Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- you will use the copy only for the purposes of research or private study
- you will recognise the author's right to be identified as the author of the thesis and due acknowledgement will be made to the author where appropriate
- you will obtain the author's permission before publishing any material from the thesis.
Ensemble models to assess the risk of exotic plant pathogens in a changing climate

A thesis
submitted in partial fulfilment
of the requirements for the Degree of
PhD

at
Lincoln University
by
Hossein Ali Narouei Khandan

Lincoln University
2014
Abstract of a thesis submitted in partial fulfilment of the requirements for the Degree of PhD in ecological modelling.

Abstract

Ensemble models to assess the risk of exotic plant pathogens in a changing climate

by

Hossein Ali Narouei Khandan

In recent decades, species distribution models (SDMs) have been widely used in many ecological, environmental and climate-change research studies to model invasive species establishment. These models associate recorded locations of species with environmental variables. Nevertheless, the few studies that attempt to model the climate suitability of plant pathogens before their arrival into a new area mainly rely on a single model projection. In this research, eleven species distribution models (in the form of three modelling approaches which include correlative and mechanistic models) were used to project the climate suitability of three target species; kiwifruit bacterial canker \((Pseudomonas syringae\ \text{pv.}\ \text{actinidia})\) (Psa), dwarf bunt of wheat \((Tellitia controversa)\) and guava rust \((Puccinia psidii)\) for New Zealand and over a global scale. The climate suitability of target species was modelled using CLIMEX as a semi-mechanistic model, MaxEnt as a presence-only correlative model and Multi-Model Framework (which includes nine correlative models). While there were similarities with regard to climate suitability for target species projected by the models over both local and global scales, there were differences in their projection with respect to the degree and extent of suitability, making it hard to select one “best” model.

All models were found to have their differences and weaknesses that are largely the result of difference in the theoretical basis and structure of each model. For example, compared with CLIMEX and the Multi-Model Framework, MaxEnt showed lower transferability of projection into new areas. Additionally, as a semi-mechanistic model, the uncertainty of CLIMEX projections was found to be increased by subjectivity in the parameter setting process. To illustrate the impact of parameter variability on the uncertainty of CLIMEX projections, a sensitivity analysis was performed on one of the target species (dwarf bunt) to
measure the effect of error in important parameters on model output. The sensitivity analysis showed that for dwarf bunt, CLIMEX outputs were very sensitive to upper temperature threshold and soil moisture parameters, which highlight that sensitivity analysis, should be an integral part of any CLIMEX modelling. For Multi-Model, despite the advantages such as calculating different performance criteria, the importance and contribution of selected variables and their influence on model output is not given.

Because of differences in model projection, a method was developed to benefit from the information provided by all the types of models, by combining the results of different model output into an ensemble, or more specifically, a consensus model. A variant of committee averaging was used where model outputs are converted to binary maps (presence- absence) which allow any kind of algorithm and output to be included. The resulting consensus model highlighted the areas where more than half of the models agreed on the climate suitability for target species establishment. Such a model that relies on agreement of model projections indicates with a level of certainty or uncertainty what is likely to happen and consequently can highlight areas, both locally and globally, that have a higher risk of target species establishment. Finally, the effect of climate change on climate suitability of target species was investigated using two scenarios (A1B and A2) for 2030 and 2090. The results showed that, the suitable areas decreased for Psa and dwarf bunt at different levels while guava rust suitability increased.

The results of this thesis confirm that models with different theoretical foundation will give dissimilar predictions, and it is difficult to determine conclusively whether one model is superior to others. Among other recommendations, I strongly advise that researchers and risk assessors should not rely on a single-model projection. If time and resources are available, an appropriate ensemble of models should be used to investigate the climate suitability of plant pathogens.

**Keywords:** Plant pathogens, kiwifruit bacterial canker (Psa), dwarf bunt, guava rust, climate suitability, Species distribution models (SDMs), CLIMEX, MaxEnt, Multi-Model Framework, correlative models, semi-mechanistic models, sensitivity analysis, consensus model, ensemble models, climate change, range expansion.
Acknowledgements

Undoubtedly, without full support of my main supervisor Associate Professor Susan Worner, I could not be able to do this PhD research. I would like to express my special appreciation to Susan for giving me and my family the opportunity to live more than three wonderful years in New Zealand and explore its culture and beautiful nature. I learnt many positive attitudes from you that doing a research is only one of them. You are an incredible person, and I owe you my whole life. I would also like to thank my wonderful co-supervisor Dr. Eirian Jones. Thank you Eirian for quick and detailed review of my writings, motivating me to be involved in conferences, applying for conference grants and your advices about childcare! You were there whenever I needed your advice. I also would like to thanks Dr. Suvi Viljanen-Rollinson my co-supervisor for her practical comments about target species. Thanks you very much Suvi for being a kind superviaor and giving me precious comments on guava rust, dwarf bunt and Psa despite being very busy, I really appreciate it.

My gratitude also goes to B3 (Better Border Biosecuritry) organisation that financially supported me more than three years to carry on this research.

In addition to the supervisory team, I got valuable advice and technical help from my officemate Senait Senay; it was a pleasure to share the office with you Senait and talking about SDMs, big bang, cosmology, politics and future of the world! Statistical advice in this research was provided by Tak Ikeda and David Saville. Thank you very much Tak Ikeda for helping me with R script and replying to my emails (and messages!). Parts of meteorological data for this research were acquired by NIWA, and I would like to thank Andrew Tait for providing the climate data; I haven’t met you Andrew, but you are truly wonderful. The new CLIMEX data were used with help of Darren Kriticos, thanks Darren for your remote help and unique figures of speech!

The species data in this thesis were acquired through email from more than 50 scientists from more than 20 counties that I would like to thank you all. I also would like to thank David Logan, Carlo van den Dijssel and Dr. Joel Vanneste and Dr. Rob Beresford from Plant
& Food Research and ZESPRI® International Limited and Giorgio Balestra from University of Tuscia, Italy or providing information regarding kiwifruit.

In addition, I like to acknowledge my friends Tasha Van Der Shelby, Jeniffer Pannel, Federico Tomassetto, Senait Senay, who supported in writing this thesis, and our lab meeting team Audrey Lusting, Mariona Roige, Ursula Torres, Maria Del Mar Rovira Capdevila and Cian O’Connor for their comments and advice. I would like to thank personnel of Bio-Protection Research centre who, in one way or another, helped me through this project: John Hampton, Andrew Holyoake, Richard Falloon, Susan Paterson, Eunice Wheal, Dianne Fyfe and everyone else that I may have forgotten.

Special thanks to my parents, sister and brothers who supported me mentally during my study. Words cannot express how grateful I am for having you all. And finally, most of all, I would like to express my deepest appreciation to my beloved wife Mozhgan and our son Arta for their endless love, and all the sacrifices that they have made on my behalf. Thank you Mozhgan for being with me all these years and supporting my every moment, I would be grateful forever.

با سیاس از همه
Table of Contents

Abstract ...................................................................................................................................... ii
Acknowledgements .................................................................................................................. iv
Table of Contents ..................................................................................................................... vi
List of Tables ............................................................................................................................. x
List of Figures ........................................................................................................................... xi

1  Chapter 1 General introduction .................................................................................... 15
   1.1 The ecological and economic importance of invasive species ......................... 15
   1.2 Pest and pathogen dispersal ................................................................................. 16
   1.3 Species distribution models (SDMs) ................................................................. 20
       1.3.1 Types of SDMs............................................................................................ 21
       1.3.2 Application of SDMs in plant pathology ................................................. 23
   1.4 Evidence of pest and pathogen invasion in New Zealand ................................... 25
   1.5 Overall objective of the thesis ............................................................................ 26
   1.6 Research Specific objectives ............................................................................. 27
   1.7 Thesis structure: ................................................................................................. 27

2  Chapter 2 Methodology ................................................................................................ 29
   2.1 Models .................................................................................................................. 29
       2.1.1 CLIMEX ...................................................................................................... 29
       2.1.2 Correlative models ...................................................................................... 33
   2.2 Data collection ..................................................................................................... 41
       2.2.1 Occurrences Data ........................................................................................ 41
       2.2.2 Climate data ............................................................................................... 42
   2.3 Further analysis using Principal Component Analysis. ..................................... 43

3  Chapter 3 Potential climate suitability for kiwifruit bacterial canker ....................... 44
   3.1 Bacterial Canker of Kiwifruit (Pseudomonas syringae pv. actinidia) ............... 44
       3.1.1 The disease ................................................................................................ 44
       3.1.2 Global Distribution .................................................................................... 45
       3.1.3 Host range .................................................................................................. 48
       3.1.4 Epidemiology ............................................................................................. 48
       3.1.5 Risk to New Zealand ............................................................................... 48
       3.1.6 Previous modelling of Psa ....................................................................... 49
   3.2 Methods ............................................................................................................... 49
       3.2.1 Psa occurrence data .................................................................................. 49
       3.2.2 CLIMEX .................................................................................................... 49
       3.2.3 Correlative models .................................................................................... 50
6 Chapter 6 Characterising model uncertainty by sensitivity analysis: the effect of interacting parameter error .......................................................... 135
   6.1 Introduction .................................................................................. 135
   6.1.1 The CLIMEX model and parameter values uncertainty ....... 136
   6.2 Methods ....................................................................................... 137
       6.2.1 Software and data ................................................................ 137
       6.2.2 Sensitivity Analysis ............................................................... 139
   6.3 Results .......................................................................................... 142
       6.3.1 Sensitivity analysis results .................................................... 142
       6.3.2 Interaction of sensitive parameters ...................................... 143
   6.4 Discussion .................................................................................... 156
       6.4.1 Presence data ........................................................................ 156
       6.4.2 Sensitivity analysis ............................................................... 156
       6.4.3 Interaction of sensitive parameters ...................................... 158
   6.5 Conclusion ................................................................................... 159

7 Chapter 7 Model Consensus .................................................................. 161
   7.1 Introduction .................................................................................. 161
   7.1.1 Selection of cut-off threshold ................................................... 163
   7.1.2 The selection of top-performing models in the Multi-Model Framework ...... 164
   7.2 Results .......................................................................................... 167
       7.2.1 Psa consensus model ............................................................ 167
       7.2.2 Dwarf bunt consensus model .............................................. 167
       7.2.3 Guava rust consensus model ... 168
   7.1.1 Evaluation of goodness of fit of consensus model versus individual models . 173
   7.2 Discussion .................................................................................... 173

8 Chapter 8 Effect of climate change on projected potential distributions of the target species ................................................................. 177
   8.1 Climate change ............................................................................. 177
       8.1.1 Emission scenarios ............................................................... 178
   8.2 Agriculture and climate change ................................................... 180
   8.3 Climate-change effects on plant pathogens ............................... 181
   8.4 Methods ....................................................................................... 185
   8.5 Results .......................................................................................... 187
       8.5.1 Psa ....................................................................................... 187
       8.5.2 Dwarf Bunt .......................................................................... 187
       8.5.3 Guava rust ............................................................................ 188
   8.6 Discussion .................................................................................... 202
   8.7 Conclusion ................................................................................... 204

Chapter 9 General discussion .................................................................. 205
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1</td>
<td>Key findings</td>
<td>205</td>
</tr>
<tr>
<td>9.2</td>
<td>Identify the current and potential global and national distributions of each of the target species using different modelling approaches.</td>
<td>206</td>
</tr>
<tr>
<td>9.2.1</td>
<td>Projection of Psa global distribution</td>
<td>206</td>
</tr>
<tr>
<td>9.2.2</td>
<td>Projection of Dwarf bunt global distribution</td>
<td>208</td>
</tr>
<tr>
<td>9.2.3</td>
<td>Projection of guava rust global distribution</td>
<td>210</td>
</tr>
<tr>
<td>9.3</td>
<td>Sensitivity of CLIMEX model</td>
<td>211</td>
</tr>
<tr>
<td>9.4</td>
<td>Building a consensus model</td>
<td>212</td>
</tr>
<tr>
<td>9.5</td>
<td>The consequences of climate change on the target species</td>
<td>213</td>
</tr>
<tr>
<td>9.6</td>
<td>Summary</td>
<td>215</td>
</tr>
<tr>
<td>9.7</td>
<td>Bioclimatic models: assumptions and limitations</td>
<td>216</td>
</tr>
<tr>
<td>9.7.1</td>
<td>Data and data sources</td>
<td>218</td>
</tr>
<tr>
<td>9.7.2</td>
<td>Predictor variables</td>
<td>219</td>
</tr>
<tr>
<td>9.8</td>
<td>Specific limitations of the models in the current study</td>
<td>219</td>
</tr>
<tr>
<td>9.9</td>
<td>Future work</td>
<td>221</td>
</tr>
<tr>
<td>9.10</td>
<td>Concluding remarks</td>
<td>223</td>
</tr>
<tr>
<td>9</td>
<td>References</td>
<td>225</td>
</tr>
<tr>
<td>10</td>
<td>Appendices</td>
<td>250</td>
</tr>
<tr>
<td>A</td>
<td>Application of genetic algorithm function of CLIMEX</td>
<td>251</td>
</tr>
<tr>
<td>B</td>
<td>List of the 19 BIOCLIM variables used in this study</td>
<td>255</td>
</tr>
<tr>
<td>C</td>
<td>Potential climate suitability of kiwifruit bacterial canker</td>
<td>256</td>
</tr>
<tr>
<td></td>
<td>Supplementary materials</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Potential climate suitability of dwarf bunt</td>
<td>261</td>
</tr>
<tr>
<td></td>
<td>supplementary materials</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Potential climate suitability of Guava rust</td>
<td>267</td>
</tr>
<tr>
<td>F</td>
<td>Predicting the potential global distribution of Pseudomonas syringae p.v. actinidiae (Psa)</td>
<td>273</td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1. Confusion matrix .............................................................................................................................. 40
Table 3.1. Parameter values used in developing CLIMEX model for Psa. .......................................................... 54
Table 3.2 The accuracy measures achieved by cross-validation (a) and bootstrapping (b) in Multi-Model framework developed for Psa.......................................................... 66
Table 4.1. Parameter values used to develop a CLIMEX model for dwarf bunt. ........................................ 86
Table 4.2. The accuracy measures achieved by cross-validation (A) and bootstrapping (B) in Multi-Model Framework developed for dwarf bunt.............................................................................. 98
Table 5.1. New Zealand indigenous species of Myrtaceae (Poole & Adams, 1994). .................. 110
Table 5.2. Parameter values used in developing CLIMEX model for Guava rust. ........................................ 116
Table 5.3. The accuracy measures achieved by cross-validation (A) and bootstrapping (B) in Multi-Model Framework developed for dwarf bunt. .......................................................... 128
Table 6.1. A summary of Parameter values suggested in other studies and parameter values used in CLIMEX modeling in this study. .................................................................................................. 138
Table 6.2. Parameter values used for sensitivity analysis of CLIMEX model on dwarf bunt. ................. 141
Table 7.1. The top five models in Multi-Model Framework (MMF) with highest AUC selected for making the consensus model from Chapter 3, 4, 5). ........................................................................ 165
Table 8.1. The best estimate and the likely range of temperature increases based on different scenario projections for 2090-2099 ............................................................................................................ 179
Table 8.2. Percentage of change in climate suitability (range expansion or range contraction) of three target species under two future climate scenarios. ........................................................................ 201

Table C. 1. Psa presence points coordinates (after removal of duplicate points). ......................... 256
Table D. 1. Dwarf bunt presence points coordinates (after removal of duplicate points). ............... 261
Table E. 1. Guava Rust presence points coordinates (after removal of duplicate points). ............... 267
List of Figures

Figure 1.1. Dispersal of fungal pathogens................................................................. 19
Figure 1.2. Two main long-distance dispersal pathways in Europe............................ 20
Figure 2.1. CLIMEX temperature index parameters.................................................... 30
Figure 2.2. The ROC curve showing an ideal accuracy (A), a typical accuracy (B) which is better than a random model and C shows a random model. ......................................................... 41
Figure 3.1. The symptoms of Psa on leaf and cane of kiwifruit plant (top-right)........... 45
Figure 3.2. Current global distribution of Psa.............................................................. 47
Figure 3.3. Change in variable importance versus distance. ........................................ 51
Figure 3.4. Climate suitability of Psa on a global scale and New Zealand projected by CLIMEX. .... 56
Figure 3.5. Weekly growth index and climate variables at Te Puke for Psa.................... 57
Figure 3.6. The Area under the Curve (AUC) of MaxEnt model. The black diagonal line indicates a random model. The chart shows MaxEnt has performed better than random.......................... 58
Figure 3.7. Jackknife of MaxEnt model on (a) regularized training data (b) jackknife on test data (c) jackknife of AUC......................................................................................................................... 59
Figure 3.8. The effect of response curves of (a) annual mean temperature (cbio01) on top and (b) annual precipitation (cbio12) at the bottom in MaxEnt model for Psa........................................ 61
Figure 3.9. The potential climate suitability of Psa on a global and New Zealand scale projected by MaxEnt .......................................................................................................................... 63
Figure 3.10. The potential climate suitability of Psa on a global and New Zealand scale projected by Multi-Model.................................................................................................................. 65
Figure 3.11. Psa occurrences plotted against the first two principal components of the environmental data extracted at locations where Psa is present. ......................................................... 68
Figure 4.1 Dwarf bunt Disease cycle. This is the disease cycle as it is the interaction of the pathogens’ life cycle and the host plant. Source: (Wilcoxson et al. 1996) ................................................................. 74
Figure 4.2. The effect of dwarf bunt on wheat. (A)......................................................... 74
Figure 4.3. Current global distribution of dwarf bunt...................................................... 76
Figure 4.4. Changes in the important variable (annual precipitation) over the distance. .......... 81
Figure 4.5. Principal component analysis on dwarf bunt presence points. The closer the points higher the climatic similarity .............................................................................................................. 84
Figure 4.6. Global potential climate suitability of dwarf bunt projected by CLIMEX model. ........ 88
Figure 4.7. Temperature index, soil moisture index and weekly growth index for dwarf bunt in Central Otago. ........................................................................................................................................ 90
Figure 6.7. The world map showing the sensitivity of model and Europe region to -25% changes of DV3 and SM3 values at the same time ................................................................. 152
Figure 6.8. The world map showing the sensitivity of model and Europe region to +25% changes of DV3 and -25% changes of SM3 values at the same time .................................................. 153
Figure 6.9. The world map showing the sensitivity of model and Europe region to -25% changes of DV3 and +25% changes of SM3 values at the same time.................................................. 154
Figure 6.10. Sensitive parameters and their interactions ................................................................. 155
Figure 7.1. Consensus model projection for Psa. The median map shows the areas in which at least half of the models agree (light green) on climate suitability for the target species .................. 169
Figure 7.2. Consensus model projection for dwarf bunt. The median map shows the areas in which at least half of the models agree (light green) on climate suitability for the target species .............. 170
Figure 7.3. Consensus model projection for guava rust. The median map shows the areas in which at least half of the models agree (light green) on climate suitability for the target species ............. 171
Figure 7.4. Comparison of consensus model with single-models based on seven performance criteria ........................................................................................................................................ 172
Figure 8.1. The main characteristics of the storylines and scenario families...................................... 179
Figure 8.2. A summary of the major influences due to changes in atmospheric composition, land use and climate on plant diseases and the level of understanding of disease and climate .......... 183
Figure 8.3. The diagram shows the process of building consensus model for each target species ...... 186
Figure 8.4. The potential climate suitability for Psa based on Scenario A1B in 2030 ......................... 189
Figure 8.5. The potential climate suitability for Psa based on Scenario A1B in 2090 ...................... 190
Figure 8.6. The potential climate suitability for Psa based on Scenario A2 in 2030 ....................... 191
Figure 8.7. The potential climate suitability for Psa based on Scenario A2 in 2090 ....................... 192
Figure 8.8. The potential climate suitability for dwarf bunt based on Scenario A1B in 2030 .......... 193
Figure 8.9. The potential climate suitability for dwarf bunt based on Scenario A1B in 2090 .......... 194
Figure 8.10. The potential climate suitability for dwarf bunt based on Scenario A2 in 2030 .......... 195
Figure 8.11. The potential climate suitability for dwarf bunt based on Scenario A2 in 2090 .......... 196
Figure 8.12. The potential climate suitability for guava rust based on Scenario A1B in 2030 .......... 197
Figure 8.13. The potential climate suitability for guava rust based on Scenario A1B in 2090 .......... 198
Figure 8.14. The potential climate suitability for guava rust based on Scenario A2 in 2030 .......... 199
Figure 8.15. The potential climate suitability for guava rust based on Scenario A2 in 2090 .......... 200

Figure A. 1. The results of genetic algorithm function in CLIMEX .................................................. 251
Figure C. 1. Pairwise correlation among all 19 variables in the background buffer of 400 km for Psa..
.................................................................................................................................................................... 258

Figure C. 2. Multivariate Environmental Similarity Surfaces (MESS) of 10 variables to the
environmental data used for training the MaxEnt model in Psa. ................................................................. 259

Figure C. 3. Marginal response curves, each of the following curves represents a different model,
namely, a MaxEnt model created using only the corresponding variable. ............................................ 260

Figure D. 1. Pairwise correlation among all 19 variables in the background buffer of 1000 km for dwarf
bunt. ........................................................................................................................................................................... 263

Figure D. 2. The ROC plot of the MaxEnt model developed for dwarf bunt................................................. 264

Figure D. 3. Multivariate Environmental Similarity Surfaces (MESS) of 10 variables to the
environmental data used for training the MaxEnt model on dwarf bunt. ................................................... 265

Figure D. 4. Marginal response curves for dwarf bunt, each of the following curves represents a
different model, namely, a MaxEnt model created using only the corresponding variable. .... 266

Figure E. 1. Pairwise correlation among all 19 variables in the background buffer of 400 km for guava
rust. ........................................................................................................................................................................ 269

Figure E. 2. Multivariate Environmental Similarity Surfaces (MESS) of 10 variables to the
environmental data used for training the MaxEnt model in guava rust...................................................... 270

Figure E. 3. Marginal response curves, each of the following curves represents a different model,
namely, a Maxent created using only the corresponding variable......................................................... 271

Figure E. 4. The ROC plot of the MaxEnt model developed for guava rust............................................... 272
Chapter 1

General introduction

1.1 The ecological and economic importance of invasive species

Enhanced volume of trade, increased human mobility and climate change are key risk factors in the introduction and establishment of non-native species to novel areas such that in the last 50 years the rate of biotic invader risk has increased extensively (Meyerson and Mooney 2007; McNeely 2001; Sandlund et al. 2001). These invaders are a major threat worldwide due to their effect on biodiversity, habitat destruction, trade disruption, loss of agricultural productivity and disease risk (Lodge et al. 2006; Pimentel et al. 2000; Vitousek et al. 1997). It has been estimated that alien invasive species cause 10-16% of crop production loss worldwide, which is a serious loss especially since the Food and Agriculture Organisation (FAO) estimated that more than 800 million people around the world do not have enough food (Pimentel et al. 2005). Furthermore, according to the same report, environmental and economic damage related to invasive species amounts to $120 billion per annum for the USA alone (Pimentel et al. 2005). Out of total crop losses in the USA due to plant pathogens (US$33 billion per annum), US$21 billion per annum (65%) is attributed to alien plant pathogens (Pimentel et al. 2005). It has been claimed that the Pimental (2005) study only accounted for a small subset of harmful species, and the real costs of invasive species is much higher and continues to grow (Lodge et al. 2006). One main way to reduce production losses of major crops is to better reduce losses due to pathogenic fungi and oomycetes (Pimentel et al. 2005). In a study by Strange and Scott (2005) on the threat of plant diseases to global food security it was suggested, the food lost to these pests would feed 8.5% of the world’s population. Such statistics highlight the necessity to find strategies to decrease the impact of plant diseases.

Plant pathologists have increasingly investigated and developed control and management measures to minimise the damage caused by plant diseases to agriculture, horticulture and forestry sectors, as well as to the conservation estate. While using pesticides is one way to cope with the impact of plant pathogens, it has been documented that pesticide run-off and leaching can be up to 15% (Devault and Pascaline 2014). Additionally, a study last century x
(dos Anjos and de Andrade 2014 cited by Pimentel and Edwards 1982) indicated that only 1% of many pesticides affect the target organism and residue may remain in the ecosystems as pollutants. There are other issues regarding pesticide application. Their effect on the build-up of resistance in the pathogen population, the difficulty targeting the pathogen at the appropriate susceptible stage and reduction in the efficiency of rhizobia, has resulted in failure of control in some instances (Fox et al. 2007). Biological control agents, as an alternative to chemical pesticides, have been used to solve pest and plant pathogen problems, but despite these methods being environmentally friendly, their effectiveness and their fate in environment is variable and requires further research (Stewart 2010; Pujol et al. 2006; Louda and O’Brien 2002). Although the current molecular, genetic, and biological studies regarding plant pathogens have increased our insight and knowledge leading to better control measures, the importance of studies which can determine the risk of establishment of these organisms in novel areas cannot be ignored.

Despite the effectiveness of current control measures at national borders to decrease the risk of plant pathogen invasion into new areas, there are evidence to indicate that some of these organisms have been able to use different pathways such as hitchhiking on objects such as shipping containers, which cannot be adequately covered by current control efforts and strategies (Derraik and Phillips 2010).

1.2 Pest and pathogen dispersal

In the book of Exodus written in 1500 B.C. the migration of an insect is mentioned as a problem and indeed migratory locusts remain a problem to this day (Camuffo and Enzi 1991). The ability of insects to travel long distances can be very important because some insects can act as vectors of plant pathogens (Hajek and St. Leger 1994). There are many records regarding insects and plant pathogens migrating to new areas including temperate and sub-tropical areas (Drake and Farrow 1988). One example is the migration of plague locusts from Australia to Tasmania (Farrow 1984). In the study by Farrow (1984) 24 species from Odonata, Hemiptera, Diptera, Lepidoptera, Coleoptera and Trichoptera orders were reported to be able to travel a minimum distance of 450 km. Farrow (1984) also reported 13 possible migration from Queensland (450 km) and one possible influx from Papua New Guinea to Willis Island a remote island in the Coral Sea at a distance of 600 km. His study showed that 73 of the 83 individuals were trapped on nights where the wind trajectories originated from
Queensland. Other insect trapping has indicated that these insects can travel 1500 km and remain alive (Farrow 1984). Furthermore, another study by Holzapfel and Harrell (1968) reports that species of *Lepidoptera* were able to migrate up to 3000 km.

It has been noted that long distance dispersal of some insects has not received enough attention, at least, not in proportion to their ecological and economic importance (Irwin et al. 1988). Some insects like corn-leaf aphids (*Rhopalosiphium maidis*) that do not persist in northern United States and Canada due to cold winters, nevertheless invade these areas seasonally from southern parts (Irwin et al. 1988). In another example, wind and temperature have been shown to be important factors in the dispersal of some scale insects; and consequently, they are responsible for the rapid establishment and colonization of a scale insect (*Icerya seychel-larum*) in the Aldabra Atoll in the Western Indian Ocean (Hill 1980).

It is suggested that also the long-distance dispersal of plant pathogens has not received enough attention given their ecological and economic importance. Unlike the situation for most macro-organisms, the eradication of plant pathogens after establishment in new regions is particularly difficult as they can rapidly and easily spread through rain, wind, soil movement, by insect vectors, on or within seeds and the action of humans (Fitt et al. 1989; Hoitink and Fahy 1986; Baker and Smith 1966).

Some organisms such as plant pathogens and insects use an atmospheric corridor to disperse to other neighbouring or distant locations by airflow or air currents (Isard et al. 2005). Using these, air-borne and vector-borne pathogens can travel large distances while soil-borne pathogens rarely travel great distances without human intervention (Brown et al. 2002). The long-distance dispersal of *Fusarium graminearum* causal agent of Fusarium head blight has also been documented where clonal isolates of the pathogen have been recovered up to 1000 meters from the release source, in the following year (Prussin et al. 2013).

Some plant pathogens are obligate biotrophs, and require a living host in order to survive and reproduce and establish. Examples of such fungal pathogens are rusts, bunts and powdery mildews. Of these some can generate enormous numbers of spores that can travel to different areas by airstreams; and using these, sometimes, depending upon the nature of the pathogen, they disperse between summer and winter habitats or cold and warm regions (Brown et al. 2002). For instance, the pathogen causing wheat stem rust (*Puccinia graminis* f. sp. *Tritici*), disperses from Mexico to Canada every year (Kolmer 2001). Similarly, *Peronospora*
that causes blue mold of tobacco both in Canada and the United States, repeatedly disperses between these two countries (Davis and Monahan 1991).

Figure 1.1 illustrates how pathogens have been able to move to remote areas. For example, both *Phytophthora infestans* the causal agent of potato late blight disease and *Mycosphaerella fijiensis* the causal agent of black Sigatoka (black leaf streak) in banana probably spread from North America to Europe through infected plant material or by people and thereafter by airborne spores. Similarly, wheat yellow rust (*Puccinia striiformis f.sp. tritici*), dispersed from Europe to Australia. Wheat stem rust (*Puccinia graminis*), coffee leaf rust (*Hemileia vastatrix*) and sugarcane rust (*Puccinia melanocephala*) have been able to disperse by direct movement (e.g. by wind) between continents (Figure 1.1). Interestingly, in some cases such as for cereal stem rust (*Puccinia graminis f.sp. tritici*) and tobacco blue mold (*Peronospora tabacina*), the periodic migration of airborne spores have been recorded as an extinction-re-colonisation cycle (Lodge et al. 2006; Brown and Hovmøller 2002). In the case of cereal stem rust (*P. graminis tritici*), the rust spores (urediospores) disperse through a fixed route which is defined as “*Puccinia path*” from Northern Mexico/Texas to US/Canadian border. The urediospores of the *Puccinia graminis tritici* move back to Mexico/Texas in autumn, at the same time that the monarch butterfly (*Danaus plexippus*) moves southward (Brown et al. 2002). This extinction-recolonisation cycle allows the pathogen to compete with the widespread varieties through rapid evolution of pathogenic forms (Lodge et al. 2006; Brown et al. 2002; Nagarajan et al. 1990). Also it has been documented that there are two main long-distance dispersal pathways in Europe that can help *Puccinia coronate* (crown rust of oats) and *Peronospora tabacina* (tobacco blue-mold) dispersal based on certain weather patterns especially during winter (Sebesta and Harder 1983). The eastern pathway extends from Turkey and Romania to Scandinavia and the western pathway from Morocco toward Scandinavia (Estlander and Kulmala 1973) (Figure 1.2).
Figure 1.1. Dispersal of fungal pathogens. Red arrows: dispersal of pathogens has occurred by windborne spores of hurricanes. Blue arrows: pathogens were transported to new areas by infected plant material, humans and by air travel. Green arrows: back and forth movement of spores. Orange dots: global spread of Sigatoka disease of banana. Image taken from Brown et al.(2002).
All these factors emphasize the increasing need for studies to evaluate the habitat suitability of high risk plant pathogens to help prevent their establishment.

1.3 Species distribution models (SDMs)

Von Humboldt et al. (1807) were the first to acknowledge the importance of climate as a factor in species distribution. Advances in computer technology together with the availability of environmental data have enabled scientists to study distribution or dispersal models in novel ways. Among these new ways, species distribution models (SDMs) associate the current distribution of target species represented by geo-referenced records with a series of environmental predictor variables (Araujo and New 2007; Austin 2007). The output can be presented as maps of habitat or climate suitability or probability of occurrence. SDMs are commonly based on niche theory (Austin 2007) and include different assumptions on how environmental factors control or affect species distribution. These models are also called habitat suitability or bioclimatic niche envelopes (Franklin 2013). In recent decades, SDMs have been used in many ecological, environmental and climate change research projects to

Figure 1.2. Two main long-distance dispersal pathways in Europe. Eastern and western routes converge toward Scandinavia. Image redrawn from Nagarajan et al. (1990).
model invasive species establishment and spread by associating recorded locations of species with environmental variables (Franklin et al. 2013; Maloney et al. 2013). The findings can help better understand various ecological questions, and results can be projected to produce habitat suitability maps for the species. Although there are shortcomings and challenges using SDM’s, such as access to adequate sampling data to build the models, different parameterization model selection, predictor contribution and evaluation strategies (Araújo and Guisan 2006), there has been increasing interest in their use in recent decades (Shaw and Osborne 2011). In particular, SDMs have been used to project the outcome of future climate scenarios on a wide range of species and for conservation planning (Guisan 2005).

1.3.1 Types of SDMs

These predictive modelling methods can be either dynamic or static (Prentice and Solomon 1991). Unlike dynamic models which model the behaviour of the system over time, the prediction provided by static models is time-independent (Beerling et al. 1995). Each type of dynamic or static model can be categorised into either mechanistic or correlative models (Beerling et al. 1995). Mechanistic models which are also called process-oriented or process-based models (Carpenter et al. 1993), are based on the fact that the distribution of every species is influenced by certain environmental limits that define the extent of species distribution (Kearney and Porter 2009). Clearly, to develop a good mechanistic model, knowledge of the life history of the target species and its functional traits or physiological response to environmental predictions is needed (Stephenson 1998). The predictor variables in mechanistic models have a direct effect on the organism rather than an indirect effect as those often used in correlative models (Robertson et al. 2003). The design of a mechanistic model can be very time-consuming and limited as physiological data are not always available or difficult to obtain for many species (Araújo et al. 2005; Robertson et al. 2003). These disadvantages have been one of the main reasons which have encouraged researchers to use correlative models as an alternative (Elith and Leathwick 2009; Pearson and Dawson 2003).

Correlative model prediction is performed by associating the current distribution of species with environmental predictor variables at known presence locations (Robertson et al. 2003). In these models, values of a series of usually abiotic variables are related to presence records or distribution, and the hyperspace of these predictor variables is classified into a probability
of presence or suitability of establishment or presence-absence zones. Correlative models are able to be used especially when the amount of data known about the physiological requirements of the target species is minimal (Webber et al. 2011; Elith et al. 2010).

Different types of correlative models have been used in ecological studies (Worner 2008; Guo et al. 2005; Robertson et al. 2003; Batchelor et al. 1997). The appropriateness of these models relies on the study question, the species under study and the available data. Presence-only models such as BIOCLIM and DOMAIN are very basic types of correlative models, in which different distances or polygons are defined around presence points to make predictions (Beaumont et al. 2005; Carpenter et al. 1993). The second type of presence-only correlative model was developed to improve the early versions of presence-only models. These models use both presence points and background data (data that are not presence points) to make inferences about species environmental needs (Senay et al. 2013). These models can better discriminate the data compared to the early versions of presence-only models and some have been used extensively. An example of this second type of presence-only model is MaxEnt which is an enhanced presence-only model that has been applied in more than 1000 studies since its first introduction in 2006 (Merow et al. 2013; Phillips et al. 2006). Other examples are PBL (Presence and Background Learning algorithm; (Li et al. 2011), ENFA (Environmental Niche Factor Analysis; (Hirzel et al. 2002) and GARP (Genetic Algorithm for Rule-set Production) (Stockwell 1999). Because all presence-only models rely on presence data they are therefore highly sensitive to sampling bias (areas which have been sampled more than other areas) (Senay et al. 2013).

The third type of correlative model aims to use presence and absence points in the modelling process. It has been suggested that despite the unavailability of true absence data, this type of modelling is the best choice (Yackulic et al. 2013). Presence-absence models benefit from statistical and machine learning methods since both presence and absence data can be used in the process of model development. It is hard to define very strict groups of correlative models as some use multiple algorithms (Senay et al. 2013) but they can roughly be considered as regression, classifiers and machine learning methods. Some of these models such as machine learning and classification models (such as support vector machine, neural networks and naïve Bayes) have recently been used in ecological studies (Worner et al. 2010; Watts and Worner 2008; Zuo et al. 2008). These models need presence and (pseudo)absence
points to fully project the species distribution. It is claimed that presence-absence models perform better than other types of model in cases where there is only a small sample of presence data available (Elith et al. 2006). The reason that presence-absence models perform better with small sample sizes and better than presence-only models is that the models generally perform better when both positives (presences) and negatives (absences) are included in the training data and can overcome some common issues such as over-prediction (a prediction higher than the actual value) and extrapolation (to project outside the range of known values) attributed to such modelling efforts (Chefaoui and Lobo 2008). It is not very easy to verify the presence of a species in a specific area, but it is much more difficult to find out if a species is truly absent in a site. Apart from an area being climatically unsuitable for a species establishment, other possible reasons which can result in absence of the species from an area can be that the species has remained undetected despite its presence, or it has not reached the site due to human intervention or natural barriers or competition or predation prevent the species establishing in an area (Senay et al. 2013; Worner et al. 2010).

Therefore, because true absences are not usually available or finding true absence data is very problematic, some methods have been used to create pseudo-absence points (Senay et al. 2013; VanDerWal et al. 2009; Thuiller et al. 2004). These methods will be discussed in the methodology chapter (Chapter 2 sections 2.1.2.1.1. and 2.1.1.1.1).

1.3.2 Application of SDMs in plant pathology

Plant pathogen modelling studies, assisted by computer programs and statistics, can be a very efficient and fast way to discover knowledge from data from different sources. These data can be analysed in novel and increasingly accurate and informative ways to assist decision-making.

The disease triangle, which is one of the basic concepts in plant pathology, considers the interaction of a susceptible host, suitable environmental conditions and a virulent pathogen necessary for disease development. Disease impact can be prevented or mitigated if one of these three elements is removed (Francl 2001). Following the incursion of a damaging plant pathogen in a new area, the first step is to determine local priority zones to control the spread of the pathogen. In such instances, studies that have already projected the habitat suitability of target pests or pathogens can be very beneficial (Worner 2008; Worner 1988).
The most-recent example which affected the New Zealand horticulture industry is kiwifruit bacterial canker caused by *Pseudomonas syringae* pv. *actinidae* (Psa) where priority zones were determined after the incursion happened. Since kiwifruit is an important economic crop in New Zealand, detection of the high-risk localities and priority zones prior to the Psa epidemic, could have helped decision-makers and authorities to better cope with the disease. Alternatively, this could have decreased the significant economic losses imposed by Psa on New Zealand horticulture industry in recent years (Greer and Saunders 2012). Such an example highlights the requirement for conducting studies to project and visualize habitat suitability for exotic plant pathogens both globally and locally to promote the design of efficient management practices to minimize the damage before a species establishes in the new area.

It has been previously stated that different models can be employed to study and monitor plant pathogen establishment and spread. In the field of plant pathology, most of the modelling studies developed are for when a pathogen has already established in the specific study/area to try to calculate and study its epidemiology (Everett and Henshall 1994; Serizawa and Ichikawa 1993; Irwin and Thresh 1990). In recent decades, despite that SDMs have been widely used in ecological, environmental and climate-change research to project invasive species establishment, these models have had limited application in plant pathogen studies (Elith et al. 2013; Susi et al. 2002). Nevertheless, there are a few studies that attempt to model the habitat suitability of plant pathogens using SDMs before their arrival into the new area (Jia et al. 2013; Venette and Cohen 2006). Initially, most such studies were limited to the application of the CLIMEX model (Brasier 1996; Sutherst and Maywald 1985). For example, Yonow et al. (2004) used CLIMEX to study the potential geographical range of *Pyrenophora semeniperda*, a pathogen causing cereal leaf spotting. Pivonia et al. (2004) also used CLIMEX dry stress index to determine the potential geographical distributions of soybean rust (*Phakopsora pachyrhizi*) and Kim and Beresford (2009) also used the CLIMEX model as a part of their study to assess the potential of dwarf bunt (*Tilletia controversa*) establishment in New Zealand. The CLIMEX climate-match region (a sub-function of the program) has also been applied to investigate the current and future potential distribution of guava rust in New Zealand (Kriticos et al. 2008; MAF 2011). With reference to application of correlative models in plant pathology, there are very limited studies which use correlative models to project the potential distribution of plant diseases (Guo et al. 2005; Batchelor et al. 1997; Yang and
Batchelor 1997). A multiple linear regression was used by Royer et al. (1991) to estimate the severity of soybean rust in the USA. They used weather parameters to assess the daily spatial development of this disease. Discriminant analysis has also been used on climate data to model at risk and risk free areas for the Ascochyta blight on chickpea (Diekmann 1992). The risk of Karnal bunt (Tilletia indica) establishment was estimated using climate data (Diekmann 1998). More recently, in a study by Klopfenstein et al. (2009) the MaxEnt model was used to predict the potential distribution of Armillaria root disease. Different studies have used different statistical methods to access the risk of guava rust (Puccinia psidii) establishment in different parts of the world (Elith et al. 2013; Kriticos et al. 2008; Magarey et al. 2007). Support vector machines (SVMs) have also been used for projecting the distribution of sudden oak death disease (Phytophthora ramorum) in the USA(Guo et al. 2005). Yang et al. (1997) used neural networks to project plant disease dynamics in general. Their study offers artificial neural networks (ANN) as a very useful approach for predicting and revealing disease patterns at different scales (spatial and temporal). ANN have also been used in other studies regarding development of plant diseases in relation to leaf wetness in wheat (Batchelor et al. 1997; Francl and Panigrahi 1997).

One common point for all the aforementioned studies is that they all try to estimate the species distribution using a single method. For example, several studies have tried to evaluate the risk of guava rust using various methods. Because the method used in each of those studies is different, discrepancy in the results is evident. On the other hand, it has been argued that since each model can offer valuable information, there are benefits from using more than one model in an ensemble of results (Poulos et al. 2012; Araújo and New 2007; Barai and Reich 1999). The ensemble concept is subject to a separate chapter (Chapter 7) in this thesis.

1.4 Evidence of pest and pathogen invasion in New Zealand

Despite being an isolated island, there are suggestions that certain pathogens have been blown over to New Zealand from neighbouring locations. Some air-borne inocula have been shown to have been transported from Australia or Tasmania to New Zealand. These include Puccinia antirrhini the causal agent of Antirrhinum rust, poplar rust (Melampsora medusae) and sunflower rust (Puccinia helianthi) (Close et al. 1978).
A recent example of a plant pathogen invasion into New Zealand is the kiwifruit bacterial canker (Psa). Psa was first reported in New Zealand in November 2010 after causing serious damage in Italy (Everett et al. 2011). Although the possible pathway of Psa still under debate, a recent study showed the bacteria from Shaanxi province in China and New Zealand were almost identical which increases the likelihood of the pathogen’s movement to New Zealand from China (Butler et al. 2013; Mazzaglia et al. 2012).

The factors that affect the establishment of particular fungal species in a geographical region can be divided into two general groups: 1) biotic factors which include the presence of potential host species, propagule number, rapid evolution, presence of competing species, and, 2) abiotic factors that essentially encompass the climate, topography and land-use (Watts and Worner 2009; Hierro et al. 2005). The early and accurate diagnoses and pathogen surveillance on local, national, and global scales are essential to enable the prediction of outbreaks and in addition, allow time for the development and application of mitigation strategies.

1.5 Overall objective of the thesis

As stated previously many SDM models are available to assess the risk of invasive species. Although some studies have used single models to evaluate the risk of a plant pathogen’s establishment, to our knowledge, there is no study that tests different models to determine which model(s) performs better (if any) for plant pathogens and how we can combine the output of the different models. The three target pathogens for this study were selected based on their importance to New Zealand agriculture and horticulture industry. The target species are *Puccinia psidii* (guava rust), *Tilletia controversa* (dwarf bunt) and *Pseudomonas syringae pv. actinidia* (Psa, bacterial canker of kiwifruit). Detailed information about the target species and their potential impact to New Zealand will be discussed in Chapters 3, 4 and 5.

It should be noted that at the beginning of this study, Psa was not present in New Zealand. The decision to model this species was made subsequently after its establishment and significant impact.

The overriding research question is, given that different models have different outputs can we improve modelling and prediction of invasive species establishment by using the strengths of different modelling approaches.
1.6 Research Specific objectives

**Objective 1:** Determine the strengths and weakness of different models to identify the current and potential global and national distributions of each of the target species.

**Objective 2:** To carry out a sensitivity analysis of the semi-mechanistic model (CLIMEX)

**Objective 3:** Determine an appropriate method for combining and interpreting the output and prediction of different models.

**Objective 4:** Use the improved methods to model the climate suitability of the target species at regional and global scales under future climate change scenarios.

1.7 Thesis structure:

**Chapter 1. Introduction.** Briefly outlines the importance of invasive species in general and more specifically with respect to plant pathology. Explanation of various dispersal mechanisms of invasive species is followed by an introduction of different types of species distribution models and their application in the field of plant pathology.

**Chapter 2. Methodology.** This chapter describes the different models which have been used to project the climate suitability of the target species. Detailed information is presented regarding the modelling process, the species occurrence data and environmental data.

**Chapter 3. Psa.** The aim of this chapter is to investigate how different models (three modelling approaches) project the climate suitability of kiwifruit bacterial canker (Psa).

**Chapter 4. Dwarf Bunt.** The aim of this chapter is to investigate how different models (three modelling approaches) project the climate suitability of dwarf bunt.

**Chapter 5. Guava rust.** The aim of this chapter is to investigate how different models (three modelling approaches) project the climate suitability of guava rust.

**Chapter 6. Sensitivity analysis of CLIMEX model.** The aim of this chapter was to evaluate the importance of performing sensitivity analysis in a process-based or semi-mechanistic model (CLIMEX). The results of a sensitivity analysis conducted on dwarf bunt using the CLIMEX are reported.

**Chapter 7. Consensus model.** This chapter investigates a method for combining and interpreting the output and predictions of multiple models. A review of the different
approaches to ensemble model results is presented and a consensus model is constructed for each target species.

Chapter 8. Future climate scenarios. The aim of this chapter is to evaluate the effect of two climate change scenarios on target species climate suitability in 2030 and 2090.

Chapter 9. General discussion and conclusions. This chapter summarises the findings of the thesis and discusses how the research outcomes contribute to the thesis objectives.

Recommendations are also given for future studies.
Chapter 2

Methodology

2.1 Models

CLIMEX, Multi-Model Framework (MMF) and MaxEnt were used to evaluate the climate suitability of target species based on their current distribution. A summary description of these models and modelling processes is mentioned below.

2.1.1 CLIMEX

Among many models developed for species distribution modelling (Guisan 2005), CLIMEX (available at http://www.ento.csiro.au/climex/bibliography.htm) is one of the earliest to be used widely across several disciplines (Legaspi Jr and Legaspi 2010; Pattison and Mack 2008; Poutsma et al. 2008; Senaratne et al. 2006; Sutherst et al. 1991; Worner 1988). CLIMEX estimates the potential geographical distribution of a species, and its response to climate based on a long-term climate database (Pattison et al. 2008; Poutsma et al. 2008; Senaratne et al. 2006). The process involves matching the species distribution and climate pattern using a calibration method rather than a statistical fitting process.

2.1.1.1 CLIMEX indices:

In CLIMEX, the potential growth and survival of a population at a given site are estimated by several indices and two constraints. The indices can be categorized into two main groups: growth-related indices and stress-related indices. The indices are calculated weekly, then these weekly indices are directly used to derive an index, or they can be combined to provide an annual value of the index. If necessary to distinguish the annual and weekly values, the subscripts of “A” and “W” are used. The growth indices and stress indices which have been used in this study are explained in brief as follows.

2.1.1.1.1 Growth-related indices:

2.1.1.1.1.1 Temperature index (TI):

In CLIMEX, the response of the species to the daily temperature cycle is described as a weekly temperature index (TIw). The index can be between 0 and 1, which when $TI_w = 1$ population growth is maximised, and population growth is zero when $TI_w = 0$. To calculate
the TIw, four parameters are used that define the range of temperature suitability. These four parameters are:

DV0= the lower temperature threshold

DV1= the lower optimum temperature threshold

DV2= the upper optimum temperature threshold

DV3= the upper temperature threshold

The values of temperature parameters (DV0-DV3) define the temperature range that a species can tolerate and survive (Figure 2.1). The temperature within DV1-DV2 shows a range where the temperature index is maximized. Species will not be able to survive or will be unable to remain active at temperatures below DV0 and above DV3 (temperature index becomes zero). To calculate the TIw, the monthly average maximum and minimum temperatures are incorporated in CLIMEX and interpolated to weekly values i.e. arithmetic mean of weekly values. From the monthly temperature index, the annual temperature index (TIA) is calculated as:

\[ TIA = 100 \sum_{i=1}^{52} \frac{TI_w}{52} \]

![Figure 2.1. CLIMEX temperature index parameters. DV0= the lower temperature threshold, DV1= the lower optimum temperature threshold, DV2= the upper optimum temperature threshold, DV3= the upper temperature threshold.](image)

2.1.1.1.2 Soil moisture index (MI):

Based on a hydrological model, the soil moisture stored in the soil from the previous week and current week’s precipitation and evapotranspiration are used to calculate the weekly soil
moisture balance (Sutherst et al. 2007). The main assumption is that soil moisture is a major factor in determining vegetation moisture content and microclimate conditions.

When soil moisture (SM) is zero, it means no soil moisture and when SM=0.5 and 1 it shows that soil moisture content is 50 % and 100% of its capacity respectively. If the soil moisture is higher than 1, it means that the soil moisture is higher than its capacity (run-off). Similar to the temperature index, the soil moisture index defines the species response to soil moisture values, and parameters are SM0, SM1, SM2 and SM3 with a similar form as for temperature parameters. The only difference is that unlike temperature, it is assumed that over a 24-hour period soil moisture is constant.

Other weekly growth indices which are included in CLIMEX but have not been used in this study (based on their relevance or importance to the target species) includes radiation index (RI), light index (LI), substrate index (SV) and diapause index (DI).

2.1.1.1.3 Growth index (GI):

Growth index can range from 0 to 1 and is calculated as:

Weekly Growth Index= TI_w×MI_w×RI_w×LI_w×DI_w×SV_w

Where TI_w, MI_w, RI_w, LI_w, DI_w, and SV_w are weekly temperature, moisture, radiation, light, diapause and substrate indices, respectively. Without reference to life cycle stage or generations of a species in a year, the weekly growth index (GI_w) can help to infer the period of seasonal growth of the species. If the stages of a life cycle occur in a distinct time of the year, then one also can infer the difference in response of the desired life stage. The overall potential for growth is described as the annual growth index (GI_A). GI_A indicates the relative abundance of a species across its range, which is determined by climate.

2.1.1.2 Stress indices:

In CLIMEX, the stress indices are used to define the limits to species survival in hostile seasonal conditions. These limits can define the species geographical distribution. The philosophy behind the stress indices is that stress indices lead to negative population growth of a species if for example the temperature is below DV0 and above DV3 or similarly if the soil moisture content is below SM0. There are four stress indices incorporated in CLIMEX that can be used according to the species studied. They are heat stress (HS), cold stress (CS), wet stress (WS) and dry stress (DS). There are also four other stress indices which consider
the interaction of the previous mentioned indices including: hot-dry stress, hot-wet stress, cold-wet stress and cold-dry stress. An accumulation rate and a threshold value define each stress index.

In CLIMEX, temperature, soil moisture, light and stress (cold, heat, wet, and dry) parameter values are selected in such a way to best represent the response of species to these variables. The Growth and Stress Indices are combined into an Eco-Climatic Index (EI, ranging from 0 to 100) which gives an overall measure of favourableness or suitability of the location or year for permanent occupation by the target species. Results can be presented as tables, graphs, or maps or be imported into other applications such as GIS (Sutherst et al. 2007; Sutherst and Maywald 2004).

The species’ climate requirements are inferred from its known geographical distribution usually in its native range or in other areas regions where the species has been established for a period of time. However, in many studies that use CLIMEX, laboratory data (e.g. developmental temperature thresholds) from the published literature are often used to determine some of the starting parameters to start the calibration process (Kim et al. 2009; Wharton and Kriticos 2004; Ekins et al. 2002; Brasier 1996).

2.1.1.2 AutoIt

One of the main disadvantages of mechanistic models such as CLIMEX is that calibration of the model is time-consuming. In the parameter fitting process in CLIMEX, the user needs to set different parameter values and run the model every single time and if the visual assessment is not satisfactory, the user will need to try new parameter values to eventually find the most appropriate value(s). This process is repeated for each temperature, soil moisture, and stress parameter, and for each threshold (upper, optimum and minimum (DV0, DV1, DV2, DV3, SM1, SM2…)). Clearly, this takes a great deal of time. To partially overcome this problem and facilitate the parameter fitting process, an AutoIt (www.autoitscript.com/site/autoit) script was developed to automate the process of parameter fitting for three species (Appendix A). Based on this script, a range of parameter values are read iteratively from a prepared excel file into CLIMEX, and the script automatically runs the CLIMEX programme repeatedly as many times as needed, and in each run the new maps and tables can be saved in the specified folder defined by the user, enabling to make visual comparisons.
2.1.1.3 CLIMEX Genetic Algorithm

A genetic algorithm is a simulation and optimization process which creates a population of parameter values and then selects the best or optimum values in each simulation. Because usually there are no data available for CLIMEX stress parameters, a genetic algorithm has been included in the latest version (Version 3.0) to fit the initial values for these parameters. The generic algorithm was used to find the stress parameters for one of the target species (dwarf bunt) and the results are presented in Appendix A. The stress parameter values calculated by the genetic algorithm were not in agreement with the findings using the iterative method. The genetic algorithm was discarded as the results were unreliable. The unreliability of genetic algorithm included in CLIMEX might possibly be the reason why it has not yet been used and reported in any published literature. The implication is that this part of the software needs improvement.

2.1.1.4 CLIMEX Modelling Process

The methodology described in Sutherst and Maywald (1985) was used to fit the growth and stress parameters. In this study, the species’ climate requirements are inferred from knowing its geographical distribution in its native range or in other regions where it has been established for a period of time. In this study, for the three target species, laboratory derived developmental temperature thresholds were used as initial estimates and the model parameters values were fine-tuned during validation. As reported by Vera et al. (2002), an EI value of 0 is considered unsuitable, 1-10 is marginal, 11-25 is suitable and greater than or equal to 26 is considered very suitable for establishment of that species. We used this classification in this study to visualize the CLIMEX map, the only difference in our maps was that suitable and very suitable classes were merged into one class and were labelled as suitable. Only in Chapter 6 where the aim was to evaluate the sensitivity of the CLIMEX model were the four categories retained (unsuitable, marginal, suitable and very suitable).

2.1.2 Correlative models

2.1.2.1 MaxEnt

MaxEnt is probably the most well-known correlative model and has been extensively used to model species distributions. Since its introduction in 2006 (Phillips et al.) more than 1000 papers in various fields have used MaxEnt as a modelling approach (Merow et al. 2013). MaxEnt is a presence-only or presence-background model and has been claimed to
outperform most of the existing modelling approaches (Merow et al. 2013; Elith et al. 2006). This model compares presence points against background sites, which are the locations where the presence or absence of the species is unknown. To show how the predictive relative probability of occurrence depends on the value of each environmental variable, a series of the response curves are created by MaxEnt. The model is also able to build complex response curves (linear and non-linear) using feature selection options. These features include, linear, quadratic, product, threshold and hinge features. The linear feature approximately matches the mean value of the predictor variable at locations that species is predicted present with the current observation of that species. In quadratic features, when used together with linear features, the variance in the predictor variable is used to match the predicted location (as presence) with current observations of that species. A product feature is the product of two continuous environmental variables. When used with linear and quadratic features, product features constrain the output distributions to have the same covariance for each pair of environmental variables as the samples (presence points). A threshold feature is calculated from continuous predictor or environmental variables. For the threshold value of x, the threshold feature can be 1 if the variable value is greater than x. The hinge feature is similar to the linear feature, except that it is constant below the threshold x.

When a set of occurrence data are available, MaxEnt estimates the average log likelihood at presence locations whereas the presence-absence models calculate the log likelihood in both presence and absence sites (log likelihood = probability of an observed outcome). In MaxEnt, the calculated log likelihood is compared with a null model, and the improvement in the average log likelihood is shown as gain bars in MaxEnt charts. The jackknife test in MaxEnt measures the importance of each environmental variable by training the model first without the variable, then with only the variable and finally with all variables.

However, as with any presence-only model, the two fundamental limitations of MaxEnt are failure to estimate species prevalence (proportion of occupied sites) and sample selection bias (the areas that are sampled more intensively than others) (Elith et al. 2006). A good explanation of MaxEnt, both with respect to its conceptual basis, its application and advantages can be found in Elith et al. (2011). In this study, recent suggestions by Merow et al. (2013) regarding background selection, feature selection and sampling bias were taken into account. Furthermore, recommendations by Syfert et al. (2013) are considered where
they concluded that correcting sampling bias will lead to major improvement in model goodness of fit. However, interestingly, Syfert et al. (2013) suggest that the choice of input variables had a minor effect on their model predictive performance that contradicts what earlier studies have reported.

2.1.1.1 Defining an appropriate background extent for use in correlative models

A major issue in species distribution modelling is the lack of absence points. A species might be absent in a location for different reasons. Simply, the species may have not been able to establish in the area because of natural barriers or human interventions. Another reason is biotic interactions such as competition, which excludes a species from an area despite it being environmentally suitable. Another reason can be the failure to detect a species which, in fact, is present in the area. For these reasons, different approaches have been proposed by modellers to overcome this problem. The proposed methods help to create pseudo-absences points in presence-absence models or defining which background point’s extent in presence-only models (Senay et al. 2013; VanDerWal et al. 2009; Thuiller et al. 2004). Following this, the question is the appropriate extent from which to select the absence or background point.

As a presence-only model, MaxEnt uses presence and background data to predict species distribution. Background samples are usually random samples of non-occurrences from the region of interest. MaxEnt creates a model with a uniform distribution of probability values over all the observations and the background sample files (so it just needs presence points). The presence-absence models instead use presence and absence points (or pseudo-absence points) to make the prediction. In presence-absence models, when real absences are not available, the pseudo-absence points are generated by the models. These pseudo-absence points which are environmentally dissimilar to presence points can be taken randomly or from a defined extent around the presence points. VanDerWal et al. (2009) suggested that the spatial extent at which background samples (or pseudo–absences) are taken should be considered carefully because it can have a significant effect on model output. He suggested defining a background buffer within some distance of an occurrence point as a solution to get the most biologically meaningful results. Although VanDerWal et al. (2009) suggested a 200 km buffer around the presence points as an optimum distance to select the background points, others had suggested performing an independent exploratory analysis to find the optimum distance and therefore, achieve the most accurate projections. In another study, Thuiller et al. (2004) found that the effect of restricting the environmental range of data can
strongly influence the response curve estimation and model performance. To overcome the problem of background selection and to avoid sample selection bias, a novel method suggested by Senay et al. (2013) was used in this research. This novel method uses Principle Component Analysis (PCA) as a method to define both spatially and ecologically, a meaningful distance to limit the background before the selection of pseudo-absence or background points. Different radii were used around the presence points to bind the background points. We tested 50 km, 100 km, 150 km, 200 km, 250 km, 300 km, 350 km, 400 km and thereafter up to 2000 km with a 100 km interval to check the changes in variable importance. For these multiple datasets, variable importance was analysed using PCA which led to identification of the variable that contributed the most to the first principal component. The variable that contributed the most was plotted against different distances and wherever the contribution of the most important variable showed obvious change (obvious increase or decrease), that distance was selected as the optimal distance in MaxEnt (as background extent) and for the Multi-Model (as pseudo-absence extent). The files were converted into an appropriate format to be used in the Multi-Model and MaxEnt program.

2.1.2.1.1 MaxEnt Variable selection

In order to minimize the correlation among the environmental variables, we removed the highly correlated variables prior to using the MaxEnt model. The candidate variables for each species were selected using a pair-wise Pearson correlation test as well as consideration of ecological relevance and knowledge of target species. The results of each Pearson test can be seen in Chapters 3, 4 and 5.

Because SDM models cannot perform well if the climate in the interested area is very different from native and invaded ranges (extrapolation), a multivariate environmental similarity surface (MESS) was performed to check if the model needs to extrapolate in the areas of interest. The MESS analysis highlights the areas where one or more variables are outside the range presented in the training data, therefore, shows those areas where prediction should be treated with caution (Elith et al. 2010)

The selected environmental variables (layers), along with the sample file that includes target species presence points, were used in MaxEnt to calculate the AUC (area under the curve) of the models. Linear features, quadratic features, product and hinge features were used during the training as the dataset is large enough to support complexity. The maximum number of
background points and number of training iterations was set at 10,000 and 5,000 respectively. Seventy-five percent of data were used to train the model and 25% were set aside to validate the model ((Huberty 1994 cited in Fielding and Bell 1997). The models were set to create the response curves and a jackknife test was performed to identify the important variables.

2.1.2.2 Multi-Model Framework (MMF)
Several correlative models have been used previously to predict the distribution of invasive species, especially when there is little knowledge about the species response to the environment (Vaclavik and Meentemeyer 2009; VanDerWal et al. 2009; Phillips et al. 2006). Most often, a single model is used, but it has been reported that there is not a perfect or “true” model (Barai et al. 1999). In this research, in addition to MaxEnt software, the aim was to apply and evaluate a number of these common methods using the Multi-Model Framework developed by the Ecological Informatics Group in the Bio-Protection Research Centre, Lincoln University (Worner et al. 2010). The program models species presence/absence data using nine different species distribution models that include: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), logistic regression (LOG), naïve Bayes (NB), classification and regression trees (CART), conditional trees (CTREE), K-nearest neighbour (KNN), support vector machines (SVM) and artificial neural networks (NNET).

2.1.1.2 Absence generation (balancing the data)
One important step is to generate absence locations. Absences are virtually never recorded. Even if they are, it is difficult to be sure those absence locations are real because in some cases there may have not been enough searching to detect the species that actually is established there, or it may not have reached the area yet ((Phillips et al. 2009);(Vaclavik et al. 2009; VanDerWal et al. 2009). The problem gets more complicated when one works with global data (in this case Worldclim data that consists of over 580,000 points) because there are a large number of potential absence locations that can cause class imbalance. Some researchers use a random selection of these absence locations to overcome the problem of class imbalance. The random selection of absence data, and the resulting selected locations, can have considerable effect on model output and consequently, may result in inaccurate interpretation. To solve this problem, the Multi-Model uses one-class support vector
machines (OCSVMs) to select appropriate absence points out of large datasets of potential locations. This method has some benefits, such as model creation in a short computational time, a high degree of accuracy and as previously suggested the ability to handle large datasets. In this study, the variable layers extracted by making a proper distance boundary around presence points (described in Section 2.1.2.1) were fed into the Multi-Model allowing it to select pseudo-absences points around the presence points (Senay et al. 2013). The application of OCSVMs in ecological studies is relatively new (Zuo et al. 2008; Drake et al. 2006; Guo et al. 2005). Instead of selecting a single best-performing OCSVM model that can result in over-fitting the data in the model, a set (ensemble) of 100 models fitted to a different sample of the data, which had the lowest prediction errors are selected. The absence locations are those where the probability of their environmental suitability was 0 in all 100 models. As there are still many possible absence locations after this analysis, these points are reduced by clustering absence locations that have similar environmental variables at a defined distance by K-means clustering with K clusters to balance the number of presence locations. For each cluster, the geographical location with environmental conditions closest to the cluster centre was chosen as the absence point for that cluster.

### 2.1.1.3 Variable selection

After selecting the balanced absence data, significant bioclimatic variables were verified and selected using a random forest analysis (Breiman 2001) and stepwise regression analysis (Thompson 1995). A random forest is a classification method that uses many (1,000) decision trees generated by a random selection of variables or features. The advantage of this algorithm is that it can indicate the importance of the variables and can help us in variable selection. The stepwise regression analysis is a very common method for selecting a good subset of variables in a short time. The stepwise regression can be done in forward or backward procedure. In the forward method, to identify the most important variables, the variables are added one-by-one and compared with the null model while the backward method starts with the null model and the least significant variables are dropped one-by-one. These two processes (random forest/stepwise regression) remove insignificant variables and therefore, improve the model fit and its validity and computational time. After selecting the appropriate environmental variables for each species, the Multi-Model uses the respective variables in all nine models.
2.1.1.4 Modelling procedure

All nine models were trained (or fitted) and tested (evaluated) using the variables that were selected in the variable selection process (Section 2.1.1.2). In order to get the best set of parameters for some models such as SVM (support vector machine) and ANN (artificial neural networks), initial parameterization was carried out followed by optimisation. The models are automatically validated, and ranks based on different performance criteria. Validation was carried out to test the ability of the models to predict new data. This was done by cross validation (partitioning data) and bootstrapping to generate independent validation data. In the cross-validation method, the data are divided into number of folds or partitions (10 folds in this case). Each time one partition or fold is set-aside and the model is fitted to the remaining data and the process is repeated 20 times. In bootstrapping, the data are partitioned and the model is usually fitted to portion of data (70% to 80%) and the rest of the data (out-of bag) are used for model validation. Model predictions were plotted globally and for New Zealand. Furthermore, the predicted maps were exported as ASCII format that can be used in GIS platforms such as ARCGIS for further analysis.

2.1.1.5 The predictive performance evaluation in binary predictions

To measure the performance of a predictive model they need to be validated. To evaluate the performance of any predictive model a set of test or evaluation data is necessary. These test data should be independent of the training data or the data that model was calibrated with. The most common method is to divide the data into two sets: training data and test data. The data can be portioned in different ways such as a fixed percentage (75% for training and 25% for test which was used in the MaxEnt model) or cross-validation and bootstrapping that are used in the Multi-Model Framework. When binary predictions are made (such as presence/absence) many performance criteria can be used to validate model performance. All these performance criteria are based on what called a confusion matrix which contains information about the actual and predicted classification made by the model. The confusion matrix is built from the frequency of each type of prediction evaluated on the test data (Table 2.1).
Table 2.1. Confusion matrix

<table>
<thead>
<tr>
<th></th>
<th>Recorded present</th>
<th>Recorded absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted present</td>
<td>TP (True positive)</td>
<td>FP (false positive)</td>
</tr>
<tr>
<td>Predicted absent</td>
<td>FN (False negative)</td>
<td>TN (true negative)</td>
</tr>
</tbody>
</table>

Where

TP = the presence points which were correctly predicted as presence.

FN = the presence points which were falsely predicted as absence.

FP = the absence points which were falsely predicted as presence.

TN = the absence points which were correctly predicted as absence.

The aim is to achieve TP and TN as high as possible. Based on the confusion matrix some criteria were calculated as follow:

Sensitivity (recall) = TP/ (TP+FN) True positive rate

Precision = TP/ (TP+FP)

Accuracy = (TP+TN)/ (TP+TN+FP+FN)

Specificity = TN/ (TN+FP) True Negative Rate

F-score = (2*recall*precision) / (recall+ precision)

For any test, there is a trade-off between the measures (e.g. low specificity = high sensitivity). This trade-off is presented as a ROC (receiver operating characteristic) curve that is usually used for binary classification, which plots true positive rates versus false positive rate (1-specificity) (Figure 2.2)
As Figure 2.2 shows, the diagonal line (C) represents a random or chance model. By increasing the accuracy of the model, the ROC curve moves towards A. Curve B shows a typical ROC curve indicating the model is better than a random model.

![Figure 2.2. The ROC curve showing an ideal accuracy (A), a typical accuracy (B) which is better than a random model and C shows a random model.](image)

### 2.2 Data collection

To perform species distribution modelling, the location data (presence points) and climate data should have spatial coordinates or in other words, be geo-referenced. The presence or occurrence data for all three target species were recorded in an Excel database that includes the name of the country, locality and spatial coordinates for the record. To transfer the location data into the modelling framework, they were saved as the text format (TXT) for the Multi-Model Framework or as the comma delimited (CSV) format for the MaxEnt model. Furthermore, the climate data were sourced from different sources that are described in detail in Section 2.2.2.

#### 2.2.1 Occurrences Data

For many organisms such as insects, plants and weeds, occurrence data are obtainable on databases such as GBIF (Global Biodiversity Information Facility) and CABI Crop Compendium (http://www.cabi.org/cpc), whereas information on exact locations of plant pathogens is rare. Probably, the most challenging part of modelling plant pathogens using
SDMs is locating the presence or occurrence data. Compared to macroscopic species, there is no universal source that provides such information. The CABI Crop Compendium gives limited information regarding plant pathogens’ distribution in general and specifically Psa and guava rust. In such databases, the names of a province and sometimes the name of a district is mentioned as a disease record. Such records at an administrative level can be problematic where there are reports from large countries with diverse climate conditions such as China, USA and where the species may not be widespread. Furthermore, no data recording Psa was found on the GBIF website. Therefore, to supplement geographical location information for target species, published papers and personal communications with researchers was used (Appendix C, D and E).

2.2.2 Climate data
Good-quality environmental data are needed for quality models. Bioclimatic variables have been frequently used in many studies such as for freshwater species, ants in New Zealand, Europe and North Africa (Kumar et al. 2014; Worner et al. 2010; Elith et al. 2006; Ward 2006; Beauchard et al. 2003). The Bioclim dataset are derived from monthly temperature and precipitation data to generate biologically meaningful variables (Hijmans et al. 2005). As with many other ecological studies, and also because the bioclim variables (19 variables) data are available on a global level (www.climond.org), these variables are used in this research based on the assumption that they are capable of reflecting the climate suitability of locations with respect to the growth and development of plant pathogens as well as other living organisms (Kriticos et al. 2012). The Bioclim data resolution prepared for this study is in 10 arc minute (0.17° ≈ 18.6 km). A list of the 19 bioclim variables used in correlative models in this study can be found in Appendix B.

2.2.2.1 CLIMEX data
In CLIMEX modelling, we used the recent updated 10’ gridded climate data of CliMond prepared by Kriticos et al. (2012) provided in CliMond website. The CliMond dataset includes temperature, rainfall and humidity values, which have been produced using the CRU and Worldclim datasets (Kriticos et al. 2012). These data are not originally included in the latest version of CLIMEX software Version 3.0). Therefore, to add new meteorological data to use in CLIMEX, the downloaded met manager files (with format of *.mm) were
copied into the data folder of the CLIMEX model. Then by opening the CLIMEX software a new simulation was created and named accordingly.

2.2.2.2 Correlative models data

For correlative models (Multi-Model Framework and MaxEnt) we used 19 bioclimatic variables with the same resolution (10') provided in the CliMond website (www.CliMond.org)). In this database, long-term temperature and rainfall provide more meaningful variables that are often used in species niche modelling efforts (Appendix B).

2.3 Further analysis using Principal Component Analysis.

A principal component analysis (PCA) was carried out to visualize the climatic similarity among presence locations for three target species. The climate data for presence locations was extracted from 19 bioclim variables in ArcGIS Desktop 10.1. A new field as presence was created in the attribute table of the file and all extracted locations were assigned 1 (as presence) to differentiate them from absence data and to better visualize them later. Data were then transferred to R software and were analysed using “princomp” command that performs a PCA. The data were transferred back into ArcGIS and the presence points were marked with different symbols to be distinct from the absence points. For the purpose of better visualisation, the pseudo-absence points were removed, instead a boundary was drawn to show the total extent of all points. The results for all three species were shown in the appropriate chapters.
Chapter 3
Potential climate suitability for kiwifruit bacterial canker

3.1 Bacterial Canker of Kiwifruit (Pseudomonas syringae pv. actinidiae)

3.1.1 The disease
The causal agent of bacterial canker of kiwifruit is Pseudomonas syringae pv. actinidiae (Psa) which is a gram-negative, obligate aerobic bacterium that occurs singly or in pairs and moves by 1-3 polar flagella. Ushiyama et al. (1993) suggested that this bacterium originated from wild Actinidia spp. and migrated to the north of Japan from unknown source, but recently its origin has been related to China (Mazzaglia et al. 2012). Until a recent Italian outbreak where Psa losses were estimated at two billion Euros, it was believed that the disease’s economic importance was relatively low because there had been no major outbreak with serious damage (Balestra et al. 2009).

Symptoms mainly appear at the beginning of spring with leaf spots becoming visible in spring, appearing brown surrounded by a bright chlorotic halo (Figure 3.1) (Renzi et al. 2012b). Floral buds may turn brown, wither and exude gum (without opening). Sepals can also be infected. One of the main symptoms of Psa on kiwifruit is translucent exudates on otherwise healthy tissues, such as canes (Figure 3.1). At the beginning of spring the droplets of exudates change to red-brown colour. Canker symptoms may also appear in mid-winter. By the end of the vine dormancy period, any wounding caused by pruning or damage can result in the development of cankers on infected vines. Necrosis may extend onto the trunk and the underlying vascular tissue making the trunk appear darkened and desiccated (Balestra et al. 2009). Twig dieback and plant wilting can also appear in the later stages of infection (Renzi et al. 2012b; Serizawa et al. 1989). Where the pathogen is active in infected

---

1 Part of results of this chapter published as:
plant material it can easily be transferred to new plants by nursery materials, particularly grafting and pruning tools (Renzi et al. 2012b).

3.1.2 Global Distribution

Bacterial canker of kiwifruit was first isolated and reported from Japan in 1984 where it infected around one-third of orchards in Shizuoka Prefecture (Serizawa et al. 1989). In the

"Figure 3.1 The symptoms of Psa on leaf and cane of kiwifruit plant (top-right)." Translucent exudate on canes of infected kiwifruit (top-left). The effect of disease on fruits of kiwifruit vines in Te Puke orchard (lower picture personal photograph by H.A Narouei Khandan 25 Jan 2012)."
late 1980s, Psa was reported in Korea (Koh et al. 1994) and thereafter a report from Yuexi province confirmed an outbreak in China (Heyuan et al. 1995). Italy is the main world kiwifruit producer and after an ephemeral Psa outbreak in 1992 (Scortichini 1994), there was a serious outbreak in 2008 both in Italy and neighbouring countries such as France (Renzi et al. 2012a; Balestra et al. 2009; Balestra et al. 2008). During the following years, Psa continued to spread into Portugal (Renzi et al. 2012a; Balestra et al. 2010), Australia (Biosecurity 2011), France (Vanneste et al. 2012b) and Spain (Balestra et al. 2011) (Figure 3.2). The most serious losses associated with Psa were recorded in New Zealand where it was first reported in the Bay of Plenty in November 2010 (Everett et al. 2011). Although two strains of Psa have been reported from New Zealand (Psa- low virulent and Psa-virulent), the most recent report stated that 77% of New Zealand’s kiwifruit orchards have been infected with Psa-virulent (KVH 2013). More recent Psa reports are for the Rize province, Turkey (Bastas and Karakaya 2012), in the O’Higginsi and Mauleu regions in Chile (Anonymous 2011), Switzerland (EPPO 2011) and Germany (EPPO 2013).
Figure 3.2. Current global distribution of Psa. The presence points in New Zealand are not shown.
3.1.3 Host range

Psa can infect both *Actinidia chinensis* and *Actinidia deliciosa* but recently its survival has been reported on *Cryptomeria japonica* (Japanese cedar) (Vanneste et al. 2012a), *Alternanthera philoxeroides* (aligator weed), *Setaria viridis* (green foxtail) and *Paulownia fortunei* (dragon tree) (all considered previously non-host plants) (Gallipoli et al. 2013). Furthermore, it has been recently reported that rice leaf-hopper (*Nephotettix bipunctatus*) and Chinese parasol tree (*Firmiana simplex*) are able to transfer and host Psa cells respectively (Gallipoli et al. 2013).

3.1.4 Epidemiology

It is reported that Psa can infect kiwifruit vines at relatively low temperatures between 10-20°C with optimum temperature of 15±3°C (Serizawa and Ichikawa 1993). Above 25°C, infection is completely inhibited although in some studies this threshold is given as 27°C (Serizawa and Ichikawa 1993). There is no detailed information regarding suitable, minimum or optimum rainfall range. All the previous studies simply report that rain is an important factor. One study however, concluded that rain is the most dominant factor in the establishment and distribution of Psa but no data were given to support this claim (Yao et al. 2001).

This pathogen can persist in infected plant materials and can also easily be transferred to new plants through nursery material, grafting instruments and pruning implements. Similar to many other pathogens, Psa can splash- and wind-disperse into new areas (Balestra et al. 2009; Madden 1997; Lindemann and Upper 1985).

3.1.5 Risk to New Zealand

Prior to starting the current study, Psa was not reported in New Zealand. Thus, the original aim of study was to evaluate the risk of establishment of Psa in New Zealand. As mentioned earlier, Psa was first reported in November 2010 in the North Island of New Zealand, Bay of Plenty region (Everett et al. 2011). Recent estimates place the cost of Psa-V (virulent strain of Psa) to New Zealand economy at $310-$410 million NZD , over the next five years, $500-$600 million NZD over the next 10 years, and $740-$885 million NZD over the next 15 years (Greer et al. 2012). Although Psa has already established in New Zealand with substantial economic damage to the horticulture industry, this study aims to better understand epidemiology of Psa and to project non-infected areas both locally and globally. Such
findings may also contribute to better management practices e.g. establishment of kiwifruit nurseries in areas which models project as risk-free or low-risk.

3.1.6 Previous modelling of Psa

Despite the substantial economic losses attributed to Psa in Italy and New Zealand in recent years, there has been no comprehensive study that evaluates global and local risk of Psa establishment in areas where kiwifruit is commercially grown. To my knowledge, only one Psa modelling study exist (EPPO 2012, Reynaud et al. (2011, cited in (EPPO 2012)) used CLIMEX software package (version 3.0, CSIRO Publishing, Melbourne, Australia) to evaluate climate suitability for Psa in Europe based on cold and dry stress parameters.

3.2 Methods

3.2.1 Psa occurrence data

As one of the goals of this study was to project climate suitability of Psa in New Zealand, the New Zealand data were not used in the process of model building. Instead, they were mainly used to evaluate and validate the models to test model predictions. The most accurate records were available for Italy. There is no published information about the environmental needs for infection of kiwifruit vines by different strains (Psa-V and Psa-LV strains), so these strains were not differentiated among the presence records. Also, studies investigating the genetic similarity of strains from different countries are currently in progress (Butler et al. 2013; Zhao et al. 2013; Mazzaglia et al. 2012; Lee et al. 2005). The original presence data (prior to removing duplicate or very close points) for Psa in this study comprised 284 points from 10 countries (Balestra et al. 2009) (Balestra et al. 2011; Balestra et al. 2010) (Everett et al. 2011; Serizawa et al. 1989; Takikawa et al. 1989; Quattrucci et al.) (Figure 3.2). There are recent reports of Psa in Switzerland, Slovenia, Turkey and Germany, but the records from these countries were not used because the presence of Psa or the exact localities were not confirmed. Duplicate points were removed based on the resolution of the data (10’ resolution) so 87 points were used in the modelling process (Appendix C).

3.2.2 CLIMEX

We used the methodology described in Sutherst and Maywald (1985) to fit the growth and stress parameters. The detailed information regarding parameters is described as follows.
The initial temperature index parameters (DV0, DV1, DV2, DV3) were set based on minimum, optimum and maximum temperature requirements of Psa reported by Serizawa et al. (1993). As there was no information available regarding the soil moisture requirement of Psa, the default temperate template of CLIMEX was used to set the initial values of soil moisture. The stress parameters were fitted so that the majority of known occurrences of Psa occurred in climatically suitable and very suitable grid cells. As there are no data available on the response of Psa to stress conditions, stress parameters were iteratively changed and the model results were compared with the known distribution. By adjusting the above-mentioned parameter values iteratively, the model was run repeatedly to achieve the highest EI values close to the known distribution of Psa and lowest EI values outside the range of Psa reported distribution. These criteria were used to obtain the closest match of suitable habitats projected by CLIMEX and the reported distribution patterns of Psa.

3.2.3 Correlative models

3.2.3.1 Background selection for correlative models (MaxEnt and Multi-Model)

To define the extent which MaxEnt and Multi-Model select pseudo-absence points, the same methodology as described in Chapter 2 (section 2.1.1.1.1) was used. Distances between 50 and 400 km at 50 km intervals and thereafter up to 2000 km at 100 km intervals were chosen. These distances were chosen to test the change in important variables i.e. maximum temperature of warmest month (variable cbio05). The variable that contributed most to the first PCA component were identified (variable cbio05), and the change in this variable was plotted against distance change. The distance at which the contribution of the important variables started to stay constant (at 300 km), was chosen as the appropriate distance to constrain the background data. Figure 3.3 shows the change in the important variable for the first principal component versus buffer distances.
3.2.3.2 MaxEnt model

3.2.3.2.1 Variable selection

To identify the relevant predictor variables for MaxEnt, the approach involved selecting the variables that were available at a global scale, which were also relevant to the ecological properties of target species. A pairwise Pearson correlation test was used to reduce multicollinearity by excluding the variables that were highly correlated (Dormann et al. 2013).

The correlation test was performed on all background points ($n=18511$) that were intended to be used in the MaxEnt model. The process resulted in selection of ten variables for which the pairwise correlation was less than 0.70 (Appendix C). Based on experience, in some cases where the two variables were correlated but were ecologically and biologically important for Psa, they were also included to candidate variables. Linear, quadratic and hinge features were used in MaxEnt and jackknife was used to identify the important variables.

Additionally, response curves were produced to show how each variable can affect the MaxEnt projection. A Multivariate Environmental Similarity Surfaces (MESS) analysis was performed to compare the environmental similarity of variables used in MaxEnt model (Appendix C) (Elith et al. 2010).

3.2.3.3 Multi-Model Framework

The procedure in the methodology Chapter 2 was used to run the Multi-Model framework. The same numbers (87) of pseudo-absence points (as presence point) were generated from a 300 km buffer around the presence points using OCSVM and k-mean clustering. Models
were trained using the presence and generated pseudo-absence points and were validated by cross-validation and bootstrapping techniques.

3.2.3.3.1 Multi-Model variable selection

As mentioned in the methodology chapter (Chapter 2), variable selection in Multi-Model was performed by random forest analysis (Breiman, 2001) and stepwise regression analysis (Thompson, 1995). For the Multi-Model Framework, the variables selected both by random forest and step-wise regressions were: cbio01 (annual mean temperature), cbio2 (mean diurnal range), cbio03 (isothermality) and cbio13 (precipitation of wettest month). As for MaxEnt, the same 300km buffer around the presence points was used allowing the model to select pseudo-absence points.

3.3 Results

3.3.1 CLIMEX

The final parameter values for the CLIMEX model for Psa can be found in Table 3.1. These values were obtained through iterative calibration until the map projection fitted the current distribution of Psa (excluding New Zealand data which were withheld for validation). The resultant parameter values were also checked for biological feasibility. Temperature values such as the lower temperature threshold \( DV_0 \) was lower than the 8-10°C that has been reported in the literature (Serizawa et al. 1993). In addition, the upper temperature threshold obtained was 28°C which was higher than reported values in available previous studies (25°C) (Cheng et al. 1995). These differences are expected as temperature thresholds reported in studies are based on experiments conducted in controlled conditions.

According to current presence locations, the CLIMEX model represented a good fit to the presence data (Figure 3.4). Central to Eastern China and Japan showed the highest EI value (meaning highest suitability) which matches with reports of Psa from these regions. For South Korea, only coastal areas in the south towards the southeast were suitable with the rest of the country being projected as marginal. The model also projected that Northern parts of Iran were highly suitable where kiwifruit is grown commercially. Kiwifruit is also grown in a few orchards in some countries such as Laos, Philippines, Cambodia, Vietnam, India and Bangladesh. Among those countries, Laos and Vietnam were predicted marginally suitable
by CLIMEX. In Europe, the Latina province, and the autonomous regions of Sicily and
Sardinia of Italy, all of Portugal, and small areas in Greece, France and Spain were projected
as highly favourable areas. Psa is already present in all these areas except Greece (Ministry of
Rural development and Food 2012). Kiwifruit is also grown in Belgium, Denmark, Cambodia
and Netherlands. Most of these areas were projected as marginal except for northern parts of
Belgium (Bruges, Gent and Antwerp) which were projected as suitable. In Turkey, some
central areas such as Mons and Waver, western parts of Izmir and northern parts of Adana
and Trabazon Rize (where kiwifruit is grown) were projected as suitable areas.

In Africa, some coastal areas in Morocco, Algeria and central areas in Ethiopia were
projected as highly suitable but kiwifruit is not currently grown in these areas. Kiwifruit
have been cultivated in South Africa only recently and CLIMEX projects areas around Cape
Town and Durban are highly suitable for long-term establishment of Psa.

The most extensive area that was projected as highly suitable for Psa was located in South
America around Buenos Aires to Mar del Plata in Argentina. Along with Argentina, the
whole of Uruguay and southern Brazil were also projected as highly suitable. In Chile
(where kiwifruit is grown), from Puerto Montt to Santiago were predicted highly suitable.
Table 3.1. Parameter values used in developing CLIMEX model for Psa in this study. The table also shows the parameter values used by the EPPO model.

<table>
<thead>
<tr>
<th>Index</th>
<th>Parameters</th>
<th>Current study values</th>
<th>EPPO values</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV₀</td>
<td>Lower temperature threshold</td>
<td>5</td>
<td>10</td>
<td>°C</td>
</tr>
<tr>
<td>DV₁</td>
<td>lower optimum temperature</td>
<td>12</td>
<td>20</td>
<td>°C</td>
</tr>
<tr>
<td>DV₂</td>
<td>upper optimum temperature</td>
<td>20</td>
<td>25</td>
<td>°C</td>
</tr>
<tr>
<td>DV₃</td>
<td>Upper temperature threshold</td>
<td>28</td>
<td>30</td>
<td>°C</td>
</tr>
<tr>
<td>SM₀</td>
<td>Lower soil moisture threshold</td>
<td>0.5</td>
<td>0.2</td>
<td>-</td>
</tr>
<tr>
<td>SM₁</td>
<td>Lower optimum soil moisture</td>
<td>0.8</td>
<td>0.5</td>
<td>-</td>
</tr>
<tr>
<td>SM₂</td>
<td>upper optimum soil moisture</td>
<td>2</td>
<td>1.5</td>
<td>-</td>
</tr>
<tr>
<td>SM₃</td>
<td>Upper soil moisture threshold</td>
<td>3</td>
<td>2.5</td>
<td>-</td>
</tr>
<tr>
<td>TTCS</td>
<td>Cold stress temperature threshold</td>
<td>5</td>
<td>2</td>
<td>°C</td>
</tr>
<tr>
<td>THCS</td>
<td>Cold stress temperature rate</td>
<td>-0.00005</td>
<td>-0.01</td>
<td>Week⁻¹</td>
</tr>
<tr>
<td>DTCS</td>
<td>Cold stress degree-day threshold</td>
<td>15</td>
<td>-</td>
<td>°C</td>
</tr>
<tr>
<td>DHCS</td>
<td>Cold stress degree-day rate</td>
<td>-0.0001</td>
<td>-</td>
<td>Week⁻¹</td>
</tr>
<tr>
<td>TTHS</td>
<td>Heat Stress Temperature threshold</td>
<td>30</td>
<td>-</td>
<td>°C</td>
</tr>
<tr>
<td>THHS</td>
<td>Heat Stress Temperature rate</td>
<td>0.0005</td>
<td>-</td>
<td>Week⁻¹</td>
</tr>
<tr>
<td>SDMS</td>
<td>Dry Stress Threshold</td>
<td>0.2</td>
<td>0.15</td>
<td>Week⁻¹</td>
</tr>
<tr>
<td>HDS</td>
<td>Dry Stress Rate</td>
<td>-0.005</td>
<td>-0.02</td>
<td>Week⁻¹</td>
</tr>
<tr>
<td>SMWS</td>
<td>Wet stress threshold (1-10)</td>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HWS</td>
<td>Wet stress rate</td>
<td>0.001</td>
<td>-</td>
<td>Week⁻¹</td>
</tr>
</tbody>
</table>

In North America, California was projected as suitable and large areas in Alabama, Louisiana, Georgia and Mississippi were projected as highly suitable and suitable but kiwifruit growing areas are centred in California around the Sacramento Valley. In Australia, limited areas in the southeast coastal areas such as Victoria were projected as highly suitable and the only confirmed report of Psa-LV comes from this area (Biosecurity 2011). In addition, Perth in Western Australia where kiwifruit is grown on a small scale was projected as highly suitable for Psa.

In New Zealand, the North Island was predicted highly suitable for Psa establishment with all areas having EI values greater than 25. Additionally, in the South Island, areas around Nelson, Christchurch and Blenheim were projected as highly suitable. The weekly growth
index chart (Figure 3.5) shows that in most days of the year, Psa can grow in the Bay of Plenty, although the most suitable periods seems to be March-April and October-December. This higher growth index in winter was in accordance with peaks of Psa reported in New Zealand (www.kvh.org.nz). In summary, all kiwifruit growing areas in New Zealand lie within the suitable category of CLIMEX confirming that New Zealand is highly suitable for long-term establishment of Psa.
Figure 3.4 Climate suitability of Psa on a global scale and New Zealand projected by CLIMEX.
Figure 3.5. Weekly growth index and climate variables at Te Puke for Psa.
3.3.2 MaxEnt

For MaxEnt, the Area under the Curve (AUC) was used to measure model performance. The AUC of the test data for the Psa model was 0.703, which indicates MaxEnt has been able to discriminate the data better than a random model (Figure 3.6).

Figure 3.6. The Area under the Curve (AUC) of MaxEnt model. The black diagonal line indicates a random model. The chart shows MaxEnt has performed better than a random.

annual temperature (cbio1), max temperature of the warmest month quarter (cbio05), mean temperature of the coldest quarter (cbio11) and minimum temperature of the coldest month (cbio06) were the most important variables respectively and provide a reasonably good fit to the training data (Figure 3.7a). Furthermore, Figure 3.7a shows that the variable annual precipitation (cbio12) decreased the gain the most when it was omitted which indicated it had the most important information that was not present in the other variables.
The light blue bars in Figure 3.7a suggest that, no variable contains significantly useful information that is not previously included within the other variables. Compared to other variables, the gain in training data experienced a slight decrease when annual precipitation (cbio12) and maximum temperature of the warmest month (cbio05) were removed from the
process of training data. This showed that in the process of training, removing annual precipitation decreased the gain most but when the annual precipitation was used by itself it did not appear to be a very important variable. The comparison of the jackknife of the test data and AUC was also very informative. The jackknife of the AUC plot shows that minimum temperature of the coldest month (cbio06) emerged as an important variable even though it was the fourth most important variable among the ten variables in the training data (Figure 3.7b). Moreover, the importance of variable annual precipitation (cbio12) was highlighted more in the gain plot, where by removing this variable resulted in the AUC of the model decreasing to 0.67 (Figure 3.7b).

The jackknife plot of test gain showed that when the model uses only temperature seasonality (cbio4) as an input variable, a negative gain was obtained (Figure 3.7c) which meant that temperature seasonality was not the best choice as a predictor variable. Figure 3.7c also shows that if some of the variables were used by themselves a very low gain was achieved. For example, if the model was based on temperature seasonality (cbio04) alone or just annual precipitation (cbio12) the test gain of the model would be low which indicates they do not significantly contribute to estimating the distribution of Psa by themselves.

Finally, response curves of some important variables such as mean annual temperature and annual precipitation showed the dependence of predicted suitability both on selected variables and on dependencies induced indicated by correlation between the selected variable and other variables (Figure 3.8a and Appendix C). For example, Figure 3.8a indicates a mean annual temperature (cbio01) of 17°C results in highest probability of Psa presence which was in accordance with the CLIMEX results for the optimum temperature range (DV1-DV2) and that reported in the literature (Serizawa et al. 1993). There was also a gradual drop in probability of Psa presence at temperatures above 17 °C where around 23°C the probability of Psa presence becomes zero which is in accordance with the report that Psa growth decreases at temperatures higher than 25°C (Serizawa et al. 1993). The response curve of annual precipitation (cbio12) indicated that more than 3500 mm rainfall per year can decrease the probability of Psa presence (Figure 3.8b).
Figure 3.8. The effect of response curves of (a) annual mean temperature (cbio01) on top and (b) annual precipitation (cbio12) at the bottom in MaxEnt model for Psa. These plots reflect the dependence of predicted suitability both on the selected variable, and on dependencies induced by correlation between the selected variable and other variables. Note that temperature values are multiplied by 10.
Based on the MaxEnt projection, central and eastern China was highly suitable for Psa establishment. Also, small areas in Japan and South Korea were predicted to have a probability of occupancy of more than 50%. Similar to the CLIMEX projection, a small area in the north of Iran was projected as highly suitable by MaxEnt but unlike CLIMEX which predicted some areas in the west of Iran as marginal, these areas were projected as suitable by MaxEnt. In Turkey, narrow strips in the southern coastal area and small areas in the north, including Trabazon Rize province were predicted as suitable (Figure 3.9).

The predictions of CLIMEX and MaxEnt did not match well in Europe, where CLIMEX predicted large areas as marginal or suitable but the MaxEnt prediction involved smaller areas of suitability in Europe compared with CLIMEX. For example, the MaxEnt model failed to predict some areas in Spain and France which are reported to be infected with Psa. For the USA, the pattern of projection was similar to CLIMEX with prediction for suitable areas in California and areas in the south-eastern states. The MaxEnt projection for South America was even more limited compared to CLIMEX where the only areas predicted as suitable were limited to areas around Santa Rosa in Argentina and areas around Santiago in Chile (Figure 3.9).

Suitable areas in Africa were limited to a very small area around Cape Town and in northern African countries, limited to Morocco and the north of Algeria. In Australia, Perth and small areas around Wagga Wagga (NSW) and Albury were projected as suitable.

The greatest discrepancy in prediction between CLIMEX and MaxEnt occurred in New Zealand where for MaxEnt, where no areas predicted to have high probability of Psa establishment. Only some areas around Napier, Palmerston North and north of Kaitaia showed the probability of 30-40% of Psa establishment (Figure 3.9).
Figure 3.9. The potential climate suitability of Psa on a global and New Zealand scale projected by MaxEnt.
3.3.3 Multi-Model Framework (MMF)

Multi-Model is an ensemble framework which produces a table of ten different performance criteria, which are separately ranked and an overall score is calculated from this ranking to find the best-performing model. Table 3.2 shows the variability of the model’s performance according to ten performance criteria calculated by cross-validation and bootstrapping. Although variability among modelling methods is expected, there was a low overall variation in the performance between and among all nine models in MMF. For Psa, the model which had the highest rank in both validation methods was KNN, the results of which are interpreted here.

The area predicted as suitable by the KNN model (K-nearest neighbour) in Asia expanded into Tajikistan, Kirgizstan, Kazakhstan and Israel. All parts of Japan and central and southern parts of Korea and northern and western parts of Iran were projected as suitable. Compared with CLIMEX and MaxEnt larger areas in Europe were projected as suitable with areas expanded into Romania, Ukraine and Hungary. For Africa, the North Africa projection shows a similar pattern to CLIMEX and MaxEnt that Morocco, and Algeria were projected as suitable but in addition, the KNN predicted northern parts of Libya as suitable. Slightly larger areas were indicated as suitable around Cape Town (Figure 3.10).

The predicted suitable areas around Uruguay were smaller than for the other two models and areas in Argentina around Comodoro Rivadavia were projected as suitable although this was not predicted by the other two models. Projection for Chile around Santiago was highly similar to MaxEnt and CLIMEX as was the North America projection in California. In Oceania, the projection for Australia was more similar to CLIMEX where Western parts around Perth and coastal areas around Victoria were suitable. Almost all parts of New Zealand except highlands were projected as suitable which was similar to the CLIMEX projections.
Figure 3.10. The potential climate suitability of Psa on a global and New Zealand scale projected by Multi-Model.
Table 3.2 The accuracy measures achieved by cross-validation (a) and bootstrapping (b) in Multi-Model framework developed for Psa. Acronyms for models are: LDA= linear discriminant analysis, QDA= quadratic discriminant analysis, LOG= logistic regression, NB= naïve Bayes, CART= classification and regression tree, CTREE= conditional tree, KNN= K-nearest neighbor, SVM= support vector machine and NNET= neural network.

(a) Cross-validation

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>rank1</th>
<th>Precision</th>
<th>rank2</th>
<th>Recall</th>
<th>rank3</th>
<th>F.score</th>
<th>rank4</th>
<th>Kappa</th>
<th>rank5</th>
<th>Specificity</th>
<th>rank6</th>
<th>TSS</th>
<th>rank7</th>
<th>Uncertainty</th>
<th>rank8</th>
<th>CV.error</th>
<th>rank9</th>
<th>auc</th>
<th>rank10</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOG</td>
<td>0.942857</td>
<td>4</td>
<td>0.926667</td>
<td>5</td>
<td>0.957778</td>
<td>4</td>
<td>0.999527</td>
<td>4</td>
<td>0.885084</td>
<td>3</td>
<td>0.928788</td>
<td>6</td>
<td>0.886566</td>
<td>4</td>
<td>0.017241</td>
<td>2</td>
<td>0.093064</td>
<td>3</td>
<td>0.955344</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td>NB</td>
<td>0.942857</td>
<td>4</td>
<td>0.944444</td>
<td>3</td>
<td>0.935556</td>
<td>5</td>
<td>0.939117</td>
<td>5</td>
<td>0.880398</td>
<td>5</td>
<td>0.93658</td>
<td>4</td>
<td>0.872136</td>
<td>6</td>
<td>0.011494</td>
<td>1</td>
<td>0.090316</td>
<td>2</td>
<td>0.970408</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>CART</td>
<td>0.895238</td>
<td>7</td>
<td>0.884957</td>
<td>7</td>
<td>0.923492</td>
<td>6</td>
<td>0.899353</td>
<td>7</td>
<td>0.7804</td>
<td>7</td>
<td>0.853463</td>
<td>7</td>
<td>0.776955</td>
<td>7</td>
<td>0.057471</td>
<td>4</td>
<td>0.10657</td>
<td>5</td>
<td>0.905007</td>
<td>7</td>
<td>64</td>
</tr>
<tr>
<td>CTREE</td>
<td>0.933333</td>
<td>6</td>
<td>0.963636</td>
<td>1</td>
<td>0.906984</td>
<td>7</td>
<td>0.929089</td>
<td>6</td>
<td>0.862984</td>
<td>6</td>
<td>0.966667</td>
<td>1</td>
<td>0.873651</td>
<td>5</td>
<td>0.068966</td>
<td>5</td>
<td>0.10824</td>
<td>7</td>
<td>0.926542</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>KNN</td>
<td>0.971429</td>
<td>1</td>
<td>0.961818</td>
<td>2</td>
<td>0.977778</td>
<td>1</td>
<td>0.968185</td>
<td>1</td>
<td>0.942099</td>
<td>1</td>
<td>0.965152</td>
<td>2</td>
<td>0.942929</td>
<td>1</td>
<td>0.034483</td>
<td>3</td>
<td>0.06482</td>
<td>1</td>
<td>0.971859</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>SVM</td>
<td>0.942857</td>
<td>4</td>
<td>0.925455</td>
<td>6</td>
<td>0.965714</td>
<td>2.5</td>
<td>0.942016</td>
<td>3</td>
<td>0.884112</td>
<td>4</td>
<td>0.931818</td>
<td>5</td>
<td>0.897532</td>
<td>3</td>
<td>0.080406</td>
<td>6</td>
<td>0.107642</td>
<td>6</td>
<td>0.939226</td>
<td>5</td>
<td>44.5</td>
</tr>
<tr>
<td>NNET</td>
<td>0.952381</td>
<td>2</td>
<td>0.941818</td>
<td>4</td>
<td>0.965714</td>
<td>2.5</td>
<td>0.952542</td>
<td>2</td>
<td>0.902946</td>
<td>2</td>
<td>0.94697</td>
<td>3</td>
<td>0.912684</td>
<td>2</td>
<td>0.143678</td>
<td>7</td>
<td>0.093511</td>
<td>4</td>
<td>0.952834</td>
<td>4</td>
<td>32.5</td>
</tr>
</tbody>
</table>

(b) Bootstrapping

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>rank1</th>
<th>Precision</th>
<th>rank2</th>
<th>Recall</th>
<th>rank3</th>
<th>F.score</th>
<th>rank4</th>
<th>Kappa</th>
<th>rank5</th>
<th>Specificity</th>
<th>rank6</th>
<th>TSS</th>
<th>rank7</th>
<th>Uncertainty</th>
<th>rank8</th>
<th>CV.error</th>
<th>rank9</th>
<th>auc</th>
<th>rank10</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOG</td>
<td>0.908667</td>
<td>4</td>
<td>0.870756</td>
<td>6</td>
<td>0.93863</td>
<td>3</td>
<td>0.910829</td>
<td>3</td>
<td>0.816098</td>
<td>4</td>
<td>0.882024</td>
<td>7</td>
<td>0.818844</td>
<td>4</td>
<td>0.063218</td>
<td>2.5</td>
<td>0.086317</td>
<td>5</td>
<td>0.958647</td>
<td>6</td>
<td>44.5</td>
</tr>
<tr>
<td>NB</td>
<td>0.909494</td>
<td>3</td>
<td>0.899604</td>
<td>4</td>
<td>0.924518</td>
<td>4</td>
<td>0.910323</td>
<td>4</td>
<td>0.817627</td>
<td>3</td>
<td>0.895</td>
<td>3</td>
<td>0.819559</td>
<td>3</td>
<td>0.04023</td>
<td>1</td>
<td>0.085646</td>
<td>4</td>
<td>0.978729</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>CART</td>
<td>0.887205</td>
<td>6</td>
<td>0.887339</td>
<td>7</td>
<td>0.85166</td>
<td>6</td>
<td>0.886648</td>
<td>6</td>
<td>0.773134</td>
<td>6</td>
<td>0.882547</td>
<td>6</td>
<td>0.775713</td>
<td>6</td>
<td>0.137931</td>
<td>6</td>
<td>0.091567</td>
<td>6</td>
<td>0.957867</td>
<td>7</td>
<td>62</td>
</tr>
<tr>
<td>CTREE</td>
<td>0.884555</td>
<td>7</td>
<td>0.892112</td>
<td>5</td>
<td>0.883301</td>
<td>7</td>
<td>0.883237</td>
<td>7</td>
<td>0.7681</td>
<td>7</td>
<td>0.888166</td>
<td>5</td>
<td>0.771468</td>
<td>7</td>
<td>0.195402</td>
<td>7</td>
<td>0.097353</td>
<td>7</td>
<td>0.96631</td>
<td>5</td>
<td>64</td>
</tr>
<tr>
<td>KNN</td>
<td>0.929329</td>
<td>1</td>
<td>0.903224</td>
<td>2</td>
<td>0.96369</td>
<td>1</td>
<td>0.931242</td>
<td>1</td>
<td>0.915656</td>
<td>1</td>
<td>0.895514</td>
<td>3</td>
<td>0.899231</td>
<td>1</td>
<td>0.123184</td>
<td>2.5</td>
<td>0.063108</td>
<td>3</td>
<td>0.977408</td>
<td>2</td>
<td>17.5</td>
</tr>
<tr>
<td>SVM</td>
<td>0.921106</td>
<td>2</td>
<td>0.903412</td>
<td>1</td>
<td>0.945285</td>
<td>2</td>
<td>0.922405</td>
<td>2</td>
<td>0.841048</td>
<td>2</td>
<td>0.897631</td>
<td>2</td>
<td>0.842916</td>
<td>2</td>
<td>0.09154</td>
<td>4</td>
<td>0.059122</td>
<td>1</td>
<td>0.973841</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>NNET</td>
<td>0.902943</td>
<td>5</td>
<td>0.901091</td>
<td>3</td>
<td>0.90865</td>
<td>5</td>
<td>0.902624</td>
<td>5</td>
<td>0.890501</td>
<td>5</td>
<td>0.898041</td>
<td>1</td>
<td>0.806601</td>
<td>5</td>
<td>0.132184</td>
<td>2</td>
<td>0.062495</td>
<td>2</td>
<td>0.969084</td>
<td>4</td>
<td>40</td>
</tr>
</tbody>
</table>
3.4 Further analysis using Principal Component Analysis

Previously (in sections 3.3.1 and 3.3.3), it was shown that CLIMEX and the KNN model (from the Multi Model) agree on suitability of New Zealand for Psa establishment. Unlike CLIMEX and MMF, the MaxEnt model failed to predict New Zealand as suitable, despite the fact that Psa has already established in most kiwifruit orchards in the North Island and has caused large economic losses. This raises the question as to why MaxEnt, a well-used model, projected a low suitability of Psa establishment in New Zealand. A principal component analysis (PCA) was carried out to visualize the climatic similarity of presence locations with New Zealand climate data. The climate data for presence locations were extracted in ArcGIS Desktop 10.1. A new field was created in the attribute table file, and all extracted presence locations were assigned 1 (as presence) and New Zealand climate points were assigned 2 to differentiate them from the background data. Data were then transferred to R software and analysed using the “princomp” function. The data were transferred back into ArcGIS and the presence points were marked with different symbols to be distinct from the absence points. The results are shown in Figure 3.11. For the purpose of better visualisation, the background points were removed.
Figure 3.11. Psa occurrences plotted against the first two principal components of the environmental data extracted at locations where Psa is present.
3.5 Discussion

With regard to the only available modelling study (EPPO 2012), our model projections coincide with respect to the climate suitability of most areas. There were differences in the parameter values obtained in these two modelling efforts although these differences did not affect the output map significantly. For example, the CLIMEX model indicated that the minimum temperature threshold (DV0) is 5°C which means the pathogen can stay active at this temperature. In the CLIMEX model presented in the EPPO report (2012), the minimum temperature threshold of 10°C (DV0) was reported which contradicts with current findings. Also, the soil moisture parameters in the current study are slightly higher which means that based on current CLIMEX model, the pathogen favours higher soil moisture content compared with the CLIMEX model presented in the EPPO report. Furthermore, in the EPPO report, the optimum temperature range is stated as 20°C to 25°C which is higher than current findings (12°C to 20°C) and those of previous studies regarding the optimum temperature range for Psa growth (Serizawa et al. 1993). Current model parameter values are more in agreement with the ranges mentioned in earlier studies (Cheng et al. 1995; Serizawa et al. 1993). Such differences can arise from different sources. The main reason can be the difference in the number of Psa records used in the model development which can significantly affect results for species that are still spreading. For example, the Spanish records have not been used in the modelling effort mentioned in the EPPO report (EPPO 2012). Occurrence data for Perth in Western Australia have also been used in the EPPO study but to date the occurrence(s) is unconfirmed and has not been included in our CLIMEX model. Despite some differences in parameter values calibrated by these two models, the output maps are similar with the EI value estimated by the EPPO model (EPPO 2012) for New Zealand ranges from 15 to 35 indicating suitable to highly suitable areas for Psa establishment, the EI values estimated by our model ranges from 50 to 90 as highly suitable. Furthermore, our model projected the north of Iran (an area which kiwifruit is grown commercially) as highly suitable but it is not projected as suitable or even marginally suitable by the EPPO study.

Regarding the results of the current study, it is clear from the maps, that CLIMEX and the KNN from the Multi-Model are a good fit to the training data. MaxEnt gave an AUC (0.70) for test data and thus was better than a random model and were reasonably good. In Asia,
where areas in northern Iran were projected as highly suitable by all models, to date, there has been no report of Psa presence. It should be noted, however, that one study mentions the presence of kiwifruit bacterial canker in Iran, but refers to the causal agent as *P. syringae pv. syringae* (Mazarei and Mostofipour 1994). In East Asia where China is the major kiwifruit producer, CLIMEX and the KNN fitted the occurrence data best. At a national level, Yunnan and Guangxi province (where kiwifruit orchards cover 500 to 1000 ha respectively) were projected as suitable by all models but so far there are no reports of Psa from these regions. These results suggest that further investigation is needed to explain why Psa has not been reported from those regions even though many areas in China are affected by Psa. It could be that Psa is present in these locations, and it has not been officially reported or that the model has made an incorrect projection. Small areas in Laos and Vietnam were predicted as marginal by CLIMEX but MaxEnt and the KNN did not predict that these locations were suitable, nor is kiwifruit grown commercially in these regions. In Europe, CLIMEX projected limited areas as suitable but mainly marginal. On the other hand, the KNN projected large areas in Europe as suitable for establishment. All three models agreed on the suitability of some countries such as Italy, Spain, Portugal, France, Turkey and Greece. The Trabazon Rize province in Turkey, from where there is one report of Psa presence (although was not used in the modelling process), is predicted as suitable by all three models. In addition, there has been a recent report from a nursery in Germany (EPPO 2013) that was not used in the modelling process but this location was successfully projected as marginal by CLIMEX and suitable by KNN. All models agreed on the suitability of a small area around Cape Town in South Africa where kiwifruit has been grown commercially in recent years.

In South America, where there have only been two reports of Psa both from Chile, all models agreed on the climate suitability of small areas around Santiago, where Psa has been reported. The discrepancy among model projections was somewhat higher with respect to the climate suitability of Brazil, although no kiwifruit is grown in these areas. For Uruguay, the KNN model projected a very small area as suitable compared to CLIMEX and MaxEnt. In USA all models agreed on the climate suitability of Sacramento and the San Joaquin valley where kiwifruit is grown commercially. However, to date there has been no report of Psa from the USA. For Australia, the prediction was limited to coastal areas in the southeast such as Victoria, where there has been a Psa-LV (low virulent strain) report. In addition, all
models agreed on the suitability of Western Australia (Perth) where kiwifruit is grown commercially on a small scale.

The data for New Zealand were left out of the model training and calibration process and were thus used for model validation in addition to validation methods that are used in modelling approaches. The general agreement between CLIMEX and the KNN from Multi-Model continued in the validation where both models predicted New Zealand as highly suitable for Psa establishment which was in accordance with the current situation of Psa establishment in that country. MaxEnt failed to predict New Zealand and some parts of Spain and France as climatically suitable, despite Psa being established in those regions.

With respect to variables that are important for Psa prediction, MaxEnt and the KNN models were in agreement about the contribution of annual mean temperature for explaining the distribution of Psa. The mean temperature of the coldest month that was selected as among the highest contributing variable by MaxEnt is in accordance with known knowledge about Psa environmental requirements (Serizawa et al. 1993; Serizawa et al. 1993; Serizawa and Ichikawa 1993). The Multivariate Environmental Similarity Surfaces (MESS) analysis (which was performed to see if there are some variables outside the range presented in the training data) did not show the New Zealand environment being different from the current Psa locations. Based on MESS analysis (Appendix C) produced by MaxEnt, only Westland in New Zealand seems to have one or more variables that are outside the range presented in training data, therefore, only the prediction in this area should be treated with strong caution and the projection for the rest areas can be trusted (Elith et al. 2010).

PCA analysis showed that among the presence points, the presence points for New Zealand were quite distinct within the multidimensional environmental space compared with other presence points (Figure 3.11). Three groups can be distinguished. Most presence points from China were in one group, European, Chilean and some points from China comprised the second group and the third group comprising New Zealand point locations which were clearly separate from the first two groups. Only two New Zealand locations in Napier where Psa-LV has been reported seem comparable to the environmental profile of European locations. Usually, for species that may not have changed their environmental range, one expects the presence profiles to be confined to a small area in the multidimensional space. The wide-spread representation of Psa may be an indication that this species is still
spreading, and that we may under-estimate the full range of its potential distribution. This result raises the question whether it is a good idea to apply presence-only models such as MaxEnt that rely on only what is currently known for species that are still spreading. Clearly, further investigation is needed.

Apart from some discrepancies among the models, all three models highlighted suitable areas in which Psa may establish. The high level of agreement between the KNN model (from Multi-Model) and CLIMEX (and in some cases MaxEnt) increased confidence with respect to model projections for novel areas. These results are particularly valuable indicators for USA, Iran, Greece, South Africa, Belgium, and Denmark, where Psa has not been reported. Additionally, despite some limitation of SDM’s, studies that attempt to characterise the environmental needs of potentially damaging biosecurity species such as Psa integrated with supporting expert knowledge can assist decision-makers in kiwifruit growing areas to plan and implement effective strategies for possible incursion of Psa.
Chapter 4
Potential climate suitability for dwarf bunt

4.1 Dwarf Bunt (*Tilletia controversa*)

4.1.1 The disease

Dwarf bunt caused by *Tilletia controversa* is mainly a winter wheat disease that can also infect barley. This disease has always been an economic concern since its first report in 1935 from Montana in the United States (Mathre 1996 cited in Young 1935). Previously it was believed that the pathogen can only infect winter wheat (Wilcoxson et al. 1996) however, according to some recent reports (Bonman et al. 2006), spring-sown wheat can also be infected by this pathogen. Dwarf bunt has a worldwide distribution but so far has not been reported from Australia, China or New Zealand. While this species has impact on yield in most years in affected countries such as the USA, Canada and Russia, the most severe occurrence of this disease occurred in Germany in the 1960s where the losses was reported as 70% (Purdy et al. 1963).

The disease cycle of dwarf bunt is shown in Figure 4.1. Teliospores in the soil germinate to produce an infection hypha, which infects the seedling pre-emergence. Mycelium from the systemic infections later in the season infects the young developing wheat spikelet, which eventually produces sori (cluster of sporangia containing teliospores). Teliospores are released from sori during the harvest period and remain as inoculum for subsequently planted wheat or barley crops (Figure 4.1). Since the spores can survive for almost a decade in the soil or plant material, significant infection is caused by soil infection rather than seed born spores.

The identification of diseased plants can be difficult before the production of the typical symptoms of narrower ears or shorter stunted stems of the infected plants. The general disease symptoms on the wheat plant include dwarfing, with infected plants producing fewer and smaller heads and an unusual odour similar to a foul fish odour (Figure 4.2). The disease cycle starts with germination of the pathogen spores (teliospore) which takes between one to three months. Similar to most other pathogens, temperature and light are the main environmental factors that influence spore germination (Purdy et al. 1963).
Since the middle of the last century control of this disease has mainly been by using resistant cultivars (Hoffmann 1982). Although, other methods such as chemical treatment of seeds has been used as an alternative, use of these chemicals was limited due to their cost along with their low performance (Jia et al. 2013; Bonman et al. 2006; Hoffmann 1982).

Figure 4.1 Dwarf bunt Disease cycle. This is the disease cycle as it is the interaction of the pathogens’ life cycle and the host plant. Source: (Wilcoxson et al. 1996).

Figure 4.2. The effect of dwarf bunt on wheat. (A) A healthy wheat plant compared to wheat plant infected with dwarf bunt. (B) Wheat plant spike infected with dwarf bunt. (C) Dwarf bunt teliospores, the main spore type which causes infection. Image credit: Peterson et al. (2009).
4.1.2 Global Distribution

In comparison with the global area of wheat cultivation, dwarf bunt has only been reported from a small total area, mainly due to the long winter snow coverage that the pathogen requires for survival (Mathre 1996). Nevertheless, yield losses can be as high as 70-80% in some cases (Holton 1941). For example, the area infected with dwarf bunt in Turkey is about 2500 ha and the losses due to this disease are 80% (Yüksel et al. 1980). In Europe, this pathogen has been reported from many countries (Figure 4.3). It was first reported in the Bavarian, Germany in the 1960s (Purdy et al. 1963). This species has not been reported as having established in Australia despite unconfirmed reports of dwarf bunt since 1910 on false barley, a type of grass (*Hordeum leporinum*). Recent investigations however have indicated that in fact it is *Tilletia trabutii* (smut of barley grass) which is present in Australia (www.padil.gov.au). In the west of the USA, dwarf bunt causes major problems in Idaho, Utah, Washington, Oregon, Montana and localized impacts on production in Wyoming and Colorado (Purdy et al. 1963). In New York, the occurrence is sporadic and minor compared to other parts of the United States. While wheat is no longer grown in two Wyoming counties, historical data shows the presence of dwarf bunt in these regions (Peterson et al. 2009). In Canada, this disease is only reported as being present in Ontario and British Columbia (where wheat is grown) but the severity is considered relatively low. Reports from South America of the incidence of this disease have been limited to Argentina and Uruguay (Wilcoxson et al. 1996). The pathogen has been reported from Algeria, Tunisia and Libya but a report of dwarf bunt from Morocco has been claimed to be unreliable (Wilcoxson et al. 1996; Ezzahiri 1992 cited in Lyamani 1990). In Asia, the disease has been reported from Iran, Iraq, Georgia, Japan, Kazakhstan, Kyrgyzstan, Syria, Tajikistan Turkey, Turkmenistan and Uzbekistan (Figure 4.3) (Özkan 2008; Goates 1996; Wilcoxson et al. 1996; Sharifnabi and Hedjaroude 1992; Mardoukhi 1989; Hoffmann 1982; Trione 1982; Yüksel et al. 1980).
4.1.3 Host range

Wheat \((Triticum\ spp.)\) is the main host of dwarf bunt in the USA, but the disease has also been reported on barley (Dewey and Hoffmann 1975). In British Colombia it has been observed on winter wheat and spelt (dinkel wheat) (Peterson et al. 2009 cited in Conners 1953). Other members of the Poaceae family can also be affected by \(T.\ controversa\) namely wheat grass, wild rye and triticale. A list of 68 Poaceae species acting as hosts of dwarf bunt has been reported by Hardison et al. (1959).

4.1.4 Epidemiology

To investigate the habitat suitability of dwarf bunt by models such as CLIMEX, a prior knowledge regarding the environmental needs of the pathogen is necessary. For this reason, existing studies which have determined these factors were reviewed in order to obtain the environmental requirements of dwarf bunt. The majority of existing epidemiological studies of dwarf bunt are limited to the mid-20th century. Dwarf bunt is reported to be confined to regions with altitudes of 300-1000 m, and years with frequent snowfalls are usually associated with serious disease outbreaks (Purdy et al. 1963). Soil compaction and shallow seed sowing also promotes \(T.\ controversa\) infection. Infection primarily occurs in the winter when plants are forming susceptible stem buds (Ian Pascoe 2010). Among the environmental variables, temperature is the most important factor which can influence germination of the pathogen. Lowther (1948) studied the effect of temperature on spore germination of dwarf
bunt. He reported that in *in vitro* assays, spore germination started after five weeks and had reached up to 50% after seven weeks in petri dishes (Lowther 1948). After 70 days incubation at 5°C germination reached up to 75% but increasing the temperature to 10°C inhibited germination with maximum germination seen after incubation at 0°C for 104 days (Lowther 1948). It also has been reported that exposure of spores to temperatures below -5°C or above 20°C results in a delay in germination compared incubation at 3°C (Gassner and Niemann 1954 cited by (Kim 2009). Studies have shown that pre-freezing spores before subsequent incubation at 5°C can also cause a delay in spore germination (Kim KS 2009 cited by Baylis 1958). Hoffman (1982) showed that the temperature range for spore germination of dwarf bunt was from -2°C as the minimum, 3 to 8°C as the optimum with no germination occurring above 15°C. The most favourable conditions for infection were temperatures between 0-8°C (maximum 10-12°C), as found under persistent snow cover. Exposure to temperatures of 15°C inhibited teliospore (syn. ustilospore) germination (Schauz and Rabie 1985). Teliospores can remain viable in the soil for 3-10 years in the absence of wheat (Zscheile 1965).

In addition to temperature, spores require exposure to light in order to germinate. Artificial light can promote germination of dwarf bunt spores (Zscheile 1965; Gassner et al. 1954). Maximum germination occurs when spores are exposed to light during the fifth or sixth week of incubation (Gassner et al. 1954). Increasing light intensity and duration of exposure can stimulate germination even under suboptimal temperatures (Kim 2009). It has also been documented that germination of spores will occur if spores are incubated in darkness after a short period of light exposure during the first stages of incubation. But, exposure to light for a long period increases the percentage germination at shorter incubation time (Baylis 1958).

Another environmental factor that has been reported to affect the growth and establishment of *T. controversa* is snow coverage. According to Mann et al. (2003), during a cold season the effect of snow on ground temperature can be considerable (Mann et al. 2003). As infection of wheat plants by *T. controversa* normally occurs on or near the soil surface, snow coverage can alter the microclimate near the soil surface (Hoffmann 1982). The difference between the air temperature and the space under the snow coverage can amount to 10°C to 20°C (Kim 2009), which means that the soil temperature can remain at 0°C when the air temperature is below the freezing point (Fitton and Brooks 1931) and this is sufficiently warm to allow *T. controversa* to infect the plant. In summary, temperature remains the main factor for
successful germination of dwarf bunt teliospores and the range of suitable temperature reported in literature are summarized as being between -2°C and 15°C with an optimum of 3 – 8°C.

4.1.5 Risk to New Zealand

New Zealand wheat is primarily grown for feed (17,755 ha) and is milled for flour (37,045) (www.far.org.nz). Some wheat grain and the by-products of flour milling, bran and pollard, are used for stock feed. Most autumn and spring wheat is grown in the South Island in the Canterbury region (www.stats.govt.nz). Most of the barley grown in New Zealand is used for the manufacture of stock feed and for malting. Exports of malting and feed barley fluctuate in response to price changes, reflecting international supply and demand. In the last 10 years (2003-2013) the average harvested area of wheat and barley was 45,000 and 59,000 hectares, respectively (www.stats.govt.nz). The average tonnage harvested per annum for both wheat and barley is around 700,000 in total (in the year ended 2012 it was around one million tonnes) (www.stats.govt.nz). The establishment of dwarf bunt in New Zealand can be very important because it can affects the quality of wheat grain (odour) and consequently can affect the livestock feeding and also could restrict export of wheat as some countries have restrictions on the import of wheat from areas where the disease occurs. The high infection level of dwarf bunt in small acreages in the USA, Germany and Turkey indicates that yield loss can be high in susceptible wheat varieties despite small cultivation areas (1% infection causes 0.8% yield loss) (Goates 1996; Mardoukhi 1989; Yüksel et al. 1980). Additionally, the ability of dwarf bunt to survive in soil for up to 10 years, increases the chances of disease outbreaks because normal crop rotation will not be as effective (Murray and Wright 2007). With dwarf bunt establishment, additional indirect costs can be imposed from using resistant varieties that may have lower yield or through the need to use systemic fungicides that increase the cost of control (Mathre 1996).

4.1.6 Previous modelling of dwarf bunt

Given the importance of dwarf bunt, there have been several attempts to model the risk of dwarf bunt establishment to global areas, mainly on local scales, using different methods. In Australia, in response to call for an national contingency plan for the possibility of a dwarf bunt incursion, the chance of dwarf bunt establishment was evaluated using CLIMEX (Murray et al. 2007). The study concluded that the chance of establishment in Australia was
In a study by Kim and Beresford (2009) the potential establishment of dwarf bunt in New Zealand and other regions of the world was evaluated. Kim and Beresford (2009) also used CLIMEX and cloud and snow coverage in their analysis. In another study in the USA, Peterson et al. (2009) used TCK-GM risk model (\textit{T. controversa} Kühn-Geophytopathological Model) to evaluate introduction scenarios to identify the regions where dwarf bunt might establish in the USA. They identified the regions where the disease may establish at a low level but where significant yield loss would not be anticipated. Peterson et al. (2009) also tried to assess the risk of dwarf bunt importation into China. They concluded that only 3.8% of the wheat growing areas in China are at risk of dwarf bunt establishment, and the potential yield loss of wheat was estimated as 1.3%. Since trade between China and the USA is high and there are already records of the presence of dwarf bunt from the USA, this has led to many studies evaluating the potential and likelihood of dwarf bunt establishment in China. In the first study, by using weather satellite data (snow coverage and ice charts), the risk of dwarf bunt establishment in China was evaluated as being extremely unlikely (Trione and Hall 1986). In contrast, using soil moisture and temperature of the surface layer of soil, the potential of dwarf bunt introduction was suggested as being considerably higher by Zhang et al. (1995). Similarly, Shuqiu et al. (1995) evaluated the potential areas of dwarf bunt establishment in China using a bio-climatic analogical distance model. In their model, they calculated the bio-climatic analogical distance between the Spokane area (Washington State) and various wheat growing areas in China. Their results showed that some wheat growing areas such as the lower reaches of Huanghe, Yangzi River and the Huaihe river valley which are major winter wheat growing locations, had very high risk of disease establishment. In a model developed by Chen et al. (2001), using geographic information system (GIS) and simulation models, the areas with high and moderate risk of dwarf bunt in China were estimated as being 19.3%. More recently, the risk of dwarf bunt establishment in China has been assessed by Jia et al. (2013). They used geo-phytopathological models and meteorological data (soil temperature and soil moisture) from about 600 weather stations in China to visualize the risk for wheat growing areas. Unlike the model developed by Peterson et al. (2009) where daily mean temperature was used, the model of Jia et al. (2013) used accumulated hourly temperature to better estimate the effect of temperature on teliospore germination. They classified the wheat growing areas into high, moderate, low and very low risk areas and concluded that 27.33 and 27.69% of wheat growing area have high and
moderate risks respectively. Although in all of these studies different environmental variables have been used to evaluate the risk of dwarf bunt establishment in different areas, all these studies appreciate the important role of temperature in dwarf bunt establishment. Further comparative studies are needed to evaluate the effect of different variables on dwarf bunt model outputs to investigate how and to what extent the used variables actually contribute to dwarf bunt establishment.

4.2 Methods

4.2.1 Dwarf bunt occurrence data

The current geographical distribution of dwarf bunt was gathered from a range of different studies (Peterson et al. 2009; Bonman et al. 2006; Mathre 1996; Trione and Ching 1971; Purdy et al. 1963). Only forty-six geo-referenced location points were available in online databases such as CABI and GBIF (including very close points which were later removed as duplicates). Most of the studies that have investigated the occurrence, biology and epidemiology of dwarf bunt are old, making it difficult and time-consuming to locate the authors to ask them to confirm the more precise location of occurrences. In many instances Google Earth 5.1.35 and ArcGIS 10 were used to extract 44 coordinates from published studies. After removing duplicate points a total of 88 points were used in the modelling process (Appendix D).

4.2.2 CLIMEX

The same procedure described in Chapter 2 (methodology), was used for setting the parameters for dwarf bunt in the CLIMEX model. As a starting point, initial parameter values for various thresholds reported in the published literature were used to fit CLIMEX parameters (Hoffmann 1982; Gassner et al. 1954). Europe is the native range for this pathogen where it is widespread. The parameters were fitted to find the closest match with dwarf bunt’s current distribution comprising its native and invaded range. The model was calibrated using 78 presence points, and the presence locations in North America (n = 11) were used to test the model. The temperate soil moisture template as a representative of Europe soil moisture (existing in the CLIMEX model) was used as a starting point to fit the soil moisture values. Heat stress was also used to exclude the persistence of dwarf bunt in areas such as southern parts of the USA (southern California and Florida), Mexico, and
northern parts of South America and northern parts of Africa. Cold stress seems only important to dwarf bunt in eastern and north-eastern parts of Russia. The resulting parameters (Table 4.1) were used to determine the risk of dwarf bunt establishment in novel areas.

4.2.3 Correlative models

4.2.3.1 Background selection for correlative models (MaxEnt and Multi-Model)

Based on the methodology outlined in the second chapter, for dwarf bunt the exploratory principal component analysis was performed to define the appropriate background extent for background point’s selection. The principal component analysis was used to identify which of the variable had the most important information. The analysis was performed for different distances shown in Figure 4.4. After performing the PCA analysis, the changes in the second component of important variable (cbio12 = annual precipitation) was plotted across different distances. In this case, this obvious change was observed in the second component at 1000 km. Therefore, the correlative model will select the background or pseudo-absence points from up to 1000 km around the presence points.

Figure 4.4. Changes in the important variable (annual precipitation) over the distance. The graph shows that after 1000 km, no significant change happened in the important variable.
4.2.3.2 MaxEnt model

4.2.3.2.1 Variable selection

The candidate variables for MaxEnt were identified using a pairwise Pearson correlation test to exclude highly correlated variables. The correlation test was performed on all background points located in the 1000 km buffer around presence points \((n = 105790)\) and variables with pairwise correlation of higher than 0.70 were discarded (Appendix D). This process resulted in selection of nine variables with low correlation: annual mean temperature (cbio1), temperature seasonality (cbio04), min temperature of the coldest month (cbio06), mean temperature of the wettest month (cbio08), mean temperature of the warmest quarter (cbio10), annual precipitation (cbio12), precipitation of the driest month (cbio14), precipitation of the wettest quarter (cbio16) and precipitation of the coldest quarter (cbio18). It was noticed that Elith et al. (2013) suggested that it is better not to use variables that are a combination of rain and temperature (such as temperature of wettest month) as e.g. it is difficult to define the precise wettest month in practice and hence it makes interpretation more difficult (Elith et al. 2013). In contrast, to this suggestion, Phillips et al. (2005) believe that when we model at a continental or global scale, although a shift in exact timing of seasonal rainfall may occur, thereby affecting monthly precipitation, this won’t affect the suitable conditions for the species and for this reason quarterly data were used in the MaxEnt model. Furthermore, it has also been suggested that using quarterly precipitation variables would also improve the model transferability to estimate the future distribution of the species under climate change scenarios because suitable conditions for target species may not depend only on precise monthly precipitation values but on rainfall consistency or wet periods (Phillips 2005).

Because the number of presence locations comprised a large sample, linear, quadratic and hinge features were used to allow more complexity in the model and jackknife was used to investigate the variable importance among the nine selected variables. These settings were used based on experience using the MaxEnt model which will affect the response curves and consequently the model output. Additionally, response curves were produced to show how each variable can affect the MaxEnt projection. A Multivariate Environmental Similarity Surfaces (MESS) analysis was performed to compare the environmental similarity of variables used in MaxEnt model (Appendix D) (Elith et al. 2010). This will highlight the areas...
where one or more variable are outside the range presented in the training data, therefore, warning will be given that predictions in those areas should be treated with strong caution (Elith et al. 2010).

### 4.2.3.3 Multi-Model Framework

The procedure in the methodology chapter (Chapter 2) was used to run the Multi-Model framework. First the duplicate presence points were removed based on the resolution of the data which led to 87 presence points. Then, the same number (87) of pseudo-absence points were generated from a 1000 km buffer around the presence points using OCSVM and k-mean clustering. Variable selection was performed by random forest and step-wise regression. Models were trained using the presence and generated pseudo-absence points and were validated by cross-validation and bootstrapping techniques.

#### 4.2.3.3.1 Multi-Model variable selection

The variables selected by the Multi-Model Framework for dwarf bunt both by random forest and step-wise regressions were: isothermality (cbio03), precipitation of the wettest month (cbio13) and precipitation seasonality (cbio15).

### 4.2.4 Further analysis using Principal Component Analysis (PCA)

A principal component analysis was performed on dwarf bunt presence points to determine their similarity in environmental space. The climate data of presence points were extracted in ArcGIS Desktop 10.1. To compare the similarity of presence points with New Zealand climate, two fields (presence and NZ) were added to the climatic data to differentiate the presence points from New Zealand data. This enabled better visualization and comparison of points of interest. As Figure 4.5 shows, some of the presence points lie close to or within the New Zealand data space, which implies the climate similarity of these presence points with New Zealand data.
4.3 Results

4.3.1 CLIMEX Results

The final parameter values obtained by the parameter fitting process in the CLIMEX model for dwarf bunt are shown in Table 4.1. The CLIMEX projection showed that the potential...
distribution of dwarf bunt extends through Europe, East Asia, North America and Oceania (Figure 4.6) giving a good fit to the current distribution of dwarf bunt. The North America occurrence data were used to test the model \( n = 11 \). For the test data, 10 out of 11 points were projected suitable to marginally suitable which indicates CLIMEX performed well for the test data. The Ontario location in the test data was projected as unsuitable. The reason CLIMEX predicted this area as unsuitable might be because the centroid of the Ontario province was used as a surrogate for species occurrence. The test data from Oregon, Utah, Idaho, Washington, New York and British Colombia are projected as suitable. Georgia, Colorado, Wyoming and Montana were projected as marginally suitable. The model suggests that southern parts of Ontario are marginally suitable for dwarf bunt establishment. In general, CLIMEX confirmed current reports by indicating the most vulnerable area for dwarf bunt establishment is parts of Europe. Parameter fitting of cold stress revealed that most parts of western China along with areas in Wyoming and Utah in USA would present the highest cold stress to this species, but Lowther (1948) and Peterson (2009) suggest that generally cold climate is not a limiting factor for dwarf bunt establishment in most parts of the world. However, the parameter fitting process showed that heat stress may limit dwarf bunt establishment in areas such as India, East Asian, Mexico, and southern California, southern Florida and northern parts of North America.
Table 4.1. Parameter values used to develop a CLIMEX model for dwarf bunt.

<table>
<thead>
<tr>
<th>Index</th>
<th>Parameters</th>
<th>Current study</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>DV0= lower threshold</td>
<td>-3</td>
<td>0°C</td>
</tr>
<tr>
<td></td>
<td>DV1= lower optimum temperature</td>
<td>3</td>
<td>0°C</td>
</tr>
<tr>
<td></td>
<td>DV2= upper optimum temperature</td>
<td>8</td>
<td>0°C</td>
</tr>
<tr>
<td></td>
<td>DV3= upper threshold</td>
<td>13</td>
<td>0°C</td>
</tr>
<tr>
<td>Moisture</td>
<td>SM0 = lower soil moisture threshold</td>
<td>0.1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>SM1= lower optimum soil moisture</td>
<td>0.5</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>SM2= upper optimum soil moisture</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>SM3= upper soil moisture threshold</td>
<td>1.6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>TTCS= Cold stress temperature threshold</td>
<td>-5</td>
<td>°C</td>
</tr>
<tr>
<td>Cold Stress</td>
<td>THCS= Cold stress temperature rate</td>
<td>-0.00005</td>
<td>Week⁻¹</td>
</tr>
<tr>
<td></td>
<td>DTCS= Cold stress degree-day threshold</td>
<td>15</td>
<td>°C</td>
</tr>
<tr>
<td>Heat stress</td>
<td>TTHS = Heat Stress temperature threshold</td>
<td>20</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>THHS= Heat Stress temperature rate</td>
<td>0.00005</td>
<td>Week⁻¹</td>
</tr>
<tr>
<td>Wet Stress</td>
<td>SMWS= Wet Stress threshold</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>HWS= Wet stress accumulation Rate</td>
<td>0.001</td>
<td>Week⁻¹</td>
</tr>
</tbody>
</table>

Among all of the locations, the highest EI values were found in Argentina (Islas Año Nueva, 64) and Chile (Punta Arenas, 52). The potential range was wider in Europe; however, for most parts the EI values were more than 26, which, by CLIMEX standards, indicate suitable climate conditions for establishment of this pathogen (Sutherst and Gunter 2005; Hoddle 2004; Sutherst 1999). The most suitable locations were aggregated around Germany, Netherlands, France, United Kingdom, Norway, Spain, Sweden and Denmark. However, so
far, there are no reports of the pathogen’s presence in the Netherlands, France or the United Kingdom. In Europe, the highest EI value was located at Wick in the UK (41), Iceland (Reykjavik 39) and Norway (Laerdal 39). However, no record of dwarf bunt in these areas exists in the literature. In North America, southern parts of Canada and northern USA were marginal, but the east and north-western USA was projected as suitable for dwarf bunt establishment. The highest EI values were determined for St Paul Island (39) in USA and Victoria in Canada (23).
Figure 4.6. Global potential climate suitability of dwarf bunt projected by CLIMEX model.
In South America, the most suitable locations were projected in Argentina (Islas Año Nuevo, 64), Chile (Punta Arena, 52 and Balmaceda, 31) and Peru (Lapos Bajo, 30). In general, the southern parts of South America were projected as very suitable for establishment of dwarf bunt.

Most parts of the African continent were predicted unsuitable for the establishment of dwarf bunt. In Africa, just northern parts of Algeria with EI = 11 was projected as suitable with the northern part of Algeria and a few locations in central Morocco fall into the marginal category. All other parts of Africa were unsuitable. The favourable locations in Asia included almost all parts of Japan and southern parts of South Korea (Pusan, 22) and Ninjiang province in China (20). Furthermore, suitable areas were restricted to small areas of Uzbekistan, Kazakhstan and northern India (Srinagar) and northern parts of Iran.

In Oceania, south-eastern Australia (such as Canberra and Cann River area) and central parts of Tasmania were categorized as suitable (Figure 4.6). For New Zealand, the suitable area in the North Island was limited to Rangitikei. The South Island was mainly predicted as suitable and marginal with central Otago being highly suitable. The suitable area extended from Marlborough to Southland and very suitable areas are limited to Waimate, central Otago and Dunedin. Most parts of the Westland region of New Zealand were predicted as unsuitable (Figure 4.6).

Figure 4.7 shows the suitable growth conditions for population growth by dwarf bunt in Central Otago. Suitable conditions for germination of dwarf bunt spores are April to September with the peak in May indicating optimal temperature and soil moisture conditions (Figure 4.7). This is in agreement with the low temperature experienced in Central Otago at this time of year providing good conditions for dwarf bunt establishment which is known to favour low temperatures.
Figure 4.7. Temperature index, soil moisture index and weekly growth index for dwarf bunt in Central Otago.
### 4.3.2 MaxEnt Results

The MaxEnt model showed a test AUC of 0.75, which is reasonable and means the model is performing better than chance. The AUC of the training data was also good (0.84) (Appendix D). The jackknife of regularized training gain showed that the most important variables were annual mean temperature (cbio01), temperature seasonality (cbio04), minimum temperature of the coldest month (cbio06), precipitation of the coldest quarter (bio19) and mean temperature of the warmest quarter (cbio10) (Figure 4.8a). Figure 4.8a also shows when the model was trained only by mean temperature of the wettest quarter (cbio08), there was almost no gain. The selected variables were also biologically important, as for example dwarf bunt favours cold climates and prefers high-altitude habitats.

Jackknife of the test data showed that the mean temperature of the warmest quarter (cbio10) achieved a negative gain on test data although it helped MaxEnt to get a good fit to the training data (Figure 4.8b). This indicates that model build with this variable is less transferable. The finding that the suitable conditions for dwarf bunt are not dependent on mean temperature of the warmest quarter is compatible with what is known about the pathogen (with the disease cycle of dwarf bunt not favoured by warm temperatures) (Peterson et al. 2009) (Figure 4.8b). The jackknife of test data re-confirms that annual mean temperature (cbio1), minimum temperature of the coldest month (bio6), temperature seasonality (cbio4) and precipitation of coldest month (cbio19) were the most important variables. These variables which helped MaxEnt to obtain a good fit to training data and also helps the model to generalize better. The jackknife of AUC shows that when annual mean temperature (cbio1) and mean temperature of the warmest quarter (cbio10) were not used, the AUC of the model increased, indicating the predictive performance of the model will improve when corresponding variables are not used (Figure 4.8c).
Figure 4.8. Jackknife of MaxEnt model on (a) regularized training data (b) jackknife on test data (c) jackknife of AUC.

Response curves of single variables on MaxEnt projection can be viewed in Figure 4.9 and
Appendix D. The response curve of annual mean temperature (cbio01) shows that when the annual mean temperature is -1.5°C, conditions starts to become suitable for growth of the pathogen but the probability of presence would be above 0.5 (more than chance) when annual mean temperature are around 5°C. Growth increases up to around 9°C and decreases afterwards. This is in agreement with the suitable range suggested by CLIMEX (3-8°C as optimum temperature range). On the other hand, the annual temperature curve also shows that above 12°C the probability of dwarf bunt presence would be lower than 0.5 (less than chance).

The response curves showed that the probability of dwarf bunt presence decreases at certain levels for temperature variables including annual mean temperature (cbio01)(Figure 4.9a), temperature seasonality (cbio04), minimum temperature of the coldest month, mean temperature of wettest quarter (cbio08) and mean temperature of warmest quarter (cbio10) (Appendix D). This can indicate the sensitivity of this species to temperature (Figure 4.9a). The probability of dwarf bunt presence also decreases when annual precipitation exceeds 500 mm (Figure 4.9b) which suggests that excessive rain depends when the rainfall occurs) and consequently very high soil moisture may not be ideal for the pathogen establishment which has been documented for some plant pathogens (Lee and Huang 1973).

Similar to CLIMEX, the most suitable area is Europe followed by areas in China, Japan, Korea, Tajikistan, Kyrgyzstan, Uzbekistan and Iran in Asia and western and central USA and small areas around Michigan are projected as suitable (Figure 4.10). MaxEnt projection does not show any suitable areas for dwarf bunt establishment in New Zealand. Only a very small area in central Otago is projected with a probability of between 0.2-0.4 for possible establishment of dwarf bunt (Figure 4.10).
Figure 4.9. The effect of response curves of a) annual mean temperature (cbio01) and b) annual precipitation (cbio12) for the MaxEnt model for dwarf bunt. These plots reflect the dependence of predicted suitability both on the selected variable, and on dependency.
Figure 4.10. Potential climate suitability of dwarf bunt in New Zealand projected by MaxEnt.
4.3.3 Multi-Model results

The variability in each model’s performance for 10 performance criteria (cross-validation and bootstrapping) is shown in Table 4.2. The QDA model that had the highest rank in both validation methods was selected for interpretation (Table 4.2). The model projection (Figure 4.11) showed that all of Europe, except for small areas in Portugal, and western Turkey, is highly suitable for dwarf bunt establishment. The projection extended towards Russia, and a large area in Russia is projected as suitable in Asia, northern Iran, small parts in Afghanistan and very small areas in China and Japan are projected as suitable. Vast areas in the USA and Canada are projected as highly suitable. The probability of dwarf bunt occurrence in Australia was projected to be low, with the exception of very small areas in New South Wales, which had a probability of 50-60% establishment. For New Zealand, a very small area in central Otago is highly suitable and the probability of dwarf bunt occurrence in the rest of the South Island (eastern coastal areas) was projected between 0.2-0.4. In summary, the global projection by Multi-Model Framework (MMF) showed larger suitable areas in Russia, USA and Canada and Europe. In addition, the projection of the Multi-Model for New Zealand is, to some extent, similar to MaxEnt projection but both are in contrast with CLMEX which projected South Island suitable to very suitable (Figure 4.11).
Figure 4.11. Global potential climate suitability of dwarf bunt projected by quadratic discriminant analysis (QDA) model.
A)

Table 4.2. The accuracy measures achieved by cross-validation (A) and bootstrapping (B) in Multi-Model Framework developed for dwarf bunt. Acronyms for models are: QDA= quadratic discriminant analysis, NB= naïve Bayes, LDA= linear discriminant analysis, LOG= logistic regression, CART= classification and regression tree, CTREE= conditional tree, KNN= K-nearest neighbor, SVM= support vector machine, and NNET= neural network.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy rank1</th>
<th>Precision rank2</th>
<th>Recall rank3</th>
<th>F.score rank4</th>
<th>Kappa rank5</th>
<th>Specificity rank6</th>
<th>TSS rank7</th>
<th>Uncertainty rank8</th>
<th>CV.error rank9</th>
<th>auc rank10</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>QDA</td>
<td>0.888889</td>
<td>1</td>
<td>0.93</td>
<td>1</td>
<td>0.935065</td>
<td>1</td>
<td>0.810823</td>
<td>1</td>
<td>0.027778</td>
<td>3</td>
<td>17.5</td>
</tr>
<tr>
<td>NB</td>
<td>0.877778</td>
<td>2.5</td>
<td>0.910556</td>
<td>2</td>
<td>0.916883</td>
<td>2.5</td>
<td>0.792641</td>
<td>2</td>
<td>0.033333</td>
<td>4</td>
<td>27.5</td>
</tr>
<tr>
<td>NNET</td>
<td>0.877778</td>
<td>2.5</td>
<td>0.853463</td>
<td>6</td>
<td>0.916883</td>
<td>2.5</td>
<td>0.770346</td>
<td>3</td>
<td>0.122222</td>
<td>7</td>
<td>17.5</td>
</tr>
<tr>
<td>KNN</td>
<td>0.866667</td>
<td>4</td>
<td>0.945455</td>
<td>7</td>
<td>0.7716</td>
<td>3</td>
<td>0.721494</td>
<td>4</td>
<td>0.157065</td>
<td>1</td>
<td>38.5</td>
</tr>
<tr>
<td>SVM</td>
<td>0.844444</td>
<td>5</td>
<td>0.880231</td>
<td>8</td>
<td>0.840191</td>
<td>8</td>
<td>0.680544</td>
<td>5</td>
<td>0.893939</td>
<td>6</td>
<td>52</td>
</tr>
<tr>
<td>LDA</td>
<td>0.822222</td>
<td>7</td>
<td>0.821624</td>
<td>8</td>
<td>0.83423</td>
<td>6</td>
<td>0.634059</td>
<td>7</td>
<td>0.680425</td>
<td>4</td>
<td>61.5</td>
</tr>
<tr>
<td>LOG</td>
<td>0.822222</td>
<td>7</td>
<td>0.855556</td>
<td>6</td>
<td>0.831522</td>
<td>7</td>
<td>0.637042</td>
<td>6</td>
<td>0.841558</td>
<td>2</td>
<td>75</td>
</tr>
<tr>
<td>CART</td>
<td>0.822222</td>
<td>7</td>
<td>0.80127</td>
<td>9</td>
<td>0.87013</td>
<td>4</td>
<td>0.824504</td>
<td>8</td>
<td>0.75671</td>
<td>9</td>
<td>75</td>
</tr>
<tr>
<td>CTREE</td>
<td>0.788889</td>
<td>9</td>
<td>0.88</td>
<td>5</td>
<td>0.737013</td>
<td>9</td>
<td>0.586675</td>
<td>9</td>
<td>0.845489</td>
<td>5</td>
<td>82</td>
</tr>
</tbody>
</table>

B)

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy rank1</th>
<th>Precision rank2</th>
<th>Recall rank3</th>
<th>F.score rank4</th>
<th>Kappa rank5</th>
<th>Specificity rank6</th>
<th>TSS rank7</th>
<th>Uncertainty rank8</th>
<th>X.632...error rank9</th>
<th>auc rank10</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN</td>
<td>0.851342</td>
<td>4</td>
<td>0.816102</td>
<td>6</td>
<td>0.90877</td>
<td>1</td>
<td>0.857679</td>
<td>4</td>
<td>0.796836</td>
<td>7</td>
<td>12.8266</td>
</tr>
<tr>
<td>SVM</td>
<td>0.850811</td>
<td>3</td>
<td>0.849715</td>
<td>4</td>
<td>0.86878</td>
<td>2</td>
<td>0.828753</td>
<td>3</td>
<td>0.845708</td>
<td>4</td>
<td>16.6666</td>
</tr>
<tr>
<td>QDA</td>
<td>0.886121</td>
<td>1</td>
<td>0.88063</td>
<td>1</td>
<td>0.846978</td>
<td>6</td>
<td>0.861444</td>
<td>1</td>
<td>0.730954</td>
<td>1</td>
<td>19.5</td>
</tr>
<tr>
<td>NNET</td>
<td>0.849432</td>
<td>5</td>
<td>0.850343</td>
<td>3</td>
<td>0.847421</td>
<td>4</td>
<td>0.697843</td>
<td>5</td>
<td>0.850504</td>
<td>3</td>
<td>43</td>
</tr>
<tr>
<td>NB</td>
<td>0.85945</td>
<td>2</td>
<td>0.865801</td>
<td>2</td>
<td>0.850149</td>
<td>5</td>
<td>0.856143</td>
<td>4</td>
<td>0.717616</td>
<td>2</td>
<td>30.5</td>
</tr>
<tr>
<td>CART</td>
<td>0.814168</td>
<td>6</td>
<td>0.809806</td>
<td>7</td>
<td>0.830351</td>
<td>8</td>
<td>0.81407</td>
<td>6</td>
<td>0.628598</td>
<td>6</td>
<td>66</td>
</tr>
<tr>
<td>CTREE</td>
<td>0.804688</td>
<td>7</td>
<td>0.836975</td>
<td>5</td>
<td>0.769599</td>
<td>9</td>
<td>0.793258</td>
<td>9</td>
<td>0.608129</td>
<td>7</td>
<td>71</td>
</tr>
<tr>
<td>LOG</td>
<td>0.801592</td>
<td>8</td>
<td>0.785771</td>
<td>8</td>
<td>0.832659</td>
<td>7</td>
<td>0.805196</td>
<td>8</td>
<td>0.602429</td>
<td>8</td>
<td>75</td>
</tr>
<tr>
<td>LDA</td>
<td>0.799667</td>
<td>9</td>
<td>0.76846</td>
<td>9</td>
<td>0.859736</td>
<td>3</td>
<td>0.888082</td>
<td>7</td>
<td>0.596761</td>
<td>9</td>
<td>76</td>
</tr>
</tbody>
</table>
4.4 Discussion

The most important part of the modelling process for species distribution models is to gather accurate data with respect to the current locations of the target species. In this study, two main data sources were used: CABI (Centre for Agriculture and Biosciences International) and the Global Biodiversity Information Facility (GBIF). The GBIF database for plant pathogens is not as rich as it is for insects or other macroscopic species, so only 10 exact presence locations exist for dwarf bunt (out of 90 locations), all of which were in Europe. On the other hand, the CABI database was not adequate as the locations were mainly recorded at the administrative regional level. All databases contain some level of error (Pascoe et al. 2012) and identification of these from the start was attempted. For example, a report from central and Western Australia in the CABI 2008-2009 Crop Protection Compendium database does not correspond to data used by Kim and Beresford (2009), who used the 2008-2009 Crop Protection Compendium (CPCI 2008). Their study included presence three points in Australia that we believe reveal a source of error in their model. These locations are not present in the current CABI database and have now been removed from the database (CPCI 2012).

The results predicting high likelihood for spread in Europe were expected, as many reports already suggest that the region is generally very suitable for establishment of dwarf bunt (Peterson et al. 2009; Trione 1982). Also, based on CLIMEX projections, some parts of New Zealand, Argentina and Southern Chile are projected as highly suitable. However, the model predicted locations in northern Africa and Argentina to be marginal, yet the pathogen is already established in these areas. Predictions of suitable areas did correspond to actual distributions in both the western and eastern USA. However, in southern parts of South America (Brazil and Chile) predicted to be very suitable, suitable or marginal, there are no reports of the occurrence of dwarf bunt in these areas.

Compared with the Kim and Beresford model (2009) for New Zealand the CLIMEX model in this study is in accordance with the predictions for some areas in the South Island such as Lake Wanaka and Lake Tekapo as being suitable, where most parts of the North Island are unsuitable or marginal for establishment of dwarf bunt. However, there are more differences between the models when the South Island is considered. The CLIMEX model in this study predicts that most parts of the South Island are suitable for the establishment of dwarf bunt. For the South Island, the model predicts that northern parts of the South Island are relatively
suitable but small parts in the south (Dunedin, Queenstown and Oamaru) are very suitable for the establishment. However, while these areas are suitable, very little wheat is grown in these parts of the country (www.stats.govt.nz) (Statistics 2013). Most of the wheat growing areas in New Zealand are located in the south-eastern part of the South Island such as Canterbury (more than 90% of wheat) from Cheviot to Oamaru, Invercargill and Balfour (www.stats.govt.nz) (Statistics 2013). In North Island, wheat is grown in southern Hawke’s Bay, Manawatu and Wairarapa. The Kim and Beresford (2009) model predicted two areas in the South Island as being suitable. These areas were around Lake Tekapo and Lake Wanaka, projected to be very suitable and suitable, respectively. While both models predict that southern parts of the South Island are suitable for dwarf bunt establishment, however if the host plant, wheat, is not grown nearby, the chance of establishment is very low (the closest wheat paddocks to Lake Tekapo would be in Fairlie, about 40 km away). The likelihood remains low even if alternative host plants are considered, such as the genera Poa, Secale and Lolium. A high density of pathogen (20,000 spores/ gram of wheat seed) is considered necessary for successful infection and establishment (Trione 1982). This level would rarely occur by random distribution of spores (Goates 1996; Trione 1982). Additionally, both current models and Kim and Beresford model (2009) generally agree on predictions for Australia where establishment potential is rated as extremely low (Pascoe et al. 2012). It seems that Kim and Beresford (2009) projected smaller suitable areas in New Zealand as suitable because they used snow cover as an important variable and ignored the role of soil moisture content in their CLIMEX model. Although we acknowledge the importance of snow cover to spore germination and development, modelling dwarf bunt with just temperature and snow coverage when there are other important factors such as soil moisture and stress indices, may under-estimate the number of suitable areas. The fact that snow coverage has been reported as an important factor which indirectly facilitates germination of spores and dwarf bunt establishment, can apply to any pathogen which tolerates low temperatures close to freezing points and should not be specific for dwarf bunt. In a similar study, Trione et al.(1986) used snow and ice charts from weather satellites to investigate the potential areas for dwarf bunt establishment in China. They concluded that it was extremely unlikely for dwarf bunt to establish in China, while in a recent study which used soil surface temperature, mean soil surface humidity and number of snow-coverage days, about 27% of the total winter-wheat growing areas in China were identified as high risk (Jia et al. 2013).
This finding contradicts Kim and Beresford (2009) and Trione et al. (Trione et al. 1986) whose models predicted a complete lack of suitable establishment areas in China. These differences highlight the importance of candidate variable selection that is further discussed in Chapter 9.

Despite some advantages of the CLIMEX model, such as the ability to use physiological data, parameters such as stress and their interaction are not usually available, which makes the parameter-fitting process subjective. Another important issue in CLIMEX is the subjectivity in climate suitability inferred from the EI value, which is somewhat arbitrarily chosen by the modeller. Different studies used different EI ranges to define the marginal, suitable and very suitable areas (Yonow et al. 2013; Venette et al. 2006). This makes it difficult to compare results among studies.

In the correlative model, each model selected different variables as being important. MaxEnt emphasized the role of temperature in disease development while the Multi-Model mainly selected variables related to precipitation. However, Multi-Model picked isothermality (cbio03) as an important variable and this variable by itself is a reflection of the mean monthly temperature, the minimum temperature of the coldest month and the annual temperature range.

On a global scale, the difference in the projections of three models was highlighted in some areas such as the USA, China and Canada. For example, unlike CLIMEX and MaxEnt, Multi-Model did not predict China as suitable or, for USA and Canada, compared to CLIMEX and MaxEnt, larger areas are predicted as suitable by Multi-Model. Additionally, in contrast to CLIMEX, MaxEnt does not predict New Zealand as a suitable location for dwarf bunt establishment while the Multi-Model projected a small area in central Otago as suitable. The Multi-Model projected larger areas in central Otago as suitable with the probability of 50-60% with similar areas projected as highly suitable by CLIMEX. It should be noted that MaxEnt model is usually used to predict at the local level, and has rarely been used on a global scale so the performance of the MaxEnt model and its ability in extrapolating to novel areas on a large scale requires more studies (Elith et al. 2013). Additionally, the principal component analysis (PCA) results give us more insight into the possibility of dwarf bunt establishment in New Zealand (Figure 4.5). As some of the dwarf bunt presence points (red dots in Figure 4.5) are close to New Zealand data therefore it can be concluded that, some
areas are climatically suitable for dwarf bunt establishment, which increases our confidence regarding the models projections for other areas that give similar suggestions.

In summary, the projections of the three models agreed on the suitability of most areas in the Northern Hemisphere, but not in the Southern Hemisphere. In New Zealand, the models agreed that there was high suitability for dwarf bunt establishment in central Otago but for the rest of New Zealand the models were not in agreement. Most autumn-sown wheat is sown from April to May in New Zealand (so would be up by May-June). Although, not much wheat growth occurs during the colder months but the host plants would be available for infection to occur. The latest statistics NZ data shows that there was 2,600 ha of wheat grown in Otago in the year ended 30 June 2013 (41,500 ha in Canterbury). Barley, triticale and 68 Poaceae species have also been indicated as hosts to \textit{T. controversa} overseas, and 5,700 ha of barley is grown in Otago, there may well be enough hosts for the pathogen to establish and spread if it were to arrive in New Zealand.
5.1 Guava rust (*Puccinia psidii*)

5.1.1 The disease

Guava rust (*Puccinia psidii*) (common names: eucalyptus rust, guava rust, Myrtaceae rust, ohi`a rust (Hawaii)) was first described from common guava in Brazil (Winter 1884) with the native range of guava rust reported as being South and Central America and the Caribbean with most reports of its occurrence coming from these regions (Coutinho et al. 1998). The first serious outbreak of guava rust was reported from nurseries in Espírito Santo province in Brazil in 1973 where it caused serious damage to young plantations of *Eucalyptus* spp., especially on *E. grandis* grown from seed sourced from South Africa (Ferreira 1981).

![Symptoms of guava rust](image)

Figure 5.1. Symptoms of guava rust (a) blight pustules on young tissues, (b) old lesions on leaves, (c,d) pustules on twigs, leaves and petioles. The arrows show dying shoot tips and infected branches (picture source: Pérez et al. 2011).
Following the first report from Brazil, guava rust appeared in South Florida in 1977 (Marlatt and Kimbrough 1979). After the discovery of guava rust in 2005 in Hawaii, the pathogen spread quickly into other countries and continents (Ramsfield et al. 2010). A recent molecular study on genotypes of \textit{P. psidii} populations in Hawaii revealed a difference in rust populations from South America (Graça et al. 2011), indicating it is unlikely that rust had come from South America. Further, since the genotype found in Hawaii was not found in South America it indicated that these populations in South America are unlikely to have been the source of the introduction to Hawaii. The wide host range of this species, its fast dispersal in Hawaii and the wide distribution of guava rust in the Southern Hemisphere has raised many concerns in Australia and New Zealand (Glen et al. 2007). Guava rust was first reported in Australia in April 2010, although based on the morphology of the urediospores and the lack of teliospores, the causal agent was primarily identified as \textit{Uredo rangelii}, which is a taxon within \textit{Puccinia psidii}. A few months later after additional laboratory studies, the causal agent was named \textit{P. psidii} (Carnegie and Cooper 2011). The first detection of guava rust in Australia was reported on three genera: \textit{Agonis flexuosa} (Western Australian peppermint), \textit{Syncarpia glomulifera} (turpentine) and \textit{Callistemon viminalis} (weeping bottlebrush) (Carnegie et al. 2010). Further experiments confirmed the ability of the pathogen to complete its life cycle on forestry Myrtaceae species such as \textit{Eucalyptus agglomerata} and \textit{E. cloeziana} (Carnegie et al. 2011). Prior to its first detection in Australia in 2010, most risk assessment studies emphasised the high risk of this pathogen to biodiversity in target regions and in particular, the Australian eucalyptus industry. Consequently, the necessity of developing strategies to tackle the possible incursion of guava rust in Australia was highlighted (Haines 2006). In New Zealand, this pathogen has potential to negatively affect the economy through the eucalyptus industry (timber, log and pulp), the feijoa industry, the manuka honey industry and nurseries that grow species of the Myrtaceae family (MAF 2011). Besides the economic values, the ecological, cultural and spiritual significance of these species is important for New Zealand. For example, the red flowers of the pohutukawa tree (\textit{Metrosideros excels}) are iconic features and important symbol for New Zealand people at home and abroad.

Guava rust has a macro-cyclic life-cycle (fungi that produce all five spores sometimes excluding pycniospores). Similar to other rust fungi, this species has several life stages and produces different types of infectious spores, including aeciospores, urediniospores,
basidiospores and teliospores. Among these, urediniospores are the most common spore that are observed, basidiospores are rarely seen. Aeciospores have only been produced in the laboratory (with none having been seen in nature) and teliospores act as the resting or overwintering spores (Glen et al. 2007). The pathogen causes infection on stems, leaves and fruit of its host especially on young growing leaves (younger than two years old) and sepals (Ramsfield et al. 2010) (Figure 5.1). In some host plants, lesions are purple in colour with spore pustules that are bright yellow to orange-yellow (Figure 5.1). Infrequently, the spores may appear as dark brown. The infection also may also cause twisting or bending of the leaves (Grgurinovic et al. 2006; Coutinho et al. 1998).

This pathogen is unusual in two ways. Firstly, unlike most rust fungi which usually have a narrow range of host plants, this pathogen has been found to have a very wide-range of hosts. Secondly, the pathogen is capable of causing severe damage especially on young plants. Currently, apart from establishing a quarantine system in infected areas, the control of guava rust is restricted to the use of chemical compounds such as Triadimenol and Chlorothalonil-based fungicides which can cause a reduction in the incidence of the rust by up to 85% (Martins et al. 2011; Ferrari et al. 1996).

5.1.2 Global distribution

Similar to other rusts, the most common way which guava rust disperses and reaches remote locations is by wind (Coutinho et al. 1998). There are also opportunities for dispersal by honey bees and other insects, infected plant material, contamination of clothes, and many human factors (Carnegie et al. 2010). Since the first report of guava rust in Brazil this pathogen has been able to move to other continents, such that there are reports from Argentina, Paraguay, Brazil, Uruguay, Venezuela, Ecuador and Colombia in South America, from Florida and Hawaii in North America, Cuba, Dominican Republic, Guatemala, El Salvador, Jamaica, Puerto Rico, Trinidad and Tobago in the Caribbean and Central America, from Japan, China, India and Taiwan in Asia (Figure 5.2). Thereafter, there were some unconfirmed reports from Indonesia and reports suspecting its presence in Taiwan and South Africa (Hardiyanto and Tridasa 2000), however, the presence of the rust in South Africa was officially confirmed in 2013 (Roux et al. 2013). Detection of the pathogen in 2010 in New South Wales of Australia (Carnegie et al. 2010) and recent report from New Caledonia in May 2013 (https://www.ippc.int/countries/pest-reports/puccinia-psidii-new-
caledonia) resulted in many studies being conducted to evaluate the likelihood of its long-term effects on the eucalyptus industry (Elith et al. 2013; Kriticos et al. 2013; Booth and Jovanovic 2012; Morin et al. 2012).

5.1.3 Host range

Guava rust can affect many plants in the myrtle family (Myrtaceae) (Coutinho et al. 1998). The Myrtaceae family is monophyletic, in other words is composed of a collection of species including the most-recent common ancestor of all those species and all the descendants of that most-recent common ancestral species. This family includes two subfamilies, 17 tribes, about 130 genera and nearly 4600 species. These species are mainly scattered in South-east Asia, Central and South America, Australasia and some in Southern Africa (Russell Haines 2006).

Some plants in this family are economically important including: *Eucalyptus* spp., *Psidium guajava* (common guava) and *Syzygium aromaticum* (clove tree). Most of the genera in Myrtaceae family are present in Australasia and South Asia and four are found in Central and South America. Among these, *Eucalyptus* spp. is very important economically as in total around eight million hectares in tropical and subtropical regions are allocated to forest plantation of *Eucalyptus* spp., and highlights the potential impact of guava rust (Coutinho et al. 1998) where any disease spreading throughout large plantations would cause a major loss to the forestry industry (Coutinho et al. 1998).

Figure 5.2. Current global distribution of guava rust.
A list of known host species of guava rust can be found in Simpson et al. (2006). Recently it has been suggested that this rust can infect more than 396 species from 65 genera (pers. comm. with Suvi L.H. Viljanen-Rollinson) (Carnegie et al. 2011; Russell Haines 2006) and this wide host range has forced many scientist and organizations to carry out concentrated research on various aspects of the biology and epidemiology of the rust around the world (Coutinho et al. 1998). For example an international project funded by the Australian government (ACIAR) to evaluate the risks to Myrtaceae posed by *P. psidii* is currently underway. There are eight species of rust that have been reported to cause infection on *Myrtaceae*, including *Phakopsora* (one species), *Physopella* (two species), *Puccinia* (two species) and *Uredo* (three species) (Russell Haines 2006). Before 2010, a genus of *Puccinia* (*Puccinia cygnorum*) had been reported from Australia (Russell Haines 2006), which has also been detected on *Astartea fascicularis* in New Zealand in the Napier Botanic Gardens in March 2006 (Glen et al. 2007). All infected and non-infected trees in the garden were removed and the rust has not been detected since, and the pathogen was considered eradicated in New Zealand in 2009.

### 5.1.4 Epidemiology

Dianese et al. (1984) found that in central Brazil there was a negative correlation between rainfall and rust on *E. grandis* (rose gum), *E. citriodora* (lemon eucalyptus) and *E. cloeziana* (Gympie messmate). The authors also concluded that there was no significant correlation between rust infection, relative humidity and temperature, but the disease established better in dry seasons. These findings are in disagreement with by another study conducted by Tessmannn et al. (2001) where they conclude that a combination of leaf wetness, temperature (above 18°C to 25°C) and relative humidity (90% or more), is the main driving force of epidemics on rose apple in central Brazil (Brasilia). In their study Tessmann et al. (2001) attributed epidemics of this species to the duration of wet conditions in the absence of light. Additionally, the authors reported that nocturnal temperature (ranging from 18°C - 25°C) over the same period of wetness is another important factor of epidemics. Tessmann et al. (2001) concluded that in the earlier studies by Dianese et al. (1984) the micro-meteorological condition of the leaf surface was not investigated and suggested that the weather data used for the analysis were sourced from a weather station distant from the experimental location.
In other studies, the optimal germination of urediospores has been reported in the presence of water, absence of light and a moderate temperature (15°C - 20°C) (De Carvalho et al. 1998; Piza and Ribeiro 1988; Castro et al. 1984). In a study by Haines and Simpson (2006), the optimum temperature for germination of urediniospores was reported as 16.1°C with the maximum temperature of 21.1°C to 25°C. In another study, the number of days with relative humidity of 90% along with temperature 18-25°C were highly correlated to spore germination (Glen et al. 2007; cited in Ruiz et al. 1989). In addition, it has been documented that for successful infection, the urediospores require an essential dormancy period. Teliospores will only infect juvenile tissue after exposure to eight-hours of darkness plus six hours surface moisture (Russell Haines 2006). Other studies have also reported that spore germination and infection of Pimenta species occurs between 13°C to 20°C (optimum at 16°C) (MacLachlan 1936 cited in Glen et al. 2007). Smith(1935 cited in Carnegie et al. 2011) concluded that high ambient humidity is needed for spore germination (urediospore) and Rayachhetry et al. (1997) showed that after inoculation of Melaleuca leaves spores are formed after 10 days, with spore release seen after 12 days. In summary, temperature has been mentioned as the important variable with the optimum temperature range suggested from 13°C to 20°C and the maximum temperature for spore germination is reported to be around 25°C while contradicting opinions exist regarding the effect of rain on guava rust development (Tessmann et al. 2001; Dianese et al. 1984).

5.1.5 Risk to New Zealand

Because the guava rust has spread throughout the Hawaiian Island chain, a distance of 600 km in less than six months (Loope 2010) and also because this rust has been observed on seedlings of ohia (Metrosideros polymorpha), a species similar to New Zealand pohutukawa (Metrosideros excels) and rata trees (Metrosideros robusta), this pathogen is considered as a high risk pathogen for New Zealand. In addition, the presence of rust in Australia and the fact that currently this rust has been isolated from 107 host plants in 30 genera increases the chance of pathogen arrival in New Zealand. Since the first report of its establishment in Australia in May 2010 and in New Caledonia in May 2013 (https://www.ippc.int/countries/pest-reports/puccinia-psidii-new-caledonia), many concerns have been raised regarding its possible damage to eucalyptus industry and other potential impacts on a wide range of Myrtaceae family that are important species in Australian
ecosystems (Carnegie et al. 2011; Carnegie et al. 2010). Based on the history of this pathogen’s
movement and its recent establishment in Australia, guava rust is considered a risk to New
Zealand biodiversity and indigenous plants belonging to Myrtaceae family (MAF 2011)
(Table 5.1). It has been suggested that four industries in New Zealand may be economically
affected by guava rust establishment: nurseries growing Myrtaceae family species, manuka
honey producers, feijoa growers and eucalyptus growers. Among these, commercial growth
of the eucalyptus trees in New Zealand is only done on a small scale compared to other tree
species, and it only constitutes 1.4% of total plantation of the country (MAF 2011). The
closure of nurseries because of infection with guava rust can also cause financial loss. Feijoa
(\textit{Acca sellowiana}) a host plant of guava rust is grown in New Zealand, and the fruits are
exported to the USA, Japan and European countries such as UK, Germany, Netherlands and
France. Although the value of feijoa market is low (NZ$1.9 million based on 2010 data)
(www.epa.govt.nz), guava rust incursion can additionally impose a cost because of control
and sanitation efforts needed. The value of manuka (\textit{Leptospermum scoparium}) honey exports
was valued at NZ$145 million in 2013 (www.stat.govt.nz). The infection of manuka trees
can affect pollen production, and consequently, the honey production will be decreased.
At an environmental level, it has been suggested that infection of the Myrtaceae family
by guava rust can potentially alter the composition of New Zealand indigenous systems.
The indigenous species of Myrtaceae in New Zealand are outlined in Table 5.1.
Table 5.1. New Zealand indigenous species of Myrtaceae (Poole & Adams, 1994).

<table>
<thead>
<tr>
<th>LEPTOSPERMOIDEAE</th>
<th>MYRTOIDEAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leptospermum scoparium</td>
<td>Lophomyrtus bullata</td>
</tr>
<tr>
<td>Kunzea ericoides</td>
<td>Lophomyrtus obcordata</td>
</tr>
<tr>
<td>K. sinclairii</td>
<td>Syzygium maire</td>
</tr>
<tr>
<td>M. albiflora</td>
<td></td>
</tr>
<tr>
<td>M. bartlettii</td>
<td></td>
</tr>
<tr>
<td>M. carminea</td>
<td></td>
</tr>
<tr>
<td>M. colensoi</td>
<td></td>
</tr>
<tr>
<td>M. diffusa</td>
<td></td>
</tr>
<tr>
<td>Metrosideros excelsa</td>
<td></td>
</tr>
<tr>
<td>M. fulgens</td>
<td></td>
</tr>
<tr>
<td>M. kermadecensis</td>
<td></td>
</tr>
<tr>
<td>M. parkinsonii</td>
<td></td>
</tr>
<tr>
<td>M. pedunculata</td>
<td></td>
</tr>
<tr>
<td>M. perforate</td>
<td></td>
</tr>
<tr>
<td>M. robusta</td>
<td></td>
</tr>
<tr>
<td>M. umbellata</td>
<td></td>
</tr>
</tbody>
</table>

5.1.6 Previous modelling attempts on guava rust

The first attempt to model the potential threat of guava rust to a new region was made by Booth et al. (2000). In their study, they conducted a preliminary assessment to identify high risk areas in Australia and the Neotropics ecozone that includes the tropical terrestrial ecoregions of both Americas and the entire South American temperate zone. They concluded that eastern coastal areas of northern NSW and Queensland were the regions with the highest risk in Australia, and these regions have been affected by guava rust since 2010 (Elith et al. 2013; Carnegie and Cooper 2011). The second study which tried to estimate the global distribution of guava rust was performed in 2007 using the NAPPFAST model by Magarey et al. (2007). Environmental factors which have been shown to be the most important for
establishment of guava rust around the world were used to generate the NAPPFAST model. Monthly data including the average daily maximum temperature (equal or less than 33°C) and the average daily minimum temperature (higher than 13°C) were included in the NAPPFAST model. Also, Magarey et al. (2007) considered between 5 to 25 wet days per month for establishment of this rust. An assumption was made that if three or more months in a year met climatic conditions, pathogen establishment would be successful. They used a 10 year weather database (1993-2002) of temperature and moisture variables (wet days per month) to run the model with the results showing that the number of years with suitable climatic conditions for infection by the pathogen is more frequent in the Caribbean and South America than other parts of the world and this finding matched the real world disease distribution pattern. The NAPPFAST model successfully projected northern and northeastern coasts of Australia as highly suitable areas for guava rust establishment but low risk of establishment was predicted for New Zealand. However, a study by Kriticos et al. (2008) to evaluate the current and future potential risk of guava rust in New Zealand, predicted high risk of guava rust establishment in the North Island and small areas in the northern South Island (Kriticos et al. 2008). A study by Elith et al. (2013) that investigated the effect of taxonomic uncertainty raised from different taxonomic assumptions, used MaxEnt to model the potential distribution of myrtle/guava rust in Australia. Although they (Elith et al. 2013) reported that among available records the difference between *U. rangelili* (a member of the guava rust complex) and *P. psidii* is not documented and it is impossible to explicitly distinguish these two pathogens, they created five datasets that reflected five different plausible interpretations of occurrence data. They concluded that taxonomy uncertainty can have a significant effect on spatial models. In summary, all previously developed models on guava risk for Australia agree on the suitability of eastern and south eastern coastal areas but the suitability of guava rust establishment in New Zealand is projected differently in available previous studies (Kriticos et al. 2008; Magarey et al. 2007). It the latest modelling study on guava rust, Kriticos et al. (2013) combined CLIMEX model with distribution of the guava rust host plants to assess the risk in Australia. The suitable areas in their model included subtropics and wet tropics areas which temperate temperature and moisture play main role in establishment of guava rust. Their model also showed the North Island of New Zealand as highly suitable for guava rust establishment.
5.2 Methods

5.1.1 Guava rust occurrence data

The current geographical distribution of guava rust was gathered from a range of different studies (Elith et al. 2013; Graça et al. 2011; Pérez et al. 2011; Klopfenstein et al. 2009; Uchida et al. 2006; Booth et al. 2000). Only three geo-referenced location points were available on GBIF online database and on CABI the data were mostly in country level. The points were gathered from literature and contacting the experts. In some instances Google Earth 5.1.35 and ArcGIS 10 were used to extract presence coordinates from published studies. After removing duplicate points a total of 144 points were used in the modelling process (Appendix E).

5.2.1 CLIMEX

The CLIMEX modelling process was repeated for guava rust using 144 presence locations. As with the other two-target species in this study, the presence data and localities available on the CABI website were not relied on because unreliable presence points were found in some cases (as explained for dwarf bunt in Chapter 4). Instead, the aim was to find the most accurate presence points on a local scale that were needed for model verification. Detailed information on guava rust occurrence was sourced from personal communication with experts, existing modelling studies and papers to be used in the CLIMEX parameter fitting process. Environmental needs of this species described earlier were used as a starting point to calibrate growth indexes (Loope 2010; Ramsfield et al. 2010; Kawanishi et al. 2009; Elith et al. 2013; MAFReport 2011; Tommerup et al. 2003). As the majority of presence points were located in South America and there was no information available regarding soil moisture requirements of the species, as a guide, the default soil moisture template (existing in the CLIMEX model) for this region was used to calibrate soil moisture parameter values. The temperature needs of the pathogen reported in a range of studies were initially used to set the temperature parameters (Tessmann et al. 2001; Coutinho et al. 1998; Piza et al. 1988). Only reports related to *Puccinia psidii* were included in the modelling process and the reports regarding *Uredo rangelii*, which has been named as the causal agent of myrtle rust in some studies, were not included in this study. The justification was that, although *Uredo rangelii* is a part of the guava rust complex, it has been shown that models developed using this
fungus project different risk areas which may have different implications for biosecurity efforts and it is recommended to be modelled separately (Elith et al. 2013).

5.1.2 Background selection for correlative models

Based on the methodology explained in Chapter 2 for guava rust, the correct background extent to be used in correlative models was first defined by Principle Component Analysis (PCA). This background extent was defined to allow Multi-Model Framework and MaxEnt model to select the pseudo-absence points (or background points in MaxEnt). The variable that contributed most to the first principal component was identified as temperature seasonality (cbio04), and the change in this variable was plotted against distance. The distance at which the contribution of the important variable started to show an obvious change, was chosen as the appropriate distance to constrain the background data (400 km) for selection of background or pseudo-absence points by correlative models. The changes in the contribution of the important variable for the first principal component versus buffer distances are shown in Figure 5.3.

![Figure 5.3](image.png)

Figure 5.3. The changes in the contribution of the important variable (cbio04) over distance. The graph shows that after 400 km, a distinct change happens in the important variable (temperature seasonality).

5.1.3 Correlative models

5.2.1.1 MaxEnt model

5.2.1.1.1 Variable selection

The candidate variables for MaxEnt were identified using a pairwise Pearson correlation which excluded highly correlated variables. The correlation test was performed using the background points located within the 400 km buffer around presence points ($n = 34803$) and
variables with pairwise correlation of higher than 0.70 were discarded (Appendix E). This process resulted in the selection of 10 variables with low correlation: annual mean temperature (cbio01), mean of monthly temperature (max temp-min temp) (cbio02), maximum temperature of the warmest month (cbio05), minimum temperature of the coldest month (cbio06), annual temperature range (cbio07), mean temperature of the wettest month (cbio08), mean temperature of the driest quarter (cbio09), precipitation of the wettest month (cbio13) precipitation of the driest month (cbio14) and precipitation seasonality (cbio15).

Because the number of presence locations comprised a large sample, linear, quadratic and hinge features were used to allow more complexity in the model. These settings were used based on their effect on model response curves and consequently model projection which was explained in Chapter 2 (Merow et al. 2013; Radosavljevic and Anderson 2013) (refer to Section 2.1.2.1). A Jackknife analysis was used to investigate the variable importance among the selected variables. Additionally, response curves were produced to show how each variable can affect the MaxEnt projection. A Multivariate Environmental Similarity Surfaces (MESS) analysis was performed to compare the environmental similarity of variables used in MaxEnt model (Elith et al. 2010). The MESS analysis highlights the areas where one or more variable are outside the range presented in the training data, therefore, shows those areas where prediction should be treated with caution (Elith et al. 2010) (Appendix E). Seventy-five percent of data were used to train/fit the model and 25% were set aside to validate the model (Fielding et al. 1997 cited in Huberty 1994).

5.1.3.1 Multi-Model Framework (MMF)
The procedure in the methodology chapter (Chapter 2) was used to run the Multi-Model framework. First the duplicate presence points were removed based on the resolution of the data which led to 144 presence points. Then, the same number (144) of pseudo-absence points were generated from a 400 km buffer around the presence points using OCSVM and k-mean clustering. Variable selection was performed by random forest and step-wise regression. Models were trained using the presence and generated pseudo-absence points and were validated by cross-validation and bootstrapping techniques.

5.1.3.1.1 Multi-Model Variable selection
In the Multi-Model Framework (MMF), the variables selected for guava rust both by random forest and step-wise regressions were: annual mean temperature (cbio01), temperature seasonality (cbio04), and mean temperature of the coldest month (cbio06), mean temperature
of the coldest quarter (cbio11), annual precipitation (cbio12) and precipitation of the driest month (cbio14).
5.3 Results

5.3.1 CLIMEX results

The parameter values used in the CLIMEX model for guava rust are described in Table 5.2.

Table 5.2. Parameter values used in developing CLIMEX model for Guava rust.

<table>
<thead>
<tr>
<th>Index</th>
<th>Parameters</th>
<th>Values</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV0</td>
<td>Lower temperature threshold</td>
<td>10</td>
<td>°C</td>
</tr>
<tr>
<td>DV1</td>
<td>lower optimum temperature</td>
<td>15</td>
<td>°C</td>
</tr>
<tr>
<td>DV2</td>
<td>upper optimum temperature</td>
<td>25</td>
<td>°C</td>
</tr>
<tr>
<td>DV3</td>
<td>Upper temperature threshold</td>
<td>34</td>
<td>°C</td>
</tr>
<tr>
<td>SM0</td>
<td>Lower soil moisture threshold</td>
<td>0.5</td>
<td>-</td>
</tr>
<tr>
<td>SM1</td>
<td>Lower optimum soil moisture</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>SM2</td>
<td>upper optimum soil moisture</td>
<td>1.5</td>
<td>-</td>
</tr>
<tr>
<td>SM3</td>
<td>Upper soil moisture threshold</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>TTCS</td>
<td>Cold stress temperature threshold</td>
<td>4</td>
<td>°C</td>
</tr>
<tr>
<td>THCS</td>
<td>Cold stress temperature rate</td>
<td>-0.005</td>
<td>Week⁻¹</td>
</tr>
<tr>
<td>DTCS</td>
<td>Cold stress degree-day threshold</td>
<td>15</td>
<td>°C</td>
</tr>
<tr>
<td>TTHS</td>
<td>Heat Stress Temperature threshold</td>
<td>35</td>
<td>°C</td>
</tr>
<tr>
<td>THHS</td>
<td>Heat Stress Temperature rate</td>
<td>0.002</td>
<td>Week⁻¹</td>
</tr>
<tr>
<td>SMDS</td>
<td>Dry Stress Threshold</td>
<td>0.35</td>
<td>Week⁻¹</td>
</tr>
<tr>
<td>HDS</td>
<td>Dry Stress Rate</td>
<td>-0.017</td>
<td>Week⁻¹</td>
</tr>
<tr>
<td>TTHD</td>
<td>Hot-dry temperature threshold (0-50)</td>
<td>35</td>
<td>-</td>
</tr>
</tbody>
</table>

Calibration of the CLIMEX model resulted in final parameters (Table 5.2), where the upper temperature threshold (DV3) is much higher (34°C) than what has been reported in the literature (25°C reported by (Tessmann et al. 2001). The upper temperature threshold (DV3) was adjusted to allow for the persistence of the pathogen in South American localities such as Quebracho and Tres Bocas in Uruguay, Paraguay and Espago Doeste in Brazil. Although the optimum range for germination of the basidiospores is 13°C-20°C (De Carvalho et al. 1998; Ruiz et al. 1989; Castro et al. 1984; MacLachlan 1936), the model accepted optimal germination up to 25°C. The cold stress temperature threshold (TTCS) was set at 4°C, thus indicating the lower temperature threshold (DV0) which starts above 10°C, does not constrain guava rust development. Soil moisture parameters were set by considering the distribution of the pathogen in Maui, Hawaii and Santa Catarina and Espirito Santo, Brazil.
Application of the dry-stress parameters resulted in the exclusion of the highlands around Petrolina, Brazil, northern Paraguay and southern Bolivia.

The CLIMEX output is a very good fit to the current distribution data. The model was validated with reports from the Caribbean, Colombia, Ecuador, Venezuela and the USA. Only two points from Argentina and one from California lie in localities projected as marginal. In Australia the model accurately predicted suitable areas in current localities and predicted areas around Sydney, Canberra and Melbourne to be highly suitable (Figure 5.4). In addition, small areas in Western Australia and China are projected as highly suitable from where there are no current reports of the pathogen.
Figure 5.4. Global potential climate suitability of guava rust projected by CLIMEX. The figure shows the North Island in New Zealand as highly suitable for guava rust establishment.
The projection for New Zealand indicated that the entire North Island (except the central highlands) were very suitable for guava rust establishment. In addition, small areas in the South Island around Nelson, Blenheim and Christchurch are projected as highly suitable (Figure 5.4). For example, Figure 5.5 indicates that for the Auckland area, the weekly growth index is optimal throughout the year with a decrease in winter indicating that the spore-germination conditions are not optimal (because of a decrease in the temperature).

For Christchurch, South Island, Figure 5.6 shows that although spore germination can happen from September to May with a decline in February, the most favourable time for optimal germination is December. Although temperatures are optimal in the early months of the year, soil moisture content is lower than SM0 which implies spore germination is impaired.
Figure 5.5. Weekly growth index and climate variables around Auckland for guava rust.
Figure 5.6. Weekly growth index and climate variables for the Christchurch area for guava rust.
5.3.2 MaxEnt model results

The candidate variables were selected by performing Pearson pairwise correlation test. Variables with a correlation value lower than 0.70 were selected to include in the MaxEnt model. The final MaxEnt model gave an AUC of 0.74 which indicated the model is better than a random model (Appendix E). The jackknife chart of regularized data identified the most important variables as mean diurnal range (mean of the monthly maximum temperature - mean of monthly minimum temperature) (cbio02), precipitation of the driest month (cbio14), maximum temperature of the warmest month (cbio05) and minimum temperature of the coldest month (cbio06).

Two variable response curves produced by the MaxEnt model are shown in Figure 5.7 and Appendix D. By increasing the precipitation of the driest quarter (cbio14), the probability of guava rust presence increases then starts decreasing at around 60 mm (Figure 5.7A). On the other hand, the response curve of maximum temperature of the warmest month (cbio05) suggests that the probability of guava presence decreases of around 30°C (Figure 5.7B; note temperature values are multiplied by 10). Also, temperature variable response curves (Appendix E) show that the probability of guava rust presence maximises when the annual mean temperature reaches around 24°C which is in accordance with the upper optimum temperature threshold (DV2) suggested by the CLIMEX model. Studying other response curves, especially precipitation variables (Appendix E), shows that when precipitation of the wettest month (cbio13) reaches around 180 mm, further increases results in the probability of guava rust presence decreasing.

The lowest gain in training data was achieved when the mean temperature of wettest month (cbio08) was used (Figure 5.8A) which indicates the lowest contribution to the modelling process. The model also trained very well by using the mean of monthly temperature (max temp-min temp) (cbio2) but the gain dramatically decreased in the test data (Figure 5.8B) which indicated that it was not a good variable for generalisation purposes. Since most of the important variables suggested by MaxEnt are temperature-based, the jackknife highlights the importance of temperature for guava rust establishment. On the other hand, the jackknife of the test data showed that although maximum temperature of the warmest month (cbio15) gave a good gain, the AUC of the model improves in the absence of this variable (light blue bars) (Figure 5.8B). The jackknife of the AUC also confirms that most of candidate variables
contribute to a better AUC score for the model (Figure 5.8C). Variables cbio5 and cbio15 gain good gain in training and test data which means they give good generalisation or transferability to new data (Figure 5.8 A, B).

The potential distribution of guava rust shown in Figure 5.9 follows the general pattern of current distribution of the pathogen in coastal areas in eastern Australia, South Africa, southern and central China, most of Brazil except highland areas, Uruguay and Bolivia. The probability of guava rust occurrence was been predicted as between 30-80% in different parts of the Caribbean (Figure 5.9). In Asia, the most suitable area was southern China. Italy, France, Spain and Portugal have the highest probabilities in Europe. Southern Florida and small areas in southern California were also projected with wide range probability of 0.3 to 0.8 (sites closer to the coast had higher probability).

The projection for New Zealand was similar to CLIMEX where the North Island had a higher probability of occurrence compared to the South Island. Auckland and northern Kaitaia have the highest suitability in New Zealand, and this also coincided with the CLIMEX results.
Figure 5.7. (A) The response curves of precipitation of the driest month (cbio14), and (B) max temperature of the warmest month (cbio05) produced by MaxEnt.
Figure 5.8. MaxEnt Jackknife charts for guava rust. A) Training, B) test and C) AUC. Variables: annual mean temperature (cbio1), mean of monthly temperature (max temp-min temp) (cbio02), maximum temperature of the warmest month (cbio05), min temperature of the coldest month (cbio06), annual temperature range (cbio07), mean temperature of the wettest month (cbio08), mean temperature of the driest quarter (cbio09), precipitation of the driest month (cbio14), precipitation of wettest month (cbio13) and precipitation seasonality (cbio15).
Figure 5.9. Potential climate suitability of guava rust projected by MaxEnt.
5.3.3 Multi-Model

In the Multi-Model, based on 10 performance criteria achieved by cross-validation and bootstrapping, the SVM (support vector machine) model showed the best score with an AUC of 0.95, and was selected as the best model for projection (Table 5.3).

The most suitable areas for guava rust establishment are: eastern coastal areas in Australia, Southern China, Vietnam, Myanmar, Thailand, Indonesia, Malaysia, Sri Lanka, Singapore and a narrow strip in western India (Figure 5.10). The projection in this part of the world followed a pattern similar to the CLIMEX and MaxEnt models, but the probability projected by MaxEnt was generally low for this area. In Europe, there were no suitable areas projected by SVM which matched the areas projected by the CLIMEX and MaxEnt models. Central Africa and eastern coastal parts of South Africa were projected as highly suitable by both Multi-Model and CLIMEX while MaxEnt projected lower suitability for this area. In South America, the projection of Multi-Model and CLIMEX is in accordance with the current distribution of guava rust while MaxEnt projection is generally low with the exception of coastal eastern areas in Brazil and Uruguay.

In New Zealand, the Multi-Model projects the suitability of northern Northland, and the Auckland area. There were small differences in the predictions for the central parts of the North Island where CLIMEX and MaxEnt projected it suitable but the Multi-Model prediction is lower in southern parts of the South Island. In addition, Northern parts of the South Island such as Nelson, Blenheim and Kaikoura were projected as suitable by MaxEnt and CLIMEX but the probability projected by Multi-Model was only 0.3.
Table 5.3. The accuracy measures achieved by cross-validation (A) and bootstrapping (B) in Multi-Model Framework developed for dwarf bunt. Acronyms for models are: QDA= quadratic discriminant analysis, NB= naïve Bayes, LDA= linear discriminant analysis, LOG= logistic regression, CART= classification and regression tree, CTREE= conditional tree, KNN= K-nearest neighbor, SVM= support vector machine, and NNET= neural network.

### A) Bootstrapping

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision rank1</th>
<th>Recall rank2</th>
<th>F.score rank3</th>
<th>Kappa rank4</th>
<th>Specificity rank5</th>
<th>TSS rank6</th>
<th>Uncertainty rank7</th>
<th>X.632 error rank8</th>
<th>auc rank9</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.888554</td>
<td>1 0.911477</td>
<td>1 0.860603</td>
<td>4 0.883794</td>
<td>1 0.776697</td>
<td>1 0.917199</td>
<td>1 0.777802</td>
<td>1 0.086806</td>
<td>1 0.087896</td>
<td>1 0.952064</td>
<td>2 14</td>
</tr>
<tr>
<td>NNET</td>
<td>0.863665</td>
<td>2 0.856423</td>
<td>2 0.871476</td>
<td>2 0.862211</td>
<td>2 0.726714</td>
<td>2 0.857486</td>
<td>2 0.728962</td>
<td>2 0.184028</td>
<td>2 0.94203</td>
<td>2 0.952353</td>
<td>1 22</td>
</tr>
<tr>
<td>KNN</td>
<td>0.846897</td>
<td>3 0.824335</td>
<td>4 0.878921</td>
<td>1 0.849323</td>
<td>3 0.693201</td>
<td>3 0.816607</td>
<td>4 0.695527</td>
<td>3 0.159722</td>
<td>4 0.127002</td>
<td>3 0.935957</td>
<td>4 32</td>
</tr>
<tr>
<td>CART</td>
<td>0.830766</td>
<td>4 0.833347</td>
<td>3 0.826532</td>
<td>6 0.826575</td>
<td>4 0.660659</td>
<td>4 0.835904</td>
<td>3 0.662435</td>
<td>4 0.246528</td>
<td>7 0.135617</td>
<td>4 0.942274</td>
<td>3 42</td>
</tr>
<tr>
<td>QDA</td>
<td>0.807753</td>
<td>5 0.777735</td>
<td>5 0.859851</td>
<td>5 0.816685</td>
<td>5 0.615236</td>
<td>5 0.750892</td>
<td>5 0.617943</td>
<td>5 0.149306</td>
<td>3 0.183524</td>
<td>5 0.910566</td>
<td>7 50</td>
</tr>
<tr>
<td>NB</td>
<td>0.798121</td>
<td>6 0.75909</td>
<td>7 0.870203</td>
<td>3 0.808887</td>
<td>6 0.596338</td>
<td>6 0.729107</td>
<td>7 0.59931</td>
<td>6 0.121528</td>
<td>2 0.196019</td>
<td>6 0.911458</td>
<td>6 55</td>
</tr>
<tr>
<td>CTREE</td>
<td>0.746117</td>
<td>7 0.770606</td>
<td>6 0.753794</td>
<td>7 0.783513</td>
<td>7 0.497733</td>
<td>7 0.475177</td>
<td>6 0.498971</td>
<td>7 0.340278</td>
<td>9 0.232909</td>
<td>7 0.912471</td>
<td>5 68</td>
</tr>
<tr>
<td>LOG</td>
<td>0.661182</td>
<td>8 0.650631</td>
<td>8 0.6877</td>
<td>9 0.664437</td>
<td>9 0.324177</td>
<td>8 0.639072</td>
<td>8 0.326733</td>
<td>8 0.25</td>
<td>8 0.328061</td>
<td>8 0.709394</td>
<td>9 83</td>
</tr>
<tr>
<td>LDA</td>
<td>0.660011</td>
<td>9 0.648176</td>
<td>9 0.691668</td>
<td>8 0.664958</td>
<td>8 0.322023</td>
<td>9 0.632995</td>
<td>9 0.324663</td>
<td>9 0.236111</td>
<td>6 0.32852</td>
<td>9 0.709973</td>
<td>8 84</td>
</tr>
</tbody>
</table>

### B) Cross-Validation

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision rank1</th>
<th>Recall rank2</th>
<th>F.score rank3</th>
<th>Kappa rank4</th>
<th>Specificity rank5</th>
<th>TSS rank6</th>
<th>Uncertainty rank7</th>
<th>X.632 error rank8</th>
<th>auc rank9</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNET</td>
<td>0.905556</td>
<td>1 0.935088</td>
<td>1 0.88462</td>
<td>3 0.90824</td>
<td>1 0.810834</td>
<td>1 0.928594</td>
<td>1 0.813215</td>
<td>1 0.15625</td>
<td>8 0.123212</td>
<td>1 0.941599</td>
<td>1 21</td>
</tr>
<tr>
<td>KNN</td>
<td>0.883333</td>
<td>2 0.88462</td>
<td>2 0.894561</td>
<td>4 0.888979</td>
<td>2 0.766046</td>
<td>2 0.872386</td>
<td>2 0.766947</td>
<td>2 0.03125</td>
<td>1 0.132668</td>
<td>3 0.933618</td>
<td>2 22</td>
</tr>
<tr>
<td>SVM</td>
<td>0.861111</td>
<td>3.5 0.849273</td>
<td>4 0.895146</td>
<td>3 0.870675</td>
<td>3 0.721045</td>
<td>3 0.825899</td>
<td>4 0.721045</td>
<td>4 0.055565</td>
<td>4.5 0.130294</td>
<td>2 0.926553</td>
<td>3 33</td>
</tr>
<tr>
<td>CART</td>
<td>0.861111</td>
<td>3.5 0.8636</td>
<td>3 0.870702</td>
<td>6 0.80598</td>
<td>4 0.720882</td>
<td>4 0.847467</td>
<td>3 0.718169</td>
<td>4 0.215278</td>
<td>9 0.168806</td>
<td>4 0.908854</td>
<td>4 45</td>
</tr>
<tr>
<td>NB</td>
<td>0.838889</td>
<td>5 0.797762</td>
<td>6 0.936199</td>
<td>5 0.859791</td>
<td>5 0.672178</td>
<td>5 0.729575</td>
<td>6 0.655774</td>
<td>5 0.052083</td>
<td>3 0.205858</td>
<td>6 0.907311</td>
<td>5 47</td>
</tr>
<tr>
<td>QDA</td>
<td>0.788889</td>
<td>6 0.750926</td>
<td>7 0.895673</td>
<td>2 0.815633</td>
<td>6 0.573022</td>
<td>6 0.674022</td>
<td>7 0.569692</td>
<td>6 0.045139</td>
<td>2 0.20857</td>
<td>7 0.904938</td>
<td>6 55</td>
</tr>
<tr>
<td>CTREE</td>
<td>0.733333</td>
<td>7 0.805996</td>
<td>5 0.641287</td>
<td>9 0.801953</td>
<td>7 0.464727</td>
<td>7 0.832942</td>
<td>5 0.465471</td>
<td>7 0.06972</td>
<td>9 0.19791</td>
<td>5 0.810618</td>
<td>7 66</td>
</tr>
<tr>
<td>LDA</td>
<td>0.705556</td>
<td>9 0.710842</td>
<td>8 0.737076</td>
<td>8 0.721181</td>
<td>9 0.409373</td>
<td>9 0.672712</td>
<td>8 0.409788</td>
<td>9 0.055566</td>
<td>4.5 0.343848</td>
<td>9 0.709289</td>
<td>8 81.5</td>
</tr>
<tr>
<td>LOG</td>
<td>0.711111</td>
<td>8 0.708453</td>
<td>9 0.757076</td>
<td>7 0.730589</td>
<td>8 0.418928</td>
<td>8 0.660948</td>
<td>9 0.418024</td>
<td>8 0.0625</td>
<td>6 0.341624</td>
<td>8 0.707996</td>
<td>9 80</td>
</tr>
</tbody>
</table>
Figure 5.10. Potential climate suitability of guava rust projected by Support Vector Machine model (SVM).
5.4 Further analysis using Principal Component Analysis.

A principal component analysis was performed on guava rust presence points to determine their similarity in environmental space. The climate data of presence points were extracted in ArcGIS Desktop 10.1. To compare the similarity of presence points with New Zealand climate, two fields (presence and NZ) were added to the climatic data to differentiate the presence points from New Zealand data. This enables better visualization and comparison of points of interest. As Figure 5.11 shows, some of the guava rust presence points lie close to or within the New Zealand climate data space (19 bioclim variables cbio1-cbio19), which implies the climate similarity of guava rust presence locations to some New Zealand data.

Figure 5.11. A principal component analysis performed on guava rust presence points showing the similarity of NZ data and occurrence data in environmental space.
5.5 Discussion

While there is a high degree of agreement among models on the suitability of areas such as the eastern coastal parts of South Africa, Australia, Caribbean and eastern coastal areas in Brazil, MaxEnt fails to project Mexico and some other parts of South America such as Argentina, Bolivia and Venezuela as suitable areas for guava rust establishment when it is known that it has established in those countries. For New Zealand, there was a good level of agreement among all three models, particularly near Auckland in the North Island. The important role of temperature in guava rust establishment was highlighted by both CLIMEX and the MaxEnt model. In CLIMEX, none of the stress parameters used in calibration of the model (Table 5.2) had any effect on guava rust establishment in New Zealand, and additionally, the model suggested that temperature plays the most important role in pathogen growth. Stress parameters in CLIMEX are one of the model’s advantages because CLIMEX can show the areas which are affected by stress parameters. For example, because all three models agree on northern Mexico and southern California as unsuitable habitat, CLIMEX calibration suggests that one possible limiting factor that makes these areas unsuitable can be the dry stress.

The response curves produced by MaxEnt can also provide good information about the environmental factors which can affect the pathogen growth. For example, the response curve showed in Figure 5.7A implies that when precipitation exceeds a certain level (around 180 mm), the probability of guava presence decreases. The effect of rain on guava rust establishment is in agreement with the claim made by Dianese et al. (1984) about the negative effect of rain on guava rust establishment.

The first attempt to map the global potential distribution of guava rust was in a study by Magarey et al. (2007). They introduced NAPPFAST as an Internet system for plant pathogen mapping studies. Their model did not project New Zealand as a suitable area for guava rust establishment. The main reason for such projection for New Zealand compared to the current models may be the result of the temperature parameter values and data time-frame which was used in their model. In their model, they used average daily maximum and minimum temperatures of 33°C and 13°C. In addition, the resolution of the data used in their study is coarser (30 minute’s ≈ 55 km) than the current study (10 minutes ≈ 18.6 km) and the time-frame is also shorter (10 years) which potentially could differently affect smaller countries when the files are rasterized in GIS platforms.
The model projections of the current study is in accordance with a more recent study by Kriticos et al. (2008) that used three approaches, including CLIMEX Compare Locations, CLIMEX Match regions and NAPPFAST to project the suitability of New Zealand for guava rust establishment. Their CLIMEX model indicated that most of the North Island is suitable for guava rust establishment (except highlands around Mount Ruapehu). In the South Island, Blenheim and some coastal areas in Canterbury region were projected marginally suitable. On the other hand, the NAPPFAST model developed by Kriticos et al. (2008) showed that the optimal time for guava rust infection in New Zealand would be between 20-120 days. However, Kriticos et al. (2008) debate that the higher number of favourable days for guava rust infection is in Central Otago compared with the other parts of New Zealand does not seem realistic. Since the pathogen favours mild to warm temperatures, one might expect that warmer parts of New Zealand would be more suitable for guava rust. This unexpected behaviour of NAPPFAST has been addressed by Kriticos et al. (2008) as a probable scramble in meteorological data for some parts of New Zealand, but we believe that the reason lies in the unique environment for this part of Otago rather than NAPPFAST performance or dataset. It should be noted that in Central Otago climate is described as land of extremes and characterized by hot and dry long summer days (daylight until 10 pm), hot days and cool evening in autumn, freezing nights in winter (-6 to 15°C) and warm springs (temperature ranges from 3 to 20°C). Therefore, by considering such microclimates across New Zealand, the prediction of Central Otago in New Zealand as suitable area is acceptable.

Based on the results from all three approaches (CLIMEX, MaxEnt and MMF) and high level of agreement among model projection for New Zealand, there is significant risk of guava rust and establishment in New Zealand if it is introduced. This causes concern for biosecurity authorities, especially when two indigenous species, pohutukawa, rata and manuka, in the Myrtaceae family exist in the areas projected as suitable by all models (Figure 5.8). In addition, the widespread distribution of pohutukawa and rata trees in New Zealand shown in Figure 5.12 (www.nzflora.landcareresearch.co.nz) indicates a high level of guava rust host plant availability which adds to the existing concern. Although none of the current or previously-developed models projected the South Island and in particular the western parts of the South Island as suitable, there might be small areas in these locations where the meso-climatic conditions (areas with moderate temperatures in the range 20-30°C) may facilitate establishment (Kriticos et al. 2008).
The Principal Component Analysis on current occurrences of guava rust globally compares them with New Zealand points. The analysis used all 19 bioclim variables, which included temperature and precipitation data and showed that presence points from Australia and Uruguay lie within or very close to, New Zealand data which indicated high likelihood of suitability in New Zealand (Figure 5.11).

Concern regarding the risk of guava rust for New Zealand has increased as a result of its establishment in Australia, which may facilitate its entrance into New Zealand. Pests, plant pathogens and biological materials are frequently dispersed by wind from Australia to New Zealand by virtue of its proximity (Viljanen-Rollinson et al. 2006; McKenzie 1998; Close et al. 1978). McKenzie (1998) highlighted that approximately a half of Puccinia species existing in New Zealand were invasive and listed eight rust species (such as poplar rust and blackberry rust) which presumably entered New Zealand from Australia via wind currents. It has also been suggested that due to the fast spread of this pathogen in Hawaii, it should be regularly monitored by trap plants in the high-risk areas following a guava rust incursion incident (Kriticos et al. 2008). In addition, more common ways of dispersal such as human interception (i.e. clothing, plant material) and nursery plant material are among pathways should always monitored carefully.

Figure 5.12. Distribution of two indigenous Myrtaceae plants, Pohutukawa and rata trees, in New Zealand (www.nzflora.landcareresearch.co.nz).
(according to per. comm. to Carnegie, nursery plant materials was the way that guava rust entered Australia).

In summary, modelling attempts prior to the establishment of guava rust in Australia successfully projected the suitability of eastern coastal areas for its establishment which indicates the usefulness of such modelling studies (Magarey et al. 2007; Haines et al. 2006). The various model projections employed in this study suggest that the distribution of guava rust host plants across New Zealand and the current New Zealand climate make it highly susceptible to guava rust. Furthermore, because the rust is already established in some parts of Australia, there is a high possibility of rapid spread via wind currents to New Zealand.
Chapter 6
Characterising model uncertainty by sensitivity analysis: the effect of interacting parameter error

6.1 Introduction

Ecologists use different statistical and software models to assess the risk and predict the potential distribution of invasive species. Most risk assessments are based on valid assumptions but for every model, there is uncertainty about parameter values used in the model (Alexander 1989). In addition, the level of uncertainty increases, especially when considering the future values of parameters. While the sensitivity of parameters should always be tested as part of the modelling process, sensitivity analysis can also be helpful to increase the awareness of decision makers about prediction uncertainty and the different risk scenarios involved (Baird 1989). Sensitivity analysis investigates the uncertainty in parameter values resulting in a range of projected possibilities based on different model outputs. Those projected possibilities may feedback to model design by challenging primary assumptions (Baird 1989). Fiasco (1983) suggested that because of the number of parameters involved, it is necessary to design a methodology to perform a sensitivity analysis in any scientific discipline (Alexander 1989). While a sensitivity analysis can lead to better understanding of a problem, it can also contribute to an improvement of the models (Pannell 1997) such that, not only recommendations offered to decision-makers would be more trustworthy and reliable, but allow decision-makers to recognise the outcomes of different assumptions and strategies. Consequently, they are more likely to implement the best possible strategy.

Sensitivity analysis can prompt new ideas and result in novel hypotheses to be tested. Additionally, after testing a model for its accuracy, sensitivity analysis can reveal existing errors or poor data (Pannell 1997). The CLIMEX model has been used widely by many researchers to model the spatial distribution of plants, insects and weeds based on climatic suitability. Despite all the recommendations, true sensitivity analysis is rarely carried out on CLIMEX. The aim of this chapter was to investigate the impact of uncertainty associated with parameter values of the CLIMEX on its output. Except for a few studies (Taylor and Kumar 2012; Vanhanen et al. 2007; Venette et al. 2006), little attention has been given to the effect of parameter changes on model
prediction. While a recent study by Taylor et al. (2012) performed a sensitivity analysis on the CLIMEX model, it does not investigate the effect of the interaction of sensitive parameters on model performance. In this study, following the development of a CLIMEX model for dwarf bunt (*Tilletia controversa*) of wheat in Chapter 4, we determined the sensitivity of the CLIMEX model to changes in parameter values for dwarf bunt and show how the interaction of sensitive parameters affected model performance.

6.1 The CLIMEX model and parameter values uncertainty

Because of parameter uncertainty, modellers calibrate or fit CLIMEX parameters in such a way to find the best match for the current distribution of the species (Pattison et al. 2008). Few studies (Taylor et al. 2012; Vanhanen et al. 2007; Venette et al. 2006) have attempted to determine how the CLIMEX model behaves in response to some form of sensitivity analysis. Of those that have, the majority focus on changes in the main parameters of temperature and soil moisture. We found only three investigations that considered the change in all parameters involved in a full baseline model as well as stress parameters (Taylor et al. 2012; Vanhanen et al. 2007; Venette et al. 2006). Venette et al. (2006) performed a sensitivity analysis on the baseline model of *Phytophthora ramorum* to determine the effect of parameter uncertainty on model predictions. Their study adjusted the temperature parameters within ±3°C of the baseline model values and for soil moisture ±25% and ±50% of baseline model values were applied to evaluate the impact of changes in these parameters on population growth of this species. However, no stress parameters were included in their sensitivity analysis. Vanhanen et al. (2007) carried out a sensitivity analysis on all parameters included in the baseline model for gypsy moth, *Lymantria dispar*, and analysed the results based on the changes in the average EI values in Europe for climate change and range shift of gypsy moth and nun moth (*Lymantria monacha*). While their study showed the sensitivity of model predictions for Europe to certain temperature parameters, they did not show or visualize significant changes in areas of Europe to illustrate which areas are more sensitive to which parameter. The most comprehensive study of sensitivity analysis of CLIMEX parameters is that carried out by Taylor et al. (2012) on modelling the potential distribution of *Lantana camara*. Although these authors developed incremental models from the baseline model in a way that reflects the most reasonable range of variables for the target species in Australia, this study did not investigate the outcome of interactions between sensitive parameters. In addition, although we acknowledge that in performing sensitivity analysis, the way of changing parameter values is up to the modeller using
prior knowledge about target species biological data, these changes need to be within a biologically reasonable range. For example, extreme changes such as those much beyond about 50% from an observed value will usually result in a dramatic effect and may be an indication that the baseline model has not been well developed in the first place. Careful consideration should be given especially when dealing with lower and upper thresholds.

Subjective selection of parameter values in CLIMEX may result in different modellers coming up with different parameters. For example the parameter values used in Kim and Beresford’s (2009) study on dwarf bunt are exactly based on the results of laboratory studies of Hoffman (1982). On the other hand, while some studies such as that by (Gassner et al. 1954) may imply that the DV0 (lower temperature threshold) for dwarf bunt can be as low as -5°C other studies indicated -2°C as the lower temperature threshold (Hoffmann 1982). Clearly when there are different temperature thresholds suggested in different studies it is necessary to consider such discrepancies to reduce the uncertainties in model prediction.

6.2 Methods

6.2.1 Software and data

CLIMEX 3.0 (Sutherst et al. 2004) was used to determine the regional suitability of dwarf bunt (*Tilletia controversa*) establishment around the world. The data included in CLIMEX 3.0 is a database of meteorological stations worldwide with 30 arc minutes resolution but we used CliMond higher resolution data at 10’ resolution (Kriticos et al. 2012; Kriticos et al. 2006). As a starting point, the initial parameter values suggested in the published literature were used to fit CLIMEX parameters (Hoffmann 1982; Gassner et al. 1954). The same model output and parameter values which were developed for dwarf bunt in Chapter 4 (refer to Table 4.1), were used to perform the sensitivity analysis. The presence data for the baseline model comprised 78 locations extracted from CABI, GBIF and consulting and communicating with experts (refer to Figure 4.3 in Chapter 4). Also a summary of parameter values suggested in different studies is summarized in Table 6.1.
Table 6.1. A summary of the parameter values suggested in other studies and parameter values used in CLIMEX modeling in this study. The parameters characterize growth requirements and the stress response for dwarf bunt (*Tilletia controversa*).

<table>
<thead>
<tr>
<th>Index</th>
<th>Parameters</th>
<th>Current study</th>
<th>(Gassner et al. 1954)</th>
<th>(Conners 1954)</th>
<th>(Hoffmann 1982)</th>
<th>(Kim et al. 2009)</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DV0= lower temp threshold</td>
<td>-3</td>
<td>-5</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>DV1= lower optimum temperature</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>DV2= upper optimum temperature</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>DV3= upper temp threshold</td>
<td>13</td>
<td>20</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Soil moisture</td>
<td>SM0 = lower soil moisture threshold</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SM1= lower optimum soil moisture</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SM2= upper optimum soil moisture</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SM3= upper soil moisture threshold</td>
<td>1.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cold Stress</td>
<td>TTCS= Cold stress temperature threshold</td>
<td>-5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>THCS=Cold stress temperature rate</td>
<td>-0.00005</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DTCS=Cold stress degree-day threshold</td>
<td>15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Heat Stress</td>
<td>TTHS = Heat Stress temperature threshold</td>
<td>20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>THHS= Heat Stress temperature rate</td>
<td>0.00005</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Wet Stress</td>
<td>SMWS= Wet Stress threshold</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HWS= Wet stress accumulation Rate</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
6.2.2 Sensitivity Analysis

In sensitivity analysis, the possible ways in which parameters may vary is studied at different levels. A single variable may be changed within certain bounds while the remaining parameters are left at the baseline level or more than one parameter and their combinations may be changed. Panell (1997) discussed that if two parameters are positively correlated it is advisable to consider high values for both at the same time. On the other hand, if two parameters are negatively related it’s better that the modeller tests high values of one parameter along with low values of the second parameter. However, in general, the modeller makes the final decision based on background knowledge about the problem, the expected behaviour of parameters in the real world and any other possible factor that he/she thinks should be considered (Pannell 1997).

When selecting the parameter changes for sensitivity analysis, a common approach is to specify the changes over an comparable sized interval to explore the possible effect of the results on the baseline model (such as increasing and decreasing the parameter by a specific percentage). The selected range is up to the modeller which can include possible outcomes of the variables or parameter and should be within a reasonable range. In this study, the baseline model parameter values were changed by ±10% and ±25% and the impact on model output was recorded. Initially, just one parameter value was changed at a time, keeping the other parameter values in the baseline model constant. To evaluate the effect of single parameters on model outputs the model simulation was repeated for 15 parameter changes (Table 6.2). To evaluate overall sensitivity of the model and to simulate error associated with multiple parameters, all of them were also adjusted randomly within the range of ±25% of baseline model values. Furthermore, to evaluate the interaction of the most sensitive parameters (DV3 and SM3) both these parameters were changed ±25% concurrently within four scenarios. The scenarios comprised: 1) DV3 and SM3 were increased by 25%, 2) DV3 and SM3 were decreased 25%, 3) DV3 was increased by 25% and SM3 decreased by 25%and 4) DV3 was decreased by 25% and SM3 increased by 25%. This required 79 model runs. The adjusted values for temperature, soil moisture and stress parameters can be found in Table 6.2. Temperature and moisture parameters were subjected to the constraint of DV0<DV1<DV2<DV3 and SM0<SM1<SM2<SM3. After each simulation, EI values calculated by the model for all locations following Vera et al. (2002) and Hoddle (2004) were classified into four categories. In other words, unsuitable (EI=0), marginal (EI=1-10), suitable (EI=11-25) and very suitable (EI≥26). In addition, we considered another interpretation suggested by Sutherst and Maywald (2005) that locations with
EI>20 as very favourable. In this study, unlike other studies (Taylor et al. 2012; Venette et al. 2006) suitable and very suitable areas were not combined because it allows the changes in each category to be better detected. To better visualize the changes, CLIMEX results were exported into ARCGIS 10.1 for mapping and further processing. Similar to other studies regarding the mapping and evaluation of CLIMEX results, change in area was used to determine model sensitivity to each parameter.
Table 6.2. Parameter values used for sensitivity analysis of CLIMEX model on dwarf bunt. The parameters were set at five levels: ±10, ±25 and random values to investigate the effect of parameter uncertainty on model predictions.

<table>
<thead>
<tr>
<th>Index</th>
<th>Parameters</th>
<th>Baseline</th>
<th>-25%</th>
<th>-10%</th>
<th>+10%</th>
<th>+25%</th>
<th>Random value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temperature</strong></td>
<td>DV0= lower threshold</td>
<td>-3</td>
<td>-3.75</td>
<td>-3.2</td>
<td>-2.7</td>
<td>-2.25</td>
<td>-2.3</td>
<td>0C</td>
</tr>
<tr>
<td></td>
<td>DV1= lower optimum temperature</td>
<td>3</td>
<td>2.25</td>
<td>2.7</td>
<td>3.3</td>
<td>3.75</td>
<td>3.47</td>
<td>0C</td>
</tr>
<tr>
<td></td>
<td>DV2= upper optimum temperature</td>
<td>8</td>
<td>6</td>
<td>7.2</td>
<td>8.8</td>
<td>10</td>
<td>7</td>
<td>0C</td>
</tr>
<tr>
<td></td>
<td>DV3= upper threshold</td>
<td>13</td>
<td>9.75</td>
<td>11.7</td>
<td>14.3</td>
<td>16.25</td>
<td>14.22</td>
<td>0C</td>
</tr>
<tr>
<td><strong>Moisture</strong></td>
<td>SM0 = lower soil moisture threshold</td>
<td>0.1</td>
<td>0.075</td>
<td>0.09</td>
<td>0.11</td>
<td>0.125</td>
<td>0.087</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>SM1= lower optimum soil moisture</td>
<td>0.5</td>
<td>0.375</td>
<td>0.45</td>
<td>0.55</td>
<td>0.625</td>
<td>0.614</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>SM2= upper optimum soil moisture</td>
<td>1</td>
<td>0.75</td>
<td>0.9</td>
<td>1.1</td>
<td>1.25</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>SM3= upper soil moisture threshold</td>
<td>1.6</td>
<td>1.2</td>
<td>1.44</td>
<td>1.76</td>
<td>2</td>
<td>1.5</td>
<td>-</td>
</tr>
<tr>
<td><strong>Cold Stress</strong></td>
<td>TTCS=Cold stress temp threshold</td>
<td>-5</td>
<td>-6.25</td>
<td>-5.5</td>
<td>-4.5</td>
<td>-3.75</td>
<td>-4.97</td>
<td>0C</td>
</tr>
<tr>
<td></td>
<td>THCS=Cold stress temperature rate</td>
<td>-0.0005</td>
<td>-0.00046875</td>
<td>-0.0004125</td>
<td>-0.0003375</td>
<td>-0.0002812</td>
<td>-0.00048617</td>
<td>Week-1</td>
</tr>
<tr>
<td><strong>Heat stress</strong></td>
<td>DTCS=Cold stress degree-day threshold</td>
<td>15</td>
<td>11.25</td>
<td>13.5</td>
<td>16.5</td>
<td>18.75</td>
<td>17.40</td>
<td>0C</td>
</tr>
<tr>
<td></td>
<td>TTHS = Heat Stress temperature</td>
<td>20</td>
<td>15</td>
<td>18</td>
<td>22</td>
<td>25</td>
<td>16</td>
<td>0C</td>
</tr>
<tr>
<td><strong>Wet Stress</strong></td>
<td>THHS= Heat Stress temperature rate</td>
<td>0.00005</td>
<td>0.0000375</td>
<td>0.000045</td>
<td>0.000055</td>
<td>0.0000625</td>
<td>0.0000455</td>
<td>Week-1</td>
</tr>
<tr>
<td></td>
<td>SMWS = Wet stress threshold (1-10)</td>
<td>2</td>
<td>1.5</td>
<td>21.85</td>
<td>2.2</td>
<td>2.5</td>
<td>1.6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>HWS = Wet stress rate</td>
<td>0.001</td>
<td>-0.00075</td>
<td>0.0009</td>
<td>0.0011</td>
<td>0.00125</td>
<td>0.0008</td>
<td>Week-1</td>
</tr>
</tbody>
</table>
6.3 Results

6.3.1 Sensitivity analysis results

The baseline model output is shown in Figure 6.1. The baseline model simulation for potential distribution of dwarf bunt was highly sensitive to 25% decrease in the upper soil moisture value (SM3 -25%), in other words a decrease from 1.6 to 1.2 (no unit), and highly sensitive to a 25% increase and decrease in upper temperature threshold (DV3) value.

With a 25% decrease in SM3 value (1.6 to 1.2) the main change occurred in the suitable category where most parts of France, Spain, Italy and the USA turned into marginal areas for establishment (Figure 6.2). In addition, the very suitable areas in Europe were mainly limited to eastern Germany, western Poland, UK, Ukraine and Spain in that the EI value did not change significantly in these areas. Suitable locations in Australia became unsuitable or marginal and most of the suitable areas in New Zealand became unsuitable. Most areas of Japan that were suitable changed to marginal but no significant changes happened in the rest of Asia. The greatest increase in EI happened in Ethiopia (Dire Dawa from 0 to 70), India Jabalpur from 0 to 53) and Tunisia (Sousse 0 to 52). The most obvious decrease in EI value happened in Argentina (Islas Ano Nuevo from 64 to 0), Chile (Punta Arenas from 52 to 0) and Falkland Islands (Stanley 49 to 0) (Figure 6.2).

Changing DV3 25% from 13°C to 16.25°C, resulted in 5 million km² increase in global suitable areas (Figure 6.3). There was a slight increase in marginal area mainly in Australia, central USA and northern parts of New Zealand. An increase in suitable areas occurred in western USA, Algeria, central China and small parts in Australia. An increase in very suitable areas was clearly evident in Europe mainly in France, Spain, Ireland, Poland, Norway, Hungary and Germany. The most dramatic change happened in Peru (Huaraz) where the EI value increased from 5 to 25 from marginal to favourable or from unsuitable to very suitable depending on whose criteria one uses. Pisayambo in Ecuador changed from 20 to 36 and in Argentina, Ushuaia changes from 44 to 58. No decrease in EI value was observed in this scenario.

As Figure 6.4 shows, a 25% decrease in DV3 from 13°C to 9.75°C, resulted in a decrease in the suitable and very suitable areas with most of these areas becoming unsuitable. The most dramatic changes happen in Europe where France, England, Germany, Iceland and Spain change from very suitable to suitable. In eastern USA a high proportion of suitable areas turn to marginal. There is
no major change in Asia or South America. In Africa, suitable areas are confined to Morocco. In New Zealand, very suitable areas change to both marginal and suitable. In this scenario, an increase in the EI value is only observed in five locations such as USA (Birmingham from 0 to 24), Australia (Albany from 0 to 11), West Indies (Plymouth from 0 to 10), Argentina (Concordia from 0 to 5) and France (Brest from 12 to 16). The dramatic decrease happened in the UK (Birmingham 31 to 0) and Peru (Lampos Bajo 30 to 1).

Increasing SM3 from 1.6 to 2 mainly resulted in an increase in very suitable areas in Ireland and England. Changes in unsuitable and suitable categories were not as noticeable (Figure 6.5) but there were major changes in some locations. The largest change in EI happened in the USA (Birmingham 0 to 35), Chile (Los Evangelis 15 to 41) and West Indies (Plymouth 0 to 24). The dramatic decrease happened in the UK (Birmingham 31 to 0 and Plymouth 20 to 0), the USA (Albany 13 to 13) and New Zealand (Gore 11 to 0).

In total the model was run 75 times and the results indicated that the baseline model was not sensitive to -25% and +25% changes in cold stress (TTCS, THCS, and DTCS), heat stress (TTHS, THHS) and wet stress parameters (SMWS, HWS). In addition the model showed no sensitivity to random values (a random number generated between lower and upper intervals of changes in of all parameters).

### 6.3.2 Interaction of sensitive parameters

To our knowledge the effect of interacting sensitive parameters has never been investigated with the CLIMEX software. Increasing both parameters to which the model was most sensitive (i.e. DV3 and SM3 by +25%) resulted in an increase in the suitable locations in France, Italy, Spain, Poland, UK and Ireland in Europe. Also, most parts of the South Island of New Zealand turned into very suitable and suitable. Small parts of South America, Japan, Australia, the USA and China also turn into very suitable (Figure 6.6). Changes in this scenario seemed very similar to the scenario of increasing DV3 alone by 25% such that some locations in Ecuador (Pisayambo 20 to 46), Peru (Huaraz from 5 to 28) and Argentina were among those where the EI value increases up to 30 units. Also Italy and New Zealand showed more suitable areas when both values were increased (Figure 6.6).

The most dramatic changes occurred when both sensitive parameters were decreased by 25% (Figure 6.7). In this scenario, in Europe, the only very suitable areas were Germany and small parts
in the west of Poland. Very suitable areas decreased in Argentina but in general the interaction did not affect southern parts of South America highly. The major changes happened in Japan with the country changed from suitable to marginal) and in New Zealand where only marginal areas were visible in the South Island (Figure 6.7). With respect to the EI value change, there was very little increase where in just in two locations are in South Africa (Queenstown and the other in France (Brest) the EI value increased by 4 units with the rest remaining unchanged or decreasing dramatically such as South Georgia (Cumberland Bay 40 to3), the UK (Lerwick 39 to 3), Iceland (Vestmannaeyjar 36 to 0) and Argentina (Islas Ano 64 to 30).

Increasing one sensitive parameter and decreasing the other one at the same time resulted in interesting findings. Increasing DV3 by 25% and decreasing SM3 by 25% simultaneously resulted in a decrease of unsuitable, suitable and very suitable areas and an increase in marginal areas globally (Figure 6.8). The marginal area increase happened in Japan, Ireland, Italy, Australia and South America. A slight decrease of very suitable was visible in England, Italy, Germany, Ireland and France but at the same time some parts of Greece, Spain, Poland, Bulgaria and Romania change from suitable to very suitable. In this scenario, an increase in EI value was not very high and the maximum increase is 14 that belong to Argentina (Ushuaia 44 to 58) and Peru (Huaraz (5 to 18). A decrease in EI value was also seen in South Georgia (Cumberland bay 40 to3), Iceland (Vestmannaeyjar 36 to 1) and the UK (Lerwick 39 to 7).

Increasing SM3 by 25% and decreasing DV3 by 25% simultaneously was the only scenario that decreased the marginal, suitable and very suitable areas and consequently an increase in unsuitable areas was evident globally (Figure 6.9). The decrease in very suitable areas mainly happened in New Zealand, China (where there are no very suitable areas any more), France, Spain, Germany, Ireland, Poland, Turkey, Ukraine and Denmark. A decrease in suitable areas was also visible in the USA, China, Iran, Algeria, Argentina, and Australia. The marginal areas decrease in Italy, Peru, Algeria, Spain, Australia and New Zealand. The biggest change in EI value happens in South Africa (Grootspruit from 0 to 53), Zaire (Barumbu from 0 to 45) and Mexico (Torreon from 0 to 40). The dramatic decrease in EI happened in Argentina (Isals Ano from 64 to 0 and Ushuaia 44 to 0), Chile (Punta Arenas from 52 to 0), Falkland Island (Stanley from 49 to 0) and the UK (Wick from 41to 0).
The difference in area from the baseline for each scenario is shown in Figure 6.10. The chart showed that the highest increase in very suitable areas happened when both sensitive parameters increased by 25%. The maximum decrease in suitable and very suitable areas happened when both sensitive parameters were decreased by 25% resulting in substantial changes in EI value from suitable and very suitable to unsuitable.
Figure 6.1. The world map showing the potential distribution of dwarf bunt (baseline model).
Figure 6.2. The world map showing the sensitivity of model and Europe region to -25% changes of SM3 values (soil moisture content) from 1.6 to 1.2.
Figure 6.3. The world map showing the sensitivity of model to +25% changes of DV3 values from 13°C to 16.25°C.
Figure 6.4. The world map showing the sensitivity of model to -25% changes of DV3 values from 13°C to 9.75°C.
Figure 6.5. The world map showing the sensitivity of model and Europe region to +25% changes of SM3 values (soil moisture content) from 1.6.
Figure 6.6. The world map showing the sensitivity of model and Europe region to +25% changes of DV3 and SM3 values at the same time.
Figure 6.7. The world map showing the sensitivity of model and Europe region to -25% changes of DV3 and SM3 values at the same time.
Figure 6.8. The world map showing the sensitivity of model and Europe region to +25% changes of DV3 and -25% changes of SM3 values at the same time.
Figure 6.9. The world map showing the sensitivity of model and Europe region to -25% changes of DV3 and +25% changes of SM3 values at the same time.
Figure 6.10. Sensitive parameters and their interactions. Chart showing the change in the area of unsuitable (EI=0), marginal (EI=1-10), suitable (EI=11-25) and very suitable (EI>25) areas from the baseline model for sensitive parameters and their interactions in eight scenarios. Zero line means no change (area in million km²).
6.4 Discussion

While laboratory findings are often used as a starting point in the CLIMEX parameter fitting process, for many parameters such as the stress indices, there are rarely data available. So modellers can never be completely sure about the validity of the stress parameters that result from calibrating the model. In addition, because no initial values are usually available for stress parameters, fitting these parameters in CLIMEX can be purely subjective, and, as a result, can increase the uncertainty about the model output.

6.4.1 Presence data

The most important part of the modelling process for species distribution models is to gather good data with respect to the current locations of the target species. In this study I paid close attention to this issue. The main sources were CABI and GBIF but unfortunately, the GBIF database for plant pathogens is not as rich as it is for insects or other macroscopic species. Only 10 exact presence locations for dwarf bunt out of 78 locations were gleaned from GBIF and all of these locations were restricted to Europe. On the other hand, we found the CABI database was not adequate as most of the locations were mainly recorded at the regional or province level. Also there are often errors in the database or updates which may not be available at the time of modelling (Pascoe et al. 2012). One example was a report of dwarf bunt from central and Western Australia in CABI 2008-9 Crop Protection Compendium. Interestingly, the baseline model did not predict this location as suitable area for establishment which is accordance with there still being no report of dwarf bunt from this location. Kim and Beresford (2009) used the 2008-2009 version of the Crop Protection Compendium (CPCI 2008) in their study in 2009 which included the false presence records for Australia which could be a possible source of error in their predictions. These locations have since been removed from the database (CPCI 2012).

6.4.2 Sensitivity analysis

Based on the results of this study it is recommended that the effect of every parameter be investigated, not only its effects on the entire model but also its effect
on regional changes that may be of importance to decision making. While a parameter may not have a large effect on overall model predictions, it is possible that some target regions may be sensitive. Sensitivity analysis can be of great value for decision makers who not only see the change in the percentage area of the locations as parameters change but, higher sensitivity of some areas to a single parameter may be indicated contributing to a better understanding of pathogen behaviour or response to the environment. For example, the model showed higher sensitivity to DV3 than SM3 but Figures 6.2 and Figure 6.8 showed the projected areas in the United Kingdom were very sensitive to change in the upper soil moisture content. Increasing the value of SM3 from 1.6 to 2 (Figure 6.5), resulted in large changes in EI values for the UK whereas the overall model projections were seen to be more sensitive to higher temperature threshold changes. Dwarf bunt has been reported from most countries in central Europe but there is no report regarding its occurrence in the UK. The baseline model predicts that almost all parts of the UK are very favourable for the establishment of this pathogen. One of the reasons dwarf bunt is absent may be that the UK has particular controls at its borders to mitigate dwarf bunt entry and establishment. However there are other possibilities. Reports of dwarf bunt from Europe dates back to the mid 20 century so it is likely this pathogen has been able to enter the country over the previous 60 years. However, one can speculate that even if the pathogen has been able to enter into the UK, it may not have been able to survive due to some limiting factors such as environmental stress. Further investigation is needed to determine the possible reasons that may have prevented dwarf bunt establishing in the UK.

Changing SM values (±10% and ±25%), results in some parts of Croatia and Albania not fitting model projections despite have been reports of the presence of dwarf bunt in those regions 50 years ago (Ul’Yanishchev 1960). To our knowledge there have been no further reports of dwarf bunt from this area. In addition, changing SM3 from 1.6 to 1.2 showed that the extent of the very suitable areas decrease dramatically but some areas such as Argentina, and some parts of Germany and Poland are not very sensitive to these changes (Figure 6.2). Such a result can be useful for decision makers, who might need to consider different strategies to prevent dwarf bunt
establishment in such locations that experience extreme changes. Additionally, such results can indicate how much care should be taken determining the value of some parameters that can have great effect on the models output. This is especially so for soil moisture, which for this study, there were no data available to guide the initial choice of a suitable range.

As described in the results, changing a single parameter can show that some locations have a high degree of sensitivity such that a small change can result in a large increase in the EI value. The best example is Islas Ano in Argentina, Punta Arenas in Brazil, Huaraz in Peru and Wick in the UK. We strongly recommend that careful attention should be given to sensitive areas, and by giving appropriate advice to decision makers, different strategies for prevention of dwarf bunt establishment can be employed for these locations.

In New Zealand, Dunedin, Queenstown, Timaru, Christchurch and Gore were among those areas that showed increased sensitivity in most of the scenarios. For example by decreasing DV3 by 25% the EI value for Queenstown changes from 19 to 7 or by decreasing SM3 by 25% the EI value of this location changed from 19 to 0.

6.4.3 Interaction of sensitive parameters

This study is the first attempt to investigate the interaction between sensitive parameters in CLIMEX. Such a study can show which of the parameters is the most dominant in each scenario. The first and second scenarios where both parameters were increased or decreased at the same time, respectively gave sensible results as increasing them both increases the suitable and very suitable areas and decreasing both of them decreases both suitable and very suitable areas. But decreasing SM3 and increasing DV3 decreased suitable and very suitable areas similar to the scenario where only SM3 was decreased, illustrating how sensitive the model was to changes in SM3. On the other hand, by increasing SM3 and decreasing DV3, the extent of unsuitable areas increased to 5 million km². Such an increase that was evident in Figure 6.8 not only comes from turning marginal areas into unsuitable areas but also from changing suitable and very suitable areas to being unsuitable. By considering the interaction of sensitive parameters, this study indicated the sensitivity of areas.
such as Australia, Tasmania, Spain, Morocco, Peru and Algeria. Also, some countries such as Italy, Argentina, Pakistan, central Asia and South Korea showed less sensitivity to most of the scenarios. In addition, most of the locations that showed high sensitivity by changing a single parameter, also showed high sensitivity when both sensitive parameters were changed at the same time (such as Islas Ano and Puntas Arenas) which can indicate the dominant effect of some parameters over other parameters. The higher sensitivity of such areas to some parameters can be very important requiring further investigation to clarify if such changes are due to error in climate data or the result of the unique nature of such localities.

6.5 Conclusion

In conclusion, sensitivity analysis can illustrate the effect of uncertainty or error in parameter measurement or selection on the interpretation of model output. This is the case no matter how simple or complex the model. With respect to CLIMEX projections, sensitivity analysis can also give insight into how sensitive some locations are to specific parameter changes which need to be considered further in both the modelling and decision making process. This study has demonstrated that sensitivity analysis should be an integral part of the process of turning scientifically relevant information into information for making decisions no matter what model is used. The sensitivity analysis of the CLIMEX model used in this study revealed how small changes in parameter values that are used in a baseline model (which is the one that interpretation of the results is usually based on) can severely affect the prediction outcomes in some regions. Another important point is that, unlike some parameters such as temperature thresholds where experimentally determined data can be used to base initial calibrations, uncertainty increases when the modeller calibrates other parameters such as soil moisture content or specific stress parameters with limited or no preliminary data available. In this latter case it is essential that the modeller test other possibilities about these more uncertain parameters especially when the aim of study is to give advice to decision makers.

Additionally, the UK results showed how sensitivity analysis can bring new questions to mind and demonstrate the opportunity to investigate unusual or
unexpected findings of the baseline model. Finally, any recommendation given to policy-makers should be given along with the consideration of uncertainties in the model enabling them to choose the best strategy to prevent target species establishment by considering all possibilities based on the most complete and realistic predictions.
Chapter 7
Model Consensus

7.1 Introduction

It was shown in the previous chapters that some models perform better than others based on chosen performance criteria. Different model projections can occur for many reasons such as model type or algorithm used, data and the target species modelled (Gallien et al. 2012; Hernandez et al. 2006). Benito et al. (2013) and Valle et al. (2013) suggest that some models outperform others and the rest are usually rejected. However, there is a debate as to which models should be considered uninformative (Richards et al. 2011; Barai et al. 1999). In fact, with respect to overall accuracy Schaffer (1994) showed that in machine learning efforts an algorithm generally is not dominated by any other algorithm. Additionally, based on the “no free lunch theorem”, regardless of performance measures, any two algorithms can perform similarly across all possible problems (Igel and Toussaint 2005; Wolpert and Macready 1997). To avoid subjectivity, the choice of model is usually done through a traditional or systematic approach. In the traditional method, models are selected based on guidelines and experience, and in a systematic approach they are compared statistically. However it is important to remember that in any modelling effort the objective is to gain insight about the problem and increase our understanding of the data (Valle et al. 2013; Worner et al. 2010). Regardless of the method used in model selection, it is always possible that discarded models will have useful information for solving a problem. Therefore, to account for uncertainties involved in various models, as well as to benefit from different models and algorithms, ensemble models have been suggested as a solution (Freckleton 2011; Araújo et al. 2007). Initially, Willard Gibbs introduced the idea of the ensemble in the field of statistical mechanics in 1878 (cited in Araújo et al. 2007). Simply put, the ensemble approach is to consider a large number of simulations of a system all at once, where each one is representative of a likely state of the real system at a particular time. Additionally,
since individual projections of each model simulation may offer some independent information, it has been observed that by combing the projections, a lower mean error is obtained compared to any individual projection (Araujo et al. 2007; Bates and Granger 1969). However it should be noted that, based on the conservation law of generalisation performance and “no free lunch theorem” described earlier, there is no guarantee that an ensemble model or some combination of models will always work better than an individual model on any given modelling problem (Zhou et al. 2002; Barai et al. 1999; Sharkey 1999). When the aim is to ensemble models, Wolpert (1992) explains that the members of an ensemble should be of different types of models and not just variations of a single model, and they should add information for building a good model. However in contrast, Kwok et al. (2013) and Buntine et al. (1992) have pointed out that members of an ensemble can also be variations of a single model. To achieve an ensemble model, theoretically, different ensemble techniques have been suggested (Araujo et al. 2007), of which some of them, such as weighted average and committee average, have been applied in several studies (Gallien et al. 2012; Marmion et al. 2009). Additionally, the results of evaluating the performance of ensemble models against a single model and occupancy models have been tested with promising results (Comte and Grenouillet 2013; Gallien et al. 2012).

In the current study, different correlative and mechanistic models were investigated which use quite different algorithms and not all give outputs in the form of probability of establishment, so simple averaging is not appropriate. To benefit from the projection of the different models, a method was needed to combine different results appropriately. When one tries to combine projections of different models the weighted or un-weighted average of projections can be used. Among available ensemble methods, the committee average method can take the un-weighted average of model projections or the average of different model output and then assign equal probabilities to each model (Prasad et al. 2006; Araújo et al. 2005; Heyuan et al. 1995). This approach has a particular advantage, because it can combine the outputs of quite different algorithms. This method was used by Gallien et al. (2012) to construct consensus models, but only for correlative models (this method also has been called “bounding box” by Araujo et al. (2007)). Because the committee average method can
combine the outputs of any kind of model such as mechanistic and correlative models, it was used in this study to construct a consensus model combining the outputs of different models. The aim is to build a consensus model that is a combination of the three mainstream modelling methods: mechanistic, presence-only and presence- (pseudo-) absence model types, which theoretically would result in a model that can benefit from each model’s strengths.

7.1 Methods

In the consensus model approach used in this study, all model outputs are transformed into binary maps (presence and absence of a species) and the agreement (consensus) among model projection of species presence and absence are presented. In other words, the final output shows the number of models which agree on the suitability of an area for establishment of the target species. To make sure that the resulting consensus performs better than any of the single model forecasts, a measure comprising the median and above (four or more) of the number of models that agree at any location was used to present the final results. This method of presentation also simplified map interpretation.

7.1.1 Selection of cut-off threshold

To convert the probability outputs of the correlative models into binary maps, selection of a threshold to determine presence or absence was needed. The issue of selecting a threshold is the subject of on-going studies and to date, several methods have been suggested (Jiménez-Valverde and Lobo 2007; Liu et al. 2005; Woolf et al. 2002). In general, this threshold can be selected in a subjective or objective way. In the subjective method, a constant threshold of 0.5 (meaning probability of 0.5) is taken as the cut-off, which is a common method in ecological studies (Gallien et al. 2012; Worner et al. 2010; Stockwell and Peterson 2002; Manel et al. 2001; Manel et al. 1999). Alternatively, many objective methods are available, which have tried to estimate the best method for selecting a threshold to maximize the agreement between current and projected distribution (Liu et al. 2013; Liu et al. 2005). Although debate around these methods still continues, one thing that researchers agree on is that the selected threshold should be based on the research goal (Jiménez-Valverde et
al. 2007). In this study, a threshold of 0.5 was selected, which has been widely used in different studies to generate binary distribution maps (Gallien et al. 2012; Manel et al. 2001; Manel et al. 1999). We applied this threshold to correlative model outputs to generate presence-absence binary maps since several authors suggest that it gives a neither conservative nor liberal projection (McPherson et al. 2004; Manel et al. 2001; Manel et al. 1999). As discussed in the methodology chapter, for CLIMEX, the eco-climatic index (EI) values range from 0 to 100 which the modeller arbitrarily can categorize into unsuitable, marginally suitable, suitable and very suitable. However, an EI value of 0 is always considered as unsuitable (absence) (Yonow et al. 2013; Hoddle 2004; Sutherst 1999), hence, EI values of greater than 0 were categorised as indicating potential presence of the species. These categories were used to convert CLIMEX output into a binary map.

### 7.1.2 The selection of top-performing models in the Multi-Model Framework

In this study, the AUC, an often-used and well-established method was used to select the top five models from the Multi-Model framework. The AUC has been criticized by Lobo et al. (2008), among others who point out that this measure of performance can be misleading especially when the absences or pseudo-absences are taken from areas which are very distant from the geographical range of the species. When there is high geographical discrimination between presence and pseudo-absence points, the chance of selecting environmentally extreme pseudo-absences increase such environmental discrimination leads to over-fitting of models.

An over-fit model could have a high AUC value but that does not necessarily mean the model is performing better and can be misleading when model performances are compared (Stokland et al. 2011). The novel three-step method by Senay et al. (2013) was implemented to avoid selection of highly discriminated pseudo-absences by selecting the pseudo-absence points from environmentally meaningful areas as well as geographical buffers around presence points. Furthermore, the Kappa index has been used in many studies but this performance criterion has also been criticized as having serious limitations as being highly dependent on prevalence (Allouche et al.
Fortunately, in all cases of choosing the top five performing models with respect to the AUC, at least four models which had the highest AUC had also highest Kappa, sensitivity, TSS and specificity, which increase the confidence in the selected models. Moreover, in the Multi-Model Framework (Chapters 3, 4 and 5), the AUC of the majority of models was excellent (AUC > 0.9), and there was no obvious difference in their performance based on the AUC score, however the selected models were also checked to confirm that they had realistic and biologically-plausible visualizations (Table 5.1) (Senay et al. 2013; Araújo et al. 2006; Elith et al. 2006).

Table 7.1. The top five models in Multi-Model Framework (MMF) with highest AUC selected for making the consensus model from Chapter 3, 4, 5). LDA= linear discriminant analysis, QDA= quadratic discriminant analysis, LOG= logistic regression, NB= naïve Bayes, CART= classification and regression tree, CTREE= conditional tree, KNN= K-nearest neighbor, SVM= support vector machine, and NNET= neural network.

<table>
<thead>
<tr>
<th></th>
<th>Psa</th>
<th></th>
<th>Dwarf Bunt</th>
<th></th>
<th>Guava Rust</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>AUC</td>
<td>Model</td>
<td>AUC</td>
<td>Model</td>
<td>AUC</td>
<td>Model</td>
</tr>
<tr>
<td>NB</td>
<td>0.978729</td>
<td>KNN</td>
<td>0.955926</td>
<td>NNET</td>
<td>0.952353</td>
<td></td>
</tr>
<tr>
<td>KNN</td>
<td>0.977408</td>
<td>SVM</td>
<td>0.952716</td>
<td>SVM</td>
<td>0.952064</td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>0.973841</td>
<td>QDA</td>
<td>0.950247</td>
<td>CART</td>
<td>0.942274</td>
<td></td>
</tr>
<tr>
<td>NNET</td>
<td>0.969084</td>
<td>NNET</td>
<td>0.949259</td>
<td>KNN</td>
<td>0.935957</td>
<td></td>
</tr>
<tr>
<td>CTREE</td>
<td>0.96631</td>
<td>NB</td>
<td>0.945185</td>
<td>CTREE</td>
<td>0.912471</td>
<td></td>
</tr>
</tbody>
</table>

Based on the methods described previously, five models per species from the Multi-Model Framework (more than half of the models available in the Multi-Model framework) with the highest AUC scores (Chapters 3, 4 and 5) were incorporated into the consensus model along with MaxEnt and CLIMEX, in the form of binary maps to give seven models in total. As probability maps produced by Multi-Model Framework included areas with no data (indicating locations where the probability is less than 0.1), those pixels were converted into zeroes. This was done using the raster calculator tool in ArcGIS 10.1 with the map algebra expression: “Con(IsNull("model"),0,"model")”. Then, based on the cut-off threshold, the raster layer of each model projection was reclassified into two classes (probability ≤0.5) and...
1 (probability >0.5). In ArcGIS 10.1, using the spatial analyst tool, Equal to Frequency, a consensus (model agreement) map was produced. This tool requires two raster files, one as “Input Value Raster” and one as “Input Raster(s)”, in this case, the seven model outputs. To make a consensus map, all Input Rasters were compared against the Input Value Raster. An Input Value Raster file was created (with the same extent and cell size of the input raster) and all pixels were assigned the value of 1, meaning we are interested in knowing for each pixel, how many models out of seven have a value of 1. The output raster (consensus map), is a raster map that shows for each pixel how many models have a value of 1 (i.e. suitable areas). Because the aim was to take the median and above for each pixel, the areas where more than half of the models agree (a value of 4 or more), were considered to be suitable.

Finally, to compare the performance of the consensus model with individual models, a confusion matrix was used. Along with presence points, the absence points generated by the Multi-Model Framework were used to calculate the false-positive rate and false-negative rates which are important elements of the confusion matrix. From the confusion matrix (discussed Chapter 2 section 2.1.1.1.5), seven performance criteria were calculated for the seven individual models and consensus model (Tables 3.2, 4.2 and 5.3). No significance test was performed to test if the consensus model performs better because the data used to construct the confusion matrix were not independent from the model fitting process and no independent data were available at the time. Despite that, the results appear to show that the consensus model fits the training data better. Additionally, the single models in MMF were validated using cross-validation which was explained earlier in Chapter 2 (2.1.1.1.4). However, a better test of the performance of the consensus model against other models would be to put aside a portion of presence data for validation of consensus model if enough are available. The results are shown in Figure 7.2.
7.2 Results

The results of consensus models for three target species were presented in Figures 7.1, Figure 7.2 and Figure 7.3.

7.2.1 Psa consensus model

The results of the consensus model for Psa indicated agreement of all seven models on the suitability of central and eastern China and the whole of Japan and small areas in the north of Iran for establishment. More than four models agreed that the main kiwifruit growing areas of Europe such as Italy, Portugal, Belgium, Spain, France and Greece were suitable for Psa establishment. In South Africa, five of the models projected very small areas in Swaziland and limited areas around Cape Town as suitable. The highest agreement among models in South America (six models) was in western Chile around Santiago and southern Uruguay, and four models agreed on the suitability of coastal areas in Argentina. In most of northern California, four or five models predicted climate suitability, with the exception of a very small area around Tacoma, Washington state which was predicted by seven models as suitable for Psa establishment (Figure 7.1). The consensus model shows that seven models agree on suitability of areas around Iowa in the USA. Small areas in south east of Australia (Victoria) and in Tasmania (western part) are projected as suitable by seven models.

For New Zealand, the whole country (except the highlands in the South Island) was predicted as suitable by six models, indicating high levels of agreement among models on the suitability of New Zealand for Psa establishment (Figure 7.1).

7.2.2 Dwarf bunt consensus model

The consensus model showed the Northern hemisphere more suitable than the Southern hemisphere for dwarf bunt establishment. The map shows that Europe and areas in western and eastern USA were the most suitable areas. Furthermore, central USA and Eastern Europe along with the UK and large areas in France were projected suitable by six models (Figure 7.2).
For New Zealand, the consensus model showed four to five models agreed that small areas in the North Island and eastern coast of the South Island are suitable for dwarf bunt establishment (Figure 7.2).

7.2.3 Guava rust consensus model

The consensus model predicted that mainly regions in the Southern hemisphere are suitable for guava rust establishment. Areas with total model agreement are in southern and Amazonas regions of Brazil, southern Paraguay and Bolivia, eastern coastal areas in South Africa, eastern Madagascar, and eastern coastal areas in Australia, Papua New Guinea, Southern China, southern Vietnam and Laos, Vanuatu and all of New Caledonia (Figure 7.3).

In New Zealand, the most suitable areas were in Northland and Auckland where all seven models agreed on climate suitability for guava rust establishment. The South Island is predicted as unsuitable except for very small areas around Nelson which four models predicted as suitable (Figure 7.3)
Figure 7.1. Consensus model projection for Psa. The median map shows the areas in which at least half of the models agree (light green) on climate suitability for the target species.
Figure 7.2. Consensus model projection for dwarf bunt. The median map shows the areas in which at least half of the models agree (light green) on climate suitability for the target species.
Figure 7.3. Consensus model projection for guava rust. The median map shows the areas in which at least half of the models agree (light green) on climate suitability for the target species.
Figure 7.4. Comparison of consensus model with single-models based on seven performance criteria. The charts show the percentage changes of accuracy measures in individual models from the consensus model.
7.1.1 Evaluation of goodness of fit of consensus model versus individual models

Figure 7.4 show that the consensus model outperformed all single-models in all performance criteria except in one case where recall (sensitivity) of CLIMEX model was higher than the consensus model. Furthermore, Figure 7.4 shows that although the consensus model generally performs better than the individual models, the level of performance varies across species and performance measures. For example, for Psa, and dwarf bunt, in some performance criteria, models such as QDA, KNN, SVM, CLIMEX and MaxEnt perform almost as well as the consensus model. For guava rust, generally CLIMEX performance is lower than the consensus model and the SVM model in most cases is as good as the consensus model.

7.2 Discussion

Because there are many different species distribution models, scientists have been challenged to try to select the "best" model and many have attempted to do exactly that (Webber et al. 2011; Townsend Peterson et al. 2007; Hernandez et al. 2006; Guisan et al. 1999). For example Guisan et al. (1999) compared the predictive power of generalized linear models (GLM) versus Canonical Correspondence Analysis (CCA) for plant distribution and concluded that GLM would provide better species specific-models. Hernandez et al. (2006) tested four different models including Bioclim, Domain, GARP and MaxEnt to evaluate the effect of sample size treatment and evaluation measures on 18 species (amphibians, birds, mammals, reptiles and insects). They concluded that, with small sample size, MaxEnt was the most capable method among the models with respect to providing good results. In contrast, in another study by Townsend Peterson et al. (2007) showed that MaxEnt models had low transferability compared to the GARP model. However, there are studies which question the selection of a single model as the “best” (Elith and Graham 2009; Marmion et al. 2009; Araújo et al. 2007). For example, Elith et al. (2009) used simulated data to show the differences among different models namely, generalized linear models (GLM), boosted regression tree (BRT), random forest (RF), genetic algorithm for rule-set prediction (GARP) and MaxEnt. They concluded that no single
model is superior to others and different models can be used based on the question asked. Marmion et al. (2009) argue that because different modelling techniques vary in their performance abilities, some consensus methods provide more robust predictions than all single models. The limitations and uncertainties of individual species distribution models such as clarification of niche concept, parameterization strategies and model selection have been addressed in different studies, but many believe that insufficient attention has been paid to the similarities or discrepancies among different models (Elith et al. 2009; Araújo et al. 2006; Hernandez et al. 2006; Loiselle et al. 2003). Often the solution has been to use model comparisons to find the best performing model (Elith et al. 2006; Lawler et al. 2006; Prasad et al. 2006). However, it’s been argued that there is no “true” or perfect model and even a model with high accuracy on independent data does not necessarily provide the most realistic representation of distributions in novel areas or time (Araújo et al. 2005; Thuiller et al. 2003). Additionally, identifying a model as the "best" often results in model uncertainties being ignored that might lead to over-confidence about model projections when making inferences about the subject of interest (Raftery et al. 1997).

Arguably, the most common method that has been employed to deal with uncertainties in model projection is model averaging (Richards et al. 2011; Symonds and Moussalli 2011; Johnson and Omland 2004). In the majority of these studies, the Akaike information criterion (AIC) is used to average models or to select the best model. The AIC is based on information theory and it minimises the information loss due to using one candidate model as a “true” model. By using AIC, multiple models can be evaluated at the same time to verify how sure one can be about a given model being the “best”. In this way, not only is the information regarding the best-fitting model used but the information about the fit of other models is also used and inferences will be based on a series of models. Each model contributes to the final model based on its degree of fit and hence the models which are the better fit to the data have a greater role compared to other models which have lower fit to the data (Freckleton 2011; Claeskens and Hjort 2008). In this way, the strength of different candidate models can be combined and presented. However, model averaging is not always practical since it can give mixed results. Recently, questions have been raised
about the methodology used and whether model averaging leads to improvement of
inferences (Richards et al. 2011; Richards 2005). For example, while a model
averaging technique proposes to deal with two types of uncertainties (uncertainty in
parameter estimates and uncertainties in the model), the effect of collinearity among
predictors on model averaging is unknown (Freckleton 2011). Because the methods
of model averaging can be questioned and since the selection of the best model is not
clear-cut and model comparisons are not always helpful (Thuiller et al. 2003),
consensus models have been proposed as a solution which use the agreement among
the models as a more realistic approach to modelling species distributions (Gallien et
al. 2012; Marmion et al. 2009; Araújo et al. 2007). The most common consensus
methods used are those where the input models are from the same category or same
type, for example all machine-learning or all correlative models (Poulos et al. 2012;
Marmion et al. 2009). In this study, models belonging to different categories were
combined in a consensus using a method proposed by Gallien et al. (2012) that is
flexible enough to combine the results of different modelling approaches. It should
be noted however, that consensus forecasting is no alternative to well-conceived
models built using high quality data. Additionally, a consensus model may not
always be better than a single-model approach because the quality of a consensus
model will still depend greatly on the quality of the forecasts by each of the single-
models (Marmion et al. 2009). It is also worth mentioning that, similar to any
forecast, a consensus forecast will never be perfect as uncertainty cannot be
eliminated, but a consensus can reduce the risk of making unrealistic decisions
because if a single model performs poorly, the effect is likely to be offset positively
by other models included in consensus model (Semenov and Stratonovitch 2010;
Araujo et al. 2007).

The results of this study show that a consensus model can perform better than all the
individual models used in this study with respect to fitting the available data based
on the different performance criteria used and the data available for the target
species. Like any forecast, the value of a consensus forecast in practice depends on
the objectives of the decision makers. In some cases, if a consensus model projects a
high false negative (lower chance of establishment than what the data suggests), then
that result might authorities investing in controlling an invasive species that may pose a high threat, despite that the. Conversely, if a consensus model projects higher false positive (higher chance of establishment than what the data suggests), then more unnecessary resources would be spent on control strategies. The consensus approach demonstrated here is likely to appeal intuitively to decision makers giving more confidence in the modelling process. Additionally, it allows us to say with a degree of confidence what would unlikely to happen.

Finally, if enough presence and absence data are available, a portion of these data should be put aside to validate the consensus model. In the current study, no true absence data were available and for presence data because a portion are already left aside during cross-classification and bootstrapping (which is a type of validation) putting another part of the presence data aside for consensus model validation would have meant each model was trained on less presence data which is not advisable.
Chapter 8
Effect of climate change on projected potential distributions of the target species

8.1 Climate change

The first assessment by the Intergovernmental panel on climate change (IPCC) in June 1990 determined that there has been an alarming increase in concentration of greenhouse gasses (GHG) since 1750 (Houghton et al. 1990). According to this assessment, the effect of such an increase would result in atmospheric imbalance and global warming. Based on the greenhouse-gas increase in the past century, the first assessment report (AR1) by the IPCC estimated a 1°C increase in temperature (temperature recorded by instruments showed a 0.5°C increase) (Houghton et al. 1990). Thereafter, the second, third and fourth assessments of climate change were released in 1996 (Houghton 1996), 2001 (McCarthy 2001) and 2006 (Pachauri and Reisinger 2007), respectively and more recently the fifth assessment report (AR5) was released between September 2013 and November 2014 (www.ipcc.ch). To simulate the response of climate to increasing greenhouse gas, numerical models such as General Circulation Models (GCMs) are used which project climate change at different spatial and temporal levels. However, notably uncertainties are involved in these simulations as our knowledge about cloud formation, hydrological phenomena and ocean circulation is not yet complete (Rosenzweig et al. 2001).

The reports prepared by intergovernmental panel on climate change (IPCC) are revised based on improvements in the models used to generate the assessments reports. These assessment reports and related climate-change scenarios are based on the driving forces of GHG emissions, namely, different technological advances, demographic and economic changes and nitrous oxide, methane carbon dioxide and sulphur emissions at regional and global scales. In total, forty scenarios have been developed in which each one belongs to one of four storylines. The IPCC Special Report on Emission Scenarios (SRES) does not include any climate initiatives such as
the United Nation convention on climate change or the Kyoto protocol (Breidenich et al. 1998). Instead, GHG emissions are increased or mitigated by different policies implemented by governments. These policies can be pollution control, resource usage management, economic developments or demographic changes.

8.1.1 Emission scenarios

A1. The scenario A1 considers a very rapid economic growth and a peak in global population in the middle of the century which declines afterward. It also is based on the assumption that an increase in cultural and social interaction is expected. This A1 storyline focus is on technological change, fossil energy sources (A1F), non-fossil energy sources (A1T), and a balance in all sources (A1B). In total, this storyline includes 17 illustrative scenarios (Figure 8.1).

A2. The A2 storyline considers a very heterogeneous future world assuming that local characteristic identities are preserved and global population increases continuously. Technical changes and economic growth are more local, fragmented and slower compared to other storylines. This storyline has just one family group and includes six illustrative scenarios (Figure 8.1).

B1. This storyline assumes a convergent future world in which the population peaks mid-century and declines afterward. The economic changes would be rapid and new, clean and efficient technologies will be introduced. The focus is on global environmental, social and economic sustainability and fairness but no further climate initiatives. This scenario has just one group and includes nine illustrative scenarios (Figure 8.1).

B2. Population increase, technological and economic development rate are lower than A2, and the diversity of these changes is more than in A1 and B1 storylines. The focus is on the local and regional scale and environmental protection. The focus is on environmental, economic and social sustainability. This scenario has just one family group and includes eight illustrative scenarios (Figure 8.1).
Table 8.1. The best estimate and the likely range of temperature increases based on different scenario projections for 2090-2099. (IPCC, 2000)

<table>
<thead>
<tr>
<th>Case</th>
<th>Best estimate</th>
<th>Likely range</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>1.8</td>
<td>1.1 - 2.9</td>
</tr>
<tr>
<td>A1T</td>
<td>2.4</td>
<td>1.4 - 3.8</td>
</tr>
<tr>
<td>B2</td>
<td>2.4</td>
<td>1.4 - 3.8</td>
</tr>
<tr>
<td>A1B</td>
<td>2.8</td>
<td>1.7 – 4.4</td>
</tr>
<tr>
<td>A2</td>
<td>3.4</td>
<td>2.0 - 5.4</td>
</tr>
<tr>
<td>A1F1</td>
<td>4.0</td>
<td>2.4 – 6.4</td>
</tr>
</tbody>
</table>

Figure 8.1. The main characteristics of the storylines and scenario families. Each one of these scenarios predicts 1°C increase in average temperature by 2050 and thereafter up to 4 and 6°C by 2100. They also project an increase in global mean rainfall from 5% to 15%. Image credit: (IPCC, 2000).
8.2 Agriculture and climate change

Vulnerability to climate change has been assessed both globally and locally, in many different fields such as the economy, ecology, environment and agriculture of the target regions (Stern 2007; Smit and Skinner 2002; Walther et al. 2002). There is now indisputable evidence that the earth’s climate is changing, and sea levels are rising. Furthermore, based on global records, May 2012 was the warmest May since 1880. Such changes can severely affect agricultural productivity and cause food insecurity. Some examples of climate events which have affected agriculture are drought in Sahel in 2000, the U.S. drought of 1988, El Niño of 1997-8 and climate extremes of 1998-2000 due to La Niña (Rosenzweig et al. 2001)

In 2011, some studies reported that the harvest failure in Niger put at risk half of the country’s population due to increase floods occurring as a result of increase in temperature and climate change (http://www.unep.org/newscentre/default.aspx?documentID=2661&articleid=8971). Monsoon failure in 2010 in a region of India destroyed half of the harvest and in Pakistan unprecedented (one in 1000 year) flooding massively affected the agricultural sector (Straatsma et al. 2010). With an increase in average temperature that has been predicted by all scenarios, overall crop yield will be decreased extensively. When considering the potential decrease in agricultural production due to climate change and fact that already around 2 billion people live in the driest areas on earth which are projected to get more dry, how to feed 9 billion people (the projected population by 2050) is challenging. Von Lampe et al. (2013) in their study of using 10 global economic model estimated a 60-111% increase in agricultural production is necessary from 2005 to 2050 to feed the projected 9 billion population at that time. The increases in food prices in 2007-2008 due to recession and local climatic events have already caused some agencies to evaluate our ability to feed the world in the coming decades (World Food Program 2009). Because developing countries are more reliant on agriculture than developed countries even the smallest effect on food production will have large effects on these countries, and it has been estimated that by 2050 the prices for maize, rice and wheat would be respectively, 131%, 78% and 67% higher than they are now. In contrast to these estimates, a study
on the effect of rising CO$_2$ concentration on crop yield, by Long et al. (2006) showed that rising CO$_2$ can have fertilization effects which can increase yield and consequently can offset the losses caused by global warming. On the other hand, Burney et al. (2010) suggest that although agriculture is responsible for 14% of greenhouse gasses emissions and has increased by up to 32% in developing countries between 1995-2005, agricultural intensification through more efficient fertilizer production and use which result in higher productivity of agricultural system has mitigated carbon emissions since 1961 (Burney et al. 2010). Furthermore, it has been suggested that economic and technological changes can contribute to mitigation of losses in agriculture in various ways (Godfray et al. 2010; FAO 2009). One way to decrease the loss of agricultural products is by better controlling and managing pests and plant pathogens that can also be affected by climate change.

### 8.3 Climate-change effects on plant pathogens

The disease triangle represents the important factors required for development of plant disease, being the presence of a virulent pathogen, a susceptible host plants and a suitable environment (Francl 2001). In addition, other factors such as human activity also contribute to disease establishment. However, in many cases, the environment is the limiting factor with optimum environmental conditions required for all stages of disease from inoculum production, germination and growth and infection by the pathogen. Such global warming or any changes in climate may affect the incidence and severity of plant diseases (Crowl et al. 2008; Burdon et al. 2006). Prior to release of the outputs of the general circulation models (GCMs) which provide predicted values of climate variables as a result of complex factors, most of the projections on the effect of climate change on plant pathogens incidence and epidemics relied on adding fixed changes in temperature and precipitation to the existing current values (Coakley et al. 1999). Even so, there are limited studies on the impact of climate change on plant pathogens but those available agree that the change will happen and it will be in three main areas: increased damage from plant pathogens, increased geographical distribution of plant pathogens and decreasing effectiveness of disease management policies (Anderson et al. 2004; Chakraborty et
al. 2000; Coakley et al. 1999). Because temperature, light and moisture are the main factors controlling growth and development of pathogens, insects and weeds, climate change will clearly affect their spatial and temporal distribution (Garrett et al. 2006; Rosenzweig et al. 2001). On the other hand, climate may affect the use of pesticides as application is highly related to the timing and intensity of rainfall and the persistence of the pesticide's chemical components is also affected by temperature (Coakley et al. 1999). Although it is not easy to test whether extreme weather conditions will increase damage by pests (Rosenzweig 2000; Yang and Scherm 1997), some attempts have correlated climate change and extreme weather conditions to plant disease epidemics, such as wheat stem rust on the U.S. Great Plains and wheat stripe rust epidemics in the U.S. Northwest (Scherni and Yang 1995; Hamilton and Stakman 1967). Although it has been predicted that temperature increases will occur globally, it is expected that the effects of global warming will be more intense in some regions such as those with a tropical climate. The reasoning is, because the range of suitable temperature for growth is narrow for tropical species, and as they are already living very close to their optimal temperature needs, the species would be more sensitive to temperature changes. This means, that while climate change has caused the extinction of some range-restricted species such as polar and mountaintop species (Parmesan 2006), some pathogens, which have been able to live at higher altitudes and in temperatures cooler than their optima, may have the chance to cause epidemics under new environmental conditions (Ghini et al. 2011; Deutsch et al. 2008). Conversely, in a study by Salinari et al. (2006) on downy mildew of grapevine under climate change, despite a decrease in precipitation, an increase in the disease pressure was simulated because the increase in temperature counterbalanced the decrease in rainfall. This is in contrast to the perception that downy mildew or foliar diseases are more affected by rainfall than temperature. Furthermore, climate change may indirectly affect pathogen development through changes in plant canopies and foliar density (and consequently, increases in leaf wetness (Huber and Gillespie 1992; Colhoun 1973). Among other plant pathogens, polycyclic pathogens may be affected more by a temperature increase because global warming may lengthen growing seasons,
consequently increasing the length of optimal time for pathogen reproduction and number of generations within a particular period of time or season. Although this can be true for some plants as well, but it’s been claimed that even a modest climate change can have a considerable impact on plant pathogens compared with plants because of their shorter life cycles, their physiological sensitivity to temperature, reproductive potential and mobility which can help them to disperse more quickly (Parmesan 2006; Davis et al. 2005; Ayres and Lombardero 2000).

The effect of climate change on plant-pathogen interactions can be positive, negative or neutral depending on the individual plant disease (Figure 8.2) (Chakraborty et al. 2000). For instance, in dry or stressful conditions a plant closes its stomata and leaf and root growth are inhibited, making the plant more vulnerable to pathogens. McElrone at al. (2005) showed that an increase in CO$_2$ level resulted in decrease in disease incidence by up to 50% by *Phyllosticta miniasome* (foliar pathogen) because the elevated CO$_2$ changed the leaf chemistry and decreased stomata opening. In

Figure 8.2. A summary of the major influences due to changes in atmospheric composition, land use and climate on plant diseases and the level of understanding of disease and climate. The understanding of disease and climate is highest at the paddock scale. Image credit: (Chakraborty 2000).
addition, by increasing the temperature and decreasing soil moisture content, the activity of some soil-borne pathogens are decreased and therefore in dry soils have less probability of reaching the root system (Kang and Banga 2013). In addition, the response of plants to climate changes or abiotic stresses may happen through gene expression or by activating general defence mechanisms, which can increase resistance to some pathogens, but increases susceptibility to other pathogens (Gessler and Pertot 2012). In barley, increased levels of CO$_2$ triggers the defence system to produce an increase in silicon and formation of a defence mechanism (papillae) at the penetration site that can protect the plant against plant pathogens (Hibberd et al. 1996). In one study, a 1°C increase in temperature and 30% rainfall increase caused a twofold increase in susceptibility of wheat and barley to stem rust compared to current conditions, but there was no change in the powdery mildew infestation levels in small grains (Jahn et al. 1996). There is already some evidence available on the effect of climate change on plant pathogens outside of the laboratory. For example, in the U.S.A., reduction in growth of Oregon pine caused by Swiss needle cast disease (*Phaeocryptopus gaeumannii*) has been correlated to elevated temperatures in winter (0.2°C-0.4°C increase per decade) and a rise in spring rainfall in coastal areas of Oregon since 1970 (0.7-1.5 cm per decade (Stone et al. 2008; Manter et al. 2005). Recent increases in the severity of blackleg disease (*Leptosphaeria maculans*) on oilseed rape by 18% in the United Kingdom has also been correlated to increased temperatures in winter, which are likely to increase further in the 2020s and 2050s. In Germany, heavy damage by Cercospora leaf spot on sugar beet has also been correlated with an increase in temperature throughout the last century (0.8°C-1°C) and as a result, the disease occurs a few days earlier than what it did last century (Pertot 2012).

Based on the importance on the potential effect of climate change on plant pathogen discussed here, the aim of this chapter is to project the climate suitability of three target species based on two climate change scenarios (A1B and A2) in 2030 and 2090.
8.4 Methods

Similar to the current climate scenario, the future climate scenario dataset was acquired from the CliMond website. The data from CliMond were used as the same variables as used for current scenarios in previous chapters to compare results and changes in climate suitability for the target species in future decades. CliMond’s future scenarios have the same conformal or orthomorphic spatial and climatological characteristics as baseline data and also downscaling of GCM data is consistent in the future climate scenario dataset. For this study, two scenarios A1B and A2, and two timelines 2030 and 2090 were used in the models. As a discussed A1B scenario predicts a world with balance in all technological and energy sources, (homogeneous representation of future climate with temperature increase) and the A2 scenario predicts a heterogeneous world in coming decades with higher increases in temperature (around 5.5°C). We did not use scenarios considering minimum changes because studies show that since 2000, the rate of CO₂ increase is already larger than what was estimated by Special Report on Emissions Scenarios (SRES projections (Beaumont et al. 2008; Canadell et al. 2007). It has also been shown that climate forcing due to CO₂ increase is happening sooner and is stronger than expected by SRES (Raupach et al. 2007). For all three models, the same methods described in the methodology section of Chapter 7 were used to model the potential distributions of the three target species under climate change, and consensus models were prepared according to the same methodology described in Chapter 7. The new climate datasets were transferred into MaxEnt and the same variables used in the current scenario were applied in the model. For the Multi-Model Framework, the new datasets were transferred into R and the same variables selected for current scenarios by random forest were implemented in the Multi-Model for the climate change scenarios. Also for CLIMEX future scenarios, the same parameter values produced for the current scenario were applied in the model to project climate suitability for target species in future decades. The procedure which was followed for making the consensus model based on future climate scenarios for each target species is shown in Figure 8.3. Also, to compare the range of the target species in current and future climate scenarios, the number of pixels which were classified as suitable were calculated (Table 8.2).
Figure 8.3. The diagram shows the process of building consensus model for each target species. The same variables used in each model for present scenarios were applied to two scenarios A1B and A2 in two timelines for 2030 and 2080. To build the consensus, the top-five performing models were selected from Multi-Model to combine with CLIMEX and MaxEnt. The process was repeated for all three species.
8.5 Results

8.5.1 Psa

Projected changes in climate suitability for Psa by the consensus model are shown in Figures 8.4 to 8.7. For Psa, in 2030, the change in both scenarios shows smaller areas suitable for Psa establishment in Europe and Central Asia, Northern China and Japan (Figures 8.4 and 8.6). In eastern USA the number of models which agree on suitability of this area increases and Florida is projected as suitable by more than four models. In the southern hemisphere, suitability extends further northwards where parts of Brazil and Uruguay become suitable and southern Argentina becomes unsuitable. A similar pattern is observed in 2090 but larger areas in Sweden, Finland and Norway become suitable for Psa establishment. The changes in this part of Europe are biologically-plausible since these countries have a colder climate but by an increase of temperature (up to 5.5°C in 2090) in 2090 in both scenarios, Psa will have more chance to establish in these areas. In South Africa, the areas around Cape Town projected as suitable by the current climate become unsuitable under climate change. This may suggest that this area is already in its marginal climate for Psa establishment, which with an increase in temperature in future decades it may become too hot for Psa establishment (Figures 8.5 and 8.7).

In New Zealand, the results show that in general, the South Island will not be suitable for Psa establishment. Instead, similar to the current scenario, the most suitable areas would be in North Island with agreement of mostly five models for this area. Noticeably, in scenario A2, northern parts of the North Island are projected as suitable by all seven models (Figures 8.6 and 8.7).

8.5.2 Dwarf Bunt

For dwarf bunt, as expected, increasing temperature toward the end of the century make large areas in Russia and North America (polar zone) suitable for dwarf bunt establishment in both scenarios and timelines (Figures 8.8 to 8.11). Also re is an increase in the number of models that agree on suitable areas in Europe. In the southern hemisphere, the main change is visible in Australia where the south east
becomes unsuitable approaching 2090 in both scenarios. Similar to current scenarios, the highest agreement is visible in Europe where mainly seven or six models agree. These changes are biologically plausible as cold stress will not be a constraint for dwarf bunt establishment in these areas. In general, compared with Psa, dwarf bunt doesn’t show high sensitivity to the changes. It seems logical as based on global circulation models, the most dramatic climate change happens in tropical areas while dwarf bunt is recorded in cold areas and therefore Europe is less affected compare to other parts of the world (Figure 8.8 to 8.11).

In New Zealand, the suitable areas for dwarf bunt extend more toward the south of South Island and in 2030, both scenarios show an increase in the number of models which agree on the suitability of central Otago compared with current climate. These changes are biologically reasonable as these areas get warmer to a level where dwarf bunt is able to establish. However, the suitability in the North Island reaches a minimum in 2090, where only four models in the climate change scenarios indicated only a very small area around Palmerston North as suitable. With a temperature increase in extreme scenarios (up to 5.5°C), the North Island which has a warmer climate would be mainly too warm for dwarf bunt establishment (Figures 8.8 to 8.11).

### 8.5.3 Guava rust

For guava rust, the suitable areas mainly remain in neotropic ecozone. However, it is noticeable that suitable areas increase toward more southern regions of the Southern hemisphere. The suitability also extends toward areas such as Mexico, USA and central Africa. The regions where suitability increases significantly southward are in South Africa, Australia, Argentina and New Zealand (Figures 8.12 to 8.15).

In New Zealand, the change in suitability is quite dramatic where almost all New Zealand becomes suitable, especially where all seven models across both scenarios and timelines agree on the suitability of the North Island. In the South Island which is projected as unsuitable for guava rust establishment, based on current climate, in future scenarios, coastal areas become suitable (Figure 8.12 to 8.15).
Figure 8.4. The potential climate suitability for Psa based on Scenario A1B in 2030.
Figure 8.5. The potential climate suitability for Psa based on Scenario A1B in 2090.
Figure 8.6. The potential climate suitability for Psa based on Scenario A2 in 2030.
Figure 8.7. The potential climate suitability for Psa based on Scenario A2 in 2090.
Figure 8.8. The potential climate suitability for dwarf bunt based on Scenario A1B in 2030.
Figure 8.9. The potential climate suitability for dwarf bunt based on Scenario A1B in 2090.
Figure 8.10. The potential climate suitability for dwarf bunt based on Scenario A2 in 2030.
Figure 8.11. The potential climate suitability for dwarf bunt based on Scenario A2 in 2090.
Figure 8.12. The potential climate suitability for guava rust based on Scenario A1B in 2030.
Figure 8.13. The potential climate suitability for guava rust based on Scenario A1B in 2090.
Figure 8.14. The potential climate suitability for guava rust based on Scenario A2 in 2030.
Figure 8.15. The potential climate suitability for guava rust based on Scenario A2 in 2090.
Table 8.2. Percentage change in climate suitability (range expansion or range contraction) of three target species under two future climate scenarios.

<table>
<thead>
<tr>
<th>Species</th>
<th>Scenario A1B</th>
<th></th>
<th>Scenario A2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2030</td>
<td>2090</td>
<td>2030</td>
<td>2090</td>
</tr>
<tr>
<td>Psa</td>
<td>-69%</td>
<td>-67.5%</td>
<td>-78.7%</td>
<td>-65.6%</td>
</tr>
<tr>
<td>Dwarf but</td>
<td>-11.2%</td>
<td>-9.2%</td>
<td>-9.2%</td>
<td>-26.6%</td>
</tr>
<tr>
<td>Guava rust</td>
<td>152.9%</td>
<td>186.3%</td>
<td>144.3%</td>
<td>159.9%</td>
</tr>
</tbody>
</table>
8.6 Discussion

Model projections under future climate scenarios across time showed an increase in climate suitability for guava rust and decrease for Psa and dwarf bunt. Range changes based on future projections are more consistent for dwarf bunt which does not seem very sensitive to climate change. Buisson et al. (2010) in a study on French stream fish species have already shown that for species with a large geographic range, change is less extreme compared to other species (dwarf bunt is already widespread in Asia, Europe, North and South America and Africa). In contrast, suitable climates for guava rust establishment increase significantly in the Southern Hemisphere and especially New Zealand. Similar changes have also been documented by Bebber et al. (2013), where they showed that in a warming world, pathogens move toward the poles. For Psa, future projections show that, similar to the current climate scenario, the North Island remains highly suitable with all models agreeing on this suitability. However, the results are debatable for central parts of the South Island and Westland. As the South Island is generally colder than the North Island, an increase in the temperature in coming decades one would expect the South Island to become more suitable for Psa establishment. The prediction that the South Island does not become suitable in future projections does not necessarily imply that temperature would be higher than species tolerance in this area as if so the North Island would not be suitable either. This uncertainty may arise from different model performance, which is either related to prevalence or species physiological properties or existing microclimates in New Zealand (Buisson et al. 2010). For example because of high reliance of the Psa on moisture, because eastern parts of the South Island get drier by temperature increase in future, the condition may not be suitable for Psa establishment (at the same time it is expected that highlands and the western parts get more rain) (Mullan et al. 2008). Undoubtedly, identification of the cause of such difference from what is expected would be very helpful for future studies. In South Africa, the areas around Cape Town which were projected as suitable by current climate scenario are no longer suitable. This result may suggest that this area is already a marginal climate for Psa establishment and with an increase in temperature in future decades, this area may become too hot for Psa establishment.

It should be noted that because the consensus model is a combination of six correlative models and one mechanistic model, the projection of the correlative models will strongly
affect the consensus model results. This becomes important when we know that the correlative models projections rely on certain relationships among variables which may change in the future (Robertson et al. 2003; Prentice et al. 1992). It has also been suggested that for that reason mechanistic models will provide more robust results compared to correlative models under future climate scenarios (Robertson et al. 2003). Despite their usefulness, mechanistic models are hard to build, time-consuming, and they require a good knowledge and background information about target species. CLIMEX as a fitted semi-mechanistic model is a good compromise but even then the functions on which it is based need to remain the same over time and space for its projections to be valid. Additionally, it is important to know that patterns of rainfall and snowfall in future scenarios are less certain compared to temperature and certain projections show that the precipitation patterns vary based on season and region. For example, although the intensity of rainfall is projected to increase on average, in some areas the annual average rainfall increases and in others it decreases (Christensen et al. 2007; Meehl et al. 2007). Therefore, variation in rainfall patterns could have different effects over a regional scale and if available, it is ideal to use future climate data that are downscaled regionally. In such data, the variations in regional topography which affect the rainfall patterns are usually considered to downscale the future global climate data.

Sinclair et al. (2010), in a discussion of the usefulness of SDMs and their role for gaining insight about the magnitude of climate-change, state that allocating scarce funding to large projects in this field should be done only if these models can give us better insights compared to basic ecological theories. Conversely, Pearson et al. (2003) argue that complex natural systems may limit the ability of a model to give robust projection, but that they are very useful to give a preliminary approximation and insight about the climate change effect in the future. Additionally, because of the importance of spatial scale, any interpretation should be carried out paying attention to existing limitations. For example, the relation of variables may change as the spatial changes which may affect the model outputs (Heikkinen et al. 2006). Additionally, the usefulness of future projections will always depend on the questions asked and the methods used (Araújo et al. 2005)

In general, it is difficult to predict the effect of global warming on species as currently the focus is on changes in one or two parameters. Although it is not hard to exactly mimic the
changes in temperature in coming decades using a GCM, simulating rainfall changes is more sensitive and likely to be temporally and spatially intermittent (Hughes 2003). Furthermore, it is necessary that models incorporate the interaction of different components of ecosystems that include the changes and adaptations that may happen with plants and/or pathogens, and changes to land-use and agricultural practices (Chakraborty 2005). Davies et al. (1998) also warn that predictions based on bioclimatic models may be misleading because we do not consider that changes in distribution of one species may influence that of other organisms (for example beneficial or antagonistic microbe populations). Additionally, the importance of using long-term data has been highlighted by Jeger et al. (2008) who suggests that model limitations can be solved by using long-term data sets on the occurrence of pathogens. Because it is expected that climatic models and emission scenarios might well diverge in the current century, it is better to include several GCMs for long-term projections to decrease the uncertainty involved using species distribution models under climate change scenarios (Beaumont et al. 2008; Meehl et al. 2007; Stott et al. 2006; Prentice et al. 1991).

Finally, it should be noted that while we produce projections for future events they may be incorrect as they haven’t happened yet, but such projections can be useful if the future projections are interpreted in their appropriate context and according to model limitations.

8.7 Conclusion

It should be noted that agreement among the models and their projection of the desired phenomena may sometimes happen by chance (Araújo et al. 2005). For this reason, it has been highlighted that even when projections are totally consistent that does not necessarily prove the model; it only means that we cannot disprove the output. When assessing the impact of climate change on species climate suitability, it would be helpful to specify the sources of uncertainties to help the managers and decision-makers to form their strategies in full awareness of projection reliability. A consensus model may go some way to fulfilling that requirement. In addition, using different scenarios may reveal the sensitivity of projections and can improve our ability to better communicate with the public and authorities who have a strategic role in biosecurity and integrated pest management.
Chapter 9
General discussion

“An ounce of prevention is worth a pound of cure.”  

-Benjamin Franklin

9.1 Key findings

Alien invasive species have been recognised as one of the main reasons for economic loss and bio-diversity change, which can lead to serious impact on a country’s human and animal health, food security and cause environmental impact (Ficetola et al. 2007). The availability of precise climate data and increased processing power of computers has allowed the effective use of species distribution models (SDMs) to assess the risk of establishment of invasive species in novel areas. Such technological change has encouraged and enabled many scientists to conduct studies to investigate methods of species distribution modelling which can assist management of these species and thereby may indicate ways to mitigate their undesirable impacts.

When an invasive species establishes, the eradication and elimination of the species is often costly (Myers et al. 2000). Control efforts are much harder if the establishment occurs over large areas or little information is known about the population biology of the species (Simberloff 2003). Therefore, to reduce the future risk from invasive species, it is important to assess, and learn from, their present and potential global distributions (Leung et al. 2002).

Many models can be used to predict the potential for establishment of an invader based on initial knowledge about its environmental needs (Richardson and Thuiller 2007; Thuiller et al. 2005). Since the superiority of any single model is in doubt and its true efficacy is difficult to fully assess, one solution is to explore the way to benefit from different model output. Accordingly, the overall aim of this research was to predict the climate suitability of the target species by three main modelling approaches, to evaluate their projections, and investigate a method to combine their outputs. With this aim in mind, four specific objectives were investigated.
9.2 Identify the current and potential global and national distributions of each of the target species using different modelling approaches.

The first objective was to determine the strengths and weaknesses of different models to identify the current and potential global and national distributions of each of the target species. The climate suitability of the three target species by CLIMEX, MaxEnt and a correlative Multi-Model Framework was investigated in Chapters 3, 4 and 5.

9.2.1 Projection of Psa global distribution

For Psa, in Chapter 3, the projections of two models (CLIMEX and Multi-Model) showed the New Zealand climate very suitable for Psa establishment. This projection was in accordance with the widespread distribution of Psa in New Zealand where 81% of kiwifruit orchards were affected by Psa by 1 May 2014 (www.kvh.org.nz/vdb/document/453). In contrast, MaxEnt which has been extensively used in different ecological studies in recent years, projected low climate suitability for Psa in New Zealand. Given that the same data were used in the Multi-Model Framework and MaxEnt, the reason why MaxEnt did not project high suitability for New Zealand may lie in the different model algorithms. MaxEnt as a presence-only model, it only accounts for what is known about the species distribution and relies highly on presence data (Phillips et al. 2009). The high reliance of presence-only models on presence data can be a disadvantage when the current presence points of the target species may not be a good representation of the species full environmental requirements because the species might be still spreading and has not yet reached its full range. For example, further analysis using principal component analysis (PCA) on the existing presence data of Psa in Chapter 3 suggested that Psa reports from New Zealand appear environmentally distinct from other presence points. In such a case, if New Zealand presence points were not used in the modelling process, some presence-only models (such as MaxEnt) won’t be able to detect locations, as in New Zealand, where the environmental conditions are outside its known environmental range. Unlike MaxEnt, other correlative models (presence-absence) used in this study were able to successfully project New Zealand as suitable. This has been confirmed in a recent study on the genomes of Psa strains from Italy, New Zealand and Chile, where it has been reported that the Psa strains from New Zealand can be distinguished from the other populations and appear to belong to a different clone (Butler et al. 2013).
Since the presence data in New Zealand were not used for model calibration or fitting of any of the approaches used in this study, this apparently caused MaxEnt to project low suitability for New Zealand. This result supports previous findings by Peterson et al. (2007) who showed that presence-only models such as MaxEnt are more sensitive to the input and occurrence data. Also it's been reported that MaxEnt model is less transferable to different times and locations compared to other models because it is highly sensitive to bias and gaps in the input data (Townsend Peterson et al. 2007). In fact, in a comparative study using different presence-only models, it has been shown that presence-only models predict more precisely for the species that have relatively narrow niches (Tsoar et al. 2007). Furthermore, to our knowledge, the MaxEnt model has mostly been used to project the spread of invasive species at a local scale to project the climate suitability of nearby areas for the invader (Radosavljevic et al. 2013; Fuller et al. 2008; Lamb et al. 2008; Pawar et al. 2007; Pearson et al. 2007). While Giovanelli et al. (2010) reported that MaxEnt performed more consistently when the calibration area and environmental space increased, their study only focused on the changes in environmental space in two countries. A recent study by Elith et al. (2013) used MAXENT to predict the climate suitability of guava rust at a continental scale. Because projection into novel environmental areas is inevitable while using global data (Kim et al. 2009), more studies are needed to test how MaxEnt performs in large scale studies.

Unlike MaxEnt, the exclusion of New Zealand data did not affect presence-pseudo-absence models used in the Multi-Model Framework. The reason can be associated with the fact that the presence-absence models learn from presence and pseudo-absence data. Actually, when true absence data are available, it has been recommended to use a presence-absence model for prediction of invasive species (Yackulic et al. 2013). Although actual absence data were not available for the target species studied here (which is a common issue in ecological studies), the selection of pseudo-absences using a three-step method (Senay et al. 2013) to define an appropriate background buffer for pseudo-absence selection, improves the projection of presence-absence models (Senay et al. 2013). While Psa presence points for New Zealand were not used in the calibration (fitting) of any of models, I recommend that because presence points in New Zealand may contain new information that can affect the model output, they should be included in future studies using (especially) presence-only models, if the purpose of the modelling is to inform eradication or monitoring efforts. Apart from
differences discussed in Chapter 3, all models agree on high climatic suitability of some countries where kiwifruit is commercially grown such as the USA, Greece and Iran. Currently there is no report of Psa from these countries. This result could provide an alert for authorities in these countries since this pathogen has spread rapidly throughout Europe in recent years (Vanneste et al. 2012b; Abelleira et al. 2012; Bastas et al. 2012).

9.2.2 Projection of Dwarf bunt global distribution

In Chapter 4, the projection of all three model approaches showed that the European climate was highly suitable for dwarf bunt establishment. The best model (QDA) in the Multi-Model projected extensive areas in the Northern hemisphere as highly suitable for dwarf bunt establishment. These areas included Russia, all of Europe, vast areas in Northern USA and Canada as. Large areas projected as suitable can be a sign of over-fitting of the model or it may indicate that dwarf bunt hasn’t yet had a chance to reach and establish in these areas. The latter premise is not tenable since dwarf bunt has been a well-known widespread plant disease for almost a century and there are known occurrence from countries such as Russia, USA and Canada but they are mainly limited to wheat-growing areas in these countries. All three modelling approaches predicted that it is unlikely the pathogen will establish in India and central and western parts of China. The projections with respect to India were largely in accordance with current information, as, to date there is no report of dwarf bunt from India. However, there was discrepancy among the three modelling approaches regarding the projection for China. Unlike the models within the Multi-Model Framework, MaxEnt and CLIMEX projected the possibility of dwarf bunt establishment in central China. The projection of MaxEnt and CLIMEX in China was similar to the recent modelling study result by Jiao et al. (2013) where factors such as temperature, snow cover, humidity and teliospore germination data were used in a geo-phytopahological model. Their risk index showed 25% chance of dwarf bunt establishment in central China and high risk of establishment in the western provinces such as Sichuan and Lhasa. Regarding the suitability of China for dwarf bunt results differ over a number of previous studies (Jia et al. 2013; Peterson et al. 2009; Zhang et al. 1995; Trione et al. 1986), probably also the result of different predictor variables used in these studies. It is noticeable that dwarf bunt has been reported mainly from areas with snow coverage in winter, but some other important factors such as precipitation have been ignored. I believe, while using variables such as snow coverage that may indirectly
affect a species establishment may improve the model outputs but ignoring other variables that can directly affect the establishment of the species (such as precipitation or temperature), may result in under-prediction or over-prediction.

Clearly, as the same climate data and presence data were used in the correlative models, MaxEnt and multi-model, such differences in projection can arise from different variables selected by these two models. The important variables selected for MaxEnt were mainly variables related to temperature and in contrast, the process of variable selection in the Multi-Model selected two variables related to precipitation and one variable related to temperature. Apart from the differences in projection of the three modelling approaches, their similar pattern highlights the major suitable areas at a continental level.

In New Zealand, the results of all three modelling approaches indicated that the South Island seemed more suitable for dwarf bunt establishment with central Otago and around Lake Wanaka projected more suitable which was similar to predictions from previous studies such as that by (Kim et al. 2009). Although wheat and barley are not grown in central Otago, the growth index parameters showed that if enough propagules (20,000 spores/ gram of wheat seed) (Goates and Peterson 1999) arrive in this area and wheat and barley are grown, then the pathogen could establish, even though, it is unlikely that such an amount of spores would arrive in an area by chance (Kim et al. 2009). In addition, Purdy et al. (1963) suggested that under natural conditions, the chance of dwarf bunt establishment on grasses (alternative hosts) is very low. In contrast, MaxEnt projected almost all of New Zealand as unsuitable and assigns very low suitability to Central Otago (but still higher that other parts). When there is such discrepancy, principal component analysis (PCA) can help. In this research PCA showed that there are similarities between some global dwarf bunt presence points and some locations in New Zealand based on a direct climate data comparison in environmental space. Such a result can add confidence regarding the suitability of New Zealand for dwarf bunt establishment as projected by CLIMEX and the Multi-model. Additionally, by performing the PCA on presence data, which can show the environmental similarity of presence points, some presence points appeared as outliers and very different from the rest. Such a result suggests that PCA can be used as a method to explore uncertainties underlying presence data and to investigate the effect of removing the outliers on model output.
9.2.3 Projection of guava rust global distribution

Modelling the potential distribution of the third target species, guava rust is reported in Chapter 5. The results of this research are of particular importance to New Zealand. Guava rust has established in Australia in recent years as predicted by previous modelling studies (Magarey et al. 2007; Haines et al. 2006; Booth et al. 2000). The establishment of the guava rust complex in Australia raises a concern about the potential establishment of this species in New Zealand, particularly because previous studies have found that some parts of the North Island might be suitable for guava rust establishment (Kriticos et al. 2008; Kriticos et al. 2003). However, as was shown for Psa and dwarf bunt, one cannot rely on results from a single model, especially when the species is still invading new areas. In this study, the projection of the three modelling approaches for guava rust has a similar pattern. This inference is further supported because all three models agreed on suitability of South Africa, Coastal areas in the Eastern Cape of South Africa where guava rust has been recently reported (Roux et al. 2013), were projected highly suitable for guava rust establishment. So far, there has been only one report from South Africa, however, if effective management strategies are not implemented, more reports of guava rust establishment and related damage would be expected.

For New Zealand, the suitability of the North Island was highlighted by all three modelling approaches. That result is important since there are many indigenous plants that are among the possible hosts of guava rust (MAFReport 2011; Poole and Adams 1994). The PCA results additionally confirmed the similarity of New Zealand climate data to that of areas globally where guava rust is present, which can increase the confidence about the suitability of New Zealand climate for guava rust establishment. To implement an effective guava rust control strategy for New Zealand in the case of an incursion, some investigation is necessary to determine if the containment strategies implemented in Australia are one of the reasons that has resulted in the mitigation of the expected effect of guava rust or if previous studies simply over-estimated such an effect.

The results of research reported in Chapters 3, 4, and 5 indicated that the New Zealand climate in some regions is suitable for establishment of the three target species. The projected extent of the area in which Psa can establish in New Zealand was larger compared
with other species studied here. Dwarf bunt is expected to be more likely to establish in the South Island and guava rust, the North Island. Such results suggest that policies and strategies need to be designed to prevent the arrival of dwarf bunt and guava rust or, alternatively, plan how to respond to their possible incursion. In general, the pattern of projection for three species was similar at the global level. At regional levels such as in New Zealand, the differences were more obvious. Also notable is that the MaxEnt model showed low probability of target species establishment globally compared to the other two models.

9.3 Sensitivity of CLIMEX model

An Ounce of Prevention, a Pound of Uncertainty

The second objective of this research was to perform a sensitivity analysis of the process-based model (CLIMEX) using one of the target species modelled by CLIMEX as a process-based model. Since parameters values are chosen using a subjective process there is always a chance that the modeller might be wrong or unsure about the accuracy or even the reality of the value of the parameters used in the fitting process. The question is, what effect does error in the important parameters, have on model output and interpretation? Such an issue is already evident in three different studies on citrus black spot using CLIMEX where the parameter values fitted in the CLIMEX model are different in each study (Roberts et al. 2013; Yonow et al. 2013; Paul et al. 2005). In Chapter 6, the CLIMEX model already developed for dwarf bunt in Chapter 4, was used to test the sensitivity of the parameter values. The results of the analysis showed that the CLIMEX model was sensitive to specific temperature and soil moisture parameter values where relatively small changes in the parameter values can lead to significant changes in the projected suitability of some regions. More importantly, some areas were more sensitive to these changes suggesting such a result should be further explored to determine why. The results can be more alarming, especially when relatively small changes in the parameter values cause obvious changes in prediction at a global level. The interaction of sensitive parameters was also investigated in this study, which has never been considered in previous studies. The model showed higher sensitivity when the interaction of sensitive parameters was investigated. Although, performing sensitivity analysis may be very time-consuming, conducting such analysis is necessary if one is testing the robustness of the model, particularly if the parameters are subjectively chosen. A sensitivity analysis can also increase our understanding about the relationship between input
and output variables, and not only detect sensitive parameters but also locations (in case of this study) and enhance communication between the modeller and decision-makers by offering more credible and understandable outputs. In this thesis, such an analysis reinforced the observation that there is no perfect model.

9.4 Building a consensus model

“Believe those who are seeking the truth. Doubt those who find it”

-Andre Gide

“Models are not like religion. You can have more than one… and you don’t have to believe them”

Daniel Pauly and Villy Christensen

The third objective of this thesis was to explore an appropriate method to combine the prediction of different models into some sort of ensemble. In all previous research, the different models and modelling approaches showed different output for the target species. In fact, models are different approximations of reality. Most often, the selection of a candidate model in a research study is usually based on factors such as the availability and resolution of the predictor variables, availability of model software and previous knowledge of the model by the practitioner. Differences in model output can occur for different reasons. Different models give different output because of the different algorithms and theory used by the models and the assumptions made in the process of model development. For example, Multi-Model variable selection is carried out using random forests compared with MaxEnt (pairwise Pearson correlation test) or the CLIMEX model where temperature and soil moisture are the main variables. Some models perform better with presence-only data such as MaxEnt and some require presence/absence data to make good predictions such as SVM and Neural Networks. In addition, in different models, species observations are used in different ways to train the models. For example, in CLIMEX, parameters are calibrated using visual assessment until the model output matches the observed data whereas for MaxEnt and Multi-Model the model performance is evaluated statistically using accuracy criteria such as AUC. Even if the same environmental data are used as input for the correlative models MaxEnt and Multi-Model, the model types gave different projections. Although it has been suggested that one way to decrease such uncertainty is to compare the results of the different models (Webber et al. 2011; Townsend Peterson et al. 2007; Robertson et al. 2003), different visualisations based on model output may not be helpful especially when the aim is to communicate with authorities and decision makers. If one can assume that each modelling
approach can offer some information that is of use, it would be very useful to show the authorities the areas where all or most models agree that the target species may establish. By presenting the results of different models in a single map, it is easier to communicate with interested parties. In an ensemble or consensus model, the strength of different SDMs is combined and the weaknesses of single models are minimized (Araujo et al. 2007). As well, one would mostly expect that the ensemble would be useful when data are very noisy (Lawrence et al. 1998). Stohlgren et al. (2010) suggested that a consensus model approach can be very useful especially in the case of invasive species that are still spreading and have not reached all suitable places because in such cases it is difficult to determine the relationship between the species and environment by only one model (Stohlgren et al. 2010). The consensus model presented in this study combines the results of the models used in Chapters 3, 4 and 5 for the target species. The hypothesis is that different model types used as building blocks of a consensus model can result in better overall understanding of the likelihood of target species establishment.

In general the consensus models did give improved fit to the observed data and there is some evidence that they gave better projections (Gallien et al. 2012; Marmion et al. 2009). In Chapter 3 it was shown that incorrect inferences may have been made if only the MaxEnt model was used to project for Psa and the consensus model showed that six models out of seven agreed on suitability of New Zealand for Psa establishment. Also in the consensus model for guava rust, all seven models agreed on the suitability of Northland in New Zealand and gives confidence about the high probability of the pathogen establishment in this area. For dwarf bunt, where previous modelling efforts show controversial results (for example) in China, the consensus model projects small areas as suitable. These areas are around the Huanghe River, Yangzi River and Huaihe River Valley that are importantly also the main wheat growing areas in China (Jia et al. 2013).

9.5 The consequences of climate change on the target species

The fourth and the last objective of this research was to use the improved methods developed in this thesis to investigate the effect of climate change on climate suitability of the target species at the regional and global scale (Chapter 8). In general, the results presented in Chapter 8 confirmed the findings by Bebber et al. (2013) which indicate that pest and plant
pathogens move pole-ward in response to climate change in future decades (Bebber et al. 2013). Guava rust climate suitability clearly increases southwards where South America, Africa, Australia and New Zealand show increased suitable areas. While current guava rust occurrences are from South America (mainly central and northern regions) the future climate change scenarios show increasing suitable areas in central parts of South America (its native range). In addition, in Africa, the suitable areas in South Africa extend to the whole country compared to the current scenario where suitable areas are mainly limited to the eastern parts of South Africa. In New Zealand, where the models for the current situation agree on the suitability of the North Island, in the future, under climate change, suitability will extend to the South Island. For the current climate scenario all seven models agreed on small areas in Northland, New Zealand as suitable. In both scenarios A1B and A2, almost the whole of the North Island are projected suitable for guava rust establishment by all seven models. This result can alert authorities to potential threat in future decades so that they can put in place appropriate biosecurity policy and strategies.

In general, the climate suitability of dwarf bunt moves northwards in the Northern Hemisphere where more suitable areas in the warming polar zone are predicted. Based on the climate-change scenarios, the changes in climate suitability for dwarf bunt are less dramatic in New Zealand. The South Island remains the main area of concern regarding dwarf bunt establishment in future scenarios. In 2090, the main change in climate suitability for dwarf bunt concerns the number of model agreements in the South Island, where four models out of seven projects the suitability of the South Island while in general this area remains suitable.

The results of climate-change scenarios for Psa show some inconsistency for New Zealand across time. Although, similar to the current scenario, the North Island remains suitable in both scenarios across both timelines (2030 and 2090), the South Island remains unsuitable. One explanation is that unlike the North Island, the South Island is expected to experience an increase in precipitation and Christensen et al. (2007) highlight that the pattern of precipitation in future climate scenarios are less certain than temperature patterns because they vary based on region and season (Christensen et al. 2007; Mullan et al. 2008). Such regional and seasonal changes are not reflected in future climate scenarios and only average rainfall intensity is included in future scenarios. The variation in rainfall can have different
implication at regional scales such as New Zealand where topography has a significant influence on rainfall patterns which can consequently affect the model outputs. Therefore, I believe that because five out of the seven models on which the consensus models is based, are selected from the Multi-Model Framework and because, in the case of PsA, the selected variables by the multi-model are mainly precipitation variables (two out of three variable), are the reasons for the resulting inconsistency of projection for New Zealand based on future climate scenarios. I recommend, for precipitation data, the downscaled future climate scenarios should be used (if available). Because in down-scaling future climate scenarios, local records are used to calculate the precipitation values (where regional topology patterns are considered), giving more accurate precipitation data than those available global future climate scenarios. To my knowledge such data were not available in the form of bioclimatic variables at the time that this research was conducted.

9.6 Summary
The use of bio-climatic models to evaluate the effect of climate change on the distribution of the species has been criticized (Lawton 2000; Davis et al. 1998; Woodward and Beerling 1997). For example, Davis et al. (1998) discuss the complexity of natural systems and point out that biotic factors that play an important role in the distribution of species might be different in the future compared to the current time. They conclude that dispersal and interactions as important factors of population dynamics must be considered in predictions. Lawton et al. (2000) point out that some factors such as major rivers, mountains and human interaction with the landscape may greatly or completely stop species from arriving at the predicted locations. Clearly such criticisms should be acknowledged but I believe that often not much is known about such interactions and taking the first step to determine climate suitability highlights possibilities and increases our knowledge that can be refined as our knowledge about individual species increases. I also believe that using known or well researched biotic interactions in modelling species distributions should not just be limited to climate change studies. If possible, all such factors should be included in any study where the objective is to predict the current or future distribution of species. Despite clear limitations, bio-climatic models can provide a good approximation of potential climate change effects on species distribution that can be useful to policy makers as long as the limitations of the models are acknowledged (Heikkinen et al. 2006).
9.7 Bioclimatic models: assumptions and limitations

“Essentially, all models are wrong, but some are useful”
– George E. P. Box

In any biological study, the “true model” can have unlimited dimensions and for this reason we cannot fully reflect full reality with limited data. In reality, the objective is to find a model that is a good approximation of available data. Clearly any model whether statistical, mechanistic or a mental model is not “truth” and as a simplification of the reality it cannot represent full reality (Taub 1993).

Every modelling approach has advantages and disadvantages and models can be categorized based on their usefulness. A range of factors should be considered when inferences are made based on the results of the SDMs. It should be always acknowledged that errors are an inevitable part of any scientific study because natural systems are not static and it is not possible to account for all factors that affect a species distribution (Araújo et al. 2005). In other words, bioclimatic models assume that a species is in equilibrium with its environment (they have fully filled their environmental niche) and they may fail to correctly project the distribution of a species invading a new area or more particularly a species that experiences a range shift in response to climate change. In cases such as climate change and invasive species this equilibrium assumption may not serve the purpose because many factors such as biotic interaction, genetic variability and evolutionary changes (Elith et al. 2009). Another important issue that should be noted is extrapolation where predictions are made in new environments. When models extrapolate, the results should be carefully evaluated because there are no records of target species available in training dataset to support the projection. Some methods have been suggested to minimize the risk of extrapolation such as estimating dispersal rate, weighting species data, models of dispersal including evolutionary change estimation in models and consensus model which can potentially reduce the differences among models and acknowledging prediction differences (Phillips et al. (2006); Pearson et al (2006); Araujo et al (2005).

Many studies on interspecies interactions have shown that functional changes in the presence or absence of other species may cause errors in bioclimatic modelling efforts (Davis et al. 1998; Crawley 1997). The effect of predation, symbiosis and competition on a species...
distribution has also been well documented. For example, in a study on three Drosophila species (flies) by Davis et al. (1998) it was found that temperature determined the outcome of competition, and the range size and abundance of the Drosophila species changed in presence of competition.

Additionally, climate change can affect the susceptibility of the host to infection caused by plant pathogens and host parasite systems. For example, Burdon et al. (2006) reported that significant changes in the host-pathogen interactions would happen over time especially for rust fungi. They report that for rust fungi in high altitudes or mountainous communities, the formation of teliospores will increase because the growing season is likely to increase in future decades. In addition, in an earlier study by Burdon and Chilvers (1974) on leaf pathogens such as *Aulographina eucalypti*, *Dichomera sp.*, *Septoria normae* attacking *Eucalyptus pauciflor* in southeast Australia, with an increase in altitude, a distinct decline was recorded in these leaf pathogens. Therefore, using bioclimatic models to make inferences based on their current distribution without considering such factors might be misleading. Despite this consideration, when bioclimatic models are used in large or macro-scale, it is likely climate has a major influence not only on the target species distribution, but also its interaction with other species (Pearson et al. 2003).

The general limitations and challenges of species distribution models have been well reviewed by Araujo and Guisan (2006). They highlighted the need for better sampling designs, improving parameterization strategies (absence estimation and type of variables used), improved model selection and predictor contribution, and improving model evaluation strategies. Some concerns such as sampling bias and design were beyond the scope of this study. However, it should be noted that unlike species of conservation interest, the target pathogens are economically very important and are monitored intensively such that the chance of not reporting the existing occurrences is very low. Additionally some concerns such as parameterization and model selection have been addressed in the Multi-Model Framework by using multiple performance criteria.

Several studies have already shown the sensitivity of niche models to sample size and the existing bias in presence data (Stockwell et al. 2002; Townsend Peterson and Cohoon 1999). The real distribution of a species is likely to be wider if the presence localities do not show the full potential of a pathogen (for example the where there are barriers to dispersal). On the
other hand, their distribution also might be narrower because of limited host distribution, competition and other biotic factors (Shaw et al. 2011). Stockwell and Peterson (2002) reported that model accuracy depends on the number of data points used in the modelling process. For example, they reported that the accuracy of machine learning models was 90% of the maximum with ten sample points and near maximum with 50 presence points. Moreover, it’s been reported that when the data/presence locations are taken from broad geographical areas, models will fit better because they may better represent the environmental conditions that limit the species distribution (Pearson et al. 2006). In every study that attempts to model the climate suitability of an invasive species, careful consideration should be given to those species that are still spreading or are in early stage of the invasion because that fact may result in under-estimation of the full potential range of the species (Václavík and Meentemeyer 2012). The reason is that available presence locations may not give a full indication of the actual potential distribution of target species, and as such modelling studies should be updated when new reports arrive. Moreover, it is important to note that the presence of a species may not be recorded for several reasons such as, 1) the difficulty detecting the species (that is an important consideration for plant pathogens due to their microscopic size), 2) evolutionary change, and, 3) the sudden appearance of new pathways aiding species spread.

9.7.1 Data and data sources

With increased understanding of the need of SDMs to formulate biosecurity plans for threatened species, the presence data for economic pests has been recorded and made available to agencies for more than a century. For example, the European and Mediterranean Plant Protection Organisation (EPPO) are providing such information as a software package (PQR) that can be freely downloaded and used. The GBIF and CABI websites are also two major sources for such information. Unfortunately, the numbers of presence locations for the target species in this study in these databases were either minimal or non-existent. Such shortage in availability of species occurrences for plant pathogens creates a challenge. The alternative is to gather such information from published journals and personal communications which can be very time-consuming (but also informative).
9.7.2 Predictor variables

The predictor variables used in this study are well-known variables in Worldclim that are derived from temperature and rainfall data that have been extensively used in modelling studies in recent years (Elith et al. 2013; Senay et al. 2013; Kriticos et al. 2012; Jarnevich and Stohlgren 2009). During the current study, a new set of bioclimatic variables (CliMond data) that includes radiation and soil moisture data were published online which can be used in future modelling studies. Collinearity and multi-collinearity which refers to the linear relationship between explanatory variables has been discussed as another source of uncertainty and methods have been suggested to deal with such issue (Dormann et al. 2013). Collinearity has less effects in machine learning methods but can affect parameter estimates more in statistical methods and lead to incorrect identification of key variables in statistical or correlative models (Elith et al. 2011; Ward 2007).

9.8 Specific limitations of the models in the current study

In addition to data quality, model type has been also reported to be one of the main sources of uncertainty of species distribution models (Dormann et al. 2008; Anderson et al. 2003). Here I try to weigh up the limitations of the models used in this research.

9.8.1.1 CLIMEX model

9.8.1.1.1 Subjectivity of parameter setting

Despite the advantages of the CLIMEX model mentioned in Chapter 2, there are some technical and general caveats associated with its use. In CLIMEX the distribution of the target species is modelled using physiological responses of the species to climate. However, the parameter setting process is purely subjective and therefore the user can influence the results greatly. For example, the effect of such subjectivity is apparent in a study on the climate suitability of citrus black spot conducted by three separate researchers where they came up with different parameter values for temperature, soil moisture and stress indices (Er et al. 2013; Yonow et al. 2013; Paul et al. 2005). Another issue which makes the comparison of similar studies conducted using CLIMEX difficult is the arbitrary categorization of the important Eco-climatic Index (EI) values used to map the outputs which may result in different visualization and interpretation. Lack of suitable and proper validation method is one of the other main disadvantages of the CLIMEX model because CLIMEX validation is
purely done though visual assessment by the user and it can be a source of high uncertainty in model output.

9.8.1.2 Growth and stress indices
Weekly and annual temperature is a component of the growth index used in CLIMEX. The direct effect of temperature on the development rate of species that have more than one generation per year (multi-voltine) is ignored. In such cases, the relationship between annual growth index and annual total population is not linear because of cumulative growth of the species (Sutherst 1999). This deficiency may become more apparent and have more effect on species such as plant pathogens that frequently have more than one generation per year. Also, only one period of each type of stress is assumed in CLIMEX which in some cases is not true as there are some locations where two favourable and two unfavourable conditions exist for each year (such as East Africa which is triggered by bimodal precipitation distributions (two seasons) (Sutherst et al. 2004).

9.8.1.3 Host plant
The distribution of many host-specific species is defined by the presence of their host and it is hard sometimes to separate the climate suitability of host plant and the species. For example in cases where the species climatic tolerance covers the full range of the host plant climate suitability there is no way to deduce the response of target species to limiting climatic factors (Sutherst et al. 2004). Nevertheless, it should be noted that for some organisms such as plant pathogens the availability of alternative host plants can affect and facilitate their establishment. Additionally, unlike correlative models, in CLIMEX, some important factors such as soil type and land-use that can affect the climate suitability of the target species cannot be incorporated (Kriticos et al. 2003).

9.8.1.2 MaxEnt Model
As discussed earlier, MaxEnt has been criticised because of its poor transferability to new regions compared to other models (Syfert et al. 2013; Townsend Peterson et al. 2007). With low transferability, the model projections into new areas and times cannot be trusted. In fact, transferability that can occur either spatially or temporally is a concern in habitat suitability modelling in general. Such models can also be affected by some factors such as sample selection bias (the areas that are sampled more intensively than others) (Phillips et al. 2009). The issue of sample bias of presence-only models has been investigated in different studies.
and the sensitivity of the presence-only models to such bias is well acknowledged, but still no clear method has been proposed to overcome such an issue and debate still continues in this regard (Fourcade et al. 2014; Merow et al. 2013; Yackulic et al. 2013; Phillips et al. 2009). In a recent study, after testing performance of five correction methods to account for sampling bias, Fourcade et al. (2014) concluded that the ability of methods to correct the sampling bias differed, as they were highly based on species, bias type and intensity. Also a study by Merow et al. (2013) offers a practical guide to MaxEnt and explain how some default settings can affect the MaxEnt model output. Merow et al. (2013) also reviewed the effects of factors such as feature selection, sampling bias, regularization and background data on MaxEnt output which have been mostly considered in this research.

9.8.1.3 Multi-Model Framework

One of the advantages of the Multi-Model Framework is the ability to test the sensitivity of the models using different performance criteria. Also, model performance is evaluated using bootstrap and cross validation which enhances the robustness of the Multi-Model Framework. Despite these advantages, the importance and contribution of the selected variables and their influence on the model output is not given in the framework. Additionally, in some cases where the performance criteria are not significantly different, the output maps of the models appear very different. Currently, in the Multi-Model Framework, model selection is performed based on ranking the performance criteria without testing whether the differences among them are significant. In such cases where the selection of the best model is based only on performance criteria ranking, which may result in incorrect inferences especially where model results are not ecologically plausible. For such reasons, the models should be supported biologically rather than just purely statistically as is often commonly done (Venette et al. 2010; Zalucki and Van Klinken 2006). Finally, it would be good to incorporate functions in the Multi-Model Framework to calculate log-likelihood which is a useful criterion and is common for model selection purposes.

9.9 Future work

Future research arising from the current study should address two main objectives. Firstly, the research should focus on the improvement of bioclimatic models to improve prediction and secondly, the efforts can then be directed to improve the prediction of plant pathogens using such models.
Correlative models have been criticized because unlike process models, they do not consider the physiology of the target species and their reliability under climate change is debated. However, mechanistic models cannot be applied because the information about the physiology of the target species is not always well-known. Therefore, one way to increase the robustness of predictions is to integrate mechanistic and correlative models. To clarify the uncertainty in prediction of species range and niche, process based models have been compared in this study as recommended by (Morin and Thuiller 2009). However, novel modelling methods also need to be explored to decrease uncertainty. Recently, the idea of the use of hybrid models in species distribution studies has been introduced, where the resulting model is based on the strengths of both mechanistic and correlative models (Dormann et al. 2012). In some of the methods for hybridising models proposed by Dormann et al. (2012), the output of one model can be used as an input of another model. For example, one can extract extra presence or absence points based on CLIMEX EI (e.g. EI> 30 as presence points and EI= 0 as absence points) values and use these presence and absence points in correlative models such as MaxEnt and the Multi-Model Framework.

Because soil moisture is a very important factor in the survival of the majority of plant pathogens, especially those that over-winter in soil, for future studies, I recommend using the new bioclim data which includes soil moisture variables (Kriticos et al. 2012; Hollaway et al. 1996). If relevant, as is often the case with plant pathogens, one should also explore the effect of taxonomic uncertainties of the pathogen on predictive performance of a SDM and the geographic range of the species (Elith et al. 2013). In this regard, two strains of Psa (virulent and low virulence) have been reported in the literature, and I recommend conducting studies to separately investigate the climate suitability of the two strains. Furthermore, the consensus map could be categorised into different areas indicating the suitability of locations for species establishment in conjunction with the presence or absence of host plants in the area. For example, the results can be shown as the areas where Psa can established and host plants are available, the areas which host species are available and Psa cannot establish and the areas that Psa can establish and host plants are not available.
9.10 Concluding remarks

- Eleven different models were assessed to investigate their prediction for three target species. The results showed that the model projections can vary even when the same variables and settings were used in the process (in case of Multi-Model Framework).

- Because, the target species are usually modelled with limited data, it is recommended to use the outputs of different methods to decrease the uncertainty in model’s projections. Such aims can be strengthened by using the different types of models available, such as correlative and mechanistic models, to benefit from the strength of each type to evaluate and test. In Chapters 3, 4 and 5, only the best-performing model was selected from the multi-model framework to compare results with MaxEnt and CLIMEX results. The results showed that all models agree on the suitability of New Zealand for Psa establishment (Chapter 3). The models also predicted that the chance of dwarf bunt establishment is high in central Otago in the South Island (Chapter 4). All three models agree on the high risk of guava rust establishment in the North Island, especially in the Northland area (Chapter 5). Noticeably, MaxEnt model predicted lower probabilities compared to other models.

Among models used in this study, the sensitivity analysis on CLIMEX parameters in Chapter 6 revealed how important it is to include the sensitivity analysis as an integral part of the modelling process. Because it is always possible that a mistake has been made in the subjective process of parameter setting, performing sensitivity analysis can provide useful information.

Despite the similarities among model projections, combining the results of seven models using the consensus model in Chapter 7 highlighted the areas with the highest agreement. Instead of merely relying on comparison of different models output that was done in Chapters 3, 4 and 5; the consensus model highlighted the areas that more than half of models agree on climate suitability of target species. In this way, the areas predicted as suitable by at least four models (or more) out of seven were presented for three target species giving some assessment of the uncertainty of model prediction.

Applying the same methods and variables as defined for the target species in current climate scenarios, the consensus model was used to investigate the effect of climate change on the
target species. The results showed that the range of all three species expanded towards the end of the century. The highest range expansion was observed for guava rust and the least was observed for dwarf bunt. The changes also confirmed a recent finding that pest and plant pathogens are moving pole-ward as the climate warms.

In conclusion, the results of this thesis confirm that models with different theoretical foundation may give dissimilar predictions, and it is difficult to determine conclusively whether one model is superior to others. I strongly recommend not relying on a single-model projection. If time and resources are available, an appropriate ensemble of models should be used to investigate climate suitability for plant pathogens.
References


Conners I L (1953). Bunt situation in winter wheat in Ontario.


232


Greer G, Saunders C (2012). The costs of Psa-V to the New Zealand kiwifruit industry and the wider community. Agribusiness and Economics Research Unit Report, Lincoln University, New Zealand: 75


Pertot I (2012). Climate change impact on plant pathogens and plant disease.


Appendices
Appendix A

Application of genetic algorithm function of CLIMEX

Figure A. 1. The results of genetic algorithm function in CLIMEX. The values calculated by the genetic algorithm included in the latest version of CLIMEX were different from what was fitted in iterative model.
Appendix A.2. AutoIt code use to automate the CLIMEX parameter fitting.

For $i = 1 To $CSV_Array[0]
    _GetMyResults($CSV_Array[$i])
Next

Func _GetMyResults($parameter, $index)
    ControlSend("Save As", ",", "[CLASS:Edit; INSTANCE:1]", "map_" & $index); saves the map by the name of "map_i"
    ControlSend("Save As", ",", "[CLASS:Edit; INSTANCE:1]", "table_" & $index); saves the table by the name of "table_i"
EndFunc

Run("C:\Program Files\CSIRO Entomology\Dymex\DxSim3.exe") ; runs the software
MouseClick("left", 830, 400, 1); clicks on welcoming window
Send("[ESC]“); closes welcoming window
WinWait('CLIMEX - Compare Locations (1 species)', ''); activates the main window
WinActivate('CLIMEX - Compare Locations (1 species)', ")
MouseClick('Left', 50, 241, 1); clicks on parameters window
Sleep(2000)
MouseClick('Left', 46, 169, 1); opens temperature Index
Sleep(2000)
MouseClick('Left', 48, 198, 2); click on temperature Index DV0 filed
Sleep(2000)
Send($parameter); changes parameter
Sleep(2000)
Send("[Enter]"); goes to the next filed to confirm entered number (1)
Sleep(2000)
Send("[ESC]“); gets out of fields
Sleep(2000)
MouseClick('Left', 471, 111, 1); closes the parameters window
Sleep(2000)
Send("^r"); Runs the model
Sleep(1000)
Sleep(35000) ;waits 27seconds for model to run; awaits for model to run fully
Send("^m")
Sleep(3000); 3 seconds shows the map

MouseClick('Left', 590, 241, 1)

Sleep(2000)

Send("[Alt]")

Send("m")

MouseClick('Left', 109, 329, 1); clicks on Export File

Sleep(2000)

Send("[BS]"); deletes the default name

ControlClick("Save As", "", "[CLASS:Button; INSTANCE:1]", ""); clicks on the empty field to type the name of file

Sleep(2000)

ControlSend("Save As", "", "[CLASS:Edit; INSTANCE:1]", "map_1"); saves the map by the name of "mapi"

Sleep(2000)

ControlClick("Save As", "", "[CLASS:Button; INSTANCE:2]", "")

Sleep(6000); waits 4 seconds to save the file

Send("[enter]"); justifies the dialog box showing the map successfully saved

Sleep("4000")

Send("[ALT]")

Sleep("1000")

Send("w"); activates the Window menu

Sleep("2000")

Send("2"); activates the Run window

Sleep("1000")

Send("^t"); creates the table

Sleep("2000")

Send("S"); opening standard table

sleep("2000")

MouseClick("left",459, 482,2)

sleep("1000")

send ("s"); sends the "save table" command

sleep("1000")

Send("[BS]"); deletes the default name

sleep("1000")
ControlClick("Save As", ",", [CLASS:Edit; INSTANCE:1], ""); clicks on the empty field to type the name of file
Sleep(1000)
ControlSend("Save As", ",", [CLASS:Edit; INSTANCE:1], "table_1"); saves the table by the name of "table_1"
Sleep(1000)
ControlClick("Save As", ",", [CLASS:Button; INSTANCE:2], "")
sleep("2000")
Send("{ALTDown}+{F4}"); closes the software
Send("{ALTUP}"); releases the ALT button
sleep("2000")
ControlClick("Dymex Simulato", ",", [CLASS:Button; INSTANCE:2], "")
EndFunc ;==>_GetMyResults
Appendix B

List of the 19 BIOCLIM variables used in this study

BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (mean of monthly (max temp - min temp))
BIO3 = Isothermality (BIO2/BIO7) (* 100)
BIO4 = Temperature Seasonality (standard deviation *100)
BIO5 = Max Temperature of the Warmest Month
BIO6 = Min Temperature of the Coldest Month
BIO7 = Annual Temperature Range (BIO5-BIO6)
BIO8 = Mean Temperature of the Wettest Quarter
BIO9 = Mean Temperature of the Driest Quarter
BIO10 = Mean Temperature of the Warmest Quarter
BIO11 = Mean Temperature of the Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of the Wettest Month
BIO14 = Precipitation of the Driest Month
BIO15 = Precipitation Seasonality (Coefficient of Variation)
BIO16 = Precipitation of the Wettest Quarter
BIO17 = Precipitation of the Driest Quarter
BIO18 = Precipitation of the Warmest Quarter
BIO19 = Precipitation of the Coldest Quarter
## Appendix C

### Potential climate suitability of kiwifruit bacterial canker

Table C.1. Psa presence points coordinates (after removal of duplicate points).

### Supplementary materials

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>146.8174</td>
<td>-38.2957</td>
<td>111.38</td>
<td>29.58</td>
<td>12.66108</td>
<td>41.45759</td>
</tr>
<tr>
<td>141.3468</td>
<td>43.06462</td>
<td>111.28</td>
<td>30.69</td>
<td>12.25039</td>
<td>42.48713</td>
</tr>
<tr>
<td>138</td>
<td>36</td>
<td>110.79</td>
<td>32.62</td>
<td>12.15639</td>
<td>45.67763</td>
</tr>
<tr>
<td>133.5</td>
<td>33.75</td>
<td>110.6625</td>
<td>29.9299</td>
<td>11.88299</td>
<td>44.28535</td>
</tr>
<tr>
<td>131</td>
<td>33</td>
<td>110.47</td>
<td>29.11</td>
<td>11.811</td>
<td>45.542</td>
</tr>
<tr>
<td>129.42</td>
<td>42.76</td>
<td>110.165</td>
<td>29.4</td>
<td>11.13719</td>
<td>45.37561</td>
</tr>
<tr>
<td>126.991</td>
<td>34.8679</td>
<td>109.79</td>
<td>30.29</td>
<td>11.119</td>
<td>44.547</td>
</tr>
<tr>
<td>124.79</td>
<td>40.71</td>
<td>109.69</td>
<td>28.26</td>
<td>10.84342</td>
<td>45.35085</td>
</tr>
<tr>
<td>123.42</td>
<td>41.83</td>
<td>109.32</td>
<td>34.15</td>
<td>9.01</td>
<td>42.03</td>
</tr>
<tr>
<td>121.42</td>
<td>29.28</td>
<td>109.03</td>
<td>34.34</td>
<td>8.173759</td>
<td>45.36737</td>
</tr>
<tr>
<td>120.46</td>
<td>33.195</td>
<td>108.6</td>
<td>34.1</td>
<td>7.851</td>
<td>44.948</td>
</tr>
<tr>
<td>118.35</td>
<td>35.1</td>
<td>108.22</td>
<td>34.16</td>
<td>7.79315</td>
<td>44.57865</td>
</tr>
<tr>
<td>117.94</td>
<td>28.45</td>
<td>106.74</td>
<td>31.86</td>
<td>5.283125</td>
<td>46.24749</td>
</tr>
<tr>
<td>117.2849</td>
<td>31.86115</td>
<td>106.55</td>
<td>29.56</td>
<td>1.52</td>
<td>44.08</td>
</tr>
<tr>
<td>117.05</td>
<td>30.52</td>
<td>106.2642</td>
<td>26.21513</td>
<td>-0.29958</td>
<td>44.70022</td>
</tr>
<tr>
<td>116.61</td>
<td>39.72</td>
<td>105.93</td>
<td>31.73</td>
<td>-0.3</td>
<td>45.9</td>
</tr>
<tr>
<td>116.52</td>
<td>31.73</td>
<td>105.59</td>
<td>30.53</td>
<td>-0.32</td>
<td>47.76</td>
</tr>
<tr>
<td>116.35</td>
<td>30.84</td>
<td>104.67</td>
<td>31.46</td>
<td>-4.22339</td>
<td>41.02428</td>
</tr>
<tr>
<td>116.14</td>
<td>39.74</td>
<td>104.0759</td>
<td>30.65165</td>
<td>-5.43333</td>
<td>43.48333</td>
</tr>
<tr>
<td>115.4</td>
<td>38.83</td>
<td>103.64</td>
<td>30.98</td>
<td>-6.06973</td>
<td>43.38855</td>
</tr>
<tr>
<td>114.48</td>
<td>29.6</td>
<td>103.59</td>
<td>31.47</td>
<td>-8.19</td>
<td>41.35</td>
</tr>
<tr>
<td>114.17</td>
<td>22.39</td>
<td>103.46</td>
<td>30.41</td>
<td>-8.4</td>
<td>41.62</td>
</tr>
<tr>
<td>113.73</td>
<td>26.74</td>
<td>101.54</td>
<td>25.47</td>
<td>-8.58</td>
<td>40.92</td>
</tr>
<tr>
<td>113.12</td>
<td>29.35</td>
<td></td>
<td></td>
<td>-8.64</td>
<td>42.02</td>
</tr>
<tr>
<td>113.01</td>
<td>25.77</td>
<td></td>
<td></td>
<td>-9.15158</td>
<td>38.75653</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>112.93</td>
<td>28.22</td>
<td>13.93376</td>
<td>41.24189</td>
<td>-71.0022</td>
<td>-34.5755</td>
</tr>
<tr>
<td>112.72</td>
<td>37.61</td>
<td>13.10813</td>
<td>46.01737</td>
<td>-71.5724</td>
<td>-35.5164</td>
</tr>
<tr>
<td>112.52</td>
<td>27.91</td>
<td>13.05889</td>
<td>41.49864</td>
<td>13.02467</td>
<td>41.29994</td>
</tr>
<tr>
<td>112.1924</td>
<td>27.45914</td>
<td>13.05788</td>
<td>41.80601</td>
<td>111.69</td>
<td>29.03</td>
</tr>
</tbody>
</table>
Figure C. 1. Pairwise correlation among all 19 variables in the background buffer of 400 km for Psa. The variables on diagonal are bio01 to bio19 (top to down) and Pearson's coefficients are in the upper right with the font-size proportional to the strength of the correlation. The variable plots are shown in lower-half (with a smoothed red line of the best fit).
Figure C. 2. Multivariate Environmental Similarity Surfaces (MESS) of 10 variables to the environmental data used for training the MaxEnt model in Psa. Areas in red have one or more environmental variables outside the range present in the training data.
Figure C. 3. Marginal response curves, each of the following curves represents a different model, namely, a MaxEnt model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. Annual mean temperature (cbio01), temperature seasonality (cbio04), minimum temperature of the coldest month (cbio06), mean temperature of the wettest quarter (cbio8), mean temperature of the warmest quarter (cbio10), annual precipitation (cbio12), precipitation of the driest month (cbio14), precipitation of the wettest quarter (cbio16) and precipitation of the coldest quarter (cbio19).
Appendix D
Potential climate suitability of dwarf bunt

Table D. 1. Dwarf bunt presence points coordinates (after removal of duplicate points).

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>67.70995</td>
<td>33.93911</td>
</tr>
<tr>
<td>20.16833</td>
<td>41.15333</td>
</tr>
<tr>
<td>1.659627</td>
<td>28.03389</td>
</tr>
<tr>
<td>-63.6167</td>
<td>-38.4161</td>
</tr>
<tr>
<td>45.03819</td>
<td>40.0691</td>
</tr>
<tr>
<td>13.48345</td>
<td>48.2083</td>
</tr>
<tr>
<td>14.48243</td>
<td>48.13371</td>
</tr>
<tr>
<td>14.55007</td>
<td>47.51623</td>
</tr>
<tr>
<td>15.82359</td>
<td>48.10735</td>
</tr>
<tr>
<td>16.28333</td>
<td>48.05</td>
</tr>
<tr>
<td>16.29254</td>
<td>48.08259</td>
</tr>
<tr>
<td>16.30659</td>
<td>48.04848</td>
</tr>
<tr>
<td>47.57693</td>
<td>40.14311</td>
</tr>
<tr>
<td>25.48583</td>
<td>42.73388</td>
</tr>
<tr>
<td>-127.648</td>
<td>53.72667</td>
</tr>
<tr>
<td>-85.3232</td>
<td>51.25378</td>
</tr>
<tr>
<td>15.2</td>
<td>45.1</td>
</tr>
<tr>
<td>16.61535</td>
<td>49.19098</td>
</tr>
<tr>
<td>9.032142</td>
<td>56.56356</td>
</tr>
<tr>
<td>10.67449</td>
<td>56.19887</td>
</tr>
<tr>
<td>-83.9071</td>
<td>34.15744</td>
</tr>
<tr>
<td>9.350134</td>
<td>48.6616</td>
</tr>
<tr>
<td>10.53807</td>
<td>48.00985</td>
</tr>
<tr>
<td>11.49789</td>
<td>48.79045</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.63667</td>
<td>51.48083</td>
</tr>
<tr>
<td>13.19101</td>
<td>51.09779</td>
</tr>
<tr>
<td>14.77361</td>
<td>51.29028</td>
</tr>
<tr>
<td>21.82431</td>
<td>39.07421</td>
</tr>
<tr>
<td>19.5033</td>
<td>47.16249</td>
</tr>
<tr>
<td>45.0759</td>
<td>37.45501</td>
</tr>
<tr>
<td>46.26821</td>
<td>37.90357</td>
</tr>
<tr>
<td>47.13621</td>
<td>35.95536</td>
</tr>
<tr>
<td>48.29505</td>
<td>38.24647</td>
</tr>
<tr>
<td>48.39882</td>
<td>33.58184</td>
</tr>
<tr>
<td>48.48511</td>
<td>36.67094</td>
</tr>
<tr>
<td>48.51625</td>
<td>34.8065</td>
</tr>
<tr>
<td>55.23483</td>
<td>37.20804</td>
</tr>
<tr>
<td>55.95</td>
<td>37.89</td>
</tr>
<tr>
<td>12.56738</td>
<td>41.87194</td>
</tr>
<tr>
<td>138.2529</td>
<td>36.20483</td>
</tr>
<tr>
<td>66.92368</td>
<td>48.01957</td>
</tr>
<tr>
<td>74.7661</td>
<td>41.20438</td>
</tr>
<tr>
<td>6.129583</td>
<td>49.81527</td>
</tr>
<tr>
<td>13.18751</td>
<td>32.87617</td>
</tr>
<tr>
<td>27.74023</td>
<td>47.87527</td>
</tr>
<tr>
<td>20.55015</td>
<td>50.758</td>
</tr>
<tr>
<td>25.42539</td>
<td>44.92899</td>
</tr>
<tr>
<td>27.72917</td>
<td>46.63833</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>105.3188</td>
<td>60.60288</td>
</tr>
<tr>
<td>21.00586</td>
<td>44.01652</td>
</tr>
<tr>
<td>14.99546</td>
<td>46.15124</td>
</tr>
<tr>
<td>-3.74922</td>
<td>41.46367</td>
</tr>
<tr>
<td>18.6435</td>
<td>60.12816</td>
</tr>
<tr>
<td>7.47267</td>
<td>46.88187</td>
</tr>
<tr>
<td>38.99682</td>
<td>34.80208</td>
</tr>
<tr>
<td>71.27609</td>
<td>38.86103</td>
</tr>
<tr>
<td>9.537499</td>
<td>33.88692</td>
</tr>
<tr>
<td>34.67736</td>
<td>37.96433</td>
</tr>
<tr>
<td>37.01788</td>
<td>39.74766</td>
</tr>
<tr>
<td>42.11667</td>
<td>38</td>
</tr>
<tr>
<td>43.36708</td>
<td>38.49417</td>
</tr>
<tr>
<td>59.55628</td>
<td>38.96972</td>
</tr>
<tr>
<td>31.16558</td>
<td>48.37943</td>
</tr>
<tr>
<td>-112.094</td>
<td>40.87</td>
</tr>
<tr>
<td>-120.554</td>
<td>43.80413</td>
</tr>
<tr>
<td>-116.742</td>
<td>43.0682</td>
</tr>
<tr>
<td>-110.363</td>
<td>46.87968</td>
</tr>
<tr>
<td>-107.29</td>
<td>43.07597</td>
</tr>
<tr>
<td>-104.782</td>
<td>39.95005</td>
</tr>
<tr>
<td>-120.74</td>
<td>47.75112</td>
</tr>
<tr>
<td>-74.006</td>
<td>40.71435</td>
</tr>
<tr>
<td>64.58526</td>
<td>41.37749</td>
</tr>
<tr>
<td>11.63667</td>
<td>52.12305</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>27.0155</td>
<td>49.58987</td>
</tr>
<tr>
<td>11.70204</td>
<td>51.48833</td>
</tr>
<tr>
<td>11.73194</td>
<td>51.45833</td>
</tr>
<tr>
<td>11.85694</td>
<td>51.49028</td>
</tr>
</tbody>
</table>
Figure D.1. Pairwise correlation among all 19 variables in the background buffer of 1000 km for dwarf bunt. The variables on diagonal are bio01 to bio19 (top to down) and Pearson’s coefficients are in the upper right with the font-size proportional to the strength of correlation. The variable plots are shown in lower-half (with a smoothed red line of the best fit).
Figure D. 2. The ROC plot of the MaxEnt model developed for dwarf bunt. The blue line indicates the AUC of the test data which is better than a random model (black diagonal line).
Figure D. 3. Multivariate Environmental Similarity Surfaces (MESS) of 10 variables to the environmental data used for training the MaxEnt model on dwarf bunt. Areas in red have one or more environmental variables outside the range present in the training data, so predictions in those areas should be treated with strong caution.
Figure D. 4. Marginal response curves for dwarf bunt, each of the following curves represents a different model, namely, a MaxEnt model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. Annual mean temperature (cbio1), temperature seasonality (cbio04), min temperature of the coldest month (cbio06), mean temperature of the wettest month (cbio08), mean temperature of the warmest quarter (cbio10), annual precipitation (cbio12), precipitation of the driest month (cbio14), precipitation of the wettest quarter (cbio16) and precipitation of the coldest quarter (cbio18).
## Appendix E

### Potential climate suitability of Guava rust

Table E.1. Guava Rust presence points coordinates (after removal of duplicate points).

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>-65.5</td>
<td>-27</td>
</tr>
<tr>
<td>-64.0667</td>
<td>-27.45</td>
</tr>
<tr>
<td>-56.1</td>
<td>-27.4</td>
</tr>
<tr>
<td>-59.5333</td>
<td>13.1667</td>
</tr>
<tr>
<td>-43.3167</td>
<td>-15.8</td>
</tr>
<tr>
<td>-44.6667</td>
<td>-14.1833</td>
</tr>
<tr>
<td>-41.1</td>
<td>-18.1</td>
</tr>
<tr>
<td>-40.1</td>
<td>-18.1</td>
</tr>
<tr>
<td>-42.8167</td>
<td>-5.08333</td>
</tr>
<tr>
<td>-42.9</td>
<td>-20.4</td>
</tr>
<tr>
<td>-40.2667</td>
<td>-19.8167</td>
</tr>
<tr>
<td>-43.3333</td>
<td>-22.95</td>
</tr>
<tr>
<td>-43.2333</td>
<td>-22.9</td>
</tr>
<tr>
<td>-49.05</td>
<td>-26.9333</td>
</tr>
<tr>
<td>-35.7167</td>
<td>-9.66667</td>
</tr>
<tr>
<td>-48.9333</td>
<td>-27.1</td>
</tr>
<tr>
<td>-34.9</td>
<td>-8.05</td>
</tr>
<tr>
<td>-51.3833</td>
<td>0.71667</td>
</tr>
<tr>
<td>-51.2833</td>
<td>-0.11667</td>
</tr>
<tr>
<td>-51.05</td>
<td>0.03333</td>
</tr>
<tr>
<td>-75.5833</td>
<td>6.25</td>
</tr>
<tr>
<td>-75.592</td>
<td>6.24115</td>
</tr>
<tr>
<td>-84.1</td>
<td>9.93333</td>
</tr>
<tr>
<td>-79.9833</td>
<td>21.8</td>
</tr>
<tr>
<td>-82.7667</td>
<td>22.81667</td>
</tr>
<tr>
<td>-74.5</td>
<td>20.35</td>
</tr>
<tr>
<td>-70.5167</td>
<td>19.21667</td>
</tr>
<tr>
<td>-70.5167</td>
<td>19.38333</td>
</tr>
<tr>
<td>-69.9</td>
<td>18.46667</td>
</tr>
<tr>
<td>-70.717</td>
<td>19.733</td>
</tr>
<tr>
<td>-78.9833</td>
<td>-2.2833</td>
</tr>
<tr>
<td>-78.4167</td>
<td>-1.4</td>
</tr>
<tr>
<td>-72.3333</td>
<td>18.53333</td>
</tr>
<tr>
<td>-76.8</td>
<td>18</td>
</tr>
<tr>
<td>-77.45</td>
<td>18.383</td>
</tr>
<tr>
<td>-77.3667</td>
<td>18.03333</td>
</tr>
<tr>
<td>-77.1167</td>
<td>18.2666</td>
</tr>
<tr>
<td>-77.7167</td>
<td>18.0333</td>
</tr>
<tr>
<td>-77.4833</td>
<td>18.16666</td>
</tr>
<tr>
<td>-77.7167</td>
<td>18.25</td>
</tr>
<tr>
<td>-76.7167</td>
<td>18.0667</td>
</tr>
<tr>
<td>-57</td>
<td>-24.5</td>
</tr>
<tr>
<td>-66</td>
<td>18.4</td>
</tr>
<tr>
<td>-66.0843</td>
<td>18.0515</td>
</tr>
<tr>
<td>-66.3914</td>
<td>18.22888</td>
</tr>
<tr>
<td>-67.0756</td>
<td>18.17455</td>
</tr>
<tr>
<td>-66.9803</td>
<td>18.18278</td>
</tr>
<tr>
<td>-66.7</td>
<td>18.16667</td>
</tr>
<tr>
<td>-61.5167</td>
<td>10.65</td>
</tr>
<tr>
<td>-60.8</td>
<td>11.2</td>
</tr>
<tr>
<td>-66.8983</td>
<td>10.49605</td>
</tr>
<tr>
<td>-96.1333</td>
<td>19.16</td>
</tr>
<tr>
<td>121.6</td>
<td>23.98333</td>
</tr>
<tr>
<td>-117.017</td>
<td>33.23657</td>
</tr>
<tr>
<td>-156.567</td>
<td>20.56667</td>
</tr>
<tr>
<td>-82.45</td>
<td>27.93333</td>
</tr>
<tr>
<td>-81.8667</td>
<td>26.63333</td>
</tr>
<tr>
<td>-157.717</td>
<td>21.35</td>
</tr>
<tr>
<td>-82.6667</td>
<td>27.76667</td>
</tr>
<tr>
<td>-80.3167</td>
<td>27.43333</td>
</tr>
<tr>
<td>-156.5</td>
<td>20.88333</td>
</tr>
<tr>
<td>-159.367</td>
<td>21.96667</td>
</tr>
<tr>
<td>-80.1833</td>
<td>25.76667</td>
</tr>
<tr>
<td>-80.05</td>
<td>26.7</td>
</tr>
<tr>
<td>-80.1333</td>
<td>26.11667</td>
</tr>
<tr>
<td>-58.45</td>
<td>-34.6</td>
</tr>
<tr>
<td>-54.6333</td>
<td>-26.4</td>
</tr>
<tr>
<td>-47.6333</td>
<td>-22.7167</td>
</tr>
<tr>
<td>-48.3167</td>
<td>-21.2666</td>
</tr>
<tr>
<td>-42.8833</td>
<td>-20.75</td>
</tr>
<tr>
<td>-51.9167</td>
<td>-23.4167</td>
</tr>
<tr>
<td>-47.0833</td>
<td>-22.9</td>
</tr>
<tr>
<td>-48.4333</td>
<td>-22.8667</td>
</tr>
<tr>
<td>-43.9333</td>
<td>-19.9167</td>
</tr>
<tr>
<td>-45</td>
<td>-21.2333</td>
</tr>
<tr>
<td>-46.6167</td>
<td>-23.5333</td>
</tr>
<tr>
<td>-47.9333</td>
<td>-15.7833</td>
</tr>
<tr>
<td>-43.7667</td>
<td>-21.2333</td>
</tr>
<tr>
<td>-51.3167</td>
<td>-30.1</td>
</tr>
<tr>
<td>-49.2333</td>
<td>-25.2833</td>
</tr>
<tr>
<td>-51.1667</td>
<td>-29.8333</td>
</tr>
<tr>
<td>-58</td>
<td>-31.4</td>
</tr>
<tr>
<td>-58.5167</td>
<td>-34.4667</td>
</tr>
</tbody>
</table>

Table E.1. Continued.
<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>-56.3333</td>
<td>-34.6667</td>
<td>-77.2975</td>
<td>18.10958</td>
</tr>
<tr>
<td>-57.3667</td>
<td>-32.4167</td>
<td>-77.4602</td>
<td>21.5217</td>
</tr>
<tr>
<td>-55.95</td>
<td>-31.6833</td>
<td>-78.1834</td>
<td>-1.83124</td>
</tr>
<tr>
<td>-55.95</td>
<td>-31.5833</td>
<td>-81.5157</td>
<td>27.6648</td>
</tr>
<tr>
<td>-55.95</td>
<td>-31.55</td>
<td>-88.8965</td>
<td>13.7941</td>
</tr>
<tr>
<td>-54.7333</td>
<td>-34.3167</td>
<td>-90.2307</td>
<td>15.7834</td>
</tr>
<tr>
<td>-55.15</td>
<td>-34.55</td>
<td>-102.553</td>
<td>23.6345</td>
</tr>
<tr>
<td>-76.717</td>
<td>18.083</td>
<td>-117.157</td>
<td>32.71533</td>
</tr>
<tr>
<td>139.87</td>
<td>34.99651</td>
<td>-155.119</td>
<td>19.649</td>
</tr>
<tr>
<td>22.9375</td>
<td>-30.5594</td>
<td>-156.47</td>
<td>20.89472</td>
</tr>
<tr>
<td>-40.3082</td>
<td>-19.1834</td>
<td>-158.053</td>
<td>21.43891</td>
</tr>
<tr>
<td>-43.2</td>
<td>-22.9</td>
<td>145.71</td>
<td>-16.92</td>
</tr>
<tr>
<td>-43.9</td>
<td>-19.8</td>
<td>145.422</td>
<td>-16.991</td>
</tr>
<tr>
<td>-43.9</td>
<td>-18.1</td>
<td>146.032</td>
<td>-17.534</td>
</tr>
<tr>
<td>-43.946</td>
<td>-19.9334</td>
<td>146.817</td>
<td>-19.257</td>
</tr>
<tr>
<td>-43.946</td>
<td>-3.41684</td>
<td>148.718</td>
<td>-20.628</td>
</tr>
<tr>
<td>-46.6388</td>
<td>-23.5486</td>
<td>150.512</td>
<td>-23.322</td>
</tr>
<tr>
<td>-54.6917</td>
<td>-32.7626</td>
<td>151.263</td>
<td>-23.851</td>
</tr>
<tr>
<td>-55.9568</td>
<td>-32.9812</td>
<td>152.701</td>
<td>-25.537</td>
</tr>
<tr>
<td>-57</td>
<td>-24</td>
<td>152.665</td>
<td>-26.19</td>
</tr>
<tr>
<td>-57</td>
<td>-27.4</td>
<td>153.09</td>
<td>-26.397</td>
</tr>
<tr>
<td>-57.7753</td>
<td>-31.9126</td>
<td>151.839</td>
<td>-26.54</td>
</tr>
<tr>
<td>-57.8645</td>
<td>-32.7952</td>
<td>150.625</td>
<td>-26.739</td>
</tr>
<tr>
<td>-60.8</td>
<td>-11.2</td>
<td>153.108</td>
<td>-27.229</td>
</tr>
<tr>
<td>-61.3285</td>
<td>10.09726</td>
<td>152.99</td>
<td>-27.485</td>
</tr>
<tr>
<td>-61.3709</td>
<td>15.4149</td>
<td>153.425</td>
<td>-28.017</td>
</tr>
<tr>
<td>-64.8964</td>
<td>18.33581</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure E. 1. Pairwise correlation among all 19 variables in the background buffer of 400 km for guava rust. The variables on diagonal are bio01 to bio19 (top to bottom) and Pearson's coefficients are in the upper right with the font-size proportional to the strength of correlation. The variable plots are shown in the lower-half (with a smoothed red line of the best fit).
Figure E. 2. Multivariate Environmental Similarity Surfaces (MESS) of 10 variables to the environmental data used for training the MaxEnt model in guava rust. Areas in red have one or more environmental variables outside the range present in the training data, so predictions in those areas should be treated with strong caution.
Figure E.3. Marginal response curves, each of the following curves represents a different model, namely, a Maxent created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. Annual mean temperature (cbio1), mean of monthly temperature (max temp-min temp) (cbio02), maximum temperature of the warmest month (cbio05), minimum temperature of coldest month (cbio06), annual temperature range (cbio07), mean temperature of the wettest month (cbio08), mean temperature of the driest quarter (cbio09), precipitation of the wettest month (cbio13), precipitation of the driest month (cbio14) and precipitation seasonality (cbio15).
Figure E. 4. The ROC plot of the MaxEnt model developed for guava rust. The blue line indicates the AUC of the test data which is better than a random model (black diagonal line).
Appendix F

Predicting the potential global distribution of *Pseudomonas syringae* p.v. *actinidiae* (Psa)

H.A. Narouei Khandan¹, S.P. Worne¹, E.E. Jones², S.L.H. Villijansen-Rollinson¹, L. Gallioli³, A. Mazzaglia⁴ and G.M. Balestra⁴

¹Bio-Protection Research Centre, Lincoln University, P O Box 85086, New Zealand
²Faculty of Agriculture and Life Sciences, Lincoln University, New Zealand
³The New Zealand Institute for Plant & Food Research Limited, Private Bag 4704, Christchurch 8140, New Zealand
⁴Department of Science and Technologies for Agriculture, Forestry, Nature and Energy (DAFNE), University of Tuscia, Viterbo, Italy
Corresponding author: hakhandan@gmail.com

Abstract The increasing spread of kiwifruit bacterial canker, caused by *Pseudomonas syringae* pv. *actinidiae* (Psa), prompted a modelling effort to assess the global and local potential risk of this species. The current potential distribution of Psa was modelled with two well-used models (CLIMEX and MaxEnt) based on available presence records and environmental data. Most discrepancies in model projections occurred for New Zealand data that was used for validation. Model projections can provide information to alert decision-makers in kiwifruit-growing regions to prepare for possible incursions of Psa. However, in this study because model findings did not agree on the New Zealand validation data, more research is necessary to achieve greater confidence on projections for novel areas. Despite that result, this study provides useful information for some kiwifruit growing countries that have not yet been affected by Psa, such as USA, Iran, Greece, Belgium, Denmark and especially South Africa, where commercial kiwifruit orchards have been planted recently.

Keywords kiwifruit bacterial disease, CLIMEX, MaxEnt, correlative model, mechanistic model.

INTRODUCTION

Bacterial canker of kiwifruit (*Pseudomonas syringae* pv. *actinidiae*, Psa) was reported from Shizuoka Prefecture, Japan in early spring 1984 where it affected 56 ha out of 162 ha of regional orchards (Serizawa et al. 1989). A few years later, in the late 1980s this pathogen was reported in Korea (Koh et al. 1994), then Yuexi China in 1990 (Cheng et al. 1995) and the central parts of Italy in 1992 (Scortichini 1994). In recent years it has appeared again but more seriously in Italy (Balestra et al. 2008; Renzi et al. 2012a) as well as in Portugal (Balestra et al. 2010; Renzi et al. 2012a), Australia (Biosecurity 2011), Spain (Balestra et al. 2011), France (Vanneste et al. 2012b), New Zealand (Everett et al. 2011) and Turkey (Bastas & Karalaya 2012).

This disease can infect kiwifruit plants belonging to different species such as *Actinidia chinensis* and *Actinidia delicosa*. Recently it has been reported that it can also survive on *Cryptomeria japonica*, a non-host plant (Vanneste et al. 2012a). Formerly, the economic importance of this disease was considered relatively low, until an outbreak in Italy (Latina...
Kiwi fruit (in the spring of 2008 caused economic losses up to 2 million Euros (Balestra et al. 2009). In New Zealand, a recent study has shown that the virulent strain of Psa (Psa-V) is likely to cost $310-$410 million over the next 5 years, increasing to approximately $500-$600 million over the next 10 years and $740-$885 million over the next 15 years (Greer & Saunders 2012). Symptoms of Psa infection are mainly visible in spring and autumn when climatic conditions are favourable. The main symptoms are well described in the literature and include the production of translucent exudate from tissues, leaf spots in spring that later change to brown, exuding gum from flowers and trunk necrosis.

The pathogen is active in infected plant material and can easily be transferred to new plants by nursery practices and implements, particularly grafting and pruning tools. Environmental factors such as wind and rain are suspected to be major vectors for introduction into new regions (Renzi et al. 2012b).

The pathogen is most virulent at relatively low temperatures between 10-20°C with an optimum temperature of 15±3°C (Serizawa & Ichikawa 1993; Cheng et al. 1995). At temperatures above 25°C (Cheng et al. 1995), spread and infection of the pathogen is completely inhibited although in one study this threshold is given as 27°C (Serizawa & Ichikawa 1993). Previous studies suggest that rain is an important factor, possibly even the most dominant factor in establishment and distribution of Psa but detailed information is lacking (Yao et al. 2001).

Psa has resulted in significant losses to several major kiwifruit producing countries, but so far there are no Psa occurrence reports from some other major kiwifruit producing countries, namely USA, South Africa, Greece and Iran. As some kiwifruit varieties (dormant mature vines) can survive from -10 up to 35°C (Morton & Dowling 1987), and available laboratory studies on environmental needs of Psa show a narrower temperature range than the host plant (Serizawa & Ichikawa 1993; Cheng et al. 1995), a modelling approach may help assess the risk of this pathogen and determine the level of climate suitability inside kiwifruit growing regions.

Such an approach may also indicate different environmental needs to what laboratory findings currently suggest. In addition, assessment of the potential geographic distribution and spread of this pathogen can inform governments, decision-makers and companies who are planning to plant kiwifruit in new areas.

Correlative and mechanistic models have not yet been widely used to model or assess potential plant pathogen distributions. Most use only one model to assess the risk of plant pathogen establishment in new regions. Early studies mainly use CLIMEX as the modelling software (Yonow et al. 2004; Beddow et al. 2013; Elith et al. 2013).

In this study the current potential distribution of Psa based on climate suitability was modelled using one semi-mechanistic model (CLIMEX) and MaxEnt as a correlative model. The aim was to not favour one approach over another, as the issues involved and the discussion has been covered in previous studies where the relative merits are still being debated (Robertson et al. 2003; Kearney et al. 2010; Webber et al. 2011). The goal was to focus on the agreement between two models and show how different model projections can contribute to a better understanding of the risk in Psa spread based on current distribution reports.

MATERIALS AND METHODS

Species occurrence data

Locating occurrence data of plant pathogens is more challenging compared to other organisms as there are few databases available that provide precise plant pathogen occurrences, and those that exist (e.g. GBIF and CABI Crop Compendium) provide little relevant information. Therefore, published papers and personal communication with their authors were used to gather the required information. The presence data comprised all locations where Psa has been reported. In total, the data comprised 211 points from 10 countries (Figure 1a). There are recent reports of Psa in Turkey and Switzerland, but these records were not used because precise localities could not be confirmed. Also, because one of the goals of this study was to project climate suitability of Psa in
New Zealand, the New Zealand data were not used in the process of model building. Instead, they were used to validate and assess model predictive performance in new areas. There is no published information about the difference in environmental needs of Psa-V (virulent strain) and Psa-LV strains (less virulent strain), so there was no attempt to differentiate these two strains in modelling efforts.

Climate data and predictor variables

For the CLIMEX model (Sutherst & Maywald 1985) 10° gridded climate data were accessed from the Climond website (Kriticos et al. 2012). For the MaxEnt model 19 bioclimatic variables with the same resolution (10°) were available from the Worldclim website (www.worldclim.org). The latter database comprises many derived variables based on long-term temperature and rainfall that are often used for species niche modelling. Twelve candidate variables were selected by use of correlation tests to eliminate highly correlated variables and by assessment of their ecological relevance, given current knowledge of the target species.

CLIMEX and MaxEnt were used to evaluate the habitat suitability of Psa based on its current range. Because species distribution models do not perform well if the climate in the target area is very different from the native and invaded ranges, a multivariate environmental similarity surface (MESS) was performed to determine the degree of model extrapolation in the areas of interest.

Modelling

MaxEnt (Phillips et al. 2006; Elith et al. 2011) and CLIMEX (Sutherst & Maywald 1985) were used to predict climate suitability of Psa based on its current occurrences around the world. MaxEnt is a presence-only model based on maximum entropy that is used in species habitat suitability modelling. This software uses environmental data together with geo-referenced location data to project the potential range of target species. A detailed description of MaxEnt has been provided by Phillips and co-workers (Phillips et al. 2006; Phillips & Dudik 2008) and the advantages and disadvantages of presence-only models have been previously discussed (Elith et al. 2011). Unlike statistical models, CLIMEX models a species distribution by selecting values for a set of parameters that describe its response to temperature, moisture and light. Four stress indices (corresponding to cold, hot, wet and dry), and in some cases their interactions, describe the extent to which the population is reduced during the unfavourable season. The growth and stress indices are combined into an Ecoclimatic Index (EI), to give an overall measure of favourableness of the location or year for permanent occupation by the target species. The EI is scaled to an integer between 0 and 100, with an EI close to 0 indicating that the location is not favourable for the long-term survival of the species. As per Vera et al. (2002), an EI value of 0 is considered unfavourable, 1-10 is marginal, 11-25 is favourable and greater than or equal to 26 is considered very favourable for establishment of that species.

Compared to MaxEnt, CLIMEX has been applied more often to model plant pathogens (Yonow et al. 2004), but there are also recent studies using MaxEnt (Elith et al. 2013) and this model has been used widely to model the distributions of other species.

CLIMEX modelling process

The methodology described in Sutherst & Maywald (1985) was used to fit the growth and stress parameters. The initial temperature index parameters (DV0, DV1, DV2, DV3) were set based on minimum, optimum and maximum temperature requirements of Psa reported by Serizawa et al. (1993). The soil moisture parameters were determined by calibrating parameters to match model output to the reported occurrence of Psa. The stress parameters were fitted so that the majority of known occurrences of Psa occurred in climatically suitable and very suitable grid cells. As there are no data available regarding the response of Psa to stress conditions, stress parameters were calibrated by comparing model results with the known distribution (Taylor & Kumar 2012). The final parameter values are shown in Table 1. Model fit was visually assessed and the criteria were to observe the closest match of suitable
Table 1 Parameter values used in CLIMEX model.

<table>
<thead>
<tr>
<th>Index</th>
<th>Parameters</th>
<th>Values</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV0</td>
<td>Lower temperature threshold</td>
<td>5</td>
<td>°C</td>
</tr>
<tr>
<td>DV1</td>
<td>lower optimum temperature</td>
<td>12</td>
<td>°C</td>
</tr>
<tr>
<td>DV2</td>
<td>upper optimum temperature</td>
<td>1</td>
<td>°C</td>
</tr>
<tr>
<td>DV3</td>
<td>Upper temperature threshold</td>
<td>27</td>
<td>°C</td>
</tr>
<tr>
<td>SM0</td>
<td>Lower soil moisture threshold</td>
<td>0.5</td>
<td>-</td>
</tr>
<tr>
<td>SM1</td>
<td>Lower optimum soil moisture</td>
<td>0.8</td>
<td>-</td>
</tr>
<tr>
<td>SM2</td>
<td>upper optimum soil moisture</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>SM3</td>
<td>Upper soil moisture threshold</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>TTCS</td>
<td>Cold stress temperature threshold</td>
<td>5</td>
<td>°C</td>
</tr>
<tr>
<td>THCS</td>
<td>Cold stress temperature rate</td>
<td>-0.00005</td>
<td>Week(^{-1})</td>
</tr>
<tr>
<td>DTCS</td>
<td>Cold stress degree-day threshold</td>
<td>15</td>
<td>°C</td>
</tr>
<tr>
<td>DHCS</td>
<td>Cold stress degree-day rate</td>
<td>-0.0001</td>
<td>Week(^{-1})</td>
</tr>
<tr>
<td>THHS</td>
<td>Heat Stress Temperature threshold</td>
<td>30</td>
<td>°C</td>
</tr>
<tr>
<td>THHS</td>
<td>Heat Stress Temperature rate</td>
<td>0.0005</td>
<td>Week(^{-1})</td>
</tr>
<tr>
<td>SDMS</td>
<td>Dry Stress Threshold</td>
<td>0.2</td>
<td>Week(^{-1})</td>
</tr>
<tr>
<td>HDS</td>
<td>Dry Stress Rate</td>
<td>-0.005</td>
<td>Week(^{-1})</td>
</tr>
<tr>
<td>SMWS</td>
<td>Wet stress threshold (1-10)</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>HWS</td>
<td>Wet stress rate</td>
<td>0.001</td>
<td>Week(^{-1})</td>
</tr>
</tbody>
</table>

Habitats projected by CLIMEX and the reported distribution patterns of Psal in areas of the world excluding New Zealand. CLIMEX was then run to project the suitable distribution in New Zealand.

**MaxEnt modelling process**

As a presence-only model, MaxEnt uses presence and background data (a sample of all other data that is not known as presence data) to predict species distribution. Van Der Wal et al. (2009) suggested that the spatial extent at which background samples (or pseudo-absences) are taken should be considered carefully. They proposed defining a background buffer within some distance of an occurrence/presence point as a solution to get the most biologically meaningful background boundary or pseudo-absence data. Although Van Der Wal et al. (2009) suggested a 200 km buffer around the presence points as an optimum distance within which to select the background points, others have suggested performing an independent exploratory analysis to find the optimum distance and therefore achieve the most accurate projections. In another study, Thuiller et al. (2004) found that the effect of restricting the environmental range of data can strongly influence the response curve estimation and model performance. To consider these issues of background selection, a novel method (S.D. Senay, Lincoln University, unpublished data) that uses Principle Component Analysis (PCA) as a method to define both a spatially and ecologically meaningful distance to limit the background data was applied. First, ArcMap 10.0 was used to create 50 km, 200 km, 250 km, 300 km and 400 km buffers around the presence points and the needed environmental layers were extracted from selected predictor variables. Using this distance buffer, the effect of the background limits on the AUC (area under the curve) of the model was tested and compared with PCA results. Both methods coincided on 250 km as the optimum buffer distance around the presence points to be used as background data for selection of pseudo-absence points. Linear features, quadratic features, product and hinge features were used during the training as the dataset is large enough to support complexity. The maximum number of background points was...
set at 10,000 and 5,000 iterations. The model was set to create the response curves and a jackknife test was used to measure variable importance.

RESULTS
The results of parameter fitting of the CLIMEX model are shown in Table 1, indicating temperature, soil moisture and stress parameters achieved by calibration or iterative adjustments. Results were checked to see if they were biologically reasonable. The lower temperature threshold, 5°C, is lower than what is reported generally in the literature (8-10°C). Also, calibration gave a higher temperature threshold of 27°C, whereas in the literature it is stated as 25°C, and this led to a slightly wider survival range for Ps a than what was expected. During calibration it was noticed that northern and eastern parts of China show high sensitivity to cold stress temperature threshold (THCS) compared to other parts of the world. Although the main goal of this study was to project the suitability of Ps a occurrence in New Zealand, results for the rest of the world are also presented and discussed.

Based on the currently available presence data, the CLIMEX results were a good fit to the current global distribution of Ps a (Figure 1b). In Asia, the areas with highest E1 values (and hence higher climate suitability) belong to central China and Japan, which coincide with documented reports from these regions. Model projections for northern parts of Iran where kiwifruit is grown but Ps a is not recorded suggest some areas are highly favorable. Portugal, the Latin province of Italy and Greece are predicted to be highly suitable. Recently there has been an increase in reports coming from northern Portugal that is affected by Ps a. Moreover, projections of both the models showed that central parts of Portugal are highly suitable areas. Regions in Africa projected as suitable by CLIMEX are mainly areas at high elevations and in the north of Africa, but that may not be of much interest because there are no or few kiwifruit orchards in those African countries. However, northern Tunisia and small areas around Cape Town and Durban are projected as highly suitable.

In South America, the highly suitable areas indicated by both models extend from southern Brazil up to Sao Paulo, the whole of Uruguay and Buenos Aires in Argentina. Also, the locations of the main kiwifruit orchards around Santiago in Chile are projected as highly suitable (two reports come from this neighboring region). With respect to North America, although there are no reports to date, both models projected that California and areas in the southeastern states are suitable to highly suitable for Ps a establishment. Kiwifruit is cultivated commercially in Sacramento and San Joaquin valley and sporadically in North Carolina. The projection for Australia is limited to southeast coastal areas, namely in Victoria, and there has been a report of Ps a-LV in this area. Also, both models predicted that small areas in the west of Australia (around Perth) are suitable, but as yet there are no official reports of establishment of Ps a there. There has been one potential report in the west of Australia (Reynaud et al. 2011, cited in EPPO 2012), but this could not be confirmed.

New Zealand was left out of the parameter fitting and calibration phase for both models so that any projections for New Zealand are independent of the parameter fitting process and could be used for validation. In general, the CLIMEX model provided a very good projection of current occurrence data for Ps a. In fact, almost all parts of the North Island of New Zealand along with small areas of Nelson were projected as highly suitable.

For MaxEnt, the MESS analysis indicated that only the Westland climate in New Zealand is climatically significantly different from climates associated with the current occurrence data, so extrapolation for this area was not recommended. The jackknife test indicated that among the temperature variables, mean annual temperature and mean temperature of the wettest and driest quarters are important variables. An analysis of the relative contribution of environmental variables showed that the mean annual temperature, mean temperature of the wettest quarter and precipitation of wettest quarter contributed most to explaining Ps a distribution.
Figure 1 (a) Current global distribution of Psa (presence data) used in CLIMEX and MaxEnt model. (b) Potential distribution of Psa modelled by CLIMEX. (c) Potential distribution of Psa modelled by MaxEnt. Red and yellow colours indicate greater climatic suitability.
(Table 2). Seventy-five percent of data were used to train/fill the model and 25% were set aside to validate the model (Hubert et al. 1994, cited by Fielding & Bell 1997). The AUC as a measure of performance of the MaxEnt model was 0.863, indicating MaxEnt projected reasonably well for the data put aside for validation. MaxEnt gave a probability of 50-70% of establishment for many areas that Psa has already established (Figure 1c). For Belgium and Denmark where kiwifruit is grown but Psa not reported, MaxEnt projection shows no or low probability of Psa occurrence whereas CLIMEX projection indicates Belgium as a suitable area. MaxEnt reflected CLIMEX predictions in general but provided different probabilities of occurrence, especially for New Zealand. MaxEnt did not predict anywhere in New Zealand as suitable for Psa establishment despite the fact that most of the kiwifruit growing areas have been badly affected.

**DISCUSSION**

In any attempt to model or characterise the suitable environment for an invasive species to establish in a new area, consideration should be given to species that are still spreading. The recorded presence locations may not always give a full indication of the actual potential distribution of target species for a number of reasons, such as the absence of disease reports due to lack of detection, an evolutionary change or a new pathway facilitating spread. These issues need to be kept in mind when interpreting projections, and further information can be acquired by analysing the presence data of the native and invaded ranges. For example, a PCA analysis showed that some of presence points for Psa in China are environmentally quite different from those in Europe. So, using the environmental conditions of native and invaded range in the models may give better insight into model projections of Psa potential distribution.

As expected, both models showed good fit to European data on which they were trained. However, there has been a recent report of the occurrence of Psa in Turkey that was not used in modelling because the exact locations were unknown (Rize Province) (Bastas & Karakaya 2012). While this area is predicted as marginal by CLIMEX, both models predict high climate suitability in western parts of Turkey and so may alert authorities in that area. Despite the fact that Psa is not known to be present in Greece (Ministry of Rural development and Food 2012), both models projected marginal to suitable areas in the southwest. As Greece produces kiwifruit commercially, special attention may be needed in these regions.

**Table 2** Estimates of relative contributions of the environmental variables used in the MaxEnt model.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Percentage contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO1 Annual Mean Temperature</td>
<td>20.1</td>
</tr>
<tr>
<td>BIO8 Mean Temperature of Wettest Quarter</td>
<td>19.5</td>
</tr>
<tr>
<td>BIO13 Precipitation of Wettest Month</td>
<td>18.7</td>
</tr>
<tr>
<td>BIO10 Mean Temperature of Warmest Quarter</td>
<td>9.8</td>
</tr>
<tr>
<td>BIO12 Annual Precipitation</td>
<td>8.9</td>
</tr>
<tr>
<td>BIO5 Max Temperature of Warmest Month</td>
<td>8.2</td>
</tr>
<tr>
<td>BIO9 Mean Temperature of Driest Quarter</td>
<td>4</td>
</tr>
<tr>
<td>BIO17 Precipitation of Driest Quarter</td>
<td>3.9</td>
</tr>
<tr>
<td>BIO3 Isothermality¹</td>
<td>2.8</td>
</tr>
<tr>
<td>BIO6 Min Temperature of Coldest Month</td>
<td>2</td>
</tr>
<tr>
<td>BIO14 Precipitation of Driest Month</td>
<td>1.7</td>
</tr>
<tr>
<td>BIO2 Mean Diurnal Range</td>
<td>0.4</td>
</tr>
</tbody>
</table>

¹Annual Mean Temperature / Mean Diurnal Range(*100).
for Portugal, most occurrence data come from northern regions, both models projected central parts of Portugal to be suitable as well. Based on these projections, it seems reasonable to expect that Psi will continue to spread into remaining kiwifruit growing areas in Portugal if no effective preventative measures are implemented (Balestra et al. 2010; Renzi et al. 2012a).

In Asia, a small area in the north of Iran was predicted as highly suitable by both models and interestingly the major kiwifruit growing areas are located in this part of Iran. Although one publication reported the presence of kiwifruit bacterial canker in Iran, it refers to *P. syringae* pv. *syringae* (Mazarei & Mostofipour 1994). In China, both models showed reasonable to good fit to occurrence data especially for northern and eastern regions. At a regional level, Hebei province (kiwifruit cultivated area 500–1000 ha) and Beijing (less that 100 ha) (G.M. Balestra, personal communication) were not projected as suitable by both models. In southern parts of China, Yunnan and Guangxi province where kiwifruit is also cultivated (500–1000 ha) were predicted as highly suitable by CLIMEX and 40–60% probability of occurrence by MaxEnt (Figures 1b & 1c), but there are no reports of Psi from these regions so far. These results suggest further investigation is needed to explain why Psi is not established in these regions. For example, maybe Psi had not yet been detected or perhaps the climate is indeed unsuitable. Small areas in Southern Laos and Vietnam where kiwifruit is grown (although not commercially) are projected as unsuitable for Psi by both models.

The largest discrepancy between the two models is clearly New Zealand, where MaxEnt projected a low probability of Psi occurrence and CLIMEX projected that the North Island and northern parts of the South Island are highly suitable. Although an attempt was made to improve the MaxEnt model by considering different scenarios with different assumptions, such as different background limits, different variables, different features and a MESS test, that did not enhance its performance.

In conclusion, despite these two well-known models being in reasonable agreement for predicting the Psi suitability of some regions overseas, significant areas of disagreement, particularly in New Zealand, indicate more research is necessary to increase confidence in the model projections. Despite the limitations of species distribution models, these efforts to characterise the environmental requirements of this dangerous biotic threat, integrated with expert opinion, can help decision-makers in kiwifruit-growing countries to design and implement effective policies to prepare for the possible occurrence of Psi. The results in this study may be particularly important for the USA, Iran, Greece, Belgium, Denmark and especially South Africa, where kiwifruit has been planted recently. For more local considerations, characterisation and projection of Psi environmental requirements can indicate to growers the location of orchards in low-risk growing areas.

ACKNOWLEDGEMENTS

The authors would like to thank the following Bio-Protection Research Centre persons for their contribution to this study: Senait Senay, Tasha Shelby, Jennifer Pannell and Federico Tomasetto, as well as David Logan from Plant & Food Research and ZESPRI International Limited for providing information regarding kiwifruit. This study was funded by B3, an unincorporated joint venture of four science agencies – Plant & Food Research, AgResearch, Scion and the Bio-Protection Research Centre at Lincoln University – and three end-user partners – the Ministry for Primary Industries, the Department for Conservation and the New Zealand Forest Owners Association.

REFERENCES


