DISEASE RISK MAPPING WITH METAMODELS FOR COARSE RESOLUTION PREDICTORS: GLOBAL POTATO LATE BLIGHT RISK NOW AND UNDER FUTURE CLIMATE CONDITIONS

by

ADAM HENRY SPARKS

B.S., Purdue University, 2000

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Plant Pathology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2009

Abstract

Late blight of potato, caused by *Phytophthora infestans*, is a pernicious disease of potatoes worldwide. This disease causes yield losses as a result of foliar and tuber damage. Many models exist to predict late blight risk for control purposes with-in season but rely upon fine-scale weather data collected in hourly, or finer, increments. This is a major constraint when working with disease prediction models for areas of the world where hourly weather data is not available or is unreliable. Weather or climate summary datasets are often available as monthly summaries. These provide a partial solution to this problem with global data at large time-steps (e.g., monthly). Difficulties arise when attempting to use these forms of data in small temporal scale models. My first objective was to develop new approaches for application of disease forecast models to coarser resolution weather data sets. I created metamodels based on daily and monthly weather values which adapt an existing potato late blight model for use with these coarser forms of data using generalized additive models. The daily and monthly weather metamodels have R-squared values of 0.62 and 0.78 respectively. These new models were used to map global late blight risk under current and climate change scenarios resistant and susceptible varieties. Changes in global disease risk for locations where wild potato species are indigenous, and disease risk for countries where chronic malnutrition is a problem were evaluated. Under the climate change scenario selected for use, A1B, future global late blight severity decreases. The risk patterns do not show major changes, areas of high risk remain high relative to areas of low risk with rather slight increases or decreases relative to previous years. Areas of higher wild potato species richness experience slightly increased blight risk, while areas of lower species richness experience a slight decline in risk.

DISEASE RISK MAPPING WITH METAMODELS FOR COARSE RESOLUTION PREDICTORS: GLOBAL POTATO LATE BLIGHT RISK NOW AND UNDER FUTURE CLIMATE CONDITIONS

by

ADAM HENRY SPARKS

B.S., Purdue University, 2000

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Plant Pathology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2009

Approved by:

Major Professor
Karen Garrett

Abstract

Late blight of potato, caused by *Phytophthora infestans*, is a pernicious disease of potatoes worldwide. This disease causes yield losses as a result of foliar and tuber damage. Many models exist to predict late blight risk for control purposes with-in season but rely upon fine-scale weather data collected in hourly, or finer, increments. This is a major constraint when working with disease prediction models for areas of the world where hourly weather data is not available or is unreliable. Weather or climate summary datasets are often available as monthly summaries. These provide a partial solution to this problem with global data at large time-steps (e.g., monthly). Difficulties arise when attempting to use these forms of data in small temporal scale models. My first objective was to develop new approaches for application of disease forecast models to coarser resolution weather data sets. I created metamodels based on daily and monthly weather values which adapt an existing potato late blight model for use with these coarser forms of data using generalized additive models. The daily and monthly weather metamodels have R-squared values of 0.62 and 0.78 respectively. These new models were used to map global late blight risk under current and climate change scenarios resistant and susceptible varieties. Changes in global disease risk for locations where wild potato species are indigenous, and disease risk for countries where chronic malnutrition is a problem were evaluated. Under the climate change scenario selected for use, A1B, future global late blight severity decreases. The risk patterns do not show major changes, areas of high risk remain high relative to areas of low risk with rather slight increases or decreases relative to previous years. Areas of higher wild potato species richness experience slightly increased blight risk, while areas of lower species richness experience a slight decline in risk.

Table of Contents

Table of Contents v

List of Figures vii

List of Tables viii

Acknowledgements ix

Dedication x

Preface xi

1 The Big Picture: Mapping Global Plant Disease Risk 1
- 1.1 Introduction . . . 1
- 1.2 Maps of disease and other ecological traits . . . 3
- 1.3 Scale and disease risk mapping . . . 4
- 1.4 Data types and data quality . . . 6
- 1.5 Spatial pattern and disease risk . . . 7
- 1.6 Disease risk maps based on meteorology . . . 8
- 1.7 Disease risk maps based on previous disease locations . . . 11
- 1.8 Disease risk maps based on host availability . . . 13
- 1.9 Disease risk maps based on climate envelope . . . 15
- 1.10 Combining factors in disease risk mapping . . . 17
- 1.11 Conclusion . . . 20

2 Using Metamodels to Adapt Hourly Time-Step, Field or Plot Scale Models to Monthly Time-Step, Global Scale Models 25
- 2.1 Introduction . . . 25
- 2.2 Materials and Methods . . . 28
- 2.3 Results . . . 30
  - 2.3.1 SimCast Daily Means and SimCast Monthly Means Fit . . . 30
  - 2.3.2 Comparison when using potato growing areas only to construct GAM 32
  - 2.3.3 Comparison of late blight risk predictions to AUDPC values . . . 34
- 2.4 Discussion . . . 36

3 Application of Metamodels for Estimating Global Disease Risk: Potato Late Blight Now and in the Future 40
- 3.1 Introduction . . . 40

3.2 Objectives 43
3.3 Materials and Methods 44
3.4 Results 45
3.4.1 Changes in global late blight risk 45
3.4.2 Changes in late blight risk for countries with chronic malnutrition and high potato priority 46
3.4.3 Changes in late blight risk for areas with wild potato species present 48
3.5 Discussion 48

Bibliography **70**

A Supplementary Maps **71**

List of Figures

1.1 Primary types of data and uses across spatial and temporal scales for disease risk prediction. 22
1.2 Hypothetical landscapes illustrating how maps of climate, historical disease, and host distributions can be used to produce disease risk maps for a continuous variable. 23
1.3 Neighborhood risk models can be used to incorporate the influence of neighboring areas in disease risk maps 24

2.1 Weather station locations used for model construction and validation 28
2.2 Fitted GAM surface for SimCast Daily Means 31
2.3 Fitted GAM surface for SimCast Monthly Means. 32
2.4 Boxplots of SimCast Daily Means fit when plotted by SimCast predictions based on hourly weather data 33
2.5 Boxplots of SimCast Monthly Means fit when plotted by SimCast predictions based on hourly weather data 34
2.6 SimCast and SimCast Daily Means blight units fitted to AUDPC values from several field plot trials. 35

3.1 1950 to 1990 global late blight risk map 51
3.2 2000 to 2020 global late blight risk map 52
3.3 2040 to 2060 global late blight risk map 53
3.4 2080 to 2100 global late blight risk map 54
3.5 2040 to 2060 global late blight risk map for resistant cultivar 55
3.6 Map of relative change in global late blight risk map from 1961-1990 to 2040-2060 56
3.7 Change in late blight risk for select countries from 1961-1990 conditions to 2040-2060 57
3.8 Late blight risk for wild potato species found in South America. 58
3.9 1961-1990 late blight risk to wild potato species 59
3.10 2000-2020 late blight risk to wild potato species 60
3.11 2040-2060 Late blight risk to wild potato species 61
3.12 2080-2100 late blight risk to wild potato species 62

A.1 Blight severities for select countries 71
A.2 Map of global regions where potato is grown 72

List of Tables

2.1 Goodness of fit of SimCast Daily Means and SimCast Monthly Means for construction and validation data sets. Results for susceptible cultivars are shown, resistant cultivar models were similar. 33

2.2 Pearson's correlation score of SimCast Daily Means and SimCast Monthly Means blight unit predictions with SimCast predictions. SimCast predicts blight units using hourly weather data. SimCast Daily Means and SimCast Monthly Means predict blight units based on daily and monthly weather data respectively. 34

3.1 Cumulative blight units and percentage change from previous time period of sum of blight units for each of the climate time periods. Blight units[89,117] are an indicator of late blight risk. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and[113 113] were provided by Robert Hijmans, University of California at Davis. 46

3.2 Change of total of blight units for time periods of 1961-1990 and 2040-2060 in countries with high potato priority and chronic malnutrition (Hunger) expressed as percentage of population chronically malnourished. Blight units[89,117] are an indicator of late blight risk. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis. 47

3.3 Blight units observed from November through April in South America where wild potato species are found. Blight units[89,117] are an indicator of late blight risk. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis. 48

Acknowledgments

I would like to thank all of the individuals who have encouraged me and supported me on this journey that has been my doctoral program. There are so many individuals who have guided me, supported me, inspired me, and pushed me to be my best along the way.

I owe a debt of gratitude to Drs. Karen Garrett, Jim Stack, Erick De Wolf, Shawn Hutchinson and Greg Forbes for their friendship and support. Their patience and guidance through this made this possible. I also thank and acknowledge Dr. Robert Hijmans for his contributions and technical assistance with this project.

I also must thank Drs. Dave Mengel, Ellsworth Christmas, and Greg Willoughby. They gave me my first opportunities to pursue research and have remained supportive long after I no longer worked with them directly.

Lastly all the wonderful individuals I have had the opportunity to work with from CIP for their assistance. Graham Thiele, Henry Juarez, Julia Zamudio, and Rubí Raymundo. Without them trips to Lima, data, and methods developed here would not have been possible.

I acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy.

I am grateful for support by USAID CIP linkage funds, by NSF Grants DEB-0516046 and EF-0525712 as part of the joint NSF-NIH Ecology of Infectious Disease program, and by the USAID for the Sustainable Agriculture and Natural Resources Management Collaborative Research Support Program (SANREM CRSP) under terms of Cooperative Agreement Award No. EPP-A-00-04-00013-00 to the Office of International Research and Development at Virginia Tech.

Dedication

Dedictated to my family and my friends who have helped and encouraged me along the way. I could not have done this without your support.

Preface

The research reported in this thesis was a project with the International Potato Center (CIP), in Lima, Peru. There was an opening for a student to collaborate with CIP scientists in an analysis of late blight risk using GIS tools. Having the technical skills to undertake this project, I readily accepted the offer. The experiences I gained from this dissertation stretch far beyond my growth as a scientist. It has been a long and winding path to reach this point, but ultimately it is rewarding to know that this project will contribute to CIP's mission of helping feed the hungry in developing parts of the world.

Chapter 1

The Big Picture: Mapping Global Plant Disease Risk

1.1 Introduction

New ecoinformatic approaches are expanding our ability to forecast ecosystem status and changes, such as evaluations of disease risk across large areas and time lines[1]. As pressures for use of resources intensify, it will be increasingly important to optimize both short-term and long-term strategies for disease management. Realistic evaluations of disease risk are a critical first step for establishing these strategies and play an important role in decision-making. Infectious plant disease, as the interaction between at least two species, offers a particularly interesting system for evaluating the status of ecological theory in support of decision-making across scales[2].

Many plant diseases exhibit a strong relationship with meteorological variables, because pathogens often have direct exposure to environmental conditions during resting stages, dispersal between hosts, and even within hosts, where they are not buffered from environmental conditions. The occurrence of plant disease is driven by three factors: a susceptible host, the presence of a competent pathogen (and vector if needed), and conducive weather[3].

Plant disease is commonly studied for its impact on crop yields in order to devise control strategies; overall yield losses of 10% due to plant disease are typical, and in severe epidemics total crop loss may result[4]. Losses of ecologically important plant species due to invasive

pathogens are another important potential impact[5], such as the decimation of the American chestnut by chestnut blight (caused by *Cryphonectria parasitica* (Murrill) Barr)[6] and the great impacts on Australian plant ecology caused by the generalist pathogen *Phytophthora cinnamomi*[7].

Because of its impact upon human activities, plant disease risk is studied in a variety of ways, where the definition of disease risk may depend upon the situation. First, disease risk can be measured as the probability that a disease or disease complex[8] is important at a particular point in space and time, where importance might be defined in terms of expected yield loss, expected introduction of mycotoxins to foods from pathogens, or other expected impacts on ecosystem integrity or ecosystem services. The capacity to assess the last impact requires the relationship between disease risk and impacts on ecosystem services to be well-understood. Second, disease risk can be defined in terms of disease intensity, as the estimated level of disease severity (e.g., the area of a plant or leaf affected by disease), or disease incidence (e.g., the amount of plants or plant parts affected by disease). This perspective emphasizes the level of disease rather than the impacts on yield or ecosystems services.

Third, disease risk can be defined in terms of the probability of an extreme event, such as unusually severe crop losses or threats to the survival of a plant species. This perspective emphasizes "worst case" scenarios and could include cases such as the introduction of an invasive pathogen into a vulnerable agricultural region[6].

The majority of plant disease epidemiologists emphasize epidemiological processes within fields, often because of the utility of this approach for the majority of stakeholders who are interested in short-term management decisions. With this synthesis I hope to provide perspective on the possibilities for expanding research scale for plant pathologists, and the analytic tools that are already available or that need to be developed. For geographic scientists already working at national and larger spatial scales, I provide a synthesis of the biological issues involved in estimating plant disease risk.

Establishing more links and shared approaches between these groups of scientists is particularly important as the scientific and policy communities grapple with an array of related problems: diminished natural resource availability per person, increased species extirpation and extinction risks under human population pressure, and global change factors such as climate change and accelerated rates of species invasions. Plant disease and its management are important components for strategies in each of these areas.

My objectives are to synthesize approaches to plant disease risk mapping. I also address how risk factors can be combined to produce better estimates, with examples of new applications and a conceptual framework for future approaches.

1.2 Maps of disease and other ecological traits

Disease maps, which have been produced for several centuries, differ from disease risk maps, in that the former are intended to communicate the current or prior occurrence of disease, rather than predicting a future risk associated with a change in the current situation. Early examples of disease maps include plague incidence in Italy in the 1690s, yellow fever in the US in the 1700s and 1800s, and perhaps most famously, John Snow's maps of cholera in London from 1854[9,10]. Snow is credited for drawing a map of deaths resulting from the 1854 Broad Street cholera epidemic in London, and as a result, removing the pump handle from the offending well. While Snow did not necessarily use the map to generate his hypothesis about the role of water, he did use his map to support his hypothesis that cholera was transmitted in the water. As Brody et al. state the mere act of seeing data arranged graphically in space yields no new understanding without the support of pathological theory[11]. This illustrates an aspect of mapping: maps can help us interpret information about disease risk, but they do not tell us how to use the information that they present.

At a larger scale, efforts to map the Earth's surface in a manner that is useful for delineating areas of biological commonality also have a long history. In 1905 Herbertson attempted to describe rational subdivisions of the Earth's surface, which were distinct from

political divisions[12]. He attempted to include many different considerations in his natural regions: physical characteristics, climate, vegetation, and even human population density. Later, Omernick defined ecoregions of the conterminous US for research and management purposes, but cautioned that the ecoregions should only be used at the resolution at which they were compiled and presented[13] . That is, there is less variation within the large ecoregions than between, however, using the definitions of ecoregions at a resolution smaller than the ecoregion itself is not suitable.

Omernik's admonition that the regions should only be used at the resolution for which they were compiled and presented is important. Information cannot be gleaned from a map at a lower resolution than the map presents, so trying to base decisions on risk at a lower resolution than the map represents will not yield satisfactory results. Regions defined by climate classification schemes, indicating the likely locations of major natural ecosystems and agricultural systems, have been used to determine the likely productivity of an area and the effects of research and development allocations on natural resources by the Consultative Group on International Agricultural Research (CGIAR) centers[14,15]. A new phase of research including CGIAR centers addresses global factors reducing crop productivity, making use of GIS (Geographic Information System) data layers (Wood, personal communication).

1.3 Scale and disease risk mapping

Maps are a powerful way to convey information quickly. For plant disease risk they are particularly appropriate because of the high level of spatial and temporal variation[16]. The contribution of disease risk maps to decision-making occurs across several scales (Figure 1.1); with relevant scales reviewed in Yang[17] and Savary et al.[8]. For policy making, formulation of strategies for quarantines or disease management, and for research prioritization, it is important to know where the combination of susceptible host, competent pathogen, and disease-conducive weather will co-occur to produce disease that will have important impacts

on crop yield or on the biological integrity of natural ecosystems. Maps for this purpose are generally at a much coarser temporal scale (months, years, or decades), and spatial scale (countries or continents). Such maps may contribute to decisions about responses to invasive pathogens, in terms of likely points of introduction, areas where climate is conducive to establishment, and the connectivity of host populations as it may influence invasion processes (e.g. Magarey et al.[18] and Margosian et al.[19]). For tactical decision-making by individual land managers or food processors who may be confronted with contaminated products (e.g., De et al.[20]), the time-frame is often much shorter (weeks or days) and the spatial scale must be smaller to retain accuracy and precision[8,17]. However, many of the same tools that are useful for regulatory decision-making are also useful for real-time decision-making processes. For example, maps of host connectivity[19] can be useful for long-term planning as well as responding to new pathogen introductions.

In ecological modeling, scale may be a challenging aspect of sampling and statistical analysis, but also a source of interest as an object of study in its own right (e.g., Borcard[21]). Scale is an important issue for applications such as plant disease forecasting under climate change, because plant disease epidemics occur at scales much smaller than that of prediction for most climate change models[22]. Information about the scale at which processes were measured is needed for comparison of studies[23]. Because the relevant scale of ecological processes for infectious diseases involves at least two organisms, knowledge of the scale of pathogen dispersal needs to be combined with information about the range of the host and the distance between host populations[23] for evaluations of large-scale processes. Models integrating three scales of host plant populations – infection sites, leaves, and plants – have been developed to examine the effects of scale on disease incidence at different spatial hierarchies[24]. This general model could be applied to other organizational scales or ecological systems making a model like this a powerful tool for investigating disease patterns at different scales.

Scaling disease ecology from fields to regions and beyond offers many challenges, not the

least being the reduced availability of experimental data at larger scales. The metacommunity concept is useful when considering plant disease since disease is a factor two or more species interacting at many scales[16]. The metacommunity concept extends metapopulation models to incorporate more directly the potential movement of multiple species among their respective populations as they interact with other species[25]. The interaction between species is influenced by the landscape within which the species occur. Characterization of a landscape's degree of fragmentation may be useful for predicting dispersal of pathogens, particularly in cases where pathogens migrate with seasonal changes or are invasive in new areas[26,27].

1.4 Data types and data quality

The type of analysis that can be carried out to produce a risk map depends on the data model that is used. Geographic information system databases may contain many different types of data; raster, vector, one, two, or three-dimensional, qualitative, categorical, or quantitative data. Risk maps can be developed using vector data or raster data models. The vector data format uses lines which have defined beginning and end points and meet at nodes, to represent spatial features, and as such is better for displaying discrete spatial data. Examples of vector data include geopolitical boundaries, points where an observation was collected, or roads. For example, Margosian et al.[19] applied network models with vector data to analyze the connectivity of the American agricultural landscape. The raster data model is effective in representing continuous data over a landscape surface. Examples of raster data include meteorological data, digital elevation models, and degree of soil erosion. The raster data model employs a regular grid to represent the area of interest and each cell is assigned a specific value. An example of a plant disease risk model that uses the raster data format is the Internet-based Wheat Fusarium Head Blight Prediction Center map[28].

Caution should be used when selecting data sources to use in a GIS. Often several data sources are used within a GIS in complex models to generate output upon which critical

and politically sensitive decisions may be based. However, spatial data of suitable quality may not be available[29]. The error associated with each data source may be propagated and multiplied through the model to the resulting risk map. Sources of uncertainty should be understood and dealt with to create useful disease risk maps, and methods have been developed to address uncertainty[30,31].

1.5 Spatial pattern and disease risk

Historically some disease risk maps were based on knowledge of prior exposure of a subject relative to the whole population[32]. These maps provided local or regional estimations of relative risk for a population, and illustrated geographical variation, but were not capable of predicting risk when the risk factors were unknown[9]. Now with more advanced ecoinformatic methods, it is becoming easier to predict the risk to individuals when the risk factors not known *a priori*[9]. The concept of exploratory relative risk mapping is predicated on the idea that disease will tend to cluster if there is spatial variation in risk, due to variation in the intensity of unknown risk factors[9]. Wu and Subbarao used cluster analysis of weather variables to identify areas in the Salinas Valley of California that were at higher risk of lettuce downy mildew[33]. Farmers can use risk maps generated by this analysis to tailor their fungicide application schedule: in areas of higher risk a conservative approach might be advised, while in areas of lower risk, fungicide use can be aggressively reduced to cut costs[33]. In some cases, the relative importance of particular sources of inoculum within landscapes can be identified. Zwankhuizen et al. explored the sources of potato late blight epidemics, concluding that potato cull piles were sources of early inoculum and organic farms were later sources of secondary inoculum[34].

Analyses of animal systems provide additional interesting examples of the incorporation of spatial pattern in modeling. In an applied example, Boender estimated that poultry farms that were closer to another farm infected with avian influenza (0-2 km) had a higher chance (1%-2%) of infection than farms that were >10 km away (0.05%)[35]. Exploratory

risk maps can help elucidate spatial patterns related to the host, agent, and environment. Berke (2005) suggested using isopleth maps to display disease risks over chloropleth maps. Chloropleth maps display aggregated data over discrete predefined regions. Isopleth maps of disease risk more useful for illustrating disease risk due to unknown factors over a region with less visual bias. However disease data are ordinarily collected as discrete spatial point data and irregularly distributed over the landscape, requiring spatial interpolation for isopleth mapping[9].

1.6 Disease risk maps based on meteorology

Long-term climate maps or short-term weather maps may be equally important for disease risk mapping, depending on the goals of the analysis. Weather is often the most rapidly changing of factors that can trigger important plant disease epidemics; thus, many disease risk maps for within-season decision-making have been developed based on weather variables[3,28]. Disease risk maps based on weather or climate variables are effective when it can be assumed that weather or climate is the limiting factor, while the host is present and susceptible to the pathogen and inoculum is not limiting. Many agricultural production systems meet this criterion at least part of the year where large areas of monocultures are grown. Disease risk maps based on weather are also effective from the point of view that even if the map-maker does not know if the host is present, but can assume that inoculum will not be limiting, individual land managers who use the maps are aware of which hosts are present in their own land. At larger time scales, prediction of the effects of climate change on plant disease risk is one important motive for evaluating and mapping the relationship between climate and disease[36,37]. Models of plant disease have been developed that can incorporate sophisticated general circulation models (GCM). But GCM output will often need to be scaled down for forecasting plant disease interactions[22]. However, in some cases coarser large-scale climate indices are better for prediction of ecological processes than local weather, when local weather has too much noise to be useful[38].

For large-scale strategic analysis of disease risk based on climate, coarser spatial and temporal evaluations are relevant. Hijmans et al. linked two plot-level disease prediction models to a climate database in a GIS to predict the severity of global late blight (caused by *Phytophthora infestans* (Mont.) de Bary) on potato (*Solanum tuberosum* L.)[39]. This novel approach to predicting disease severity made it possible to compare and contrast the differences in predicted disease risk and actual management practices derived from survey data. The results of studies such as this can lead to improved interpretation of survey data[39]. An example of a risk map for potato late blight based on point-by-point analyses of temperature and humidity is given in Figure 2. These late blight forecasting systems use the daily average temperature, hours of leaf wetness (estimated using relative humidity), and the host's level of resistance to estimate disease risk[39].

For smaller-scale tactical decision-making within a season, disease risk maps may still be generated for larger regions, allowing individual land managers to reference their own locations on the map. Thomas et al. described a GIS created to map disease and insect risk, and crop cultural requirements using ground-based weather from local stations, plant-stage measurements, and remotely sensed imagery for commercial crop decision-making in California, Washington, Oregon, Idaho, and Arizona[40]. Since the relationship between weather variables and risk is different for different diseases, such an integrative approach requires many different models to predict risk for the entire disease complex. Using intuitive color-coded maps to indicate risk increases the speed with which the growers can learn to use the maps and evaluate updated maps distributed via the Internet[23]. When combined with disease scouting reports, disease risk maps can be useful for letting pest managers know what diseases have been sighted, where disease is likely to occur, and where inoculum is likely to be present[23].

Another example where short-term risk maps for individual farmer decision-making are useful is Fusarium head blight (FHB), primarily caused by *Gibberella zeae* (Schwein.) Petch (anamorph: *Fusarium graminearum* Schwabe). FHB has proven difficult to control in

wheat (*Triticum aestivum* L. em. Thell) in North America, where epidemics can cause severe losses through a direct reduction in yield and loss of wheat quality[20]. Mycotoxins produced by *G. zeae*, such as deoxynivalenol, necessitate that infected kernels be removed from the harvest, increasing cleaning costs. To disseminate updates regarding FHB risk, an Internet-based risk map was developed for 24 states east of the Rocky Mountains (http://www.wheatscab.psu.edu/index.html)[28]. Maps such as this provide farmers with real-time information to make management decisions and allow grain handlers and food producers to prepare for the potential for mycotoxin contaminated grain before the grain is delivered to them[20].

When it can be assumed that inoculum is not a limiting factor, disease severity may be predicted using meteorological data for an area where a host is known to occur. In a model to predict soybean rust (caused by *Phakopsora pachyrhizi* Syd.) severity during the soybean *Glycines max* (L.) Merrill) growing season in Brazil, precipitation explained 85-93% of the variation in disease severity at the end of the season[41]. Such models work well for alerting farmers to scout for disease development or, in the case of soybean rust, to prepare for fungicide application since the host can be assumed to be susceptible.

Frequently, the interaction of temperature and moisture is used to predict plant disease occurrence[28]. These data are often readily available, though often with a coarse resolution both geographically and temporally, and they may vary in quality depend upon the region where the data were collected[42]. For many biological processes, temperature has a large impact on disease development and temperature data are generally readily available at some scale. However, moisture in the form of relative humidity and free leaf wetness are critical for infection by some pathogens as well[43]. Relative humidity is well defined, but it is not as easy to define leaf wetness duration or to gather data to characterize leaf wetness, because various portions of the canopy may be wet or dry at the same time[43]. These challenges have been met by plant pathologists through several methods for measuring leaf wetness or estimating it from more commonly available measures[43–45].

The above applications of climate and weather for the generation of disease risk maps are typical in that risk is generated point-by-point without incorporating information about the risk at neighboring points. This approach is illustrated (Figure 1.2) for a scenario where there is a single climate variable of importance and a climate threshold below which there is no risk and above which risk increases linearly with the increasing climate variable. This approach may be adequate for many applications, but an obvious climatic component that motivates inclusion of relationships between locations is wind speed and direction, in terms of both average values and maxima[46]. Incorporating wind in models provides a link between previous disease locations and predictions of future disease locations (e.g. Isard et al.[47,48]).

1.7 Disease risk maps based on previous disease locations

Observations of historical or current disease intensity may be one of the most useful predictors of future disease risk, particularly when the pathogen and host were not limiting during previous evaluations. However, most observations of disease intensity have been collected on a large spatial resolution and have generally not been mapped, *per se*. Examples of disease data collections include Diagnostic Compendia from the American Phytopathological Society, which often include a brief statement about the range where the disease is known to occur, the International Virus Database of the International Committee on Taxonomy of Viruses, http://www.ncbi.nlm.nih.gov/ICTVdb/index.htm, and the Systematic Mycology and Microbiology Laboratory Index of Fungi Database,
http://nt.ars-grin.gov/fungaldatabases/fungushost/fungushost.cfm.

Other compendia such as the Centre for Agricultural Bioscience International data sets do include maps or the capability to directly generate maps from their datasets of known occurrences. Such databases related to the observed distribution of host and pathogen species have been used to evaluate hypotheses such as the enemy release hypothesis, which states that upon introduction to a foreign area a plant species experiences a decrease in regula-

tion by herbivores and other natural enemies leading to a rapid increase in abundance and distribution[49]. The surprising absence of maps of disease occurrence and intensity for most important diseases may be a function of the intensive effort that must be made to conduct a survey of disease. Mapping host occurrence, especially in agricultural monocultures, is much easier on a macro scale where there is often a single host species per field. In contrast, even a single plant species may have several different diseases. Even when disease severity or incidence data are recorded, it may not be completely clear where the disease occurs geographically. The failure to detect disease during limited sampling events provides little confidence that the disease is absent in the surveyed area. Maps of previously known locations for plant disease can be useful for strategies, such as decision-making about the introduction of new germplasm. However, maps of past disease occurrence are not useful for immediate decision-making within a field, as they are too coarse temporally for the current situation.

Data such as the historical extent of a disease's occurrence may be available, typically in the form of single points of data that can be represented as a simple dot map or the data can be interpolated or extrapolated to represent a much larger area. Sporadic disease occurrence can require a careful repeated approach to sampling, and misidentification can occur. The quality of this type of data is often lower than for meteorological or host data, and it is generally rarer. Despite the potential problems with disease occurrence data, it is useful in helping to point out locations where other factors may influence disease that are not encompassed by the host availability or meteorological data. For example, soil type may be important in disease risk for some soil-borne pathogens. Figure 1.2 illustrates a scenario where a disease has been surveyed and found to be present in two geopolitical units, but, as is typical for this type of data, another large geopolitical unit was not surveyed, or during the survey disease presence data only is taken omitting absence data. The lack of survey data for one geopolitical unit indicating presence or absence is often an issue when attempting to construct maps or models.

Pathogens outside plant hosts are sometimes sampled directly, which, combined with information about disease, host availability, and meteorological conditions, can provide very useful information about the conditions that support the occurrence of disease. Pathogen data provides information about whether inoculum is a limiting factor. It also can be used for short-term predictions about the geographic distribution of disease risk; for example, combined with other predictor variables, trapping soybean rust spores during the disease's annual progression from the southern USA to the northern USA can help in estimating disease risk[48]. One of the most important characteristics of pathogen populations is their competence for overcoming resistance in plant populations. New strains of pathogens that can infect previously resistant plants are often mapped at a coarse resolution. For example, new strains of the wheat stem rust pathogen (UG99) are mapped as they spread through and out of Africa to other wheat production regions[50]. In addition to information about the locations of disease or pathogens, information about vectors may also be useful for estimating disease risk. Peterson et al. used migratory birds to model the range of West Nile virus[51]. The risk of Stewart's disease of maize (*Zea mays*) can be forecast by predicting the winter survival of the corn flea beetle (*Chaetocnema pulicaria* Melsheimer) populations which transmit the bacterial pathogen (*Pantoea stewartii* subsp. *stewartii* (Smith) Dye)[52], and aphids are routinely monitored as a virus disease risk indicator for virus diseases in potato seed tuber production, http://aphmon.csl.gov.uk/.

1.8 Disease risk maps based on host availability

Host availability data often may be difficult to locate for use so fewer disease risk models have been based on host availability. The presence of host species is a logical requirement for disease risk. Reliable host availability data can be attained from some sources for agricultural species maps, such as the US National Agriculture Statistics Service (NASS), and the USDA Natural Resources Conservation Service Plants Database (http://plants.usda.gov). Figure 1.2 illustrates a scenario where the risk of disease is minimal for low levels of the host,

but increases rapidly to achieve a high level of risk when host availability is no longer limiting. Often the collection methods for host availability data are similar to those for disease occurrence data, and may be incomplete, or only positive occurrences may be recorded while negatives are unknown or difficult to interpret. Another important consideration is the timing of host availability; annual plant species may be available only part of the year, and even perennial species may be available in a susceptible growth stage only part of the time.

The effect of small-scale host heterogeneity has been illustrated dramatically in the case of rice blast (causal agent *Magnaporthe grisea* (Barr)) management through mixtures of higher-value susceptible rice and lower-value resistant rice[53]. The effects of host heterogeneity are not as consistent in all systems (e.g., Garrett et al.[54]); in some cases the characteristics of the host-pathogen system can be used to predict the magnitude of effects[55,56]. Information about host availability can be evaluated point-by-point (field-by-field), but also lends itself to evaluation of the potential for pathogen movement through the host landscape[57,58].

Graph theory is becoming more commonly used in ecological applications[59]. In the graph theory framework, disease can be modeled as traveling along paths (edges) between host individuals or populations (nodes). Jeger et al. suggested the use of epidemiological models based on networks to study individual hosts as a set of vertices[60]. This helps elucidate management strategies which provide effective control through better understanding of how the disease spreads over the host landscape. Networks can be used to study the effects of a fragmented landscape on disease spread compared to a homogeneous landscape[60]. Using a network modeling approach in a GIS, Margosian et al. have applied graph theory to the US agricultural system[19]. County-level USDA NASS data about crop abundance were used to construct networks of four major crops in the US, maize, soybean, wheat, and cotton[19]. Determining the scale at which crop species plantings become discontinuous (disconnected) enough to disrupt pathogen transmission can help target regions where a rapid response may slow or stop disease movement, and inform policy to enhance agricultural landscape

heterogeneity.

Networks based on pathogen movement between host populations can be improved by including information about pathogen starting locations and about how climate or weather influences rates of reproduction and movement. If neither pathogen availability at starting points nor climate or weather suitability is limiting, it may be reasonable to assume that a network based on host availability will predict where disease will develop. Information about the role of different host species at different time points may also improve risk estimation (e.g., Isard et al.[48]).

1.9 Disease risk maps based on climate envelope

Climate envelope models (also known as species distribution models and as ecological niche models) combine information about previously mapped occurrence of a species with information about the climate (and potentially other) characteristics of those areas of occurrence, in order to predict additional locations where the species could potentially survive.

Most of these models focus on the use of presence or absence data from the species' known range in combination with other environmental data known or thought to have an effect on the species survival in its native range[61]. Common statistical approaches used in climate envelope modeling include machine learning methods such as Maximum Entropy, Boosted Regression Trees, as well as General Additive Models and General Linear Models (often logistic) and distance measures[62].

Climate matching models such as CLIMEX[63] were designed to determine species' likely distribution or abundance and elucidate factors that limit the species' distribution[23]. CLIMEX includes a global meteorological database, is process driven[64] and has been applied to predict new locations for plant pathogens[65,66]. Desprez-Loustau et al. used CLIMEX to examine the potential effects of climate change on forest pathogenic fungi in Europe, noting that studies such as this cannot account for pathogen adaptation to climate change[67]. Other approaches to map the likelihood of an invasive species' establishment include the North

Carolina State University – United States Department of Agriculture (USDA) Animal and Plant Health Inspection Service (APHIS) Plant Pest Forecasting (NAPPFAST) system[18]. The NAPPFAST system is used by the USDA APHIS for mapping the potential geographic ranges of exotic pathogens and insect pests. It offers a web interface for the user to link meteorological and geographic templates for biological modeling[18].

Often models of species distribution based on climate are accurate in predicting current species distributions. Models like these may be useful because in many cases the distribution of a species is not systematically sampled to provide presence/absence data and absence data cannot be inferred with certainty. It is also important to realize that these models often assume biotic interactions are unimportant[68]. These methods have not been fully tested when applied to forecasting invasive species or the effects of climate change[69]. Do biotic interactions play an important enough role to help to define the organism's niche? If so, then the niche may be different in an exotic range versus the native range. A species may become invasive in a new geographic area due to lack of natural enemies[49,68].

Predicting plant disease is different than predicting the occurrence of a single species because infectious disease is always an interaction between two or more organisms. Both a suitable host and a competent pathogen (and vector if necessary) must be present, combined with conducive weather conditions to support the occurrence of disease. Because pathogen species have shorter generation times than plant species, they have the potential to adapt to new conditions much more quickly than the host species. When including the host in a map of disease risk, the host population may or may not be homogeneous in its susceptibility to disease, contributing to spatial and temporal variation in disease occurrence[16].

These mapping approaches can be used to quickly create information for disease control, bringing risk areas into focus for closer inspection. NAPPFAST has been designed to quickly create a scenario of first guess risk maps at the expense of resolution. These systems also rely mainly upon climate information to derive the disease risk at a certain point rather than incorporating the presence of the host, diseased host, inoculum levels, or pathogen

dispersal. Incorporating real-time movement or inoculum levels into climate-matching maps could enhance their accuracy. These systems are not suited for pathogenic species where climate is not the key determinant for the species distribution.

1.10 Combining factors in disease risk mapping

Climate, historical disease, and host maps may each be useful alone for predicting risk when the other factors are not limiting. As discussed above, these factors have also been combined in various ways. Climate envelope modeling is based on use of historical disease occurrence, where the boundaries of historical occurrence in one region was assumed to be defined by an environmental factor, so that the same factor can be used to predict occurrence in new regions when inoculum and host are no longer limiting. Other approaches to risk mapping may use host species abundance along with meteorological variables. Meentemeyer et al. used a rule-based model to describe forests in California that were likely to be affected by sudden oak death (SOD), caused by *Phytophthora ramorum* (Werres, De Cock & Man), for monitoring purposes[70]. This is a somewhat different approach, in that many models previously used host presence only to indicate where the pathogen species could occur. In this model, risk based on host abundance is weighted by experimental data and expert input. The model matched ground-truthed observations, but there were many areas mapped as high risk that did not have any current disease observations. Meentemeyer et al. suggested that this indicates that large areas are in danger of future SOD epidemics[70]. But as they point out, typical data limitations complicate interpretation.

Pathogen dispersal models may be combined with meteorological conditions for disease risk assessment[46,48]. Pan et al. introduced a model for the prediction of disease caused by wind-dispersed pathogens using a particle dispersal model, Hybrid Sing Particle Lagrangian Integrated Trajectory Model (HYSPLIT_4), and a regional climate prediction model, the Penn State/National Center for Atmospheric Research Mesoscale Model (MM5)[71]. This approach can be applied to other plant pathogens that are aerially transported and could

be used for an early warning and detection system for soybean rust.

Remote sensing approaches are also being applied to disease risk mapping and offer the potential for automating otherwise labor-intensive sampling procedures. For example, remote sensing has been used to identify and analyze oak wilt disease (causal agent *Ceratocystis fagacearum* Bretz) Hunt) Everitt[72] and rice blast[73]. Landscape features can be linked to disease risk estimates as well[74,75]. Ultimately it may be possible to identify patterns in remote sensing data that lead to reliable inference about the presence of plant disease, in which case such data could be used to strengthen risk mapping based on historical occurrence of disease[76]. Applications in precision agriculture may make use of newly generated maps of disease intensity within management areas to guide the appropriate application of pesticides[77].

In general, multiple disease risk maps based on individual factors can be combined in different fashions to create a single map. First I describe potential point-by-point analyses that combine the host, disease, and climate data layers, in an approach analogous to decision-making rules based on evaluation of risk factors. Methods for combining factors include a *limiting risk factor* approach where the risk at a point is based on the lowest risk identified by any of the risk maps included in the creation of the final combined risk map. For example, if the host is rare in a region, this may result in low risk despite highly conducive climatic conditions. Assuming the system is understood fairly well and the maps are accurate, this would point out areas known to be of higher risk. This approach would be motivated by a desire to protect against a type I error in the sense that it would avoid identifying high risk where there is an indication that risk is lower based on at least one risk factor. Second, using the *highest risk factor* would mean that the risk at a point is based on the highest risk identified by any of the risk maps (Figure 1.3). This method is a good evaluation of the worst case scenario, if limiting factors are removed or if limiting factors were misunderstood. This approach would be motivated by a desire to protect against a type II error in the sense that it would avoid identifying low risk where at least one factor indicates that risk is higher.

A third method, *weighted risk factors*, evaluates the risk at a point as a weighted mean of the risk identified by the single-factor risk maps. This assumes that enough is known about the system to place greater weight on some potential risk factors than others; or that some risk factors are measured with substantially more error than others, where those with higher error would be given lower weight. Bayesian analyses may be used to incorporate different types of knowledge about the risk factors (e.g., Meentemyer et al.[70]).

Another approach for combining risk factors on a point-by-point basis is through mechanistic modeling, such as the use of disease progress curves[78], when enough is known with confidence about the system to formulate a more detailed expression of disease risk as a function of the component factors (Figure 1.2). Where there is sufficient knowledge, this approach may give much more realistic estimates of risk. In this case, the risk at a point is a known (potentially complex) function of the risk factors, based on knowledge about the nature of the biological interactions between the factors. For example when evaluating disease risk landscapes in a growing season, historical pathogen distributions may provide model input relevant to likely regional inoculum loads, and so likely starting values for epidemics (Figure 1.2). Weather conditions may also contribute to inoculum loads through impacts on the probability of overwintering of pathogens and vectors, and will have an important impact on the rates of disease increase during periods of infection. Host availability will also influence the rate of disease increase, and will determine the carrying capacity for disease in a region.

Once a disease risk map has been developed on a point-by-point basis, it may be possible to improve it by incorporating risk neighborhood analyses. If network models can be applied, this would be one way to capture the influence that the level of risk at one point can have on neighboring points. As another example, Wimberly et al. incorporated spatial autocorrelation or spatial heterogeneity into predictive disease models for the geographic distribution of tick-borne pathogens[79]. This was more effective than using a standard logistic regression and could be useful in predictions for plant diseases that exhibit strong spatial

patterns. A location with neighboring sites of high disease risk can reasonably be expected to exhibit a higher disease risk than if it were surrounded by sites of low disease risk[79]. A general approach to risk neighborhood analysis could be to map regions at the interface between higher and lower risk areas as having a risk influenced by both areas (Figure 1.3). The size of the affected area at the interface would be a function of the probability of pathogen dispersal across the relevant distance at the time scale being considered. This approach incorporates the potential effects of inoculum produced in a higher risk area on lower risk neighboring areas[80]. Ultimately the structure of risk neighborhoods at multiple scales may also be incorporated in a more mechanistic manor, as pathogen dispersal and its response to host landscapes and climate landscapes are better understood.

1.11 Conclusion

The future for plant disease risk mapping will be quite exciting, as the technologies develop further. Advances in remote sensing will facilitate evaluation of large-scale epidemic processes. Advances in genomics may also support a number of improvements in epidemic forecasting[37]. For example, genomic information about emerging pathogen species may improve prediction of their responses to patterns of meteorological variables and host distributions in a form of genome matching' with other pathogens analogous to climate matching approaches. The synthesis of ecoinformatic approaches with new pyrosequencing approaches for characterizing microbial communities (e.g., Roesch et al.[81]) will also open up an interesting research area. As pyrosequencing becomes more accessible so that characterization of microbial communities can become a typical feature of any ecological study, it will be fascinating to consider global patterns of microbial communities. This new type of information will inform analyses of global plant disease risk as well, through analysis of pathogens, themselves, and the other microbes which influence pathogens' short-term fitness and long-term adaptation, and the induced and acquired resistance status of plants. One of the most important challenges will be optimizing use of this data to help us in our long battle

with plant pathogens that damage crops and invade new ecosystems. Meeting this goal will also require continuing analyses of well-known but under-studied relationships such as the relationship between disease severity and losses in plant productivity and fitness[8].

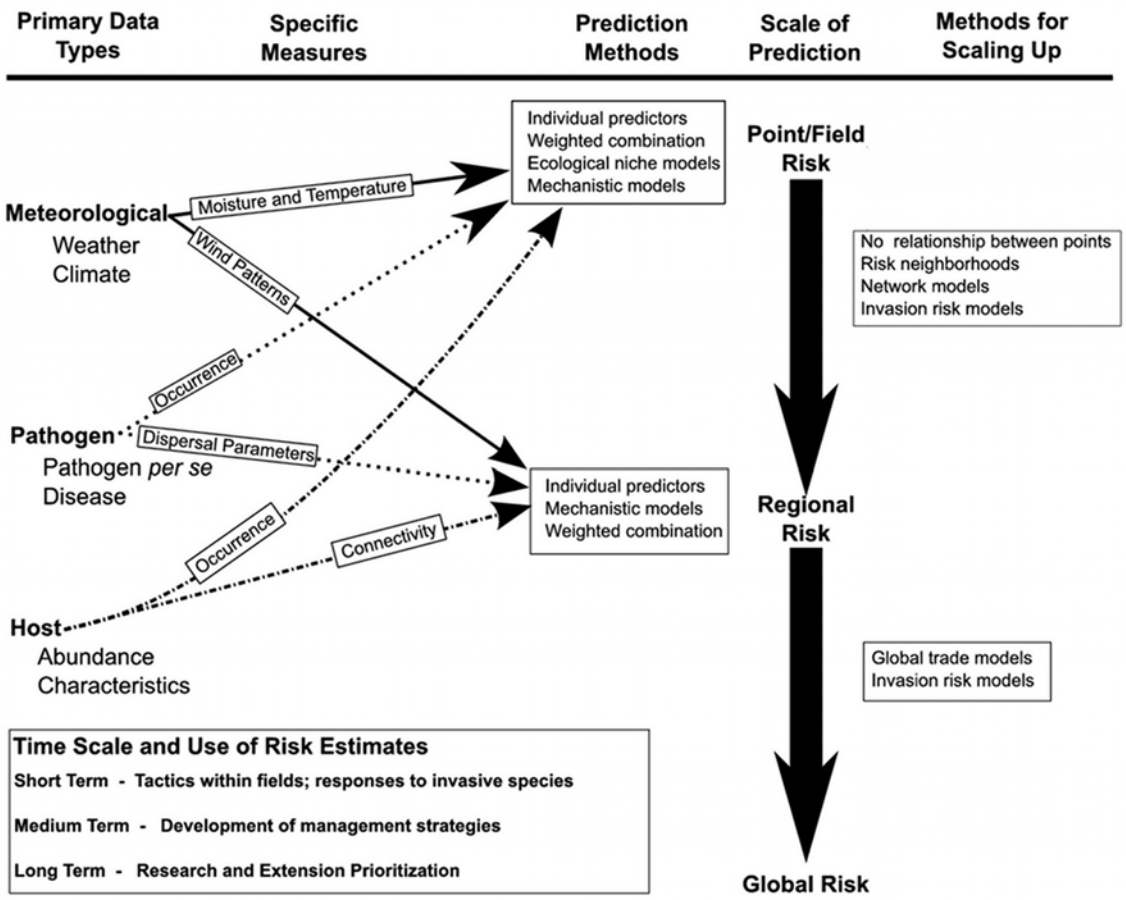

Figure 1.1: Primary data types and their use in disease risk prediction across spatial and temporal scales. Prediction models are often used with small field or plot weather data to make determinations about risk at a single 'point'. These risk estimates can then be plotted with estimates for other points for a more complete regional or global risk map. Given the predictor variables, no further relationship is assumed between points for such an analysis. Other types of analyses, such as risk neighborhoods and weighted network models, provide ways to incorporate direct risk effects from surrounding areas.
Moisture and temperature most often are useful in local small-scale predictions due to variance. Wind patterns are useful when dispersal by wind over longer distances is important. Pathogen occurrence at the local level determines whether disease can occur. Dispersal, by wind, water or other methods are useful for models in this example for illustrating regional or global movement and risk. High or low host abundance can influence inoculum loads on a point or field scale. Regional or global abundance and connectivity are useful for determining the risk of movement and infection by a pathogen through the population.

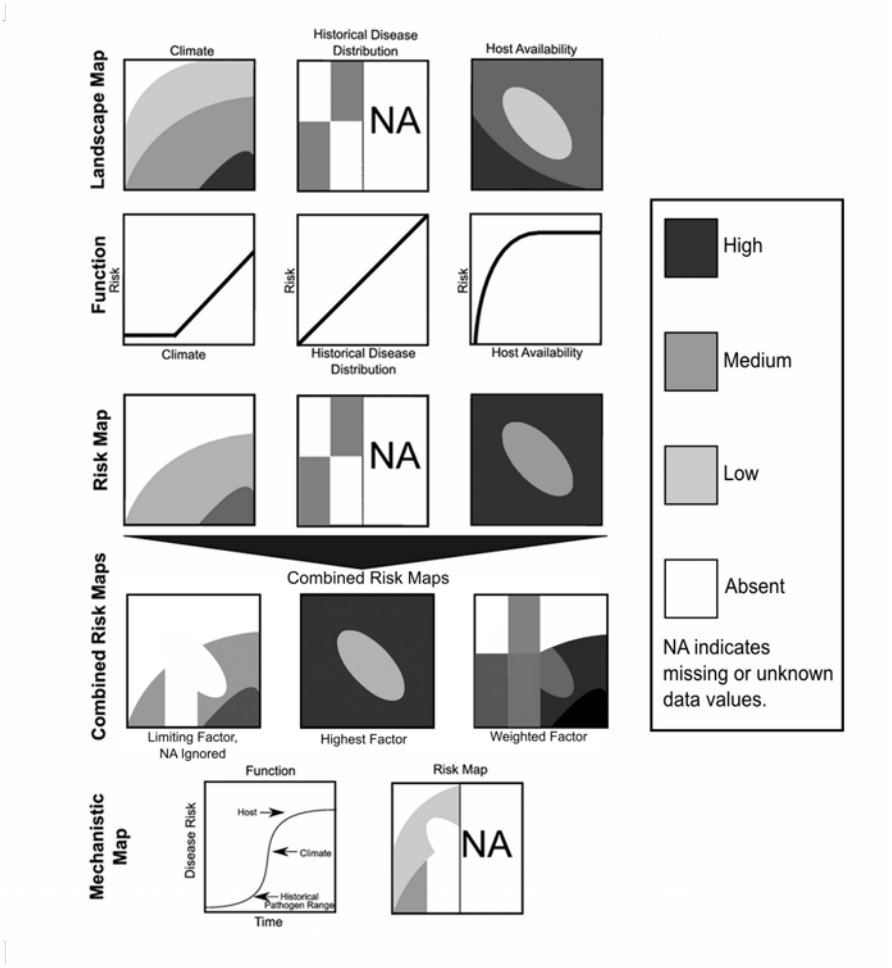

Figure 1.2: Hypothetical landscapes illustrating how maps of climate, historical disease, and host distributions can be used to produce disease risk maps for a continuous variable for a raster data format. Darker and lighter shading indicate higher and lower values of variables, respectively. NA indicates information is not available for that area. The first column illustrates a scenario where disease risk increases with increasing values of a climate variable once a threshold level of the climate variable has been reached. The second column illustrates a scenario where historical disease distribution has been assessed in some geopolitical units and is used as a direct estimate of future disease risk. The third column illustrates a scenario where disease risk rises rapidly with host availability and reaches a plateau. The lower part of the figure illustrates approaches for preparing combined-factor disease risk maps from single-factor disease risk maps. The limiting factor approach is based on identifying the lowest risk level among the factors at each point. The highest risk factor approach is based on identifying the highest risk level among the factors at each point. The weighted factor approach is based on weighting factors depending on their known importance or the coincidence with which they were estimated. In this example, the different risk types were weighted equally when information was available. When enough is known about the system, combined-factor disease risk maps can be developed using mechanistic models.

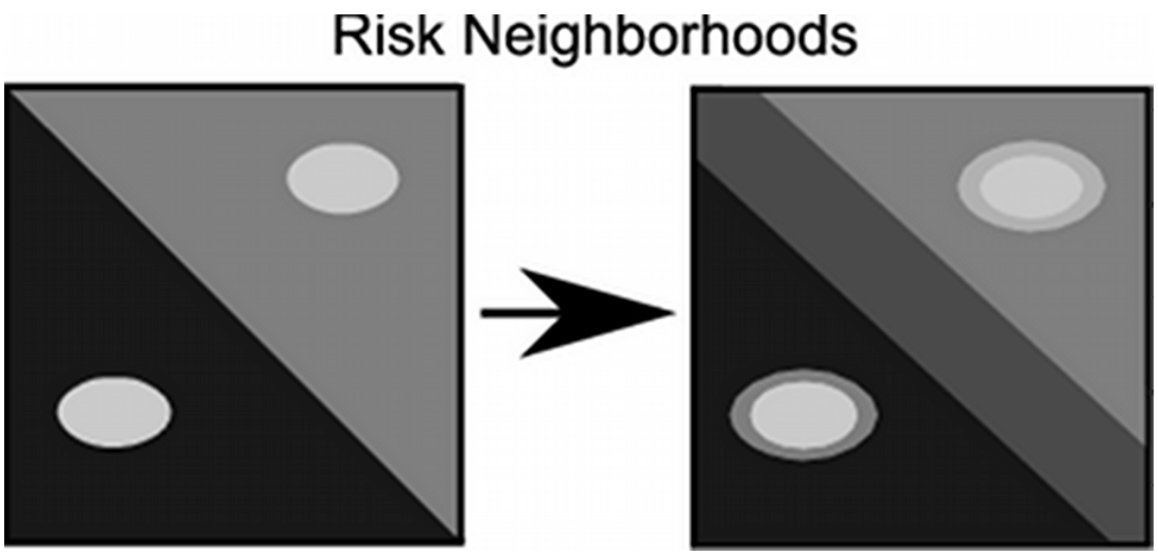

Figure 1.3: Neighborhood risk models can be used to incorporate the influence of neighboring areas in disease risk maps, as illustrated in these hypothetical risk landscapes. Darker and lighter shading indicate higher and lower values of variables, respectively. Where areas of high risk (dark) meet low risk (light) the risk is estimated to be a function of both areas as a gradient rather than a sharply defined line between areas of risk.

Chapter 2

Using Metamodels to Adapt Hourly Time-Step, Field or Plot Scale Models to Monthly Time-Step, Global Scale Models

2.1 Introduction

Many plant disease forecasting models exist for purposes of assisting in disease control efforts by producers[28,82–84]. The majority of models developed by plant disease epidemiologists emphasize short-term forecasting models because they are most useful to the majority of stakeholders interested in management decisions. The models I have developed are useful for long-term decision-making for research resource allocations, and other policy decisions where it may be necessary to look several years into the future across large areas (e.g., countries, continents, or global). Due to the detailed input data required by most disease forecasting models, weather data availability may be a limiting factor for disease risk models in some areas. I have adapted an existing disease forecasting model for large-scale, long-term decisions about policy and research priorities using readily-available gridded weather data sets such as WorldClim[42] or CRU CL 2.0[85]. This framework for disease forecasting allows use of available weather data sets when quick responses are needed to address problems such as exotic pathogen introductions, or long-term problems such as potential changes in risk

under climate change.

Potato late blight (causal agent *Phytophthora infestans* (Mont.) De Bary.) forecasting models recognize the importance of temperature and moisture in disease development, but use different combinations of these variables for forecasting. The earliest of these types of predictive models that was useful for predicting late blight risk were the Dutch Rules postulated by Van Everdingen in 1926[86]. Since then many different models have been developed for late blight infection in many different regions. Wallin developed a system to predict initial late blight and the subsequent spread using relative humidity and temperature to calculate seasonal accumulation of disease severity values, numbers assigned based on specific combinations of intervals of relative humidity greater than 90% and the average temperature during those periods[87]. The Blitecast system is based on daily rainfall and maximum and minimum temperatures to forecast late blight occurrence[88]. Fry et al. developed SimCast which included modifications to include cultivar resistance[89]. Grünwald et al. further refined the SimCast model for potato varieties with moderate to high resistance and validated it in a highland tropic setting, which demonstrates that it is possible to apply a model developed in a temperate climate more broadly[82]. Hijmans et al. estimated the number of sprays globally necessary to control late blight by developing a method to utilize two disease forecast models, Blitecast and SimCast, with a climate database in a geographic information system (GIS)[39].

Other models that forecast plant disease risk for large regions for risk analysis and planning have been developed. Margosian et al. created models to predict the connectivity of the American agricultural system for pathogen movement[19]. This model is based on a network analysis of the amount of host available on a county basis using USDA National Agriculture Statistics Services data[19]. Magarey et al. developed the NAPPFAST Internet system for weather-based mapping of plant pathogens, a system which uses generic templates to help the user make a disease risk map when biological information regarding the pathogen of interest may be limited[18].

Short-term disease forecasting models mainly rely upon fine scale (e.g., hourly or sub-hourly) weather data for input. These data are often easily collected locally or provided through outreach efforts which make these models easy to use. Grünwald et al. noted that in the Toluca Valley of Mexico, because conditions are typically conducive to late blight, using a rain gauge was the only necessity in predicting fungicide applications using their fungicide units table, something local farmers could easily do. Daily or monthly time step weather data are readily available for large scales (e.g. continental or global) data sets, and climate data associated with climate change scenarios are often provided as monthly averages. There are two ways to overcome this. One possibility is to estimate fine scale weather data based on coarser scale weather data. However, this approach may not be feasible for monthly weather data to create a fine scale data set such as hourly; this approach is commonly referred to as "weather data downscaling". Alternatively, new metamodels that predict disease risk directly based on coarser resolution weather data can be developed using large fine resolution weather data sets for model development and testing.

In response to this need, I have developed a metamodel which adapts an existing potato late blight forecasting model, SimCast, for use with weather data with a large, monthly, time step. SimCast was developed to predict the number of fungicide applications needed for management with a contact, non-systemic pesticide[89]. SimCast estimates the disease risk based on hourly temperature (T) and relative humidity (RH) inputs. Blight units, a measure of disease risk, are accumulated each hour when RH is greater than 90%. The interaction of temperature and cultivar resistance determine the blight unit value accumulated.

My overall objective was to evaluate the impact of adapting the SimCast model to predict the risk of late blight based on daily and monthly weather data. My first objective for this project was to develop disease prediction models based on daily and monthly weather means and compare to results based on hourly weather data. The second objective was to compare the blight unit predictions of models constructed from weather data sets specific to potato growing regions with models constructed with a data set that represents a broad range

climate types. The third objective was to compare late blight risk predictions based on hourly, daily, and monthly weather averages to observed late blight severity data sets from four countries.

2.2 Materials and Methods

The first objective was to develop disease prediction models for use with coarse scale weather data and compare to fine-scale model output. To do this, I needed a data set with wide geographic coverage, hourly reporting, and extensive data quality control. To meet these criteria, I used the National Climatic Data Center Hourly United States Weather Observations (HUSWO) 1990-1995 CD-ROM[90] containing hourly, georeferenced weather observations from 262 National Weather Service stations nationwide (Figure 2.2). Data from the 247 stations reporting hourly temperature, relative humidity, and precipitation were selected. Blight units for each location were predicted for susceptible and resistant cultivars using the table from Grünwald et al.[82].

Figure 2.1: Locations of weather stations within the US Hourly Weather Station Weather Observation data set. Alaska, Hawaii, Guam, and Puerto Rico are not shown but were included in construction and validation data sets.

Daily and monthly weather average data sets were created from the hourly observations.

Because SimCast reports blight units as daily values no further modifications to the blight units were necessary to create daily blight unit values. To create monthly blight unit values, SimCast blight unit values were averaged to create a daily average blight unit value for every location for all 12 months for each year of the data.

I developed two generalized additive models (GAM)[91] that use daily or monthly time step weather data to predict daily or monthly blight unit accumulation, respectively. The MGCV package[92–94] was used to construct these models in R. Subsets of daily weather averages with 236,653 observations, monthly weather averages with 7,773 observations were created. Corresponding blight unit data sets from 1990 through 1992 calculated from hourly weather data using SimCast and averaged to daily or monthly values were created.

The model SimCast Daily Means used daily time step weather to predict daily blight unit accumulation and had the form

$$z_i \sim f(x_i,\, y_i,\, k = w) + \epsilon_i \tag{2.1}$$

where z is the response variable, blight units, $f(x, y)$ is the smoothed function of the interaction of temperature and relative humidity, w is 150. The second model, SimCast Monthly Means, used monthly time step weather values to predict the average daily blight unit accumulation for a given month and had the same form as SimCast Daily Means.

The first form of model validation was to create a data subset for the same locations during a different set of years than the construction data set, from 1993 to 1995, and then to correlate the performance of SimCast Daily Means and "observed" blight unit estimates from SimCast. SimCast Monthly Means outputs were also compared to monthly blight unit values created from averaging SimCast blight unit estimates to create monthly time step data using Pearson's correlation.

The second objective was to compare the performance of these models when constructed from weather data sets more or less specific to potato growing regions. A map from Hijmans et al. was used to determine the location HUSWO station locations within potato growing regions in the US[95]. Weather stations in the potato growing areas or within a distance of

10 kilometers were selected and a subset of weather data from these stations was created for use in GAM model construction as detailed in objective one.

The third objective was to compare SimCast Daily Means and SimCast Monthly Means output to disease severity observations from several countries. For the third objective, late blight severity and hourly weather data from 19 cultivar, site-year combinations in Israel, Mexico, Peru, and United States were provided by the International Potato Center (CIP). SimCast was used to predict blight units using these weather data. At locations where there were two different heights for weather stations collecting data, the weather station closest to 50cm above the ground was selected based on the weather station placement in Grünwald et al.[82]. Daily and monthly time step data were created for use with SimCast Daily Means and SimCast Monthly Means, respectively. SimCast Daily Means and SimCast Monthly Means were used to predict blight units for each of the 19 cultivar, site-year locations, and the areas under the disease progress curve (AUDPC)[3], were calculated from disease severity observations for each location. Linear regression was used to find the fit of SimCast Daily Means and SimCast Monthly Means output to the SimCast output based on hourly data, and the fit to the AUDPC for each location.

2.3 Results

2.3.1 SimCast Daily Means and SimCast Monthly Means Fit

Objective 1: Disease prediction models were developed based on daily and monthly weather means and compared to results based on hourly weather data. The model fit for construction and validation data sets for susceptible cultivars exhibited little difference (Table 3.1). For both the construction and validation set predicted blight unit values were under predicted by the GAMs when compared to SimCast output. The fit of both SimCast Daily Means and SimCast Monthly Means was similar (Figure 2.4, Figure 2.5). SimCast Daily Means under predicts the number of blight units that will be accumulated when compared to daily blight unit estimates from SimCast based on hourly weather data. SimCast Monthly

Means exhibits a positive relationship with monthly averaged blight unit values (Table 3.2). Results from potato growing regions were similar with a slightly better fit and lower AIC value (Table 2.3.1). Results were similar for models for resistant potato varieties (data not shown).

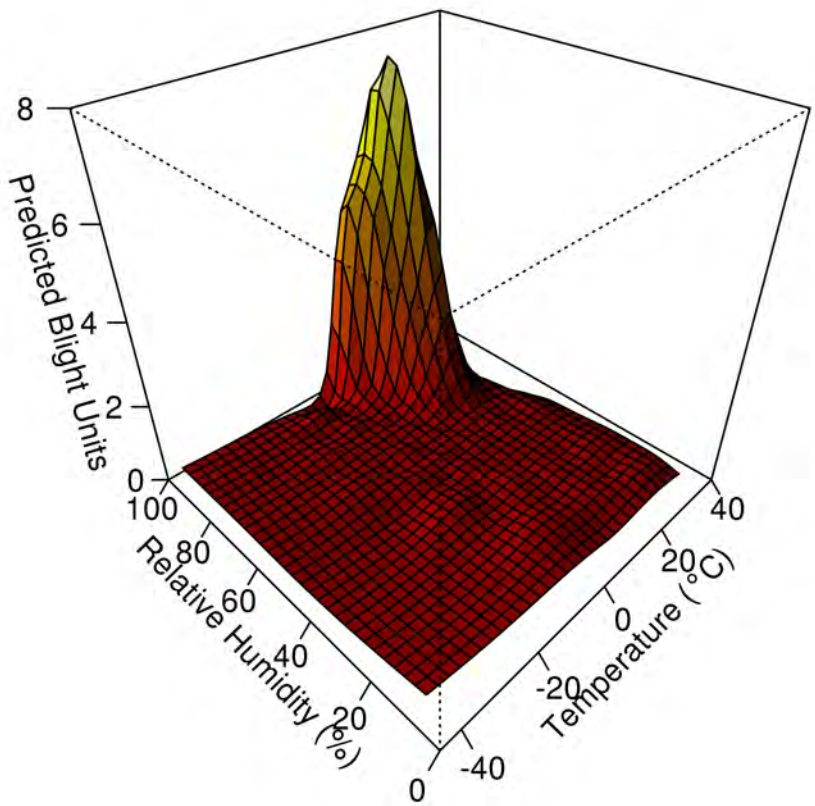

Figure 2.2: Fitted GAM surface for SimCast Daily Means. The smoothed interaction of relative humidity and temperature predicts blight units using daily weather data values.

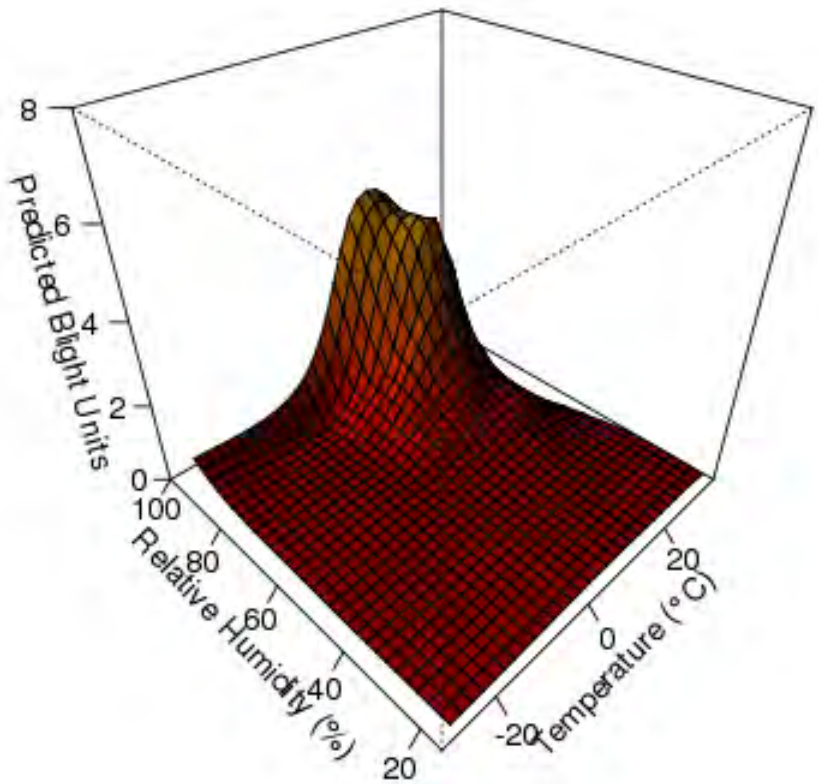

Figure 2.3: Fitted GAM surface for SimCast Monthly Means. The smoothed interaction of relative humidity and temperature predicts blight units using monthly weather data values.

2.3.2 Comparison when using potato growing areas only to construct GAM

Objective 2: compare the performance of these models when constructed from weather data sets more or less specific to potato growing regions and seasons versus the whole weather data set. The models showed little difference in performance when created from the whole US weather data set or just potato growing regions of the US. SimCast Daily Means tended

Table 2.1: Goodness of fit of SimCast Daily Means and SimCast Monthly Means for construction and validation data sets. Results for susceptible cultivars are shown, resistant cultivar models were similar.

	Daily Model Predictions				Monthly Model Predictions			
	R2	GCV[a]	AIC[b]	p-value	R2	GCV	AIC	p-value
All US	0.62	1.69	795995	<0.01	0.78	0.24	10882	<0.01
US potato regions	0.66	1.35	48314	<0.01	0.83	0.138	1314	<0.01

[a]Generalized Cross Validation score, and [b]Akaike Information Criterion are used in model selection, a lower score indicates a better model fit.

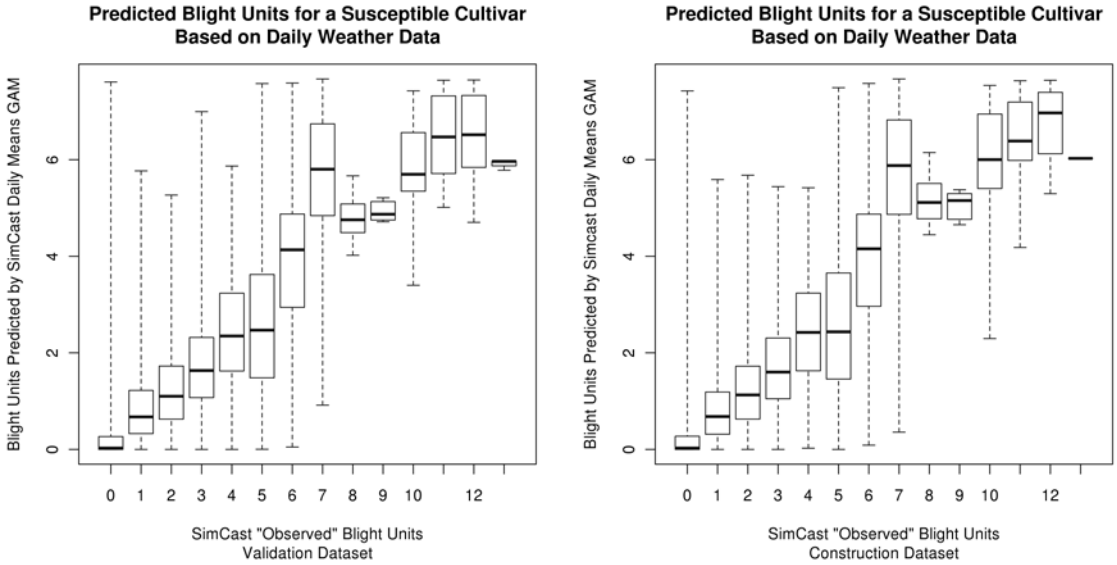

Figure 2.4: Boxplots of SimCast Daily Means fit when plotted by SimCast predictions based on hourly weather data, validation dataset (L) and construction dataset (R).

to under predict more than SimCast Monthly Means. Similarly the application of a model created using the whole US data set when applied to just potato growing regions showed little difference from a model created using the whole US data set and applied to the whole US.

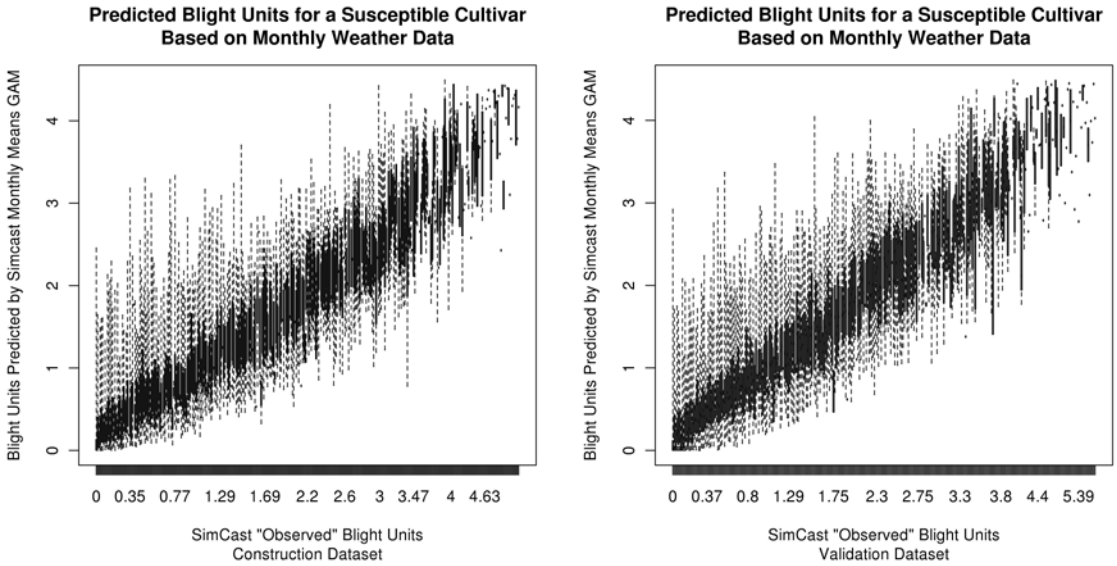

Figure 2.5: Boxplots of SimCast Monthly Means fit when plotted by SimCast predictions based on hourly weather data, validation dataset (L) and construction dataset (R). Hourly data from SimCast was averaged for a month and then data from SimCast Monthly Means was plotted against this to test fit.

Table 2.2: Pearson's correlation score of SimCast Daily Means and SimCast Monthly Means blight unit predictions with SimCast predictions. SimCast predicts blight units using hourly weather data. SimCast Daily Means and SimCast Monthly Means predict blight units based on daily and monthly weather data respectively.

	Daily		Monthly	
	Construction	Validation	Construction	Validation
All US	0.82	0.84	0.89	0.89
US Potato growing regions	0.76	0.74	0.91	0.91

2.3.3 Comparison of late blight risk predictions to AUDPC values

The third objective was to compare late blight risk predictions based on hourly, daily, and monthly weather averages to observed late blight severity data sets from several countries.

SimCast Daily Means and SimCast Monthly Means produce blight unit estimates which have a positive relationship with the AUDPC of late blight observations in Israel, Mexico, Peru, and the US. The blight units predicted by SimCast have an of R2-value of 0.31

and p>0.01 for SimCast blight units and R2-value of 0.27 and p>0.01 for blight units predicted by SimCast Daily Means when regressed on AUDPC values calculated from field observations.

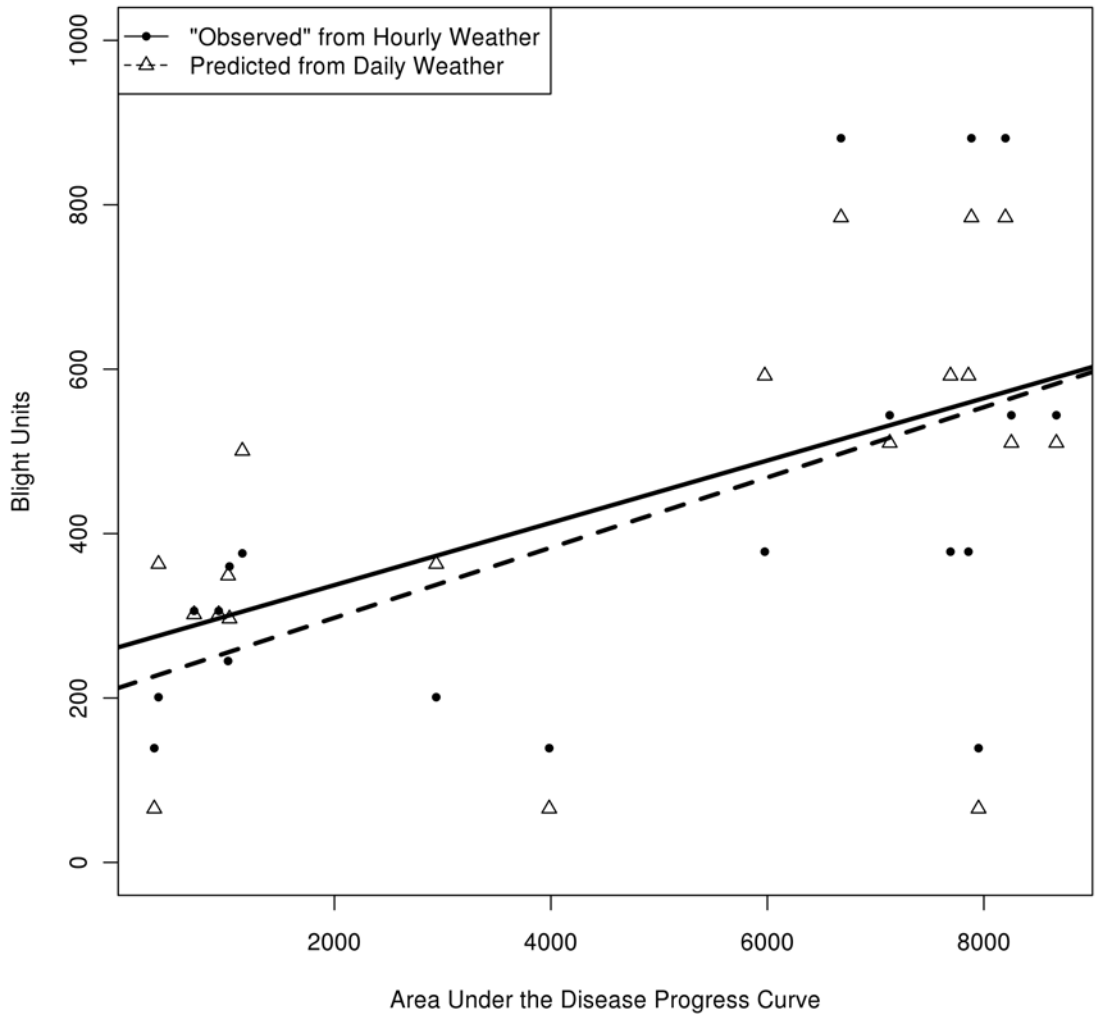

Figure 2.6: SimCast and SimCast Daily Means blight units fitted to AUDPC values from several field plot trials.

2.4 Discussion

SimCast Daily Means and SimCast Monthly Means predictions were closely correlated to output from SimCast using hourly time step weather data estimated from daily time step weather. SimCast Daily Means and SimCast Monthly Means follow the same trend as SimCast predictions but underpredict slightly. However, there is a significant relationship ($p < 1e\text{-}16$) between the blight units predicted by SimCast and blight units predicted by both SimCast Daily Means and SimCast Monthly Means respectively. SimCast Daily Mean's greater under-prediction than SimCast Monthly Means could be due to the lack of the influence of having multiple days of high RH and T together. SimCast Monthly Means would capture this because it uses a monthly average. The monthly data used to construct SimCast Monthly Means has less variability in it due to the effects of averaging. These two factors could contribute to underprediction.

Even though the models based only on potato-growing regions had slightly higher R-squared and lower AIC values, they represent smaller variety of climate types and the performance is similar to the models created using the whole US data set. Because of the similar performance between models constructed from data representing the whole US climate data set or a subset of potato growing regions, I chose to use the models constructed using the whole US weather data set. I felt that this approach represented a broader type of climates and was more suitable for use with global predictions.

The SimCast model as originally described by Fry et al.[89], and modified by Grünwald et al.[82] includes a table of fungicide units which are used in conjunction with blight units to calculate the number of fungicide applications necessary to control late blight. These fungicide units are based on chlorothalonil, a contact fungicide, which degrades over time and washes off with precipitation[89]. With the increasing use of systemic fungicides, such as mefanoxam or metalaxyl, fungicide unites will not be relevant for all scenarios. Because blight units capture the biological aspects of the disease, they are useful when comparing different areas for late blight risk rather than calculating the amount of fungicide applied

to control disease. In cases where the environment is always conducive, fungicide units are useful for predicting the number of sprays. The purpose of these metamodels however is to predict the level of relative risk based on environmental parameters, T and RH.

The number of fungicide applications possibly could be calculated using only blight units in the original SimCast model but more often it was a combination of blight units and fungicide units which indicated a fungicide application was necessary[89]. Also, a new fungicide unit table which accounts for reduced efficacy of the fungicide in the plant based on time since the last application could be derived for use with systemic fungicides. Combining these two measures could be used to create estimates of fungicide applications necessary to control late blight.

SimCast lends itself well to being adapted to models like SimCast Daily Means and SimCast Monthly Means because it is based on a table of values that indicate likely disease risk, blight units. Other systems for which this approach might work well include leaf spot of peanut, causal agent *Cercosporidium personatum*[96], and fire blight of pear and apple, causal agent *Erwinia amylovora*[97] because of their model structure similarity to SimCast. Both use thresholds which, when crossed for defined time periods trigger a response in the model. Adapting systems that use time steps other than hourly would be a similar approach. The original model would need to be used to create disease risk or severity estimations that had corresponding weather data. This combination would then need to be averaged to daily or monthly values as desired and GAMs created to use coarse scale weather data.

Methods do exist to estimate fine-scale weather data using algorithms such as described by Cesaraccio et al.[98], though this method requires more calculation and detailed information regarding the time of year, and sunrise and sunset times. Map results must then be interpolated to create a complete surface. Hijmans et al. used an approach similar to this to estimate the number of fungicide applications necessary to control late blight globally[39]. However, this approach required time intensive calculations to estimate hourly time step weather values from daily weather observations for use in SimCast and Blitecast.

The use of monthly time step weather data for model input as described in this paper gives greater flexibility in the data formats that may be used. Text files can be used to generate specific point or plot disease severity estimates or environmental raster layers can be used to generate maps as output. Additionally, models capable of using daily and monthly weather data further extend the usefulness of these models.

Tools such as CLIMEX[63], GARP[99], and BIOCLIM[100] are commonly used to predict species distributions using climate data in conjunction with presence/absence data. These tools use climate similarity to match areas that may be suitable for a species to become established, based on presence and absence data for the species of interest. Tools such as SimCast, and SimCast Daily and Monthly Means are based on known responses of *P. infestans* to weather and by extrapolation, climate.

Scaling issues occur when trying to model epidemics using weather data that covers large areas or spans of time because disease is driven by localized events. Rainfall patterns or the previous crop in an agricultural field may result in disease development, but rainfall or previous crop data may be too fine to be detectable at a larger spatial resolution grid cell sizes. This would lead to a missed risk prediction at a large scale when using a model. Similarly, when weather data is averaged it removes the extreme events that may trigger disease development, thus disease may occur due to events that are below the temporal resolution threshold of detection for the model.

Using coarser spatial scale data has been examined in other applications and found to be useful as well. Guisan et al. examined the effects of environmental layer grain size on species distribution models[101]. As spatial resolution coarsened there was an overall decrease in model performance. However, an improvement, no change, or degradation of the models' performance can be obtained using coarse grain environmental layers, other confounding factors beyond grain size appear to play a role. The species being modeled, and the region being mapped both affected the model performance as the grain coarsened. As the models were applied across greater spatial scales, interactions that occurred at a plot-level scale did

not necessarily occur or in some cases other interactions began to occur that affected the model output.

This type of modeling, using only weather data, to map risk can be combined with other methods of disease risk prediction. Margosian et al. present a network model which illustrates the connectivity of the US agricultural landscape[19]. Using GIS methods, these types of data could be combined for a more complete picture of host interconnectedness and disease risk. Where both climate and connectivity indicate high risk, there may be "hot spots" for disease to spread quickly.

Other approaches have been taken to model disease risk over large areas. Magarey et al. developed the NAPPFAST Internet system for weather-based mapping of plant pathogens, however, this system uses generic templates rather than being tailored from risk models specific to a particular disease[18]. Thomas et al. created a weather-based information system for commercial growers in California, Washington, Oregon, Idaho, and Arizona which integrated ground-based weather data and plant-growth measurements to determine disease and insect risks[40]. A model to predict wheat head scab or head blight, caused by *Fusarium spp.*, has been deployed on the Internet as a map indicating risk of wheat scab to assist in timing fungicide applications[28].

This type of metamodel is useful for modeling the effects of climate change on plant disease because of the coarse temporal and spatial resolution of climate change data, even when downscaled. It also will be useful for modeling large-scale disease risk relatively quickly since the computing power required to use coarser temporal and spatial scales is much less than fine-resolution data models. Estimates of relative risk can quickly be generated for a given area and further investigation can be pursued if necessary.

Chapter 3

Application of Metamodels for Estimating Global Disease Risk: Potato Late Blight Now and in the Future

3.1 Introduction

At least 800 million people have too little access to food[4]. Plant diseases such as late blight of potato, proximate cause *Phytophthora infestans* (Mont) de Bary, can cause losses of crops further intensifying this problem. It is estimated that US$5 billion are lost due to late blight on an annual basis[102]. Late blight can be controlled through chemical applications[103] and planting resistant potato varieties and other cultural practices including planting dates to avoid wet weather and mixtures[54,104].

Many plant pathogens exhibit a strong relationship with weather patterns. Infectious plant disease occurs due to the interaction of three factors, a favorable environment, a susceptible host, and a competent pathogen (and vector if needed)[3]. This is known as the disease triangle by plant pathologists. Because the environment, or weather, has such a strong influence on plant disease, disease forecasting models are typically constructed using weather observations as input variables.

The interaction of plant pathogens with the host and environment make plant disease

models less straightforward than most animal or insect presence, absence or population models. Plant pathogens, especially fungi and oomycetes, have optimal temperature and moisture requirements for disease to occur. While temperature data is often widely available, moisture data used for plant disease prediction models, usually relative humidity, is not as commonly available.

Because of the important role that environment plays in plant disease development, climate change is likely to affect plant disease[6,36,37] and the impact of plant disease on crop production[15]. Projections and estimations of the effect of future climate on plant disease are common[6,15,36,37,105]. However, creating predictions for the effect of climate change on plant disease are complicated due to issues of scale[22,106].

Models representing plant, pathogen and weather interactions are frequently constructed to predict disease risk for tactical, within season, or strategic between season, short-term control purposes in agricultural or horticultural settings. Somewhat less frequently models are constructed for long-term management decisions e.g. breeding programs, or strategic research prioritization[8]. Ordinarily short-term, tactical within-season models focus on a small spatial scale, field or small region or area. Long-term decisions are made for large regions, countries or continents and the time-period upon which they are based is months or years, not days or weeks.

Efforts to create plant disease risk models for application to large temporal and spatial scales have been developed for different uses. Hijmans et al. mapped global potato late blight severity, predicting the number of fungicide applications necessary to control disease. They found that the estimated number of fungicide sprays in many areas did not correlate to the observed sprays[39]. It was then easy to identify areas that could benefit from host resistance and increased access to fungicides using maps. Magarey et al. developed the NAPPFAST system to map exotic pests risk and identify the potential for establishment, possible entry points, pathway analysis and commodity risk assessment[18]. In a similar fashion Margosian et al. created a network model to analyze the connectivity of the American agricultural

landscape to assess plant disease risk by analyzing host connectivity[19]. Maps generated by these two aforementioned methods are useful for quarantine efforts when a new pathogen is detected, spread can be predicted, and pathogen movement across a large landscape understood.

Climate change effects are likely to be observed on a long time line relative to most crop-production, short-term, tactical within-season plant disease models; the effects will take years to be observed. Climate change model outputs are a larger time-step (e.g., monthly) relative to output from short-term plant disease models (e.g. daily). Breeding programs and other research efforts require long-term decisions. Climate change will likely affect these decisions.

The Intergovernmental Panel on Climate Change (IPCC) developed long-term climate change scenarios[107]. The scenarios are used in analyses of possible climate change impacts and options to mitigate climate change. Because of the complexity of future greenhouse gas (GHG) emissions, different scenarios were developed to provide alternative images of what might occur with future climate conditions. It is highly unlikely that any one of the emission scenarios will be the actual outcome[107].

Four qualitative storylines were developed which the scenarios, or families, are based on. The A1B storyline and scenario family describes a future with very rapid economic growth, global population that peaks mid-century and declines thereafter, and new more efficient technology that is rapidly introduced where similar improvement rates apply to all energy sources and end use technologies[107].

Climate models, based on well-established physical principles have been demonstrated to reproduce recent climate observations and past climate changes. Atmospheric-Ocean General Circulation Models provide credible quantitative estimates of future climate change, particularly at continental and larger scales. Several climate models exist with differing values for atmosphere, ocean, sea ice, coupling, and land features. For a table of selected model features I refer the reader to Table 8.1 in Randall et al.[108].

Nearly 200 wild potato species occur in the Americas[109]. These relatives of cultivated potatoes (*Solanum tuberosum*) can be a valuable genetic resource for cultivated potato breeding programs[110]. *Phytophthora infestans* and closely related *Phytophthora andina* are capable of infecting these wild potato species, causing disease. Climate change could lead to increased disease pressure on wild potato species in areas where resistance to late blight may not occur in the wild populations but inoculum is present from cultivated potato fields.

Because of data availability issues and the large time-steps at which long-term decisions are made most tactical decision models may not be appropriate because of the small time-step that is often required for use with these models. One way of dealing with this issue is through the use of metamodels. Metamodels are "a model of the model"[111]. Using this method I created two metamodels based on the version of SimCast[89] modified by Grünwald et al.[82], see SimCast Daily Means, and SimCast Monthly Means, for more on the model development (Section 2.2).

Simcast Daily Means and Simcast Monthly Means models were created for use with coarse temporal scale weather data in response to a lack of the fine-scale weather data often required for plant disease risk evaluations. These models use daily and monthly resolution, respectively, for temperature and relative humidity to predict 'blight units', a measure of potato late blight risk from SimCast. The applications of larger-than field scale models are typically different than traditional in-field, or within-season models[8]. Large scale disease risk predictions such as those generated by SimCast Monthly Means are useful for policy decisions or making research decisions for future efforts.

3.2 Objectives

My objectives were to map global late blight risk under current and climate change scenarios for resistant and susceptible varieties. Once these maps were generated I used them to extract data and evaluate changes in disease risk globally and for locations where wild potato species are indigenous, and for changes in disease risk for countries where chronic

malnutrition is experienced.

3.3 Materials and Methods

10 arc minute (344 km^2) current climate observations, mean temperature and relative humidity, from 1961 to 1990 were downloaded from the University of East Anglia Climatic Research Unit (CRU), website: http://www.cru.uea.ac.uk, as gridded datasets[85]. Raster files with empty cell values were created in R using the Raster package[112] and then monthly data were extracted from the CRU data and used to create monthly global climate surfaces for mean temperature and relative humidity observations.

Future climate scenario A1B data downloaded from the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset website: http://www-pcmdi.llnl.gov/ipcc/info_for_analysts.php,[113] were downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM were provided by Robert Hijmans, University of California at Davis. Mean temperature was calculated from maximum and minimum temperature. Late blight risk was estimated using SimCast Monthly Means for the 1950-2000 average, current conditions, and three future 20-year time periods, 2000-2020, 2040-2060 and 2080-2100. For each month of each year, the average temperature and relative humidity of the three climate models was generated in R using the Raster package[112].

The predict function from the R Raster package was used to predict late blight risk for both observed and future predictions using the SimCast Monthly Means fitted GAM surface to generate raster file outputs[112].

A map of worldwide potato growing areas, provided by Robert Hijmans and CIP, was used to remove non-potato production areas from SimCast Monthly Means global blight unit predictions.

A file of potato growing seasons was provided by Robert Hijmans. Potato growing seasons were calculated using the ECOCROP 1 database[114] and a R script to predict the

first day of the month in which the planting date produced the highest yield. Growing seasons were assumed to be three months long. Blight units were summed at a moving three month interval beginning with January, February and March for the January growing season, and so on to generate a summary late blight risk value for 12 growing seasons.

The potato growing season raster file was converted to a polygon shape file and individual growing season polygon selections were created. The growing season summary blight unit files were clipped by the appropriate growing season polygon. This resulted in separate raster layers that represented blight units for each three-month growing season only. These were then stacked to create output maps of global late blight risk or used to extract risk values for different countries or locales for further analysis.

A map potato priority was provided by CIP. Potato priority was determined by production area (hectares) for each country in relation to the respective country's population. A map of malnutrition values based on FAO's estimates[115] of chronic malnutrition was also provided by CIP. These two shape files were used to determine countries with high potato priority and high malnutrition for further analysis.

A shape file based on surveys of wild potato species occurrence in the Americas was provided by Robert Hijmans and CIP. This map was used to create a file of species richness using Diva-GIS v.5[116]. The species richness file was used with the SimCast Monthly Means results as a mask of wild potato species occurrence and as a measure of where late blight risk changes in relation to the areas of greatest wild potato species richness.

3.4 Results

3.4.1 Changes in global late blight risk

Global late blight risk in current potato growing regions increases slightly from the 1961-1990 climate conditions to the 2000-2020 time period predictions. From the 2000-2020 time period to to 2060-2080 time period late blight in current potato growing regions decreases by 2%. From 2040-2060 to 2080-2100, late blight in current potato growing regions is reduced

by 4% (Table 3.1). For resistant varieties late blight units are reduced from 522332 for the 1961-1990 time period to the 2040-2060 time period in which blight units total 252958, a -107% drop in risk for resistant cultivars. The amount of blight units accumulated does change. However, the same general patterns in relative risk can be observed time periods. That is, areas of high late blight risk do not become extremely low risk compared to other high risk areas and vice-versa.

Under all climate time-periods used, late blight risk is greatest in south-east Brazil, northern Europe, especially United Kingdom and Ireland, south-eastern Africa and Madagascar, parts of the Himalayas, north-east Asia, and the Highland Tropics of South America. These areas were confirmed by personal communication with Greg Forbes (Centro Internatial de la Papa), Robert Hijmans (University of California at Davis), and Eduardo Mizubuti (Universidade Federal de Viçosa).

Table 3.1: Cumulative blight units and percentage change from previous time period of sum of blight units for each of the climate time periods. Blight units[89,117] are an indicator of late blight risk. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and[113 113] were provided by Robert Hijmans, University of California at Davis.

	Sum Blight Units	Change
1961-1990	489400	
2000-2020	495174	1.1%
2040-2060	485469	-2.0%
2080-2100	466007	-4.0%

3.4.2 Changes in late blight risk for countries with chronic malnutrition and high potato priority

Countries with a high priority for potato production and having high levels of chronic malnutrition are predicted to have high levels of late blight risk. Only Pakistan remains relatively low when compared with the other countries with chronic malnutrition and high potato priority (Figure 3.7).

Changes in late blight risk from 1961-1990 to 2040-2060 for the countries with high potato priority and chronic malnutrition differed. Late blight risk in Malawi and Bolivia is to projected decrease by greater than 10% while Nepal is projected to experience the greatest increase in late blight risk, 18% more blight units to 180 blight units (Table 3.2, Figure 3.8). Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

Table 3.2: Change of total of blight units for time periods of 1961-1990 and 2040-2060 in countries with high potato priority and chronic malnutrition (Hunger) expressed as percentage of population chronically malnourished. Blight units[89,117] are an indicator of late blight risk. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

	Blight Units			
Country	1961-1990	2040-2060	Change	Hunger
Bangladesh	140	143	2.0%	37%
Bolivia	83	73	-12.2%	23%
China	126	123	-2.4%	13%
Columbia	296	313	6.0%	12%
India	103	94	-9.1%	22%
Kenya	88	88	0.5%	41%
Korea, DPR	288	271	-5.8%	48%
Madagascar	341	351	2.7%	39%
Malawi	256	204	-20.3%	37%
Nepal	153	180	18.0%	21%
Pakistan	18	17	-7.9%	19%
Peru	117	128	9.1%	19%
Rwanda	253	270	6.6%	37%
Uganda	228	226	-1.0%	28%

3.4.3 Changes in late blight risk for areas with wild potato species present

Late blight risk for areas of South America where the greatest wild potato species richness occurs decreases slightly under the future climate scenario used in this study. However, a shift occurs, areas of greater wild potato species richness are at higher risk than areas of lower species richness under climate change effects (Figure 3.8).

Table 3.3: Blight units observed from November through April in South America where wild potato species are found. Blight units[89,117] are an indicator of late blight risk. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

Blight Units	1961 to 1990	2000 to 2020	2040 to 2060	2080 to 2100
0 to 5	135	135	134	107
5 to 10	130	130	128	128
10 to 15	108	113	112	113
15 to 20	50	52	54	57
Total	**423**	**430**	**428**	**405**

3.5 Discussion

From the data sets used, it appears that late blight risk in current potato growing areas will increase in the near future and then decrease. Decreases in late blight risk, the sum of all annual blight units for potato growing areas, are exhibited between the three predicted climate model based results. The weather data observations from 1961-1990 could predict a different level of blight units because it is the average of one set of weather observations. The predicted climate data are averages of three climate model outputs and this may have a smoothing effect on the data, removing extreme observations. However, this does remove areas where the models are not in agreement as to weather patterns and the blight unit patterns are similar to what the observed weather data predicted. It appears that late

blight risk is not as affected for resistant varieties as for susceptible varieties. Resistant varieties decrease by a sum total of -236363 blight units per growing season between the 1961-1990 growing season and the 2040-206 growing season.

Most countries where chronic malnutrition is a problem experience a decrease or very slight increase in late blight risk. The exception, Nepal, experiences an increase of 27 blight units (15%), respectively, by 2060. The global decrease in the amount of blight units accumulated does not occur in all areas.

There is a shift in late blight risk in areas where wild potato species are found in South America. Most wild potato species are found in the tropical highlands[109]. The grid cell size of 10 arc minutes is coarse for the changes of elevation that occur in montane regions and may not adequately detect the nuances of weather affected by peaks and valleys.

The assumptions were that the rainy season when these plants are actively growing would not change. Precipitation is the most difficult weather variable to predict under climate change. I did not have any reason to anticipate that it would change for this area. However, the alpine regions where most of these species are found are difficult to accurately represent at a 10 arc minute spatial resolution due to rapid changes in elevation. It would be beneficial to use .5 arc second data for this area only to see what effects may occur. While computationally impractical for the entire global dataset, such an approach could be used a smaller subset such as a region of interest.

The factors used to predict late blight risk are weather factors. The models that were used to predict future scenarios predict climate. Climate data represent weather means, but not the typical variation of weather. Late blight risk will likely be affected differently across years as a result of changing weather patterns not represented by climate.

The model only predicts risk under conditions when inoculum is present. Therefore, it is an indicator of the the host and environment interaction only in the presence of inoculum. For areas where inoculum loads are low but environmental conditions are favorable, the model will incorrectly predict an elevated risk. Because this model cannot account for an

increase in risk as inoculum builds over the growing season, the values are a simple sum of a growing season three months long, as determined to be optimal for potato yield. Future modifications could include a function that would cause risk to increase as the time passed; a growth curve model could be applied here. The inoculum from one month would lead to a higher initial level in the following month. This approach would not account for high inoculum loads early in the growing season followed by lower loads because of the temporal resolution being used. The assumptions only allow for an increasing inoculum load.

Another effect caused by inoculum load that this model is not capable of predicting includes a "risk neighborhood". As inoculum loads increase nearby this will increase risk in other areas with a conducive environment and susceptible host. Spatio-temporal models such as those developed by Skelsey et al. are being developed that can model regional interactions of inoculum loads[58] . However, the time-frame in which these models work is much shorter than the time-frame for which SimCast Monthly Means was designed.

In areas where irrigation is used, this model will not be appropriate, because growing season was based on a rainfed potato crop. This could exclude areas where irrigation is used to raise potatoes. Because SimCast Monthly Means is based on SimCast, a model which is successful in areas of high-humidity and rainfall, it might not be suitable for semi-arid potato growing environments such as potato-producing areas of the Pacific-Northwest (Idaho, Oregon, and Washington)[118]. A different model that uses similar variables as Henderson et al. could be incorporated and applied to only applicable semi-arid areas where potatoes are grown[118].

Even using an "optimistic" climate change scenario such as the A1b scenario we see changes in late blight risk. With the framework that is now in place it will be possible to quickly make comparisons with other emission scenarios to see what effects higher fossil fuel emissions could have on late blight.

It could be possible to use this metamodel to estimate late blight risk based on daily or monthly data to estimate late blight risk using remotely sensed satellite data to continuously

update blight risk. This type of information could provide real time late blight risk at a glance, globally.

Metamodels like SimCast Monthly Means give plant pathologists new tools for the ongoing battle with plant disease. The ability to quickly estimate relative risk globally using readily available weather data is a useful tool. Previously efforts to do this were computationally and time intensive. SimCast Monthly Means is capable of producing localized results in a few minutes and a global estimate for one month in about an hour. Further advances hopefully will improve the model accuracy in the future by incorporation of potential shifts in potato growing areas due to climate change giving a more complete estimate of the impact that late blight may have on potato in years to come.

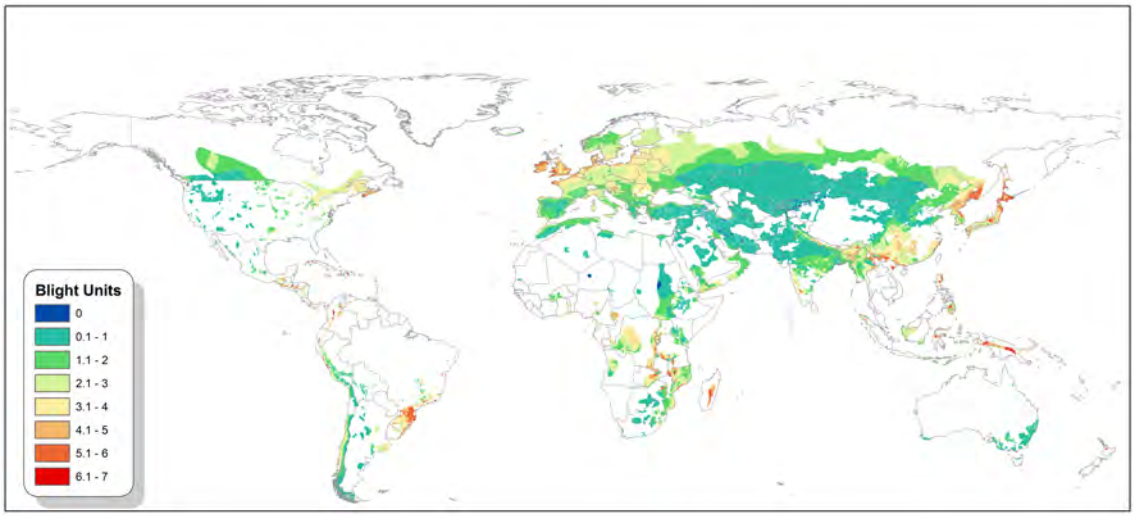

Figure 3.1: Map one of four of susceptible potato cultivar global late blight risk. Global late blight risk from 1961 to 1990 for a susceptible potato cultivar, expressed as the sum of blight units for the highest yielding three month growing season per locality for potato growing areas only. Blight units[89,117] are an indicator of late blight risk. Gridded weather data from New et al.[85] was used to create climate surfaces for use with SimCast Monthly Means.

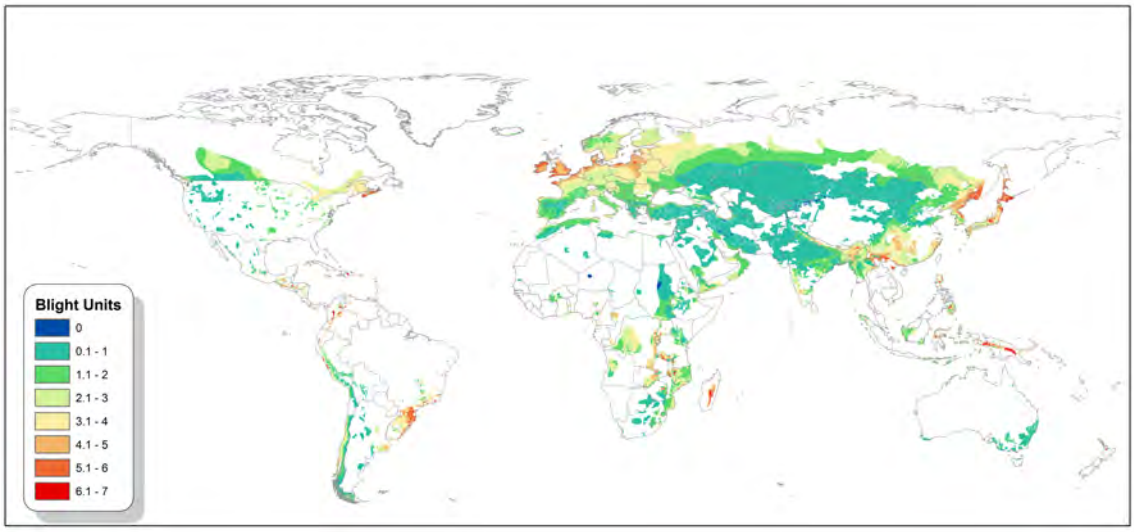

Figure 3.2: Map two of four of susceptible potato cultivar global late blight risk. Global late blight risk from 2000 to 2020 for a susceptible potato cultivar, expressed as the sum of blight units for the highest yielding three month growing season per locality for potato growing areas only. Blight units[89,117] are an indicator of late blight risk. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

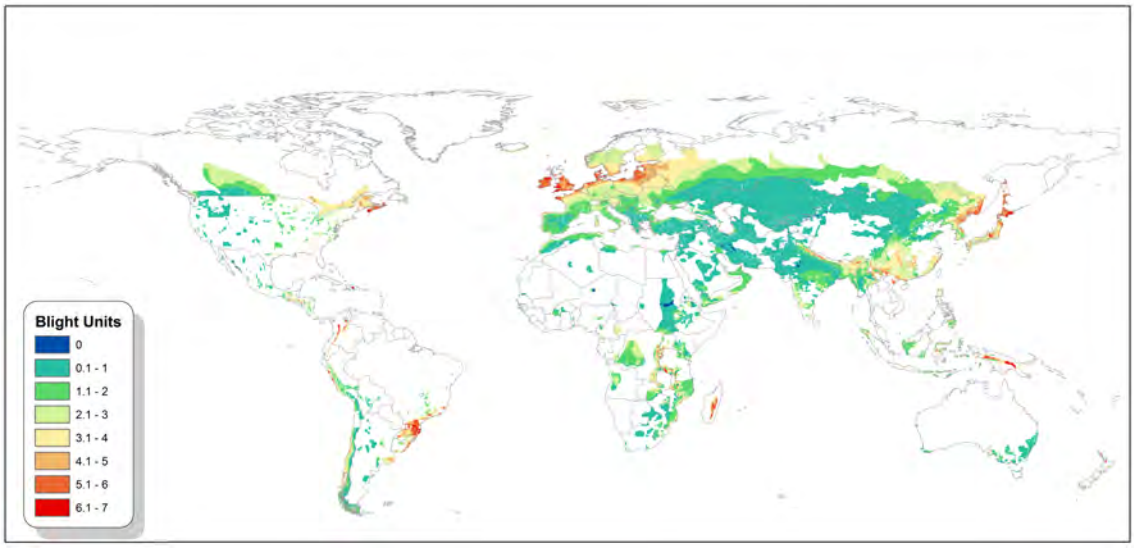

Figure 3.3: Map three of four of susceptible potato cultivar global late blight risk. Global late blight risk from 2040 to 2060 for a susceptible potato cultivar, expressed as the sum of blight units for the highest yielding three month growing season per locality for potato growing areas only. Blight units[89,117] are an indicator of late blight risk. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

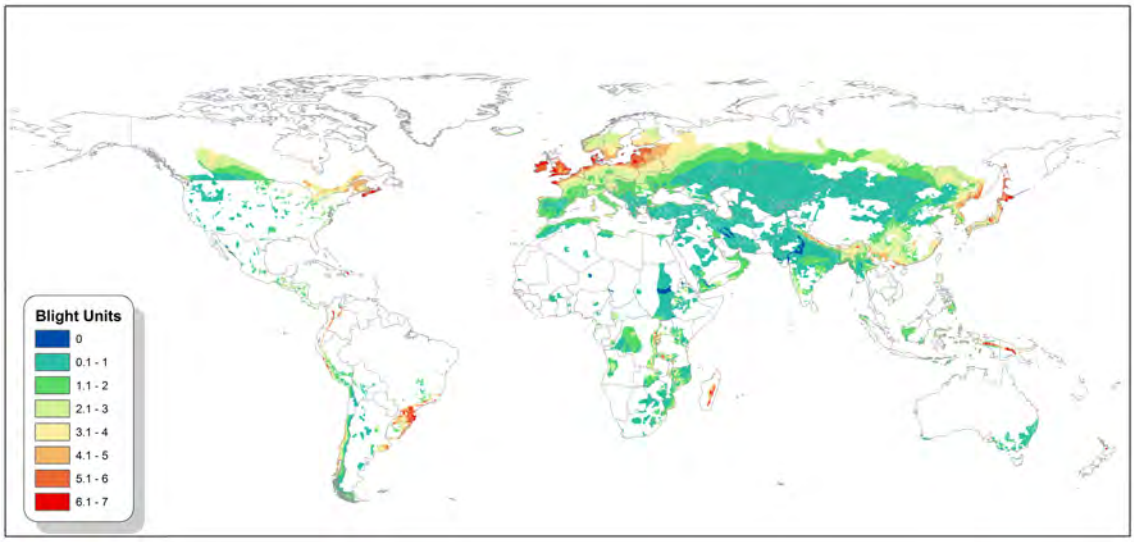

Figure 3.4: Map four of four of susceptible potato cultivar global late blight risk. Global late blight risk from 2080 to 2100 for a susceptible potato cultivar, expressed as the sum of blight units for the highest yielding three month growing season per locality for potato growing areas only. Blight units[89,117] are an indicator of late blight risk. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

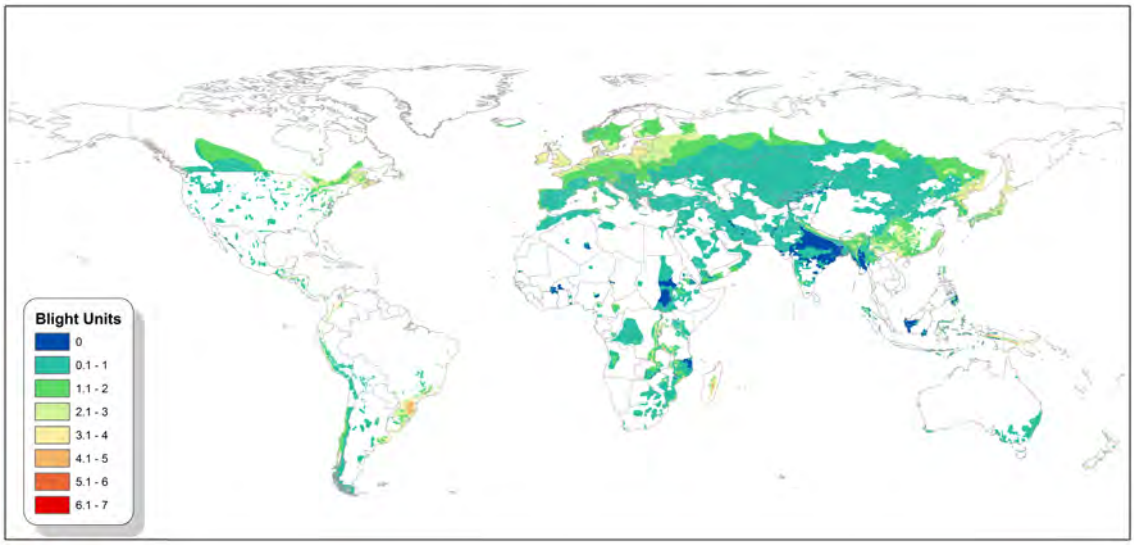

Figure 3.5: Global late blight risk from 2040 to 2060 for a resistant potato cultivar, expressed as the sum of blight units for the highest yielding three month growing season per locality for potato growing areas only. Blight units[89,117] are an indicator of late blight risk. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

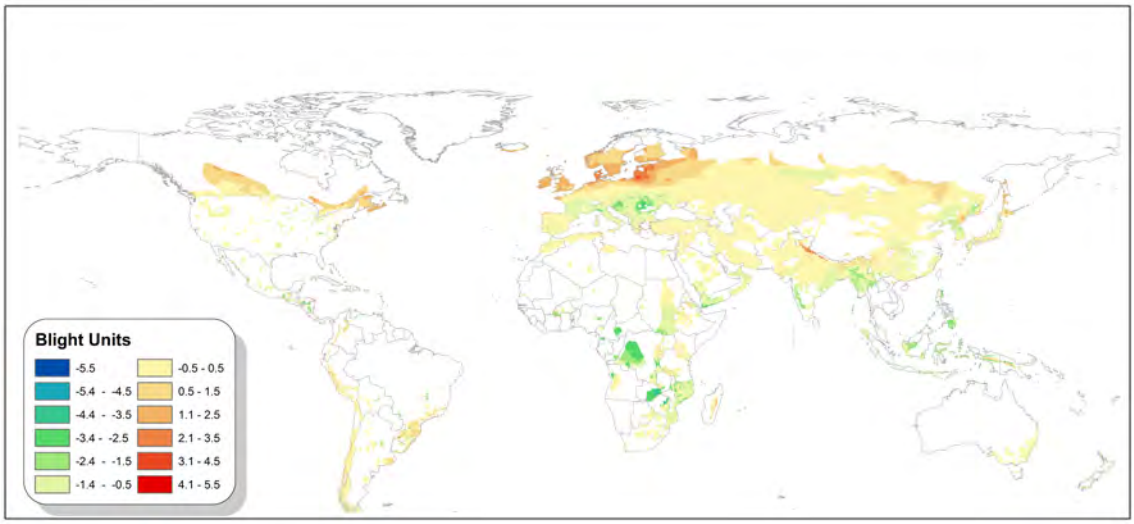

Figure 3.6: Change in global late blight risk from the 1961-1990 time period to projected 2040 to 2060 climate conditions for a susceptible cultivar, expressed as the sum of blight units for the highest yielding three month growing season per locality for potato growing areas only. Blight units[89,117] are an indicator of late blight risk. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

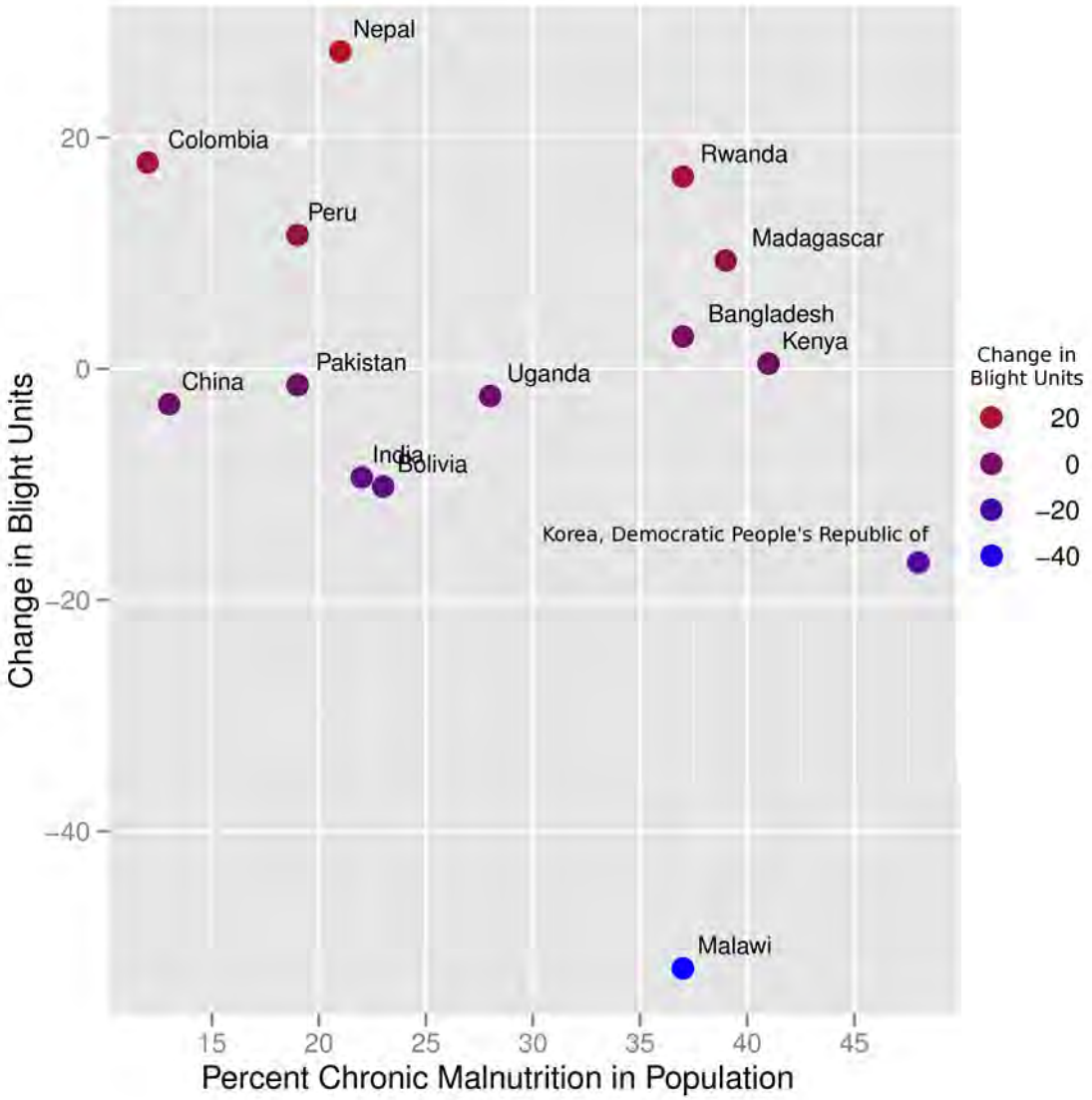

Figure 3.7: Change in blight units normalized by area for countries with malnutrition and a high priority on potato production, from 1961-1990 conditions to 2040-2060 predictions. Values may change due to shifts in potato production area suitability and climate suitability changes. Future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis. Malnutrition data is from FAO[115] and provided courtesy of CIP.

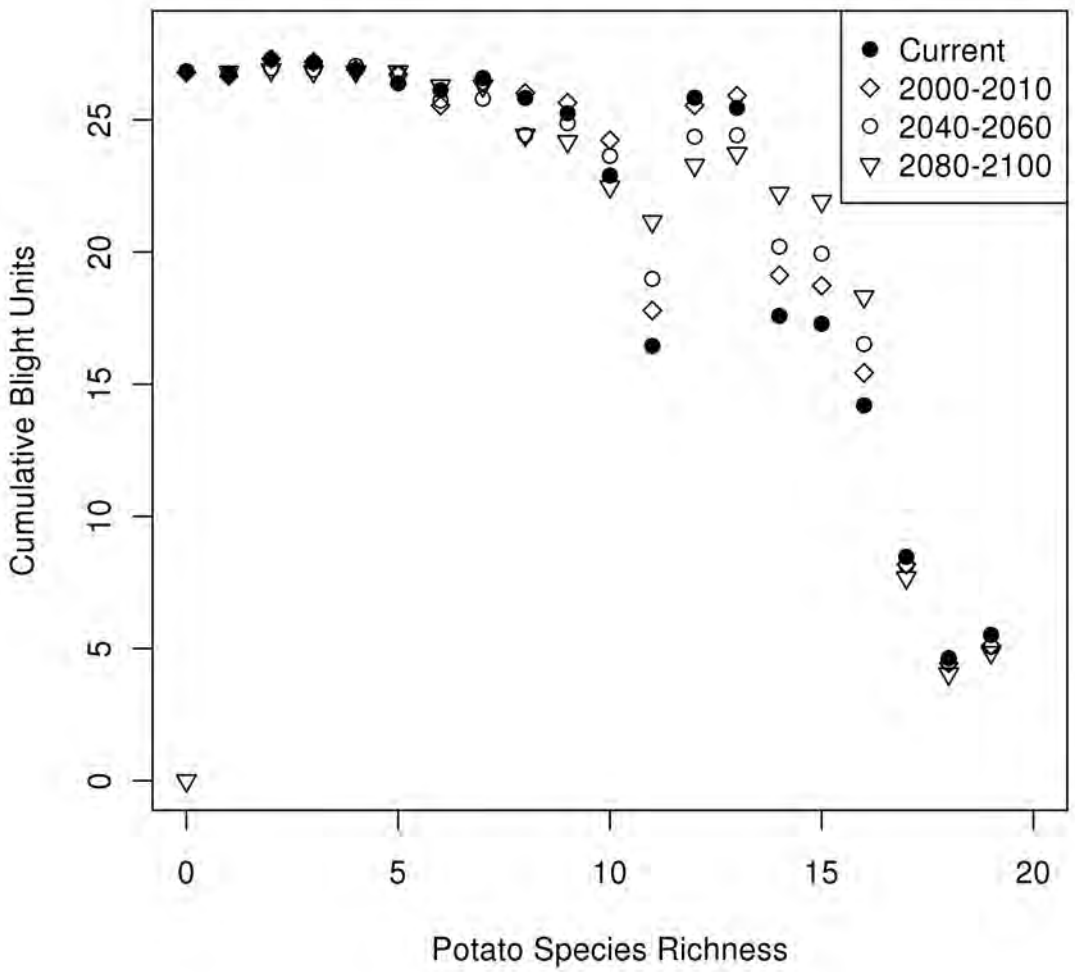

Figure 3.8: Late blight risk to wild potato species found in South America. Potato species richness is the number of wild potato species observed in a grid cell neighborhood. Species in Central and North America were excluded from these observations. Blight units[89,117] are an indicator of late blight risk. Wild potato species occurrence and future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

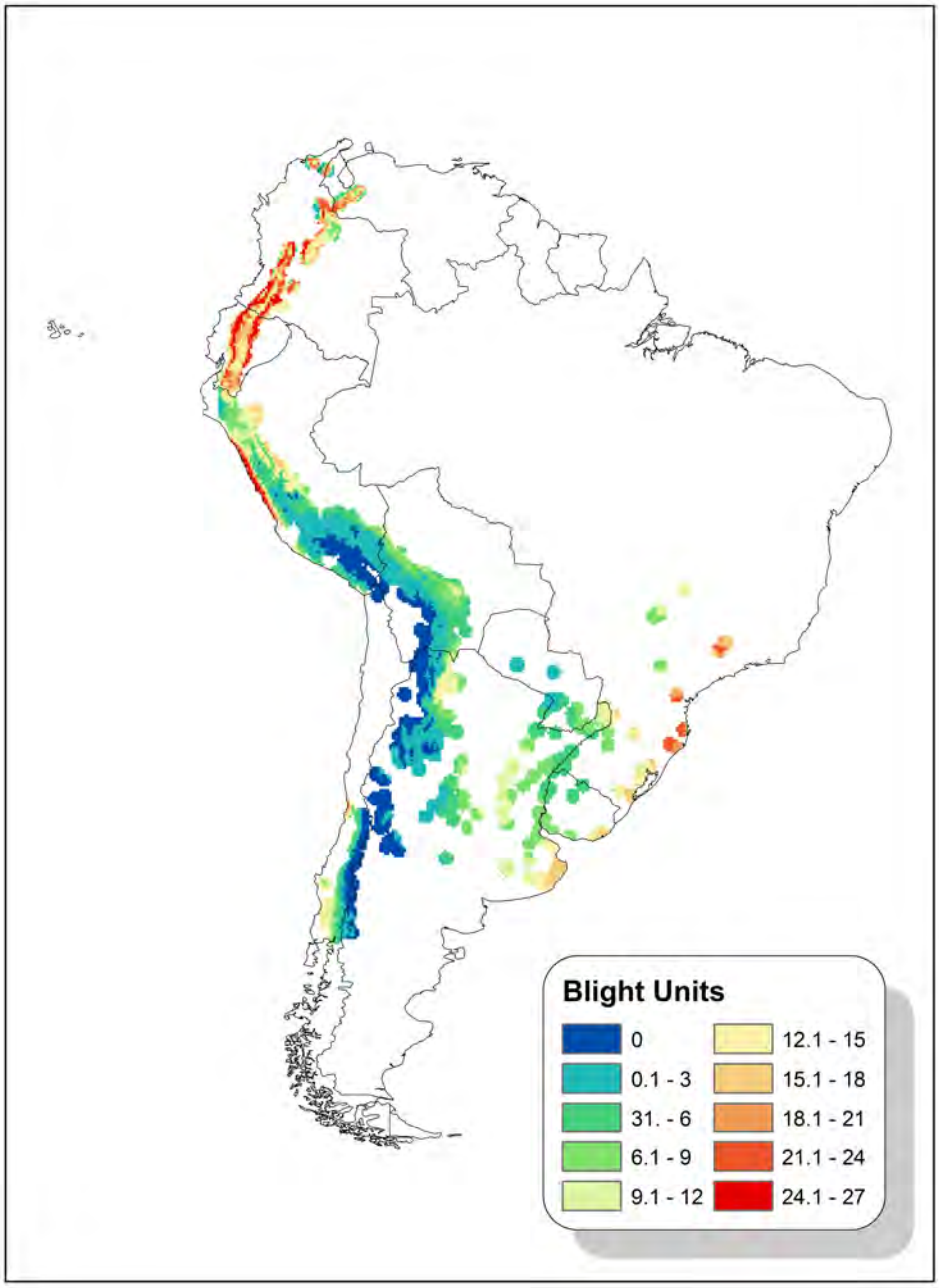

Figure 3.9: Map one of four of late blight risk to wild potato species. Late blight risk to wild potato species for 1961-1990 time period for the months of November through April. Blight units[89,117] are an indicator of late blight risk. Wild potato species occurrence and future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

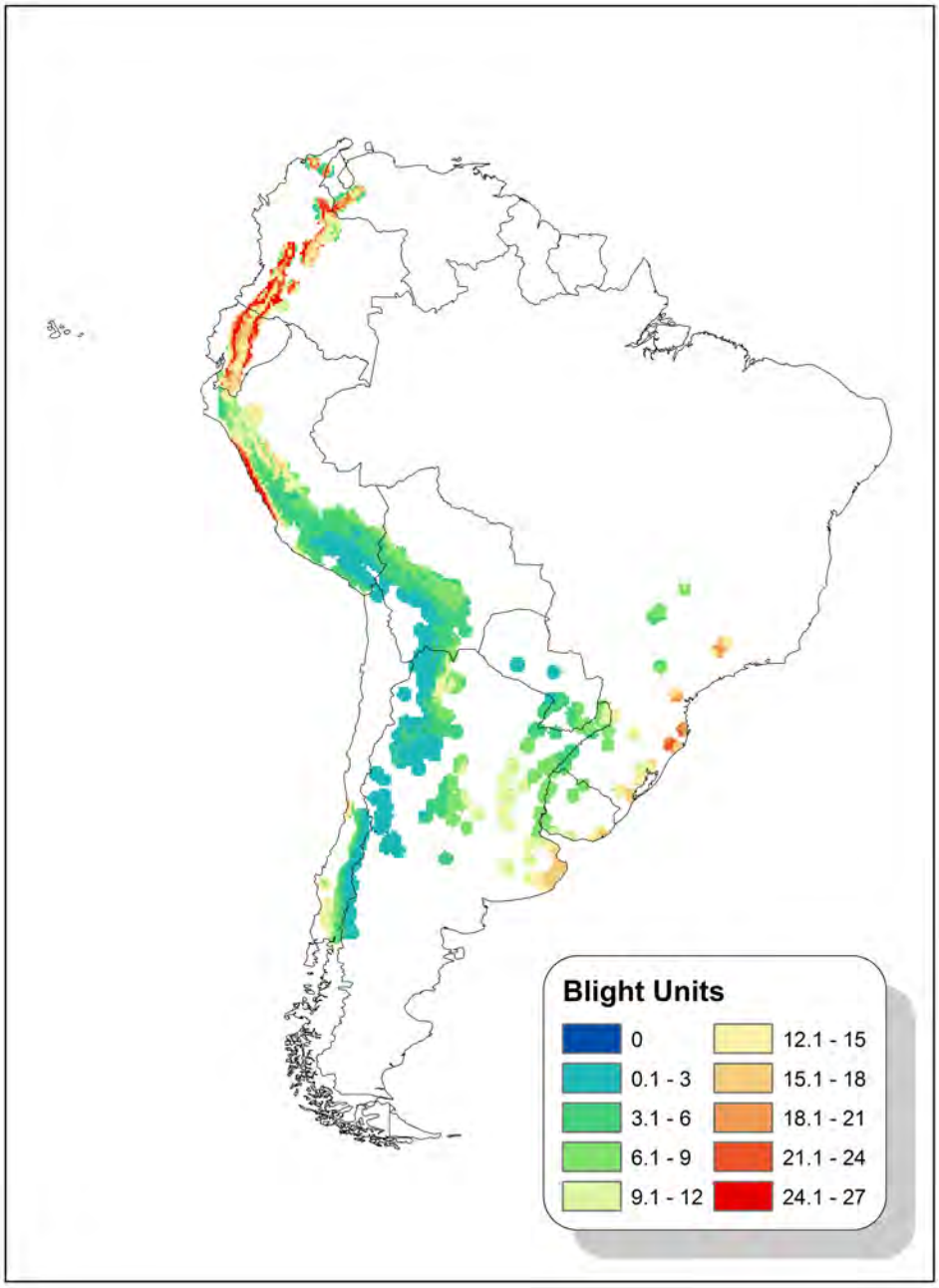

Figure 3.10: Map two of four of late blight risk to wild potato species. Late blight risk to wild potato species for the 2000-2020 time period for the months of November through April. Blight units[89,117] are an indicator of late blight risk. Wild potato species occurrence and future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

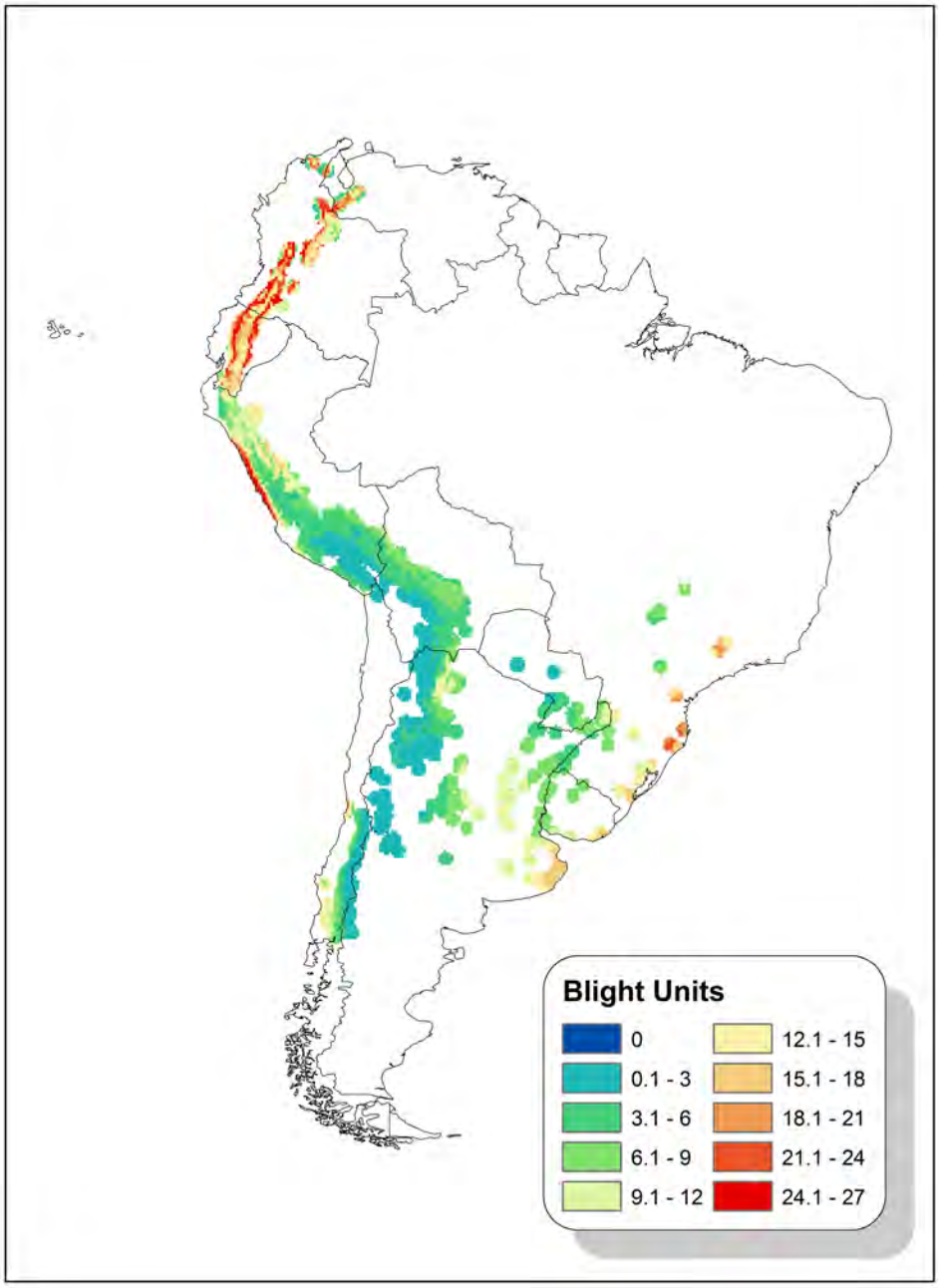

Figure 3.11: Map three of four of late blight risk to wild potato species. Late blight risk to wild potato species for the 2040-2060 time period for the months of November through April. Blight units[89,117] are an indicator of late blight risk. Wild potato species occurrence and future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

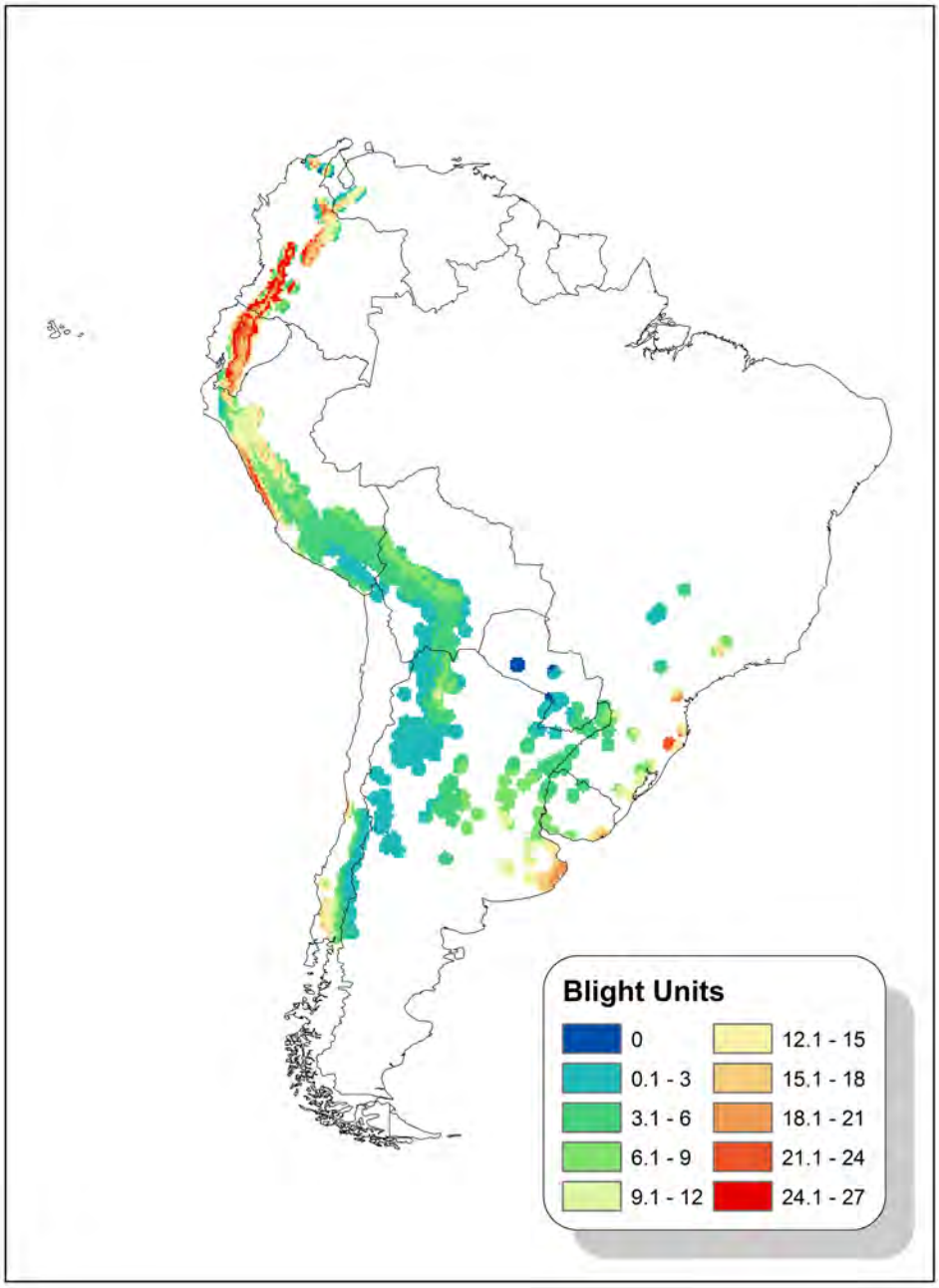

Figure 3.12: Map four of four of late blight risk to wild potato species. Late blight risk to wild potato species for the 2080-2100 time period for the months of November through April. Blight units[89,117] are an indicator of late blight risk. Wild potato species occurrence and future climate scenario A1B data downscaled to 10 arc minutes for three global climate change models, MIROC3.2 (hires), INM-CM3.0, and GISS-AOM[113] were provided by Robert Hijmans, University of California at Davis.

Bibliography

[1] J. S. Clark, Science **293**, 657 (2001).

[2] S. A. Levin, Ecology **73**, 1943 (1992).

[3] L. V. Madden, G. Hughes, and F. van den Bosch, *The Study of Plant Disease Epidemics*, The American Phytopathological Society, APS Press St. Paul, Minnesota, 2007.

[4] R. N. Strange and P. R. Scott, Annual Review of Phytopathology **43**, 83 (2005).

[5] J. J. Burdon, P. H. Thrall, and L. Ericson, Annual Review of Phytopathology **44**, 19 (2006).

[6] P. K. Anderson et al., Trends in Ecology & Evolution **19**, 535 (2004).

[7] D. M. Cahill, J. E. Rookes, B. A. Wilson, L. Gibson, and K. L. McDougall, Australian Journal of Botany **56**, 279 (2008).

[8] S. Savary, P. S. Teng, L. Willocquet, and F. W. Nutter, Annual Review of Phytopathology **44**, 89 (2006).

[9] O. Berke, Preventive Veterinary Medicine **71**, 173 (2005).

[10] T. Koch, *Cartographies of Disease: maps, mapping, and medicine*, ESRI Press, Redlands, CA, 1st edition, 2005.

[11] H. Brody et al., The Lancet **356**, 64 (2000).

[12] A. J. Herbertson, The Geographical Journal **25**, 300 (1905).

[13] J. Omernik, Annals of the Association of American Geographers **77**, 118 (1987).

[14] S. Wood and P. G. Pardey, Agricultural Systems , 13 (1997).

[15] R. J. Hijmans, American Journal Of Potato Research **80**, 271 (2003).

[16] J. J. Burdon and P. H. Thrall, American Naturalis **153**, S15 (1999).

[17] X. B. Yang, European Journal of Plant Pathology **115**, 25 (2006).

[18] R. D. Magarey et al., Plant Disease **91**, 336 (2007).

[19] M. L. Margosian, K. A. Garrett, J. M. S. Hutchinson, and K. A. With, BioScience **59**, 141 (2009).

[20] E. D. De Wolf, L. V. Madden, and P. E. Lipps, Phytopathology **93**, 428 (2002).

[21] D. Borcard and P. Legendre, Ecological Modelling **153**, 51 (2002).

[22] R. C. Seem, Canadian Journal of Plant Pathology **26**, 274 (2004).

[23] W. W. Turechek, European Journal of Plant Pathology **115**, 53 (2006).

[24] L. Willocquet and S. Savary, Phytopathology **94**, 883 (2004).

[25] M. A. Leibold et al., Ecology Letters **7**, 601 (2004).

[26] K. A. With, Risk Analysis **24**, 803 (2004).

[27] M. Plantegenest, C. Le May, and F. Fabre, Journal of The Royal Society Interface **4**, 963 (2007).

[28] E. D. De Wolf and S. A. Isard, Annual Review of Phytopathology **45**, 203 (2007).

[29] P. A. J. Van Oort and A. K. Bregt, Risk Analysis **25**, 1599 (2005).

[30] M. Crosetto and S. Tarantola, International Journal Geographical Information Science **15**, 415 (2001).

[31] G. Wang, G. Z. Gertner, S. Fang, and A. B. Anderson, Photogrammetric Engineering & Remote Sensing **71**, 1423 (2005).

[32] J. F. Bithell, Statistics in Medicine **9**, 691 (1990).

[33] B. M. Wu and K. V. Subbarao, Plant Disease **89**, 90 (2005).

[34] M. J. Zwankhuizen, F. Govers, and J. C. Zadoks, Phytopathology **88**, 754 (1998).

[35] G. J. Boender et al., PLoS Computational Biology **3**, e71 (2007).

[36] S. M. Coakley, H. Scherm, and S. Chakraborty, Annual Review of Phytopathology **37**, 399 (1999).

[37] K. A. Garrett, S. P. Dendy, E. E. Frank, M. N. Rouse, and S. E. Travers, Annual Review of Phytopathology **44**, 489 (2006).

[38] T. B. Hallett et al., Nature **430**, 71 (2004).

[39] R. J. Hijmans, G. A. Forbes, and T. S. Walker, Plant Pathology **49**, 697 (2000).

[40] C. S. Thomas, P. W. Skinner, A. D. Fox, C. A. Greer, and W. D. Gubler, Journal of Nematology **34**, 200 (2002).

[41] M. E. Del Ponte, C. Godoy, X. Li, and X. Yang, Phytopathology **96**, 797 (2006).

[42] R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, International Journal of Climatology **25**, 1965 (2005).

[43] L. Huber and T. J. Gillespie, Annual Review Of Phytopathology **30**, 553 (1992).

[44] K. S. Kim, M. L. Gleason, and S. E. Taylor, Plant Disease **90**, 650 (2006).

[45] P. C. Sentelhas and T. J. Gillespie, Theoretical and Applied Climatology **91**, 205 (2008).

[46] D. E. Aylor, Ecology **84**, 1989 (2003).

[47] S. A. Isard, S. H. Gage, P. Comtois, and J. M. Russo, BioScience **55**, 851 (2005).

[48] S. A. Isard, J. M. Russo, and A. Ariatti, Aerobiologia **23**, 271 (2006).

[49] C. E. Mitchell and A. G. Power, Nature , 625 (2003).

[50] E. Stokstad, Science **315**, 1786 (2007).

[51] T. Peterson, D. A. Vieglais, and J. K. Andreasen, **3**, 27 (2003).

[52] P. D. Esker, J. Harri, P. M. Dixon, and F. W. Nutter Jr, Plant Disease **90**, 1353 (2006).

[53] Y. Zhu et al., Nature **406**, 718 (2000).

[54] K. A. Garrett, R. J. Nelson, C. C. Mundt, R. E. J. Chacón, and G. A. Forbes, Phytopathology **91**, 993 (2001).

[55] K. A. Garrett and C. C. Mundt, Phytopathology **89**, 984 (1999).

[56] C. C. Mundt, Annual Review of Phytopathology **40**, 381 (2002).

[57] P. Skelsey, W. A. H. Rossing, G. J. T. Kessel, J. Powell, and W. van der Werf, Phytopathology **95**, 328 (2005).

[58] P. Skelsey, G. J. T. Kessel, W. A. H. Rossing, and W. V. D. Werf, Phytopathology **99**, 290 (2009).

[59] D. L. Urban and T. Keitt, Ecology **82**, 1205 (2001).

[60] M. J. Jeger, M. Pautasso, O. Holdenrieder, and M. W. Shaw, New Phytologist **174**, 279 (2007).

[61] A. T. Peterson and D. A. Vieglais, BioScience **51**, 363 (2001).

[62] J. Elith et al., Ecography **2**, 129 (2006).

[63] R. Sutherst and G. Maywald, Agriculture, Ecosystems and Environment **13**, 281 (1985).

[64] T. Yonow, D. J. Kriticos, and R. W. Medd, Phytopathology **94**, 805 (2004).

[65] S. Pethybridge, M. Nelson, and C. Wilson, Australasian Plant Pathology **32**, 493 (2003).

[66] S. Pivonia and X. B. Yang, Plant Disease **88**, 523 (2004).

[67] M. Desprez-Loustau et al., Canadian Journal of Plant Pathology / Revue Canadienne de Phytopathologie **29**, 101 (2007).

[68] J. M. Jeschke and D. L. Strayer, Annals of the New York Academy of Sciences **1134**, 1 (2008).

[69] R. J. Hijmans and C. H. Graham, Global Change Biology **12**, 2272 (2006).

[70] R. Meentemeyer, D. Rizzo, W. Mark, and E. Lotz, Forest Ecology and Management **200**, 195 (2004).

[71] Z. Pan et al., Plant Disease **90**, 840 (2006).

[72] J. H. Everitt, D. E. Escobar, D. N. Appel, W. G. Riggs, and M. R. Davis, Plant Disease **83**, 502 (1999).

[73] T. Kobayashi, E. Kanda, K. Kitada, K. Ishiguro, and Y. Torigoe, Phytopathology **91**, 316 (2000).

[74] O. Holdenrieder, Trends in Ecology & Evolution **19**, 446 (2004).

[75] D. Goodin et al., Global Ecology and Biogeography **15**, 519 (2006).

[76] H. E. Nilsson, Annual Review Of Phytopathology **33**, 489 (1995).

[77] J. West et al., Annual Review Of Phytopathology **41**, 593 (2003).

[78] C. A. Gilligan, New Phytologist **115**, 223 (1990).

[79] M. C. Wimberly, A. D. Baer, and M. J. Yabsley, International Journal of Health Geographics **14**, 1 (2008).

[80] A. G. Power and C. E. Mitchell, The American Naturalist **164**, S79 (2004).

[81] L. Roesch et al., IMSE **1**, 283 (2007).

[82] N. J. Grünwald, G. R. Montes, H. L. Saldana, O. A. R. Covarrubias, and W. E. Fry, Plant Disease **86**, 1163 (2002).

[83] P. A. Paul and G. P. Munkvold, Phytopathology **94**, 1350 (2004).

[84] M. E. Del Ponte and P. D. Esker, Scientia Agricola **65**, 88 (2008).

[85] M. New, D. Lister, M. Hulme, and I. Makin, Climate Research **21**, 1 (2002).

[86] L. J. Beaumont, A. J. Pitman, M. Poulsen, and L. Hughes, Global Change Biology **13**, 1368 (2007).

[87] J. R. Wallin, Forecasting tomato and potato late blight in the north-central region, in *Phytopathology*, page 37, 1951.

[88] R. A. Krause, L. B. Massie, and R. A. Hyre, Plant Disease Reporter **59**, 95 (1975).

[89] W. E. Fry, A. E. Apple, and J. A. Bruhn, Phytopathology **73**, 1054 (1983).

[90] US-EPA and NOAA, Hourly united states weather observations 1990-95, CD-ROM, 1997.

[91] T. Hastie and R. Tibshirani, Statistical Science **1**, 297 (1986).

[92] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2009, ISBN 3-900051-07-0.

[93] S. N. Wood, Journal of the Royal Statistical Society (B) **70**, 495 (2008).

[94] S. Wood, *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC, 2006.

[95] R. J. Hijmans, American Journal Of Potato Research , 403 (2001).

[96] R. M. Cu and P. M. Phipps, Phytopathology **83**, 195 (1993).

[97] E. Billing, Phytopathology **97**, 1036 (2007).

[98] C. Cesaraccio, D. Spano, P. Duce, and R. L. Snyder, International journal of biometeorology **45**, 161 (2001).

[99] D. Stockwell and D. Peters, International Journal Geographical Information Science **13**, 143 (1999).

[100] H. A. Nix, *Atlas ofAustralian Elapid Snakes*, Australian Government Publishing Service, Canberra., 1986.

[101] A. Guisan, C. H. Graham, J. Elith, and F. Huettmann, Diversity and Distributions **13**, 332 (2007).

[102] H. S. Judelson and F. A. Blanco, Nature reviews. Microbiology **3**, 47 (2005).

[103] J. M. Stein and W. W. Kirk, Crop Protection **21**, 575 (2002).

[104] F. Pilet, G. Chacon, G. A. Forbes, and D. Andrivon, Phytopathology **96**, 777 (2006).

[105] H. Scherm, Canadian Journal of Plant Pathology **273**, 267 (2004).

[106] R. C. Seem, R. D. Magarey, J. W. Zack, and J. M. Russo, Environmental Pollution **108**, 389 (2000).

[107] IPCC (Intergovernmental Panel on Climate Change), Special report on emissions scenarios, Technical report, IPCC, 2000.

[108] D. Randall et al., *Climate Models and Their Evaluation*, chapter 8, pages 589–662, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2007.

[109] R. J. Hijmans and D. M. Spooner, American Journal of Botany **88**, 2101 (2001).

[110] S. Jansky, Plant Breeding Reviews **19**, 69 (2000).

[111] J. P. C. Kleijnen, *Statistical tools for simulation practitioners*, Marcel Dekker, Inc., New York, NY, USA, 1986.

[112] R. J. Hijmans and J. van Etten, *Raster: Geographic analysis and modeling with raster data*, 2009, R package version 0.9.8-16/r600.

[113] G. A. Meehl et al., Bulletin of the American Meteorological Society **88**, 1383 (2007).

[114] *ECOCROP: The crop environmental requirements database and the crop environment response database*, Food and Agriculture Organization of the United Nations (FAO), 2003.

[115] FAO, The state of food insecurity in the world 2002, Rome.

[116] R. J. Hijmans, G. L, A. Jarvis, O. R, and M. P, *DIVA-GIS*, v 5.4.0.1 edition.

[117] N. J. Grünwald, O. A. Rubio-Covarrubias, and W. E. Fry, Plant Disease **84**, 410 (2000).

[118] D. Henderson, C. J. Williams, and J. S. Miller, Plant Disease **91**, 951 (2007).

Appendix A

Supplementary Maps

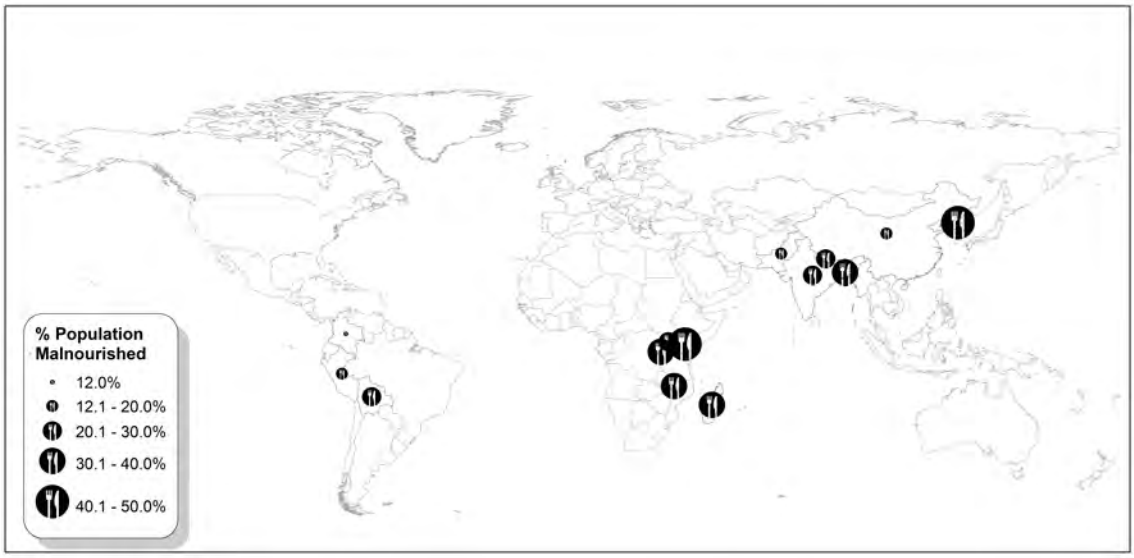

Figure A.1: Blight unit severity normalized by area for countries with malnutrition and a high priority on potato production. Countries with high priority for potato production have a high ratio of hectares of potato production to population. Symbols are relative in size to the percent of a particular country's population which suffers chronic malnutrition. The People's Democratic Republic of Korea, and Malawi both experience much malnutrition. Columbia is the lowest shown around 12% chronic malnutrition. Data from FAO[115] and International Potato Center (CIP), Lima, Peru.

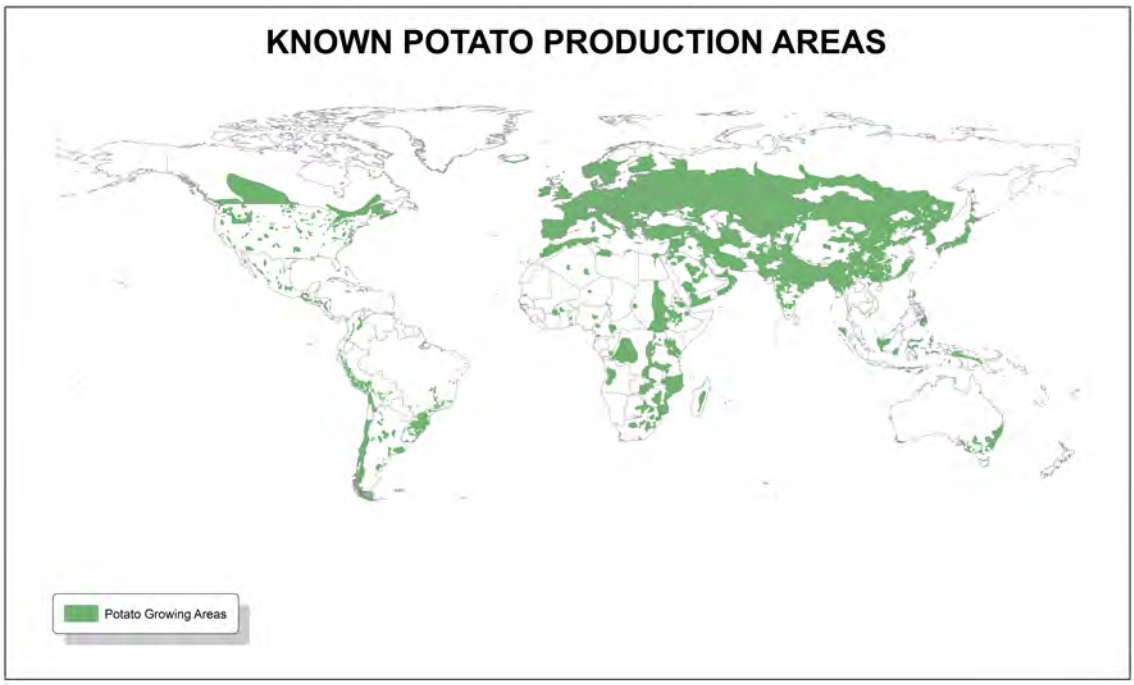

Figure A.2: Global map of areas where potato is known to be grown from International Potato Center (CIP), Lima, Peru, database on global potato production. Data provided by CIP.