table 4.15 in section 4.4.3.1 shows the scoring coefficients for each risk factor on each of the four retained PC axes. An overall weighting for each risk factor was calculated as a weighted sum of the coefficients for the factor across the retained PCs. The weighting was simply the proportion of the variance explained by each PC axis. This takes into account the different explanatory power of each factor on each PC axis. For example, *host range* has most of its influence (highest coefficient) on axis PC1(and the coefficient is also positive), *Host range* has smaller negative coefficients on the remained PCs, therefore its overall scoring coefficient/weighting was partly balanced by those negatives.

The weighting for a risk factor is calculated from the scoring coefficients obtained from PCA in Table 4.15. For the convenience of explaining the calculating process, Table 4.16 below substitutes the scoring coefficients with a1 ... g4 and the percentage of variance explained with μ.

Table 4.16 Symbols substituted for scoring coefficients of principal components in Table 4.15

Risk Factors	PC1	PC2	PC3	PC4
Climate suitability	a1	a2	a3	a4
Host Range	b1	b2	b3	b4
Dispersal potential	c1	c2	c3	c4
Economic impact	d1	d2	d3	d4
Environmental impact	e1	e2	e3	e4
Pathway	f1	f2	f3	f4
Trading Partner	g1	g2	g3	g4
Variance Explained	µ1	µ2	µ3	µ4

First of all, the weighting of an original risk factor on a certain PC was calculated, it is the product of the scoring coefficient for the original risk factor (a1 ... g4 in Table 4.16) multiplied by the percentage of variance (µ in Table 4.16) explained by that PC. The overall weighting of an original risk factor was then calculated by summing up the weightings across all retained PCs. Table 4.17 shows this process.

Table 4.17 Calculating process of factor weightings

Risk Factors	**Weighting on PC1**	**Weighting on PC2**	**Weighting on PC3**	**Weighting on PC4**	**Factor weighting**
Climate suitability	a1*μ1	a2*μ2	a3*μ3	a4*μ4	|Sum(a*μ)|
Host Range	b1*μ1	b2*μ2	b3*μ3	b4*μ4	|Sum(a*μ)|
Dispersal potential	c1*μ1	c2*μ2	c3*μ3	c4*μ4	|Sum(a*μ)|
Economic impact	d1*μ1	d2*μ2	d3*μ3	d4*μ4	|Sum(a*μ)|
Environmental impact	e1*μ1	e2*μ2	e3*μ3	e4*μ4	|Sum(a*μ)|
Pathway	f1*μ1	f2*μ2	f3*μ3	f4*μ4	|Sum(a*μ)|
Trading Partner	g1*μ1	g2*μ2	g3*μ3	g4*μ4	|Sum(a*μ)|

Replacing the formulae in Table 4.17 with the PCA results in 4.4.3.1 (numbers in Table 4.15), the following weightings for the original risk factors were obtained (Table 4.18)

Table 4.18 Weightings for original risk factors

Risk Factors	**Weighting on PC1**	**Weighting on PC2**	**Weighting on PC3**	**Weighting on PC4**	**Factor weighting**
Climate suitability	-0.024013	-0.003585	0.1387152	-0.003246	0.10787122
Host Range	0.2871828	-0.013986	-0.022633	-0.007721	0.24284334
Dispersal potential	0.1508315	0.0076408	0.0441637	0.009769	0.21240499
Economic impact	0.1552008	0.0386922	0.0196846	0.0097894	0.22336697
Environmental impact	-0.008919	-0.001733	-0.003574	0.1010893	0.08686302
Pathway	-0.018308	0.1161607	4.467E-05	-0.001802	0.09609532
Trading Partner	-0.00975	0.1142925	-0.00813	-0.001649	0.09476425

The absolute value of the sum in the above calculation is used in this study to represent the factor weighting[16]. It is the magnitude rather the sign that indicates the importance of the original variables on the PC axes.

The weightings of the risk factors are shown in Table 4.19 in descending order, while the standardised weightings are weightings adjusted so that the sum of the factor weightings equals 1.

Table 4.19 Risk factor weightings in PCA scores

Risk factor	**Weighting**	**Standardised weighting**
Host Range	0.24284334	0.23
Economic impact	0.22336697	0.21
Dispersal potential	0.21240499	0.20
Climate suitability	0.10787122	0.10
Pathway	0.09609532	0.09
Trading Partner	0.09476425	0.09
Environmental impact	0.08686302	0.08

Table 4.19 suggests that *host range*, *economic impacts,* and *dispersal potential* are the three more important risk factors amongst the seven.

Species PCA score

Based on the original score for each risk factor and the weighting obtained above, a species PCA score can be obtained. A PCA score for a species is

---

[16] There was no negative value obtained in this table.

defined as the summation of the weighted scores for each original risk factor over the retained PCs. A PCA score can be expressed as following:

$$Score_{PCA} = \sum Weighting * FactorScore$$

Using the weightings obtained above (Table 4.18), a species PCA can be derived as follows:

$$\begin{aligned} Score_{PCA} = {} & 0.1079 * ClimateScore + 0.2428 * HostScore + 0.2124 * DispersalScore \\ & + 0.2234 * EconomicScore + 0.0869 * EnvironmentalScore + 0.0961 * PathwayScore \\ & + 0.0948 * TradingScore \end{aligned}$$

PCA scores for all species in the original dataset were calculated. Figure 4.10 shows the histogram of PCA scores for 252 pest species. These species were also categorised by a low/ medium/ high risk classification based on a simple aggregation of the original variables used in Black and Abdallah (1997).

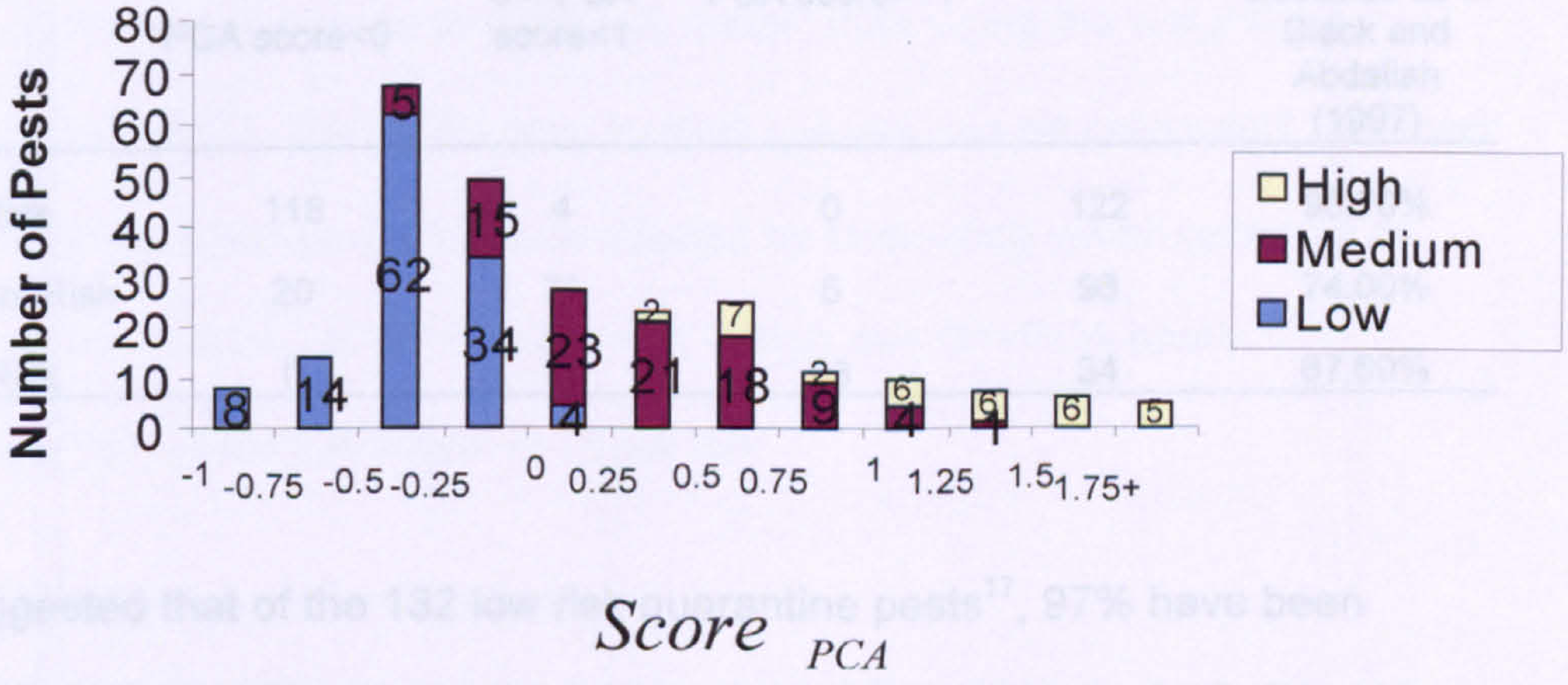


Figure 4.10 Histogram of PCA score

Figure 4.10 shows that the majority of high-risk quarantine pests (as identified by simple averaging) have PCA scores of 1 or beyond, the majority of low-risk quarantine pests identified by simple average have negative PCA scores, while the majority of medium-risk quarantine pests (identified by averaging) have PCA scores between 0 and 1.

Setting the cut-off PCA scores for low risk and high risk at 0 and 1 could reasonably identify low, medium, and high-risk pests, as classified by score aggregation by Black and Abdallah (1997). The number of quarantine pests having been identified by the cut-off PCA scores of 0 and 1 are shown in Table 4.20.

Table 4.20 The number of quarantine pests

	Number of PCA score<0	Number of 0<=PCA score<1	Number of PCA score>=1	Total	% Being classified as in Black and Abdallah (1997)
Low Risk	118	4	0	122	96.70%
Medium Risk	20	71	5	96	74.00%
High Risk	0	11	23	34	67.60%

It suggested that of the 132 low risk quarantine pests[17], 97% have been identified by negative PCA score. Of the 96 medium risk quarantine pests, 74% have been identified by positive PCA score yet less than 1. Of the 34 high-risk quarantine pests, 68% have been identified by PCA score greater than 1.

---

[17] As classified by Black and Abdallah (1997).

The above analysis was not intended to compare which method gave a better result of classification of pest risk. Rather, it indicated that in a large data set, there were some boundary cases. For example, there seem to be two 'boundary' PCA scores in Figure 4.10: -0.25 (between low and med) and 0.5 (between med and high). Between –0.25 and the cut-off PCA score 0, there are mixed cases of low and medium risk pests; between 0.5 and the cut-off PCA score 1, there are a mixed cases of medium and high risk pests. So ideally, the next step would have been to ask the assessors to revisit these 'boundary' cases, and ask them how likely they would modify the original assessments, i.e. whether these boundary cases in fact pose similar risks, some of which originally have been classified high and some medium.

A further application of PCA would be it is possible to conclude the overall level of risk of a species without combining the scores. It works like this: first of all, a score is assigned to each risk factor; then using the weightings derived from PCA, the PCA score for that species can be calculated; the level of risk for the species could be concluded by examining which category its PCA score fits into, i.e. PCA score >=1 - High risk; 0<=PCA score <1 - Medium risk; and PCA score <1 - Low risk.

4.4.3.4 Principal Components Analysis of different taxonomic groups

In this section, PCA was applied to different taxonomic groups. The scoring coefficients of PCs for different taxonomic groups are reported in Table 4.21 to Table 4.25.

Bacteria

Table 4.21 shows the scoring coefficients of the PCs for bacteria.

Table 4.21 Scoring coefficients of principal components for bacteria

	PC1	PC2	PC3
Climate suitability	0.0159	**0.9279**	0.0247
Host Range	-0.03	0.0261	**0.9322**
Dispersal potential	**0.4879**	-0.2787	0.0548
Economic impact	**0.5972**	0.0543	-0.1909
Pathway	**0.5247**	0.2262	0.0127
Trading Partner	**0.359**	-0.0812	**0.3013**
Eigenvalue	3.0759	1.1268	0.8402
Variance explained	51.27%	18.78%	14.00%

In this case three components explained 84% of the total variation. The first component represents mainly *economic impact, pathway, dispersal potential, and trading partner*, explaining 51.3% of the overall variance; the second component represents mainly *climate suitability*, explaining 18.8% of the overall variance; the third component represents mainly *host range*, explaining 14% of the overall variance.

It is worth noting that the *environmental impact* does not come into the PCs, indicating *environmental impact* does not have any explanatory power in explaining the overall variance of bacteria risk factors. This is true by examining the original scores of this factor, which reveals that score 1 (low) was given to *environmental impact* to all the bacteria. It is safe to suggest that *environmental impact* is a redundant factor in assessing the risk of bacteria.

Fungi

Table 4.22 shows the scoring coefficients of PCs for fungi.

Table 4.22 Scoring coefficients of principal components for fungi

Risk Factors	PC1	PC2	PC3	PC4
Climate suitability	-0.0447	0.0787	**0.8663**	0.1152
Host Range	-0.0398	**0.8771**	0.0706	-0.0873
Dispersal potential	-0.0161	-0.1301	0.1355	**0.8353**
Economic impact	0.1722	0.3208	-0.1823	**0.4469**
Environmental impact	**-0.5138**	0.1507	**-0.4045**	0.2669
Pathway	**0.6545**	-0.1189	-0.1642	0.1008
Trading Partner	**0.5236**	0.2604	-0.0483	0.0178
Eigenvalue	1.8727	1.204	1.1991	1.159
Variance Explained	26.75%	17.20%	17.13%	16.56%

Four PCs explained 77.6% of the total variance for fungi. The first component represents mainly *pathway*, *trading partner* and *environmental impact*, explaining 26.8% of overall risk factor variance. A high value for PC1 occurs when high scores for pathway and trading partner are combined with a low score for environmental impact. The second component represents mainly *host range*, explaining 17.2% of total risk factor variance; the third component represents mainly *climate suitability* and *environmental impacts*, explaining 17.1% of overall risk factor variance; the fourth component represents mainly *dispersal potential and economic impacts*, explaining 16.6% of overall risk factor variance.

Insects

Table 4.23 shows the scoring coefficients of PCs for insects.

Table 4.23 Scoring coefficients of principal components for insects

Risk factor	**PC1**	**PC2**	**PC3**	**PC4**
Climate suitability	**0.5083**	**-0.3007**	**0.3195**	0.1827
Host Range	**0.6937**	0.0932	-0.3335	-0.0552
Dispersal potential	-0.0654	0.0501	**0.8434**	-0.0416
Economic impact	**0.4849**	0.0979	0.2624	-0.0961
Environmental impact	-0.0086	0.0215	-0.0221	**0.9739**
Pathway	0.1313	**0.6503**	0.0107	0.0114
Trading Partner	-0.0604	**0.6823**	0.0772	0.0633
Eigenvalue	1.8665	1.7243	1.2218	1.0231
Variance Explained	26.66%	24.63%	17.45%	14.62%

Four PCs explained 77.6% of the total variance for insects. The first component represents mainly *climate suitability, host range, and economic impact*, explaining 26.7% of overall risk factor variance; the second component represents mainly *pathway* and *trading partner*, explaining 24.6%

of total risk factor variance; the third component represents mainly *dispersal potential* and *climate suitability*, explaining 17.5%% of overall risk factor variance; the fourth component represents *environmental impact*, explaining 14.6% of overall risk factor variance.

Nematodes

Table 4.24 shows the scoring coefficients of PCs for nematodes.

Table 4.24 Scoring coefficients of principal components for nematodes

Risk factor	**PC1**	**PC2**	**PC3**	**PC4**
Climate suitability	-0.0064	-0.0067	**0.946**	-0.0232
Host Range	-0.1443	**0.7123**	0.0762	0.1275
Dispersal potential	0.1019	**0.5948**	-0.2178	-0.1663
Economic impact	0.2447	0.3651	0.2251	-0.0123
Environmental impact	0.0122	0.0126	-0.0206	**0.9766**
Pathway	**0.6554**	0.0048	-0.0264	0.0396
Trading Partner	**0.6922**	-0.073	0.0025	0.0005
Eigenvalue	2.239	2.2298	1.0943	1.0363
Variance Explained	31.99%	31.85%	15.63%	14.80%

Again, four PCs explained 94.3% of the total variation for nematodes. The first component represents mainly *pathway* and *trading partner*, explaining 31.99% of overall risk factor variance; the second component represents mainly *host range* and *dispersal potential*, explaining 31.8% of total risk factor variance; the third component represents mainly *climate suitability*, explaining 15.63%

of overall risk factor variance; the fourth component represents mainly *environmental impact*, explaining 14.8% of overall risk factor variance.

The results of *climate suitability* and *environmental impact* for nematodes are rather similar to that of all species (4.4.3.2) in that PC3 and 4 are largely influenced by a single original risk factor. This might also suggest that *climate suitability* and *environmental impact* are largely either uncorrelated with or unaffected by other risk factors. Again, by examining the biology of nematodes, this makes sense as many nematodes are parasites to the root and climate may not play a big role in it. Also, its effect on environment may not as obvious as others such as insects.

Viruses

Table 4.25 shows the scoring coefficients of PCs for viruses.

Table 4.25 Scoring coefficients of principal components for virus

Risk Factors	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
Climate suitability	0.0065	0.0131	0.0207	**0.9841**	0.0137
Host Range	-0.0235	-0.0388	**0.9423**	0.0245	-0.0172
Dispersal potential	-0.0041	-0.0093	-0.0146	0.0135	**0.9878**
Economic impact	0.1557	**0.6038**	0.2444	-0.0444	0.0137
Environmental impact	-0.0939	**0.7855**	-0.1671	0.0457	-0.0246
Pathway	**0.7123**	-0.1034	-0.1273	0.1164	-0.1052
Trading Partner	**0.6775**	0.0772	0.0868	-0.1144	0.1094
Eigenvalue	1.6618	1.4444	1.0885	1.0158	1.0078
Variance explained	23.74%	20.63%	15.55%	14.51%	14.40%

Here, five PCs were needed to explain 88.8% of the total variance for viruses. The first PC represents mainly *pathway* and *trading partner*, explaining 23.74% of overall risk factor variance; the second PC represents mainly *economic impact* and *environmental impact*, explaining 20.63% of total risk factor variance; the third PC represents mainly *host range*, explaining 15.55% of overall risk factor variance; the fourth PC represents mainly *climate suitability*, explaining 14.51% of overall risk factor variance, while the last PC represents mainly *dispersal potential*, explaining 14.4% total risk factor variance.

Again, PC3, 4 and 5 mainly represent a single risk factor: *host range, climate suitability* and *environmental impact*, respectively. This can be explained by the biology of virus as well. Many viruses are relatively host specific so host range does not vary that much among species. Climate condition does not affecting virus directly, it may affect some vectors and subsequently affects the viruses. Some viruses are latent and perhaps for this reason, the environmental impact is not that obvious.

4.4.3.5 Risk factor weightings for various taxonomic groups

In this section, the standardised risk factor weightings for different taxonomic groups have been calculated with the same method discussed in section 4.4.3.3.

The weightings for the seven risk factors for different taxonomic groups are shown in Table 4.26, while the three pest risk factors with the highest weightings are in bold.

As in 4.4.3.3, the absolute value of coefficient was used here to represent the factor weighting, because the purpose of applying PCA in this study was to find out the weighting/relative importance of each risk factor. It is considered that it is the magnitude of the eigenvector rather than the sign that indicates the importance/weighting of a risk factor (original variable).

Table 4.26 Standardised risk factor weightings for taxonomic groups

Risk Factor	Bacteria	Fungi	Insects	Nematodes	Virus
Climate suitability	0.14	**0.16**	0.13	0.11	**0.14**
Host Range	0.09	0.13	0.13	**0.16**	0.13
Dispersal potential	0.15	0.13	0.12	0.13	0.13
Economic impact	**0.22**	**0.14**	**0.17**	**0.18**	**0.19**
Environmental impact	0	0.13[18]	0.12	0.11	0.11
Pathway	**0.24**	0.14	**0.18**	**0.16**	0.12
Trading Partner	**0.16**	**0.17**	**0.15**	0.15	**0.18**

The result suggests that there are no apparent similarities in terms of risk factors weightings across all the pest taxonomies. For example, *economic impact, pathway* and *trading partner* are more important in assessing the risk for bacteria. While for virus, *economic impact, trading partner* and *climate suitability* are more important. These seem to suggest that it is difficult to find a general pattern of weighting that suits all the pest categories.

This result complements the result from correlation analysis, which also suggests that it is not plausible to develop a generalised scheme for different taxonomic groups (see 4.2.4).

However, if the top three important factors are considered in each taxonomic group, *economic impact* emerges as an important factor across all groups. The result also matches that of the correlation analysis (see 4.2.4).

---

[18] The absolute value is used here instead of the original negative number, as explained in the calculation method, Section 4.4.3.3.

Trading partner is anther important factors for all groups except nematodes. This can be explained by that nematodes are normally associated with root and soil, sometimes seed, and in the Tanzanian data, such trades were minor.

Climate suitability stands out as having more influence for fungi and viruses than other groups. This factor may affect these groups more through their vectors.

4.4.4 Conclusions

In this Section, Principal Components Analysis was used to deal with the multicollinearity among the risk factors and to place different weights on different risk factors for the overall risk assessment.

The results from the PCA analysis do suggest that weightings can be derived for individual risk factors by applying statistical techniques to historical data of pest introductions and invasions, previous PRA cases, or expert opinion.

Historical data and previous case studies do not necessarily apply to new situations; however, these can provide at least a starting point for new pests.

By putting different weights to different risk factors, the more important risk factors can be identified, and the less important risk factors, i.e. risk factors contribute less to the overall risk assessment, are filtered out.

Based on the weightings thus derived, a quick summary scheme could be developed, which will give a quick and preliminary idea of pest risk rating.

A further application of PCA would be it is possible to conclude the overall level of risk of a species without combining the individual risk scores.

However, the results shown that different taxonomic groups have different weighting patterns, it is therefore not plausible to develop a generic weighting pattern for different pest categories.

4.5 USING EMPIRICALLY-DERIVED RULES TO EXPRESS RISK: AN APPLICATION OF GENETIC ALGORITHMS

4.5.1 Introduction

In the previous sections, the relative importance, i.e. weighting, of different risk factors was discussed and it was concluded that not all the risk factors were equally important when assessing pest risks. PCA was used to reduce the number of risk factors and to incorporate risk factor weightings into the risk assessment. This Section is devoted to a technique using automated machine learning to help identify patterns of risk in different situations that may help the risk analysts in their tasks. However, each PRA case has elements of novelty in it; the study presented here is intended to find out some approximate rules with available data.

Machine learning uses algorithms that can aid in the discovery of rules and patterns in sets of data and can self-adapt as more data becomes available. Machine learning has already been widely used in risk analysis, such as financial risk analysis, and medical risk analysis, it has also been widely used in agricultural data, such as soybean disease diagnosis (Michalski *et al.*, 1982).

However, machine-learning approaches have not been applied to PRA[19]. In the context of PRA, machine-learning approaches are desirable in cases where pest risks cannot be defined well, except by examples. In such cases, humans can specify the input/output pairs, but the relationship between the inputs and outputs are unknown. The machine learning approaches can automatically adjust the internal structure of its data analysis process to generate approximate results for the given tasks.

Another advantage of machine learning techniques is that they can easily adapt to deal with new situations or situations not previously encountered. This is important in pest risk analysis in that new quarantine pests emerge frequently, hence it is crucial that an approach can be revised to incorporate new knowledge and generate new hypotheses.

Finally, the machine learning approaches can be used to extract important knowledge, relationships and correlations, which may have been hidden in the pest risk analysis data.

By using machine-learning techniques, an exploration was made of what circumstances the pest risk would be assessed 'high' or 'low' according to the observation and evaluation of individual risk factors. In other words, the techniques were used to find a rule (or rules) for pest risk assessment, i.e. whether pest risk can be determined by some particular function of the component risk factors. In addition, an attempt was made to see whether this

---

[19] It is meant at the time of this study (2002). Neural networks were since applied to evaluate establishment potential, see 2.4.2.

enabled a reduction in the number of factors to be considered in the pest risk analysis process.

In this section, the machine-learning software, BEAGLE (Forsyth, 1981), has been employed to analyse the data from the initial pest risk assessments for Tanzania[20] to find out the rules for determining a high-risk situation. This section is organised as follows: Section 4.5.2 introduces the concepts of Machine Learning and Genetic Algorithms; Section 4.5.3 describes the data used for the machine learning process and the machine-learning programme to implement machine learning in this chapter – BEAGLE; Section 4.5.4 reports the results from the machine learning process, while Section 4.5.5 concludes.

4.5.2 Concepts of machine learning, genetic algorithms, and BEAGLE

4.5.2.1 Machine learning

Machine learning refers to a system capable of the autonomous acquisition and integration of knowledge, which can aid in the discovery of rules and patterns in sets of data. This capacity to learn from experience, analytical observation, and other means, results in a system that can continuously self-improve and thereby offer increased efficiency and effectiveness. Mitchell (1997) formally defined Machine learning as: "A program learns from experience E with respect to some class of tasks T and performance measure

---

[20] See Chapter 4 correlation for data description.

P, of its performance at task T, as measured by P, improves with experience E."

As a broad subfield of artificial intelligence, machine learning is concerned with the development of algorithms and techniques that allow computers to "learn". At a general level, there are two types of learning: inductive, and deductive. Inductive machine learning approaches create computer programs by extracting rules and patterns out of (usually very large) data sets. Deductive machine learning methods create additional rules from a set of known facts and rules.

The focus of this study was inductive learning. In inductive learning, the machine learning system is given a set of training and testing examples. It learns from the training examples and defines the hypothesis for them. Then, the learning system uses the test set to evaluate the rules that come out of the training set.

There are three main types of inductive learning:

(1). Supervised learning in which both the inputs and the outputs of an example can be observed;

(2). Reinforcement learning where the learning agent is given an evaluation of its action but not told the correct action; and

(3). Unsupervisized learning where the learning agent can only observe the inputs, but has no information about the output.

The type of machine learning used in this chapter is supervised inductive learning, as both the inputs (risk ratings for individual risk factors) and outputs (classfication of pest risk) are observed.

4.5.2.2 Genetic Algorithms

The Genetic Algorithm (GA) is a type of Evolutionary Computation devised by John Holland (Holland, 1975). It is a model of machine learning based on a genetic/evolutionary metaphor of survival of the fittest and evolution in nature, and derives solutions (rules and patterns) in an evolutionary manner.

The main idea of GA is to maintain a population of candidate solutions, represented by a set of character strings that are analogous to the chromosomes in DNA. The individuals in the population then go through a process of simulated "evolution", and evolve through competition to the best set of solutions by controlling the variation to improve the performance of the learning programme. This population of candidate solutions undergoes crossover and mutation processes to adapt to the new environment (i.e. the set of data used to train the algorithm), and their ultimate goal is to become the fittest (i.e. offer the best explanation of the data with which it is presented).

When a genetic algorithm is implemented, it starts with a population of randomly generated solution candidates; then continues with the following cycle: evaluate the fitness of all of the individuals in the population in terms of an objective function; create a new population by performing operations such as crossover, fitness proportionate reproduction, and mutation on the individuals whose fitness has just been measured; and then discard the old population and iterate using the new population. One iteration of this process is referred to as a generation. The first generation (generation 0) of this process operates on a population of randomly generated individuals. From there on, the genetic operations, in concert with the fitness measure, operate to improve the population.

This evolution process includes four basic components:

- Selection: selecting individuals for reproduction according to their fitness (objective function value);
- Crossover: merging the genetic information of two individuals; if the individuals are chosen properly, two good parents produce some good children;
- Mutation: in real evolution, the genetic material can by changed randomly by erroneous reproduction or other deformations of genes. In genetic algorithms, mutation is realized as a random deformation of the individuals with a certain probability;
- Sampling: creating a new population from the previous one and its offspring.

There are three types of genetic algorithm programming systems: application oriented systems, algorithm oriented systems, and GA programming tool kits (Filho *et al.*, 1994). While algorithm oriented systems and GA Tool Kits are programming systems or programming blocks designed for system developers and supporting specific genetic algorithm programming, the application oriented systems are designed for users wishing to utilize GA in specific applications, without having to acquire detailed knowledge of the workings of genetic algorithms (See Fliho *et al.*, 1994 for more details).

PC/Beagle and XpertRule Gensys, are two application oriented GA programming systems. They are also rule-finders using GA to generate new rules to expand the knowledge base in the applied areas. While XpertRule Gensys is targeted to solve scheduling and design applications using optimization techniques, PC/Beagle applies machine-learning techniques to create a set of new rules for classifying examples. PC/Beagle was used in this study, further details of which were discussed in the following section (4.5.2.3).

4.5.2.3 BEAGLE

In this study, the machine learning process is implemented by a machine learning software, BEAGLE, which stands for Biology Evolutionary Algorithm Generating Logical Expression. BEAGLE was developed by Richard Forsyth in 1986 and was one of the first systems to use genetic algorithms for deriving rules from data. It is a supervised machine learning system incorporating several advanced ideas from the field of artificial intelligence. It has been successfully used to produce the knowledge base for a number of different applications, e.g. a system to classify glass fragment evidence in forensic science (Evett and Spiehler, 1987).

BEAGLE is in fact a rule finder system, which examines a set of examples and uses GA to create a set of rules for classifying these examples and other examples of the same types. BEAGLE not only tests hypotheses as do conventional statistical packages, but it also proposes the hypotheses to be tested. Its distinctive feature is the use of an evolutionary induction strategy to advise new discrimination rules.

4.5.3 Method

In this study BEAGLE has been used to conduct automatic machine learning on the Tanzanian data. It is applied to the entire potential quarantine pests group (252 species), as well as different taxonomic groups, i.e. fungi, bacteria, virus and virus-like organisms, insects, phytoplasma, nematodes and mites.

Data for each group were randomly split into a training set and a test set using a 50%: 50% split each time.

Each pest species has eight attributes: attribute 1 = score for *climate suitability*; attribute 2 = score for *host range*; attribute 3 = score for *dispersal potential*; attribute 4 = score for *economic impact*; attribute 5 = score for *environmental impact*; attribute 6 = score for *pathway*; attribute 7=score for *trading partners*. The possible domain values for the first seven attributes (all risk factors) are: 1 = low risk, 2 = medium risk, 3 = high risk. Attribute 8 is the goal/target attribute, specifying whether the pest was of high risk (3), medium risk (2) or low risk (1) based on pest risk assessment in Black and Abdallah (1997).

The study aims to derive rules specifying under what conditions of the other attributes, the goal attribute would assume a given value, in this study, high risk. The targets are set as follows:

Attribute 8 =3

The objective therefore was to derive rules of the form:

IF conditions THEN overall risk = high

For taxa containing high risk species, targets were set as: Risk = High, but for taxonomic groups without high-risk species, such as viruses and virus-like organisms, phytoplasma and mites groups, the targets were set as Risk=low.

Six runs of BEAGLE were conducted for each group.

For each group, 200 generations were used per run to derive rules.

After a set of rules generated, BEAGLE would then evaluate the crude success rate of the rule or rules using a fitness function.

The fitness function used in this study to evaluate the quality of each set of rules is based on the following four different types of results that can occur for a prediction:

- True positive (TP) – the rule predicts that the overall risk is high and it is.

- False positive (FP) – the rule predicts that the overall risk is high and it is not.
- True negative (TN) – the rule predicts that the overall risk is low and it is;
- False negative (FN) – the rule predicts that the overall risk is low and it is not.

The success rate was also calculated to evaluate the performance of the rules generated.

For each dataset, the machine learning process has been repeated six times to examine rule consistency, the variation in the success rate and to select the fittest and most information efficient rule combinations.

The rule combinations with the best predicting success rates were selected as the fittest rules; while the rule combinations involving the least risk factors (requiring the least information) yet with reasonable success rate were selected as the most information efficient rules[21].

4.5.4 Empirical results

4.5.4.1 Rules for all quarantine pests[22]

The machine learning process has been firstly conducted for all the 252 quarantine pests, with overall risk being calculated by summation of all risk

---

[21] All the discussions are made in the context of potential quarantine pests for Tanzania.

[22] All quarantine pests here refer to the potential quarantine pest evaluated by Black and Abdallah (1997) for Tanzania.

scores, i.e. Beagle was being used to see if some particular combination of risk factors (expressed as a rule or rules) was associated with a high score sum. The target was set as: overall risk = high. The results are shown in Table 4.27.

Table 4.27 Rules for all species with overall risk rating as target expression: (overall risk= high)

Rule found	Rule interpretation	Crude success rate
1st RUN: two rules found		96.69%
1. ((ECONOMIC + HOST) > 4.0000)	1. Scores for Economic and Host are both High, or one is High, the other is Medium.	
2. ((PATHWAY + DISPERSAL) > 4.0000) and/or (CLIMATE = 3)	2. Either score for Climate is High or scores for Pathway and Dispersal are both High or one is High, the other is Medium.	
2nd RUN: one rule found		93.70%
1. ((DISPERSAL + ECONOMIC) > 4.0000) and (PATHWAY >1.0000) and (HOST = 3)	1. Score for Host is High and scores for Economic and Dispersal are both High, or one is High, the other is Medium. Score for Pathway is medium or high.	
3rd RUN: two rules found		95.83%
1. ((CLIMATE+ DISPERSAL)>4 and (ECONOMIC + HOST) > 4.5000))	1. Scores for Climate and Dispersal are both High, or one is High, the other is Medium, and scores for Economic and Host are both High, or one High, the other is Medium.	
2. ((TRADING + PATHWAY) > 3.0000)	2. At least one of the scores for Trading and Pathway is High, or both High, or both Medium.	
4th RUN: three rules found		97.62%
1. ((ECONOMIC + HOST) > 4.4000)	1. Scores for Economic and Host are both high, or one is High, the other is Medium.	
2. ((CLIMATE + PATHWAY) > 3.5000)	2. At least one of the scores for Climate and Pathway is High, or both High, or both Medium.	
3. ((TRADING + DISPERSAL) > 4.0000)	3. Scores for Trading and Dispersal are both High, or one is High, the other is Medium.	
5th RUN: three rules found		100.00%
1. ((PATHWAY + CLIMATE) > 3.5000) and/or (HOST=3)	1. Score for Host is High and scores for Pathway and Climate are both High, or one is High, the other is Medium.	
2. ((ECONOMIC + DISPERSAL) > 4.4000)	2. Scores for Economic and Dispersal are both High, or one is High, the other is Medium.	
3. (TRADING > 2.2500)	3. Score for Trading is High	
6th RUN: two rules found		96.83%
1. ((HOST + ECONOMIC) > 4.0000)	1. Scores for Economic and Host are both high, or one is High, the other is Medium.	
2. ((PATHWAY + CLIMATE) > 3.2553) and ((ENVIRONMENTAL + DISPERSAL) > 3.5000)	2. At least one of the scores for Climate and Pathway is High, or both High, or both Medium and at least one of the scores for Environmental and Dispersal is High, or both High, or both Medium	
Average Success Rate		96.78%

Column 1 shows the rules found through the machine learning, column 2 gives the interpretation of the rules. Column 3 shows the success rate of the rule combination, giving the probability of the target expression being true when all the rules are true. For example, in the first run, two rules have been found:

Rule 1: ((ECONOMIC + HOST) > 4.0000),

suggesting that high-risk pests seem to have medium to high-risk ratings for *economic impacts* and *host range*;

and

Rule 2: (((PATHWAY + DISPERSAL) > 4.0000) > and/or (CLIMATE = 3)

indicating that high-risk pests tend to have either medium to high-risk ratings for *pathway* and *dispersal potential* and/or higher than average risk ratings for *climate suitability.*

For pest risk to be classified as high, both these rules have to be true. Success rate of the first run is 96.69%, indicating that when both rules are true, the probability of the pest being high risk is 96.69%.

The table above shows the success rates of the six runs of BEAGLE ranged between 93.7% and 100%. The average crude success rate was 96.78%. While looking in more detail, the results also indicated that not all the risk factors were used in the rules found. For example, amongst the two rules found in the first run, only five risk factors out of the total seven were used in the two rules, while in the 5th run six risk factors were used in the three rules found.

The required scores of risk factors for each of the six rule sets are presented in Table 4.28. In other words, Table 4.28 gives the scores of each risk factor required for each rule set to identify high risk pests.

Table 4.28 Possible values for risk factors in each run

	1st Run	2nd Run	3rd Run	4th Run	5th Run	6th Run
Climate	3		3, 2	3, 2, 1	3, 2, 1	3, 2, 1
Dispersal	3, 2	3, 2	3, 2	3, 2	3, 2	3, 2, 1
Economic	3, 2	3, 2	3, 2	3, 2	3, 2	3, 2
Environmental						3, 2, 1
Host	3, 2	3	3, 2	3, 2	3	3, 2
Pathway	3, 2	3, 2	3, 2, 1	3, 2, 1	3, 2, 1	3, 2, 1
Trading			3, 2, 1	3, 2	3	
Success Rate	96.70%	93.70%	95.80%	97.60%	100%	96.80%

Note: 3: High Risk; 2: Medium Risk; 1: Low Risk; Blank means that the risk factor didn't not appear in the rules found. The possible values are not of course independent, so for example if in Run 3, if trading equals 1, pathway must equal 3 (see Table 4.27)

It can be seen that although the rule sets found in the six runs were different from each other, four risk factors: *economic impact*, *host range*, *dispersal potential*, and *pathway*, appeared in all the six runs, and assumed similar risk ratings. On the other hand, *environmental impact* appeared only once in rule set found in the 6th run; *trading partners* appeared in three of the six runs; and *climate suitability* appeared in four of the six runs.

This result allied to that of the PCA, both suggested that *environmental impact* does not contribute much to the PRA.

While the initial risk assessment by Black and Abdallah (1997) was based on the sum of scores for all of the seven risk factors, the absence of the *environmental impact* in the rule sets found in five of the six runs indicated that the inclusion of the *environmental factor* in the risk assessment was largely redundant.

Fittest rule set

The rule set with the best success rate was found in the 5th RUN, with a success rate of 100%. In this run, a set of three rules were found as follows:

Rule 1: (((PATHWAY + CLIMATE) > 3.5000) and/or (HOST=3)), indicating that high-risk pests tend to have high score for *host range*, while *pathway* and *climate* tend to have medium or high risk ratings;

Rule 2: ((ECONOMIC + DISPERSAL) > 4.4000), suggesting that scores of high risk pests for *economic impact* and *dispersal potential* are both High, or one is High, the other is Medium;

Rule 3. (TRADING > 2.2500), suggesting that high-risk pests tend to have high-risk ratings for *trading partners.*

The most efficient rule set

The rule set with the least risk factors was found in the second run, with only four risk factors being used, yet with a satisfactory 93.7% success rate. The rule found in the second run was as follow:
(((DISPERSAL + ECONOMIC) > 4.0000) and ((PATHWAY > 1.0000) and (HOST = 3.0000))), suggesting that high risk pests tend to have high risk rating for *host range*, and medium to high risk ratings for *dispersal potential, economic impacts,* and *pathway.*

The results from this study showed a high degree of consistency with that from PCA (see 4.4.3.3). Risk factors that yield high weightings from PCA also appear to be the predictive factors in Beagle rules. For example, *host range* had the highest weighting in the PCA study, it is also the most important factor in this study (HOST = 3.0000). Other risk factors appeared in the most efficient rules were also had higher weightings in the PCA study, i.e. *economic impact, dispersal potential.*

Risk factors that tended to have low weightings in PCA also appeared less in Beagle rules. For example, *environmental impact* is such factor.

That two quite different techniques yield similar results lends credence to the conclusions about the relative importance of risk factors

The application of the rules

Both rule sets have a very good success rate. Comparing the most efficient rule set with the fittest one, this set didn't include *climate suitability* and *trading partner*, suggesting that removing these two risk factors reduced the success rate by less than 7%. Neither of the two rule sets included *environmental impacts*.

Both rule sets can be used to predict under what circumstances (risk factor scores) the risk of a species being introduced into a new area would be high. While the predication of the first rule set is more accurate, the second rule set uses less information yet maintains a reasonable success rate.

It was considered whether risk assessment could be simplified by applying these rules, especially when resources and time are limited. These rules could enable the simplification of risk assessment by only considering the individual risk scores. It was simple yet relatively reliable, as the rules found here were based on relatively large number of real RA cases.

4.5.4.2 Rules for various taxonomic groups

Machine learning processes have also been performed on different taxonomic groups in order to find rules for those particular taxonomic groups: bacteria, viruses or virus-like organisms, insects, nematodes, phytoplasma, and mites[23]. The results revealed that different factors seemed to be of different importance for different taxa, which may due to the differences of biology and the nature of trading pathways associated with different groups.

However, no biological meaningful rules were found for the phytoplasma or mites groups, as they did not have enough observations to enable meaningful machine learning process. The machine learning results for the rest of the taxonomic groups are shown in Table 4.29.

For each run, only the rule with the highest success rate is shown in the table. The most efficient rule for each taxonomic group was selected from the six runs. This was taken to be the one that had the least risk factors and the highest success rate. If the number of factors and success rate were similar, the rule appearing most frequently in the six runs was chosen.

---

[23] Although the analysis was applied to mites, of which there were only 4 species, no meaningful rules were expected.

Table 4.29 Rules for different taxonomic group

		Fungi	Bacteria	Viruses and virus-like organisms	Insects	Nematodes
		Target: Risk=High	Target: Risk=High)	Target: Risk=Low)	Target: Risk=High	Target: Risk=High
1st Run	Rule	((PATHWAY+DISPERSAL)>5.00)	((ECONOMIC +CLIMATE)>4.0000)	(ECONOMIC<= (HOST<=2.0000))	((ECONOMIC+TRADING)>3.5)	(ECONOMIC >1.7500)
	Success Rate	89.74%	66.67%	79.40%	83.33%	100.00%
2nd Run	Rule	((TRADING+DISPERSAL)>5.500)	((ECONOMIC+CLIMATE)>4.00)	(ECONOMIC=((HOST+CLIMATE) <3.2032))	(4.0000 <= (PATHWAY+CLIMATE))	(1.0000 <ECONOMIC)
	Success Rate	89.74%	100.00%	79.40%	81.08%	100.00%
3rd Run	Rule	((DISPERSAL+PATHWAY)>5.00)	((ECONOMIC+HOST)>4.5000)	(3.0000 >= (CLIMATE +HOST))	((CLIMATE+PATHWAY)>3.500)	(ECONOMIC>ENVIRONMENTAL)
	Success Rate	89.74%	83.33%	75%	81.08%	100%
4th Run	Rule	((DISPERSAL+PATHWAY)>5.00)	((CLIMATE+ECONOMIC)>4.50)	(ECONOMIC=(CLIMATE<=2.000))	((PATHWAY+CLIMATE)>3.500)	(ECONOMIC >=1.2500)
	Success Rate	89.74%	100.00%	82.10%	81.08%	100.00%
5th Run	Rule	((TRADING +PATHWAY)>5.0000)	((HOST+ECONOMIC)>4.5000)	((CLIMATE+ECONOMIC)<3.5000)	((PATHWAY+CLIMATE)>3.500)	(ENVIRONMENTAL<ECONOMIC)
	Success Rate	89.74%	83.33%	82.10%	81.08%	100.00%
6th Run	Rule	((DISPERSAL+PATHWAY)>5.00)	((ECONOMIC+CLIMATE)>4.00)	(ECONOMIC+CLIMATE)<3.5000	((CLIMATE+PATHWAY)>3.000)	(ECONOMIC>1.6039)
	Success Rate	89.74%	100.00%	82.10%	81.08%	100.00%
Fittest	Rule	((DISPERSAL+PATHWAY)>5.00)	((ECONOMIC+CLIMATE)>4.0)	(ECONOMIC+CLIMATE)<3.5000	((ECONOMIC+TRADING)>3.5)	(1.0000 <ECONOMIC)
	Success Rate	89.74%	100.00%	82.10%	83.33%	100.00%
Efficient	Rule	((DISPERSAL+PATHWAY)>5.00)	((ECONOMIC+CLIMATE)>4.0)	(ECONOMIC+CLIMATE)<3.5000	((CLIMATE+PATHWAY)>3.5)	(1.0000 <ECONOMIC)
	Success Rate	89.74%	100.00%	82.10%	81.08%	100.00 %

Fungi

For fungi group, three different rules were generated through six runs of BEAGLE with success rate of 89.74%. The fittest and the most information efficient of them all was:

(DISPERSAL+PATHWAY)>5.0000

suggesting that the risk associated with the introduction of a fungal disease into Tanzania would be high if the risk scores for *dispersal potential* and *pathway* were both high, indicating that *dispersal potential* and *pathway* were the two most important risk factors in identifying high risk fungi.

For this group, *dispersal* and *pathway* were two predictive factors. This seems to be true as whether the spores are highly mobile and reproductive, could influence the *dispersal potential*, together with the frequency/volume of the trade, these two factors could determine whether the associated risk is high.

At first glance, the result seems different from that of correlation analysis (4.2.2.1), which indicated that the overall risk was affected by *economic impact* the most, followed by *trading partners, host range, pathway, dispersal potential* and *climate suitability.* However, as discussed in section 4.2, *economic impact* can be represented by other factors; *trading partner* is highly correlated with *pathway*; *host range* and *climate suitability* may be represented by *dispersal potential.* If these relationships were taken into account, the result of *dispersal* and *pathway* being the two predictive factors can be explained. This also indicated that BEAGLE worked well by extracting hidden information from a dataset.

Bacteria

For Bacteria group, two rules were generated from 6 runs of BEAGLE. The fittest and the most information efficient rule was:

(ECONOMIC+CLIMATE)>4.0000

suggesting that the risk associated with the introduction of a bacteria into Tanzania would be high if at least one of the risk scores for *economic effects* and *climate* was high, and risk score for the other was at least medium. The success rate for this rule was 100%, indicating that these two risk factors were critical to the risk classification.

This could be interpreted as this: the risk would be high if the bacterium causes severe damage to yield/value and loss of market, or it is present in both tropical and temperate climates.

Viruses and virus like diseases

As there are no high-risk species in virus and virus like organisms according to Black and Abdallah's (1997) risk assessment, the target expression was set as Risk= low. There were five rules generated from six runs of machine learning, two of them had very similar meaning. The fittest and most information efficient rule was:

(ECONOMIC+CLIMATE)<3.5000

with a success rate of 82.1%. This suggested that the risk associated with the introduction of virus and virus like organisms into Tanzania would at most be low if scores for both *economic impact* and *climate* are low, or one is low, the other is medium.

This result is also slightly differed from that of correlation analysis (4.2.2.2), which indicated that the overall risk was affected by *economic impact* and *trading partner* the most, followed by *climate suitability and environmental impact*. However, *economic impact* is strongly correlated with *environmental impact* and *trading partner*. If these relationships were taken into account, the result of *economic impact* and *climate suitability* being the two predictive factors can be explained.

Insects

For the insect group, two rules for high-risk situation were generated through the six runs of machine learning process. The fittest and most efficient rule was:

((ECONOMIC+TRADING)>3.5000)

indicating that the risk associated with the introduction of an insect into Tanzania was high if both risk scores for *economic impact* and *trading partner* were medium or at least one of the risk scores was high. The other rule:

(Pathway+Climate)>3.5

was information efficient as well, yet with less accuracy.

For this group, *economic impact* and *trading partner* are the most predictive in determining the risk. Basically, the assessor would consider the damage an insect could cause to yield/value and loss of market; and where the trading partner is from: neighbouring country, other African country or a country from another continent.

This result matches fairly well with that of correlation analysis (4.2.2.3), which indicated that the overall risk was affected by *economic impact* the most, followed by *pathway, host range, climate suitability and trading partner.* However, *pathway and trading partner are* strongly correlated, *climate* and *host range* are also strongly correlated with *economic impact.* If these relationships were taken into account, the result of *economic impact* and *trading partner being* the two predictive factors can be explained.

Nematodes

For the nematode group, five rules were generated through six runs of machine learning process, yet four of these five different rules were in fact having the same meaning. The fittest and most efficient rule was:

(ECONOMIC>1.0)

with a success rate of 100%. It suggested that the risk associated with the introduction of a nematodes species into Tanzania would be high if the risk score of *economic impact* was medium or high. The other risk factors were not so important in identifying high risk for nematodes.

It is also slightly differed to that of correlation analysis (see 4.2.2.4), which indicated that overall risk is affected by *economic impact* the most, followed by *pathway, host range, dispersal potential* and *trading partner.* However, closer examination reveals that *pathway, host range, dispersal potential* and *trading partner* are all highly correlated with *economic impact.* This may explain that why *economic impact* alone can be the predictive factor.

4.5.5 Conclusions

This section has described an application of a genetic algorithm (GA) technique, which is useful in predicting the risk of an organism being introduced into and becoming established in a new area.

The study used data of risk assessments for 252 species and subset data on different taxonomic groups, each with seven risk factors. Rules were evolved using GA software for all species combined and for subset taxonomic groups to forcast high risk situations. The results showed a crude success rate between 66.67% and 100% in forcasting high risk situation.

The fact that the majority of the rules found for each taxonomic group are quite similar in either format or meaning suggests that the machine learning process is stable and the results are robust.

The fact that different taxa groups had different predictive factors may be due to the differences of biology and the nature of trading. This complements the results from correlation analyses.

The results have also shown that the rules sets found for all potential quarantine pests as a group involved typically five risk factors, while the rule sets found for individual taxonomic groups involve at most three risk factors, in fact the majority of rules found involve only two risk factors. This suggests that not all the information is needed to assess the potential risk associated with the introduction of a pest, and the information needed to assess the potential risk associated with the pest within a particular taxonomic group could be further reduced.

The rules obtained with BEAGLE included a subset of the risk factors and could distinguish high-risk situations with relatively high accuracy. This suggests that simplification of risk assessment is possible.

While the prediction of the fittest rule set is more accurate, the most efficient rule set requires fewer numbers of risk factors, thus less information is needed, yet still maintains a reasonable success rate. Which rule set to choose depends on the cost/benefit balance of accuracy and resources required, i.e. the cost of a risk assessment vs. the consequence of getting it wrong.

These rules provide a good indication of what risk factor scores could lead to high risk situation. In some circumstances, for example, only limited resources (e.g. personnel or information) are available, or a quick decision needs to be made, these rules could be very beneficial.

However, it must be kept in mind that every PRA case has its novelty elements and these rules are only intended to provide a quick and approximate idea of what level of pest risk the risk scores would lead to. Furthermore, simplification of pest risk assessment is not to be taken lightly, as the consequence of a wrong assessment could be very serious.

There is an interesting parallel with the PCA work in that factors with high weightings turn out to be predictive factors in BEAGLE rules, and factors with low weightings in PCA appear less in Beagle rules. That two quite different techniques yield similar results lends credence to the conclusions about the relative importance of risk factors.

Further research will be useful at increasing the reliability of the GA to predict high-risk situations by adding more risk factors and by increasing the sample size. Other machine learning technique, e.g. neural networks have also been applied to pest risk analysis to predict establishment potential (see Ch2).

CHAPTER 5 CONCLUSIONS AND FUTURE NEEDS

The objective of this study is to explore the structural nature of the pest risk assessment process in plant quarantine decision-making and to explore approaches that could lead to the development of new methods for practical PRA in line with the requirement of "scientific justification" by WTO and FAO.

This thesis first reviewed the legislative and economic background to pest risk analysis and its significance in phytosanitary decision-making in the context of the growing international trade. Then an attempt was made to examine the international guidelines and various approaches developed over the years and identify some problematic areas.

The aim thereby was to develop methodologies for risk assessment as the major component of risk analysis and to explore different approaches in risk assessment. This chapter summarises the main findings of this study and draws out its implication for pest risk assessment methodology.

5.1 SUMMARY OF LITERATURE

The international guidelines and the regional standards on PRA were reviewed, as well as the various approaches by some countries.

The establishment of SPS Agreement motivated the Secretariat of the International Plant Protection Convention to develop international standards for phytosanitary measures and even made a revision of the convention itself necessary (FAO 1997).

Some Regional Plant Protection Organisations such as EPPO and NAPPO have also established PRA guidelines or schemes, which followed the general principles of the ISPMs but are more sophisticated and operable.

Member countries of WTO and IPPC are encouraged to apply the ISPMs. However, recognising the diversities of resources, circumstances, target pests or commodities between different countries, it is understandable that different countries may employ different approaches to PRA, provided they are in line with the ISPMs. Various methods e.g. qualitative or quantitative approaches are used to fit the different purposes. Some examples of national PRA approaches were discussed in Chapter 2.

Various national and non-governmental bodies have indeed made efforts in developing PRA methodologies. Yet it is found that the current practices and methods exhibited a number of characteristics that left scope for improvement, and it is necessary to examine a number of problem areas in relation with PRA, which is discussed at the end of this chapter.

In 2002, a review paper was presented on the approaches to pest risk assessment, which summarised the methodologies at the time and the problems involved in different approaches (Zhu *et. al.* 2002).

5.2 SUMMARY OF THE RESEARCH RESULTS AND DISCUSSIONS

In Chapter 3, a structure was proposed for the components of the pest risk. New methodologies to assist in PRA have been described and subjected to some evaluation with case studies. The pest risk assessment stage was divided into two steps: Pest risk identification (PRI) and pest risk evaluation (PRE).

The application of mind mapping was discussed, which proved to provide a means to identify the pest risk factors (PRI) that may lead to a pest introduction, in order to avoid important factors being overlooked, and potentially, to reduce ambiguity and increase transparency.

However, the limitation of mind mapping is that it only allows a hierarchical approach, where factors can be disaggregated into different numbers of levels. This can be sometimes misleading in that more factors under a heading might be taken to imply greater importance. However, this is not true, if we take the weighting results from Delphi study, for example, *abiotic habitat* was not disaggregated further, but it has a weighting of 0.21; whereas at the same hierarchy level, *intrinsic characteristics of the pest* was further disaggregated into 8 risk factors, but it only has a overall weighting of 0.18.

Some approaches to the evaluation of risk (PRE) were proposed, which facilitate the scoring of risk factors, and the subsequent weighting and combining of risk scores.

A method for introducing weighting to risk assessment is a novelty in PRA. Incorporating weighting into PRA was further investigated and several methods developed, which included subjectively assigned weighting and Delphi technique-derived weighting. Principal component analysis-derived weighting was discussed in Chapter 4.

The weightings obtained from a Delphi study can help provide a better understanding of the different importance of each risk factor. The result of such a weighting also provides a less biased or more objective opinion because it is derived from a wide range of risk analysts and quarantine officials, who brought in different experience, knowledge and background. However, the feasibility of this approaches might become an issue because of the cost and availability of experts.

Results of weightings from this study can provide a starting point for a risk analyst to commerce his/her own analysis, should he/she wish to incorporate weighting into the assessment. The weightings can be slightly adjusted to reflect his/her own perception of the importance of each risk factor, as long as the sum of the weightings remains equalling to 1.

Metrics for combining risk scores into an overall risk value were also explored, compared and evaluated. Weighted average and high and low biased weighted average were proven to be better, reflecting the perceived quarantine status, at least for those cases examined.

Chapter 4.1 presented the results of the analyses of the correlations and interaction between risk factors, which revealed that some risk factors were highly correlated with others; some were relatively independent.

Risk factors of different taxonomic group of pests correlate and interact differently.

When involving large numbers of risk factors, some very strong correlations would always exist, which means there is redundancy in the assessment process; those highly correlated factors might be combined or redefined or modelled by others.

Risk factors may be classified into two different types: primary factors and secondary factors. A primary factor is a risk factor that has a direct effect on but is not affected by, or is independent of others. A secondary factor is a risk factor affected by or derived from other factors. This seems to have further applications. One possibility would be to develop a summary scheme with all primary factors, which may provide an approximate and quick estimation of the risk level. However, it is still important to find out what role the secondary factor play, i.e. how do they modify the result. On the other hand, a cluster of inter-correlated factors might as a group form a primary factor. Due to the scope of this thesis, this is not studied further, but may warrant future study.

Another implication of the result of correlation analysis is that it would be a useful first step for the application of Bayesian nets in PRA. In a recent study of Bayesian application by Holt, some nodes didn't depend on others whilst others were conditional on other nodes. Establishing the conditional factors would be a useful first step for this (Holt 2009 pers. comm.).

In Chapter 4.2, it was considered whether all the factors were indeed contributing the same degree to the outcome of a risk assessment, and how they contributed to the overall risk rating. The results from cluster analysis suggested that the means and variances of risk factor scores could be effectively used to determine the importance of risk factors in terms of contributing to the overall risk assessment and distinguishing the level of overall risk.

Based on the seven-cluster result, also taking into account the risk factors falling into each clusters, the application of different clusters of risk factors for different purposes was proposed:

- Risk factors for preliminary assessment, which could be used to decide whether an organism has the characteristics of a quarantine pest.
- Risk factors for determining the level of risk; this group of factors could distinguish the level of risk more efficiently, which could be used as the key risk factors to determine whether a pest risk is high or low.
- Redundant risk factors; they either contribute little to the assessment of the level of risk or are dependant to or can be expressed by other risk factors. Therefore they could be eliminated from the scheme.

By grouping risk factors for different purposes, risk assessment could be potentially simplified, but rigor still could be retained.

There are some interesting similarities between the results from the cluster analysis and that of the Delphi study. For example, some risk factors suggested to be eliminated, obtained low weightings in the Delphi study; some factors suggested to be used for determining the level of risk, obtained high weightings in the Delphi study. This gives the author more confidence in this proposed approach, especially as the results were derived from different data: one was based on real PRA cases; the other was based on expert opinion on risk factors without specifying a pest.

Further study is needed to evaluate the applications proposed above. The evaluation would need to answer the following questions:

- Whether an organism would be characterised as a quarantine pest by only using risk factors in Group 1; and
- Whether it would achieve similar results of overall risk by only using factors in Groups 2 and 3.

In chapter 4.3, it was proposed to use Principal Component Analysis (PCA) to deal with the high correlation among the risk factors and to place different weights on different risk factors for the overall risk assessment.

The results from the PCA analysis do suggest that weightings can be derived for individual risk factors by applying statistical techniques to historical data of pest introductions and invasions, previous PRA cases, or expert opinion.

Historical data and previous case studies do not necessarily apply to new situations; however, these can provide at least a starting point for new pests.

By putting different weights to different risk factors, the more important risk factors can be identified, and the less important risk factors, i.e. risk factors contributing less to the overall risk assessment, are eliminated. Therefore risk assessment can be simplified without compromising the rigor.

However, it is difficult to develop a generic weighting pattern for different pest categories, since different taxonomic groups have different weighting patterns. Based on the weightings thus derived, a quick summary scheme may be developed, which will give a quick and preliminary idea of pest risk rating. Results from PCA match well with those of correlation analysis.

Chapter 4.4 described an application of a genetic algorithm (GA) technique, which is useful in predicting the risk of an orgamism being introduced into and become established in a new area. It is also a useful way to reflect the assessors' perception of risk using a simplified set of conditions.

The fact that the majority of the rules found for each taxonomic group are quite similar in either format or meaning suggests that the machine learning process is stable and the results are robust. The fact that different taxa groups had different predictive factors may due to the differences of biology and the nature of trading.

The results have also shown that the rules sets found for all potential quarantine pests as a group involved typically five risk factors, whereas the rule sets found for individual taxonomic groups involve at most three risk factors, actually the majority of rules found involve only two risk factors. This suggests that not all the information is needed to assess the potential risk associated with the introduction of a pest, and therefore the information needed to assess the potential risk associated with the pest within a particular taxonomic group can be further reduced. This may due to the different biological characteristics of the taxonomic groups, i.e. some factors within a taxonomic group have similar scores so that they could not be used to distinguish the level of risk. For example, *dispersal potential* did not appear in the rules for insect group, perhaps because the insects assessed have similar dispersal mechanism. It might also be that less complicated rules were needed to distinguish cases when the data sets were smaller.

The results showed a crude success rate between 66.67% and 100% in forcasting high risk situations.

The rules obtained with BEAGLE were based on a subset of the risk factors and could distinguish high-risk situations with high accuracy. This suggests that simplification of risk assessment is possible.

There is an interesting parallel between BEAGLE results and PCA study, in that factors with high weightings turn out to be predictive factors in BEAGLE rules, and factors with low weightings in PCA appear less in BEAGLE rules.

That two quite different techniques yield similar results lends credence to the conclusions about the relative importance of the risk factors.

Further research will be useful at increasing the reliability of the GA to predict high-risk situations by adding more risk factors and by increasing the sample size. Other machine learning techniques might also be investigated.

5.3 GENERAL ISSUES AND DISCUSSIONS

In common with risk assessment in some other disciplines, a number of problems must be overcome, e.g. subjectivity, uncertainty, non-quantifiable variables, and the need to integrate information into a simple statement of risk. It is difficult to define the risk of pest introduction, which is not only dependent on the biological characteristics of the pest but is also related with many human activities, as well as involving many uncertainties.

During the course of this study, it became apparent that there was a lack of methods to handle economic and social criteria. It is sometimes possible to attach monetary values to some aspects of economic consequences. Adequate attention is not, however, paid to the assigning of values to environmental and social consequences, as well as some aspects of economic consequences. Further work in this area will be invaluable for risk assessment practice. Furthermore, in PRA practice such techniques of integrating biological, economic, environmental and social impacts that have been developed are not widely available. Methodologies designed to meet such demands need to be simple and practical as well as meaningful. Further research in this area will be extremely useful.

Lack of data remains a big issue in risk assessment. Even when the risk factors are identified, it may not be possible to make an evaluation on the probability of the risk due to lack of data.

The approaches to risk scoring discussed in this study do not allow for a statement of uncertainty, some techniques, e.g. using a fuzzy number to represent the risk score, was considered but not explored in this study; this might be worthwhile for future study.

The author also has reservation about the terminology used in some PRA schemes on qualitative and quantitative assessment. In the author's view, it is wrong to call an assessment "quantitative or semi-quantitative" simply by introducing score/number into an assessment. There is no fundamental difference between a 1-3 or 1-9 scoring system and the qualititative description of high, medium and low risk, as none of those is based on observation data.

Methodology for a summation or combination of individual risk scores and assessments under different criteria needs to be dealt with. This study proposed several formulae for the combination of the individual risk scores, but further application and evaluation with more case studies will be beneficial.

Overlap in information suggests that a) clear definition of risk factors is needed; and b) redundancy/simplification of risk factors is possible. This is particularly important when there are more risk factors and ambiguities involved.

Further research will be useful at increasing the reliability of the research presented in Chapter 4, e.g. weightings derived from PCA and the used of GA to predict high-risk situations, by adding more risk factors and increasing the sample size for some pest groups.

Introducing weighting into risk assessment to differentiate the degree of the importance for risk criteria and individual risk factors is also an area that warrants future research. This study put forward some methodologies to assign weighting to a risk score to enable the integration of weighting into risk assessment. Future work to evaluate these methodologies and the application to more and larger risk assessment data will be advantageous.

Weightings and rules differ for different group of pests; it may be impossible or meaningless to develop a universal pattern. However, it may be possible to develop patterns for different pest group either by taxonomy or host.

PRA is based on predictions. The technical work involves estimating the biological, economic and social impacts that may be caused by the introduction of a new pest. In the case of some impacts, predictions may be based on some mathematical formulae. For others, such as some environmental and social impacts, numerical analysis might not be possible. Regardless of how predictions are derived, though, they are not facts and should not be presented as such.

5.4 CONCLUSIONS AND FUTURE NEEDS[1]

Organisational and interdisciplinary cooperation. Risk assessment has been the subject of much less research in plant quarantine than in a number of other disciplines. It has been left largely to the plant quarantine authorities themselves to devise workable schemes. PRA would benefit from a more academic framework and the involvement of different stakeholders such as research scientists and industry. PRA is an essentially multidisciplinary activity combining environmental science, economics, mathematics and biology. Some aspects have received less attention than others and in particular, guidelines for economic impact assessment need to be developed or reviewed.

Simplifying risk assessment approach. Simplified approaches that maintain the rigour of risk assessment without sacrificing necessary detail and depth is needed to accelerate the phytosanitary decision-making procedure. Zhu *et al.* (2002a) proposed some approaches that showed that simplification of risk assessment was possible. Such approaches would be particularly attractive to developing country trading partners who may have severe resource

---

[1] This section is largely based on a review paper the author presented at the BCPC conference 2002 (Zhu *et al.*, 2002b)

limitations. It is inevitable that simplification will lead to loss of accuracy and perhaps a central question concerns where the balance between simplification and accuracy should lie.

Incorporating weighting into PRA. Ideally, weighting should come from historical data of the pests that have already been introduced to new areas. Previous pest introductions and invasions can provide valuable information for PRA but previous data do not necessarily apply to new situations. Some common ground certainly exists in the weightings appropriate for different pest groups and these can provide at least a starting point for new pests.

Improving quantitative analysis. Due to the diversity and large quantity of information involved in PRA, it is extremely difficult to collate it to provide an overall pest risk assessment. Limiting a quantitative assessment to a few risk factors (which currently, is often the case) might lead to errors. Methods should be developed for risk ranking and scoring, as well as combining risk scores. Also, it might be asked whether some risk factors cannot be quantified at all. If so, how are they to be recognised in the final risk assessment?

Finally, a lesson learned from probabilistic study of the space shuttle was that *"conservative estimates should not be mixed with probabilities that represent mean future frequencies of failures. Otherwise the results are meaningless and possibly counterproductive"* (Pate-Cornell and Dillon, 2001). Applying this to pest risk assessment, when combined risk factors with a high degree of uncertainty with those can be more accurately assessed, the overall result may be equally questionable. Instead the uncertainty should be explicit and open for scrutiny.

PAGE

NUMBERING

AS ORIGINAL

REFERENCES

Adler, M., and E. Ziglio. 1996. Gazing into the Oracle: The Delphi Method and its Application to Social Policy and Public Health. Jessica Kingsley Publishers, London.

Anderson, D. R., D. J. Sweenly, and T. A. Williams. 1997. An Introduction to Management Science. West Publishing Company, Minneapolis/St. Paul.

Anon. 1999. Executive Order 13112. USA. 1999.

Anon. 1999b. Commission Decision of 26 May 1999 on Emergency Measures Against the Dissemination of Anoplophora Glabripennis (Motschulsky) as Regards China (except Hong Kong). Official Journal of the European Communities: L137/45.

Anon. 2000. Plant Protection Act. SEC 402. USA. June 2000.

Armstrong, R. A., D. Nochlin, and T. D. Bird. 2000. Neuropathological heterogeneity in Alzheimer's disease: A study of 80 cases using principal components analysis. Neuropathology 20: 31-37.

Baker, R. H. A. 1996. Developing a European Pest Risk Mapping System. EPPO Bulletin 26: 485-494.

Baker, R. H. A., A. MacLeod, and C. E. Sansford. 1999. Pest risk analysis: The UK experience. ANPP - Fifth International Conference on Pests in Agriculture. 7-9 December 1999. Montpellier, pp 119-126

Baker, R. H. A., C. E. Sansford, C. H. Jarvis, R. J. C. Cannon, A. MacLeod, and K. F. A. Walters. 2000. The Role of Climatic Mapping in Predicting the Potential Geographical Distribution of non-Indigenous Pests Under Current and Future Climates. Agriculture, Ecosystems & Environment 82: 57-71.

Baker, R. H. A., Cannon, R. J. C., & MacLeod, A. 2003. Predicting the potential distribution of alien pests in the UK under global climate change: *Diabrotica virgifera virgifera*. In Proceedings of the BCPC International Congress - Crop science and technology 2003, pp. 1201-1208. Alton, UK, BCPC Publications. 1236 pp.

Baker, R., R. Cannon, P. Bartlett, and I. Barker. 2005. Novel Strategies for Assessing and Managing the Risks Posed by Invasive Alien Species to Global Crop Production and Biodiversity. Annals of Applied Biology 146: 177-191.

Black, R., and A. Sweetmore. 1995a. Appropriate diagnostic systems for less developed countries. EPPO Bulletin 25: 383-391.

Black, R., and A. Sweetmore. 1995b. Crop Protection in the Developing World. British Crop Protection Council, Farnham, England.

Black, R., and A. Sweetmore. 1995c. Plant Quarantine: A Primer for Development Workers. Natural Resources Institute.

Black R., Roshan Abdallah. 1997. Analysis of Plant Quarantine

Black, R. 2000. Lecture notes for the MSc Course: Plant quarantine and sustainable agriculture (Natural Resources Institute, University of Greenwich).

Black, R. 2003. The Legal Basis for Controls on Importation of Animal and Plant Material into the United Kingdom. Environmental Law Review 5: 179-192.

Bourke, P. M. A. 1964. The Emergence of Potato Blight, 1843-46. Nature 203: 805-808.

Brasier, C. M. 2008. The Biosecurity Threat to the UK and Global Environment From International Trade in Plants. Plant Pathology 57: 792-808.

Bryant, P. J. 2002. Biodiversity and Conservation. A Hypertext book. University of California, Irvine.

CAB International. 2000. Crop Protection Compendium. CAB International. Wallingford, UK.

Calinski, R.B. and Harabasz, J. 1974. A dendrite method for cluster analysis. Communications in Statistics 3: 1-27.

Chapin, J. K., and M. A. L. Nicolelis. 1999. Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. Journal of Neuroscience Methods 94: 121-140.

Claiborne, W. 1997. Unwanted Serpents Enter Hawaii, Snakeless Island Fights Onslaught. The Washington Post, Volume 117 Issue 33, August 23, 1997.

Coakley, S. M. Scherm, H. Sukumar Chakraborty. 1999. Climate change and plant disease management. Annual Review of Phytopathology. 1999. 37: 399-426.

Coleman, M. E., and H. M. Marks. 1999. Qualitative and quantitative risk assessment. Food Control 10: 289-297.

Cook, D., and W. Proctor. 2007. Assessing the threat of exotic plant pests. Ecological Economics 63: 594-604.

Daehler, C. C., Denslow, J. S., Ansari, S. and Kuo, H. C.. 2004. A Risk-Assessment System for Screening Out Invasive Pest Plants from Hawaii and Other Pacific Islands. Conservation Biology 18.

Dawber, T. R., and W. B. Kannel. 1961. Susceptibility to coronary heart disease. Mod Concepts Cardiovasc Dis 30: 671–77.

DEFRA. 2003. Review of non-native species policy: Report of the working group. Department for Environment, Food and Rural Affairs, London.

Dobesberger, E. J. 2002. Multivariate Techniques for Estimating the Risk of Plant Pest Establishment in New Environments. The North American Plant Protection Organization: International Symposium on Pest Risk Analysis. March 2002. Purto Vallarta, Mexico.

Dohoo, I. R., C. Ducrot, C. Fourichon, A. Donald, and D. Hurnik. 1997. An Overview of Techniques for Dealing With Large Numbers of Independent Variables in Epidemiologic Studies. Preventive Veterinary Medicine 29: 221-239.

EPPO. 1997. Guidelines on Pest Risk Analysis: Pest Risk Assessment Scheme. EPPO, Paris.

EPPO. 1998a. Guidelines on Pest Risk Analysis: Pest Risk Analysis to Decide Immediate Action to Be Taken on Interception of a Pest in an EPPO Country. EPPO, Paris.

EPPO. 1998b. Guidelines on Pest Risk Analysis: Check-list of Information Required for Pest Risk Analysis (PRA). EPPO, Paris.

EPPO. 1999. Plant Quarantine Retrieval System. EPPO, Paris.

EPPO. 2002. Pest risk analysis on detection of a pest in an imported consignment. EPPO/OEPP Bulletin 32: 235-239.

Evans, H. F., A. MacLeod, and R. H. A. Baker. 1998. Pest Risk Assessment of Anoplophora Glabripennis. Forest Research / CSL, York.

Everett, R. A. 2000. Patterns and pathways of biological invasions [News and Comment]. Trends in Ecology and Evolution 15: 177-178.

Everitt, B. S., and G. Dunn. 1992. Applied Multivariate Data Analysis. Oxford University Press, New York.

Everitt, B. S. 1993. Cluster Analysis. Edward Arnold, London.

Evett, I. W., and E. J. Spiehler. 1987. Rule Induction in Forensic Science. Central Research Establishment, Home Office Forensic Science Service, Aldermaston, Reading, Berkshire RG7 4PN.

Fedra, K. 1998. Integrated Risk Assessment and Management: Overview and State of the Art. Journal of Hazardous Materials 61: 5-22.

Ferson, S., W. Root, and R. Kuhn. 1998. RAMAS Risk Calc: Risk Assessment with Uncertain Numbers, Applied Biomathematics, New York.

Filho, J. L. R., P. C. Treleaven, and C. Alippi. 1994. Genetic-Algorithm Programming Environments. Computer 1994: 28-43.

for International Development Project ZA0186. Chatham,

Forrest, S. 1993. Genetic Algorithms: Principles of Natural Selection Applied to Computation. Science 261: 872-878.

Forsyth, R. 1981. BEAGLE - A Darwinian Approach to Pattern Recognition. Kybernetes 10: 159-166.

Forsyth, R. 1989. Machine Learning, Principles and Techniques. Chapman and Hall Computing, New York.

Frank, J. H., and E. D. Mccoy. 1995. Introduction to Insect Behavioural Ecology: The Good, The Bad, and The Beautiful: Non-Indigenous Species in Florida. The Florida Entomologist 78: 21-35.

Gratz, N. G., R. Steffen, and W. Cocksedge. 2000. Why aircraft disinfection? Bulletin of the World Health Organization 78: 995-1004.

Gray, G. M., J. C. Allen, D. E. Burmaster, S. H. Gage, J. K. Hammitt, S. Kaplan, R. L. Keeney, J. G. Morse, D. W. North, J. P. Nyrop, A. Stahevitch, and R. Williams. 1998. Principles for Conduct of Pest Risk Analyses: Report of an Expert Workshop. Risk Analysis 18.

Griffin, R. 2002. PRA: A Global Perspective, Present and Future. The North American Plant Protection Organization: International Symposium on Pest Risk Analysis. March 2002. Purto Vallarta, Mexico.

Haimes, Y. Y. 1991. Total Risk Management. Risk Analysis 11: 169 - 171.

Haimes, Y. Y. 2004. Risk Modeling, Assessment, and Management. Wiley.

Henneberry T. J., M. R. V. O., and P. Anderson. 2001. History, current status, and collaborative research projects for Bemisia tabaci. Crop Protection 20: 709 - 723.

Hertz, D. B. 1964. Risk Analysis in Capital Investment. Harvard Business Review.

Hertz, D. B., and H. Thomas. 1983. Risk analysis and its applications. Wiley, Chichester; New York.

Holland, J. H. 1975. Adaption in Natural and Artifical Systems. University of Michigan Press.

Holt, J., R. Black, and R. Abdallah. 2006. A rigorous yet simple quantitative risk assessment method for quarantine pests and non-native organisms. Annals of Applied Biology 149: 167-173.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. Journal of Education Philosophy. 1933. 24: 417-41.

Http://www.planetark.org. 2001. Asian beetle costs US hundreds of millions, Daily News Story.

IPPC. 1995a. ISPMs. No. 2. Guidelines for pest risk analysis. FAO, Rome.

IPPC. 1995b. ISPM No. 5: Glossary of phytosanitary terms. FAO. Rome.

IPPC. 1997. 1997 - International Plant Protection Convention (New Revised Text). FAO. Rome.

IPPC. 1999. Pest Risk Analysis for Quarantine Pests. FAO, Rome.

IPPC. 2001. ISPMs. No. 11. Pest risk analysis for quarantine pests. FAO, Rome.

IPPC. 2005. ISPM No. 03. Guidelines for the export, shipment, import and release of biological control agents and other beneficial organisms. FAO, Rome.

IUCN. 2000. IUCN Guidelines for the Prevention of Biodiversity Loss Caused by Alien Invasive Species. The International Union for Conservation of Nature. Nairobi, Kenya.

James, W. C. 1974. Assessment of Plant Diseases and Losses. Annual Review of Phytopathology 12: 27-48.

Jarvis, C. H., and R. H. A. Baker. 2001a. Risk assessment for nonindigenous pests: 1. Mapping the outputs of phenology models to assess the likelihood of establishment. Diversity & Distributions 7: 223-235.

Jarvis, C. H., and R. H. A. Baker. 2001b. Risk assessment for nonindigenous pests: 2. Accounting for interyear climate variability. Diversity & Distributions 7: 237-248.

Jones, D. 2007. Arguments for a low risk of establishment of Karnal bunt disease of wheat in Europe. European Journal of Plant Pathology 118: 93-104.

Kahn, R. P. 1979. A Concept of Pest Risk Analysis. OEPP Bulletin 9: 119-130.

Kannel, W. B., T. R. Dawber, A. Kagan, N. Revotskie, and J. Stokes. 1961. Factors of risk in the development of coronary heart disease—six year follow-up experience. The Framingham Study. Ann Intern Med 55: 33–50.

Kaplan, S., and B. J. Garrick. 1981. On the Quantitative Definition of Risk. Risk Analysis 1: 11-27.

Kareiva, P., and M. Quinlan. 2002. Review of Methodologies for Environmental Risk Assessments, The North American Plant Protection Organization: International Symposium on Pest Risk Analysis. March 2002. Purto Vallarta, Mexico.

Kaufman, L. a. R., P. J. 1990. Finding groups in data. John Wiley & Sons, New York.

Kizlinski, M. L., D. A. Orwig, R. C. Cobb, and D. R. Foster. 2002. Direct and indirect ecosystem consequences of an invasive pest on forests dominated by eastern hemlock. Journal of Biogeography 29: 1489-1503.

Liebhold, A. M., W. L. Macdonald, D. Bergadahl, and V. C. Mastro. 1995. Invasion by exotic forest pests: a threat to forest ecosystems. Forest Science 41: 1-49.

Liebhold, A. M., T. T. Work, D. G., McCullough, and J. F. Cavey. 2006. Airline Baggage as a Pathway for Alien Insect Species Invading the United States. American Entomologist 2006: 48-54.

Lowe, S., M. Browne, S. Boudjelas, and M. DePoorter. 2000. 100 of the World's Worst Invasive Alien Species A selection from the Global Invasive Species Database. The Invasive Species Specialist Group.

MacLeod, A., H. F. Evans, and R. H. A. Baker. 2002. An Analysis of Pest Risk from an Asian Longhorn Beetle (*Anoplophora glabripennis*) to Hardwood Trees in the European Community. Crop Protection 21: 635-645.

MacLeod, A., and R. H. A. Baker. 2003. The EPPO pest risk assessment scheme: assigning descriptions to scores for the questions on entry and establishment. EPPO/OEPP Bulletin 33: 313-320.

MacLeod, A., Evans, H. F., & Baker, R. H. A. 2003. The establishment potential of *anoplophora glabripennis* in Europe. [Das Ausbreitungspotenzial von *Anoplophora glabripennis* in Europa.] *Nachrichtenblatt Des Deutschen Pflanzenschutzdienstes, 55*(4), 83-84.

MacLeod, A. 2007. CSL Pest Risk Analysis for *Aceria tulipae.* Central Science Laboratory, UK.

McDowell, R. 2002. Introduction to Quantitative Analysis, The North American Plant Protection Organization: International Symposium on Pest Risk Analysis. March 2002. Purto Vallarta, Mexico.

Michalski, R. S., J. H. Davis, V. S. Bisht, and J. B. Sinclair. 1982. PLANT/DS: an Expert Consulting System for the Diagnosis of Soybean Diseases, pp. 133-138, Proceedings of the 1982 European Conference on Artificial Intelligence, Orsay, France.

Mitchell, T. M. 1997. Machine Learning. McGraw-Hill, Burr Ridge, IL.

Mitchell, T. M. 2006. The Discipline of Machine Learning. Department of Machine Learning Working Paper, School of Computer Science, Carnegie Mellon University.

Novak, R. 1992. The Asian Tiger Mosquito, Aedes albopictus. Wing Beats 3: 5.

OEPP/EPPO. 1997. EPPO Standards PM 5/3(1) Guidelines on Pest Risk Analysis: Pest Risk Assessment Scheme. OEPP/EPPO Bulletin 27: 281-306.

Oliveira M. R. V., Henneberry T. J., and Anderson P.. 2001. History, current status, and collaborative research projects for Bemisia tabaci. Crop Protection 20: 709-723.

Orlinski, A. D. 2006. Outcomes of the EPPO project on quarantine pests for forestry. EPPO/OEPP Bulletin 36: 497-511.

Orwig, D. A. 2002. Ecosystem to rRegional Impacts of Introduced Pests and Pathogens: Historical Context, Questions and Issues. Journal of Biogeography 29: 1471-1474.

Oryang, D. 2002. Probabilistic Scenario Analysis (PSA), a Methodology for Quantitative Risk Assessment. The North American Plant Protection Organization: International Symposium on Pest Risk Analysis. March 2002. Purto Vallarta, Mexico.

Paté-Cornell, M. E., and R. L. Dillon. 2001. Probabilistic Risk Analysis for the NASA Space Shuttle: A Brief History and Current Work. Reliability Engineering and System Safety 74: 345-352.

Peacock, L. Worner, S. 2006. Using analogous climates and global insect pest distribution data to identify potential sources of new invasive insect pests in New Zealand. New Zealand Journal of Zoology. 2006. 33: 2, 141-145.

Peacock, L. Worner, S. Sedcole, R. 2006. Climate variables and their role in site discrimination of invasive insect species distributions. Environmental Entomology. 2006. 35: 4, 958-963.

Peacock, L. Worner, S. Pitt, J. 2007. Application of artificial neural networks in plant protection. Bulletin OEPP/EPPO Bulletin. 2007. 37: 2, 277-282.

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine. 1901. 2 (6): 559–572.

Pegg, J. R. 2003. U.S. Losing War Against Invasive Species. Environment News Service December, 2003.

Pemberton, R. W. 2000. Predictable risk to native plants in weed biological control. Oecologia 125: 489-94.

Pest Risks for Tanzania. Final technical report, Department

Pfeiffer, J. 1968. New look at education. Odyssey Press, Poughkeepsie, NY.

Pimentel, D., L. Lach, R. Zuniga, and D. Morrison. 2000. Environmental and Economic Costs of Nonindigenous Species In the United States. BioScience 50: 53-65.

Pimentel, D., S. McNair, J. Wightman, J. Janecka, C. Simmonds, C. O'Connell, E. Wong, L. Russel, J. Zern, and T. Aquino. 2001. Economic and Environmental Threats of Alien Plant, Animal and Microbe Invasions. Agriculture, Ecosystems & Environment 84: 1-20.

Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the Environmental and Economic Costs Associated With Alien Invasive Species in the United States. Ecological Economics 52: 273– 288.

Pitt, J., J. Régnière, and S. Worner. 2007. Risk assessment of the gypsy moth, Lymantria dispar (L), in New Zealand based on phenology modelling. International Journal of Biometeorology 51: 295-305.

Porter, S. D., and D. A. Savignano. 1990. Invasion of Polygyne Fire Ants Decimates Native Ants and Disrupts Arthropod Community. Ecology 71: 2095-2106.

Rappaport, R. A. 1967. Pigs for the Ancestors. Yale University Press, New Haven.

Rizzo, D. M., and M. Garbelotto. 2003. Sudden oak death: endangering California and Oregon forest ecosystems. Front Ecol Environ 1: 197–204.

Rowe, W. D. 1990. Understanding Uncertainty. Risk Analysis 14: 743-750.

Saaty, T. L. 1980. The Analytic Hierarchy Process. McGraw Hill, NY.

Sansford, C. 2002. Quantitative versus Qualitative: Pest Risk Analysis in the UK and Europe including the European and Mediterranean Plant Protection (EPPO) system. The North American Plant Protection Organization: International Symposium on Pest Risk Analysis. March 2002. Purto Vallarta, Mexico.

Schrader, G., and J. G. Unger. 2003. Plant Quarantine as a Measure Against Invasive Alien Species: The Framework of the International Plant Protection Convention and the plant health regulations in the European Union. Biological Invasions 5: 357-364.

Sequeira, R. A. 2002a. Evaluation of the Impact of Exotic Species on the Environment: the IPPC Supplementary Standard to Pest Risk Analysis Guidelines. The North American Plant Protection Organization: International Symposium on Pest Risk Analysis. March 2002. Purto Vallarta, Mexico.

Sequeira, R. A. 2002b. Role of Spatial Analysis in Meeting Phytosanitary Challenges including Risk Analysis. The North American Plant Protection Organization: International Symposium on Pest Risk Analysis. March 2002. Purto Vallarta, Mexico.

Sgrillo, R. B. 2005. Quantitative Analysis Tools for Phytosanitary Measures: a Perspective from South America, International conference on sanitary and phytosanitary (SPS) risk assessment methodology. Washington, D. C.

Simberloff, D., and M. ALEXANDER. 1994. Issue paper on biological stressors. Ecological Risk Assessment Issue Papers, EPA/630/R-94/009, U.S. EPA, Washington DC.

Simberloff, D., and P. Stiling. 1996. How Risky is Biological Control? Ecology 77: 1965-1974.

Simberloff, D. 1996. Impacts of Introduced Species in the United States. CONESQUENCES: Volume 2, Number 2 1996. http://www.gcrio.org/CONSEQUENCES/vol2no2/article2.html

Simberloff, D., and M. ALEXANDER [eds.]. 1998. Assessing Risks to Ecological Systems From Biological Introductions (excluding Genetically Modified Organisms). Blackwell.

Simberloff, D. 2000a. Introduced Species: The Threat to Biodiversity & What Can Be Done. American Institute of Biological Sciences, An ActionBioscience.org original article, http://www.actionbioscience.org/biodiversity/simberloff.html

Simberloff, D. 2000b. Global climate change and introduced species in United States forests. The Science of The Total Environment 262: 253-261.

Simberloff, D. 2004. A Rising Tide of Species and Literature: A Review of Some Recent Books on Biological Invasions. BioScience 54: 247-254.

Simberloff, D. 2005. The Politics of Assessing Risk for Biological Invasions: the USA as a Case Study. Trends in Ecology and Evolution 20: 216-222.

Smith, C. S., W. M. Lonsdale, and J. Fortune. 1999. When to Ignore Advice: Invasion Predictions and Decision Theory. Biological Invasions 1: 89-96.

Smith, I. M., D. G. McNamara, P. R. Scott, and M. Holderness. 1997. Quarantine pests for Europe. CAB INTERNATIONAL, Wallingford, UK.

Smith, I. M. 2000. New Constraints to Plant Quarantine. Proceedings of the BCPC Conference - Pests & Diseases 2000. Brighton, UK.

Smith, R. M., R. H. A. Baker, C. P. Malumphy, S. Hockland, R. P. Hammon, J. C. Ostojá-Starzewski, and D. W. Collins. 2007. Recent non-native invertebrate plant pest establishments in Great Britain: origins, pathways, and trends. Agricultural and Forest Entomology 9: 307-326.

SPSS Inc. 2001. SPSS version 11. SPSS Inc. South Wacker Drive, Chicago,

Stanaway, M. A., M. P. Zalucki, P. S. Gillespie, and C. M. Rodriguez. 2001. Pest risk assessment of insects in sea cargo containers. Australian Journal of Entomology 40: 180-192.

StataCorp. 2001. Stata statistical software: Release 7.0. Stata Corporation. College station, TX.

Strong, D. R., and R. W. Pemberton. 2000. Biological Control of Invading Species--Risk and Reform. Science 288: 1969-1970.

Stynes, B. 2002. PEST RISK ANALYSIS: METHODS AND APPROACHES, The North American Plant Protection Organization: International Symposium on Pest Risk Analysis. March 2002. Purto Vallarta, Mexico.

Sutherst, R. W., G. F. Maywald, and W. Bottomley. 1991. From CLIMEX to PESKY, a Generic Expert System for Pest Risk Assessment. OEPP Bulletin 21: 595-608.

UK: Natural Resources Institute.

Venette, R., and J. Gould. 2006. A pest risk assessment for Copitarsia spp., Insects associated with importation of commodities into the United States. Euphytica 148: 165-183.

Vose, D. 2000. Risk analysis: A Quantitative Guide. Wiley & Sons, London.

Warner, F. 1992. Chapter 1 Introduction, Risk: Analysis, Perception and Management: Report of A Royal Society Study Group. The Royal Society, London.

Watler, D. 2002. Qualitative Analysis, Methodologies and Applications in Canada. The North American Plant Protection Organization: International Symposium on Pest Risk Analysis. March 2002. Purto Vallarta, Mexico.

White, P. C. L., A. E. S. Ford, M. N. Clout, R. M. Engeman, S. Roy, and G. Saunders. 2008. Alien invasive vertebrates in ecosystems: pattern, process and the social dimension. Wildlife Research 35: 171–179.

Wilcove, D. S., D. Rothstein, J. Dubow, A. Phillips, and E. Losos. 1998. Quantifying threats to imperiled species in the United States. BioScience 48: 607-615.

WTO. 1995. Agreement on the Application of Sanitary and Phytosanitary Measures. World Trade Organisation. Geneva.

Yang, X. B. Dowler, W. M. Royer, M. H. 1991. Assessing the risk and potential impact of an exotic plant disease. Plant Disease. 1991. 75: 10, 976-982.

Zhu, L., J. Holt, and R. Black. 2000. New Approaches to Pest Risk Analysis for Plant Quarantine, Proceedings of the BCPC Conference - Pests & Diseases 2000 1, 495-498. Brighton, UK.

Zhu, L., R. Black, and J. Holt. 2001. Novel Approaches to Pest Risk Assessment: Uncertainty and Quantification. . In: The Crop Protection Compendium. Global Module. 2001 Edition. CAB International: Wallingford.

Zhu, L. Black, R. and Holt J. 2002a. Investigating the nature of pest risk: towards a simplified quantitative assessment. The North American Plant Protection Organization: International Symposium on Pest Risk Analysis. March 2002. Purto Vallarta, Mexico

Zhu, L., R. Black , and J. Holt. 2002b. International developments in pest risk analysis for phytosanitary decision making: a review of methodologies for pest risk assessment. Proceedings of the BCPC Conference - Pests & Diseases 2002 pp. 911-916. Brighton, UK.

-

APPENDIX 1. AMBIGUITY IN PRA

AMBIGUITY IN PRA GUIDELINES

Effective pest risk assessment requires a thorough understanding of the components of the problem involved. It is felt that the main pest risk criteria (geographical and regulatory criteria, i.e. introduction/spread potential and economic importance) in PRA as first defined under ISPM2 by FAO in 1995 were very general and sometimes hard to assess directly. The FAO guidelines for PRA (ISPM2 and ISPM11) and the EPPO pest risk assessment scheme suggested some more details of what should be taken into account under the main criteria (IPPC, 1999; EPPO, 1997) (see Chapter 2). The contention here is that some risk factors are still either too broad or too ambiguous to assess directly and could sometimes give controversial outcomes from the same information.

The concern about ambiguities in the current PRA guidelines stemmed from a discussion with Dr Baker of CSL, an UK representative in the EPPO PRA working panel, in 2000. Some research was done on the existing EPPO pest risk assessment scheme (EPPO 1997) based on several PRA case studies and prepared a list of ambiguities to be discussed in an EPPO PRA panel meeting. This list of ambiguities in the EPPO PRA scheme covered mainly technical terms and ambiguous words or sentences. Moreover, it was felt that the consistency of the use of terms needed to be improved in the current schemes

AMBIGUITY IN RISK FACTOR IDENTIFICATION

Some further ambiguities were also identified during the PRA case studies.

For instance, the ambiguity of some questions in the EPPO risk assessment scheme was evident from conducting a PRA on *Anoplophora glabripennis* (Asian longhorn beetle, ALB) for Europe during a PRA exercise at the Central Science Laboratory (CSL) in 2000 (Zhu, unpublished). One question (question 1.5b) in the EPPO scheme was: "*how likely is the pest to survive existing cultivation or commercial practices? Not likely = 1, very likely = 9*[1]". Available data and personal experience indicated that ALB was very likely to survive existing cultivation practices, but not likely to survive commercial practices, e.g. commercial wood processing. Thus there was a high risk of it surviving existing cultural practices, whereas there was a low risk of it surviving commercial practices. Whether this question should be given a high or a low risk score was thus unclear. This example illustrated one of the ambiguities in the current EPPO scheme. Another ambiguous example question in the EPPO scheme was 1.6 "*How likely is the pest to survive or remain undetected during existing phytosanitary procedures?*" A note provided as guidance read:

> "existing phytosanitary measures (e.g. inspection, testing or treatments) are most probably being applied as a protection against other (quarantine) pests; the assessor should bear in mind that such

---

[1] More explanation on 1-9 scale see section "scoring risk factors"

> measures could be removed in the future if the other pests were to be re-evaluated. The likelihood of detecting the pest during inspection or testing will depend on a number of factors including:
>
> a. Ease of detection of the life stages, which are likely to be present. Some stages are more readily detected than others, for example insect adults may be more obvious than eggs;
>
> b. location of the pest on the commodity - surface feeders are more readily detected than internal feeders;
>
> c. symptom expression - many diseases may be latent for long periods, at certain times of the year, or may be symptomless in some hosts or cultivars and virulent in others;
>
> d. distinctiveness of symptoms - the symptoms might resemble those of other pests or sources of damage such as mechanical or cold injury;
>
> e. the intensity of the sampling and inspection regimes;
>
> f. distinguishing the pest from similar organisms.
>
> (not likely = 1; very likely = 9)"

It is argued here that this question would be better disaggregated into at least three smaller questions. The first could be a combination of notes *a, b* and *c*, i.e. *How easy to detect the pest?* This would consider the ease of detection of the relevant life stage, location on the commodity and the symptoms. The second could consider note *e: the intensity of the sampling and inspection.* The third could concern *the distinctiveness from other organisms or other source of symptom,* which is a combination of notes *d* and *f.*

AMBIGUITY IN PEST RISK ASSESSMENT

Another type of ambiguity was that controversy could arise from different interpretations of the same information. For instance, in the above PRA exercise on Asian longhorned beetle (ALB), some differences of opinion occurred between the pest risk analysts at CSL and the author, when assessing the risk associated with reproductive strategy within establishment potential (in general, the more effective the reproduction, the higher the pest risk). Both parties used the same information - long life cycle period and survival inside wood. The CSL staff held that the long life cycle period might enable ALB to reproduce during the long time or long distance of transportation, and survival inside wood meant it would be well protected; therefore a relative high risk score should be given. The author, on the contrary, regarded the long life cycle period as reducing the frequency of reproduction. Moreover, from the author's direct experience, ALB was not very likely to have both male and female individuals on a single piece of wood packaging material at the same time. Packaging material was regarded as an important pathway of the introduction of ALB. By this argument, the risk of reproduction in a new area was deemed low. This example illustrated that the same information could be interpreted in different ways according to the experience of the personnel involved.

APPENDIX 2. SOFTWARE MINDMANAGER

THE SOFTWARE MINDMANAGER

Based on the principles of the mind mapping method, MindManager, a tool for creative thinking, was developed from extensive research into how people receive and process information. MindManager combines the traditional pen-and-paper Mind Mapping method with modern technology and makes it possible to mind map on a PC (MindManager, MindJET).

THE STRUCTURE OF A MIND MAP

Figure 3.11 shows the structure and naming convention of a mind map as it is used in a typical MindManager application. Every thought uses a separate branch. Follow-up thoughts are then attached as sub-branches. The original branch is identified as a topic. This branching can be continued as long as it needed and even rearranged at a later point. The major issues in the problem or project are the Main Topics directly linked to the central title box of the mind map. During the development of the mind map, a complete tree forms, which shows all the thoughts and especially the relationships between those thoughts (MindManager, MindJET).

However, this software does have its shortcoming as it forces a rather hierarchical type of thinking, where interactions between components are harder to show, and arguably, therefore, tend to be ignored.

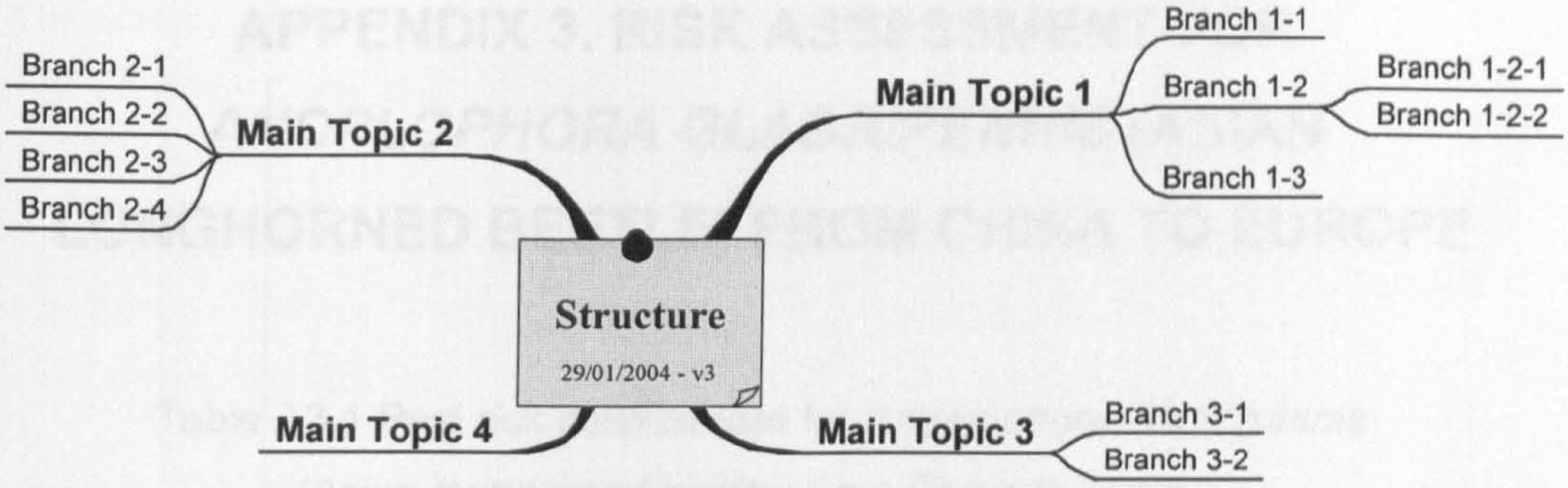


Figure A2.1 The structure of a Mind Map

APPENDIX 3. RISK ASSESSMENT FOR *ANOPLOPHORA GLABRIPENNIS* (ASIAN LONGHORNED BEETLE) FROM CHINA TO EUROPE

Table A3.1 Pest risk assessment for *Anoplophora Glabripennis* (Asian longhorned beetle) from China to Europe

Section A: Pest categorization	
Geographical criteria	The pest doesn't occur in the PRA area
Potential for establishment	There are host plants in the PRA area
	The pest doesn't need an alternate host plant
	The pest doesn't require a vector
Potential economic importance	The pest in its present range causes significant damage
The pest could present a risk to the PRA area	

Section B: Quantitative evaluation	**Score**	**Statement**
1. Probability of introduction		
Entry		
1.1 Number of pathways	2	Packaging material (dunnage)
1.3b Likelihood of association with pathway	6	Wood process
1.4 Concentration of the pest on the pathway	3	Low population on commodities
1.5b Likelihood of survival existing cultivation or commercial practices	6	Likely to survive existing cultivation, but not very likely to survive commercial practices
1.6 Likelihood of survival or remain undetected	6	Obvious symptom expression but difficult to detect, either because of life stage or sampling.
1.7b Likelihood of survival in transit	9	Long life cycle and remain within wood
1.8 Likelihood of multiplication during transit	1	Not likely
1.9 Movement along the pathway	2	Lack of information, assume a low medium volume
1.10 Commodity distribution	5	Lack of information, assume a medium range of distribution
1.11 Width of spread in time of commodity arrival	8	Perhaps the whole year
1.12b Likelihood of transfer to a suitable host	6	Likely to transfer to private gardens and amenity plantings
1.13 Intended use of Commodity	8	Wide spread of commodities
Establishment		
1.14 Host plant species	4	Some trees
1.15 Host plant extension	6	Wide distribution
1.19 Wild plant aids dispersal or maintenance	9	Very likely

1.20 Similarity of climatic conditions	8	
1.21 Similarity of other abiotic factors	8	
1.22 Competition	8	Not likely to have competition from existing species
1.23 Natural enemies	9	Lack of effective natural enemy
1.25 Control measures for other pests	8	Very difficult to control
1.26 Reproductive strategy	5	Long life cycle and not very likely to introduce both male and female in the same time
1.27 Populations for establishment	3	Low population is not likely to become established
1.28 Probability of eradication in the PRA area	4	Easy to eradicate but difficult to discover the infection
1.29 Genetic adaptability	5	Non-polymorphic, no information on mutation rate
1.30 Introduction to a new area	2	Only two cities in USA
2. Economic impact assessment		
2.1 Economic importance in existing area	4	Loss of forest products either in yield or in quality, as well as the change of packaging material
2.2 Environmental damage in existing area	8	Attacking protection forests, amenity plantings
2.3 Social damage in existing area	5	Additional costs on research, advices, publicity, etc.
2.4 Endangered area in the PRA area	5	Europe
Spread potential		
2.5 Spread by natural means	2	Very slow
2.6 Spread by human activities	7	With commodities
2.7 Containment of spread in the PRA area	7	Easy to contain by destroying the infected trees, but difficult to discover the infection
2.8 Direct effects on crop yield/quality	7	Causing leaves drop, branches or tree fall; indirect damage e.g. the wounds may easily be attached by secondary pests and diseases
2.9 Effects on producer profits	7	There aren't many forestry industries
2.10 Effects on consumer demand	5	No direct effects,
2.11Efffects on export market	6	May cause changes for packaging material
2.12 Other costs	7	Research, advice, phytosanitary measures etc.
2.13 Likelihood of environmental damage in the PRA area	4	Damage to forest, private and amenity lands
2.14 Likelihood of social damage in the PRA area	3	Tourism
2.15 Natural enemies	9	No effective natural enemies
2.16 Ease of control	8	Very difficult to control
2.17 Disruption of control measures to other pests	1	Not likely

2.18 Side-effects of control measures	5	Causing pollution, danger to human health
2.19 Pest resistance	1	Lack of information
The overall risk score for *Anoplophora glabripennis*	242	
The averaged risk score for *Anoplophora glabripennis*	5.5	
The outcome of the pest risk assessment: a MEDIUM to LOW risk to Europe		

APPENDIX 4. DELPHI STUDY - A SURVEY FOR PEST RISK ANALYSIS: WEIGHTING RISK ELEMENTS ACCORDING TO IMPORTANCE

Lihong Zhu

Natural Resources Institute, University of Greenwich, UK

If this questionnaire applies to a specific pest or PRA, please indicate in box below.

☐

1. Main Risk Criteria

1.1 *the following risk factors are taken into account while assessing Introduction/spread Potential in Pest risk analysis (PRA[2]).*

Please rank orders beginning with the most important one, score 1,2 or 3 in the boxes. If you think some of them are equal, give the same score.

a. Entry potential ☐

b. Establishment potential ☐

c. Spread potential ☐

1.2 *Which factor do you think more important for Economic Consequences in PRA?*

Please rank orders. If you think they are equal, give the same score.

a. Direct pest effect (e.g. crop losses, costs of control measures etc.) ☐

b. Indirect pest effect (e.g. effects on markets, environmental & social effects, capacity to act as a vector for other pests etc.) ☐

---

[2] The process of evaluating biological or other scientific and economic evidence to determine whether a pest should be regulated and the strength of any phytosanitary measures to be taken against it [FAO, 1995; revised IPPC, 1997]

*For other definitions, please see: Glossary of Phytosanitary Terms (FAO, ISPMs No. 5)

http://www.fao.org/WAICENT/FAOINFO/AGRICULT/AGP/AGPP/PQ/En/Publ/ISPM/ispms.htm

http://www.eppo.org/html/glossary.html

2. Elements of the Risk Factors

Please rank orders beginning with the most important element (score 1). If you think some of them are equal, give them the same score. If you think some of them are unnecessary, leave the box (es) blank. Please specify any other risk elements you think important for this criterion and rank them.

21 The following risk elements are taken into account while assessing Entry Potential in PRA.

a. Pathway ☐

b. Pest survival during transport or transit ☐

c. Possibility of remaining undetected ☐

d. Frequency of interception ☐

e. Intended use of the commodities ☐

f. Others (please specify) ☐

2.1.1 From 2.1a, the following elements are taken into account while assessing the risk of Pathways.

a. Number of pathways ☐

b. Ability of association with pathways ☐

c. Frequency of movement along pathways ☐

d. Quantity of movement along pathways ☐

e. Duration of the pest stage or state being associated with pathways ☐

f. Others (please specify)

2.1.2 *From 2.1b, the following elements are taken into account while assessing the risk of Survival during transport or transit.*

a. Capacity to withstand starvation/remain viable ☐

b. Proximity of origin to the PRA area ☐

c. Speed of transport ☐

d. Condition of transport ☐

e. The life stage likely to be transported ☐

f. Duration of life cycle in relation to time in transport ☐

g. Measures applied throughout the pathways ☐

h. Others (please specify) ☐

2.2 *The following risk elements are taken into account while assessing Establishment Potential after entry in the PRA area.*

a. Biotic factors for establishment ☐

b. Abiotic factors for establishment ☐

c. Intrinsic characteristics conducive to establishment ☐

d. Survival under current cultural practices & control measures ☐

e. Frequency of introduction/spread to a new area from the origin ☐

f. Others (please specify) ☐

2.2.1 *From 2.2a, the following elements are taken into account while assessing the suitability of Biotic Factors for establishment in the PRA area.*

a. Presence of host, vector ☐

b. Number of host/vector species ☐

c. Density of hosts/vectors ☐

d. Distribution of hosts/vectors ☐

e. Ability of the pest being associated with host/vector ☐

f. Competition with the existing species ☐

g. Natural enemies ☐

h. Others (please specify) ☐

2.3 *The following risk elements are taken into account while assessing Spread Potential after establishment in the PRA area.*

a. Spread by natural means ☐

b. Spread by human activities ☐

c. Biotic suitability for pest spread in the PRA area ☐

d. Abiotic suitability for pest spread in the PRA area ☐

e. Others (please specify) ☐

2.4 *The following elements are taken into account while assessing Direct Pest Effects on Economic Impact.*

a. Value of host plants attacked ☐

b. Host plant damage (type, frequency & quantity) ☐

c. Crop losses (in yield and grade) ☐

d. Ability of populations to reach damaging densities ☐

e. Rate of pest spread (including number and distribution of host plants) ☐

f. Effect on existing production practices (e.g. changes to producer costs or input demands, quality changes to products, loss of a domestic agricultural sector etc.) ☐

g. Cost of control measures ☐

h. Others (please specify) ☐

3. Rules for Scoring Risks

3A. Please tick the approximate boxes for 3.1 to 3.3.

3.1 How far should the proximity of the origin to the PRA area be considered as conducive to high risk of pest entry?

a. <=500 km ☐

b. <=1000 km ☐

c. <=2000 km ☐

d. Other (please specify) ☐

3.2 From 2.1d, if the Frequency of interception within a certain time (e.g. per year) can be used to show the risk of pest introduction, how many times should interception occur before it is taken to indicate HIGH risk of entry?

a. 1-2 times ☐

b. 3-5 times ☐

c. 6-10 times ☐

d. More than 10 times ☐

e. Other (please specify) ☐

3.3 From 2.2e, if the Frequency of the occurrence of introduction to a new area from the origin can be used to show a risk of pest Introduction. How many times should introduction occur before it is taken to indicate HIGH risk of introduction?

a. 1-2 times ☐
b. 3-5 times ☐
c. 6-10 times ☐
d. More than 10 times ☐
e. Others (please specify) ☐

3B. Please rank the following pest characteristics for 3.4 to 3.5. Please specify other characteristics you think important for pest establishment.

3.4 The following characteristics are taken into account while assessing the Reproduction Potential of the pest in the new environment.

a. Parthenogenesis/asexual ☐
b. Typical timing of life cycle ☐
c. Number of generations per year ☐
d. Reproductive strategy (e.g. sex ratio, egg number/number & type of spore stages, hatching ratio etc.) ☐
e. Others ☐

Which of the above is the most important one? (Please give the letter) ☐

3.5 The following genetic characteristics are taken into account while assessing the Adaptability Potential of the pest to the new environment?

a. Polymorphic/more than one spore stage ☐
b. Polyphagous/large number of host species ☐

c. Pesticide resistance ☐

d. Adaptations to survival (e.g. diapause/resting stage) ☐

e. Others ☐

Which of the above is the most important one? (Please give the letter) ☐

4. Risk Management

There are some phytosanitary measures appropriate to reduce the risk of pest introduction.

Please rank beginning with the most effective one (score 1). If you think some of them are equal, give the same score. If you think some of them are unnecessary, leave the box blank. Please specify any other measures you think effective for reducing pest introduction and rank them.

a. Phytosanitary certification ☐

b. Pest-free area ☐

c. Pre-clearance ☐

d. Quarantine treatment ☐

e. Post entry quarantine ☐

f. Road/border inspection ☐

g. Others (please specify) ☐

5. Please add any risk elements you think necessary in PRA below and to rank them if you wish:

Please add any comments about the questionnaire below:

Contact Details

Name:

Title:

Organisation:

Email:

Tel: Fax:

Address:

Please return the completed questionnaire to:

Ms Lihong Zhu
Natural Resources Institute
University of Greenwich
Central Avenue, Chatham Maritime
Kent, ME4 4TB
UK

Or email to: l.zhu@gre.ac.uk

APPENDIX 5. WEIGHTINGS DERIVED FROM EXPERT SURVEY BY DIFFERENT TRANSFORMATIONS

Figure A5.1 – A5.4 show the weightings derived from expert survey by the following transformations on the rankings:

Figure A5.1: Linear transformation $Wr_i = (9 - rk_i)$

Figure A5.2: Squared transformation $Wr_i = (9 - rk_i)^2$ and

Figure A5.3 and A4.4: Exponent transformation $Wr_i = e^{(a(9 - rk_i))}$,

where $a = 0.4$ or 0.8

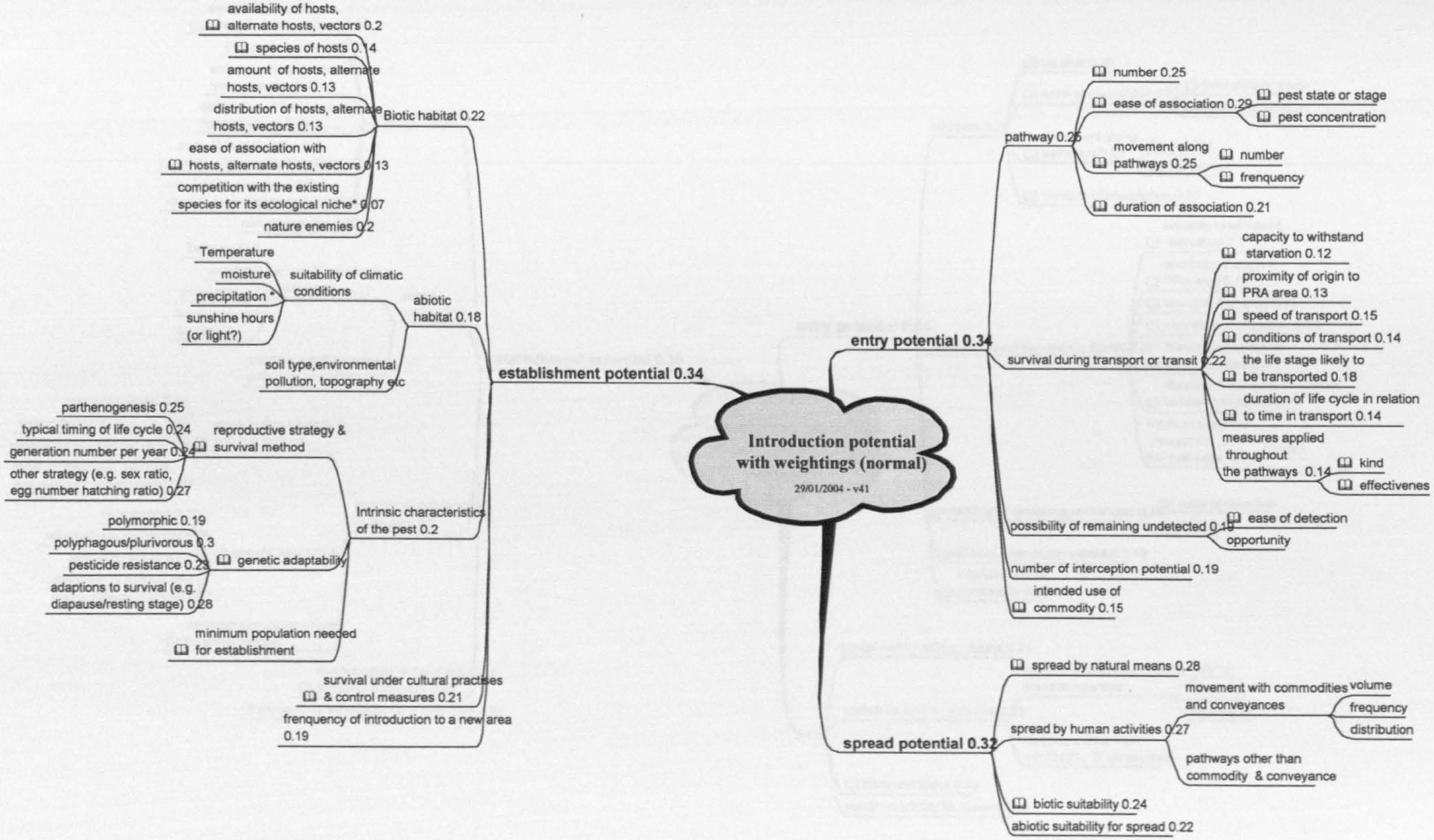


Figure A5.1 Weightings from linear transformation $Wr_i = (9 - rk_i)$

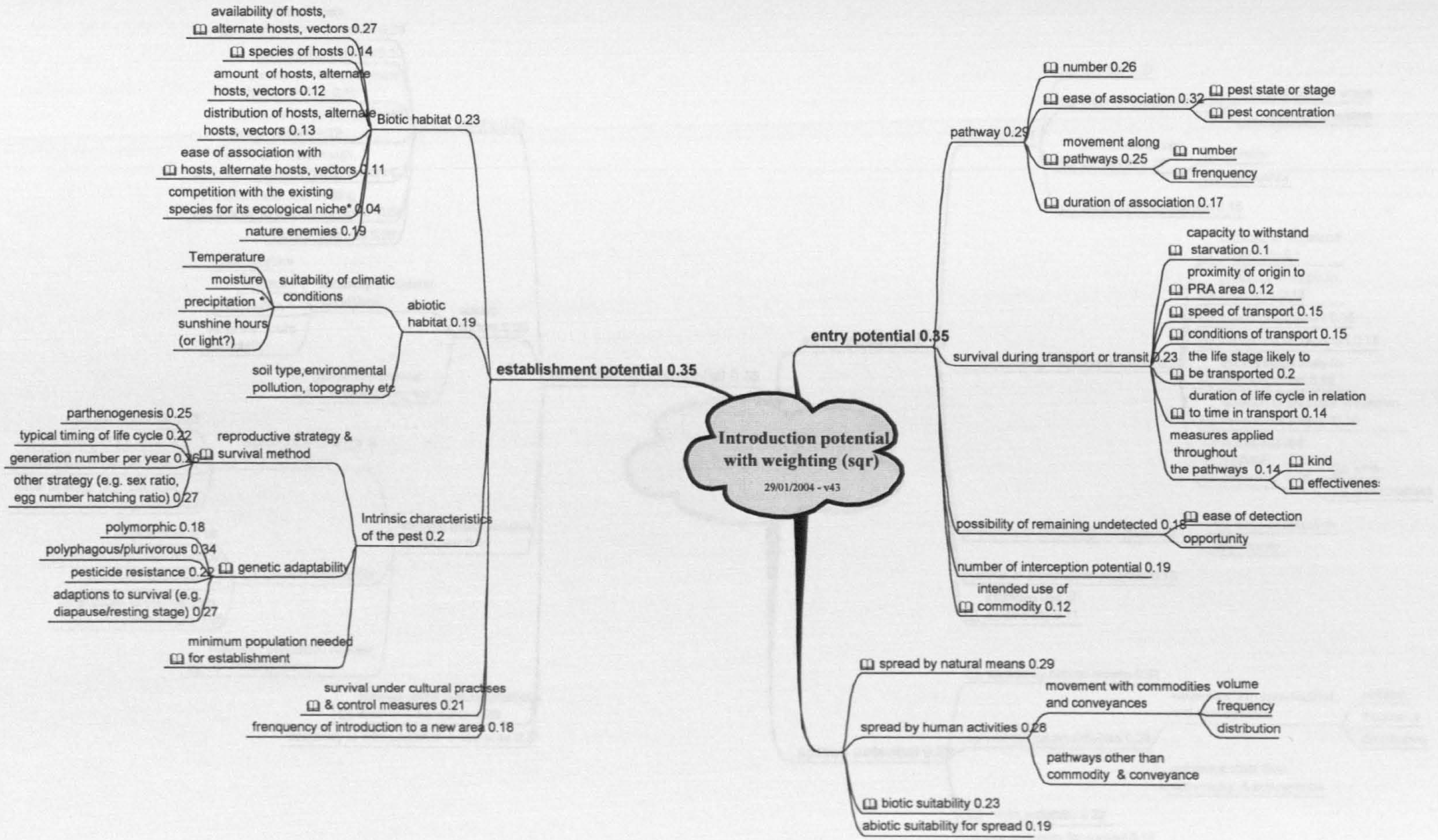


Figure A5.2 Weightings from squared transformation $Wr_i = (9 - rk_i)^2$

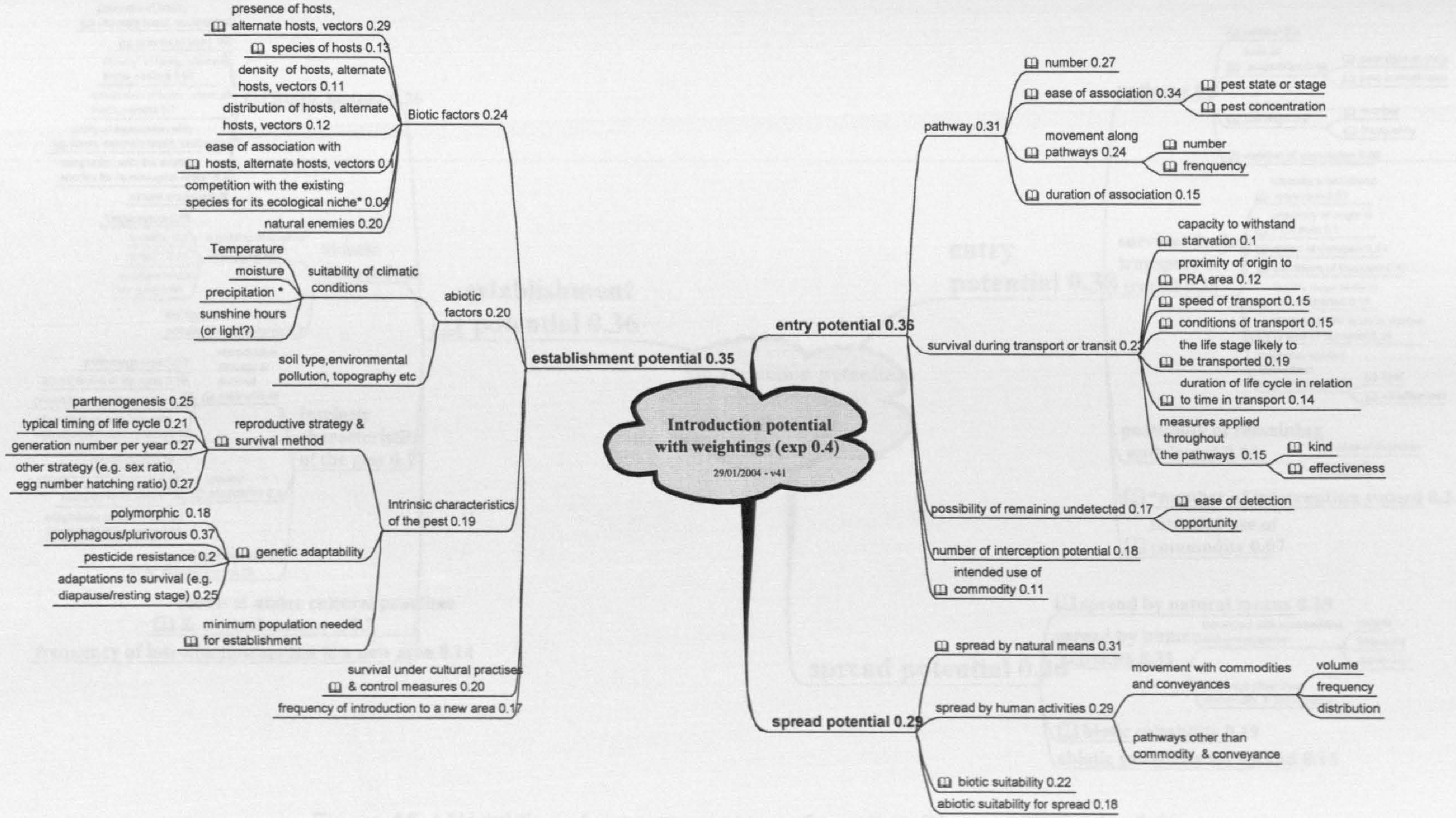


Figure A5.3 Weightings from exponent transformation $Wr_i = e^{(a(9-rk_i))}$ (a=0.4)

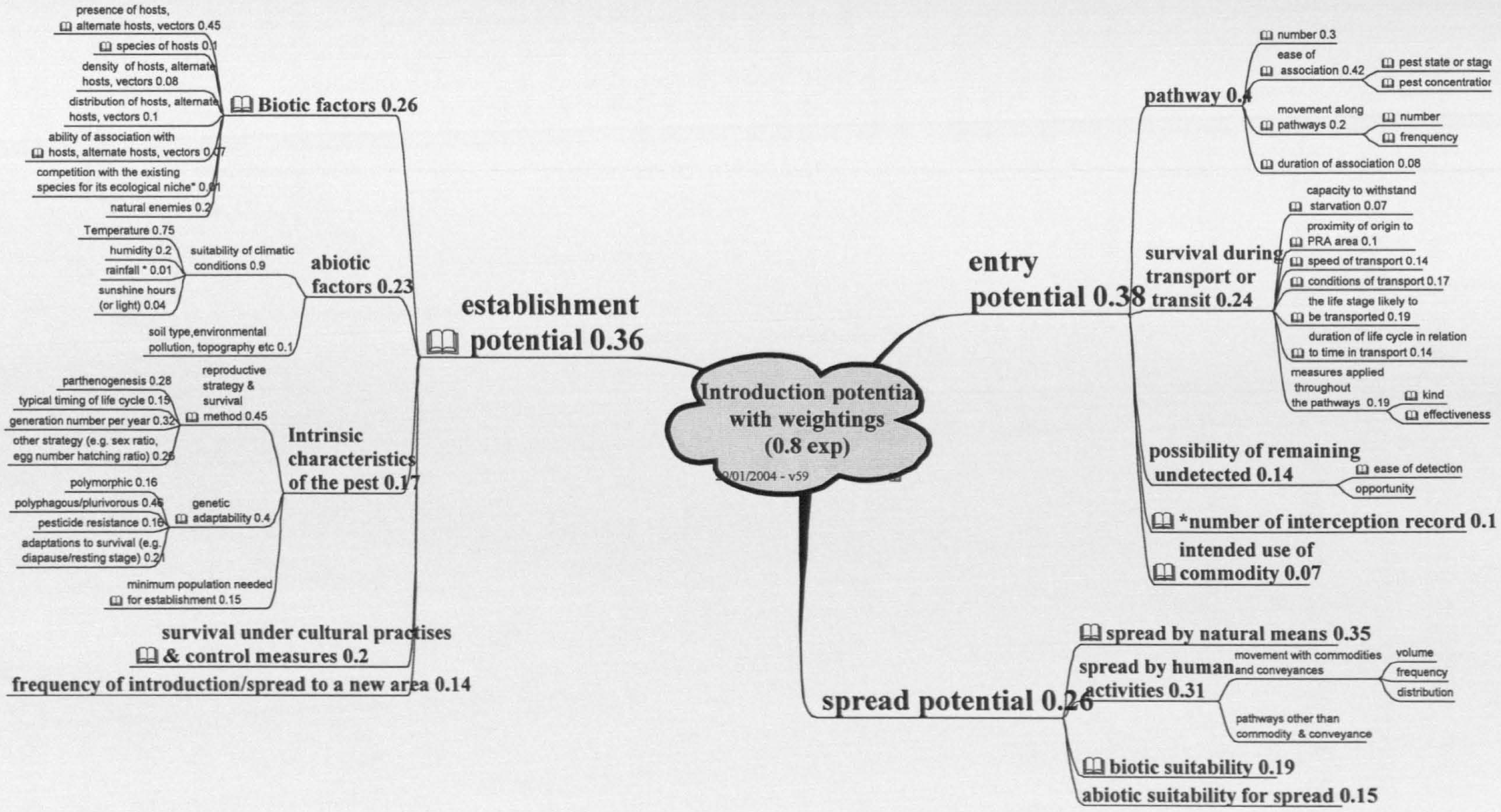


Figure A5.4 Weightings from exponent transformation $Wr_i = e^{(a(9-rk_i))}$ (a=0.8)

APPENDIX 6. EPPO FORESTRY PESTS DATA: RISK SCORES FOR QUESTIONS IN THE EPPO PEST RISK ASSESSMENT SCHEME (MCNAMARA 2002)

Table A6.1 EPPO PRA data: risk scores and quarantine status

Risk factor	Pest code													
	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1.1 Number of pathways	5	3	3	4	4	3	2	3	3	4	4	5	4	3
1.3b Likelihood of association with pathway	2	6	4	5	7	5	6	5	2	5	7	7	4	6
1.4 Concentration of the pest on the pathway	2	4	5	7	7	7	2	3	2	2	4	5	3	5
1.5b Likelihood of survival existing cultivation or commercial practices	7	8	7	5	7	7	8	8	5	7	7	7	8	8
1.6 Likelihood of survival or remain undetected	8	9	8	8	9	9	3	9	7	9	8	8	9	9
1.7b Likelihood of survival in transit	4	7	7	8	8	9	8	9	1	9	8	8	9	9
1.8 Likelihood of multiple during transit	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1.9 Movement along the pathway	7	4	7	4	6	3	3	8	7	7	7	7	7	7
1.10 Commodity distribution	8	7	5	4	6	8	5	7	7	8	8	8	8	8
1.11 Width of spread in time of commodity arrival	9	9	8	5	7	6	3	9	9	5	8	9	9	9
1.12b Likelihood of transfer to a suitable host	3	6	6	6	1	6	8	6	6	4	4	7	5	4
1.13 Intended use of Commodity	2	6	2	7	3	3	9	4	2	3	3	3	3	3
1.14 Host plant species	8	8	2	2	1	8	9	6	5	4	4	8	4	4
1.15 Host plant extension	8	8	6	8	6	8	9	8	8	8	8	8	8	8

1.19 Wild plant aids dispersal or maintenance		9		9	8	9	9	2	2	1	1	1	2	2
1.20 Similarity of climatic conditions	7	3	6	7	7	9	4	8	8	8	9	7	7	5
1.21 Similarity of other abiotic factors	9	5	7	7	7	7	8						7	6
1.22 Competition	8	8	9	8	6	6	5	5	8	9	9	8	5	5
1.23 Natural enemies	7	8	8	7	7	9	7	8	8	9	9	9	8	9
1.24Eenvironment aiding establishment	5		5	5	5	5						5		
1.25 Control measures for other pests	7	8	8	8	8	8	8	8	8	8	9	8	8	8
1.26 Reproductive strategy	3	3	3	5	2	6	5	3	2	6	4	4	3	6
1.27 Populations for establishment	4	6	6	3	3	6	5	6	2	5	3	3	6	6
1.28 Probability of eradication in the PRA area	9	8	8	8	5	9	8	7	8	9	9	9	8	8
1.29 Genetic adaptability	7	4	2	6	5	7	4	1	5	1	2	1	1	1
1.30 Introduction to a new area	4	4	1	3	1	1	1	1	1	1	1	1	5	1
2.1 Economic importance in existing area	9	6	8	6	7	6	7	6	6	8	8	4	6	6
2.2 Environmental damage in existing area		4	4	5	6	5			5	4	4	2	2	5
2.3 Social damage in existing area	8	6	7	5	5	4	2	2	2	4	4	1	3	3
2.4 Endangered area in the PRA area	6	3	6	5	4	6	3	7	4	7	9	7	7	6
2.5 Spread by natural means	7	3	5	5	1	5	7	5	7	3	3	6	7	5
2.6 Spread by human activities	6	6	6	6	7	7	6	5	5	6	6	5	6	6
2.7 Containment of spread in the PRA area	7	5	5	5	4	6	7	5	6	9	9	8	8	8
2.8 Direct effects on crop yield/quality	6	6	8	3	6	6	6	6	6	7	8	6	6	6
2.9 Effects on producer profits	6	3	6	4	5	5	5	3	4	5	6	6	5	5
2.10 Effects on consumer demand	4	3	6	4	5	5	5	3	3	2	2	2	5	5
2.11Efffects on export market	6	6	6	6	6	6	4	6	6	6	6	4	6	6
2.12 Other costs	5	2	3	5	5	5	3	2	2	3	5	3	3	3

2.13 Likelihood of environmental damage in the PRA area		4	4	4	5	5		4	4	6	7	2	2	5
2.14 Likelihood of social damage in the PRA area	6	3	3	3	3	4	2	4	4	6	7	1	2	2
2.15 Natural enemies	8	8	8	3	6	8	4	8	5	8	8	8	6	6
2.16 Ease of control	8	8	8	7	5	7	5	6	5	7	7		8	8
2.17 Disruption of control measures to other pests	5	2	3	5	5	3	2	2	4	2	2		2	2
2.18 Side-effects of control measures	5	1	2	4	4	4	6	2	2	4	4		2	2
2.19 Pest resistance	1	5		3	3	4	5		2	1	1	2	3	3

Blank cell in the table are factors not given a score in the assessment. They are adjusted by assigning a score 5 to keep them neutral.

Pest code:

B = *Dendrolimus sibiricus*

C= *Aeolesthes sarta*

D = *Xylotrechus altaicus*

E = *Scolytus morawitzi*

F = *Ips subelongatus*

G = *Dendroctonus ponderosae*

H = *Tetropium gracilicome*

I = *Ips hauseri*

J = *Dasychira albodentata*

K = *Erannis jacobsoni*

L = *Malacosoma parallela*

M = *Melanophila guttulata*

N = *Sphinx morio*

O = *Hylobius albosparsus*

APPENDIX 7. A CASE STUDY USING THE PROPOSED RISK ASSESSMENT METHODS FOR FALL WEBWORM (*HYPHANTRIA CUNEA*): ITS INTRODUCTION POTENTIAL

To determine some of the potential difficulties likely to be faced in PRI, a quarantine pest *Hyphantria cunea* (fall webworm), which the author had been working on for several years in China, was chosen as the first example to analyse in detail. The objective of this risk assessment on introduction exercise is to evaluate the introduction risk from Liaoning province (in the Northeast of China) to Beijing (the Capital of China).

The fall webworm, *Hyphantria cunea* (Drury), is a native of North America, and feeds on many species of deciduous forest, shade, and fruit trees. In China, it was first discovered in Dan Dong City, Liaoning Province in 1979. Since then it has spread to all the other cities in Liaoning province and several neighbouring provinces, such as Hebei Province, Tianjin, and Shandong province. The spread of fall webworm has brought significant economic and environmental damage to these areas. At the time of this PRA exercise, Beijing was still free from fall webworm. However, Beijing was located in the same geographical region, which has the ecosystem deemed to be suitable for the survival, establish, and spread of fall webworm, as Liaoning. Additionally, Beijing is only about 700km away from the capital city of Liaoning Province, which are connected by highways and railroad. There were also significant movements of commodities between Beijing and Liaoning.

This PRA exercise evaluated the introduction potential of fall webworm from Liaoning province to Beijing. It involved the following work:

- Identifying the risk factors of introduction potential by mind mapping;
- Subjectively assessing the risk scores (adopting 1-9 scale scoring system) at the most disaggregated level of risk factor structure in the mind map;
- Assigning weightings (0-1) to each of the risk factors at all levels according to their perceived importance.

The result of the pest risk assessment exercise is shown in Figure A7.1, which outlines the scores and weightings of the risk factors. As shown in Figure 3.10, the risk factors for the introduction potential of fall webworm to Beijing were structured as a hierarchy from general to specific. At the most general level, three factors were identified: entry potential, establishment potential, and spread potential. Each of these three risk factors was assigned a weighting, indicating the importance of each factor to the final risk assessment. Of the three general risk factors, entry potential and establishment potential were each given a weighting of 0.4, spread potential was given a weighting of 0.2.

Each of these general level risk factors was further divided into more specific risk factors. For example, establishment potential was further divided into biotic factors (including hosts, alternative hosts, and vectors), abiotic factors, biological characteristics of the pest, survival under cultural practise and control measure, and intended use of commodity. Each of these factors was also granted a weighting, indicating their importance to the assessment of establishment potential. Of these factors, biotic factors and biological characteristics of the pest were considered to be the most important risk factors contributing to establishment potential.

These factors, if necessary, were again further divided into specific risk factors. For example, biotic factors were further divided into six fundamental risk factors:

- Presence of hosts, alternative hosts, and vectors;
- Number of host species;
- Density and distribution of hosts, alternative hosts, and vectors;
- Ease of association with hosts, alternative hosts, and vectors;
- Competition with the existing species for its ecological niche.
- Natural enemies

Each fundamental factor was given both a risk score in a 1-9 scale and a weighting. For example, the presence of hosts, alternative hosts, and vectors was given a score between 7 and 9, and a weighting of 0.2; the number of host species was given a risk score of 9 and a weighting of 0.2, etc.

Using a weighted average and taking into account the hierarchical structure of the risk analysis process, an overall risk score for the introduction of fall webworm from Liaoning to Beijing was derived as between 5.7 and 7.2 using high and low biased weighted averaging, indicating the risk of introducing fall webworm to Beijing tend to be high or high medium[3]. Entry potential, establishment potential, and spread potential accounted for 41%-43%, 37%-41%, and 18%-19% of the overall risk score respectively. This does not depend on the number of questions in each category, although the number of questions within a category will affect

---

[3] Fall webworm is now reported in Beijing..

the weightings assigned to each question. Under this structure, each question within a category is assigned a weighting, such that the weightings sum to 1.

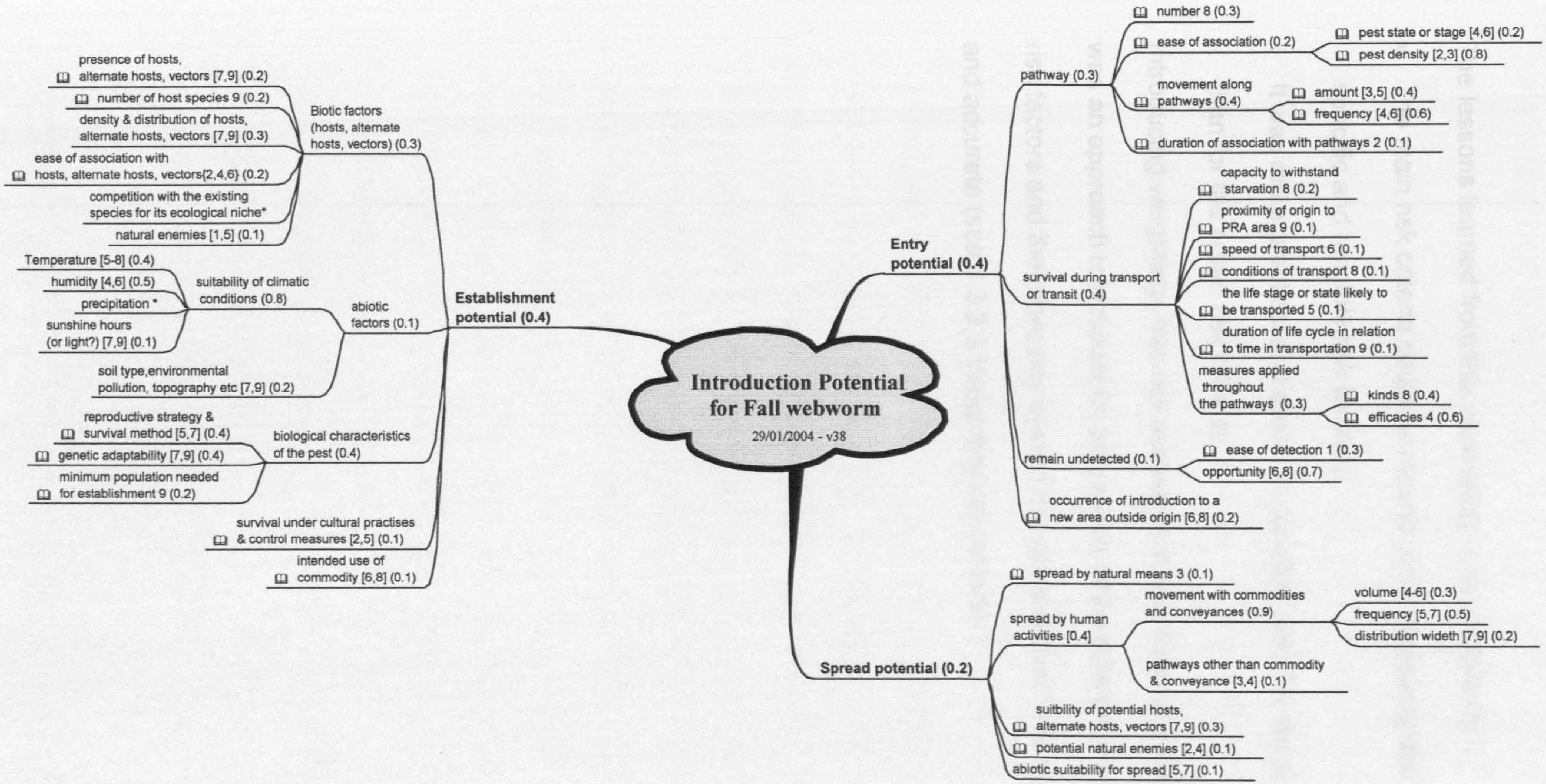


Figure 7.1 Introduction potential of fall webworm (*Hyphantria cunea* Drury)

The lessons learned from this case study were as follows:

- The main risk criteria could be meaningfully disaggregated into more specific and tangible risk factors;
- It was more practical and easier to consider risks for the specific factors than for the broad risk criteria;

Introducing weightings into risk assessment was a novel practice in PRA. It was an approach to account for perception of the different importance of the risk factors and therefore any overall risk value derived was more meaningful and accurate (see 3.3.3.3 Weighting risk factors).

APPENDIX 8. CORRELATION AND INTERACTION ANALYSES OF THE TANZANIAN DATA[4]

Correlation analysis and consideration of the likely interactions between risk factors suggested that:

- *Environmental impact* was the only risk factor that was not significantly correlated with others, which suggested that environmental impact was an independent factor. All other six risk factors were significantly correlated with each other
- Correlations between *climate suitability* and others (except *environmental impact*) were relatively low: only three significant correlations at 0.01 level and two significant correlations at 0.05 level, which suggested that it related less with others. *Climate suitability* may have had a direct effect on *host range*, *dispersal potential* and *economic impact*, but may not have had direct effect on pathway and trading partners.
- *Host range* was significantly correlated with all other risk factors except *environmental suitability*. The highest correlation was with *economic impact*, followed by *dispersal potential, trading partners, climate suitability* and *pathways. Host range* may have had a direct effect on all other correlated risk factors except *climate suitability.*
- *Dispersal potential* was significantly correlated with all other risk factors except *environmental impact*. The highest correlation was with *climate suitability,* closely followed by *host range,* trading *partners and pathway.*

---

[4] In this section, the correlations of the risk factors were analysed and the dependence or cause and effect were further examined. This was based on the author's judgments.

Dispersal potential may have had a direct effect only on *economic impact*, and may have been affected by all other correlated risk factors.

- *Economic impact* had relatively high significant correlation with all other risk factors except *environmental impact.* The three equally highest correlations were with *host range*, *dispersal potential* and *pathway*, followed by *trading partners* and *climate suitability. Economic impact* may have been directly affected by all other correlated risk factors.
- *Pathway* was less significantly correlated with *climate suitability.* It is significantly correlated with all other risk factors except *environmental impact.* The highest correlation was with *trading partners*, followed by *economic impact*, *dispersal potential* and *host range*. *Pathway* may have had a direct effect on *economic impact* and *dispersal potential*, and may have been directly affected by *host range* and *trading partners*. There may have been no direct interaction between *climate suitability* and *pathway*.
- *Trading partners* was less significantly correlated with *climate suitability.* It was significantly correlated with all other risk factors except *environmental impact.* The highest correlation was with *pathway*, followed by *economic impact*, *host range* and *dispersal potential*. *Trading partners* may have had a direct effect on *economic impact*, *pathway* and *dispersal potential*, and may have been directly affected by *host range*. There may have been no direct interaction between *climate suitability* and *trading partners*.
- Examining the correlations between overall risk and the risk factors, it was interesting to note that *economic impact* (0.74) had the biggest effect on the overall risk, followed by *host range* (0.68), *dispersal potential* (0.63), *pathway* (0.59) and *trading partners* (0.59), *climate suitability* (0.48),

environmental impact (0.27) appeared to have the least effect on the overall risk.

APPENDIX 9. CORRELATION AND INTERACTION ANALYSES OF THE EPPO DATA[5]

Correlation analysis and interaction analysis suggested that:

- *Intended use of commodity* was not correlated to any other risk factors and therefore is the most independent one.
- *Resistance to plant protection products* was another fairly independent factor that only less significantly negatively correlated to *climate similarity.*
- *Environmental damage in PRA area* was less significantly positively affected by *number of pathways* and *environmental damage in existing area.*
- *Surviving in transit* was slightly correlated to *effect of natural enemies on establishment* and *other costs due to introduction.*
- *Ease of being eradicated* was slightly significantly positively affected by *host species in PRA area* and *competition with existing species*; it was also slightly positively correlated to *social damage in existing area.*
- *Association with pathway* was slightly significantly positively affected by *pest concentration* and negatively by *wild plants aiding dispersal*; it also had a slight significant positive affect on *spread by human assistance in PRA area.*
- *Competition with existing species* had slightly significant positive effect on *ease of being eradicated* and *social damage in existing area*; it was also slightly significantly positively correlated to *social damage in PRA area.*

---

[5] In this section, the correlations of the risk factors were analysed and the dependence or cause and effect were further examined. This was based on the author's judgments.

- *Commodity distribution* was slightly significantly positively correlated to *duration of consignment arrival* and *movement along pathway* and negatively to *impact on control of other pests.*
- *Impact on control of other pests* was negatively affected by *spread by human assistance in PRA area*; it was also negatively slightly significantly correlated to *commodity distribution* and *duration of consignment arrival.*
- *Other side-effects* was positively affected by *economic loss in existing areas, direct effect on yield or quality in PRA area, effect on producer's profit and other costs due to introduction*; it was also positively yet less significantly correlated to *similarity of other abiotic factors*.

APPENDIX 10. CORRELATION ANALYSIS OF EPPO RISK FACTORS (WITH 15 PRA CASES)

RISK FACTORS RELATED TO ENTRY POTENTIAL

- *Number of pathway* had a significant positive effect on *economic loss in existing area* and *spread by human assistance in PRA area.*
- *Pest concentration* was negatively significantly correlated to *wild plants aiding establishment.*
- *Surviving existing control practices* was significantly correlated to *remaining undetected.*
- *Remaining undetected* was *positively significantly correlated to surviving existing control practices* and *effect of natural* enemies on spread.
- *Multiplying in transit* had a positive significant effect on *frequency of introduction* and *spread by human assistance in PRA area*; was positively significantly affected by *reproductive strategy, low population becoming established* and *genetic adaptability*; it was also positively correlated to *movement along pathways* and negatively correlated to *effect of natural enemies on spread.*
- *Movement along pathways* had a positive significant effect on *spread by human assistance in PRA area* and was positively significantly correlated to *multiplying in transit.*
- *Duration of consignment arrival* had a positive significant effect on *effect of natural enemies on establishment* and was negatively significantly correlated to *genetic adaptability.*

- *Transfer to suitable host* was positively significantly affected by *available host species* and *host extension in PRA area.*

RISK FACTORS RELATING TO ESTABLISHMENT POTENTIAL

- *Available host species in PRA area* had a positive significant effect on *host extension in PRA area* and *transfer to suitable host.*
- *Extension of host in PRA area* was positively significantly affected by *available host species in PRA area*, and had a positive significant effect on *transfer to suitable host.*
- *Wild plants aiding dispersal* had a negative significant effect on *pest concentration* and *direct effect on yield and quality in PRA area*; it was also negatively significantly correlated to *effect on producer profit.*
- *Climate suitability* had a positive significant effect on *extension of endangered area* (0.79); it also had a negative significant effect on *effect of natural enemies on establishment.*
- *Similarity of other abiotic factors* had a positive significant effect on *ease of containing of spread.*
- *Effect of natural enemy on establishment* was positively significantly affected by *duration of consignment arrival,* and was negatively significantly affected *climate suitability* and *genetic adaptability.*
- *Existing control measures* had a negative significant effect on *spread by natural means in PRA area* and *spread by human assistance in PRA area;* it was positively significantly affected by *effect of natural enemies on spread* and was negatively significantly affected by *genetic adaptability.*

- *Reproductive strategy* had a positive significant effect on *multiplying in transit* and *spread by human assistance in PRA area;* it had a negative significant effect on *effect of natural enemies on spread*; it was also positively significantly correlated to *genetic adaptability.*
- *Low population being established* had a positive effect on *multiplying in transit.*
- *Genetic adaptability* had a positive significant effect on *multiplying in transit, spread by natural means in PRA area* and *spread by human assistance in PRA area*; it had a negative significant effect on *effect of natural enemies on establishment, effect of natural enemies on spread* and *surviving existing control measures;* it was also positively significantly correlated to *reproductive strategy* and was negatively significantly correlated to *duration of consignment arrival.*
- *Introduction frequency* was positively significantly affected by *multiplying in transit.*

RISK FACTORS RELATING TO CONSEQUENCES OF INTRODUCTION

- *Economic loss in existing area* was positively significantly affected by *number of pathways*; had a positive significant effect on *direct effect on yield or quality in PRA area, effect on producer profit* and *other cost due to introduction*; it was also positively significantly correlated to *spread by human assistance in PRA area.*
- *Environmental damage in existing area* was negatively significantly correlated to *effect on consumer demands.*

- *Social damage in existing area* had a positive significant effect on *social damage in PRA area.*
- *Extension of endangered area* was positively significantly affected by *climate suitability.*
- *Spread by natural means in PRA area* was positively significantly affected by *genetic adaptability* and negatively significantly affected by *existing control measures.*
- *Spread by human assistance in PRA area* was positively significantly affected by *number of pathways, multiplying in transit, movement along pathway, reproductive strategy* and *genetic adaptability*; it was negatively significantly affected by *existing control measures*; it was also positively significantly correlated to *economic loss in existing areas* and *effect on consumer demands.*
- *Ease of containing of spread* was positively significantly affected by *similarity of other abiotic factors.*
- *Direct effect on yield and quality in PRA area* was positively significantly affected by *economic loss in existing areas* and had a positive significant effect on *effect on producer's profits*; it was also negatively significantly correlated to *wild plants aiding dispersal.*
- *Effect on producer profit* was positively significantly affected by *economic loss in existing areas, direct effect on yield or quality in PRA area* and *effect on consumer demands*; it was also negatively significantly correlated to *wild plants aiding dispersal.*
- *Effect on consumer demands* had a positive significant effect on *effect on producer's profit* and *social damage in PRA area*; it was positively

significantly correlated to *spread by human activities* and was negatively correlated to *environmental damage in existing area.*

- *Effect on export market* was positively significantly affected by *other costs due to introduction* and *ease of being controlled.*
- *Other costs due to introduction* was positively significantly affected by *economic loss in exiting areas* and had a positive significant effect on *effect on export market.*
- *Social damage in PRA area* was positively significantly affected by *social damage in existing areas* and *effect on consumer's demands.*
- *Effect of natural enemies on spread* had a positive significant effect on *surviving existing control measures in PRA area*; was negatively significantly affected by *reproductive strategy* and *genetic adaptability*; it was also positively significantly correlated to *remaining undetected* and was negatively correlated to multiplying in transit. *Ease of being controlled* had a positive significant effect on *effect on export market.*

APPENDIX 11. MATHEMATICS OF PRINCIPAL COMPONENTS ANALYSIS

MATHEMATICS OF PRINCIPAL COMPONENT ANALYSIS

Suppose there are M potential quarantine pests, each with N risk factors. The pests risk attributes can be organised as follow:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_N \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{pmatrix}$$

where x_{ij} denotes the risk scores given by the risk assessors for the pest j in regard to the *ith* risk factors. PCA seeks to find a set of new variables $Y = (Y_1 \quad Y_2 \quad \ldots \quad Y_P)^T$, which are linear combinations of X as follows:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_P \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2N} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \alpha_{P1} & \alpha_{P2} & \cdots & \alpha_{PN} \end{pmatrix} \times \begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_N \end{pmatrix}$$

where α_{ij} is the weight value that reflects the contribution of X_j to Y_i, the *ith* principal component, satisfying:

$$\sum_{i=1}^{N} \alpha_{ij}\alpha_{jk} = 0 \qquad j \neq k$$

$$\sum_{i=1}^{N} \alpha_{ij}\alpha_{jk} = 1 \qquad j = k$$

α_{ij} is also termed factor l's.

As Principal Components are linear combinations of variables, the coefficient of each variable is scoring coefficient on the principal components.

The new variables $Y = (Y_1 \quad Y_2 \quad \ldots \quad Y_P)^T$ [6] are themselves uncorrelated, but retain maximally the variance of observations. This linear combination can be found by solving the following eigensystem subject to the above constraints:

$$(C - \lambda I)A = 0$$

where λ are the eigenvalues and $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$, C is the covariance matrix, and A are the eigenvectors.

The ratio of variance explained by the first q principal components can be expressed as

$$R_q^2 = \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_q}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \qquad q \leq p$$

---

[6] T: transpose

The first Principal Component accounts for the largest share of the total variance, the second principal component accounts for the largest share of the remained variances, and so on.

HOW MANY PRINCIPAL COMPONENTS (PCS) TO RETAIN

The use of more components increases the model's explanatory power, but does not achieve model simplification. In contrast, using or choosing fewer components results in reduced explanatory power for the model. In deciding how many factors to retain and extract, there are generally three tests. The first test is the scree test. The scree test is a graphic method for determining the number of factors. The eigenvalues are plotted in the sequence of the principal factors. The number of factors is chosen where the plot levels off to a linear decreasing pattern. The second test is proposed by Everitt and Dunn (1992) suggesting to discard all components accounting for less than $(70/n)\%$ of the overall variance, where n is the number of PCs. The third test is proposed by Hotelling (1933) suggesting to keep the first few PCs that explains more than 85% of the total variance.

FACTOR LOADINGS

Factor loading is a term used to refer to factor pattern coefficients or structure coefficient, which multiply with PCs to produce measured variables. Furthermore, it represents the correlations between the original variables and the new principal components. Mathematically, it can be shown as below:

$$X=\begin{pmatrix}X_1\\X_2\\ \cdot\\ \cdot\\ \cdot\\X_N\end{pmatrix}=\begin{pmatrix}\alpha_{11}&\alpha_{12}&\dots&\alpha_{1N}\\ \alpha_{21}&\alpha_{22}&\dots&\alpha_{2N}\\ \cdot&&&\\ \cdot&&&\\ \cdot&&&\\ \alpha_{P1}&\alpha_{P2}&\dots&\alpha_{PN}\end{pmatrix}^{-1}\times\begin{pmatrix}Y_1\\Y_2\\ \cdot\\ \cdot\\ \cdot\\Y_P\end{pmatrix}=\begin{bmatrix}\beta_{11}&\beta_{12}&\dots&\beta_{1P}\\ \beta_{21}&\beta_{22}&\dots&\beta_{2P}\\ \cdot&&&\\ \cdot&&&\\ \cdot&&&\\ \beta_{N1}&\beta_{N2}&\dots&\beta_{NP}\end{bmatrix}\times\begin{bmatrix}Y_1\\Y_2\\ \cdot\\ \cdot\\ \cdot\\Y_P\end{bmatrix}$$

where β_{ij} is the factor loading or pattern coefficient for factor X_i on principal component factor Y_j.

The new axes, or dimensions, are uncorrelated with each other, and are selected according to the amount of the total variance that they describe. Normally this results in there being a few large axes accounting for most of the total variance, and a large number of small axes accounting for very small amounts of the total variance. These small axes are normally discounted from further consideration, so that the data set having P correlated variables has been transformed to a data set having N uncorrelated axes, or principal components, where N is usually considerably less than P.

The fact that the m axes are uncorrelated is often a very useful property if further analysis is planned. The fact that N is less than P introduces a parsimony that is more often than not describable in any scientific discussion. But much attention focuses on the relationship of the principal components to the original variables - which of the original axes contributed the largest variance to each of the principal components?