

## **ABSTRACT**

TAKEUCHI, YU. An Integrated Spatial Analytic Framework to Manage Invasive Species in Regulatory Phytosanitary Applications (Under the direction of Dr. Frank Koch and Dr. Stacy Nelson)

Non-native pests cause economic and ecological damage to managed and natural U.S. forests and agricultural landscapes. Many insects and diseases are currently under regulatory control in an effort to prevent and manage outbreaks. However, it is difficult to control and minimize the damage once a non-native pest is established. Quick detection and response are required to mitigate invasive species when they are introduced into the United States. Alternatives to rapid response efforts include port inspection, surveillance and monitoring, shipment treatments and pre-clearance programs; however, the international nature of this approach makes it highly complex, operationally difficult and challenging to coordinate.

To advance more timely responses to pest threats, I developed an integrated spatial analytic framework named Spatial Analytic Framework for Advanced Risk Information Systems (SAFARIS) to manage invasive species for phytosanitary management agencies. The SAFARIS is designed to provide a seamless environment for pest predictive models. It supports pest forecast models and tools for researchers, risk analysts, decision/policy makers, rapid-responders, and land managers in need of streamlined and tractable system. The system was designed to assist pest surveys, pest forecasts, pest risk analyses, emergency response, and economic analyses. It allows users to include a variable number of drivers as well as models and can be modified for specific pests in a short time without sacrificing critical elements that serve as the foundation for a science-based

process, especially consideration of epidemiology, population dynamics, impact and overall risk dynamics, including characterization of uncertainty.

The SAFARIS has three main components: 1) databases containing pest model input drivers, 2) databases containing pest biology information and expert-derived parameters; and 3) pest forecast models and related analytic tools. The key characteristics of SAFARIS include 1) streamlining algorithms for data pre-processing, 2) plug and play approach to data drivers and models, 3) algorithmic and data transparency, 4) enabling quantitative evaluation of sources of uncertainty associated with pest forecasts, and 5) integration of climate change effects into pest predictions.

I developed SAFARIS by establishing a base framework, identifying and collecting key bioclimatic data drivers and biological information and integrating pest forecast models and analytic tools within the framework. The current system is designed to grow with the inclusion of more models. It already includes phenology models, insect population dynamics models, and suitable/unsuitable-conditions mapping tools as well as associated data drivers (bioclimatic drivers, pest biology, and host information).

The system was evaluated by assessing the likelihood of Asian gypsy moth (AGM) (*Lymantria dispar* L.) introduction into the conterminous United States from Asian countries. The likelihood of introduction was determined within the SAFARIS environment by analyzing AGM phenology, population dynamics, human-assisted spread, and host availability. Despite the large quantity of materials transported into U.S. ports through Western coastal areas, the highest likelihood of introduction was identified in the Eastern United States.

To incorporate uncertainty evaluations within SAFARIS, I integrated Mean-Variance Frontier (MVF) uncertainty evaluation tools. Although there are many sources of uncertainty associated with pest forecasts, I focused on uncertainty associated with natural variation and model uncertainty.

I estimated AGM climate suitability in the conterminous United States for future conditions (2016-2090) and incorporated uncertainty associated with climate change into pest forecasts. The results were used to prioritize locations (U.S. counties) for management and to assist science-based decisions and allocation of resources. The results were compared with traditional pest forecasts for decision making (no uncertainty incorporation) to illustrate the benefit and importance of including uncertainty information in pest forecasts.

© Copyright 2016 Yu Takeuchi

All Rights Reserved

An Integrated Spatial Analytic Framework to Manage Invasive Species in Regulatory  
Phytosanitary Applications

by  
Yu Takeuchi

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Forestry and Environmental Resources

Raleigh, North Carolina

2016

APPROVED BY:

---

Dr. Frank Koch  
Co-Chair of Advisory Committee

---

Dr. Stacy Nelson  
Co-Chair of Advisory Committee

---

Dr. Ron Stinner

---

Dr. Ron Sequeira

## **DEDICATION**

To my father and mother who gave me a wonderful life

## **BIOGRAPHY**

Yu Takeuchi was born and raised in Tokyo, Japan. Despite living in the middle of Tokyo, she showed great interests in outdoor activities and decided to move to the United States to study Forestry. The rest of her life is still progressing.

## ACKNOWLEDGMENTS

This dissertation was completed because of the love and care my parents took with me. They passed away during the final phases of my Ph.D. study. I miss them terribly and hope that these words of gratitude resonate in those dimensions where my wishes and dreams reside.

I would like to express sincere appreciation to my committee members, Dr. Frank Koch, Dr. Stacy Nelson, Dr. Ron Stinner, and Dr. Ron Sequeira for encouragement and support during my Ph.D study. The journey has been a very tough one; however, you have made it possible for me to continue my study and to write my dissertation. I would not have completed my study if all of my committee members were not there to patiently listen to my dilemmas and provided guidance through difficult steps. Because of our long discussions, I have gained a lot of ideas and opportunities to allow me to remain in the academic environment where I hope to thrive.

I also would like to acknowledge people who provided me a helping hand and countless discussions to support the development of the main tangible outcome of my dissertation, SAFARIS (Spatial Analytic Framework for Advanced Risk Information Systems). In particular, I thank Dr. Glenn Fowler, Mr. Ignacio Baez, Dr. Ryan Boyles, Ms. Heather Aldridge, Ms. Talitha Molet, Ms. Sarah Marnell, Dr. Consuelo Arellano, and Ms. Ann Joseph. In addition, I would like to thank friends and colleagues from PERAL and CIPM for providing me moral support, especially during odd and long hours.

Finally, I thank all my family members – my grandmother, brother, niece, sister-in-law, uncles, and aunts – who have given me unconditional love throughout all these years. I

have gone through the most difficult time in my life during my Ph.D. years, but my family members have been always there to provide me comfort.

## TABLE OF CONTENTS

LIST OF TABLES .....	x
LIST OF FIGURES .....	xi
CHAPTER 1. Alien species in the United States .....	1
1.1 Introduction .....	1
1.2 Global Changes and Trends .....	8
1.2.1 Historical Climate Patterns and Climate Change.....	8
1.2.2 Demographic Changes .....	14
1.2.3 International Travelers .....	16
1.2.4 World Trade and Agricultural Commodity Trade Trends .....	19
1.3 USDA APHIS.....	25
1.3.1 Animal and Plant Health Inspection Service .....	25
1.3.2 Monitoring Invasive Species: Archival Systems at APHIS.....	29
1.4 Trends of Introduced Alien Species in the United States.....	31
1.4.1 Methods.....	31
1.4.2 Interception Data.....	32
1.4.3 Overall Statistics for Introduced Species .....	33
1.4.4 Nonindigenous arthropod species .....	39
1.4.5 Analysis of trends associated with nonindigenous plant pathogens .....	50
1.4.6 Analysis of trends associated with non-indigenous weeds.....	53
1.4.7 Analysis of trends associated with non-indigenous mollusks.....	55
1.5 Examination of correlations between invasive species data and selected independent variables.....	56
1.6 Discussion, Summary, and Conclusions .....	65
1.7 References .....	70
CHAPTER 2. Review of Insect Forecast Systems .....	76
2.1 Introduction .....	76
2.2 Invasive species forecast system .....	77
2.2.1 Brief literature review on forecast systems.....	78
2.2.2 Pattern Matching and fitness-driven models.....	79

2.2.3	Phenology models.....	86
2.2.4	Population dynamics.....	92
2.2.5	Pathway and spread model.....	94
2.3	Historical and Current Insect Forecast Systems in USDA APHIS .....	97
2.3.1	CLIMEX .....	97
2.3.2	NAPPPFAST (NCSU-APHIS Plant Pest Forecasting System).....	103
2.3.3	MaxEnt (VisTrail SAHM).....	107
2.3.4	USPest.....	110
2.4	Model comparisons .....	113
2.4.1	Input variables.....	114
2.4.2	Outputs, post processes, and additional analysis .....	117
2.4.3	Output Comparison.....	119
2.5	Uncertainty .....	122
2.5.1	Epistemic uncertainty.....	123
2.5.2	Linguistic uncertainty .....	127
2.5.3	Evaluation of Uncertainty.....	130
2.6	Summary and Discussion .....	132
2.7	References .....	134
<b>CHAPTER 3. Development of a Spatial Analytic Framework for Advanced Risk Information Systems .....</b>		
		<b>146</b>
3.1	Introduction .....	146
3.1.1	Review of frameworks for pest forecasting.....	147
3.1.2	Pest forecasting and information management in the regulatory agency, USDA APHIS PPQ .....	148
3.2	Concept of a Spatial Analytic Framework for Advanced Risk Information Systems 150	
3.2.1	Description of the components and integration of SAFARIS .....	155
3.2.2	Targeted Users and Audience for SAFARIS.....	156
3.2.3	Determination of the SAFARIS Components.....	156
3.3	Components of SAFARIS .....	158
3.3.1	Information Streams and Data Drivers for Pest Forecasts.....	160
3.3.2	How SAFARIS supports key analysis categories and post processing for pest forecast models.....	189

3.3.3	Integration within the Spatial Analytic Framework for Advanced Risk Information Systems.....	205
3.4	Results .....	216
3.4.1	Development of the web-based phenology model.....	218
3.4.2	Development of the web-based climatologically suitable and unsuitable mapping models .....	221
3.4.3	Testing of SAFARIS.....	223
3.5	Discussion .....	235
3.6	References .....	239
CHAPTER 4.	Uncertainty Associated with Pest Forecasts.....	250
4.1	Introduction .....	250
4.2	Methods.....	256
4.2.1	Climate Data .....	256
4.2.2	Gypsy moth distribution .....	257
4.2.3	Species Distribution Model.....	258
4.2.4	Uncertainty Evaluation .....	261
4.3	Results .....	262
4.4	Discussion .....	306
4.5	References .....	310
CHAPTER 5.	Risk Assessment for Asian Gypsy Moth ( <i>Lymantria dispar</i> ) using the Spatial Analytic Framework for Advanced Risk Information System.....	314
5.1	Introduction .....	314
5.2	Biological Information .....	318
5.3	Analytic Methods .....	322
5.3.1	Spatial Analytic Framework.....	322
5.3.2	Phenology Model.....	323
5.3.3	Population Trends Model.....	327
5.3.4	Modeling Human Assisted Movement and spread.....	339
5.3.5	Asian Gypsy Moth Likelihood of Introduction .....	342
5.4	Results .....	343
5.5	Discussion .....	369
5.6	References .....	371
APPENDICES	.....	377

Appendix A.	Equations to estimate heat units (daily degree days).....	378
Appendix B.	CLIMEX parameter values for <i>Epiphyas postvittana</i> developed by Lozier & Mills (2011) and He et al. (2012).....	395
Appendix C.	A life table of Oriental fruit fly ( <i>Bactrocera dorsalis</i> ).....	397
Appendix D.	Bioclim variables.....	398
Appendix E.	Bioclimatic variables and standard deviations used for MaxEnt models	405

## LIST OF TABLES

Table 1-1 Current and estimated future world population by regions (Source: (UN 2012))..	16
Table 1-2 U.S. major trading partners since 1970 (U.S. Census Bureau, Foreign Trade).....	24
Table 1-3 Number of accidentally introduced alien species first reported in the United States from 1980 to 2014 (Sources: NANIAD and NPAG).....	36
Table 1-4. Correlation matrix between annual establishment number and U.S. trade data....	59
Table 1-5. Correlation matrix between annual establishment number and environmental factors.....	61
Table 1-6. Correlation matrix between annual establishment number and human factors.....	61
Table 1-7. Summary of the re-parameterized species-import models .....	63
Table 3-1. Five GCMs that were selected for current SAFARIS system (MACA 2016) ....	210
Table 4-1. Relative influence of the variable contribution to MaxEnt for historical climatic data.....	274
Table 4-2. Average test AUC for 60 MaxEnt runs and associated standard deviation .....	282
Table 5-1. Male gypsy moth ( <i>Lymantria dispar</i> L.) degree-day parameters for each stage of development.....	326
Table 5-2. Female gypsy moth ( <i>Lymantria dispar</i> L.) degree-day parameters for each stage of development.....	327

## LIST OF FIGURES

Figure 1-1. World average annual temperature difference between 1961-1990 and 1991-2014.....	11
Figure 1-2. Average annual precipitation from 1991 to 2014 .....	13
Figure 1-3. Average annual precipitation difference between average annual precipitation of 1961-1990 and 1991-2014.....	14
Figure 1-4 World Population (Source: (UN 2012)).....	16
Figure 1-5. Number of World Travelers by year since 1995.....	17
Figure 1-6. Number of U.S. inbound and outbound travelers .....	19
Figure 1-7. Global trade values (WTO 2015).....	21
Figure 1-8. The values of world trade, US, and China .....	21
Figure 1-9. US import and export values of all goods since 1948.....	22
Figure 1-10. U.S. agricultural trade values (Source: USDA FAS GATS, 2015) .....	22
Figure 1-11. U.S. agricultural trade volumes (Source: USDA FAS GATS, 2015) .....	23
Figure 1-12 Invasive Species Management Organizations within USDA.....	28
Figure 1-13 Interception records with U.S. agricultural commodity import volume .....	33
Figure 1-14. Number of alien species (arthropods, plant pathogens, mollusks, and weeds) introduced to the United States since 1980 (Source: NANIAD and NPAG) .....	37
Figure 1-15. Number of alien species (Arthropods, Mollusks, Plant Pathogens, and Weeds) introduced to the United States since 1980 (Sources: NANIAD and NPAG).....	38
Figure 1-16. Accumulation of immigrant arthropod species (Source: Sailer, 1978).....	39
Figure 1-17. Number of arthropods accidentally introduced to the United States since 1980 (Sources: NANIAD and NPAG).....	41
Figure 1-18. Order of the immigrant species introduced before 1978 (Source: (Sailer 1978)) .....	43
Figure 1-19. Order of the immigrant species accidentally introduced to the United States since 1980 (Sources: NANIAD and NPAG) .....	44
Figure 1-20. Geographic Regions used for categorization of immigrant species origin (Sailer 1978).....	46
Figure 1-21. Origin of immigrant arthropods before 1978 (Sailer, 1987).....	48
Figure 1-22. Origin of immigrant arthropods accidentally introduced into the United States after 1980 (Sources: NANIAD and NPAG) .....	49
Figure 1-23. Comparison of immigrant arthropod origins between prior to 1978 and post 1980.....	50
Figure 1-24. Types of non-indigenous plant pathogen species introduced to the United States since 1980 .....	51
Figure 1-25. Origins of alien plant pathogens introduced after 1980 (Source: NPAG) .....	53
Figure 1-26 Family of alien weed species in the United States.....	54
Figure 1-27. Origins of alien plant pathogens introduced after 1980 (Source: NPAG) .....	55
Figure 1-28 Origin of mollusks recorded as established in the United States after 1980 in NPAG.....	56

Figure 1-29. The relationships between cumulative import values and number of established alien arthropods.....	64
Figure 2-1. Diurnal temperature shapes based on four methods (average, double triangulation, single sine, and double sine) of calculating degree-day accumulation (Wilson and Barnett 1983).....	91
Figure 2-2. Map of environmental suitability for LBAM resulting from CLIMEX Lozier and Mills (2011) .....	102
Figure 2-3. Predicted potential global distribution of LBAM by CLIMEX. Red color represents EI values ranging from 0-100, which represent the suitability of a location for LBAM establishment (He et al. 2012). .....	102
Figure 2-4. Climate match analysis by NAPPFAST for areas at risk for LBAM establishment. The results are reported in terms of frequency of years from 1999 to 2008 where enough degree days accumulated for LBAM to complete $\geq$ three generations and non-lethal minimum daily temperatures $\geq$ $-16^{\circ}\text{C}$ occurred (Fowler et al. 2009a).....	106
Figure 2-5. Light brown apple moth suitability predicted by using MaxEnt with different sets of input drivers.....	110
Figure 2-6. Annual accumulated degree days for LBAM calculated by USPest based on 2009-2014 climate data.....	113
Figure 3-1. The framework environment.....	160
Figure 3-2. Environmental suitability analysis approaches .....	192
Figure 3-3. Pathways and dispersal means of alien species from infested areas to new colonization sites.....	196
Figure 3-4. Population dynamics analysis in a closed system (no immigrants and emigrants) .....	199
Figure 3-5. Three types of economic impact evaluation analyses (Soliman et al. 2010) ....	202
Figure 3-6. Diagram showing the processes of estimating uncertainties associated with pest forecast outputs within SAFARIS .....	205
Figure 3-7. Models and associated data that are available at web-based SAFARIS .....	215
Figure 3-8. Web interface for degree-day model within SAFARIS .....	219
Figure 3-9. Web interface for degree-day model parameter settings in SAFARIS .....	220
Figure 3-10. Template for setting phenology model parameters.....	221
Figure 3-11. Temperature condition tool that indicates true or false based on the temperature criteria(s) for user defined number of consecutive days.....	222
Figure 3-12. Temperature tool that indicates number of days that user defined criteria met.....	222
Figure 3-13. Precipitation condition tool that indicates true or false based on the criteria users specify.....	223
Figure 3-14. Precipitation tool that indicates accumulated precipitations in millimeters.....	223
Figure 3-15. Flow chart diagram explaining analysis processes within SAFARIS.....	225
Figure 3-16. Potential number of <i>Bactrocera dorsalis</i> generations based on 2014 PRISM climatology .....	226
Figure 3-17. Potential number of <i>Bactrocera dorsalis</i> generations per year based on 2014 METDATA climate data.....	227

Figure 3-18. Potential number of <i>Bactrocera dorsalis</i> generations per year based on 2014 TopoWX climate data .....	228
Figure 3-19. The suitable areas for permanent establishment of Oriental fruit fly ( <i>Bactrocera dorsalis</i> ) based on the PRISM daily average temperatures .....	229
Figure 3-20. Host species distributions for <i>Bactrocera dorsalis</i> generated from 2015 NASS Cropland Data Layer.....	230
Figure 3-21. Forecast on accumulated degree-days for <i>B. dorsalis</i> based on downscaled 2016-2040 CanESM2 data.....	232
Figure 3-22. Forecast on accumulated degree-days for <i>B. dorsalis</i> based on downscaled 2041-2065 CanESM2 data.....	233
Figure 3-23Figure 3-24. Differences in annual accumulated degree-days for <i>B. dorsalis</i> estimated with historical climate data and forecasted climate data .....	234
Figure 4-1. Average annual temperature at RDU airport (35.633°N, -077.383°W) (NOAA 2016a) .....	253
Figure 4-2. (A) Total global mean radiative forcing for the four RCP scenarios based on the Model for the Assessment of Greenhouse-gas Induced Climate Change (MAGICC) energy balance model (Note that the actual forcing simulated by the CMIP5 models differs slightly between models). (B) Time series of global annual mean surface air temperature anomalies (relative to 1986-2005) from CMIP5 concentration-driven experiments. (C) Multi-model ensemble average of annual mean surface air temperature change (compared to 1986-2005 base period) for 2016-2035 and 2081-2100, for RCP 2.6, 4.5, 6.0 and 8.5. (IPCC 2013).....	255
Figure 4-3. Forecasted annual mean temperature for 2066-2090 based on GFDL-ESM2G RCP 4.5 and 8.5 models and expected changes in annual mean temperature from base period of 1990-2005.....	265
Figure 4-4. Forecasted annual precipitation for 2066-2090 based on GFDL-ESM2G RCP 4.5 and 8.5 models and expected changes in annual precipitation from base period of 1990-2005. ....	266
Figure 4-5. Forecasted maximum temperature of the warmest month for 2066-2090 based on GFDL-ESM2G RCP 4.5 and 8.5 models and standard deviations among years.....	268
Figure 4-6. Survey trap locations for European gypsy moth .....	269
Figure 4-7. Gypsy moth positive locations used in Maxent to estimate habitat suitability ..	271
Figure 4-8. Probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G 1990-2005 data.....	278
Figure 4-9. Probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP4.5 averaged data by three time periods (2016-2040, 2041-2065, and 2066-2090).....	279
Figure 4-10. Probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP8.5 averaged data by three time periods (2016-2040, 2041-2065, and 2066-2090).....	280
Figure 4-11. Predicted future habitat for gypsy moth based on 2066-2090 data from MACA downscaled GFDL-ESM2G RCP 8.5 .....	281

Figure 4-12. Probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP4.5 data (mean and randomly generated climatic variables based on mean and standard deviation by time periods).....	285
Figure 4-13. Probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP8.5 data (mean and randomly generated climatic variables based on mean and standard deviation by time periods).....	286
Figure 4-14. Predicted future habitat for gypsy moth for 2066-2090 data estimated with mean and randomly generated climate data from MACA downscaled GFDL-ESM2G RCP 4.5 and RCP 8.5 .....	287
Figure 4-15. Maximum and minimum probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP4.5 data (mean and randomly generated climatic variables based on mean and standard deviation by time periods) .....	290
Figure 4-16. Maximum and minimum probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP8.5 data (mean and randomly generated climatic variables based on mean and standard deviation by time periods) .....	291
Figure 4-17. Scatter plot between probability and standard deviation at county level derived from MaxEnt model with GFDL ESM2G RCP 8.5 2066-2090 data (mean and randomly generated climatic variables).....	295
Figure 4-18. Relative rank estimated by mean-variance frontier approach and coefficient of variation for GFDL ESM2G RCP 8.5 2066-2090 data (mean and randomly generated climatic variables).....	296
Figure 4-19. Suitability of gypsy moth for 2066-2090 based on GFDL ESM2G RCP 8.5 model.....	298
Figure 4-20. Scatter plot between probability and standard deviation at county level derived from MaxEnt model with GFDL ESM2G RCP 8.5 2066-2090 data (mean climatic variables).....	299
Figure 4-21. Relative rank estimated by mean-variance frontier approach and coefficient of variation for GFDL ESM2G RCP 8.5 2066-2090 data (mean climatic variables)...	300
Figure 4-22. Scatter plot between probability and standard deviation at county level derived from MaxEnt model with GFDL ESM2G RCP 4.5 2066-2090 data (mean and randomly generated climatic variables).....	303
Figure 4-23. Relative rank estimated by mean-variance frontier approach and coefficient of variation for GFDL ESM2G RCP 4.5 2066-2090 data (mean and randomly generated climatic variables).....	304
Figure 5-1. Quarantine Areas for European Gypsy Moth ( <i>Lymantria dispar</i> ) in the continental United States as of April, 2015 .....	316
Figure 5-2. Asian gypsy moth detections in the United States .....	317
Figure 5-3. National forest type groups in the continental United States.....	321
Figure 5-4. Forest biomass (mg per hectare) in the continental United States.....	322
Figure 5-5. Monthly average temperature for winter months developed by using PRISM daily Tmean data.....	333

Figure 5-6. The effect of temperature on the developmental rates of GM larvae (data from Casagrande et al. (1987)) .....	334
Figure 5-7. The effect of temperature on the developmental rates of GM pupae (data from Casagrande et al. (1987)) .....	335
Figure 5-8. The effect of temperature on daily mortality rate at larval stage (data from Casagrande et al. (1987)) .....	336
Figure 5-9. Mean percentage hatch at different number of days that egg mass had been exposed to temperature at 5°C (from Keena (1996)).....	337
Figure 5-10. Mean hatching rate estimates based on the experiments by Keena (1996).....	338
Figure 5-11. Population dynamic model at each life stage of gypsy moth.....	339
Figure 5-12. Annual accumulated degree-days at lower and upper threshold temperatures of 3°C and 38°C .....	345
Figure 5-13. Expected gypsy moth eclosion week based on 2009-2013 historical temperature data.....	346
Figure 5-14. Expected gypsy moth pupation week based on 2009-2013 historical temperature data.....	347
Figure 5-15. Expected gypsy moth emergence week based on 2009-2013 historical temperature data.....	348
Figure 5-16. Expected emergence week based on 2009 historical climate data and GM biology parameters derived from experiments by Casagrande et al. (1987) .....	353
Figure 5-17. Expected emergence week based on 2010 historical climate data and GM biology parameters derived from experiments by Casagrande et al. (1987) .....	354
Figure 5-18. Temperature induced mortality rate during larval stage .....	355
Figure 5-19. Temperature induced mortality rate during pupal stage .....	356
Figure 5-20. Egg hatching rate predicted by exposure to cold climate conditions.....	357
Figure 5-21. Gypsy moth suitability expressed with relative population growth.....	358
Figure 5-22. European gypsy moth quarantine area in Maine (map created by USDA APHIS PPQ (2011)) .....	359
Figure 5-23. The commodity volume flows by trucks from three AGM at-risk ports (Los Angeles, Seattle, and Charleston) .....	363
Figure 5-24. Most likely routes from Seattle, Los Angeles, and Charleston to common destinations .....	364
Figure 5-25. The relative traffic volume expressed between 0 and 1 along the most likely routes from at-risk ports.....	365
Figure 5-26. The host species availability determined by tree species types and density along most likely routes.....	366
Figure 5-27. The relative likelihood of AGM introduction within the contiguous United States from Los Angeles, Seattle, and Charleston by truck movements .....	367
Figure 5-28. Likelihood of Asian gypsy moth introduction based on phenology model, temperature driven demographic model, and human assisted movement model. ....	369
Figure A-1. Six possible relationships can exist between the daily temperature cycle and the upper and lower thresholds (UCIPM 2016).....	379

Figure A-2. Mean diurnal range (Bioclim 2) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model .....	405
Figure A-3. Mean diurnal range (Bioclim 2) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model .....	406
Figure A-4. Minimum temperature of the coldest month (Bioclim 6) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model .....	407
Figure A-5. Minimum temperature of the coldest month (Bioclim 6) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model .....	408
Figure A-6. Mean temperature of the wettest quarter (Bioclim 8) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model .....	409
Figure A-7. Mean temperature of the wettest quarter (Bioclim 8) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model .....	410
Figure A-8. Precipitation seasonality (Bioclim 15) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model .....	411
Figure A-9. Precipitation seasonality (Bioclim 15) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model .....	412
Figure A-10. Precipitation of the warmest quarter (Bioclim 18) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model .....	413
Figure A-11. Precipitation of the warmest quarter (Bioclim 18) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model .....	414
Figure A-12. Precipitation of the coldest quarter (Bioclim 19) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model .....	415
Figure A-13. Precipitation of the coldest quarter (Bioclim 19) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model .....	416
Figure A-14. Shuttle Radar Topography Mission Elevation data for the contiguous United States .....	417

## CHAPTER 1. Alien species in the United States

### 1.1 Introduction

The topic of my study revolves around establishing a forecasting system for invasive species of regulatory concern. This chapter will review patterns and trends associated with invasive species for the past few decades. Other studies have explored patterns associated with pest interceptions at US ports (Haack 2001, Work et al. 2005, Haack 2006, McCullough et al. 2006, Koch et al. 2011) and European ports (Kenis et al. 2007, Eschen et al. 2015). Here, I will focus on exploring trends associated with actual introductions (i.e., establishment and spread) of new species of regulatory concern in the US. This baseline understanding will serve to anchor my work on developing a forecasting framework to characterize the behavior and impact of invasive species.

Invasive alien species cause significant economic and environmental damage. Pimentel et al. (2005) estimated that invasive species losses in the United States amount to \$120 billion per year; although others like Nghiem (2013) claimed that whereas damage was significant, the methods used by Pimentel et al. were crude and did not consider the real nature of damage that occurs in the dimensions of time and space, such that the present value of future losses may be a fraction of the number claimed by frequently cited authors like Pimentel et. al. (2005). Although I agree that the methods of Pimentel et al. are crude, I note that the budgets of Federal agencies like USDA Animal and Plant Inspection Service (APHIS), Agricultural Research Service (ARS) and U.S. Geological Survey (USGS) that focus on invasive species are upwards of one billion dollars per year (USDA 2015), and formal economic analyses of damage by selected species like fruit flies and emerald ash

borer have reported economic damage in the billions of dollars for several species (Vo et al. 2014). Some of the damage caused by these species is related to direct negative effects on their host species, while sometimes the damage is related to the efforts to combat the pests. Insects and diseases such as European gypsy moth (*Lymantria dispar* Linnaeus) and chestnut blight (*Cryphonectria parasitica* (Murrill) Barr) have indeed caused serious damage and altered U.S. forests and ecosystems permanently. Importantly, independent of how primitive the approaches used may be, most scientists, including authors like Nghiem (2013), still recognize the value of forecast models in establishing more rational estimates of losses and overall impacts.

Alien pests are often introduced unintentionally because of trade linked to globalization (Hulme 2009, Aukema et al. 2010, Koch et al. 2011, Nghiem et al. 2013). Today, people and commodities tend to move invasive species more quickly and over longer distances than non-human-assisted processes.

Since 1970, the volume of global imports has increased four-fold (Hulme 2009) and is expected to continue increasing, especially since newly established treaty agreements such as the Trans Pacific Partnership (TPP) and Transatlantic Trade and Investment Partnership (TTIP) (The Economist 2015) are expected to deeply impact agricultural trade (USDA FAS 2015b). In the United States, the total value of agricultural trade (imports and exports) was \$226 billion in 2015 (USDA FAS 2015a). It is a challenge to protect natural resources and prevent additional introductions of new alien species because of the tension between safeguarding (including regulatory) efforts, human-assisted movement and natural spread.

The very notion of regulating the movement of products is a relatively new concept. Although introduction of alien species is well understood to be linked with international trade (Levine and D'Antonio 2003, Hulme 2009), we also observe that the massive increase in invasive species introductions recorded by Sailer (1978) coincides well with the establishment of the USDA by President Abraham Lincoln in 1862. If there is causality associated with that correlation, it may be linked to the intense support USDA gave to foreign exploration and importation of plant species that might be adapted and exploited in the United States.

Each year, a large number of previously unrecorded species arrives in the United States. However, only a small fraction of those species become established (Williamson and Fitter 1996, Liebhold and Tobin 2008, Aukema et al. 2010) and a smaller number becomes economically or environmentally detrimental. This is due to many factors, but some are likely key. For example, unintentionally imported agricultural pests need to find suitable host species, suitable environments and arrive in sufficient numbers to overcome genetic bottlenecks and other abiotic limitations.

First, the important barrier to successful establishments of agricultural pests in new environments is the location of suitable hosts. Although the likelihood of a given alien species finding suitable host materials varies with species and is different each time, finding suitable hosts is strongly linked to the type of trade. Agricultural commodities for consumption have lower likelihood of introduction of new species because life cycles are destroyed or disrupted by the disposition of the commodity (consumption or disposal). In contrast, propagative materials which, by definition, arrive as viable plant materials and are

destined for planting, have a higher likelihood of introducing pests because in the absence of effective treatments, the life cycles of associated invasive species are unlikely to be disrupted and they are often transported and distributed in large numbers (Kenis et al. 2007, Liebhold et al. 2012).

Second, alien species must have environmentally suitable conditions in order to establish in a new area and be capable of spreading from newly colonized locations. For example, most areas in the continental United States are not suitable for tropical species; therefore, alien species introduced from extant ranges in the tropics are likely to only become established in limited areas (e.g., south Florida, Mexican Gulf Coast, and parts of California).

Third, a sufficient population is required for establishment of an invader. Low-density populations can simply become extinct due to inimical random variation (Liebhold and Tobin 2008). Also, many biological factors, such as finding mates, predators, and lack of genetic diversity, can cause Allee effects (Liebhold and Tobin 2008). The population of an alien species arriving through trade frequently may not attain a propagule pressure necessary for successful establishment (Liebhold and Tobin 2008).

An increasingly important pathway is the introduction of alien species through non-agricultural commodities. Solid wood packing material is a common example but is far from an isolated one. Solid wood packing materials are commonly used in freight shipments for packing support, bracing and dunnage. Many forest species, such as Asian longhorned beetle (*Anoplophora glabripennis* (Motschulsky)) and the emerald ash borer (*Agrilus planipennis* Fairmaire), are likely to have been introduced to the United States through solid wood packing materials (Poland et al. 1998). As a result of the introduction of Asian longhorned

beetle, ISPM 15 (International Standards for Phytosanitary Measures: ISPM) was developed by the International Plant Protection Convention (IPPC) (IPPC 2013). This document requires wood packing materials to be treated by heat treatment or methyl bromide for international trade (IPPC 2013). Other non-traditional potential pathways for invasive species spread include shipping containers; transportation vessels; materials that remain exposed for long periods at staging areas or warehouses before shipping (e.g., recycling products, construction materials); movement of materials and people related to military operations; hobbyists, adventure travel, collectors; humanitarian aid; academic research; and illegal activities, including smuggling (Zhang et al. 2007, Fowler et al. 2009).

Some introduced alien agricultural pests can cause significant economic and environmental damage. Citrus greening disease (*Candidatus Liberibacter asiaticus*), which was first detected in Florida in 2005, had caused more than \$3.63 billion damages to the state's citrus industry by 2012 (Nordlie 2012), and it is now affecting citrus production elsewhere in the United States. The Asian longhorned beetle (ALB) (*Anoplophora glabripennis* Motschulsky) was first discovered in 1996 and has since been under official control. Despite USDA APHIS being able to successfully eradicate ALB in New Jersey and Chicago areas, this invasive insect is still spreading and currently distributed in New York, Massachusetts, and Ohio (Rose 2014). More uncontrolled spread has characterized the movement of the emerald ash borer (EAB) (*Agrilus planipennis* Fairmaire). EAB was first detected in Michigan in 2002. Currently, EAB has spread to 25 states (Arkansas, Colorado, Connecticut, Georgia, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Hampshire, New Jersey, New York,

North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, and Wisconsin). More than 53 million native ash (*Fraxinus* sp.) were killed by EAB as of 2007, and a simulation model predicted the discount cost of treatment, removal and replacement of ash trees by 2019 to be \$10.7 billion (Kovacs et al. 2010). The classic example is chestnut blight disease (*Cryphonectria parasitica* (Murrill) Barr). It was introduced to the United States in the early 1900s and completely destroyed American chestnut (*Castanea dentate* Marsh.), a dominant hardwood forest species in eastern United States (Anagnostakis 1987).

As global trade and human travel increase each year, the likelihood of introducing new invasive species increases (Hulme 2009, Aukema et al. 2010, Koch et al. 2011, Nghiem et al. 2013). Understanding the historical patterns of introductions could help estimate imminent threats and prevent further introductions of devastating pests into the United States. Several studies have been conducted to examine historical patterns of introductions on alien species (Sailer 1978, Mack 2003, Aukema et al. 2010). Sailer (1978) is of particular interest to this study because he established a baseline and sought to identify trends and patterns associated with invasive species. His study focused on immigrant arthropod species and examined trends by taxonomy, origin, and economic importance. His study indicated that the introduction of alien species into the continental United States was closely related to trade patterns; the number of insects introduced each year corresponded with the amount of foreign trade (Sailer 1978). In addition, Sailer pointed out that the majority of the insects introduced into the United States prior to 1977 originated from Europe, which was the largest commerce partner through that time (Sailer 1978). Similar studies have been conducted focusing on

wood-boring forest insects (Koch et al. 2011), forest insects and pathogenic species (Aukema et al. 2010) and plants (Mack 2003).

In addition, Levine & D'Antonio (2003) parameterized the relationships between alien species establishment data from Sailer (1978) and plant pathogens and mollusks as reported by the Office of Technology Assessment and merchandise import values in the United States by three models (1. Log-log species-area model, 2. Log-linear species-area model, and 3. Michaelis-Menten model). Levine & D'Antonio (2003) then estimated how many non-indigenous species (insects, plant pathogens, and mollusks) could be introduced to the United States between 2000 and 2020 by using those three models. The results from the three models varied significantly. For example, the log-log species-area model predicted 1,535 insect species introduced from 2000 to 2020, while the log-linear species-area model and the Michaelis-Menten model predicted 492 and 115 insect species introductions, respectively.

This study will characterize recent introduction patterns and contrast those to the baseline established by Sailer (1978). For this study, we have expanded the taxonomic groups. We have thus included plant pathogens, mollusks, and weeds in addition to arthropods. This type of trend analysis will help regulatory agencies identify phytosanitary risks associated with alien species invasions and to prevent/minimize new introductions to the United States. Also, it helps us understand the baseline for the development of my forecasting framework which is the topic of the main body of my research.

For this section, I particularly focused on changes in climate patterns (temperature and precipitation), demographic patterns, international traveler patterns, and global trade

patterns. I explored surface relationships (association or correlation) with introductions. I integrated and analyzed information from three main sources, 1) PestID (interception records at U.S. ports of entry), 2) North American Non-Indigenous Arthropod Database (NANIAD), and 3) New Pest Advisory Group Database (NPAG). The PestID and NPAG databases are accessible through cooperative agreements and are managed by USDA APHIS. The NANIAD database was developed by K.C. Kim at Pennsylvania State University as a product of a cooperative agreement with USDA APHIS PPQ. The NANIAD data are currently available at the USDA National Agricultural Library. No one database had complete information and extensive efforts were made to address data gaps and work towards a standardized presentation of the information, as I will detail shortly.

Finally, I evaluated alien species trends in the United States and relationships among alien species introductions and global parameters to identify the factors that most contribute to alien species introductions to the United States.

## **1.2 Global Changes and Trends**

### **1.2.1 Historical Climate Patterns and Climate Change**

Climate patterns have been changing gradually over time. The changes have been caused by numerous factors, but one of the causal factors is the increase of greenhouse gases, especially carbon dioxide (CO<sub>2</sub>). Since the Industrial Revolution period in the 18<sup>th</sup> century, human activities, such as burning plant matter, burning coal, consumption of fossil fuels and deforestation, have caused significant increases of CO<sub>2</sub> in the atmosphere. In average, CO<sub>2</sub> concentrations measured in 2005 were 35% higher than the ones observed during the

Industrial Revolution period (1760 - 1840) (Azar and Rodhe 1997, UCAR 2016). Recent increase in CO<sub>2</sub> concentration in the atmosphere (since 1960) is often plotted in a figure that shows a monotonic increase (Mauna Loa Observatory CO<sub>2</sub> observations from 1960 to 2010: (NOAA 2015b). The most recent observations show that the trends have not changed (NOAA 2015c).

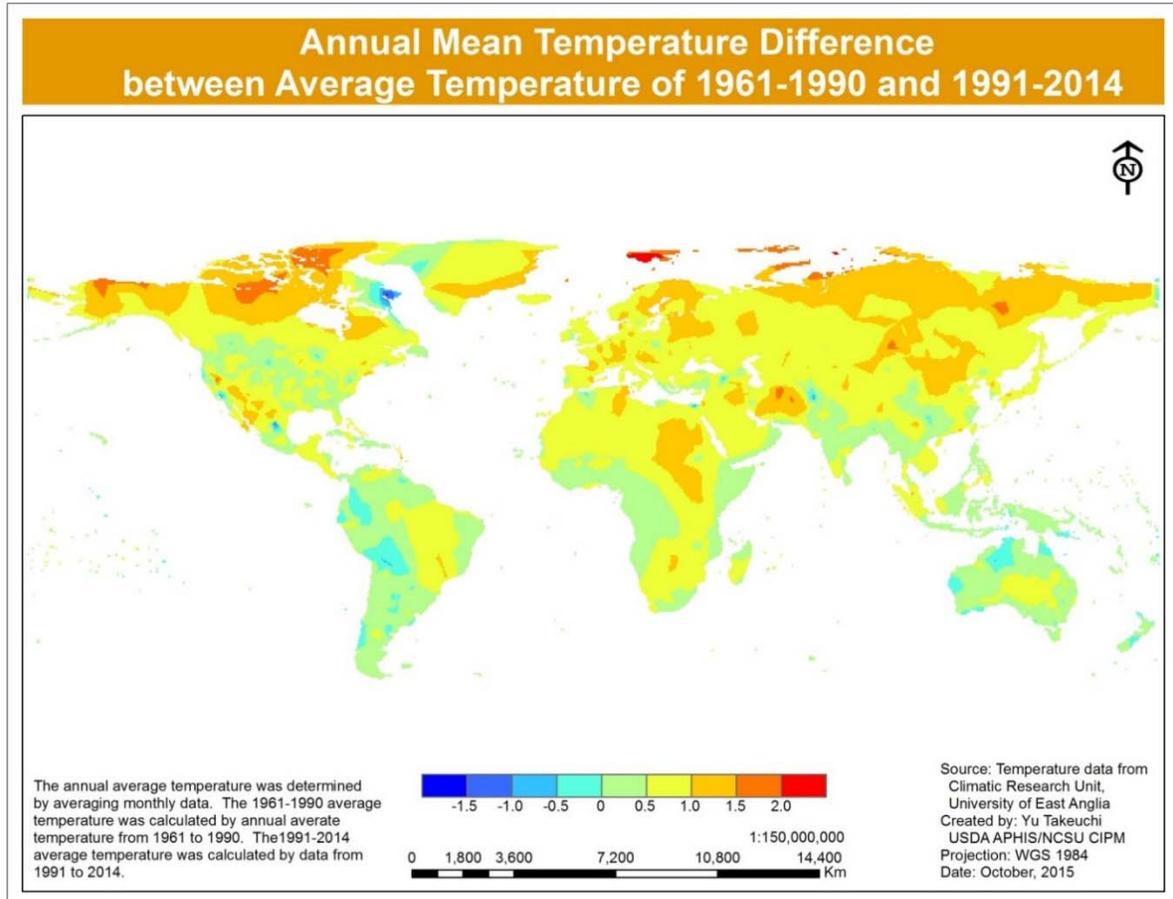
Climate change includes changes in air temperature, precipitation, sea level, and vegetation. Perhaps most importantly, the latest patterns (frequencies, amplitude) are already departing from expectations based on long-term observations. Although their exact responses vary, species behave differently under different environments. Global warming changes habitats and distributions. Warm conditions may favor cold intolerant species, and they may expand their suitable habitats (Parmesan 2006, Kurz et al. 2008). Insect species tend to increase their growth rates under warmer conditions, allowing insects to have more generations in a year and/or higher population sizes (Ziter et al. 2012, Srygley 2014).

Gridded data depicting monthly average temperatures across the world were obtained from the Climatic Research Unit at the University of East Anglia (CRU TS v. 3.23: <http://www.cru.uea.ac.uk>) and were further averaged by each year. The annual mean temperatures were then averaged from 1961 to 1990 and from 1991 to 2014. Although the number of years in these two time periods are different, the annual average temperature for more recent years (1991 to 2014) was compared to prior years (1961 to 1990) because CLIMEX, one of the commonly used pest forecast system, uses climatic variables derived from observations from 1961 to 1990. In addition, the latest available CRU TS v. 3.23 was 2014. Figure 1-1 shows the differences between the average annual mean temperature of

1961-1990 and 1991-2014. The light orange, orange and red colors indicate more than 1°C hotter in 1991-2014 compared to 1961-1990, while blue colors indicate colder conditions in 1990-2014 compared to 1961-1990.

Cumulative observations show that the majority of the areas in the world are experiencing slight temperature increases (up to 1°C increase, with the majority of areas showing no to small increases). The areas that exhibited greater temperature increases (greater than 1°C increase) are in northern Alaska, northern Canada, northern Europe, northern Russia, Inner Mongolia and Mongolia, Sudan, and Eastern Iran to Western Afghanistan. On the other hand, some areas showed colder environments in 1961-1990 compared to 1991-2014 including Bolivia, northern and western coastal areas of Australia, and Canterbury, New Zealand.

Most of the United States exhibited a slight increase or slight decrease (-0.5 to 1°C) in the average annual temperature from 1991 to 2014 compared to the one from 1961 to 1990, except the Alaska and Sierra Nevada mountain ranges. Both mountain ranges experienced a greater than 1°C increase in 1991-2014.

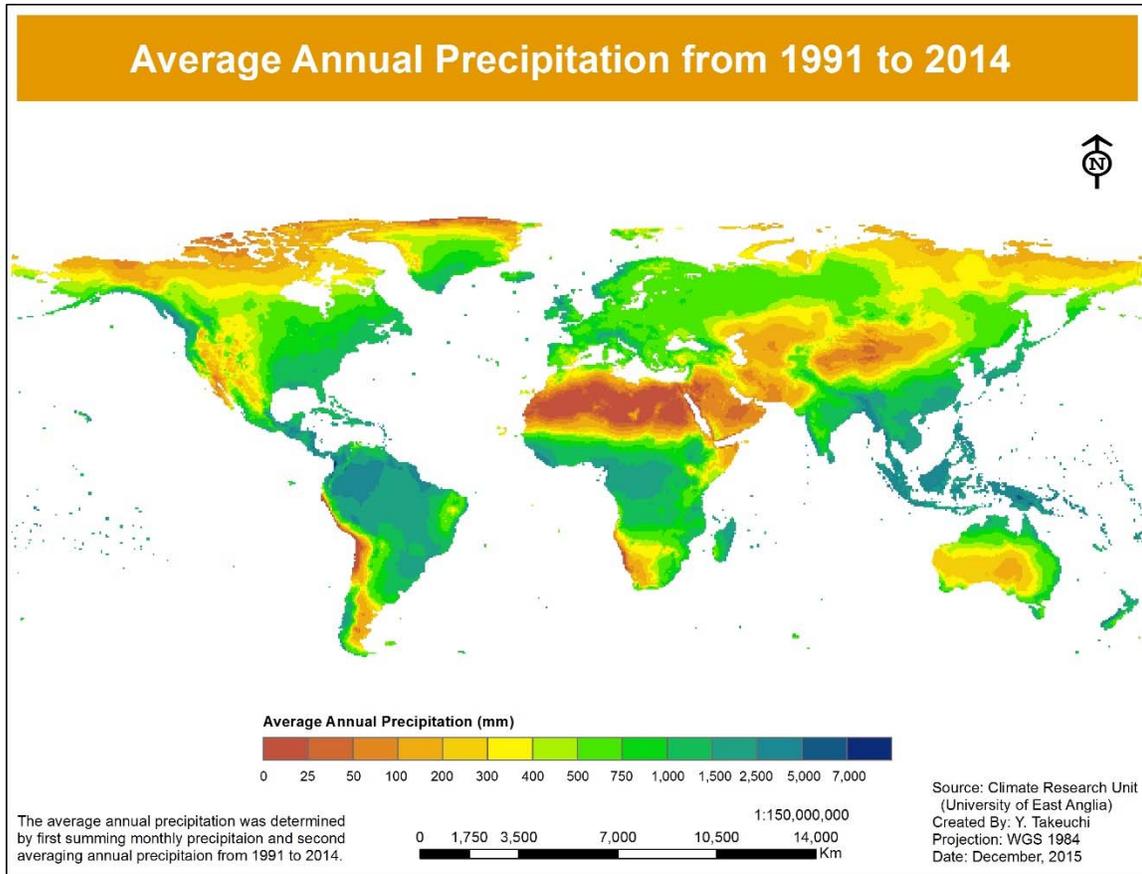


**Figure 1-1. World average annual temperature difference between 1961-1990 and 1991-2014**

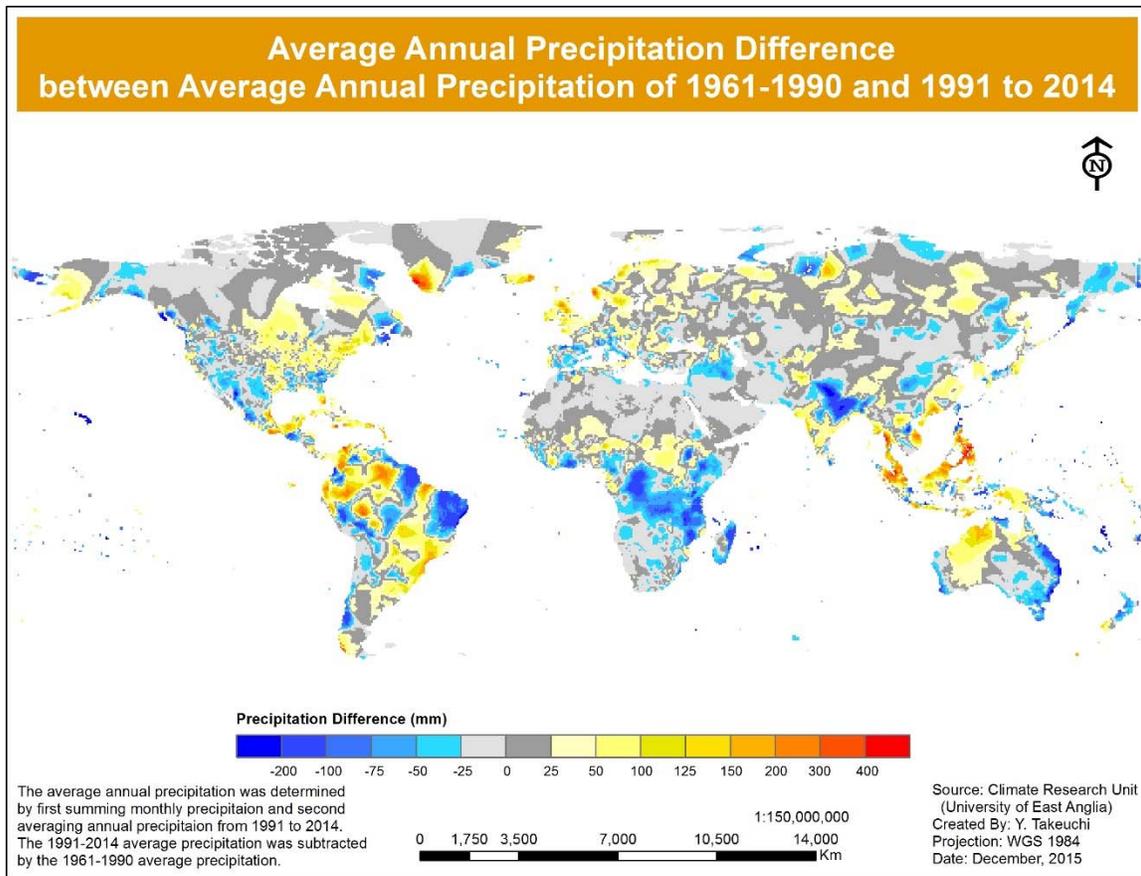
Gridded data depicting monthly precipitation were obtained from the Climatic Research Unit at the University of East Anglia (CRU TS v. 3.23: <http://www.cru.uea.ac.uk>), and were further summed by each year. The annual precipitation measurements were then averaged from 1961 to 1990 and from 1991 to 2014. Figure 1-2 shows the average annual precipitation in millimeters across the world based on the precipitation from 1991 to 2014. Areas in brown or orange colors indicate very little precipitation. Those areas include northern North America (northern Alaska and northern Canada), western United States, coastal areas of Peru, northern Chile, southern Argentina, southwest Southern Africa, northern Africa (Sahara Desert), Arabian Peninsula, Central Asia, northern East Asia,

northern Russia, and southwest Australia. On the other hand, areas receiving large amounts of precipitation are indicated in dark green to blue colors. Those areas are the northwest coast of North America, northern South America, Central Africa, and Southeast Asia.

The total amount of precipitation is expected to change under climate change. Figure 1-3 indicates differences in average annual precipitation between 1961-1990 and 1991-2014. The grey areas indicate small changes in the amount of precipitation. The blue areas indicate more than 25mm reduction, while the yellow to red areas indicate areas with more than 25mm increase. The eastern United States, northern South America, southern Brazil, Southeast Asia and Western Australia received more precipitation in recent years compared to the 1961-1990 period.



**Figure 1-2. Average annual precipitation from 1991 to 2014**



**Figure 1-3. Average annual precipitation difference between average annual precipitation of 1961-1990 and 1991-2014**

### 1.2.2 Demographic Changes

The current world population is just over 7 billion people (US-Census-Bureau 2012). This number is projected to grow to 9.3 billion by 2050 (UN 2012) at an annual growth rate of 0.92% from 2000 (UN 2004). Although the world population is projected to increase at this steady rate through 2050, the growth rate is lower compared to 1950 to 2000, a period during which the world population increased from approximately 2.5 billion to over 6 billion (Figure 1-4).

Food production has also changed in recent years. Food production has outpaced population growth mainly due to the development and use of improved plant varieties,

improvement of fertilization, agricultural land increase, more effective control and management of pests, improved strains of livestock and poultry, and wider use of nutritionally balanced feeds (Gilland 2002). Despite a significant human population increase, the global supply of food calories per capita has also been increasing (Gilland 2002). Based on a Food Balance Sheet by Food and Agriculture Organization of the United Nations, the per capita global food supply was 2200 kilocalories/day in 1961 and 2512 kilocalories/day in 1981 (FAO 2012). More recent data indicated that the per capita global food supply increased to 2796 kilocalories/day in 2007 (FAO 2012).

The greatest current challenge is to estimate how long food production can keep pace with population growth (Gilland 2002) and how food can be supplied globally. Currently, Asia is the most populated region in the world (Table 1-1), and is estimated to remain the most populated region in 2100. Africa and Oceania are anticipated to have significant population increases, while Europe is expected to decrease in population from 2000 to 2100. From 2000 to 2100, Europe's share of the world population is projected to be nearly halved (12 to 7%), while Africa's share is expected to more than double (13 to 35%). Due to significant population increases in Africa, the Asian population will represent 45% of the world population (60% in 2000) by 2100. These demographic changes should alter agricultural commodity trade patterns between the United States and foreign countries, which may result in changing pest invasion pathways.

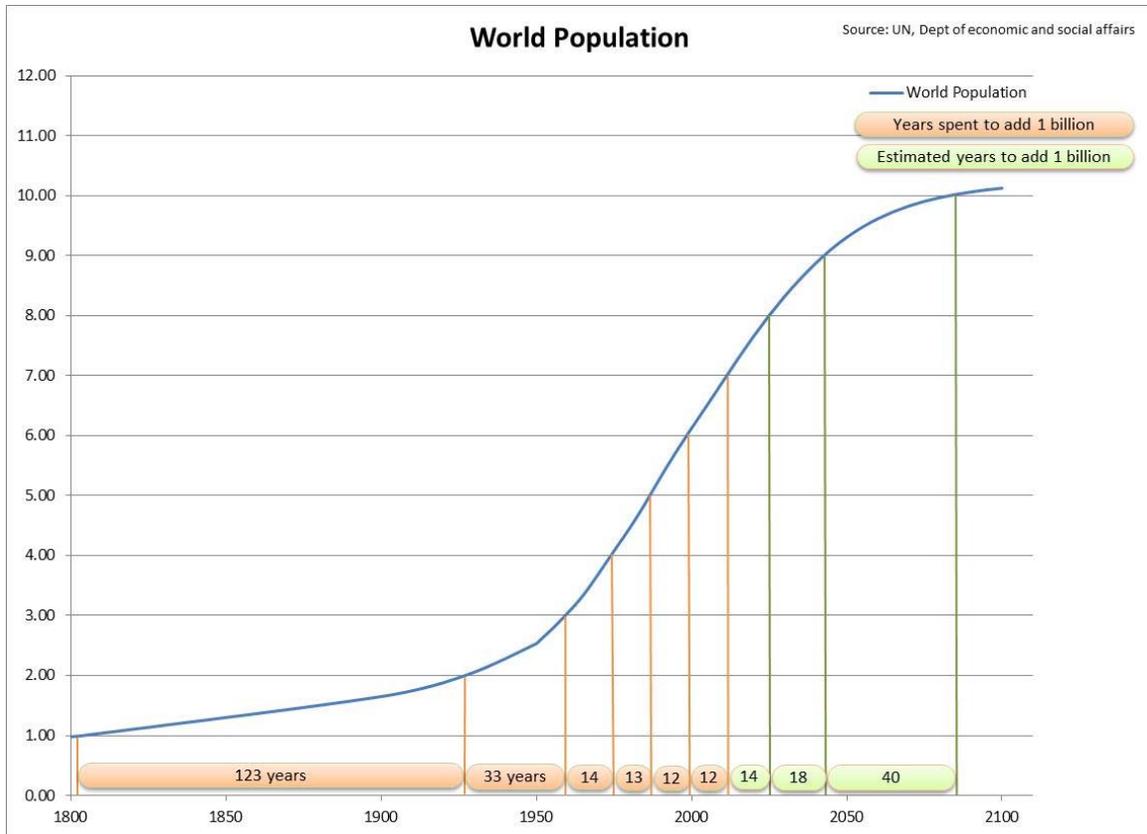


Figure 1-4 World Population (Source: (UN 2012))

Table 1-1 Current and estimated future world population by regions (Source: (UN 2012))

Region	2000 (thousands)	2010 (thousands)	2100 (thousands)	Change from 2000 to 2100
Africa	811,101	1,022,234	3,574,141	340.7%
Asia	3,719,044	4,164,252	4,596,224	23.6%
Europe	726,777	738,199	674,796	-7.2%
Latin America & Caribbean	521,429	590,082	687,517	31.9%
North America	313,289	344,529	526,428	68.0%
Oceania	31,130	36,593	65,819	111.4%
World	6,122,770	6,895,889	10,124,926	65.4%

### 1.2.3 International Travelers

The number of people who travel to a country other than that of their residence for a period not exceeding 12 months, and whose main purpose in visiting is other than an activity

remunerated from within the country visited, has increased significantly since 1995 (The World Bank 2015). In 1995, there were approximately 658 million people who traveled out of their usual residence for a period less than 12 months, while there were over 1.2 billion travelers in 2013 (The World Bank 2015) (Figure 1-5).

The number of global travelers has been increasing at a constant rate. A linear regression line (Figure 1-5) indicates that there were approximately 30 million more people traveling each year since 1995 ( $R^2 = 0.98$ ).

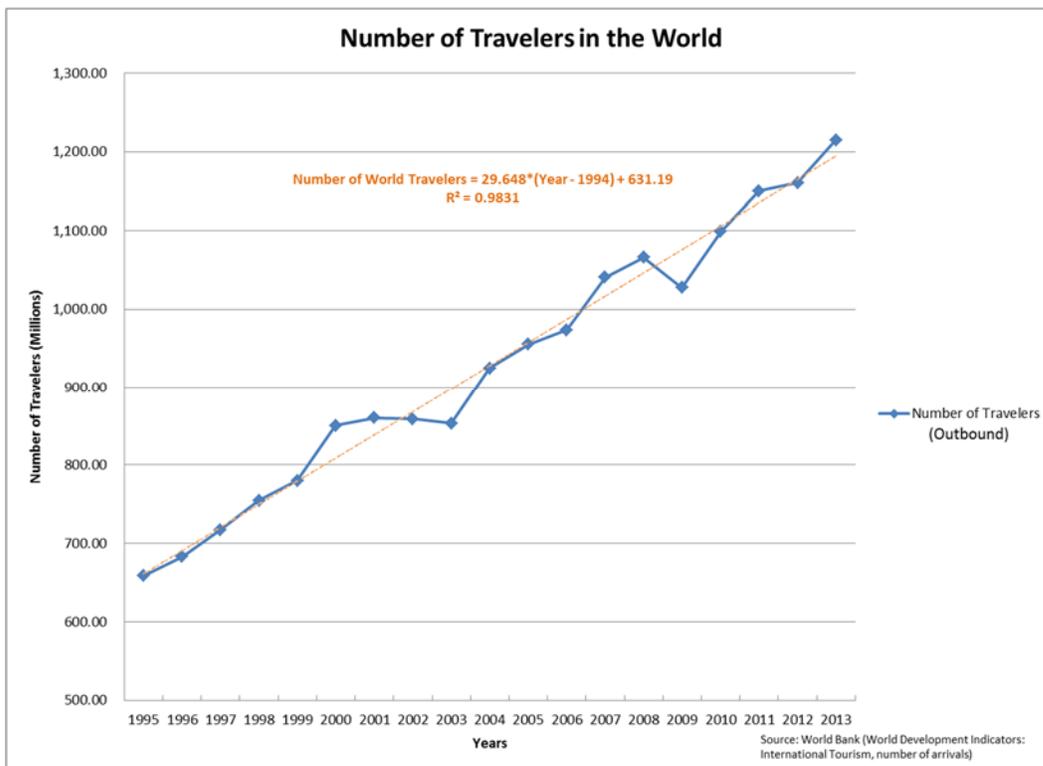


Figure 1-5. Number of World Travelers by year since 1995

US travelers (outbound) and foreign travelers coming into the United States have been also increasing since 1995 (Figure 1-6) (The World Bank 2015). The outbound number in Figure 1-6 represents the number of departures that US residents made from the US to other countries for any purpose other than a remunerated activity in the country visited. The

number of US outbound travelers fluctuates each year and does not show a strong linear increase since 1995 ( $R^2 = 0.61$ ), although there is a clear increasing trend. Approximately 51 million US residents traveled outside of the United States in 1995, while over 68 million US residents traveled outside of the United States in 2014.

Foreign nationals traveling to the United States (inbound) have also increased since 1995 (Figure 1-6). There was a first peak in 2000, and then the number of foreigners traveling to the United States dropped until 2003. This might be result of the terror attacks of September 11, 2001 (Cornwell and Roberts 2010). Since then, foreign travelers have increased at a steady and rapid rate, other than a slight drop in 2009; a linear regression line ( $R^2 = 0.96$ ) fit to the data shows an increase of 2.7 million more people per year since 2003. Until 2010, the number of U.S. residents traveling was always greater than the number of foreigners traveling into the United States. In 2011, the number of inbound foreigners exceeded outbound US travelers for the first time, and it has been more foreign travelers since 2011.

Approximately 140 million passengers entered the United States in 2014 assuming that US residents who travel overseas return to the United States in the same year. Since 1997, over 100 million passengers entered the United States each year except 2003 (97.43 million passengers in 2003).

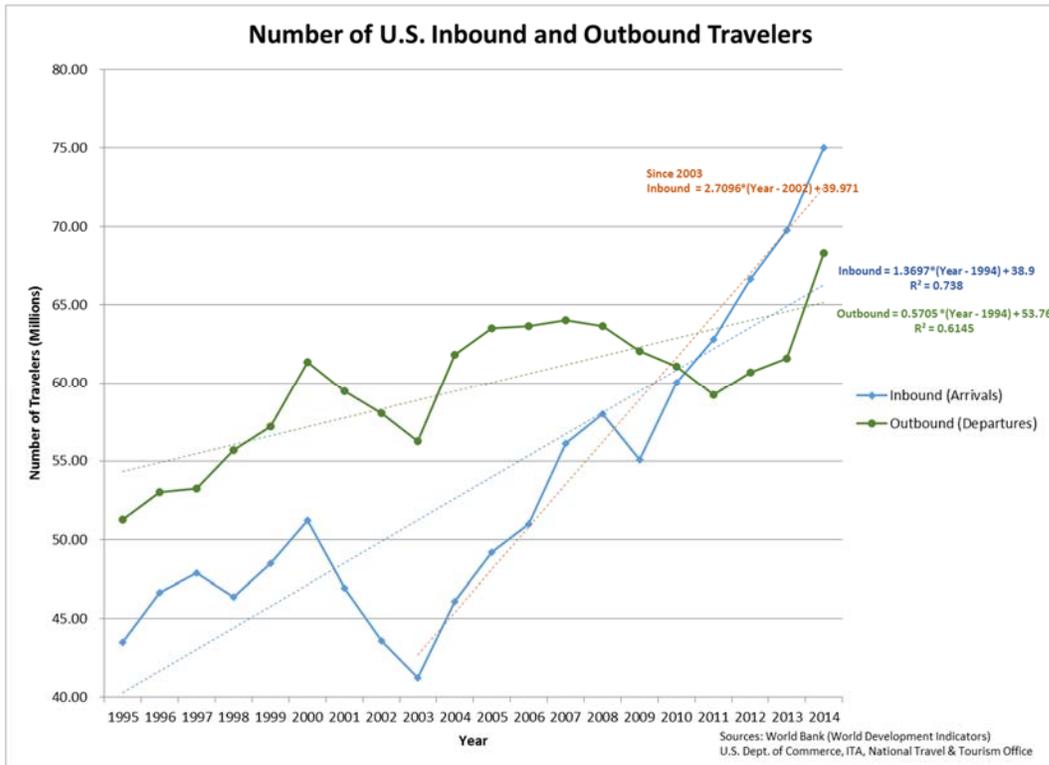


Figure 1-6. Number of U.S. inbound and outbound travelers

### 1.2.4 World Trade and Agricultural Commodity Trade Trends

Global trade has been increasing as economies have grown since 1800 (Hulme 2009). However, global trade has increased most rapidly since 1970 (Figure 1-7). The value of world trade has increased almost 300-fold since 1950. The US trade trend (Figure 1-8) and world agricultural trade trend (Figure 1-7) are similar to the world trade trends for all goods. US trade has expanded nearly 200-fold since 1950. The United States has exported slightly more than it imported until 1975; since 1975, the United States has been importing more materials than exporting. In 2014, the United States imported and exported goods valued at US\$2.412 trillion and US\$1.62 trillion, respectively (WTO 2015). Trade of agricultural commodities has also significantly increased since 1970. Agricultural commodity exports were slightly more than imports in the United States until early 1970 (USDA FAS 2015a)

(Figure 1-10); since 1972, agricultural commodity exports have increased rapidly and are significantly more than imports in the United States. It has been the opposite with the fresh fruit and vegetable trade. The values of imports and exports of fresh fruits and vegetables were similar until the early 1980s (USDA FAS 2015a). Since then, the United States has been importing significantly more valued fresh fruits and vegetables than exporting them. In 2014, United States imported more than US\$16 billion of fresh fruits and vegetables, while it exported approximately US\$7 billion of fresh agricultural commodities.

Trends in agricultural commodity volumes have been similar to trends observed for agricultural commodity values (Figure 1-11), but the magnitude of the difference between exports and imports has been different. In terms of values, total agricultural commodity exports have been greater than imports since 1967. However, the percent differences between import and export values have ranged from a low of 8.4% to a high of 152.2% in 1981 (USDA FAS 2015a). . On the other hand, fresh fruit and vegetable import values have been greater than export values in the United States for most years since 1967. The years that fresh fruit and vegetable exports exceeded than imports were 1974, 1976, 1977, and 1980. By contrast, in terms of fresh fruit and vegetable volume, imports have always been greater than exports. The differences between fresh fruit and vegetable import and export volumes were small until late 1970s, but have since become quite large. For example, the difference between imports and exports in 1975 was 0.79 million metric tons. In 2014, United States imported 17.28 million metric tons of fresh fruits and vegetables, yet exported 5.32 million metric tons.

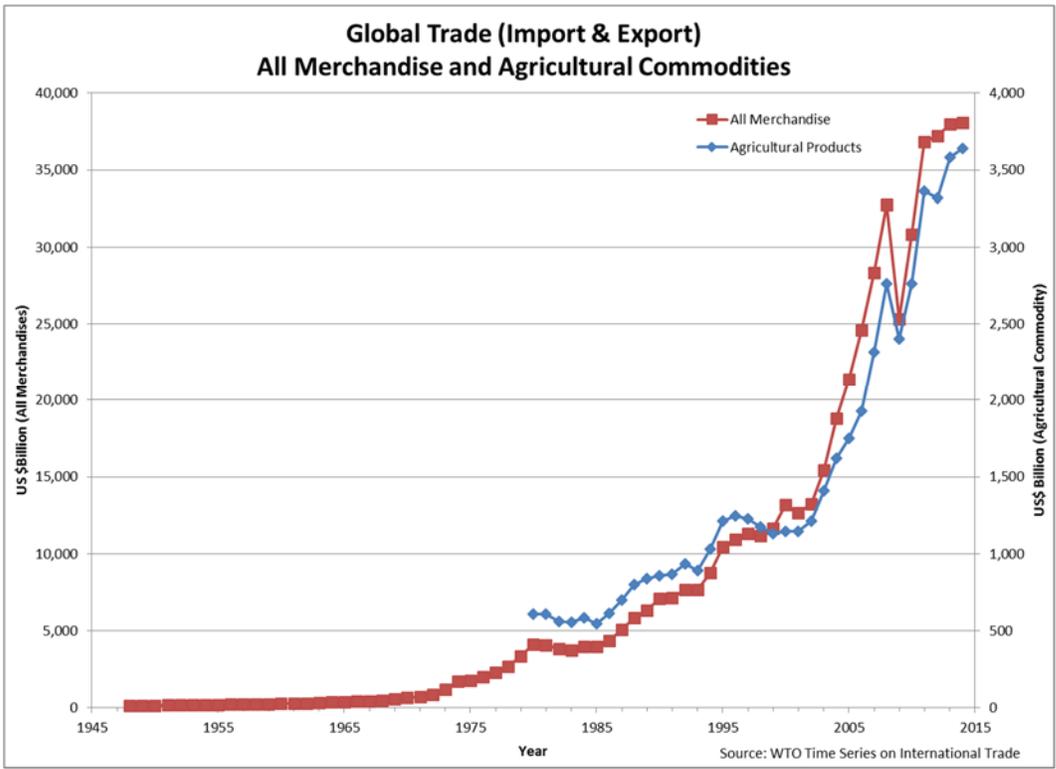


Figure 1-7. Global trade values (WTO 2015)

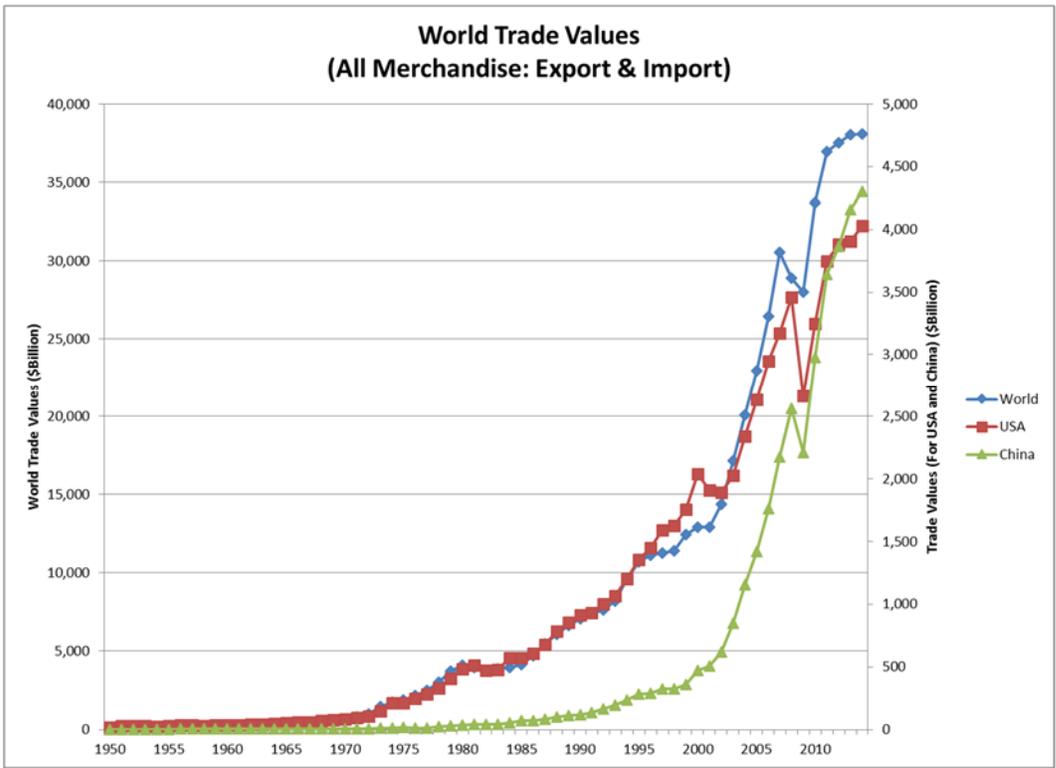


Figure 1-8. The values of world trade, US, and China

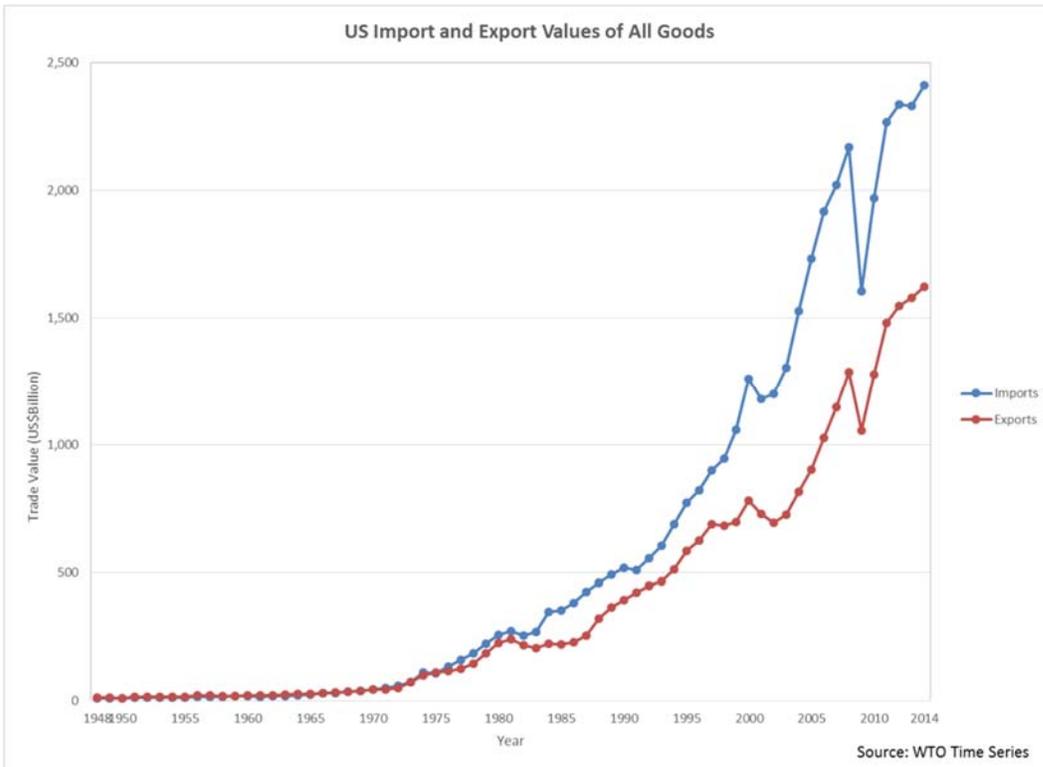


Figure 1-9. US import and export values of all goods since 1948

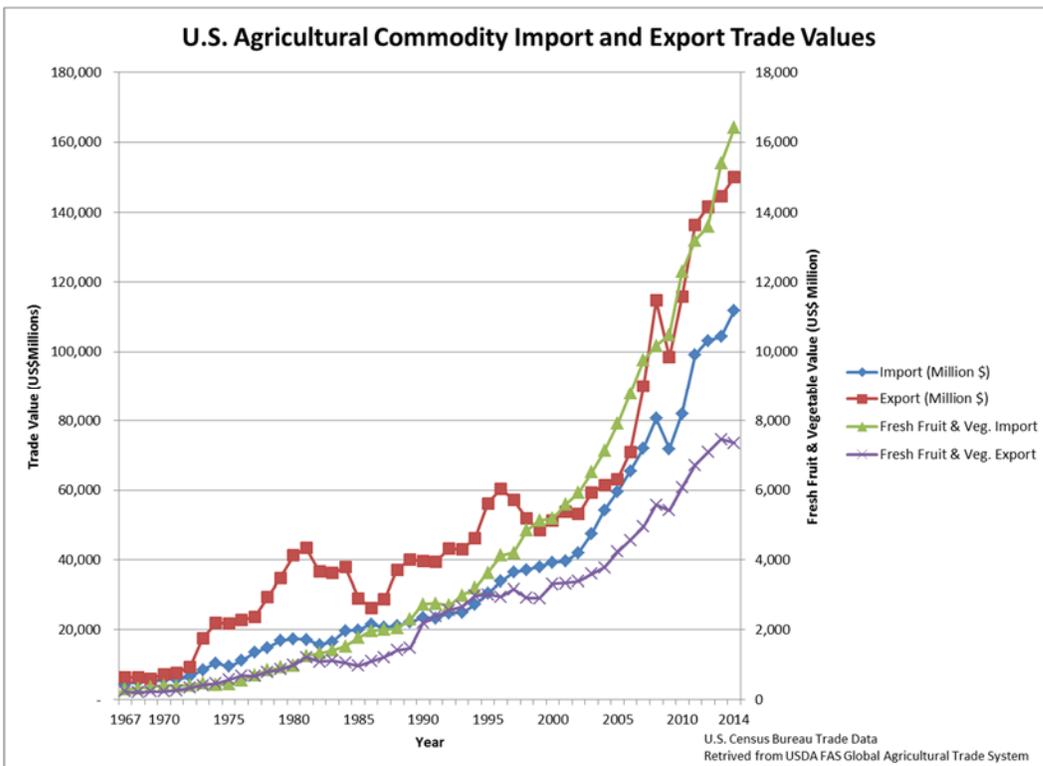


Figure 1-10. U.S. agricultural trade values (Source: USDA FAS GATS, 2015)

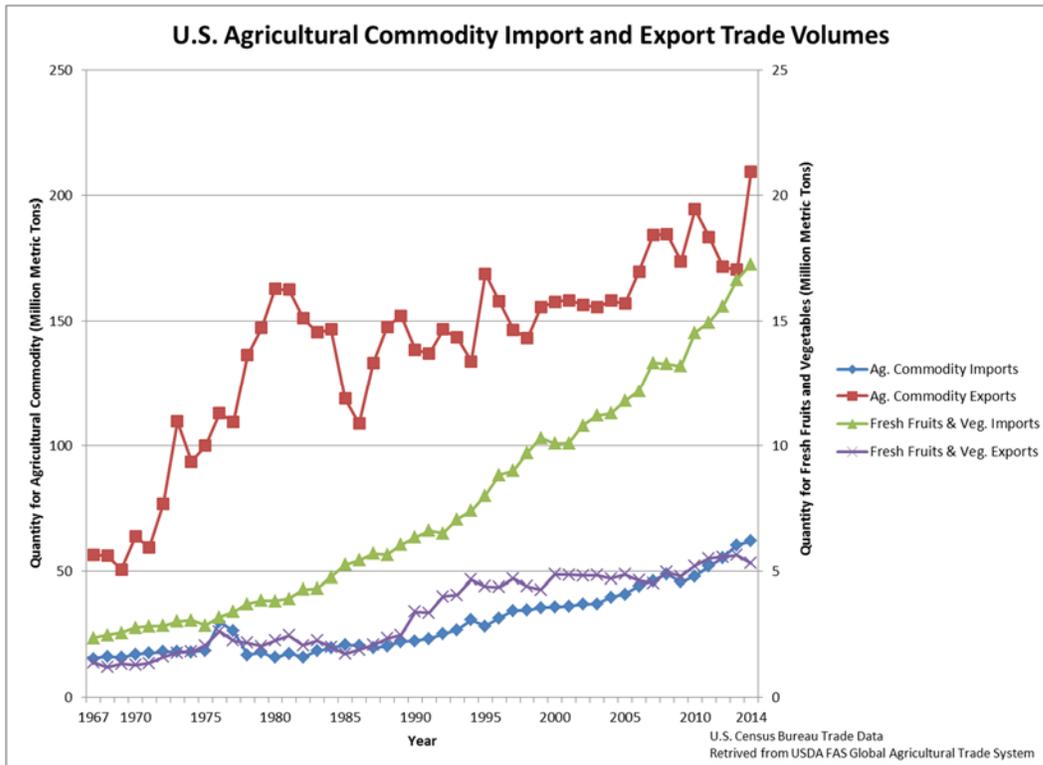


Figure 1-11. U.S. agricultural trade volumes (Source: USDA FAS GATS, 2015)

The primary US trading partners have also shifted in terms of their relative importance (as indicated by total volume and value). Historically, the biggest partners to the United States were European countries (Sailer 1978). After World War II, trade patterns started to change, and Asian countries such as Japan became important partners. Particularly since 1970, trade with Asian countries has grown sharply, while trade with European countries has grown more slowly (Table 1-2). The most dramatic change has been in the U.S. trade relationship with China, which ranked as the 24<sup>th</sup> trading partner in 1980 but was the second largest trade partner by 2010. Trade with India has also increased significantly in recent decades. In contrast, South American countries such as Colombia, Argentina, and Peru, were strong trade partners in 1970, but they were no longer among the top 25 U.S. trading partners in 2010.

The United States currently trades with over 230 countries; however, over 80 percent of trade happens with partners in the top 25. Although trade with the European Union (EU) accounts for over 20% of total U.S. trade, the top three U.S. trading partners in 2011 were Canada, China, and Mexico. Trade with South and East Asian countries such as South Korea, Japan, Taiwan, and Singapore remains strong, but Russia, Nigeria, and Saudi Arabia have also become important trade partners in recent years.

These trade increases are linked to expanding world economies but also facilitated infrastructure development (e.g., roads, rails, airports, seaports) and political reform (increased free trade) (Rana 2004, Hulme 2009). International trade has facilitated alien species movements between the countries and infrastructure expansion has assisted rapid and further invasions to larger areas (Hulme 2009).

**Table 1-2 U.S. major trading partners since 1970 (U.S. Census Bureau, Foreign Trade)**

Country	Rank in 1970	Rank in 1980	Rank in 1990	Rank in 2000	Rank in 2010	2010 Trade (\$billion)
Canada	1	1	1	1	1	526.75
China	N/A	24	10	4	2	456.82
Mexico	5	3	3	2	3	393.38
Japan	2	2	2	3	4	181.03
Germany	3 <sup>1</sup>	4	4	5	5	130.59
United Kingdom	4	5	5	6	6	98.19
South Korea	17	13	7	7	7	87.72
France	7	7	8	9	8	65.32
Taiwan	15	9	6	8	9	61.89
Brazil	12	16	17	14	10	59.38

---

<sup>1</sup> Includes both Federal Republic of Germany (West Germany) and Democratic Republic of Germany (East Germany).

Country	Rank in 1970	Rank in 1980	Rank in 1990	Rank in 2000	Rank in 2010	2010 Trade (\$billion)
Netherlands	8	14	11	13	11	53.99
India	19	35	27	25	12	48.78
Singapore	38	23	12	10	13	46.44
Venezuela	10	10	18	17	14	43.36
Saudi Arabia	56	6	15	22	15	42.97
Italy	6	11	9	12	16	42.72
Ireland		47	30	16	17	41.12
Belgium	9 <sup>2</sup>	15	14	18	18	41.01
Malaysia	36	26	20	11	19	39.98
Switzerland	14	21	19	23	20	39.82
Nigeria	N/A	N/A	24	29	21	34.58
Israel	22	33	25	21	22	32.28
Russia	N/A	N/A	31 <sup>3</sup>	31	23	31.70
Thailand	44	38	23	19	24	31.67
Hong Kong	13	17	13	15	25	30.87
Philippines	21	27	26	20	33	
Australia	11	20	16	24	26	
Spain	16	N/A	21	28	29	
Sweden	18	N/A	22	26	34	
South Africa	20	N/A	33	34	37	
Colombia	23	N/A	29	30	27	

### 1.3 USDA APHIS

#### 1.3.1 Animal and Plant Health Inspection Service

Responsibility for preventing and responding to invasive species threats in the United States is distributed among different federal and state agencies that work jointly to coordinate

---

<sup>2</sup> Belgium and Luxembourg

<sup>3</sup> Recorded as USSR.

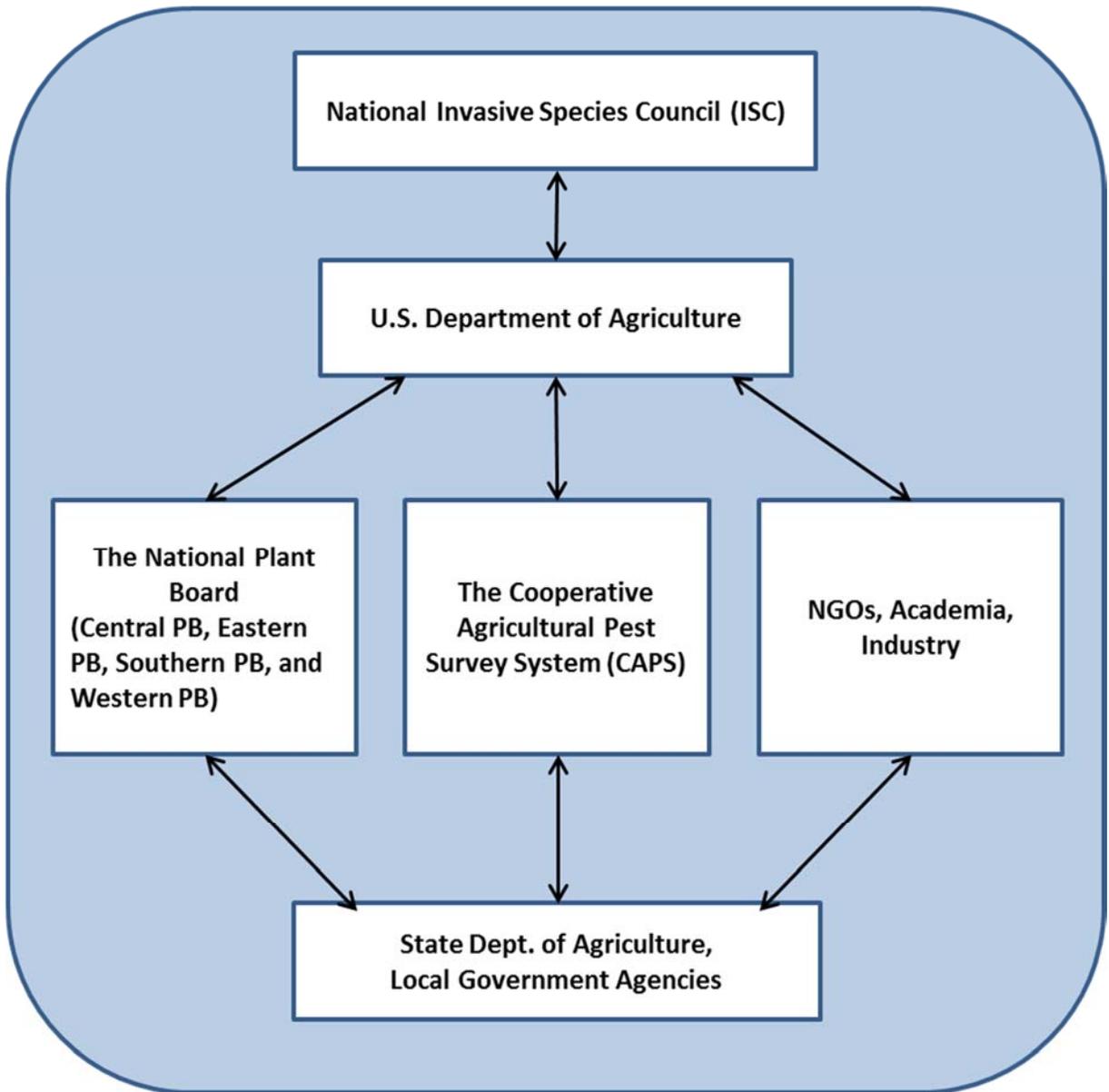
public and private efforts. The National Invasive Species Council (NISC) is an interagency council that helps to coordinate and ensure complementary, cost-efficient and effective federal activities regarding invasive species ([www.invasivespeciesinfo.gov](http://www.invasivespeciesinfo.gov)). The council members include the Secretaries of Agriculture, Commerce, Interior, State, Defense, Homeland Security, Treasury, Transportation, and Health and Human Services and administrators of the Environmental Protection Agency, U.S. Agency for International Development, the U.S. Trade Representative, and the National Aeronautic and Space Administration. The NISC produces a National Management Plan (NMP) every two years to address all federal actions on invasive species. The NISC tends to focus on pests that require intervention at the multi-agency level. Based on the framework established by the NISC, there are efforts within individual agencies (e.g., USDA APHIS) that have specific statutory authority and responsibility for invasive species.

The chart below shows key agencies and organizations that play a vital role in invasive species management and the relationship among them within the U.S. Department of Agriculture (Figure 1-12). In this section, the role and responsibility of USDA APHIS is our focus because it is the lead Federal regulatory agency.

The USDA APHIS mission is to protect the health and value of American agriculture and natural resources. APHIS responsibilities are to manage invasive species and to prevent spread and new introductions of alien invasive species. Plant Protection and Quarantine (PPQ) is a program within APHIS that safeguards agriculture and natural resources from risks associated with the entry, establishment, or spread of plant pests and noxious weeds to ensure an abundant, high-quality, and varied food supply. PPQ focuses on detection,

diagnostics (identification), agricultural trade, emergency response, and methods development (research).

In addition to safeguarding from invasive species at the interagency level, the Plant Protection Act (PPA) was established in 2000. The PPA, which is managed by USDA APHIS, regulates the movement of plants, plant products, biological control organisms, noxious weeds, articles, and means of conveyances to prevent and manage invasive species. At the State level, federal agencies cooperate with State Departments of Agriculture to manage and respond to alien species emergencies. The Cooperative Agricultural Pest Survey (CAPS) system is a partnership between all 50 states and the USDA to detect and monitor alien pests of economic significance. Recently, the CAPS program has shifted its strategy from being “pest-specific” to a “commodity” targeting basis. These regulations and programs have shown different degrees of effectiveness in preventing new introductions into the United States.



**Figure 1-12 Invasive Species Management Organizations within USDA**

USDA APHIS is funded through a system that includes Congressional Appropriations, user fees, and Commodity Credit Corporation (CCC) funds. The Plant Protection Act authorizes APHIS to utilize funds appropriated by Congress to support its core activities. The appropriations by Congress are voted on each year, and specific amounts are communicated as part of the annual Federal Congressional budgetary cycles. User fees

are collected from international air passengers. International air passengers arriving in the United States are subject to inspections to ensure they possess legal entry and immigration documents, and do not bring in contraband items, such as illegal drugs, counterfeit goods, harmful pests, and prohibited agricultural products. With the creation of the Department of Homeland Security (DHS) in 2003, customs, immigration, and agricultural inspections were integrated into one program led by Customs and Border Protection (CBP). The collected fees are for inspections and co-administered by CBP, Immigration and Customs Enforcement (ICE), and APHIS. The CCC was created to stabilize, support, and protect farm income and prices, as well as to help maintain balanced and adequate supplies of agricultural commodities and aids in their orderly distribution. To fulfill its mission, APHIS must be able to react immediately when pest or disease outbreaks pose a serious economic threat to American agriculture. The Secretary of Agriculture has the authority to transfer funds from the CCC to APHIS for emergency control programs.

### **1.3.2 Monitoring Invasive Species: Archival Systems at APHIS**

#### **US Interception Data (PestID):**

Fresh agricultural commodity imports to the United States are regulated by U.S. Department of Agriculture, Animal and Plant Health Inspection Service (USDA APHIS) and are inspected at the ports of entry by Department of Homeland Security, Customs and Border Protection (DHS CBP) except if it involves propagative materials (“plants for planting”) in which case USDA APHIS is responsible for inspections. Inspection of agricultural commodities was previously part of APHIS responsibilities, but after establishment of DHS in 2003, most inspection responsibilities were transferred to CBP under DHS but USDA

APHIS retained the scientific oversight, control of pest identification stations and inspection of some cargo. Agricultural commodities imported into the US are inspected mainly by CBP agricultural specialists.

USDA APHIS keeps records of pest interceptions at U.S. port of entries. The database is commonly known as PestID. This database contains interception date, inspection location, pathway (e.g., airport, maritime), origin, destination, pest type, pest scientific name (often times genus level, but some of them are identified to species), inspected host species, inspected host part (e.g., seed, leaf, fruit, flower), where inspected (e.g., mail, permit cargo, baggage), and number of pests found. This database only keeps records on positive findings. The inspectors are inspecting more materials than recorded. Since APHIS does not record negative findings in the database, there is no indication of volume of the materials being inspected at the ports of entry.

### **North American Non-Indigenous Arthropod Database**

The North American Non-Indigenous Arthropod Database (NANIAD) was developed by K.C. Kim. This project was funded by USDA APHIS in the early 1990s. The NANIAD includes approximately 2400 non-indigenous insect species. Although the latest record date in this database was in 1991, the majority of the species in this database were established in the United States prior to 1983.

### **New Pest Advisory Group**

The New Pest Advisory Group (NPAG) is maintained by the Plant Epidemiology and Risk Analysis Laboratory (PERAL), which is part of USDA, APHIS, Plant Protection and Quarantine (PPQ), Center for Plant Health Science and Technology (CPHST). The NPAG

assess alien plant pests that are new or imminent threats to U.S. agriculture or the environment and recommends appropriate actions to the PPQ Deputy Administrator. Its database records historic data of non-indigenous species. Currently, the database contains over 1000 species of arthropods, plant pathogens, weed, and mollusks (as of December, 2015).

## **1.4 Trends of Introduced Alien Species in the United States**

### **1.4.1 Methods**

The number of interception records by year was obtained from the PestID database. The number of interception records was summarized by calendar years since 1980. In some cases, there was more than one specimen of the same pest species found on a single commodity. We counted such cases as one interception record.

Information about alien species introduced to the United States was gathered from NANIAD and NPAG databases. The number of alien species newly introduced to the United States since 1980 was summarized by combining and analyzing the NANIAD and NPAG databases. In addition, both databases contain information about the first detection date in the United States, the first detection location in the United States, and the origin or native region of the species. The NPAG database includes newly introduced species as well as imminent and intercepted species. Imminent and intercepted species were not considered as successfully introduced species, so those species were not included in this study. Often, information about origin and the location and date of first detection in the United States were missing from both databases. In such cases, I estimated the origin of a given pest based on

the native range of the species or the country from which we believe the pest originated prior to introduction into the United States. As for the missing information about introduction date and location, those information were estimated based on the first public record or the oldest literature citation that recorded its presence in the United States.

#### **1.4.2 Interception Data**

The number of interception records was plotted with import values for agricultural commodities (Figure 1-13). The numbers of interception records in 2003 and 2004 was much lower compared to 2002; this reduction may have been caused due to transition efforts as hundreds of staff migrated from APHIS to DHS in 2003. Interception records increased from 2006 to 2010 and decreased from 2010 to 2013. The interception number increased slightly from 2013 to 2014; however, the interception number in 2014 was less than the number in 2009.

Interception records at U.S. ports of entry may fluctuate with many factors and may not be a perfect indicator of the number of alien species introductions into the United States. Inspection effort changes due to administrative changes, budget, inspector availability, seasons, and locations. However, interception data seem to indicate the number of pest outbreaks at global scale. For instance, Asian gypsy moth was intercepted significantly more from Japan in 2012 and 2013 when Japan was having outbreaks of AGM in those two years (PestID 2015).

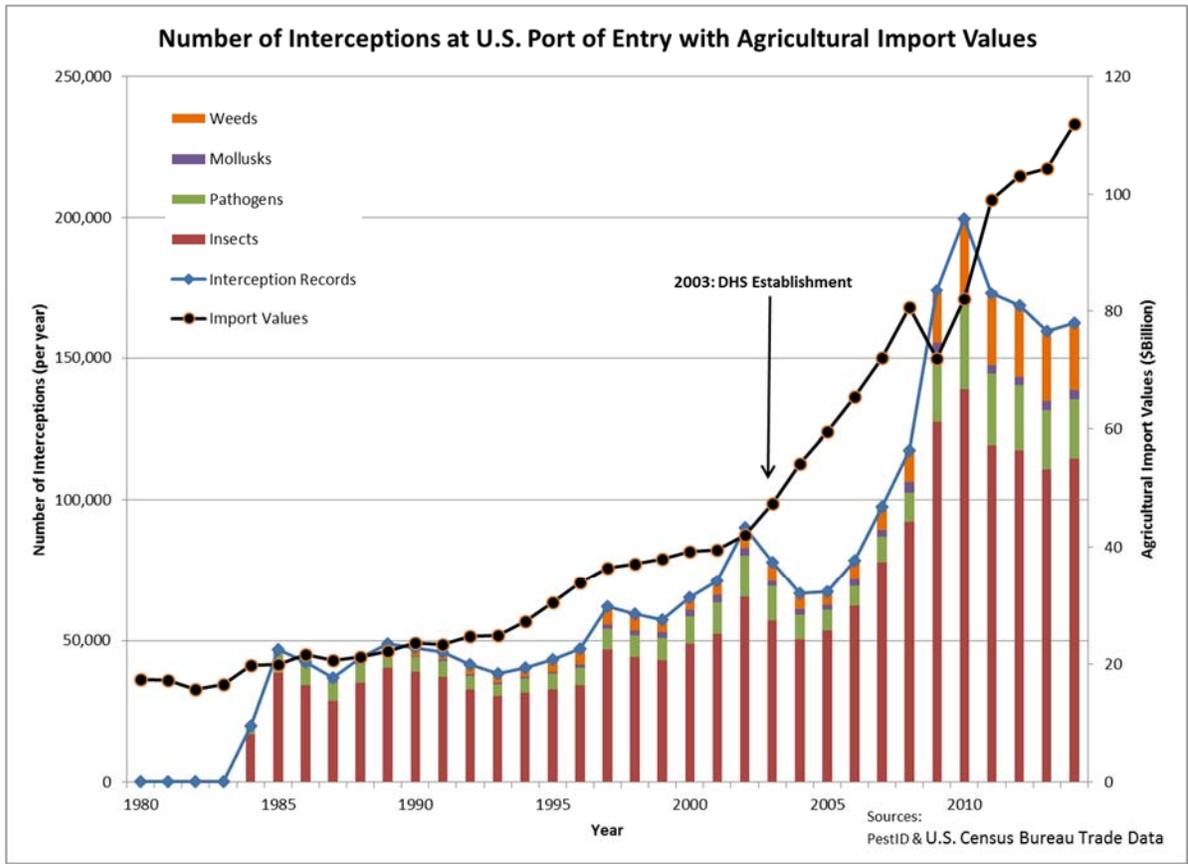


Figure 1-13 Interception records with U.S. agricultural commodity import volume

### 1.4.3 Overall Statistics for Introduced Species

The United States can be characterized by complex movements of humans, domestic and international trade. There have been several studies that estimate the number of alien species established in the United States, but most studies focus on specific fauna (e.g., arthropods, pathogens, weeds), on specific time periods, or on specific ecological categories (e.g., forest species) (Sailer 1978, Mack 2003, Aukema et al. 2010). Approximately 1,300 alien arthropods had been established in the United States by 1978 (Sailer 1978). Recent studies indicated that 450 arthropods and 16 pathogens had colonized U.S. forests (Aukema et al. 2010), and over 2,500 flora species have become naturalized. However, none of the

studies has reported comprehensive non-indigenous species for all weeds and fauna of phytosanitary significance in the United States.

NANIAD recorded only arthropod species, while NPAG includes arthropods, pathogens, mollusks, and weeds. The NANIAD database contains over 2,400 exotic arthropod species introduced in as early as 1,565 although it has not been updated since 1991. The NPAG database currently has over 1,000 non-indigenous species records.

There were missing data on the first detection date and location in both NANIAD and NPAG. Missing data were completed in my study using literature search and expert opinion. Yet, there were still some data that were impossible to obtain at this time. Those species were still included in the total number as newly introduced alien species since 1980 but were excluded from the data analysis by introduced state or by origin depending on the types of missing data.

The majority of the reported species are arthropods, followed by plant pathogens, weeds, and mollusks. These species belong to 17 phyla, 25 classes, 74 orders, 195 families, and 513 genera (

Table 1-3). The numbers in parentheses in the Species row indicate the number of organisms that were only identified to genus or family level.

**Table 1-3 Number of accidentally introduced alien species first reported in the United States from 1980 to 2014 (Sources<sup>4</sup>: NANIAD and NPAG)**

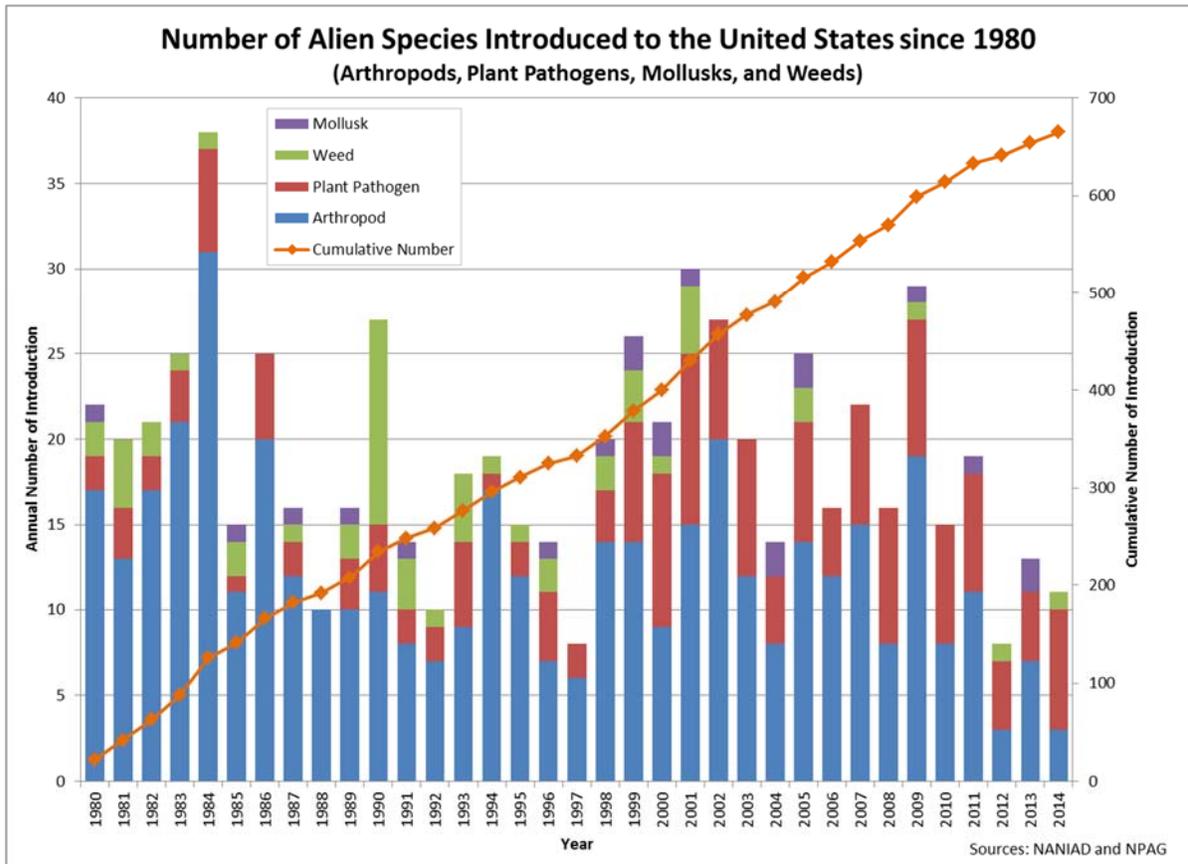
	Arthropod	Mollusk	Plant Pathogen <sup>5</sup>	Weed	Total
Phylum	1	1	10	5	17
Class	2	1	17	5	25
Order	17	4	30	21	74
Family	108	10	51	26	195
Genus	353	16	98	46	513
Species	431 (26)	20 (1)	160 (9)	54 (4)	665

The number of alien species introduced to the United States was plotted by year (Figure 1-14). The NANIAD database included a large number of arthropods that were intentionally introduced to the United States as beneficial species or biological control agents. Those species were excluded from this analysis because this study focused on analyzing alien species that were accidentally introduced to the United States by weather events, trades, and human movements. Also, the NPAG database does not include any alien species that are introduced to the United States intentionally. Although the number of alien species recorded varied from year to year, there was a constant alien species introduction ( $R^2=0.9967$ ) each year. On average, 19 new pest species were recorded as introduced (entered, established and spread) in the United States each year.

---

<sup>4</sup> There were some records which were not identified to the species level.

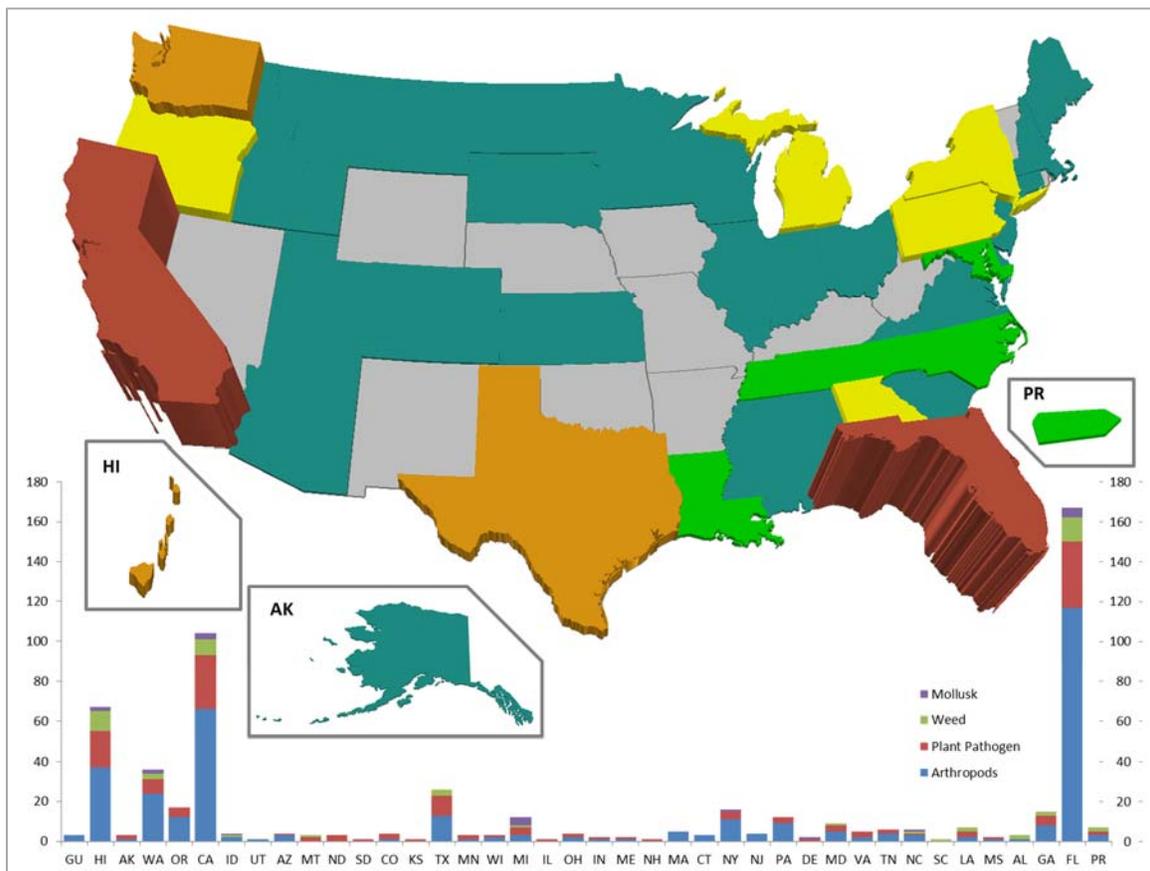
<sup>5</sup> Nematodes are considered as plant pathogens.



**Figure 1-14. Number of alien species (arthropods, plant pathogens, mollusks, and weeds) introduced to the United States since 1980 (Source: NANIAD and NPAG)**

To gain initial insights into the spatial distribution of newly introduced species, the alien species recorded since 1980 were mapped based on their first detection locations (Figure 1-15). Florida had the largest number of alien species. The state is a well-known hub for trade and human movement, especially from Caribbean, Central American, and South American countries. California and Hawaii ranked second and third, respectively. Both states have high volumes of international trade and human movement, mainly from Asia and the Pacific Islands. Other significant states were Washington and Texas, which border Canada and Mexico, respectively. There is also a large volume of commodity and human movement via highways and railways within North America. In the United States,

trucking is the most frequently used mode of freight transportations, accounting for 60 percent of the freight weight (USDOT 2006). Trucking and rails account for 77% of all freight transportation in the United States. Since the 1980s, U.S. rail intermodal volume has increased more than 4-fold. Trucks transported 11.5 billion tons and 13.2 billion in 2002 and 2012, respectively (USDOT 2007, 2013). Both trucks and rails moved approximately 41.6 million tons of freight a day in 2012. The U.S. Department of Transportation forecasted that freight transportation by trucks and rails would increase to 21.5 billion tons (59 million tons per day) by 2040. These massive movements of freight within the United States could create entry locations for invasive alien species via international trade and human movements.



**Figure 1-15. Number of alien species (Arthropods, Mollusks, Plant Pathogens, and Weeds) introduced to the United States since 1980 (Sources: NANIAD and NPAG)**

#### 1.4.4 Nonindigenous arthropod species

##### Analysis of trends associated with the number of arthropods introduced to the United States:

Sailer's study (1978) revealed that there was a slow rate of introduction until 1860 but it increased significantly thereafter (Figure 1-16). The exponential increase from 1860 to 1920 is thought to have been caused by increases in commerce (Sailer 1978). The introduction rate after 1920 slowed compared to the rate between 1860 and 1920. The rate was relatively constant from 1920 to 1978 (Sailer 1978). Sailer (1978) believed this 1920 change was the result of the Plant Quarantine Act of 1912, which was not fully implemented until 1920. According to Sailer (1978), there were on average 12 species introduced into the U.S. since 1920, and 3 out of those 12 species were intentionally introduced as beneficial species.

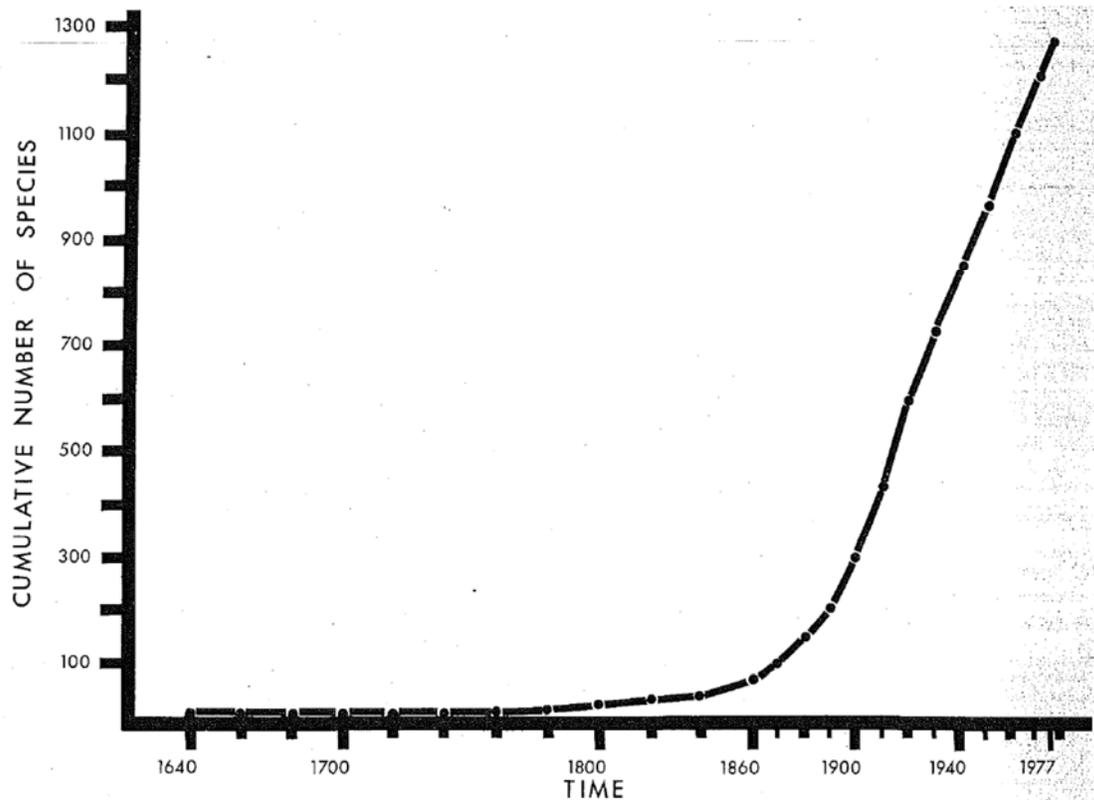


Figure 1-16. Accumulation of immigrant arthropod species (Source: Sailer, 1978)

The number of arthropods accidentally introduced into the United States was examined based on NPAG and NANIAD database data. Figure 1-17 shows the results. The number of new alien arthropod introductions varied each year; however, the rate of increase in the cumulative number of alien species from 1980 to 2014 was steady. The numbers of introductions in 2012 and 2014 were small compared to other years. This may be due to the NPAG database not being updated yet. When the NPAG team determines some newly introduced species are not causing significant damage, those species are analyzed and added into the database at a much later time. Still, the rate of unintentionally introduced arthropods was linear at 12 species per year since 1980 ( $R^2=0.9929$ ).

Despite the significant increase in agricultural commodity trade volume since 1980, the number of arthropods introduced to the United States shows only a slight increase, especially when compared to the findings of Sailer (1978) where an average nine accidentally introduced arthropods were reported per year. One possible explanation is due to improvement in technology and effective treatments prior to entering the United States, which may be effectively preventing some introductions. APHIS and CBP collected \$255 million in FY2001 and \$594 million in FY2014 as user-fee collections. APHIS currently plans to increase the user fees to \$748 million for FY2016 to prevent economic and environmental damages caused by alien species. Another theory is that the number of inspectors (agricultural quarantine inspection (AQI) staff) has not been increased according to import trade volumes. There were approximately 1,560 AQI personnel in 1985 and 6,800 AQI staffs in 2012. The number of AQI personnel has increased approximately 4.3-fold since 1985, while the volume of agricultural commodity imports has increased 2.6-fold (3.2-

fold for fresh fruit and vegetable import volumes). Also, as globalization increases, all countries around the world get tightly connected. Because so many commodities have been moved around the world each day, the species pool of successful invaders begins to be depleted.

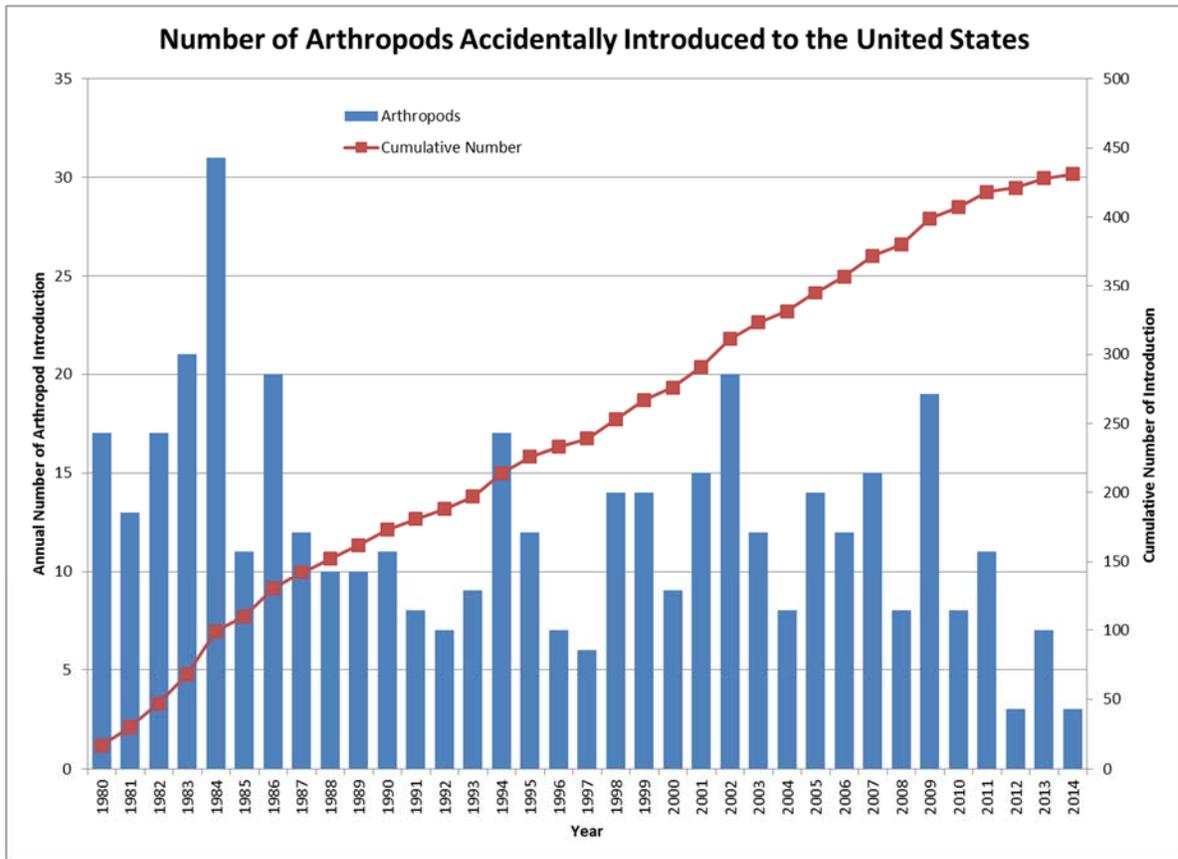
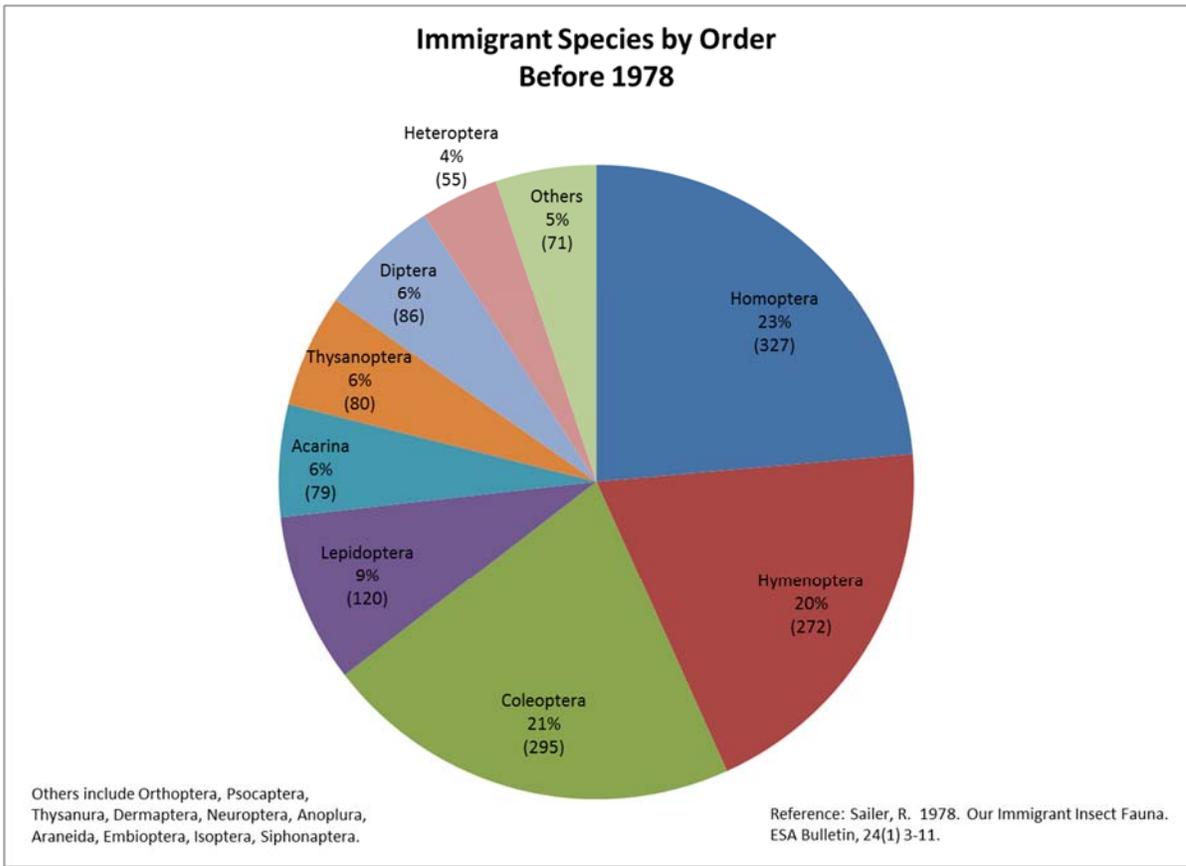


Figure 1-17. Number of arthropods accidentally introduced to the United States since 1980 (Sources: NANIAD and NPAG)

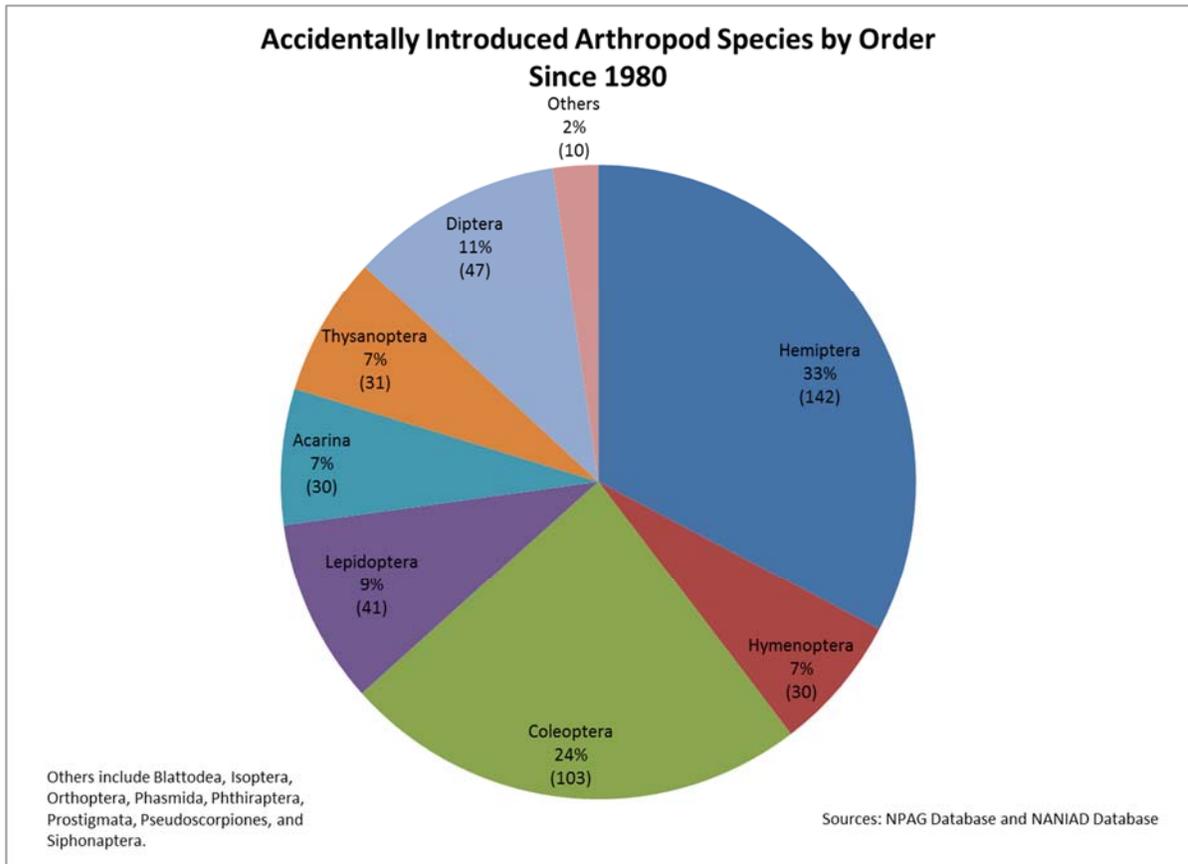
**Analysis of taxonomic trends of established nonindigenous arthropod species:**

Based on Sailer’s (1978) study, Homoptera, Hymenoptera, and Coleoptera were the major three orders for non-indigenous species introduced to the United States (Figure 1-18). Sailer also examined the historical proportional changes of immigrant species by order. In 1820, the majority of immigrant species were Coleoptera (90%). Between 1860 and 1930,

Homoptera was the dominant order of introduced arthropod species. Homoptera species (now categorized as Hemiptera) have piercing and sucking mouthpart and feed on terrestrial plants by withdrawing saps (Gullan and Cranston 2014, Meyer 2015). The major families in Homoptera are Cicadidae (cicadas), Cicadellidae (leafhoppers), Cercopidae (spittlebugs), Psyllidae (psyllids), Aleyrodidae (whiteflies), Aphididae (aphids), and Coccidae (soft scale insects). These insects are easily transported via nursery stock. Sailer (1978) suggested that increased demand for nursery stock from Europe caused Homoptera species introductions into the United States. By 1970, the majority of alien species established in the United States belonged to Hymenoptera. Hymenoptera species include ants (Formicidae), parasitic wasps (Ichneumonidae, Braconidae, Encyrtidae, Eulophidae, and Trichogrammatidae), and bees (Aphidae). Sailer (1978) suggested that this was a result of introducing beneficial insects as biological controls.



**Figure 1-18. Order of the immigrant species introduced before 1978 (Source: (Sailer 1978))**



**Figure 1-19. Order of the immigrant species accidentally introduced to the United States since 1980 (Sources: NANIAD and NPAG)**

The majority of the alien species discovered in the United States after 1980 belong to the Hemiptera order followed by Coleoptera species (Figure 1-19). Homoptera and Heteroptera were merged into Hemiptera (Gullan and Cranston 2014); therefore, major alien species established in the United States still remained as Homoptera (or Hemiptera) species. Hymenoptera species were significantly reduced compared to immigrant species before 1978. This was due to not including purposely introduced species for the data after 1980, while Sailer (1978) included all beneficial species in his analysis.

Even though NANIAD database included both deliberately and accidentally introduced species, I did not include deliberately introduced species since I did not have a

source for how many beneficial species were introduced to the United States after 1991. Thus, trends for deliberately introduced species (mostly biological control agents) are not a focus of this study and remain part of future characterizations.

As summarized in Figure 1-19, the database shows that Hemiptera and Diptera species after 1980 increased to 33% from 27% and to 11% from 6%, respectively. Hemiptera and Diptera species are easily transported with nursery stock, propagative materials, fresh fruits, and fresh vegetables. In terms of volume, import of fresh fruits and vegetables to the United States were more than 4-fold in 2014 compared to 1980. Coleoptera species also represented a slightly larger proportion of arthropod introduction after 1980. Increase in global trade has stimulated the usage of solid wood packing materials (Lidsky 2003), which may explain the increase in Coleoptera species introductions into the United States compared to the period before 1978.

**Analysis of origins of established non-indigenous arthropod species:**

Sailer (1978) categorized the origin of non-indigenous immigrant arthropods into eight zoogeographic regions (Figure 1-20). Those regions are 1) Nearctic, 2) Neotropical, 3) Eastern Palearctic, 4) Oriental, 5) Western Palearctic, 6) Ethiopian, 7) Australian, and 8) Oceania.



**Figure 1-20. Geographic Regions used for categorization of immigrant species origin (Sailer 1978)**

Independent of whether Sailer's is the best geographic categorization, I have retained that template to enable a comparison between the study by Sailer (1978) and the current study. Before 1978, the majority of alien species came to the United States from Europe because the major trading partners were European countries and trade with other countries was limited. Sailer (1978) reported that more than half of arthropods introduced into the United States before 1978 were Western Palearctic origin (Figure 1-21). Asia was represented in three regions, Eastern Palearctic, Oriental, and Australian-Oriental. Before 1978, approximately 21% of alien arthropods introduced into the United States originated in Asia. Despite its proximity, arthropods arriving from Neotropical region accounted for only 13% of total.

The origins of arthropod species were categorized using the best available information into these 8 regions described, but the process was not perfect. Some pests were reported in the scientific literature or database as native to two or sometimes three regions (e.g., Eastern Palearctic and Oriental regions, Eastern and Western Palearctic regions). For

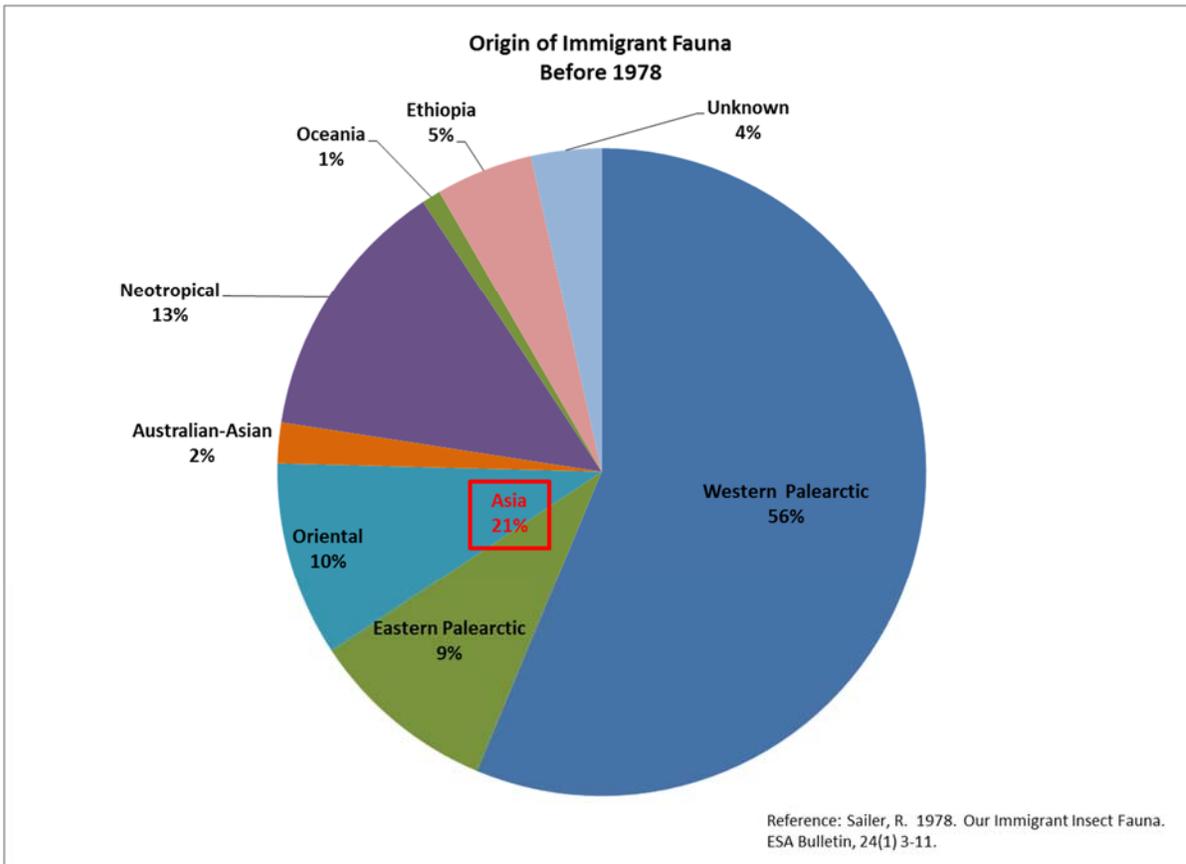
those species whose native ranges were reported from more than one zoogeographic region, their native ranges were described by combining regions.

After 1980, the pattern of immigrant arthropod origins changed. Europe (Western Palearctic) was still the major region from which new arthropods were introduced; however, European native arthropods were only 33% of all accidentally introduced insects.

The proportion of European origin insects was reduced to almost half compared to introductions prior to 1978. Insects originating from Neotropical regions have become more common; Neotropical native insects accounted for 20% after 1980. Another region that grew in prominence as a source was Asia.

Eastern Palearctic origin insects actually decreased 2% after 1980. In contrast, the Oriental region represented only 10% of total non-indigenous species before 1978 but increased to 16% after 1980. In total, 32% of newly introduced arthropod species since 1980 originated in Asia (including the Eastern Palearctic, Oriental Western-Eastern Palearctic, Eastern Palearctic- Oriental, Oriental-Australian, and Oriental-Oceania regions), while Asia only represented 21% of the total before 1978.

The origins of non-indigenous arthropod species were further analyzed by categorizing into five regions: Europe, Asia, Oceania (Australasia and Oceania), America (North and South), and Africa (Figure 1-23). If introduced arthropod species were native to multiple regions, then each region was counted as one, meaning one species could be counted multiple times. Figure 1-23 clearly indicates that there was a significant decrease for the species that were native to Europe, while there were significant increases in all other regions.



**Figure 1-21. Origin of immigrant arthropods before 1978 (Sailer, 1987)**

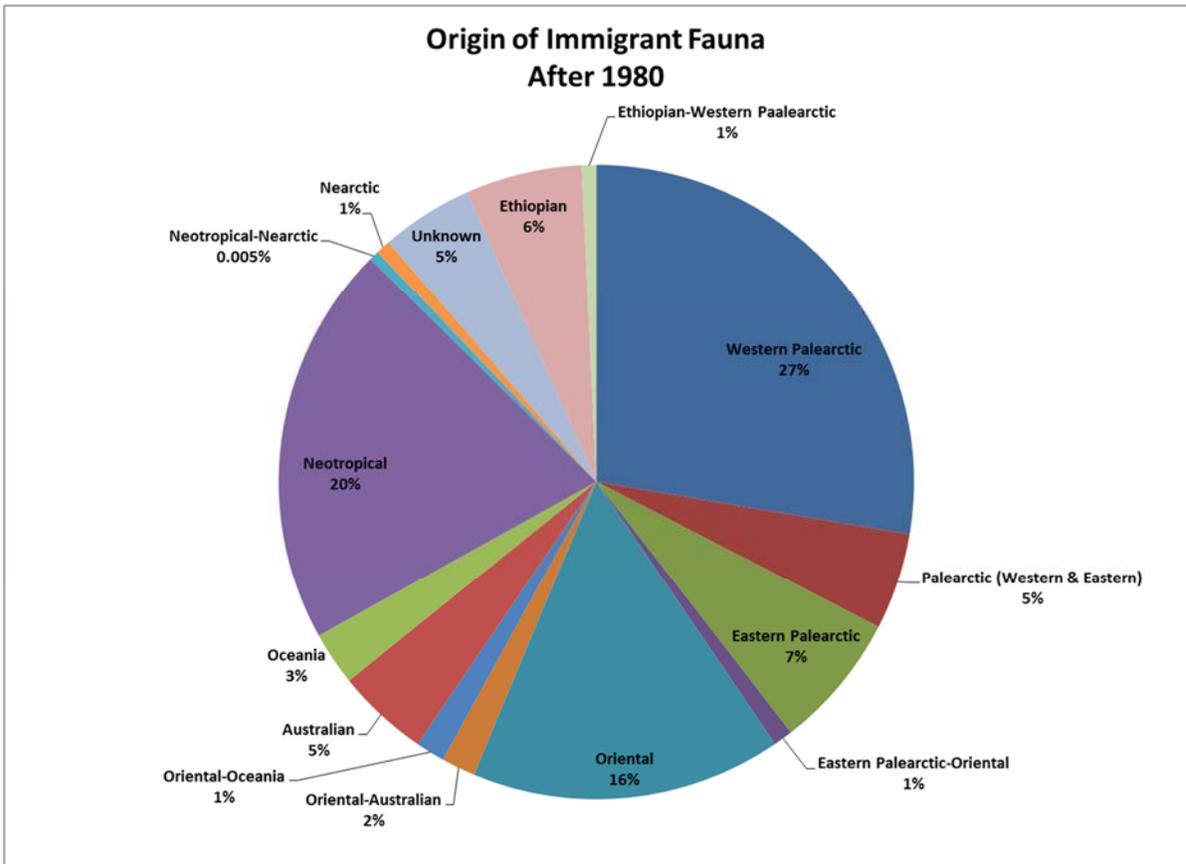


Figure 1-22. Origin of immigrant arthropods accidentally introduced into the United States after 1980  
(Sources: NANIAD and NPAG)

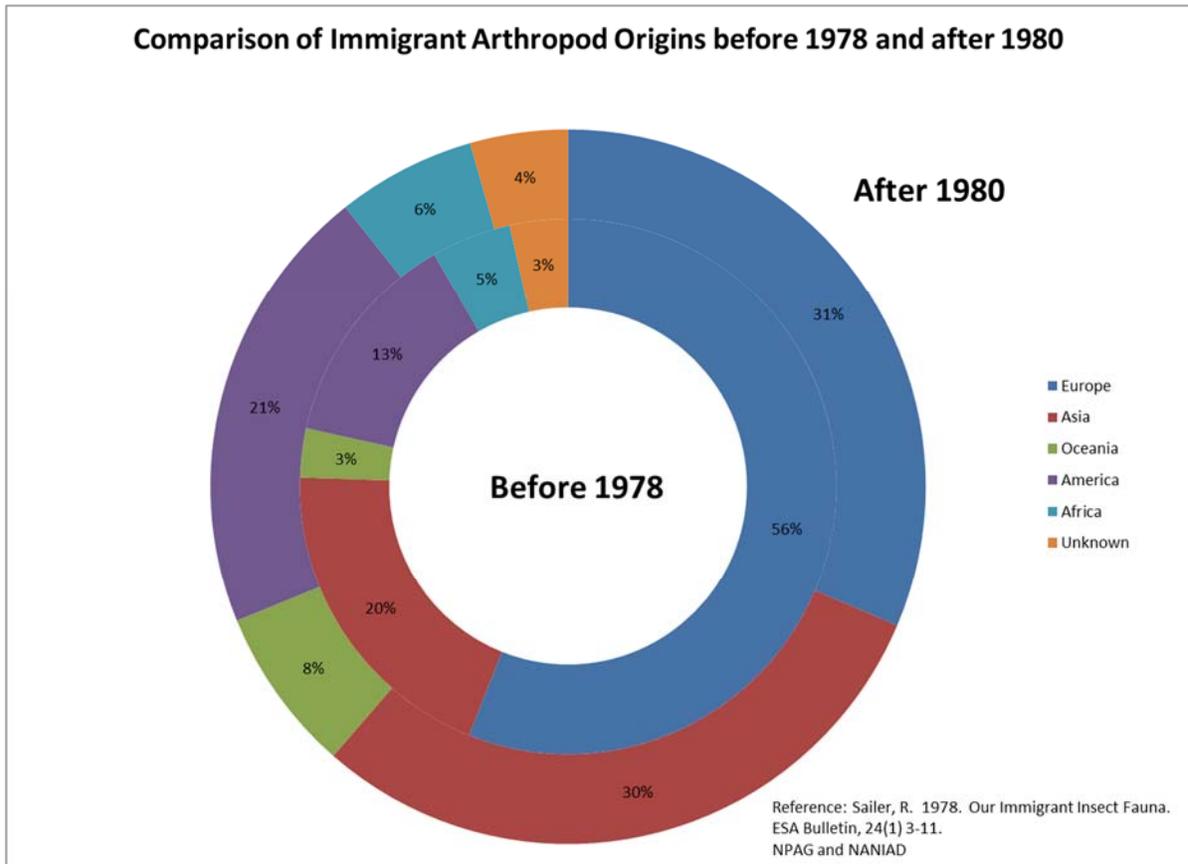


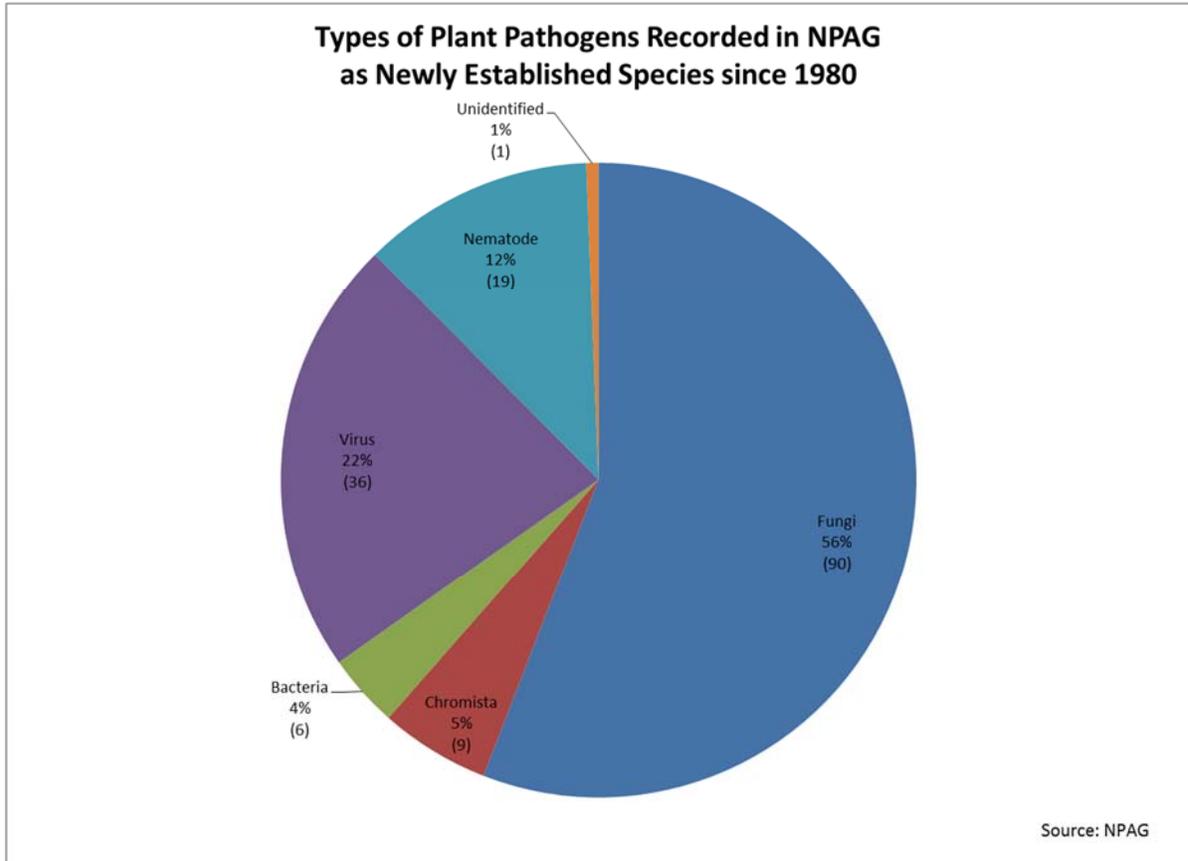
Figure 1-23. Comparison of immigrant arthropod origins between prior to 1978 and post 1980

#### 1.4.5 Analysis of trends associated with nonindigenous plant pathogens

Plant pathogen records in NPAG database were analyzed. Both plant pathogenic species and nematodes were included in this category. The earliest record in the NPAG database was 1922, but plant pathogen records before 1980 were limited (27 records). Any records before 1980 were excluded from this study.

There were approximately 160 alien plant pathogens and nematodes introduced to the United States from 1980 to 2014 (

Table 1-3). Although 160 unique species have been identified as newly introduced plant pathogen species to the United States, 9 species have been only identified to genus or higher level. On average, there were 4.5 species established each year. The most common plant pathogens were fungi followed by viruses. The types of plant pathogens are summarized in Figure 1-24.



**Figure 1-24. Types of non-indigenous plant pathogen species introduced to the United States since 1980**

The origins of plant pathogens recorded in the NPAG database were not straightforward to determine. This was because the classification of plant pathogenic fungi has significantly changed as molecular technologies have developed and because commonly used phylogenetic analyses have taken advantage of advanced and rapid molecular-based

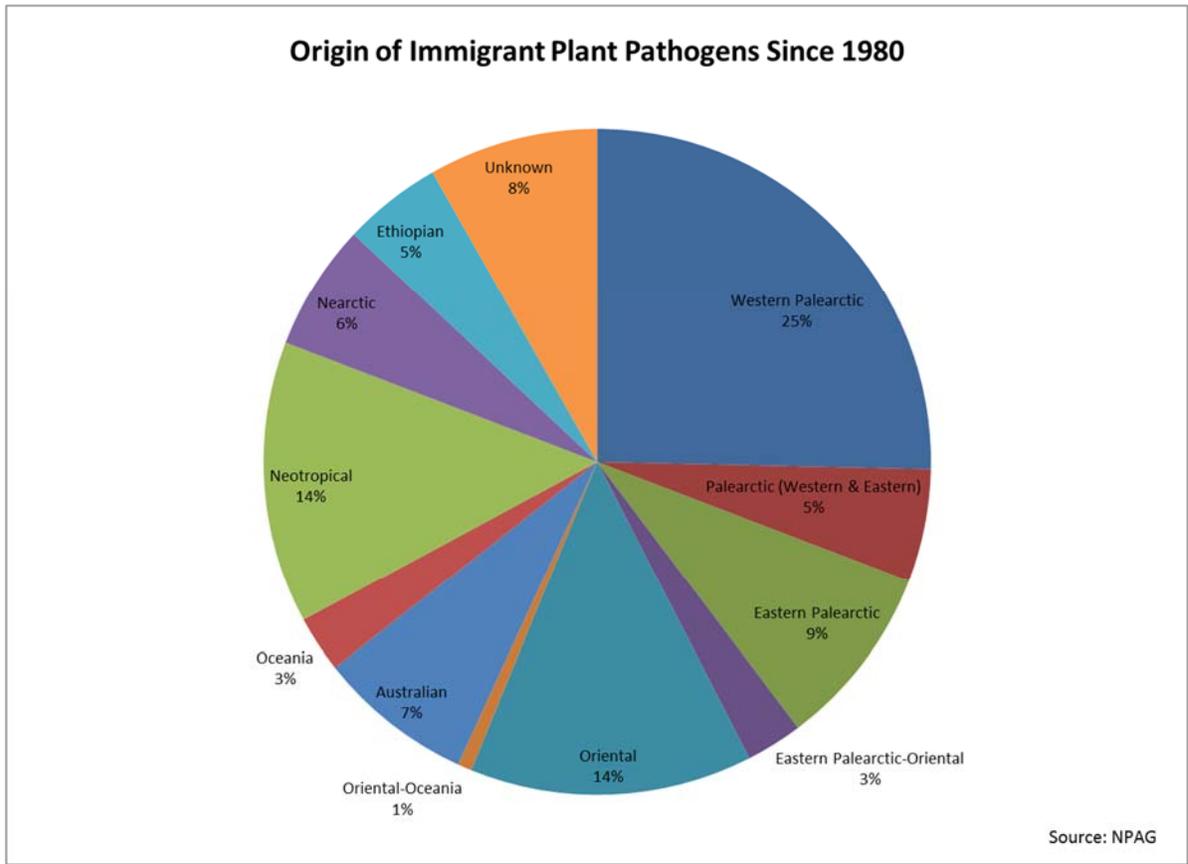
classifications. As a result, the same species has often been called different names, or one name has represented several species.

For consistency, the origins of plant pathogens were determined according to their current names. Origin was determined based on the oldest literature found for the specific species if its native range was not easily identified. Most plant pathogenic species came from the Western Palearctic (25%) (Figure 1-25).

Asian countries, represented by two regions in this analysis (Eastern Palearctic and Oriental), accounted for 27% of alien plant pathogens. In addition, 5% of plant pathogens newly introduced to the United States came from Western-Eastern Palearctic.

Other significant regions of origin of alien plant pathogens were Australian (Australia and New Zealand) and the Neotropical region (Caribbean, Central America, and South America). Some species, recorded as Nearctic in origin, were never recorded in the United States before but were also never recorded or identified anywhere outside the Nearctic region. This suggests that they may be newly evolved species; therefore, they were considered as non-indigenous species.

There were several records which were not identified to the species level. Due to lack of scientific evidence, the origins of those species were categorized into an 'Unknown' class.



**Figure 1-25. Origins of alien plant pathogens introduced after 1980 (Source: NPAG)**

#### **1.4.6 Analysis of trends associated with non-indigenous weeds**

Although Mack (2003) reported over 2,500 species of non-indigenous weed species as naturalized in the United States, a very small number of weeds are reported in the NPAG database. In addition to Mack (2003), the PLANTS Database administered by USDA Natural Resources Conservation Service (<http://plants.usda.gov/>) lists over 4,043 weed species as introduced to the contiguous United States, 1,427 species introduced to Hawaii, and 907 species introduced to Puerto Rico (USDA NRCS 2015). This NPAG database listed 54 weeds as introduced species since 1980, belonging to 26 families (Figure 1-26). The largest proportion of alien weeds belong to the Poaceae family.

Unlike arthropods and plant pathogens, very few weed species (4%) introduced to the United States originated in the Western Palearctic (Figure 1-27), although some other weed species were identified as “Western-Eastern Palearctic” and “Western Palearctic-Oriental”. Therefore, Western Palearctic species could represent up to 11%. Asian species represented 38% of non-indigenous weed species (combining all Asian regions). Neotropical and Ethiopian species represented 24% and 15% of introduced weed species into the United States, respectively.

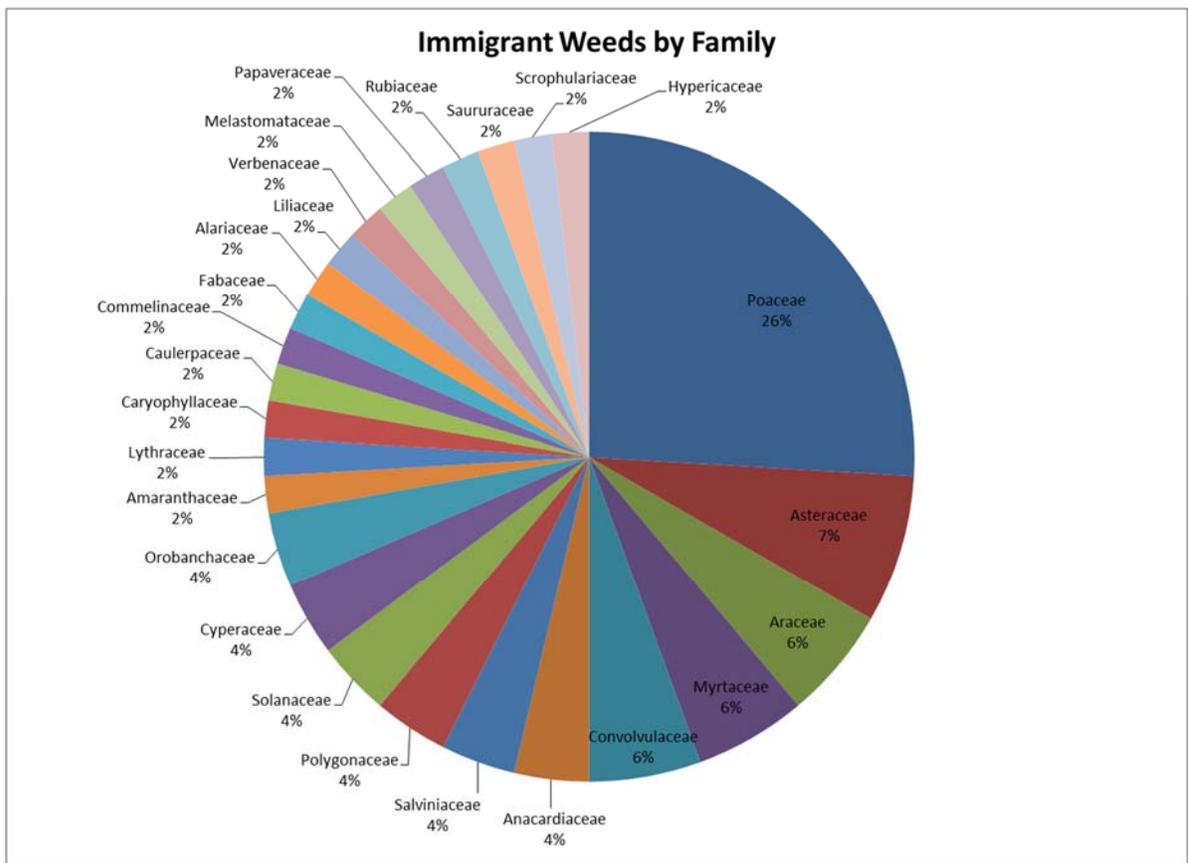


Figure 1-26 Family of alien weed species in the United States

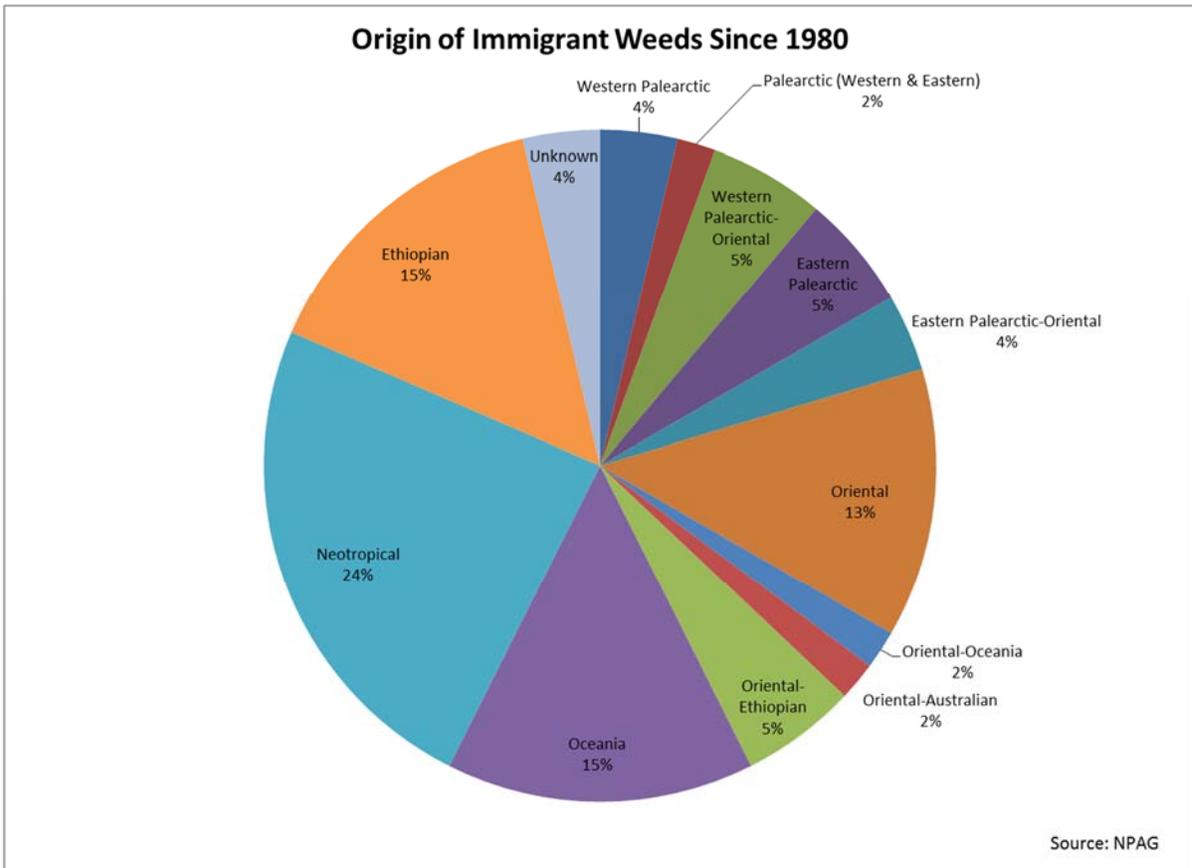


Figure 1-27. Origins of alien plant pathogens introduced after 1980 (Source: NPAG)

#### 1.4.7 Analysis of trends associated with non-indigenous mollusks

Mollusk data were limited in the NPAG database; only 20 species were recorded as established mollusk species in the database. Among those, one species was only identified to the genus level.

The greatest number of mollusks detected in a year was 1999 and 2000. The majority of mollusks introduced to the United States originated from Europe (Figure 1-28). Among 20 mollusk introductions, 5 establishments were in Florida. The states of Michigan, California, Washington, and Hawaii reportedly had a total of 4 introductions, 3 introductions, 2 introductions, and 2 introductions, respectively.

Due to the limited number of records, the NPAG database may not be the best source to analyze trends and determine the patterns of alien mollusk species introductions into the United States. Other sources will be investigated in the future.

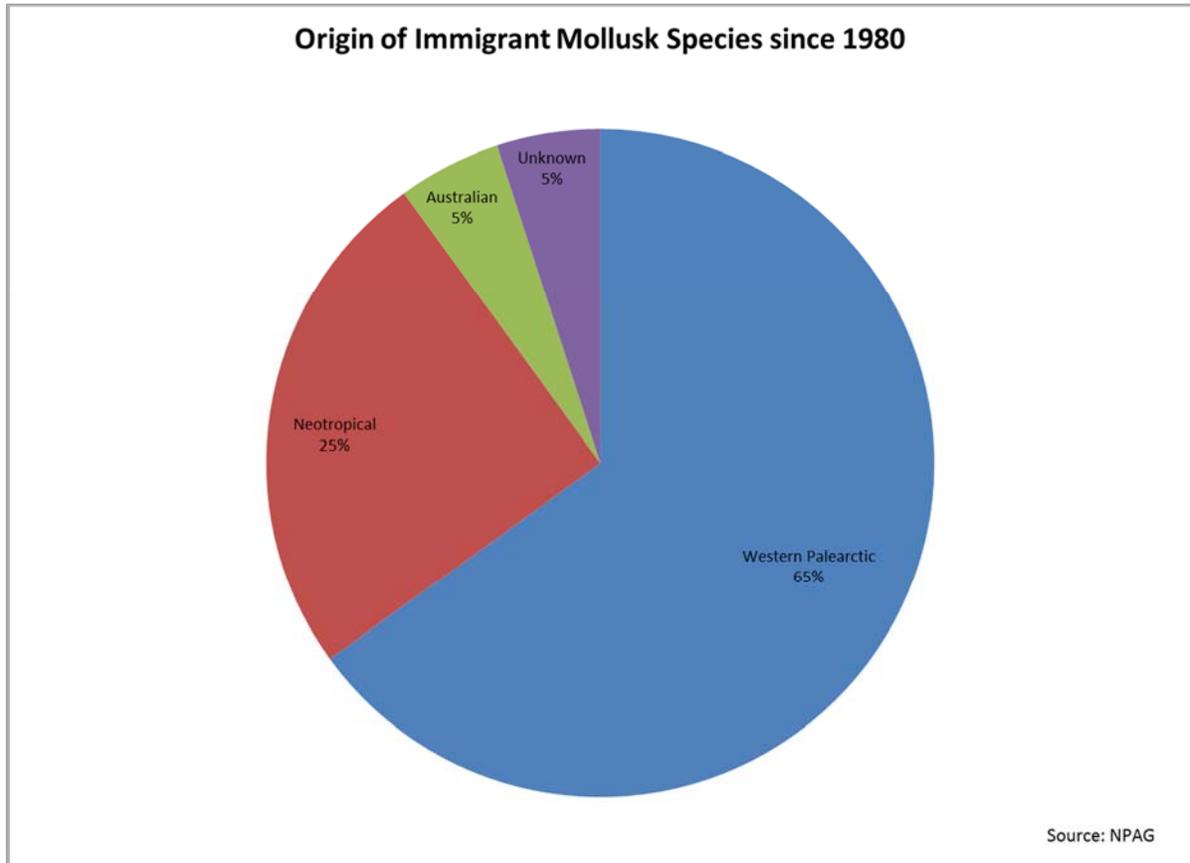


Figure 1-28 Origin of mollusks recorded as established in the United States after 1980 in NPAG

### 1.5 Examination of correlations between invasive species data and selected independent variables

Linear relationships between introduction of alien species, environmental factors, and human activities were analyzed. Pearson's correlation coefficients were determined by using a JMP multivariate analysis tool. The variables examined against introduction of alien species were the following:

1. Number of interception records at U.S. ports of entry (PestID 2015),
2. World population (UN 2012),
3. Number of foreign travelers into United States (U.S. inbound travelers) (The World Bank 2015),
4. Number of U.S. travelers to overseas (U.S. outbound travelers) (The World Bank 2015),
5. Number of international travelers (sum of U.S. inbound and outbound travelers),
6. U.S. trade import values (WTO 2015),
7. U.S. trade export values (WTO 2015),
8. U.S. agricultural commodity import values (USDA FAS 2015a),
9. U.S. agricultural commodity export values (USDA FAS 2015a),
10. U.S. agricultural commodity import volumes (USDA FAS 2015a),
11. U.S. agricultural commodity export volumes (USDA FAS 2015a),
12. U.S. fresh fruit and vegetable import values (USDA FAS 2015a),
13. U.S. fresh fruit and vegetable export values (USDA FAS 2015a),
14. U.S. fresh fruit and vegetable import volumes (USDA FAS 2015a),
15. U.S. fresh fruit and vegetable export volumes (USDA FAS 2015a),
16. U.S. annual average temperature (NOAA 2015a),
17. U.S. annual average maximum temperature (NOAA 2015a),
18. U.S. annual average minimum temperature (NOAA 2015a), and
19. U.S. annual average precipitation (NOAA 2015a).

In addition to U.S. annual average climatic data, annual average climatic data for California and Florida obtained from NOAA National Centers for Environmental

Information (NOAA 2015a) were examined since those two states had the highest rates of alien species introductions.

Correlations between annual establishment and trade data are displayed in Table 1-4. The number of annual establishments, including all arthropods, plant pathogens, weeds, and mollusks, had a very weak negative linear relationship with all the trade variables examined in this study.

Correlations between annual establishment and environmental factors were examined next (

Table 1-5). Again, there was no indication of a strong relationship between establishment rate and environmental factors.

Finally, correlations among annual establishment, number of interception records, and human factors were examined (Table 1-6). Human factors included world population, U.S. inbound travelers, U.S. outbound travelers, and overseas travelers coming into the United States. There was no indication that interception records and human factors had any relationship with non-indigenous species establishment in the United States. I included U.S. agricultural commodity import values in the matrix in order to examine if agricultural imports had any relationships with interception records and human factors.

U.S. agricultural commodity import values had strong linear relationships with three factors: 1) number of interception records at U.S. ports of entry, 2) world population, and, 3) number of overseas passengers. Interestingly, U.S. agricultural commodity import values had weaker linear relationships with number of U.S. inbound travelers and number of U.S. outbound travelers. However, when those two factors were combined, there was a strong relationship. The number of overseas passengers also had a strong positive linear relationship with the number of interceptions at U.S. ports of entry, suggesting that the number of pests coming into the United States increases as the number of overseas passengers increases. There were over 2.5 million interception records by December 31, 2014. Among those records, almost 1.2 million interceptions were on passenger baggage.

**Table 1-4. Correlation matrix between annual establishment number and U.S. trade data**

Characteristics	1	2	3	4	5	6	7	8	9	10	11
(1) Establishment	1.00										
(2) US Import (\$)	-0.22	1.00									
(3) US Export (\$)	-0.28	0.98	1.00								

(4) US Ag Import (\$)	-0.26	0.97	0.99	1.00							
(5) US Ag Export (\$)	-0.31	0.91	0.96	0.97	1.00						
(6) US Ag Import Volume	-0.24	0.98	0.99	0.97	0.92	1.00					
(7) US Ag Export Volume	-0.17	0.77	0.79	0.78	0.82	0.74	1.00				
(8) US Fresh Ag Import (\$)	-0.25	0.97	0.99	0.99	0.97	0.98	0.78	1.00			
(9) US Fresh Ag Export (\$)	-0.29	0.97	0.99	0.98	0.96	0.98	0.77	0.98	1.00		
(10) US Fresh Ag Import Volume	-0.23	0.98	0.98	0.96	0.91	0.99	0.75	0.97	0.98	1.00	
(11) US Fresh Ag Export Volume	-0.23	0.84	0.86	0.78	0.73	0.88	0.63	0.80	0.87	0.88	1.00

---

**Table 1-5. Correlation matrix between annual establishment number and environmental factors**

Characteristics	1	2	3	4	5	6	7	8	9	10	11	12	13
(1) Establishment	1.00												
(2) US Annual Avg Temp	0.03	1.00											
(3) US Annual Avg Tmax	0.00	0.97	1.00										
(4) US Annual Avg Tmin	0.06	0.94	0.82	1.00									
(5) US Annual Avg Precip	0.07	-0.38	-0.56	-0.11	1.00								
(6) FL Annual Avg Temp	0.03	0.42	0.39	0.43	0.07	1.00							
(7) FL Annual Avg Tmax	0.11	0.47	0.49	0.41	-0.12	0.89	1.00						
(8) FL Annual Avg Tmin	-0.03	0.32	0.25	0.39	0.19	0.93	0.68	1.00					
(9) FL Annual Avg Precip	-0.19	-0.21	-0.31	-0.06	0.42	0.05	-0.35	0.36	1.00				
(10) CA Annual Avg Temp	-0.28	0.19	0.23	0.11	-0.25	-0.16	-0.16	-0.13	0.02	1.00			
(11) CA Annual Avg Tmax	-0.27	0.24	0.34	0.10	-0.42	0.02	0.05	0.00	-0.10	0.94	1.00		
(12) CA Annual Avg Tmin	-0.26	0.07	0.03	0.13	0.07	-0.37	-0.42	-0.26	0.21	0.86	0.63	1.00	
(13) CA Annual Avg Precip	0.02	-0.03	-0.17	0.15	0.51	-0.34	-0.41	-0.24	0.30	-0.24	-0.52	0.22	1.00

**Table 1-6. Correlation matrix between annual establishment number and human factors**

Characteristics	1	2	3	4	5	6	7
(1) Establishment	1.00						
(2) US Ag Import (\$)	-0.26	1.00					
(3) Interception	-0.21	0.92	1.00				
(4) World Population	-0.22	0.92	0.85	1.00			
(5) US Inbound Travelers	-0.09	0.48	0.46	0.13	1.00		
(6) US Outbound Travelers	-0.15	0.79	0.68	0.95	-0.05	1.00	
(7) Oversea Passengers	-0.19	0.95	0.84	0.87	0.56	0.80	1.00

Increased human activity at a global scale is likely to increase introductions of new invasive species into the United States (Hulme 2009, Aukema et al. 2010, Koch et al. 2011,

Nghiem et al. 2013); however, previous parameterizations of the relationships between trade value and established species through time have shown that the relationships are non-linear (Levine and D'Antonio 2003, Koch et al. 2011). Levine and D'Antonio (2003) used three different approaches to estimate the number of alien species introduced by using trade data as drivers. Koch et al. (2011) re-parameterized two of the models by updating the trade values to estimate number of establishments on forest species. I wanted to re-visit their approaches and explore whether additional data yields results consistent with their reports and thus validate their observations. Alternatively, if the additional evidence showed large deviations, alternative or modified models might be appropriate. Although mollusks and plant pathogen introductions were parameterized in addition to arthropods (Levine and D'Antonio 2003), I only evaluated arthropods in this study.

The approach used by Levine and D'Antonio (2003) can be summarized as the equations below:

- 1) Log-log equation:  $N = cI^k$
- 2) Log-linear equation:  $N = \log I * k + c$
- 3) Michaelis-Menten equation:  $N = (N_{\max} * I) / (B + I)$

Where N = species richness (accumulated number of alien species)

c = constant

I = cumulative imports in dollar

k = rate of species increase due to trade imports

$N_{\max}$  = upper bound of Michaelis-Menten model

B = accumulated import value at half of  $N_{\max}$

Their findings from log-log, log-linear, and Michaelis-Menten models forecasted 1,535, 492, and 115, respectively, new insect species introduced to the United States for the period 2000 to 2020 (Levine and D'Antonio 2003). Although Levine and D'Antonio's models included deliberately introduced species, the NPAG database recorded 155 arthropods species from 2000 to 2014. The log-log and log-linear models overestimated, while Michaelis-Menten underestimated, the number of arthropod introductions recorded in NPAG.

I re-parameterized the log-linear and Michaelis-Menten models by obtaining trade data from U.S. Census Bureau and World Trade Organization (WTO) (U.S. Bureau of the Census 1975, WTO 2015) and arthropod accidental introduction data from Sailer (1978), NANIAD, and NPAG from 1920 to 2000 (Table 1-7). The re-parameterized models were validated by examining the new arthropod introduction records from 2001 to 2014.

**Table 1-7. Summary of the re-parameterized species-import models**

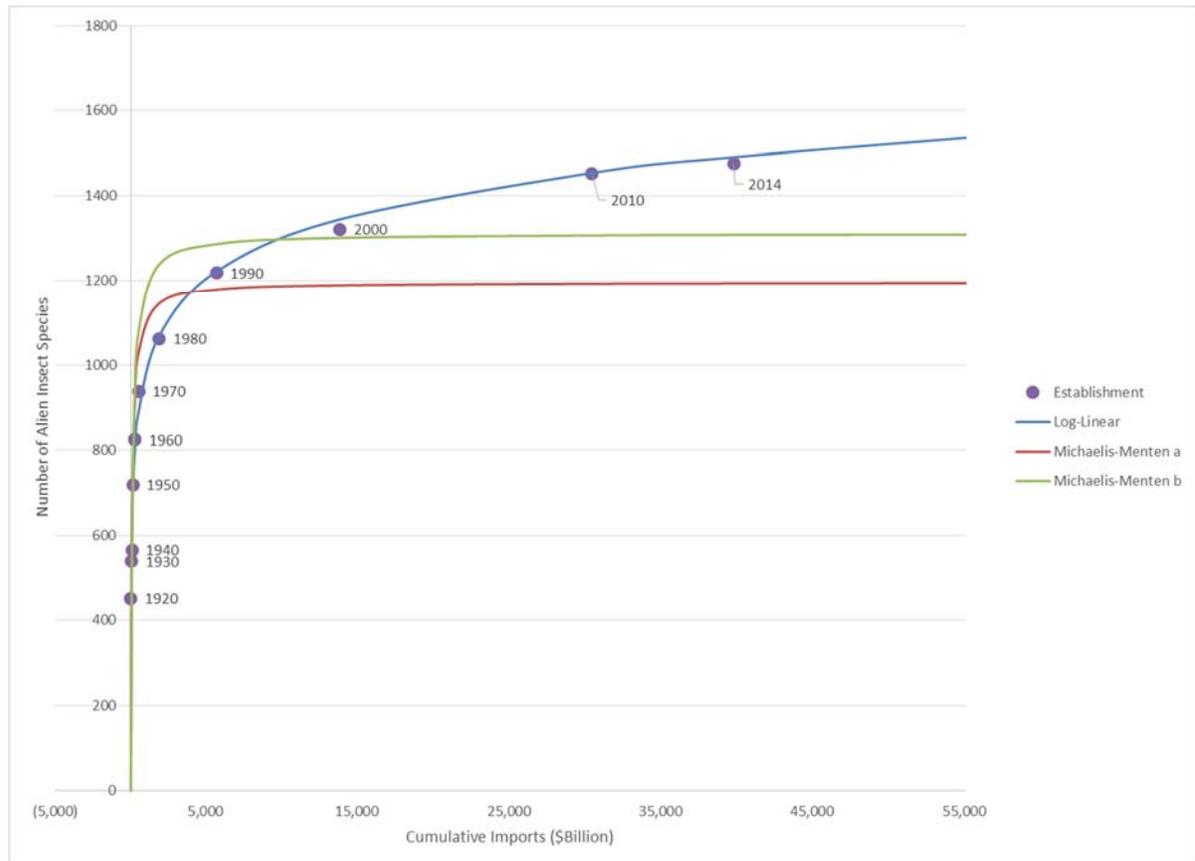
Model	Parameterized Equation (Import values in \$billion)	R <sup>2</sup>	Number of Established Arthropods Prediction from Model (Actual: NPAG)		
			By 2010	By 2014	By 2020
Log-linear	$N = \text{Log } I * 318.16 + 27.631$	0.992	1454 (1452)	1491 (1476)	1538
Michaelis-Menten <sup>a</sup>	$N = (1195.83 * I) / (82.94 + I)$	0.921	1193 (1452)	1193 (1476)	1194
Michaelis-Menten <sup>b</sup>	$N = (1311.36 * I) / (110.55 + I)$	0.902	1307 (1452)	1308 (1476)	1309

<sup>a</sup>: Parameterization by using data from 1920 to 2000

<sup>b</sup>: Parameterization by using data from 1920 to 2014

The log-linear model predicted arthropod introductions from 2010 to 2014 better than the Michaelis-Menten model (Figure 1-29). The problem with the Michaelis-Menten model was the maximum value of species number within the equation. The recent significant increase in trade imports not affecting the rate of arthropod introductions to the United States

forced the B value in the Michaelis-Menten model to correspond to a low import value. Even when the input data were expanded to 2014, the parameterization did not improve. Both parameterizations resulted in  $N_{max}$  to be less than observed.



**Figure 1-29. The relationships between cumulative import values and number of established alien arthropods**

The accuracy of predictions depends on both model fitness and accuracy of trade expectations. The global Gross Domestic Product (GDP) is expected to increase at a 3 to 4% annual rate from 2015 to 2020 (Laudicina and Peterson 2015). The number of alien arthropod species established into the United States by 2020 was estimated by assuming international trade imports from 2015 to 2020 will increase at the same rate as GDP. In 2014, U.S.

imports were valued at \$2.412 trillion; thus, the United States is likely to import over \$16 trillion in commodities from 2015 to 2020.

The log-linear model predicted that there would be 1,538 accidentally introduced arthropod species in the United States by 2020, while the Michaelis-Menten model predicted 1,194 and 1,309 species for the model parameterized by using data until 2000 and until 2014, respectively. Based on a log-linear model, there will be 62 more species by 2020, resulting in approximately 10 new species a year.

The current introduction rate for arthropod species based on the establishments from 1980 to 2014 is 12 species. Future trends for alien species and global trade are uncertain. However, the available evidence suggests that in the near future, United States may start seeing a reduction in the number of new invasions. This observation is consistent with findings in other studies (Levine and D'Antonio 2003, Koch et al. 2011).

## **1.6 Discussion, Summary, and Conclusions**

Non-indigenous species to the United States were analyzed to determine the diversity of species, the annual number of species introduced, the origin of the species, and the entry locations in the United States based on two databases, NANIAD and NPAG. There were several challenges to integrating the databases. First, the NANIAD database has not been updated since 1991. All analyses for recent year introductions were based on the NPAG database. Second, NANIAD included only arthropod species. There was no database available other than the NPAG database to gather information on plant pathogens, mollusks,

and weeds. Third, the NPAG database contained new detections and imminent data, while the NANIAD database was more comprehensive.

The NANIAD database included species introduced for biological control, but the NPAG database did not record this information. This is due to nature of NPAG, its purposes and objectives associated with invasive pests. The NPAG process is normally initiated by detection of new species (only accidental, not deliberate). Also, if the detection is detected too late (i.e., already widespread in the United States at the time of first report), then it is no longer considered a NPAG issue. These species are often not recorded in the NPAG database. Much of our efforts involved updating information in the databases with corrections and updates from current literature.

A key purpose for this study was to determine immigrant species patterns and how they correlated with changes in international trade patterns. Although it is of great value to know the country of origin of these species, it was challenging to make such determinations, as I have discussed.

As Sailer (1978) and others (Hulme 2009, Aukema et al. 2010, Koch et al. 2011, Nghiem et al. 2013) predicted, introduced alien species were observed to increase since 1980. The most noticeable change in recent alien species introductions (after 1980) and the ones before 1978 was the origin (geographic source) of the alien species. As for arthropods, there was a strong trend in the observations that native ranges of alien species in the United States were shifting to Asian countries and increasingly away from Europe as foreign trade patterns continue to emphasize trade with Asian countries.

International trade is expected to increase more in the near future. Now that the Trans Pacific Partnership (TPP) and Transatlantic Trade and Investment Partnership (TTIP) agreements are in place, the United States expects to trade more with Asian countries, Australia, New Zealand, and the European Union. In addition, the Panama Canal expansion project is ongoing (Trefis Team 2015). Currently, the maximum size of a vessel which can pass through Panama Canal is 106 feet wide by 1050 feet long, and 41.2 feet deep (Spivak 2011). This size vessel is known as a Panamax vessel. After completion of the Panama Canal expansion project, the maximum size of a vessel will be expanded to 160 feet wide by 1400 feet long (Spivak 2011).

The capacity of a Post-Panamax vessel is increased almost three-fold. Such vessels can accommodate 12,000 TEU (20-foot-equivalent unit) while a Panamax vessel carries 4,500 TEU (Spivak 2011). Even bigger vessels called Triple E or Ultra Large Container Vessels (14,501+TEU) already exist to transport commodities between Europe and Asia. Although there are no U.S. seaports that can handle Triple E vessels, seaports around the world are being upgraded to expand the ports as well as infrastructure from/to the ports to compete with other ports.

In addition, Asian exporters may discharge their containers at Mexican Pacific Coast ports (specially, Manzanillo or/and Lazaro Cardenas), and Mexican exporters transport Asian originated commodities into the United States via the Panama Canal and railroads (MARAD 2013). These expansion projects are expected to alter maritime paths at a global scale and significantly change the transportation patterns between Asian countries and Eastern United States (MARAD 2013, Jervis 2015).

There have been approximately 19 species introduced to the United States each year since 1980. The annual accidental introduction rate for arthropods since 1980 was 12 species, which was only 3 species more than Sailer (1978) reported for immigrant species before 1978. Since the volume of international trade imports before 1978 was less than one quarter of the volume today and held relatively steady for a long time, the volume of agricultural commodity imports did not likely influence the rate of alien species introduction into the United States for a period of decades prior to 1978.

This observation is probably due to more stringent regulatory measures placed in recent years. It is certain that federal agencies are focusing on safeguarding and preventing alien invasive species using more efficient systems and better monitoring and diagnostic tools. Still, the NPAG database may not be sufficiently comprehensive to draw strong conclusions regarding immigrant species.

Also, many pests may not be detectable if the population levels are low. Furthermore, many plant pathogens have latent periods that it may take several years before becoming visible to human. The first detection is often reported several years after its introduction (USDA FS 2012, Inghilesi et al. 2013).

Spread patterns of alien species often suggest that maximum impacts happen years after initial introduction because spread over suitable areas at the continental scale rarely occurs until several decades have passed. Moreover, some alien species only exhibit invasive behavior after a long period of adaptation to their new environment (USDA FS 2012). Many non-indigenous species may have been already established within the United

States due to increase in international trades but have not been detected in the field yet. In the next decade, we may start detecting large number of alien species establishments.

Nevertheless, while it is very likely that the invasive species record is incomplete, it is unlikely that latent periods and inability to detect low level populations would have a very large impact on a record that spans many decades.

Another key observation of this analysis was that trade volumes were more related to number of non-indigenous species outbreaks than establishments. For example, *Bactrocera dorsalis* and *Ceratitis capitata* have been detected in California, Texas, and Florida numerous times. However, those species have not been successfully established, mainly due to eradication efforts by USDA APHIS.

The number of detections in the field or the number of fruit fly outbreaks is most related to volumes of fresh fruit and vegetable imports from certain countries and climate conditions of both exported countries and the United States. More investigation is needed to further examine the relationship between outbreaks of non-indigenous species and trade patterns.

USDA APHIS and other agencies like CBP continue to be challenged to better understand trade patterns, human movement patterns, and climate change to safeguard American agriculture and natural resources. Forecast models should consider recent patterns that suggest species originating in Asia are more likely to be part of current and expected invasive cohorts.

Although some scientists have stated that non-indigenous forest species would continue to become established in the United States regardless of current regulatory measures

(Aukema et al. 2010), U.S. government agencies can implement new findings and technologies to build more robust forecast models, treatment measures, and integrated pest management practices to protect natural resources.

This chapter has examined past baseline studies, exemplified by Sailer (1978). However, a key goal of my study is to establish a pest forecasting framework. The information and trends described here are powerful instruments to ensure multi-level verification and validation of eventual forecasting tools. The data systems and trends established and described here will be the foundation for ensuring that forecast tools and models have underlying biological information to support them.

## 1.7 References

- Anagnostakis, S. L. 1987. Chestnut blight - The classical problem of an introduced pathogen. *Mycologia* **79**:23-37.
- Aukema, J. E., D. G. McCullough, B. Von Holle, A. M. Liebhold, K. Britton, and S. J. Frankel. 2010. Historical accumulation of nonindigenous forest pests in the continental United States. *Bioscience* **60**:886-897.
- Azar, C., and H. Rodhe. 1997. Targets for stabilization of atmospheric CO<sub>2</sub>. *Science* **276**:1818-1819.
- Cornwell, D., and B. Roberts. 2010. The 9/11 terrorist attack and overseas travel to the United States. Department of Homeland Security, Office of Immigration Statistics, Washington, D.C.
- Eschen, R., A. Roques, and A. Santini. 2015. Taxonomic dissimilarity in patterns of interception and establishment of alien arthropods, nematodes and pathogens affecting woody plants in Europe. *Diversity and Distributions* **21**:36-45.
- FAO. 2012. FAOSTAT. Food and Agriculture Organization of the United Nations. <http://faostat.fao.org/site/342/default.aspx>.

- Fowler, G., Y. Takeuchi, R. Sequeira, G. Lougee, W. Fussell, M. Simon, A. Sato, and X. Yan. 2009. Pathway-initiated pest risk assessment: Asian gypsy moth (Lepidoptera: Lymantriidae: *Lymantria dispar* (Linnaeus)) from China into the United States on maritime shipments. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Raleigh, NC.
- Gilland, B. 2002. World population and food supply: Can food production keep pace with population growth in the next half-century? *Food Policy* **27**:47-63.
- Gullan, P. J., and P. S. Cranston. 2014. *The insects: an outline of entomology*. Fifth edition. Wiley Blackwell, Chichester, West Sussex, UK ; Hoboken, NJ
- Haack, R. A. 2001. Intercepted Scolytidae (Coleoptera) at U.S. ports of entry: 1985-2000. *Integrated Pest Management Reviews* **6**:253-282.
- Haack, R. A. 2006. Exotic bark- and wood-boring Coleoptera in the United States: recent establishments and interceptions. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **36**:269-288.
- Hulme, P. E. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization [electronic resource]. *Journal of applied ecology* **46**:10-18.
- Inghilesi, A. F., G. Mazza, R. Cervo, F. Gherardi, P. Sposimo, E. Tricarico, and M. Zapparoli. 2013. Alien insects in Italy: Comparing patterns from the regional to European level. *Journal of Insect Science* **13**:13.
- IPPC. 2013. ISPM 15: Regulation of wood packing material in international trade. International Plant Protection Convention, Rome, Italy.
- Jervis, R. 2015. U.S. ports prepare for Panama Canal expansion. *USA Today*. USA Today, Tysons Corner, VA.
- Kenis, M., W. Rabitsch, M. A. Auger-Rozenberg, and A. Roques. 2007. How can alien species inventories and interception data help us prevent insect invasions? *Bulletin of Entomological Research* **97**:489-502.
- Koch, F. H., D. Yemshanov, M. Colunga-Garcia, R. D. Magarey, and W. D. Smith. 2011. Potential establishment of alien-invasive forest insect species in the United States: where and how many? *Biological Invasions* **13**:969-985.
- Kovacs, K. F., R. G. Haight, D. G. McCullough, R. J. Mercader, N. W. Siegert, and A. M. Liebhold. 2010. Cost of potential emerald ash borer damage in U.S. communities, 2009-2019. *Ecological Economics* **69**:569-578.

- Kurz, W. A., C. C. Dymond, G. Stinson, G. J. Rampley, E. T. Neilson, A. L. Carroll, T. Ebata, and L. Safranyik. 2008. Mountain pine beetle and forest carbon feedback to climate change. *Nature* **452**:987-990.
- Laudicina, P., and E. Peterson. 2015. Global economic outlook 2015-2020. ATKearney, Washington, D.C.
- Levine, J. M., and C. M. D'Antonio. 2003. Forecasting biological invasions with increasing international trade. *Conservation Biology* **17**:322-326.
- Lidsky, M. 2003. Importation of solid wood packing material: Final environmental impact statement - August 2003. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD.
- Liebhold, A. M., E. G. Brockerhoff, L. J. Garrett, J. L. Parke, and K. O. Britton. 2012. Live plant imports: the major pathway for forest insect and pathogen invasions of the US. *Frontiers in Ecology and the Environment* **10**:135-143.
- Liebhold, A. M., and P. C. Tobin. 2008. Population ecology of insect invasions and their management. Pages 387-408 *Annual Review of Entomology*. Annual Reviews, Palo Alto.
- Mack, R. N. 2003. Plant naturalizations and invasions in the eastern United States: 1634-1860. *Annals of the Missouri Botanical Garden* **90**:77-90.
- MARAD. 2013. Panama Canal expansion study, Phase I Report: Developments in trade and national and global economics. U.S. Department of Transportation (DOT), Maritime Administration (MARAD).
- McCullough, D. G., T. T. Work, J. F. Cavey, A. M. Liebhold, and D. Marshall. 2006. Interceptions of nonindigenous plant pests at US ports of entry and border crossings over a 17-year period. *Biological Invasions* **8**:611-630.
- Meyer, J. 2015. General Entomology: Hemiptera suborder Heteroptera. North Carolina State University, Raleigh, NC. Last accessed January, 2016, <https://www.cals.ncsu.edu/course/ent425/library/compendium/heteroptera.html>.
- Nghiem, L. T. P., T. Soliman, D. C. J. Yeo, H. T. W. Tan, T. A. Evans, J. D. Mumford, R. P. Keller, R. H. A. Baker, R. T. Corlett, and L. R. Carrasco. 2013. Economic and environmental impacts of harmful non-indigenous species in Southeast Asia. *Plos One* **8**:9.
- NOAA. 2015a. Climate at a glance. Department of Commerce, National Oceanic and Atmospheric Administration, National Centers for Environmental Information, Asheville, NC. Last accessed <http://www.ncdc.noaa.gov/cag/>.

- NOAA. 2015b. NOAA in situ carbon dioxide (CO<sub>2</sub>) measurements. U.S. Department of Commerce, National Oceanic & Atmospheric Administration, Earth System Research Laboratory, Boulder, CO. Last accessed December, 2015, <http://www.esrl.noaa.gov/gmd/obop/mlo/programs/esrl/co2/co2.html>.
- NOAA. 2015c. Trends in atmospheric carbon dioxide. U.S. Department of Commerce, National Oceanic & Atmospheric Administration, Earth System Research Laboratory, Boulder, CO. Last accessed December, 2015, <http://www.esrl.noaa.gov/gmd/ccgg/trends/>.
- Nordlie, T. 2012. Citrus greening costs \$3.63 billion in lost revenues and 6.611 jobs, new UF study shows. University of Florida, Gainesville, FL. <http://news.ufl.edu/archive/2012/01/citrus-greening-costs-363-billion-in-lost-revenues-and-6611-jobs-new-uf-study-shows.html>.
- Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. Pages 637-669 Annual Review of Ecology Evolution and Systematics. Annual Reviews, Palo Alto.
- PestID. 2015. Pest Identification Database (PestID). United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine.
- Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics* **52**:273-288.
- Poland, T. M., R. A. Haack, and T. R. Petrice. 1998. Chicago joins New York in battle with Asian longhorned beetle. *Newsletter of the Michigan Entomology Society* **43**:15-17.
- Rana, K. A. A. 2004. World Trade Report 2004. World Trade Organization, Geneva, Switzerland.
- Rose, R. 2014. Asian longhorned beetle response guidelines. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD.
- Sailer, R. I. 1978. Our immigrant insect fauna. Pages 3-11 *Bulletin*.
- Spivak, J. 2011. The battle of the ports. American Planning Association.
- Srygley, R. B. 2014. Effects of temperature and moisture on Mormon cricket reproduction with implications for responses to climate change [electronic resource]. *Journal of insect physiology* **65**:57-62.

- The Economist. 2015. Why everyone is so keen to agree new trade deals (2015, October 6).  
The Economist.
- The World Bank. 2015. World Development Indicators: United States. *in* T. W. Bank, editor.,  
Washington, DC.
- Trefis Team. 2015. The potential impact of the Panama Canal expansion on Union Pacific.  
Forbes. Forbes, Inc., New York, NY.
- U.S. Bureau of the Census. 1975. Historical statistics of the United States, colonial times to  
1970, Bicentennial Edition. U.S. Bureau of the Census, Washington, D.C.
- UCAR. 2016. How much carbon dioxide (and other kinds of greenhouse gas) is already in  
the atmosphere? University Corporation for Atmospheric Research, National Center  
for Atmospheric Research. Last accessed January, 2016,  
[https://www2.ucar.edu/climate/faq/how-much-carbon-dioxide-and-other-kinds-  
greenhouse-gas-already-atmosphere](https://www2.ucar.edu/climate/faq/how-much-carbon-dioxide-and-other-kinds-greenhouse-gas-already-atmosphere).
- UN. 2004. World population to 2300. United Nations, Department of Economic and Social  
Affairs, Population Division, New York, USA.
- UN. 2012. Population. United Nations, Department of Economic and Social Affairs.  
[http://esa.un.org/unpd/wpp/unpp/panel\\_population.htm](http://esa.un.org/unpd/wpp/unpp/panel_population.htm).
- US-Census-Bureau. 2012. U.S. & World population clocks. United States Census Bureau,  
Washington, DC. Last accessed April, 2012,  
<http://www.census.gov/main/www/popclock.html>.
- USDA. 2015. FY 2015 Budget summary and annual performance plan. United States  
Department of Agriculture, Washington, D.C.
- USDA FAS. 2015a. Global Agricultural Trade System. United States Department of  
Agriculture, Foreign Agricultural Service. <http://apps.fas.usda.gov/GATS/>.
- USDA FAS. 2015b. Trans-Pacific Partnership (TPP). United States Department of  
Agriculture, Foreign Agricultural Service, Washington, D.C. Last accessed  
December, 2015, <http://www.fas.usda.gov/topics/trans-pacific-partnership-tpp>.
- USDA FS. 2012. Non-native invasive species best management practices: Guidance for the  
U.S. Forest Service Eastern Region. United States Department of Agriculture, Forest  
Service, Washington, D.C.
- USDA NRCS. 2015. The PLANTS Database. United States Department of Agriculture,  
Natural Resources Conservation Service, National Plant Data Team.  
<http://plants.usda.gov>.

- USDOT. 2006. Freight in America: A new national picture. U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, Washington, D.C.
- USDOT. 2007. Freight facts and figures 2007. U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, Washington, D.C.
- USDOT. 2013. Freight facts and figures 2013. U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, Washington, D.C.
- Vo, T., Y. Takeuchi, and L. Garrett. 2014. Emerald ash borer, *Agrilus planipennis* (Fairmaire): an analysis of the regulatory program. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Raleigh, NC.
- Williamson, M., and A. Fitter. 1996. The varying success of invaders. *Ecology* **77**:1661-1666.
- Work, T. T., D. G. McCullough, J. F. Cavey, and R. Komsa. 2005. Arrival rate of nonindigenous insect species into the United States through foreign trade. *Biological Invasions* **7**:323-332.
- WTO. 2015. Time Series on International Trade. World Trade Organization.  
<http://stat.wto.org/StatisticalProgram/WSDBStatProgramHome.aspx?Language=E>.
- Zhang, R. Z., Y. C. Li, N. Liu, and S. D. Porter. 2007. An overview of the red imported fire ant (Hymenoptera : Formicidae) in mainland China. *Florida Entomologist* **90**:723-731.
- Ziter, C., J. A. Newman, and E. A. Robinson. 2012. Climate change and voltinism in Californian insect pest species: sensitivity to location, scenario and climate model choice [electronic resource]. *Global change biology* **18**:2771-2780.

## CHAPTER 2. Review of Insect Forecast Systems

### 2.1 Introduction

Climate and weather-driven forecast systems play important roles for a wide range of applications. Pest forecast systems are widely used by farmers not only to predict suitable timings for planting, spraying pesticides, and harvest schedules but also to manage cost effective production. Pest forecast systems are also useful for regulatory agencies to create early warning systems as well as to control invasive pests that are already established in the United States.

Historically, the U.S. Department of Agriculture, Animal and Plant Health Inspection Service (USDA APHIS) examined pest risks solely based on climate variables, particularly temperature, although many other factors influence ecological and economic impacts caused by invasive species (Fowler et al. 2009a, Gutierrez et al. 2010). Temperature is a strong factor for growth and reproduction; therefore, characterizing climate data to estimate species suitability has been widely applied in biological conservation studies. However, this “climate envelope” approach may be misleading because it does not include dispersal and interactions, which are important elements of population dynamics (Davis et al. 1998). In addition to the epidemiological rationale and the scientific literature that suggests that a climate envelope alone is limited in its forecasting abilities, there is empirical evidence that using such approaches can not only be misleading but create a lot of unrest in the public (Burke 2008, LBAMspray.com 2008, Clark 2015, Geissinger 2016). The popular press citations noted are just some examples of the hundreds of protests, including entomology

professors at the University of California, that disputed the results presented by USDA APHIS regarding the impact of a pest (Kay 2008).

There are a number of forecast systems and techniques available to predict plant pest species suitability and potential distributions (Elith et al. 2006, Jarnevich and Reynolds 2011). Choosing an appropriate forecast system is crucial yet difficult. Forecasting pest behaviors in natural environments is complex and involves many types of uncertainty; therefore, it is difficult to determine which forecast system(s) would give the most realistic results. For this reason, an ensemble modeling approach is commonly used to predict pest suitability, especially for newly invading pests whose biology are not well known (Stohlgren et al. 2010). Stohlgren et al. (2010) suggested that ensemble models could be useful in risk analysis of a newly established species because the species has not yet expanded its population to all suitable habitats, which makes it difficult to understand the relationship between the environment and the species.

In this chapter, commonly used forecast systems for insect species are reviewed. Also, uncertainties associated with insect forecasts and predictions are discussed. This chapter summarizes forecast approaches suitable for regulatory applications such as in USDA APHIS, i.e., applications to manage invasive alien species as well as uncertainties, all of which need to be addressed by the forecast systems.

## **2.2 Invasive species forecast system**

Pests cause significant economic and environmental damage to natural resources. USDA APHIS PPQ currently have 15 arthropod, 1 mollusk, 2 nematode, and 12 plant

pathogen eradication or containment programs (as of January, 2016) (USDA APHIS 2016b). Nearly 200 other pests are targeted for surveys every year (Cooperative Agricultural Pest Survey (CAPS) program) (USDA APHIS 2016a).

Each program has several management functions: survey, control, and public outreach. For example, official control programs for *Anoplophora glabripennis* Motschulsky (Asian Longhorned Beetle (ALB)) began in fiscal year 1997, for *Phytophthora ramorum* Werres, de Cock & Man in't Veld (Sudden Oak Death (SOD)) in fiscal year 2000, and for *Agrilus planipennis* Fairmaire (Emerald Ash Borer (EAB)) in fiscal year 2002. The USDA provided over \$420 million (\$249 million for ALB, \$61 million for SOD, and \$112 million for EAB) for those three programs by 2006, primarily through the Commodity Credit Corporation fund and appropriations to APHIS's emerging plant pest program (Bertoni 2006). The overall allocation to APHIS PPQ (including the plant-pest interception activities by DHS-CBP funded by PPQ through its user fees) from all sources in 2016 approaches \$1 billion. Therefore, identifying suitable and potential spread areas becomes important as a foundation to manage unwanted pests.

### **2.2.1 Brief literature review on forecast systems**

A pest forecast system is defined as a system to predict pest-suitable areas and pest behavior. Some systems estimate potential species distribution or habitat niche (Stohlgren et al. 2010, Morisette et al. 2013), while others evaluate timing of biological events (e.g., phenology models) (Sevacherian et al. 1977), population growth (e.g., population dynamics) (Liebhold and Bascompte 2003, Liebhold and Tobin 2008, Gutierrez et al. 2010), pest spread (Meentemeyer et al. 2011), and economic damages (Wittwer et al. 2005, Soliman et al. 2010,

Cook et al. 2011). Many researchers and regulatory decision makers now use multiple approaches to estimate risks of invasive species.

### **2.2.2 Pattern Matching and fitness-driven models**

Predicting pest suitable areas implies identifying a pest's potential distribution and habitat niche. This is particularly important for resource managers to either promote or control the species of concern (Thuiller et al. 2005, Morissette et al. 2013). There are several approaches for identifying pest suitable areas. One approach is to predict the suitability based on the mathematical relationship between physiological conditions and current distribution areas. This approach is often called climate matching. Climate matching approaches are useful and important when the biology of the species is not well known or studied.

A second approach is to understand aspects of the biology of the pest, such as required temperatures for growth (optimal, threshold high, and threshold low), biological conditions (e.g., soil moisture, precipitation), and dispersal means. In this approach, a model predicts suitable areas based on the specific biology of the pest.

There are number of techniques used in ecological niche modeling. Ecological niche modeling predicts potential species habitat distribution by developing mathematical algorithms between known distributions and environmental conditions (Peterson 2003). Several mathematical algorithms are commonly used to identify species suitable areas: MaxEnt, Genetic Algorithm for Rule-set Prediction (GARP), Boosted Regression Trees, Random Forest, Generalized Linear Model (GLM), and Multivariate Adaptive Regression

Splines (MARS) (Rodda et al. 2007, Morisette et al. 2013). Commonly used mathematical approaches to identify suitable habitat areas are discussed in this section.

### **MaxEnt:**

MaxEnt, or Maximum Entropy, is an approach to estimate the probability distribution of a species from presence-only data. This approach was first used with biological specimens to estimate potential distributions. This idea became particularly interesting because many of the biological data were only presence-only data, especially historical museum and herbaria data and data from poorly surveyed areas (Phillips et al. 2005).

Phillips et al. (2005) listed several advantages and disadvantages of the MaxEnt approach. The advantages include that 1) it requires only presence data, 2) it can utilize both continuous and categorical data, 3) deterministic algorithms have been developed to converge the maximum entropy probability distribution, 4) the MaxEnt probability distribution has a concise mathematical definition, 5) over-fitting can be avoided, 6) the outcome is continuous, and 7) the MaxEnt approach can be applied to species presence/absence data by using a conditional model (Phillips et al. 2005). The disadvantages are that 1) it is not as mature a statistical method as other methods such as GLM and generalized additive models (GAM), 2) it does not predict well outside of the study area because it uses an exponential model for probabilities, and 3) MaxEnt is not available in standard statistical packages (Phillips et al. 2005).

A number of studies have estimated potential habitat niches of invasive species by using MaxEnt. The approach has been used for plant species (Stohlgren et al. 2010,

Jarnevich and Reynolds 2011), insect pests (Dupin et al. 2011, Lozier and Mills 2011, Kumar et al. 2014), and plant pathogens (Narouei Khandan et al. 2013).

MaxEnt analysis is run by using MaxEnt software for species habitat modeling. Users can download the software free for educational and research purposes at <http://www.cs.princeton.edu/~schapire/maxent/>. MaxEnt is also part of the Software for Assisted Habitat Modeling (SAHM) package for VisTrails developed by US Geological Survey (USGS), Fort Collins Science Center. It is also free for research purposes.

### **Genetic Algorithm for Rule-set Prediction (GARP):**

Genetic Algorithm for Rule-set Prediction (GARP) is an ecological niche modeling technique that relates ecological characteristics of known distribution areas to characteristics of other areas (i.e., where occurrence has not been documented) to predict which of them could maintain the species population (Stockwell and Peters 1999, Adjemian et al. 2006). GARP was originally developed by David Stockwell at ERIN Unit of Environment Australia and was later modified at the San Diego Supercomputer Center.

GARP is a two-step process that generates rules to describe presence and absence in variable space (Blackburn 2010). This rule set is applied to generate spatially explicit predictions of presence and absence (Blackburn 2010). GARP is known to predict well with small samples of presence-only data (Stockwell and Peterson 2002).

GARP was once uncommon in invasive pest species analyses (Peterson et al. 2003), but in recent years, it has become one of the most commonly used ecological niche modeling approaches for approaches. Some examples of potential geographical distribution predictions by using GARP are *Anoplophora glabripennis* in North America (Peterson et al.

2004), *Ceratitis capitata* and *C. rosa* in Africa (De Meyer et al. 2008), *Phenacoccus solenopsis* in China (Wang et al. 2009), and *Linepithema humile* at global scale (Roura-Pascual et al. 2004). Evaluation of Asian longhorned beetle (*A. glabripennis*) in North America based on a few known distribution points from its native range (40 occurrence records) by Peterson et al. (2004) revealed the Great Lakes region (Ohio, south Michigan, and northeast Indiana) as a critical area. This result coincided with a new infestation from Clermont County, Ohio in 2011 (USDA APHIS 2015). However, some studies showed that GARP models over-predicted species' geographical niches (Peterson et al. 2002, Larson et al. 2010).

GARP can be run by the software called DesktopGARP. It was developed by Scachetti-Pereira at University of Kansas Biodiversity Research Center in collaboration with the Reference Center for Environmental Information in Campinas, Brazil. DesktopGARP can be downloaded at <http://www.nhm.ku.edu/desktopgarp/Download.html> and installed on Windows platform. Users must supply coordinate information (latitude and longitude) of the species locations and select environment layers that would be used to define environmental conditions of the distribution. DesktopGARP includes environmental layers such as elevation, slope, and aspect (Scachetti-Pereira 2014). The results can be output as maps in bitmap, ascii raster grid, or ESRI Arc/Info grid formats (Scachetti-Pereira 2014).

### **Boosted Regression Trees:**

Boosted regression trees (BRT) is another modeling technique to predict species suitability. It is a relatively new technology and has been rarely used in ecological studies

(Elith et al. 2008). BRT is a combination of two algorithms: regression trees (decision trees) and boosting builds (a collection of models).

Decision trees are statistical multivariate analyses to determine classifications. Regression trees fit the mean response for an observation in the appropriate class (Elith et al. 2008). The decision tree approaches are valued because they are easily visualized, can be simple, and can accommodate any type of data as predictors (e.g., numeric, binary, categorical data) (Elith et al. 2008).

Boosting is a machine learning algorithm to combine the performance of many weak classifiers to create an accurate set (Friedman et al. 2000, Elith et al. 2008). Boosting algorithms vary based on the types and setting of the data (Elith et al. 2008). For BRT, the first regression tree reduces the loss function, and subsequent steps focus on the residuals (Elith et al. 2008). BRT requires presence and absence records of species with dependent variables. It can be run by using R project for statistical computing (Elith and Leathwick 2016) or USGS SAHM.

### **Random Forest:**

Random Forest (RF) is a machine learning algorithm that ensembles classifiers (Breiman 2001). Breiman (2001) defines random forest as a classifier consisting of a collection of tree-structured classifiers where the  $k$ th tree, a random vector, is independent of the past random vectors and each tree casts a unit vote for the most popular class at input  $x$ . In other words, each node is split by using the best among a subset of predictors randomly chosen at that particular node (Breiman 2001, Liaw and Wiener 2002). Since a large number of trees are grown with RF, this counterintuitive strategy produces limited generalization

error and does not overfit the data (Breiman 2001, Liaw and Wiener 2002, Prasad et al. 2006).

Prasad et al. (2006) pointed out that RF could be a “black box” since the individual trees could not be examined separately. However, RF does provide performance metrics to evaluate the results.

RF is a relatively new technique in biological applications (Prasad et al. 2006, Cutler et al. 2007) but is one of the commonly used species distribution models (SDMs). RF has often used to estimate potential distribution on invasive plant species (Cutler et al. 2007, Stohlgren et al. 2010, Naidoo et al. 2012); however, it has been applied to predict potential spread of forest pathogens (Philibert et al. 2011) and to model arthropod diversity based on environmental factors (Bowser and Morton 2008).

The randomForest R package developed by Breiman and Cutler provides an R interface to run RF (Liaw and Wiener 2002). It is also one of the five SDMs that are provided through USGS SAHM package (Talbert and Talbert 2012).

### **Generalized Linear Model (GLM):**

A Generalized Linear Model (GLM) is a type of linear regression model that does not force data into unnatural scales, allowing for non-linearity and non-constant variance structures in the data (Guisan et al. 2002). The model is built based on the relationship between the mean of the response variable and the linear combination of the explanatory variables (Guisan et al. 2002). Data used in GLMs can be from several types of probability distributions, such as normal, binomial, Poisson, negative binomial, and gamma (Guisan et

al. 2002). Thus, GLM are flexible and suited for an ecological modeling (Guisan et al. 2002).

GLM has been used to identify the potential distributions of many categories of invaders (e.g., insect, plant pathogens, mammals, and plants) as well as other environmental phenomena (Meentemeyer et al. 2008, Roura-Pascual et al. 2009, Vaclavik and Meentemeyer 2009). For example, GLM was used to predict potential distribution areas of Argentine ants (*Linepithema humile*) in the Iberian Peninsula (Roura-Pascual et al. 2009), sudden oak death (*Phytophthora ramorum*) potential dispersal areas in California (Meentemeyer et al. 2008, Vaclavik and Meentemeyer 2009), and potential wildfire under climate change (West et al. 2016).

GLM requires both presence and absence distribution locations. GLM is one of the methods included in the USGS SAHM package (Talbert and Talbert 2012). It can be also run by using the glm function in R.

### **Multivariate Adaptive Regression Splines (MARS):**

A Multivariate Adaptive Regression Splines (MARS) is a flexible regression model that works by fitting piecewise logistic regressions (Friedman 1991). MARS is a nonparametric regression that makes no assumptions about the underlying functional relationship between the dependent and independent variables (Friedman 1991). It breaks the range of each predictor variable into subsets using ‘knots’ and allows the slope of the fitting linear segments between pairs of knots to vary while ensuring the full fitted function is without breaks (Elith and Leathwick 2007). Model fitting is a fairly fast process and produces smoother and less coarse-grained outputs than other methods (Prasad et al. 2006,

Elith and Leathwick 2007). However, it tends to be excessively influenced by the local nature of the data (Prasad et al. 2006).

MARS has been used to predict habitat niches for freshwater fish species, birds, mammals, reptiles, plants, and trees (Leathwick et al. 2006, Prasad et al. 2006, Elith and Leathwick 2007, Stohlgren et al. 2010). MARS requires both presence and absence distribution locations (Elith and Leathwick 2007). MARS is one of the SDMs included in USGS SAHM package (Talbert and Talbert 2012). It can be also run by using the earth package in R (Milborrow 2016).

### **2.2.3 Phenology models**

A phenology model predicts timing of an organism's developmental stages. Most insects are poikilothermic; therefore, environmental temperatures directly affect the rate of insect growth and development if other resources are not limited (e.g., water, food) (Gullan and Cranston 2010). These kinds of models are often called "degree day" models because a key calculation is a standardized estimate of heat exposure (heat units or degree days) for the organism of interest. Many pest species including most insects require a certain amount of heat exposure to develop from one stage to another. This required heat over time is expressed in physiological time because each insect requires a different amount of heat for development and because environmental temperature varies each day (Baskerville and Emin 1969, Gullan and Cranston 2010). Physiological time is the cumulative developmental time multiplied by the temperature above a declared developmental threshold; therefore, physiological time is commonly expressed as degree-days ( $D^{\circ}$ ) (Baskerville and Emin 1969, Gullan and Cranston 2010).

The original idea of organisms developing as a function of heat unit accumulation is 300 years old and reportedly originated with the French scientist Rene de Reaumur (Reaumur 1736, Fraisse et al. 2011). I note that the notion of growth as influenced by temperature has been well understood and well reported. There is no question that de Reaumur was very clear not only about the fact that growth was influenced by temperature but also that this fact was likely to be useful as a tool. The quote below was translated directly from the original French volumes.

*“... among butterfly of the same species, there are those that remain more or less time under the form of chrysalis. This depends on the season in which the caterpillar is transformed. This fact merited more than being emphasized. It really merited that we be attentive to the singular consequences that we can deduce. It also merited conducting the experiments that are invited by these observations. These observations led us directly to reflect that the composition of the body of the insect, its machines and apparatus is no less admirable than that with which we ourselves are put together with. We can almost prolong or reduce according to our will the duration (of the insect life cycle). Thus we can make the course of the life of the insect to be twice, three times, four times, etc. longer than it would have heretofore known to be possible relative to the behavior of members of its species. We can also reverse this pattern (prolong or reduce the cycle) without hurting the insect, without destroying them. We can reduce considerably the course (duration) of its life. That is to say, we can put this insect in a state that will make the duration (of the life cycle) relatively short and note that the same sequence of operations can also occur with a duration of time that would be much longer. We can, I say, come up to from observations and consequences to*

*this conclusion by the observations which have taught us that a caterpillar does not remain in the summer but a few weeks under the form of chrysalis. Whereas another butterfly of the same species will remain during several months the same form of chrysalis if one should observe the sequence in the fall. From that, there is certain degree of heat that can render the growth of the butterfly very fast. There is certain degree of cold unit that renders the growth very slow. The heat and cold temperatures will influence in such a way to increase or decrease the growth of insects. Thus, the entire life of any such animal is no more than sequence of degree of growth. And (similarly) a sequence or degree of non-growth. It is curious to see if we can continue to develop this idea not only to prolong and to reduce the duration of the life of insects that links successively caterpillars, chrysalis, and butterfly. Thus, we can in the same way ask about the life of insects in general. Whether they follow the same transformations or not. These are the kind of research activities that were the focus of my first dissertation. We should be excited to push this research further that we have up until now because of the curious knowledge that is promised. In fact, they seem to be promising specific tools (and/or usefulness) (for other species) in the same way that we find to be the case for this last (studied) species.” (Reaumur 1736)*

Thus, the entire apparatus of temperature dependent growth and insect development forecasts has a long history and one that was formally launched three hundred years ago.

The phenology or so-called degree-day model is commonly used in insect management. Understanding the timing of a given pest stage is important because economic losses are often associated with specific insect stages and applications of management practices are more effective at specific insect stages (Dennis et al. 1986, Herms 2013).

Although there are several methods to calculate degree-days, a degree-day model is fundamentally a function of daily minimum, daily maximum and base temperature (insect growth threshold temperature) (Herms 2013). Common methods to calculate degree-days are 1) average method, 2) modified average method, 3) triangulation method, and 4) sine wave method (Baskervi and Emin 1969, Allen 1976, Zalom et al. 1983, Roltsch et al. 1999, Herms 2013).

The “average” method calculated by the following equation (Roltsch et al. 1999, Herms 2013) (Figure 2-1 A.):

$$\text{Degree day} = \frac{(\text{Daily Maximum Temp} + \text{Daily Minimum Temp})}{2} - (\text{Minimum Threshold Temp})$$

Although this method is easy to calculate, it generates some of the least precise results (Roltsch et al. 1999). This method works best if daily minimum temperature is above the threshold temperature. Since this method does not take into account the length of time that daily temperature is above the threshold temperature, accumulated degree-days could be overestimated or underestimated (Zalom et al. 1983).

A modified average method was developed to take account of situations where daily minimum temperature was below the threshold temperature or daily maximum temperature was above the maximum threshold temperature (Herms 2013). When daily minimum temperature is below the minimum threshold temperature, the threshold temperature is used for daily minimum temperature.

*Degree day =*

$$\frac{(\text{Daily Maximum Temp} + \text{Minimum Treshold Temp})}{2} - (\text{Minimum Threshold Temp})$$

When the daily maximum temperature is above the maximum threshold temperature, the maximum threshold temperature is used for daily maximum temperature.

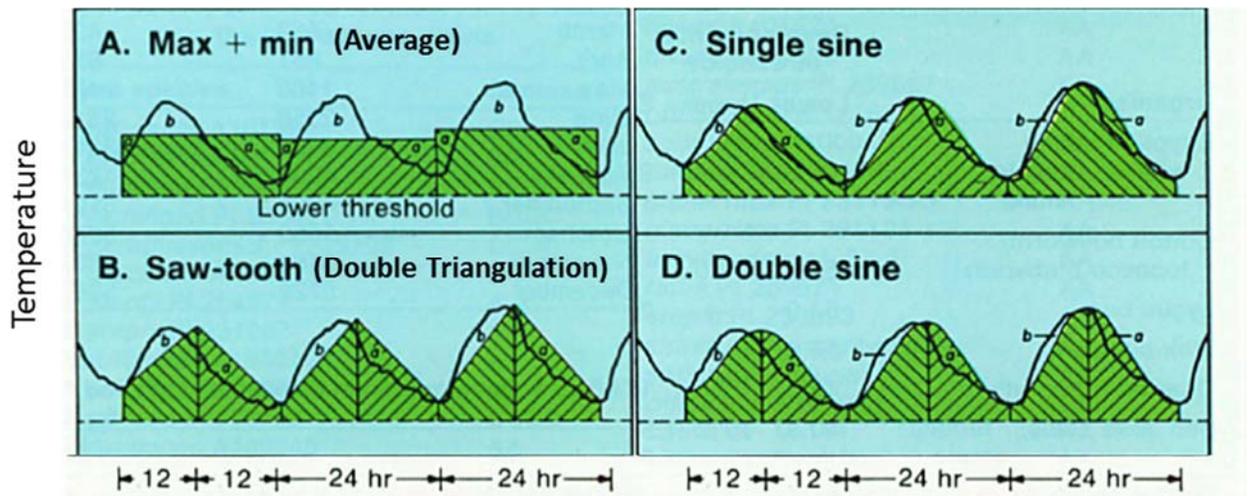
*Degree day* =

$$\frac{(\text{Maximum Threshold Temp} + \text{Daily Minimum Temp})}{2} - (\text{Minimum Threshold Temp})$$

There are two commonly used triangulation calculations (Figure 2-1 B.). One is called “single triangulation” and the other is “double triangulation”. The single triangulation method creates an equilateral triangle based on daily maximum and minimum temperature over a 24-hour period and then calculates the triangle area between the two thresholds as degree-days (Lindsey and Newman 1956). The triangulation calculation is still straightforward; however, one major problem is that the single triangulation method uses the same minimum temperature to calculate both halves of the triangle (Sevacherian et al. 1977) (Figure 2-1 B.). Sevacherian et al. (1977) modified the triangulation degree-day method to calculate the first half of triangle with morning minimum temperature and the second half of triangle with the next day’s minimum temperature, as Allen (1976) proposed for the double sine wave method. This method is called the double triangulation method.

The sine wave method assumes that the trend of daily temperature follows the form of sine wave (Allen 1976). There are two types of sine wave methods (Figure 2-1 C., D.). The single sine wave method fits a sine curve based on daily minimum and maximum temperature and calculates the area under the curve above the minimum threshold as degree-days (Baskervi and Emin 1969). In contrast, the double sine wave method fits a sine curve through morning minimum temperature and afternoon maximum temperature to calculate the first half of the day and fits different a sine curve with the afternoon maximum temperature

and the minimum temperature of the next day to calculate the other half of the day (Allen 1976, Roltsch et al. 1999).



**Figure 2-1. Diurnal temperature shapes based on four methods (average, double triangulation, single sine, and double sine) of calculating degree-day accumulation (Wilson and Barnett 1983)**

Degree-day models have been developed under the assumption that daily temperature profiles can be expressed by a specific geometric shape (Roltsch et al. 1999). With this assumption, sine wave methods are commonly used for degree-day calculation because daily temperature tends to increase and decrease in sine wave shape. However, sine wave degree-day calculations are complex, especially with horizontal, vertical, and intermediate cut-offs. When the sine wave methods were developed, calculation with computers was still costly and was not common. Although the triangulation method is simpler to calculate than the sine wave method, both single triangulation and single sine wave methods estimated degree-days relatively well (Roltsch et al. 1999). The triangulation method performed well during winter and early spring months, but only in combination with certain cut-off techniques (Roltsch et al. 1999). The equations for sine wave and triangulation with horizontal, vertical, and intermediate cut-offs are listed in Appendix A.

In regulatory applications, phenology models are commonly used in pest management and phytosanitary risk assessments (Nietschke et al. 2007, Fowler et al. 2009a, Fowler et al. 2009b). Since a phenology model estimates timing of pest events (e.g., eclosion, adult emergence, plant germination), it is often used to specify periods of control measures (e.g., biological control, chemical control) (Sevacherian et al. 1977), to target survey periods (Morris et al. 1956, Fowler et al. 2009b), and to estimate potential number of generations (Bryant et al. 1997, Fowler et al. 2009a).

#### **2.2.4 Population dynamics**

A population is a group of individuals that live in the same geographical area. Studies of population dynamics focus on changes in population density over time (Liebhold and Bascombe 2003). A current population is often calculated by adding births and immigration but subtracting deaths and emigration from a previous population. In other words, population change rate is determined by birth rate, immigration rate, mortality rate, and emigration rate.

The population of a species increases under optimal conditions due to higher birth rate, possibly higher immigration rate moving to more favorable conditions, and lower mortality rate. Conversely, the population decreases when mortality rate and emigration rate become high and birth rate becomes low due to growth limiting factors, such as food availability and temperature.

There are several factors, both biotic and abiotic, affecting population densities. There are two categories of factors that affect population densities: density-independent factors and density-dependent factors. Density-independent factors impact all individuals,

regardless of the overall population density. Generally, those factors are climate conditions. For example, extreme cold temperatures kill all fruit flies regardless of the population density of fruit flies within an area of interest. On the other hand, density-dependent factors are features that affect an insect population based on its population size. Common examples are predation, parasitism, disease, and resource availabilities. For example, the rate of ambrosia fungus inoculations (*Raffaelea* sp.) in red bay (*Persea borbonia* (L.) Spreng.) and sassafras (*Sassafras albidum* (Nuttall) Nees) is likely to be affected by the population of ambrosia beetles (*Xyleborinus gracilis* Eichhoff, *Ambrosiodmus obliquus* LeConte, and *Xyleborus glabratus* Eichhoff) because ambrosia beetles are known to inoculate ambrosia fungi (Mueller et al. 2005, Fraedrich et al. 2008).

Another density-dependent factor that affects insect populations is called the Allee effect. Warder Clyde Allee first observed that, at higher population densities, goldfish in a tank increased population faster than at lower population densities (Allee 1931). This is mainly due to not being able to find mates, inbreeding depression, and lack of cooperative feeding at low densities (Courchamp et al. 1999, Liebhold and Tobin 2008); consequently, Allee effects often lead to extinction.

The Allee effect applies to insect populations as well, and an insect population must be above a minimum population threshold to remain viable (Allee 1931, Courchamp et al. 1999, Liebhold and Tobin 2008). The Allee effect is particularly important during the establishment phase of invasive species and can be a critical determinant of the invasiveness of a species (Liebhold and Tobin 2008). However, a variety of factors causes Allee effects,

and it is challenging to understand the combination and degree of Allee effects for a given insect species (Liebhold and Tobin 2008).

Pest population dynamics models can be complex because pest populations are affected by a number of factors such as Allee effects, weather, hosts, natural enemies, and niche conditions (Wallner 1987, Elkinton and Liebhold 1990). However, it is important to understand how populations increase and decrease over time to predict outbreaks. The populations of Gypsy moth (*Lymantria dispar* Linnaeus), non-native forest pests that have invaded North America, have probably been the most intensively studied. Gypsy moth population densities have been evaluated by investigating predator population levels (Elkinton et al. 1996), correlated with weather conditions by using Moran's effect (Peltonen et al. 2002, Allstadt et al. 2015), evaluated by developing an insect-pathogen (nuclear polyhedrosis virus) dynamics model (Dwyer and Elkinton 1993, Dwyer et al. 2000), studied by examining gypsy moth diet and nutritional factors of hosts (Rossiter 1991), and correlated with oak defoliation (Williams et al. 1991). In addition, Allee effects and low-density population effects of gypsy moth have been evaluated to understand the factors regulating gypsy moth populations in the field (Liebhold et al. 2000, Liebhold and Bascompte 2003).

### **2.2.5 Pathway and spread model**

Pests are introduced to new environments purposefully, accidentally or through natural (not human mediated) spread. Although a number of studies have indicated that unintentional introduction of alien species is linked to globalization (Hulme 2009, Aukema et al. 2010, Koch et al. 2011, Nghiem et al. 2013), the rate of recent unintentional arthropod species introduction remained similar to the findings by Sailer (1978) (Chapter 1). However,

the methods (pathways) of pest introduction to new environments have changed.

Traditionally, insect pests were often introduced through the movement of agricultural commodities (Niemela and Mattson 1996, Sequeira 2002b, Tyler and Hodges 2002, Areal et al. 2008). In recent years, alien invasive species such as *Anoplophora glabripennis* Motschulsky (Asian longhorned beetle: ALB) and *Agilus planipennis* Fairmaire (emerald ash borer: EAB) are believed to have been introduced to the United States through usage of solid wood packing materials (Poland and McCullough 2006, Haack et al. 2010). Hitchhiker type pests (e.g., Asian gypsy moth) can travel to the United States without having host species present (Fowler et al. 2009b).

In addition to pest pathways through human assisted activities, pests have ability to disperse due to their own dispersal abilities and weather events. The introduction of soybean rust (*Phakopsora pachyrhizi* Sydow) into the continental United States was likely due to hurricanes and tropical storms, notably Hurricane Ivan in 2004 (Schneider et al. 2005). Not only plant pathogens but also insects can be moved for a long distance with weather events. African locusts (*Schistocerca gregaria*) has been reported to travel long distances (~5,000 kilometers) with the aid of hurricanes (Richardson and Nemeth 1991). Other examples are the Old World bollworm (*Helicoverpa armigera* (Hübner)), which can migrate 250 km per night (Feng et al. 2010), and the first instar larvae of the Asian gypsy moth (*Lymantria dispar* Linnaeus), which can balloon several kilometers with wind depending on the population density and food availability (Zlotina et al. 1999).

There are a number of general approaches to model pathways and dispersal of pests. The Pratique (a framework developed to improve management of phytosanitary issues

mainly for European Union) provides five types of pathway/spread analysis techniques. Those five models are 1) Simple Logistic Growth (SLG), 2) Logistic Growth with Economic Values (LG-Econ), 3) Radial Rate Expansion (RR), 4) Random Radial Rate Expansion (Rand-RR), and 5) population growth and dispersal modeling (Robinet et al. 2011). The advantages of the Pratique models are that they are well documented and provide clear guidance, and some approaches (e.g. SLG, RR) are simple to run. However, some models require secondary model outputs e.g., CLIMEX outputs for Ecoclimatic Index and Growth Index (Robinet et al. 2011); therefore, the users must have access to CLIMEX models. The population growth and dispersal model is the combination of population growth within cells and a probability distribution of dispersal distance (Robinet et al. 2011). This calculation can be a deterministic or a stochastic process (Robinet et al. 2011).

Another well-known stochastic dispersal modeling approach for invasive species include a Markov chain Monte Carlo (MCMC) (Cook et al. 2008, Meentemeyer et al. 2011). A Markov chain is a mathematical model for stochastic systems that is governed by a transition probability (Geyer 2011). The current state in a Markov chain depends on the previous state (1<sup>st</sup> order Markov chain). Monte Carlo simulation is applied to build the probability of outcomes by random samples. MCMC estimates the potential pest distribution areas by using a Markov chain sampling method.

This stochastic approach has been used to forecast the potential dispersal areas of plant pathogens (Cook et al. 2008, Meentemeyer et al. 2011) and insects (Koch et al. 2009). Meentemeyer et al. (2011) used MCMC to predict how and where sudden oak death

(*Phytophthora ramorum* Werres, de Cock & Man in't Veld) could spread from its current distributions in California.

### **2.3 Historical and Current Insect Forecast Systems in USDA APHIS**

USDA APHIS prepares risk assessments of quarantine pests to evaluate the potential risks and economic impacts caused by plant pests. Some risk assessments were prepared using simulation models with geo-referenced data (Sequeira 1999). Until the early 2000s, USDA APHIS did not consistently use pest forecast systems; therefore, the analysts used approaches that were available in an ad hoc manner. There are four well-known pest forecast systems that have been used within APHIS; however, many early assessments were done by developing pest specific models and an “un-branded” approach, often simply described as “epidemiological” (Sequeira 1999).

In this section, four systems that USDA APHIS has previously used are reviewed by evaluating *Epiphyas postvittana* Walker (light brown apple moth: LBAM) as an example. Those systems are 1) CLIMEX, 2) NAPPFAST, 3) VisTrail SAHM (MaxEnt), and 4) USPest. Some of the results produced by these models have been published as noted below and some are recreated here for the first time.

#### **2.3.1 CLIMEX**

CLIMEX is one of the most well-known forecasting systems widely used to predict potential distributions of biological organisms (e.g., weeds, insects, and animals) (Rodda et al. 2007, Lozier and Mills 2011, Narouei Khandan et al. 2013). CLIMEX was developed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Entomology in

1985 (Sutherst et al. 2007). CLIMEX was originally developed to predict exotic species survival in Australia. Currently, CLIMEX is widely used globally by scientists, universities and governments.

CLIMEX is a software program, which must be installed on individual computers to run the models. It includes a global meteorological database consisting of approximately 3,000 locations covering the entire globe. Each location contains weekly temperature and precipitation data. The models can run at those weather station point locations or on gridded weather data, which are interpolated across those weather stations. Users are able to modify and update weather data through the CLIMEX MetManager interface. Weather data can be updated with the most recent data, and new weather stations can be added into the database by users. CliMond (Global Climatologies for Bioclimatic Modeling) currently provides raster-based weekly climate data for CLIMEX (<https://www.climond.org/Default.aspx>). The CliMond dataset was generated from climate data from 1961 to 1990 at 10' resolution and 30' resolution. Users are also able to incorporate irrigation and climate change effects into the climate data. However, only a single irrigation amount can be applied to an entire study area (not area specific), although different irrigation amounts can be entered by season (winter and summer) or by month. Climate change scenarios are also applied in a similar way. Different effects in temperature, precipitation, and evaporation can be applied by season, but those effects can only be applied spatially in limited ways: either the same effect is applied to the entire study area or by 0.1 latitude gradient. In other words, the forecasts derived from General Circulation Models cannot be integrated into the CLIMEX system.

CLIMEX is a climate-based model and consists of two distinct tools to predict potential distribution. The first type is CLIMEX mechanistic simulation models, and the second type is CLIMEX match climate models (Sutherst et al. 2007). CLIMEX match climate models run based on the current distribution of the species of interest, without any biological information. It strictly compares climate conditions for a target area with the known distributions of the species. CLIMEX mechanistic simulation models enable users to parameterize models based on the biology of the species (e.g., optimal temperature, growth threshold temperature, degree day, moisture requirement, stresses caused by cold, heat, wetness and dryness, diapause, and lay length). The mechanistic models simulate population dynamics by evaluating suitable climate for growth based on species growth rate and climate data. Therefore, parameterization of those mechanistic simulation models is critical to the result. Parameter fitting in CLIMEX requires substantial skill, and estimating them based on accurate biological data, distribution maps, abundance and seasonal activities results in more reliable predictions. Users first fit parameters based on known biological information and examine the output (climatologically preferred areas) based on the current distribution. Users need to adjust parameters until the output distributions are similar to the known distributions.

CLIMEX expresses overall suitability with a measurement called the Ecoclimatic Index (EI). The species-specific model constricts a climatological simulation for predicting a pest's distribution based on its biology and environmental stress indices, and characterizes the suitability with EI. The EI value is a function of the annual growth index and annual stress index (Sutherst et al. 2007). The growth index is determined by temperature, moisture, radiation, substrate (e.g., soil type, topography, host availability, suitability for a parasitic

organism), light, and diapause. The stress index is determined by cold, heat, dryness, wetness and a combination of temperature and moisture stresses. The EI is scaled between 0 and 100, with EI values close to 0 indicating that the location is not suitable for long-term survival of the species, while EI values above 30 represent favorable conditions for the species (Rafter et al. 2008). The EI model can be calibrated so that its predictions give the best agreement with where the pest naturally occurs.

The advantages of CLIMEX include many factors described here. The system is well known and has been widely used by scientists for many years. There are a large number of publications (more than 100) describing the use of CLIMEX to predict species suitability. Therefore, parameters used for the studies are often easily found through existing literature and easily shared with others. CLIMEX mechanistic models also provide stress values (e.g., cold, hot, wet, dry, cold-wet, hot-dry); therefore, it is easy to understand specific environmental conditions that lower the EI value at each location or each cell. It also has disadvantages. CLIMEX is a commercial product; users must buy the software to run models. There is an advantage of using commercial products like CLIMEX because technical support is available if needed. CLIMEX-DYMEX Suite 4 currently costs US\$1315 for a single license. Models included in CLIMEX are not modifiable; however, DYMEX enables users to modify models and create new models. CLIMEX-DYMEX does have a maintenance cost at an annual price of \$265, but this is optional. Another disadvantage is default CLIMEX climate data are at point level. Although CLIMEX recommends using CliMond climate dataset, which is in raster format, the data are still treated as points in the CLIMEX system. CLIMEX treats raster data as point data by calculating at a centroid of

each cell. Therefore, post-processing is required to convert into raster format outputs even when the model is run with raster format climate data. Another issue is that CliMond data are the averaged climate data from 1961 to 1990. Where running models with more recent climate data is appropriate, user must update the climate data into specific format periodically.

Several studies have predicted the potential global distribution of LBAM (light brown apple moth) by using CLIMEX “Compare locations (1 species)”. Lozier and Mills (2011) used known biological information of LBAM with known distributions from Australia, New Zealand, Hawaii, United Kingdom, and California to predict potential LBAM establishment at a global scale (Figure 2-2). A similar study was done by He et al. (2012), who parameterized an LBAM CLIMEX model based on LBAM biological data and calibrated with its current distribution in its native range (Australia) and in New Zealand where the moth has been established for more than 100 years (Figure 2-3). Then, the model was validated by using recent introductions of LBAM in Caledonia, UK, Hawaii, and California. The parameters used by Lozier and Mills (2011) and He et al. (2012) are listed in Appendix B.

The model parameterizations from those two studies were slightly different but similar. The major difference was that the model parameterized by He et al. (2012) was more moisture tolerant than the one by Lozier and Mills (2011). The results from the two studies are shown in Figure 2-2 and Figure 2-3. There are some differences between those two studies. The model parameterized by He et al. (2012) predicted the eastern United States was more suitable. On the other hand, the model by Lozier and Mills (2011) predicted

favorable suitability in Florida while the model by He et al. (2012) predicted unsuitable conditions in Florida.

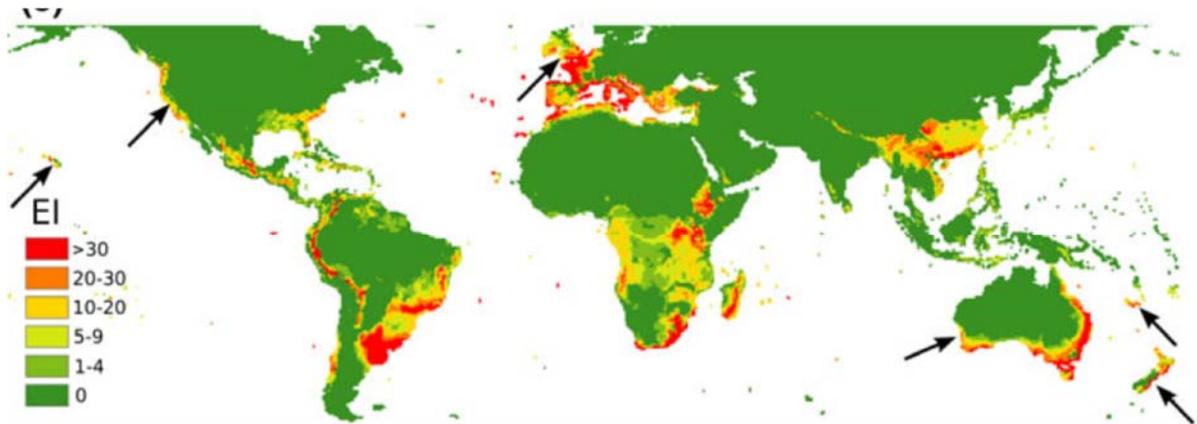


Figure 2-2. Map of environmental suitability for LBAM resulting from CLIMEX Lozier and Mills (2011)

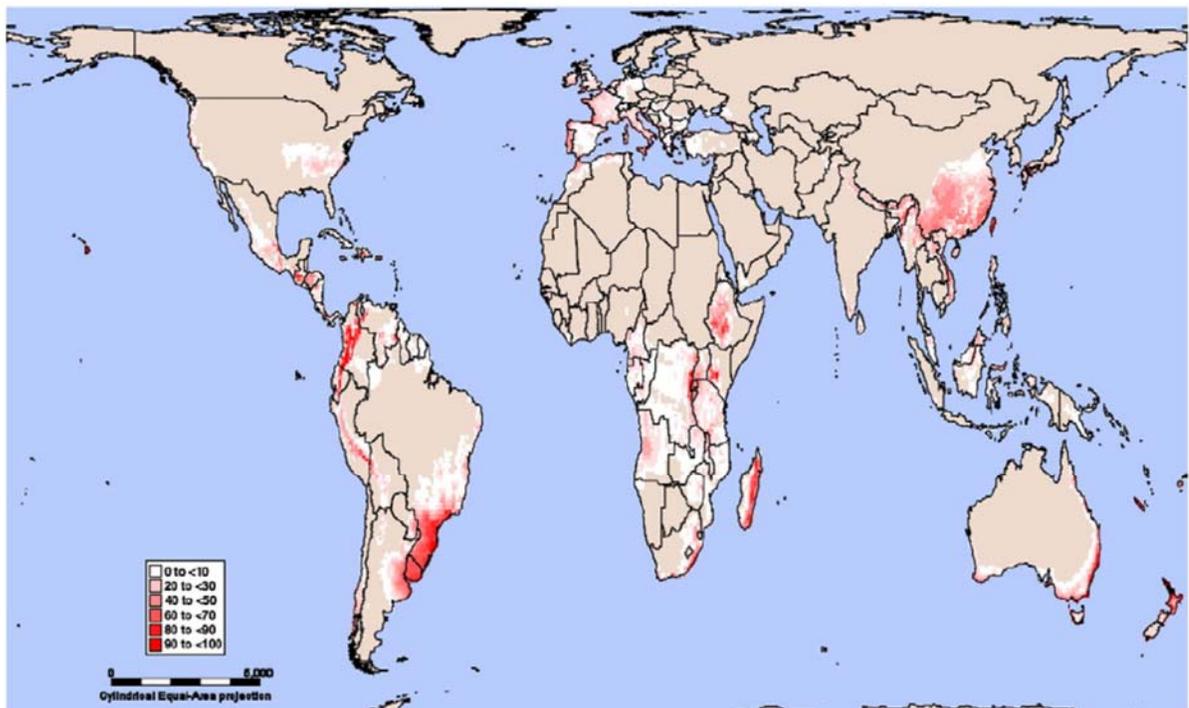


Figure 2-3. Predicted potential global distribution of LBAM by CLIMEX. Red color represents EI values ranging from 0-100, which represent the suitability of a location for LBAM establishment (He et al. 2012).

### **2.3.2 NAPPFAST (NCSU-APHIS Plant Pest Forecasting System)**

NCSU-APHIS Plant Pest Forecasting System (NAPPFAST) is a web-based forecasting model using geo-referenced climatological weather data (Magarey et al. 2007). It was a commonly used pest forecast system by APHIS PPQ until 2014. The concept underlying NAPPFAST was initially established by the late Jack Bailey, professor at North Carolina State University (NCSU), and was also influenced by the work of Sequeira (Sequeira 1999, 2002a). Since 2002, the system was developed, maintained, and updated by APHIS in partnership with NCSU. NCSU in turn established sub-contractual agreements with a group called ZedX Inc. ZedX Inc. is an agricultural information technology company that manages weather data and programs web-interfaced modeling applications. NAPPFAST is originally designed to meet the needs of USDA APHIS Plant Protection and Quarantine; however, it is currently not used by APHIS PPQ due to funding and model functionality limitations

NAPPFAST was designed as a password protected website to support applications by USDA APHIS and its registered cooperators. Registered users were able to access the system through the Internet. NAPPFAST had three models: phenology, infection, and climate match. The NAPPFAST phenology model enabled users to specify base and maximum temperature and allowed inputting degree day ranges at specific stages. The NAPPFAST infection model allowed users to enter temperature parameters (minimum, maximum, and optimal), leaf wetness period (minimum and maximum), and required precipitation. The phenology model was mainly used for modeling arthropods and weeds, while the infection model was used for plant diseases. Climate match model was used for all

species because it only required climate condition criteria without inputting species biology information.

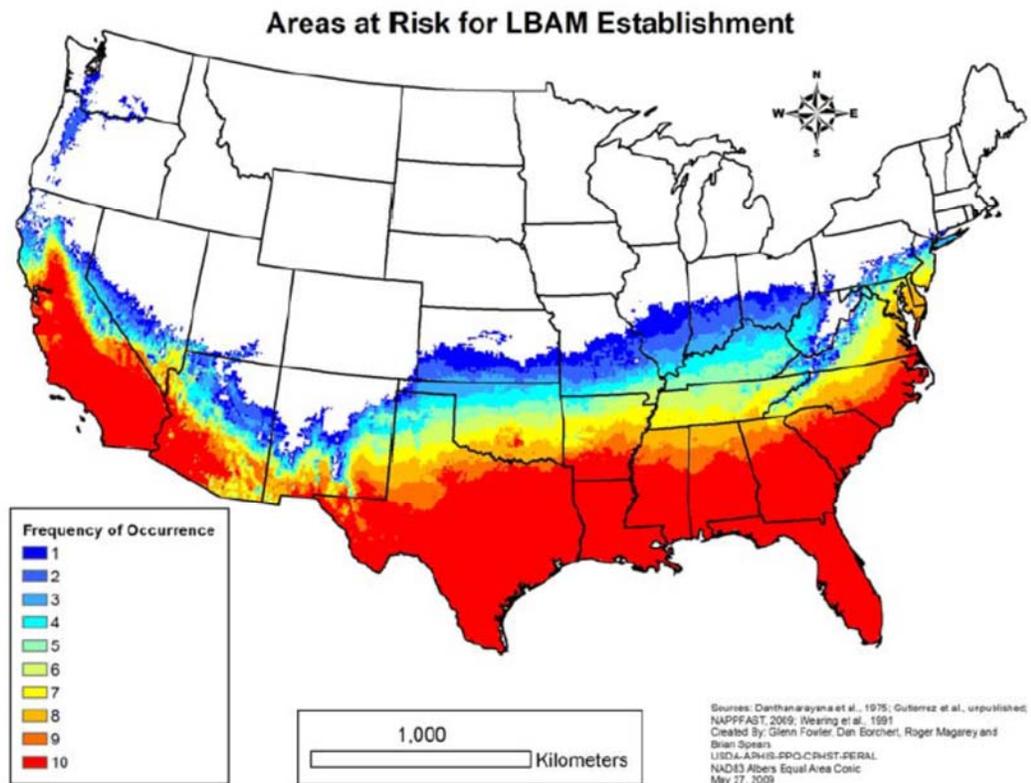
NAPPFASST could generate graphs and maps at North America and global scales. The resolution of the output map depended on the regions. It created 10-km resolution maps for North America, while it generated 32-km resolution maps for global areas (Magarey et al. 2007). Users were able to create two types of maps based on pest biology information; average history maps and probability maps. An average history map indicated the averages for the specific days user specified. For example, when a user selected degree day as a variable and set the date period as June 1 to June 30, the average history map displayed the average of degree days during June 1 to June 30 for the last 10 or 30 years. The probability map displayed the frequency of occurrence based on 10, 20, or 30 years of weather data. Frequency of occurrence could be classified into 5 or 10 classes.

The advantages of NAPPFASST were as follows. First, it was a web-based system, so users did not need to install software on individual computers. Users always had access to the system from any computer as long as there was Internet access. Second, the climate data were processed, maintained, and provided to the NAPPFASST system by ZedX, Inc. Collecting and processing weather data often requires considerable time, and it also must be updated regularly. The availability of global climate data at appropriate spatial and temporal scales for PPQ activities is a challenge, but ZedX provided adequate climate datasets at a global scale. Third, NAPPFASST included both insect and plant pathology models. Fourth, NAPPFASST outputs are already geo-referenced raster data and required no post-processing for further spatial analysis (i.e., ArcGIS, GRASS compatible formats). The main

disadvantage of NAPPFAST was that users had no access to the input climate data (“raw” data). The systems provided historical average and probability results based on the ZedX pre-processed archived climate data. For example, if the user wanted to examine the number of days that exceeded 30°C within a certain period of the time, the user would specify the number of days (e.g., 2 days, 3 days) but could only obtain the probability that a certain area would have temperatures greater than 30°C for certain days. The second disadvantage was that users did not have an access to the models proper. The users did not know how the degree-days were calculated or how the climate match model was evaluating the match index based on the user input criteria. The third disadvantage of NAPPFAST was the cost of the product (Berger and Sequeira 2015). It was only available through cooperative agreements and modification and additional tools and functions resulted in new cost structures that became unsustainable. Importantly, the results of some of its output were targeted in the scientific community (Gutierrez et al. 2010) as problematic and specific changes needed to address but the concerns were not considered within budget.

APHIS PPQ often used NAPPFAST models to predict insect phenology, potential number of generations per year, and environmental suitability for pest introduction and establishment (Fowler et al. 2009a, Fowler et al. 2009b, Newton et al. 2009, Magarey et al. 2015a). The timing of *Lymantria dispar* (Asian gypsy moth) adult emergence was estimated in China and Japan by using the NAPPFAST phenology model in order for PPQ to establish offshore trapping programs at Chinese and Japanese ports during adult emergence periods (Fowler et al. 2009b). The potential economic impact caused by LBAM was estimated by using the NAPPFAST phenology model to identify the areas where three or more generations

could occur in the continental United States (Fowler et al. 2009a). The map below (Figure 2-4) indicates the probability of three LBAM generations or more ( $\geq 2221$  degree-days) based on the sufficient yearly degree day accumulations at a lower threshold temperature of  $7.5^{\circ}\text{C}$  (Danthanarayana 1975, Fowler et al. 2009a). Since LBAM is not known to survive temperatures below  $-16^{\circ}\text{C}$ , the areas where the minimum temperature becomes colder than  $-16^{\circ}\text{C}$  at least one day during a year are excluded from potential establishment areas in Figure 2-4.



**Figure 2-4. Climate match analysis by NAPPFAST for areas at risk for LBAM establishment. The results are reported in terms of frequency of years from 1999 to 2008 where enough degree days accumulated for LBAM to complete  $\geq$  three generations and non-lethal minimum daily temperatures  $\geq -16^{\circ}\text{C}$  occurred (Fowler et al. 2009a).**

### **2.3.3 MaxEnt (VisTrail SAHM)**

The U.S. Geological Survey, Invasive Species Science team has developed Software for Assisted Habitat Modeling (SAHM), which includes MaxEnt. The purposes of SAHM are to expedite habitat modeling as well as to maintain records of the various input data, pre- and post-processing steps, and modeling options incorporated in the construction of a species distribution model (Talbert and Talbert 2012). The SAHM is a module developed within the VisTrails system.

The SAHM has five components, which are all processed automatically once the users set the variables. Those five components are 1) input data, 2) preprocessing data, 3) preliminary model analysis & decision, 4) models, and 5) outputs. The main advantages of the SAHM system include automated data preprocessing, correlation results among input variables, no special software installation to run SAHM, and averaged output by the number of simulations the users specify. The output also includes a standard error raster layer, which indicates the level of uncertainty.

The data preparation for any spatial analysis can be time-consuming. The one disadvantage of utilizing user-friendly software (e.g., ArcMap) to run a spatial analysis is that the users are often not aware of projection and raster resolution of the raw data. Naïve GIS users may not realize that conducting a spatial analysis with different projections and resolution sizes may result in inaccurate outputs. The SAHM preprocessing data component automatically corrects spatial properties for the spatial analysis. However, since this is an automated system and generates all the input variables in the same spatial extent, raster

resolution, and projection, this process may take several hours, especially if the users are running global data.

Estimating habitat suitability by MaxEnt can be subjective. Since MaxEnt allows users to input of any type of data, the users make decisions on what input variables to use based on their experience and knowledge. The SAHM package includes correlation analysis among all the input variables; therefore, the users are able to determine the similarities between some variables. Highly correlated variables ( $r > 0.8$  or  $r < -0.8$ ) are often deleted from the analysis (Jarnevich and Reynolds 2011). Selecting input variables can be subjective and may introduce judgment-subjective uncertainty. In addition, the MaxEnt outputs are often significantly different based on the input data (example in Figure 2-5). Users can run the model several times with different input variables until the output results seem reasonable. This judgment is based on the user's skill and knowledge, and this could be difficult if the users have limited knowledge, especially for exotic species that are not known well.

The habitat suitability for *Epiphyas postvittana* (Light Brown Apple Moth) was estimated by using the MaxEnt approach (Figure 2-5). The LBAM survey data from 2009 were used in this model with 19 Bioclim variables derived from 2009 climate data, which were created by Climate Source, Inc., Spatial Climate Analysis Service (SCAS) at Oregon State University. These 2-km resolution Bioclim data were developed by the USGS, Fort Collins Science Center. After correlation analysis, some variables were deleted due to high correlation rates with other data. Model 1 was developed by using the following input variables: annual mean temperature, minimum temperature of the coldest month, and annual

precipitation. Model 2 was developed by using the following input variables: annual mean temperature, iso-thermality, minimum temperature of the coldest month, mean temperature of the coldest quarter, annual precipitation, precipitation of the driest quarter, and precipitation of the warmest quarter.

Models 1 and 2 resulted in significantly different pest suitability. The result from Model 1, which was predicted by using 3 climate parameters, had a high probability of LBAM suitability for the entire south region. On the other hand, the result from Model 2 forecasted high suitability only in southern Texas and Florida. Comparing standard deviations revealed that Model 1 had an overall smaller standard deviation compared to Model 2. This may be because fewer parameters were used in the model, resulting in less variation when combining input data. In this exploratory study, only climatological data were used in the model. Both models were validated with survey points. The probabilities in the San Francisco area, where LBAM had been intensively surveyed, were from 0.5 to 0.7 in both models. Both models performed fairly well in southern California (Los Angeles and San Diego regions). Model 1 resulted in above 0.9 probabilities for the locations where LBAM was found, whereas Model 2 resulted in approximately 0.8 for LBAM positive locations.

MaxEnt generates suitability of pest establishment based on current distribution points as well as variance (e.g., standard deviation, error) associated with the output. This value indicates how well or poorly the model performed (uncertainty associated with model performance).

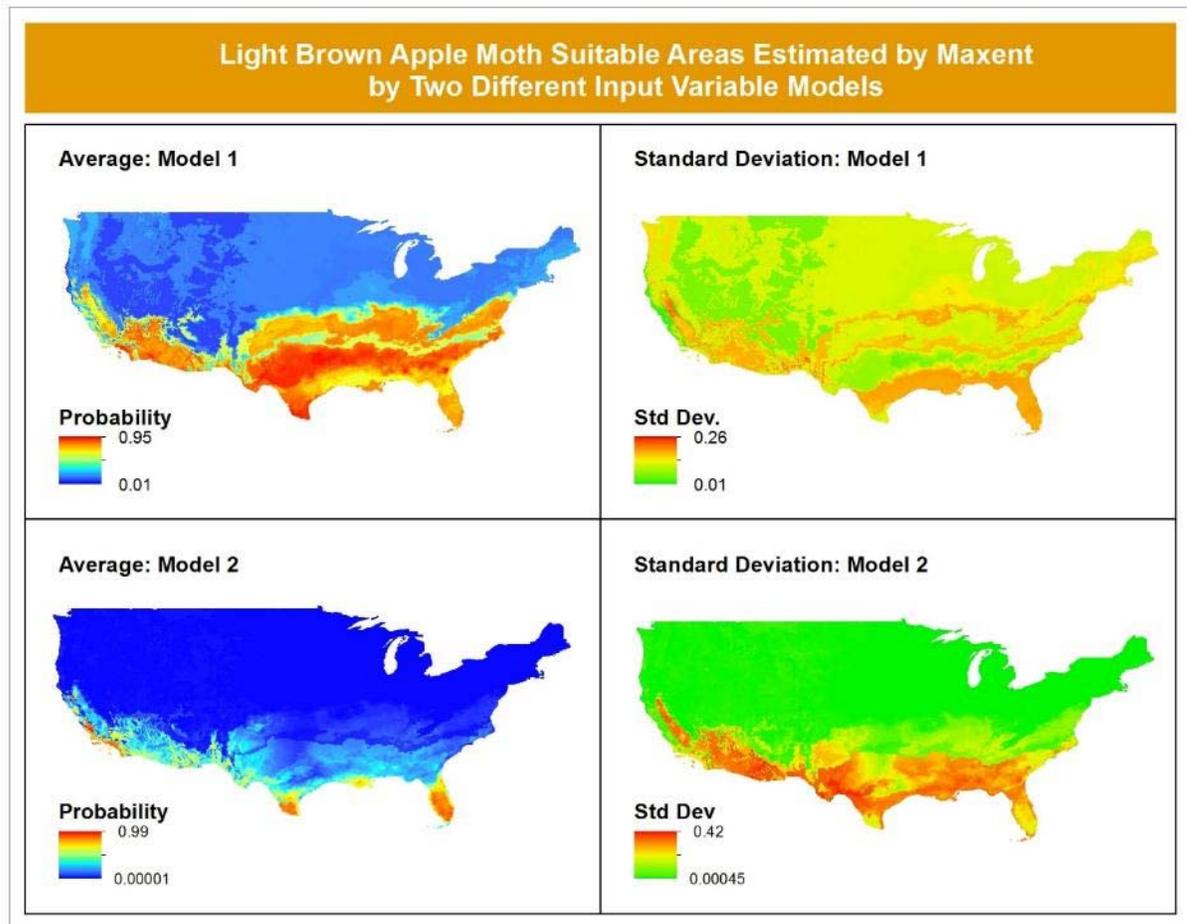


Figure 2-5. Light brown apple moth suitability predicted by using MaxEnt with different sets of input drivers

### 2.3.4 USPest

The website [www.USPest.org](http://www.USPest.org) was established to provide IPM pest and plant disease models and forecasting for agricultural, pest management, and plant biosecurity decision support in the United States (Coop 2016). USPest.org links US weather data from over 24,000 weather stations and models organism phenology with Degree-Day models. This project was funded by numerous USDA National Institute of Food and Agriculture (NIFA) grants, USDA Risk Management Agency (RMA) Pest Information Platform for Extension and Education (ipmPIPE) grants, National Plant Diagnostic Network (NPDN) grants,

Western IPM Center and Oregon Statewide IPM funds, and local and regional commodity grants. Degree-day (DD) models by USPest.org are hosted by the Integrated Plant Protection Center of Oregon State University.

The weather data used in USPest.org are derived from National Climate Data Center (NCDC) of National Oceanic and Atmospheric Administration (NOAA) and AgriMet by the Department of Interior, Bureau of Reclamation (Coop 2016). Online tools allow users to download weather data by region through the USPest.org website; however, USPest.org recommends that weather data should be obtained through original data sources such as NCDC and AgriMet (Coop 2016).

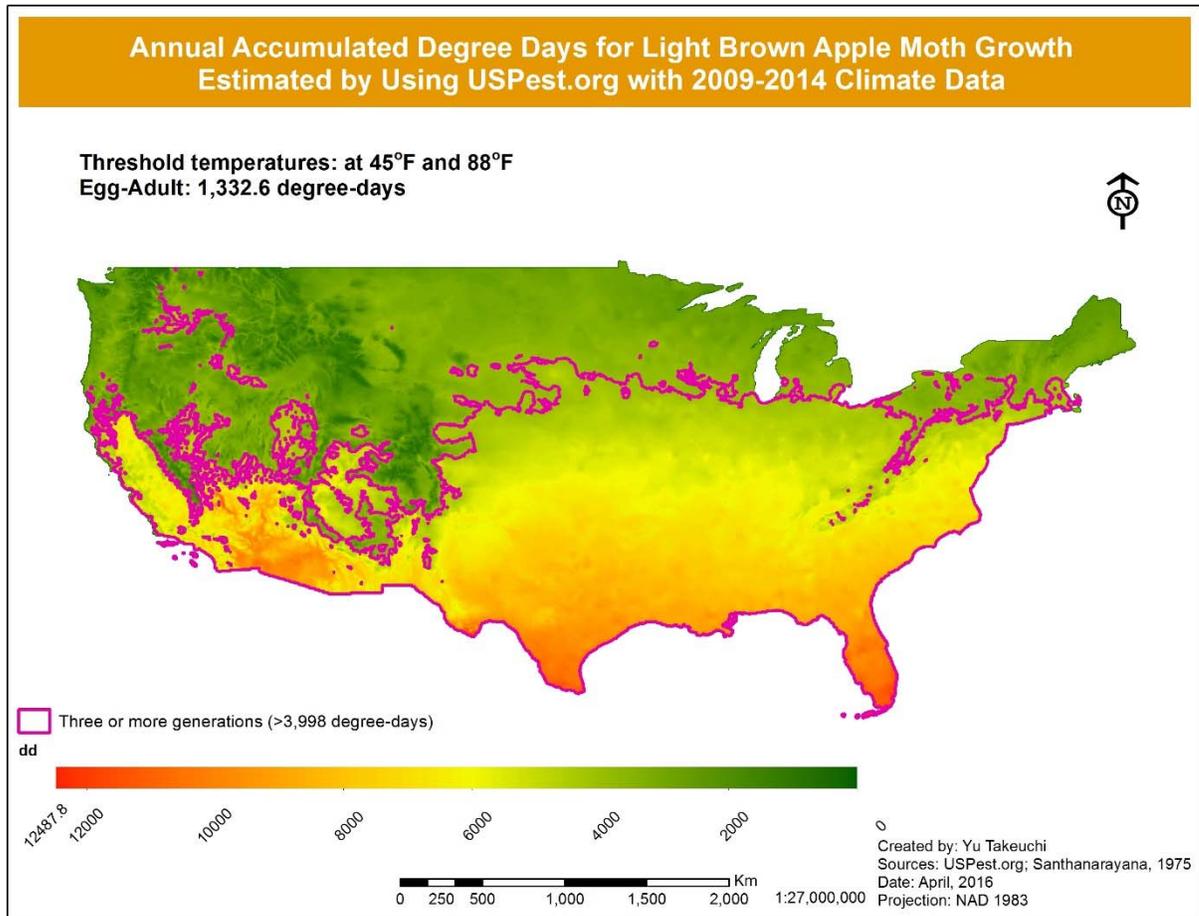
Currently, USPest.org hosts 101 plant, insect, and plant pathogen DD models (as of June, 2016). Users are able to review DD model parameters by species and to run the models online. The DD models can be run at a weather station location or for a seamless region (e.g., northwest states, western states, southeast states). If DD model parameters for specific pests are not available, users are able to supply lower and upper threshold temperatures. USPest offers multiple ways (e.g., single sine, double sine) to calculate degree-days for DD. It also offers different cut-off methods (e.g., horizontal and intermediate); however, these features are only available when using DD at the weather station level.

USPest.org is a good resource because it offers a database containing key life table information for agricultural pests, weather data, and tools to calculate degree-days. It works fine when the user is selecting a single weather station or small regional areas (e.g., NW 5 states). However, if a large area is selected, model run time is normally quite long, and it often does not return results (the system crashes). Also, DD models for seamless data require

lower and upper threshold temperature in Fahrenheit instead of Celsius. The users are able to export maps generated from USPest. The map contains accumulated degree-days during the user specified time period.

A phenology model for LBAM was run in USPest.org with climate data from 2009 to 2014. The original idea was to run the phenology model for 10-year climatology but reduced to 6 years because USPest.org server did not support to run for 10 years. I selected single sine with horizontal cutoff method to calculate accumulated degree-days. Although the lower threshold temperature for LBAM was estimated as 7.5°C (45.5°F) (Danthanarayana 1975), USPest.org only allowed users to select whole number in Fahrenheit for lower and upper threshold temperatures. Therefore, the lower threshold temperature for LBAM was set for 45°F. Also, upper threshold temperatures for LBAM were reported as 31.3°C for egg stage and 31.5°C for larval and pupae stages (Gutierrez et al. 2010). The upper threshold temperature for LBAM was set for 88°F in USPest.org.

Figure 2-6 indicates average annual accumulated degree days at lower and upper threshold temperature of 45°F and 88°F. The Southern US, on average, accumulated approximately 10,000 degree-days in a year. LBAM requires approximately 3998 degree-days (°F) to complete three generation (egg to egg) (Danthanarayana 1975, Fowler et al. 2009a). The areas delineated with pink lines in Figure 2-6 indicate areas could potentially have three or more generations per year based on accumulated degree-days.



**Figure 2-6. Annual accumulated degree days for LBAM calculated by USPest based on 2009-2014 climate data**

## 2.4 Model comparisons

Comparing models has been the subject of many efforts (e.g., the International Pest Risk Research Group) (Venette et al. 2010, Magarey et al. 2015b). In general, models are run with a specific target species. The outputs of the models are then compared with reference to known (“ground truth”) parameters. Thus, the reference points might be the known distributions of a pest. Models might then be compared with reference to how well they capture the known distributions. The model that best approximates what is known is

generally regarded as “better”. This is an obvious simplification that focuses on exercises with pest forecasting models associated with phytosanitary applications. The topic of comparing models is covered extensively in the peer-reviewed literature.

To avoid attribution of performance (better or worse) to a given model, one must isolate effects associated with data drivers and parameter estimates from the effects of the model proper. This “blocking” of effects is often ignored and the differences observed in the output cannot be attributed or traced back to the source. Attributing differences to the model specifically becomes difficult unless one approaches model comparisons by holding all drivers (e.g., weather drivers and biological parameters) constant. Thus, the comparison here attempts to identify differences in approaches and attempt model contrasts that hold drivers and parameters constant to the extent possible.

#### **2.4.1 Input variables**

In the previous section, four forecast systems were reviewed. Each system has some advantages and disadvantages. Before comparing outputs derived from those four models used, input variables for each system are reviewed.

All models examined, except MaxEnt, in the previous section require weather data as input variables. Although MaxEnt can be run without supplying weather data, weather data is commonly used to predict insect suitability because weather conditions are primary drivers for insect growth. However, the formats of the weather data required are different. Users can supply their own climate data and update to more recent historical weather data for CLIMEX, MaxEnt, and BioSim models. Users are unable to update and modify weather data on the USPEst.org or in the NAPPPFAST systems.

#### **2.4.1.1 CLIMEX**

CLIMEX stores weekly climate data with location ID number. CLIMEX default weather data are at weather stations and contains approximately 3000 locations covering global space. Alternative data source is Global Climatologies for Bioclimatic Modeling (CliMond) (<https://www.climond.org/>). CliMond provides weekly climate data for CLIMEX, which were normal based on 1960 to 1999. CliMond CLIMEX datasets come with two different raster sizes, 10Km and 30Km. Although CliMond data are formatted as raster layer, CLIMEX only recognizes climate data with location ID number. Therefore, raster formatted climate data are treated as point dataset by assigning a unique location ID number to each CliMond data cell. Users can also supply and update input weather data by using MetManager tool in CLIMEX; however, processing weather data into CLIMEX format is a time-consuming process and must be periodically updated to reflect current weather data.

#### **2.4.1.2 NAPPFASST**

NAPPFASST maintained and updated weather data periodically by Zed-X, Inc. The Zed-X, Inc. obtained global weather data from various datasets from NOAA including Real-Time Mesoscale Analysis, and Climate Forecast System Reanalysis by National Center for Environmental Prediction to generate seamless global scale daily weather data (Magarey et al. 2015a). Zed-X Inc. provided 10-kilometer raster format for United States and 38-kilometer raster data for global through NAPPFASST (Magarey et al. 2015a). Users did not have access to weather data to view climate data in NAPPFASST system.

### **2.4.1.3 MaxEnt VisTrail SAHM**

The MaxEnt model can be run without climate data; however, analysts commonly use climate data to estimate suitability of a pest in MaxEnt because climate is the primary element in pest growth. Therefore, MaxEnt does not have specific format requirements for climate data. The most commonly used climate data for MaxEnt are Bioclim data. There are 19 core Bioclim variables, such as annual temperature, mean diurnal temperature range, maximum temperature of warmest week, and minimum temperature of coldest week (Hijmans et al. 2005, Kriticos et al. 2012, O'Donnell and Ignizio 2012, Kriticos et al. 2014) (Appendix D), but currently there are 40 known Bioclim variables. The Bioclim 19 data can be downloaded from WorldClim ([www.worldclim.org](http://www.worldclim.org)). The WorldClim generated 19 Bioclim variables at 4 different spatial resolutions (30-seconds, 2.5 arc-minutes, 5 arc-minutes, and 10 arc-minutes) based on current climatology (1950-2000) and CMIP5 general circulation models (Hijmans et al. 2005). The CliMond dataset also provides Bioclim 40 variables generated by averaged historical climate data (1961-1990) and two general circulation models (CSIRO and MIROC-H) for two greenhouse gas emission scenarios (A1B and A2) at 10 arc-minutes (approximately 18.6 kilometer) and 30 arc-minutes (approximately 50) resolutions (Kriticos et al. 2012, Kriticos et al. 2014).

### **2.4.1.4 USPest**

USPest uses near-real time temperature data provided from AgriMet, HydroMet, National Weather Service, Interagency Remote Automatic Weather Stations, Snotel, and others in addition to historical average climate data (clim84: 1971-200 Normals) and monthly maximum and minimum temperature from PRISM (Coop 2016). PRISM Climate Group

belongs to Oregon State University, Northwest Alliance for Computational Science and Engineering. PRISM Climate Group creates daily and 30-year annual and monthly normal of average temperature, minimum temperature, maximum temperature, and precipitation for Contiguous United States. PRISM 30-year normals are average monthly and annual conditions from 1981 to 2010 at 4-kilometer and 800-meter resolution formats. PRISM also provides daily data since 1981 in ascii and BIL format. PRISM updates climate data daily and makes near-real time data available to the public, although they do spend another 6 months to do quality control and adjust outputs by adding new data that are not available at near-real time.

#### **2.4.2 Outputs, post processes, and additional analysis**

The suitability for light brown apple moth (*Epiphyas postvittana* Walker) was predicted using 4 models (CLIMEX Compare Location model, NAPPFAST Degree-Day model, MaxEnt, and USPest). Results from those models may be required for post-processes for further spatial analysis or statistical analysis. In this section, format of the output as well as required processing to make the output into a standard raster format are discussed.

##### **2.4.2.1 CLIMEX**

The CLIMEX compare location model expresses overall suitability of locations by values called the Ecoclimatic Index (EI) and generates stress indices (e.g., cold stress, hot stress, dry stress, wet stress). The stress index outputs are based on pest biology information that a user supplies in the model. Each location has an EI value with associated stress values; therefore, users are able to examine limiting factors for pest growth for specific locations.

Outputs from all CLIMEX models are exported in tabular format regardless of input climate data (e.g., point, raster). It allows users to choose the variables to be included in the output table. Those variables include latitude, longitude, altitude, EI, stress indices, growth index, and number of weeks in a year that had a positive growth index.

CliMond CLIMEX datasets include fishnet shapefiles to represent raster cells. With location ID, the tabular output values can be linked to convert to raster layers. There is a large number of cells within the global CliMond dataset, especially for 10' resolution data. Therefore, it often takes a long time (several hours) to link, process and convert the data into raster format. Since all outputs from CLIMEX include latitude and longitude (centroid for a cell), results can be converted to points then interpolated across points instead of linking to the fishnet shapefile to convert to a raster layer.

#### **2.4.2.2 NAPPFAST**

The NAPPFAST models generated probabilities or frequencies of the specific value that a user specified in the models. For example, in LBAM analysis, Fowler et al. (2009a) were trying to identify the areas where 3+ generations of LBAM could be possible based on accumulated physiological time. Since LBAM is known to require approximately 740 degree-days to complete one generation (Danthanarayana 1975), The NAPPFAST model examined the areas that could accumulate greater than 2,221 (740.3 x 3) degree-days based on historical climate data and calculated the frequency (=probability, number of years that met this condition out of 10 or 20 or 30 years).

Outputs from NAPPFAST are zip files containing a tiff raster layer. The raster layer is georeferenced to the WGS 1984 projection. Thus there is no post processing procedure required with NAPPFAST output to be converted into a georeferenced dataset.

#### **2.4.2.3 MaxEnt in VisTrail SAHM**

The MaxEnt model generated habitat suitability and standard deviation associated with the results in raster format. It provides an ascii file, expressing probability of presence with values between 0 and 1 (Phillips 2016). Since overall probability of presence is combined results from all the iterations, MaxEnt summarizes the model performance for overall and each run. The summary report is also part of the outputs from MaxEnt.

#### **2.4.2.4 USPest.org**

The USPest.org degree-day model generates accumulated degree days for the user specified time period calculated with the lower and upper threshold temperatures that the user specifies. The output results can be downloaded as ascii formatted files; therefore, it does require conversion into raster format from ascii.

Since the results are expressed in accumulated degree-days, the users must translate degree-days into suitability based on the criteria. If a risk analyst defines the risk areas based on three or greater generations of LBAM annually, then the areas that have accumulated degree days greater than 2,221 are defined as suitable.

### **2.4.3 Output Comparison**

CLIMEX, NAPPFAST, and USPest use pest biology information, while MaxEnt uses presence data to predict pest suitability. MaxEnt is a machine learning technique to predict potential habitat niche based on the environmental factors and pest presence locations.

Although both NAPPFAST and USPest use phenology models, the outputs from those two systems were different. NAPPFAST provided probability or frequency based on the user inputs for output. If a user wanted to estimate the timing of adult insect emerging at specific time period, a user needed to supply required degree days for adult emergence and the specific time period. In this case, the NAPPFAST output would be the probability of adult emergence during that specified time period. NAPPFAST was able to generate the frequency of the specific event happening during a certain time periods based on the last 10, 20, or 30 years.

USPest generated accumulated degree-days for the pest. The potential number of generations in a year can be calculated by accumulating degree-days each year and dividing by the required degree-days for one generation. The specific pest event timing can be estimated by accumulating degree-days at a daily or weekly intervals to examine when the required degree-days would be accumulated from a specific date.

MaxEnt generated an estimate between 0 and 1 for the probability of presence, while CLIMEX generated an index (Ecoclimatic Index: EI) between 0 and 100, representing pest establishment. Any positive EI values have some potential of pest establishment although it is known that areas with EI values greater than 30 are considered highly suitable.

The environmental suitability of *Epiphyas postvittana* (light brown apple moth: LBAM) was examined by using four different systems (CLIMEX, NAPPFAST, MaxEnt, and USPest). Outputs from MaxEnt were significantly different based on the selection of input variables; therefore, the users must select input variables based on variable contribution to MaxEnt model and importance to pest biology. MaxEnt Model 1 predicted that Southern US

was highly suitable, while MaxEnt Model 2 predicted that Florida, southern Texas, and coastal area of California were highly suitable. Both Model 1 and 2 predicted that Northern US were unlikely to be suitable for LBAM growth with a low standard deviation. Two CLIMEX models (Lozier and Mills 2011, He et al. 2012) predicted environmental suitability of LBAM on a global scale, but the results were slightly different. The model parameterized by Lozier & Mills (2011) indicated that east Florida could be suitable area for LBAM, while the model by He et al. (2012) resulted in Florida as unlikely to be suitable for LBAM. Both models predicted that coastal areas of California (where LBAM was established) and coastal areas of the Carolinas (Lozier and Mills 2011, He et al. 2012) were highly suitable.

NAPPPFAST and USPest.org generated accumulated degree-days for LBAM growth. Since the output from NAPPPFAST indicated the likelihood of LBAM having 3 or more generations per year, the USPest result was also converted to indicate the potential areas that could have three or more generations based on accumulated degree-days. The USPest predicted larger areas than NAPPPFAST, but this was probably due to the NAPPPFAST having an additional cold mask (areas where daily minimum temperature becomes colder than -16°C at least a day during a year) applied to the degree-day results.

Outputs from those models were not easily compared because all models used climate data in different formats and from different time periods. Therefore, it was impossible to determine whether the differences were due to differences in climate data or model performance.

## 2.5 Uncertainty

“Essentially, all models are wrong, but some are useful” (Box and Draper 1987). One tries to model precisely with all available information, sophisticated computation methods, and a high performance computer, but there are still uncertainties associated with predictions. Unfortunately, this is not avoidable for reasons explored here. The important things are to understand errors associated with models and outputs and to reduce uncertainty. There are many scientists who have studied uncertainties associated with predictions and have examined how to deal with them (Elith et al. 2002, Regan et al. 2002, Regan et al. 2005, Koch et al. 2009).

There are several types of uncertainty. Uncertainty can be derived from data collection, forecast model, unknown fact, and so on. Elith et al. (2002) and Regan et al. (2002) identified types and sources of uncertainty and ways to deal with the uncertainties. Uncertainty was classified into two categories: epistemic uncertainty and linguistic uncertainty (Elith et al. 2002, Regan et al. 2002). Epistemic uncertainty has six types: measurement error, systematic error, natural variation, inherent randomness, model uncertainty, and subjective judgment (Elith et al. 2002, Regan et al. 2002). Linguistic uncertainty has five types: vagueness, context dependence, ambiguity, indeterminacy of theoretical terms, and underspecificity (Elith et al. 2002, Regan et al. 2002). In this section, each type of uncertainty is reviewed and summarized with respect to how it could be related with the proposed integrated risk modeling framework.

## **2.5.1 Epistemic uncertainty**

### **2.5.1.1 Measurement error**

Measurement error is due to imperfections in measuring equipment and observational techniques (Elith et al. 2002, Regan et al. 2002). In some cases, spatial analysts may not have control of field experiments or laboratory observations to understand this type of uncertainty. However, the data that are often used in pest predictive models may have inherited measurement errors. Thus pest simulations and forecasts are not immune from measurement error. For example, weather station data from NOAA may have measurement errors due to inaccurate thermometers and misreporting of the data. Regan et al. (2002) suggested that this type of uncertainty could be treated by statistical techniques using multiple measurements or by reporting the measurement with bounds (confidence intervals). Data obtained from multiple sources should be compared, and confidence intervals should be created around the data if possible.

The proposed framework should be collected from as many climate data sources available to accommodate the errors associated with climate data. Multiple climate data sources make it possible to create the bounds and understand the variabilities that are associated with climate data sources and processing methods (e.g., interpolation). In addition, pest biological information should be collected from as many sources as possible. Pest species develop adaptations to different environments and would not behave in exactly the same way even when the same species are inhabiting in the same environment. In many cases, a single pest may have differing subpopulations at different locations and different

biological requirements, growth rates, and mortality rates (e.g., gypsy moth life cycle studies by Sheehan (1992), Russo et al. (1993), Johnson et al. (1983), and Carter (1992)).

### **2.5.1.2 Systematic error**

Systematic error is due to bias in the measuring equipment or the sampling procedure (Elith et al. 2002, Regan et al. 2002). This type of error can also occur due to categorizing data into classes (e.g., frequencies at a certain scale) (Elith et al. 2002, Regan et al. 2002). Scale becomes an important and a key issue in ecological modeling because analyzing biological features should be done at an appropriate scale. An appropriate scale is one where the input data have sufficient data points to represent a given scale. Elith et al. (2002) noted that errors in digital elevation models (DEMs) might be small at a global scale but large at a local scale. The only way to reduce this type of error is to recognize a bias in the experimental procedure (Regan et al. 2002).

In a pest forecast framework, especially for National Plant Protection Organizations, the analysis needs be performed at both global and local scales. The reason is that different management and/or regulatory actions need be taken at different scales when dealing with invasive species. Ideally, the appropriate scale is clearly defined based on the objectives; however, local scales in the framework can be only defined based on the availability of data. Information becomes extensive, expensive and hard to obtain at certain scales and certain resolutions. Reasonable data scales and resolutions (that balance availability, desired detail and practicality) need to be identified at the beginning based on the availability of data, among other reasons, so that adequate analysis is done to reduce systemic error caused by scale. However, as noted I recognize that I have to strike a balance between my desire to be

comprehensive and maximize the resolution of the data (very fine grain resolution) and the need to have manageable data sets that can be used practically in management and regulatory settings.

### **2.5.1.3 Natural variation**

Natural variation occurs in systems that change with respect to time, space, and other variables, which are difficult to predict (Elith et al. 2002, Regan et al. 2002). This type of uncertainty occurs because it is not possible to build error-free models (Elith et al. 2002). Although Elith et al. (2002) suggested a way to deal with natural variation errors (creating confidence intervals around the quantity in question), this is unlikely to be addressed directly in the proposed framework. The reasons are: 1) it is unavoidable and cannot be adequately modeled based on knowledge, and 2) intervals are only appropriate for numerical uncertainty (Regan et al. 2002).

Uncertainty associated with natural variation should be addressed by understanding what the likely sources of natural variation are and providing descriptive analyses of this kind of uncertainty. More specifically, I will aim to discuss what I know to be important and what I know to constitute gaps in my current knowledge as respect to the issue of natural variation (“known knowns vs. known unknowns”).

### **2.5.1.4 Inherent randomness**

Inherent randomness is mentioned in the paper by Regan et al. (2002), where they state that this type of uncertainty is unlikely to occur in ecology and conservation biology. The uncertainty in ecology and conservation biology probably occurs due to incomplete information, not inherent randomness (Regan et al., 2002).

#### **2.5.1.5 Model uncertainty**

Model uncertainty occurs due to the necessary process of simplifying the representation of biological and ecological processes (Elith et al. 2002, Regan et al. 2002). This type of uncertainty arises because only variables and processes that are relevant and prominent are used in the model (Regan et al. 2002). Some variables might indirectly influence the outcome; however, those variables would not be included in the model if they are not fully understood. Also, constructing models can cause uncertainty. Ecological processes are not mathematical in nature, yet mathematical models are used to represent those processes (Regan et al. 2002). Selecting the right form of mathematical model is subjective, and it can lead to model uncertainty.

This type of uncertainty is one of the most important factors that need to be managed within the framework. Model uncertainty cannot be eliminated; however, it can be evaluated by several different methods.

#### **2.5.1.6 Subjective judgment**

Subjective judgment uncertainty is caused by data interpretation (Elith et al. 2002, Regan et al. 2002). Data are often judged or even developed based on expert opinions. This kind of uncertainty can be treated by applying a degree of belief about an event in the form of subjective probability although degrees of belief themselves are often uncertain (Regan et al. 2002).

Subjective judgment is unavoidable, but it is reducible. The data and outcome are interpreted based on knowledge and experience. Examining and understanding input data

and biology of the insects would increase the accuracy of interpretation and reduce uncertainty caused by subjective judgment.

## **2.5.2 Linguistic uncertainty**

### **2.5.2.1 Vagueness**

Vagueness is a type of linguistic uncertainty that is caused by the very nature of languages (Elith et al. 2002, Regan et al. 2002). Biology and its processes many times cannot be perfectly defined by words, causing vagueness uncertainty. Defining vocabularies clearly is the first step to reducing vagueness; however, in some cases, it would not completely reduce it. For example, when classifying U.S. ports by arrival of Asian gypsy moth infested ships, the ports with 5 or more infested ships arrived are categorized as high risk, while the ports with 1 to 4 infested ships are categorized as medium risk. The ports with 4 ships are very differently categorized from the ports with 5 ships when in fact there may not be a significant difference between them. Yet the language used suggests that there may be more information in the description than may be sometimes justified by the actual values.

Vagueness uncertainty has been very problematic in plant protection organizations (Griffin 2016). The word “invasive species” has become so popular now that national plant protection organizations have slightly different definitions of “invasive species” (Griffin 2016). In addition, “invasive species”, “alien species”, “invasive alien species”, “quarantine pests”, and “regulated pests” are terms used daily in PPQ, but they are treated as interchangeable such that risk analysts as well as decision makers no longer recognize the distinctness of their precise definitions (Griffin 2016). The key words and vocabularies should be first well defined, and linguistic vagueness should be minimized by awareness and

communication efforts. Also, when categorizing data into classes, each class should be significantly different from other classes and the language used should map appropriately to the values that give rise to the linguistic expression. However, since vagueness is created because of language, I acknowledge that it is almost impossible to completely eliminate it. Scientists have been studying how to deal with such uncertainty. Regan et al. (2002) mentioned several techniques (e.g., supervaluations, fuzzy logic and fuzzy set theory, intuitionistic logic, three-value logic, paraconsistent logic, model logic, and rough sets) and stated that the supervaluational approach was the most appropriate technique to deal with this uncertainty.

### **2.5.2.2 Context Dependence and Ambiguity**

Context dependence uncertainty is the result of not specifying the context in which a proposition is to be understood (Regan et al. 2002). Ambiguity is the result of some words having more than one meaning (Elith et al. 2002, Regan et al. 2002). The only way to deal with context-dependent uncertainty is to specify context (Regan et al. 2002).

The context and words should be defined clearly so that users will not misunderstand the context represented as the outcome. The model results needs to be clearly displayed and explained in order to deliver clear messages on the model outputs.

### **2.5.2.3 Underspecificity**

Underspecificity happens when there is unwanted generality in data (Elith et al. 2002, Regan et al. 2002). Regan et al. (2002) and Elith et al. (2002) give an example of underspecificity as historical data that were often too general (e.g., species distribution specified at region and city level) but recent data were very specific (e.g., GPS recording for

species distribution). Many of the data managed by USDA APHIS are only be recorded at general scales. Port locations can already be ambiguous (e.g., Port Environs), and Asian gypsy moth true distributions around foreign ports remain unclear or can only be estimated based on the information observed from small-scale maps. Also, the New Pest Advisory Group database as well as the North American Non-Indigenous Arthropods Database only record states as first detection locations.

Regan et al. (2002) state that the best way to deal with underspecificity is to provide the narrowest possible bounds on estimates given the data and make available all of the information behind such statements. Risk analysts should make the best judgment based on their knowledge to narrow information if possible. However, analysis within the framework should be able to accommodate all potential events.

#### **2.5.2.4 Indeterminacy of theoretical terms**

Indeterminacy uncertainty of theoretical terms is due to indeterminacies in the theoretical terms themselves (Regan et al. 2002). Some of our theoretical terms have the potential for ambiguity even though the terms are not ambiguous at the current stage (Regan et al. 2002). Regan et al. (2002) state that theoretical indeterminacy should be dealt with by making conscious decisions about the future usage, but unfortunately it cannot be dealt with in the present. This can be only done when new information or usage of the word comes into effect (Regan et al. 2002). An approach to this problem is the continuous review of the framework and associated theoretical terms to ensure that currency is maintained as terminology evolves. I hope to be able to be associated with this area in the future and will

strive to maintain currency by making periodic evaluation of terminology and theoretical terms part of a continuous improvement cycle.

### **2.5.3 Evaluation of Uncertainty**

All predictions have uncertainty. Because uncertainty is a key characteristic of any prediction, any forecast needs to develop a strategy to manage and communicate this characteristic. In fact, I argue that uncertainty is the “currency” and *raison d’etre* for those that engage pest forecasts and risk analyses.

A model itself includes uncertainty at several levels because ecological processes cannot be perfectly modeled by using theories and mathematics. Also, data (e.g., input data) involve uncertainty. There are uncertainties associated with collecting data and processing data. Importantly, many times users are not aware of such uncertainties. Elith et al. (2002) reported that prediction of habitat suitability involved the following uncertainties: 1) model uncertainty arising from incomplete systems, 2) data uncertainty arising from measurement error, systematic error, model uncertainty, and subjective judgment, 3) natural variation, and 4) under-specificity in historical records.

There are two main strategies to address uncertainties in pest predictions. One is to reduce uncertainty, and the other is to quantify and display uncertainty.

There are a number of ways to reduce uncertainty associated with pest prediction. Since uncertainty is mainly caused by lack of scientific data and information, additional knowledge can be obtained through expert opinions (Drescher et al. 2013). The other way to reduce uncertainty is to adopt an ensemble modeling approach (Stohlgren et al. 2010). Ensemble modeling adds robustness and consistency of model performance by neutralizing

the errors across the study areas (Tuv et al. 2009, Stohlgren et al. 2010). Another key method is to use a pseudo-absence distribution approach (Senay et al. 2013). Senay et al. (2013) tested four methods to generate pseudo-absence distribution points and concluded that the model performance was improved when pseudo-absence points were generated based on geographic extent, environmental factors, and k-means clustering.

While reducing uncertainty to make a prediction more accurate is important, communicating uncertainty is as important as reducing it. Elith et al. (2002) summarized several techniques to quantify uncertainty in prediction. One way to quantify is to estimate confidence intervals around predictions. Another method is to use Bayesian Model Averaging (BMA) by averaging over a number of plausible computing models. A third method is to use sensitivity analyses that allow users to quantify the uncertainties in species data and explanatory variables to explore the extent of their effect on predictions.

Displaying outcomes in GIS has become popular in the past two decades as noted by Elith et al., (2002), and a map is one of the best ways to represent uncertainty. Elith et al. (2002) introduced two classes of map visualization: static maps and dynamic maps. Predictions with upper and lower bounds can be visualized in static maps (Elith et al. 2002). Elith et al. (2002) also indicated that categorizing uncertainty into groups in the static maps may be useful. Another method to visualize uncertainty is by using dynamic maps. Dynamic maps can be animated by combining several static maps. Or each run of the simulations can be expressed as a dynamic map. Yemshanov et al. (2015) demonstrated two approaches (mean-variance frontier concept and second-degree stochastic dominance rule) to evaluate uncertainties that could be incorporated into pest forecast results. These approaches

could provide additional guidance to decision makers to prioritize pest control options (Yemshanov et al. 2015).

In order to reduce uncertainty and to be consistent throughout with the International Panel on Climate Change (IPCC) working group, IPCC defined clear guidance on how to report and treat uncertainty in their studies (IPCC 2010). This guidance recommends the usage of terminology to reduce linguistic uncertainty and provides specific uncertainty levels associated with commonly used terms to express uncertainties.

## **2.6 Summary and Discussion**

Commonly used pest forecast models were reviewed in this chapter. I also critically reviewed four models/tools that USDA APHIS PPQ has used previously to illustrate the similarities and differences among those models. The major problem was that I could not compare those four models or could not state which models performed the best (or worst) because I did not know if the differences in outputs originated due to the models or input data drivers or lack of LBAM biological information. Previously, USDA APHIS has not evaluated uncertainty associated with pest predictions. Some levels of uncertainty in pest predictions should be reported so that decision makers have insight into the confidence in the predictions.

For analysts and risk assessors, all forecast models involve uncertainty (Yemshanov et al. 2015). As I reviewed above, there are multiple sources of uncertainty associated with pest model outputs. Although there are some uncertainties that we cannot control or cannot correct, uncertainty within the model output should be traced back as much as possible so

that limitations with the results can be communicated better and can be corrected by obtaining more information.

I first developed and then proposed to USDA APHIS to adopt a “framework” approach (a plug-in approach where different models can be hosted simultaneously by a consistent modeling environment) to develop transparent and consistent pest forecast systems while at the same time using the framework approach to allow for the continuity of scientific progress in pest forecasting science. The framework can provide ready-to-use pest models with minimum user inputs because commonly used data drivers are already formatted and plugged into pest models and can easily expand additional models and data within the framework. It also provides means to store and share pest models and data. For example, a GIS-based model for wheat karnal bunt (*Tilletia indica*) was developed on a desktop computer environment in the late 1990’s (Sequeira 1999); however, the model was not distributed to others or modified for other species. This is probably because a spatial model with large-size associated input data (e.g., nationwide climate data) was not easily shared.

The idea of an analytic pest forecasting framework for GIS-based plant pest models was insinuated when the wheat karnal bunt model was developed (Sequeira 1999) but was not explored further. As the technology and availability of data have advanced in recent years, the needs and requirements for a spatial analytic framework for pest forecast models and system also changed. In the next chapter, the proposed framework is explained in detail.

## 2.7 References

- Adjemian, J. C. Z., E. H. Girvetz, L. Beckett, and J. E. Foley. 2006. Analysis of Genetic Algorithm for Rule-Set Production (GARP) modeling approach for predicting distributions of fleas implicated as vectors of plague, *Yersinia pestis*, in California. *Journal of Medical Entomology* **43**:93-103.
- Allee, W. C. 1931. *Animal aggregations: a study in general sociology*. The University of Chicago Press, Chicago, IL.
- Allen, J. C. 1976. A modified sine wave method for calculating degree-days. *Environmental Entomology* **5**:388-396.
- Allstadt, A. J., A. M. Liebhold, D. M. Johnson, R. E. Davis, and K. J. Haynes. 2015. Temporal variation in the synchrony of weather and its consequences for spatiotemporal population dynamics. *Ecology* **96**:2935-2946.
- Areal, F. J., J. Touza, A. MacLeod, K. Dehnen-Schmutz, C. Perrings, M. G. Palmieri, and N. J. Spence. 2008. Integrating drivers influencing the detection of plant pests carried in the international cut flower trade. *Journal of Environmental Management* **89**:300-307.
- Aukema, J. E., D. G. McCullough, B. Von Holle, A. M. Liebhold, K. Britton, and S. J. Frankel. 2010. Historical accumulation of nonindigenous forest pests in the continental United States. *Bioscience* **60**:886-897.
- Baskervi, G. I., and P. Emin. 1969. Rapid estimation of heat accumulation from maximum and minimum temperatures. *Ecology* **50**:514-&.
- Baskerville, G. L., and P. Emin. 1969. Rapid estimation of heat accumulation from maximum and minimum temperatures. *Ecology* **50**:514-&.
- Berger, P., and R. Sequeira. (2015). NAPPFAST. Personal communication to on 11/4/2015, from Meeting.
- Bertoni, D. 2006. Recent infestations and continued vulnerabilities at ports of entry place U.S. forests at risk. Government Accountability Office, Washington, DC.
- Blackburn, J. K. 2010. Integrating geographic information systems and ecological niche modeling into disease ecology: a case study of *Bacillus anthracis* in the United States and Mexico. Pages 59-88 in K. P. O'Connell, E. W. Skowronski, A. Sulakvelidze, and L. Bakanidze, editors. *Emerging and Endemic Pathogens*. Springer, Dordrecht, The Netherlands.
- Bowser, M. L., and J. M. Morton. 2008. Monitoring and modeling terrestrial arthropod diversity on the Kenai National Wildlife Refuge. Pages 1-23 in 2008 Forest Inventory

- and Analysis (FIA) Symposium. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Parc City, UT.
- Box, G. E. P., and N. R. Draper. 1987. Empirical model-building and response surfaces. John Wiley & Sons, Inc, New York, NY, USA.
- Breiman, L. 2001. Random forests. *Machine Learning* **45**:5-32.
- Bryant, S. R., C. D. Thomas, and J. S. Bale. 1997. Nettle-feeding nymphalid butterflies: temperature, development and distribution. *Ecological Entomology* **22**:390-398.
- Burke, G. 2008. Officials suspend Calif. aerial spraying program. *The Washington Post*. The Washington Post.
- Carter, M. R., F. W. Ravlin, and M. L. McManus. 1992. Effect of defoliation on gypsy moth phenology and capture of male moths in pheromone-baited traps. *Environmental Entomology* **21**:1308-1318.
- Clark, S. 2015. Court hears appeal for light brown apple moth environmental review. *Santa Cruz Sentinel*. Santa Cruz Sentinel, Scotts Valley, CA.
- Cook, A. R., G. J. Gibson, T. R. Gottwald, and C. A. Gilligan. 2008. Constructing the effect of alternative intervention strategies on historic epidemics. *Journal of the Royal Society Interface* **5**:1203-1213.
- Cook, D. C., L. R. Carrasco, D. R. Paini, and R. W. Fraser. 2011. Estimating the social welfare effects of New Zealand apple imports. *Australian Journal of Agricultural and Resource Economics* **55**:599-620.
- Coop, L. 2016. USPEST.ORG. Last accessed February, 2016, <http://www.uspest.org/>.
- Courchamp, F., T. Clutton-Brock, and B. Grenfell. 1999. Inverse density dependence and the Allee effect. *Trends in Ecology & Evolution* **14**:405-410.
- Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, and K. T. Hess. 2007. Random forests for classification in ecology. *Ecology* **88**:2783-2792.
- Danthanarayana, W. 1975. Bionomics, distribution and host range of light brown apple moth, *Epiphyas postvittana* (Walk) (Tortricidae). *Australian Journal of Zoology* **23**:419-437.
- Davis, A. J., L. S. Jenkinson, J. H. Lawton, B. Shorrocks, and S. Wood. 1998. Making mistakes when predicting shifts in species range in response to global warming. *Nature* **391**:783-786.

- De Meyer, M., M. P. Robertson, A. T. Peterson, and M. W. Mansell. 2008. Ecological niches and potential geographical distributions of Mediterranean fruit fly (*Ceratitis capitata*) and Natal fruit fly (*Ceratitis rosa*). *Journal of Biogeography* **35**:270-281.
- Dennis, B., W. P. Kemp, and R. C. Beckwith. 1986. Stochastic-model of insect phenology - estimation and testing. *Environmental Entomology* **15**:540-546.
- Drescher, M., A. H. Perera, C. J. Johnson, L. J. Buse, C. A. Drew, and M. A. Burgman. 2013. Toward rigorous use of expert knowledge in ecological research. *Ecosphere* **4**:1-26.
- Dupin, M., P. Reynaud, V. Jarosik, R. Baker, S. Brunel, D. Eyre, J. Pergl, and D. Makowski. 2011. Effects of the training dataset characteristics on the performance of nine species distribution models: application to *Diabrotica virgifera virgifera*. *Plos One* **6**:11.
- Dwyer, G., J. Dushoff, J. S. Elkinton, and S. A. Levin. 2000. Pathogen-driven outbreaks in forest defoliators revisited: building models from experimental data. *American Naturalist* **156**:105-120.
- Dwyer, G., and J. S. Elkinton. 1993. Using simple-models to predict virus epizootics in gypsy-moth populations. *Journal of Animal Ecology* **62**:1-11.
- Elith, J., M. A. Burgman, and H. M. Regan. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling* **157**:313-329.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**:129-151.
- Elith, J., and J. Leathwick. 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions* **13**:265-275.
- Elith, J., and J. Leathwick. 2016. Boosted regression trees for ecological modeling. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>.
- Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* **77**:802-813.
- Elkinton, J. S., W. M. Healy, J. P. Buonaccorsi, G. H. Boettner, A. M. Hazzard, H. R. Smith, and A. M. Liebhold. 1996. Interactions among gypsy moths, white-footed mice, and acorns. *Ecology* **77**:2332-2342.

- Elkinton, J. S., and A. M. Liebhold. 1990. Population-dynamics of gypsy moth in North America. *Annual Review of Entomology* **35**:571-596.
- Feng, H. Q., F. Gould, Y. X. Huang, Y. Y. Jiang, and K. M. Wu. 2010. Modeling the population dynamics of cotton bollworm *Helicoverpa armigera* (Hubner) (Lepidoptera: Noctuidae) over a wide area in northern China. *Ecological Modelling* **221**:1819-1830.
- Fowler, G., L. Garrett, A. Neeley, R. Margarey, D. Borchert, and B. Spears. 2009a. Economic analysis: risk to U.S. apple, grape, orange and pear production from the light brown apple moth, *Epiphyas postvittana* (Walker). United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Center for Plant Health Science and Technology, Plant Epidemiology and Risk Analysis Laboratory, Raleigh, NC.
- Fowler, G., Y. Takeuchi, R. Sequeira, G. Lougee, W. Fussell, M. Simon, A. Sato, and X. Yan. 2009b. Pathway-Initiated Pest Risk Assessment: Asian Gypsy Moth (Lepidoptera: Lymantriidae: *Lymantria dispar* (Linnaeus)) from China into the United States on Maritime Shipments. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Raleigh, NC.
- Fraedrich, S. W., T. C. Harrington, R. J. Rabaglia, M. D. Ulyshen, A. E. Mayfield, J. L. Hanula, J. M. Eickwort, and D. R. Miller. 2008. A fungal symbiont of the redbay ambrosia beetle causes a lethal wilt in redbay and other Lauraceae in the southeastern United States. *Plant Disease* **92**:215-224.
- Fraisse, C. W., J. Bellow, and C. Brown. 2011. Degree days: heating, cooling, and growing. Agricultural and Biological Engineering Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL.
- Friedman, J., T. Hastie, and R. Tibshirani. 2000. Additive logistic regression: a statistical view of boosting. *Annals of Statistics* **28**:337-374.
- Friedman, J. H. 1991. Multivariate adaptive regression splines. *Annals of Statistics* **19**:1-67.
- Geissinger, S. 2016. Apple moth spray plan draws Capitol protesters. *East Bay Times*. Bay Area News Group, Walnut Creek, VA.
- Geyer, C. 2011. Introduction to Markov chain Monte Carlo. Pages 3-48 in S. Brooks, A. Gelman, G. L. Jones, and X. L. Meng, editors. *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton, Florida.
- Griffin, R. (2016). Exotic species from a regulatory/trade policy perspective. Personal communication to on February 4, 2016, from Doors between Floors.

- Guisan, A., T. C. Edwards, and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* **157**:89-100.
- Gullan, P. J., and P. S. Cranston. 2010. *The Insects: an outline of entomology*. 4th edition. Wiley-Blackwell, Malden, MA.
- Gutierrez, A. P., N. J. Mills, and L. Ponti. 2010. Limits to the potential distribution of light brown apple moth in Arizona-California based on climate suitability and host plant availability. *Biological Invasions* **12**:3319-3331.
- Haack, R. A., F. Herard, J. H. Sun, and J. J. Turgeon. 2010. Managing invasive populations of Asian longhorned beetle and Citrus longhorned beetle: a worldwide perspective. Pages 521-546 *Annual Review of Entomology*. Annual Reviews, Palo Alto.
- He, S., S. P. Worner, and T. Ikeda. 2012. Modeling the potential global distribution of light brown apple moth *Epiphyas postvittana* (Lepidoptera: Tortricidae) using CLIMEX. *Journal of Asia-Pacific Entomology* **15**:479-485.
- Hermes, D. A. 2013. Using degree-days and plant phenology to predict pest activity. IPM of Midwest Landscapes: Tactics and tools for IPM. University of Minnesota, Minneapolis, MN.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**:1965-1978.
- Hulme, P. E. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization [electronic resource]. *Journal of applied ecology* **46**:10-18.
- IPPC. 2010. Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. Intergovernmental Panel on Climate Change (IPCC).
- Jarnevich, C. S., and L. V. Reynolds. 2011. Challenges of predicting the potential distribution of a slow-spreading invader: a habitat suitability map for an invasive riparian tree. *Biological Invasions* **13**:153-163.
- Johnson, P. C., D. P. Mason, S. L. Radke, and K. T. Tracewski. 1983. Gypsy moth, *Lymantria dispar* (L.) (Lepidoptera: Lymantriidae), egg eclosion: degree-day accumulation. *Environmental Entomology* **12**:929-932.
- Kay, J. 2008. Experts question plant to spray to fight moths. SFGate, San Francisco, CA.
- Koch, F. H., D. Yemshanov, M. Colunga-Garcia, R. D. Magarey, and W. D. Smith. 2011. Potential establishment of alien-invasive forest insect species in the United States: where and how many? *Biological Invasions* **13**:969-985.

- Koch, F. H., D. Yemshanov, D. W. McKenney, and W. D. Smith. 2009. Evaluating critical uncertainty thresholds in a spatial model of forest pest invasion risk. *Risk Analysis* **29**:1227-1241.
- Kriticos, D. J., V. Jarošik, and N. Ota. 2014. Extending the suite of bioclim variables: a proposed registry system and case study using principal components analysis. *Methods in Ecology and Evolution* **5**:956-960.
- Kriticos, D. J., B. L. Webber, A. Leriche, N. Ota, I. Macadam, J. Bathols, and J. K. Scott. 2012. CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution* **3**:53-64.
- Kumar, S., J. Graham, A. M. West, and P. H. Evangelista. 2014. Using district-level occurrences in MaxEnt for predicting the invasion potential of an exotic insect pest in India. *Computers and Electronics in Agriculture* **103**:55-62.
- Larson, S. R., J. P. DeGroot, L. C. Bartholomay, and R. Sugumaran. 2010. Ecological niche modeling of potential West Nile virus vector mosquito species in Iowa. *Journal of Insect Science* **10**:110.
- LBAMsray.com. 2008. Defending the Central Coast from biochemical aerial spraying. LBAMsray.com. Last accessed April, 2014, <http://www.lbamspray.com/>.
- Leathwick, J. R., J. Elith, and T. Hastie. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling* **199**:188-196.
- Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. *R News* **2/3**:18-22.
- Liebhold, A., and J. Bascompte. 2003. The Allee effect, stochastic dynamics and the eradication of alien species. *Ecology Letters* **6**:133-140.
- Liebhold, A., J. Elkinton, D. Williams, and R. M. Muzika. 2000. What causes outbreaks of the gypsy moth in North America? *Population Ecology* **42**:257-266.
- Liebhold, A. M., and P. C. Tobin. 2008. Population ecology of insect invasions and their management. Pages 387-408 *Annual Review of Entomology*. Annual Reviews, Palo Alto.
- Lindsey, A. A., and J. E. Newman. 1956. Use of official weather data in spring time - temperature analysis of an Indiana phenological record. *Ecology* **37**:812-823.
- Lozier, J. D., and N. J. Mills. 2011. Predicting the potential invasive range of light brown apple moth (*Epiphyas postvittana*) using biologically informed and correlative species distribution models. *Biological Invasions* **13**:2409-2421.

- Magarey, R. D., D. M. Borchert, G. A. Fowler, and S. C. Hong. 2015a. The NCSU/APHIS Plant Pest Forecasting System (NAPPFASST). Pages 82-96 in R. C. Venette, editor. Pest Risk Modeling and Mapping for Invasive Alien Species. CAB International, Wallingford, UK.
- Magarey, R. D., M. Colunga-Garcia, J. A. Simpson, T. B. Sutton, G. A. Fowler, and D. M. Borchert. 2007. NAPPFASST: an internet system for the weather-based mapping of plant pathogens. *Plant disease: an international journal of applied plant pathology* **91**:336-345.
- Magarey, R. D., L. Newton, S. Hong, Y. Takeuchi, D. Christie, C. Jarnevich, L. Kohl, M. Damus, S. I. Higgins, L. Miller, K. Castro, G. Cook, J. Kartesz, and A. Koop. 2015b. Comparison of four modeling tools for the prediction of potential distribution for non-indigenous weeds in the United States. Manuscript submitted for publication.
- Meentemeyer, R. K., B. L. Anacker, W. Mark, and D. M. Rizzo. 2008. Early detection of emerging forest disease using dispersal estimation and ecological niche modeling. *Ecological Applications* **18**:377-390.
- Meentemeyer, R. K., N. J. Cunniffe, A. R. Cook, J. A. N. Filipe, R. D. Hunter, D. M. Rizzo, and C. A. Gilligan. 2011. Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in California (1990-2030). *Ecosphere* **2**:24.
- Milborrow, S. 2016. Package 'earth'. R Foundation for Statistical Computing, Vienna, Austria. <https://cran.r-project.org/>.
- Morisette, J. T., C. S. Jarnevich, T. R. Holcombe, C. B. Talbert, D. Ignizio, M. K. Talbert, C. Silva, D. Koop, A. Swanson, and N. E. Young. 2013. VisTrails SAHM: visualization and workflow management for species habitat modeling. *Ecography* **36**:129-135.
- Morris, R. F., F. E. Webb, and C. W. Bennett. 1956. A method of phenological survey for use in forest insect studies. *Canadian Journal of Zoology* **34**:533-540.
- Mueller, U. G., N. M. Gerardo, D. K. Aanen, D. L. Six, and T. R. Schultz. 2005. The evolution of agriculture in insects. Pages 563-595 *Annual Review of Ecology Evolution and Systematics*. Annual Reviews, Palo Alto.
- Naidoo, L., M. A. Cho, R. Mathieu, and G. Asner. 2012. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. *Isprs Journal of Photogrammetry and Remote Sensing* **69**:167-179.
- Narouei Khandan, H. A., S. P. Worner, E. E. Jones, S. L. H. Villjanen-Rollinson, L. Gallipoli, A. Mazzaglia, and G. M. Balestra. 2013. Predicting the potential global distribution of *Pseudomonas syringae* pv. *actinidiae* (Psa). *New Zealand Plant Protection* **66**:184-193.

- Newton, L., G. Fowler, A. Neeley, R. Schall, and Y. Takeuchi. 2009. Pathway-initiated pest risk assessment: *Geosmithia* sp. and *Pityophthorus juglandis* movement from the Western United States into the Eastern United States., United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Raleigh, NC.
- Nghiem, L. T. P., T. Soliman, D. C. J. Yeo, H. T. W. Tan, T. A. Evans, J. D. Mumford, R. P. Keller, R. H. A. Baker, R. T. Corlett, and L. R. Carrasco. 2013. Economic and environmental impacts of harmful non-indigenous species in Southeast Asia. *Plos One* **8**:9.
- Niemela, P., and W. J. Mattson. 1996. Invasion of North American forests by European phytophagous insects: legacy of the European crucible? *Bioscience* **46**:741-753.
- Nietschke, B. S., R. D. Magarey, D. M. Borchert, D. D. Calvin, and E. Jones. 2007. A developmental database to support insect phenology models. *Crop Protection* **26**:1444-1448.
- O'Donnell, M. S. O., and D. A. Ignizio. 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States. U.S. Geological Survey 691, <http://pubs.usgs.gov/ds/691/>.
- Peltonen, M., A. M. Liebhold, O. N. Bjornstad, and D. W. Williams. 2002. Spatial synchrony in forest insect outbreaks: roles of regional stochasticity and dispersal. *Ecology* **83**:3120-3129.
- Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology* **78**:419-433.
- Peterson, A. T., M. Papes, and D. A. Kluza. 2003. Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science* **51**:863-868.
- Peterson, A. T., V. Sanchez-Cordero, C. Ben Beard, and J. M. Ramsey. 2002. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. *Emerging Infectious Diseases* **8**:662-667.
- Peterson, A. T., R. Scachetti-Pereira, and W. W. Hargrove. 2004. Potential geographic distribution of *Anoplophora glabripennis* (Coleoptera : Cerambycidae) in North America. *American Midland Naturalist* **151**:170-178.
- Philibert, A., M. L. Desprez-Loustau, B. Fabre, P. Frey, F. Halkett, C. Husson, B. Lung-Escarmant, B. Marcais, C. Robin, C. Vacher, and D. Makowski. 2011. Predicting invasion success of forest pathogenic fungi from species traits. *Journal of applied ecology* **48**:1381-1390.
- Phillips, S. 2016. A Brief Tutorial on Maxent. AT&T Research.

- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2005. Maximum entropy modeling of species geographic distributions. *Ecological Modeling* **190**:231-259.
- Poland, T. M., and D. G. McCullough. 2006. Emerald ash borer: invasion of the urban forest and the threat to North America's ash resource. *Journal of Forestry* **104**:118-124.
- Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **9**:181-199.
- Rafter, M. A., A. J. Wilson, K. A. D. Wilmot Senaratne, and K. Dhileepan. 2008. Climatic-requirements models of cat's claw creeper *Macfadyena unguis-cati* (Bignoniaceae) to prioritise areas for exploration and release of biological control agents. *Biological Control* **44**:169-179.
- Reaumur, R. d. 1736. Dissertation to inform the history of insects. Volume 2. Continuation of the history of caterpillars and butterflies; and of the history of insects enemies of caterpillars. (Memoires pour servir a l'histoire des insectes. Tome Second. Suite de l'histoire des chenilles et des papillons; et l'histoire des insectes ennemis des chenilles.) Imprimerie Royale, Paris, France.
- Regan, H. M., Y. Ben-Haim, B. Langford, W. G. Wilson, P. Lundberg, S. J. Andelman, and M. A. Burgman. 2005. Robust decision-making under severe uncertainty for conservation management. *Ecological Applications* **15**:1471-1477.
- Regan, H. M., M. Colyvan, and M. A. Burgman. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* **12**:618-628.
- Richardson, C. H., and D. J. Nemeth. 1991. Hurricane-borne African Locusts (*Schistocerca gregaria*) on the Windward Islands. *GeoJournal* **23**:349-357.
- Robinet, C., H. Kehlenbeck, and W. Van der Werf. 2011. A report comparing the advantages and disadvantages of different approaches for creating a generic integrated model for pest spread and impacts. PRATIQUE.
- Rodda, G. H., R. N. Reed, and C. S. Jarnevich. 2007. Climate matching as a tool for predicting potential North American spread of brown treesnakes. *in* G. W. Witmer, W. C. Pitt, and K. A. Fagerstone, editors. USDA National Wildlife Research Center Symposia, Managing Vertebrate Invasive Species, Ft. Collins, CO.
- Roltsch, W. J., F. G. Zalom, A. J. Strawn, J. F. Strand, and M. J. Pitcairn. 1999. Evaluation of several degree-day estimation methods in California climates. *International Journal of Biometeorology* **42**:169-176.

- Rossiter, M. C. 1991. Environmentally-based maternal effects: a hidden force in insect population dynamics? *Oecologia* **87**:288-294.
- Roura-Pascual, N., L. Brotons, A. T. Peterson, and W. Thuiller. 2009. Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. *Biological Invasions* **11**:1017-1031.
- Roura-Pascual, N., A. V. Suarez, C. Gomez, P. Pons, Y. Touyama, A. L. Wild, and A. T. Peterson. 2004. Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. *Proceedings of the Royal Society B-Biological Sciences* **271**:2527-2534.
- Russo, J. M., A. M. Liebhold, and J. G. W. Kelley. 1993. Mesoscale weather data as input to a gypsy moth (Lepidoptera, Lymantriidae) phenology model. *Journal of Economic Entomology* **86**:838-844.
- Sailer, R. I. 1978. Our immigrant insect fauna. Pages 3-11 *Bulletin*.
- Scachetti-Pereira, R. 2014. Desktop Garp users manual. University of Kansas Biodiversity Research Center, Lawrence, KS.
- Schneider, R. W., C. A. Hollier, H. K. Whitam, M. E. Palm, J. M. McKemy, J. R. Hernandez, L. Levy, and R. DeVries-Paterson. 2005. First report of soybean rust caused by *Phakopsora pachyrhizi* in the continental United States. *Plant Disease* **89**:774-774.
- Senay, S. D., S. P. Worner, and T. Ikeda. 2013. Novel three-step pseudo-absence selection technique for improved species distribution modeling. *Plos One* **8**:16.
- Sequeira, R. 1999. Safeguarding production agriculture and natural ecosystems against biological terrorism. Pages 48-67 *in* T. W. Frazier and D. C. Richardson, editors. *Food and Agricultural Security: Guarding Against Natural Threats and Terrorist Attacks Affecting Health, National Food Supplies, and Agricultural Economics* The New York Academy of Science, New York, NY.
- Sequeira, R. 2002a. Agricultural risk assessment. Pages 139-156 *in* G. J. Hallman and C. P. Schwalbe, editors. *Invasive Arthropods in Agriculture*. SCIENCE PUBLISHERS, Inc., Enfield, New Hampshire.
- Sequeira, R. A. 2002b. Availability of risk management analysis for the importation of clementines from Spain. Pages 18578 -18579 *in* U. A. PPQ, editor. *Federal Register Notice*, Washington, DC.
- Sevacherian, V., V. M. Stern, and A. J. Mueller. 1977. Heat accumulation for timing *Lygus* (Hemiptera-(Heteroptera)-Miridae) control measures in a safflower-cotton complex. *Journal of Economic Entomology* **70**:399-402.

- Sheehan, K. A. 1992. User's guide for GMPHEN: gypsy moth phenology model. United States Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. General Technical Report NE-158,
- Soliman, T., M. C. M. Mourits, A. Lansink, and W. van der Werf. 2010. Economic impact assessment in pest risk analysis. *Crop Protection* **29**:517-524.
- Stockwell, D., and D. Peters. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* **13**:143-158.
- Stockwell, D. R. B., and A. T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* **148**:1-13.
- Stohlgren, T. J., P. Ma, S. Kumar, M. Rocca, J. T. Morisette, C. S. Jarnevich, and N. Benson. 2010. Ensemble habitat mapping of invasive plant species. *Risk Analysis* **30**:224-234.
- Sutherst, R. W., G. F. Maywald, and D. Kriticos. 2007. CLIMEX version 3, user's guide. Hearne Scientific Software Pty Ltd, Melbourne, Australia.
- Talbert, C. B., and M. K. Talbert. 2012. User manual for SAHM package for VisTrails. US Geological Survey, Fort Collins Science Center, Ft. Collins, CO.
- Thuiller, W., D. M. Richardson, P. Pysek, G. F. Midgley, G. O. Hughes, and M. Rouget. 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology* **11**:2234-2250.
- Tuv, E., A. Borisov, G. Runger, and K. Torkkola. 2009. Feature selection with ensembles, artificial variables, and redundancy elimination. *Journal of Machine Learning Research* **10**:1341-1366.
- Tyler, P. S., and R. J. Hodges. 2002. Phytosanitary measures against larger grain borer, *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae), in international trade. *Integrated Pest Management Reviews* **7**:279-289.
- USDA APHIS. 2015. Asian longhorn beetle eradication program. United States Department of Agriculture, Animal and Plant Health Inspection Service, Riverdale, MD.
- USDA APHIS. 2016a. Pest detection. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD. Last accessed January, 2016, <https://www.aphis.usda.gov/aphis/ourfocus/planthealth/pest-detection>.
- USDA APHIS. 2016b. Plant pests and diseases programs. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and

- Quarantine, Riverdale, MD. Last accessed January, 2016, <https://www.aphis.usda.gov/aphis/ourfocus/planthealth/plant-pest-and-disease-programs/pests-and-diseases>.
- Vaclavik, T., and R. K. Meentemeyer. 2009. Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling* **220**:3248-3258.
- Venette, R. C., D. J. Kriticos, R. D. Magarey, F. H. Koch, R. H. A. Baker, S. P. Worner, N. N. G. Raboteaux, D. W. McKenney, E. J. Dobesberger, D. Yemshanov, P. J. De Barro, W. D. Hutchison, G. Fowler, T. M. Kalaris, and J. Pedlar. 2010. Pest risk maps for invasive alien species: a roadmap for improvement. *Bioscience* **60**:349-362.
- Wallner, W. E. 1987. Factors affecting insect population dynamics: differences between outbreak and non-outbreak species. *Annual Review of Entomology* **32**:317-340.
- Wang, Y., S. Wu, and R. Zhang. 2009. Pest risk analysis of a new invasive pest, *Phenacoccus solenopsis*, to China. *Chinese Bulletin of Entomology* **46**:101-106.
- West, A. M., S. Kumar, and C. S. Jarnevich. 2016. Regional modeling of large wildfires under current and potential future climates in Colorado and Wyoming, USA. *Climatic Change* **134**:565-577.
- Williams, D. W., R. W. Fuester, W. W. Metterhouse, R. J. Balaam, R. H. Bullock, and R. J. Chianese. 1991. Oak defoliation and population-density relationships for the gypsy moth (Lepidoptera, Lymantriidae). *Journal of Economic Entomology* **84**:1508-1514.
- Wilson, L. T., and W. W. Barnett. 1983. Degree-days: an aid in crop and pest management. *California Agriculture* **37**:4-7.
- Wittwer, G., S. McKirdy, and R. Wilson. 2005. Regional economic impacts of a plant disease incursion using a general equilibrium approach. *Australian Journal of Agricultural and Resource Economics* **49**:75-89.
- Yemshanov, D., F. H. Koch, and M. Ducey. 2015. Making invasion models useful for decision makers; incorporating uncertainty, knowledge gaps, and decision-making preference. Pages 206-222 in R. C. Venette, editor. *Pest Risk Modeling and Mapping for Invasive Alien Species*. CAB International, Wallingford, UK.
- Zalom, F. G., P. B. Goodell, L. T. Wilson, W. W. Barnett, and W. J. Bentley. 1983. Degree-days: the calculation and use of heat units in pest management., University of California, Division of Agriculture and Natural Resources. Leaflet 21373,
- Zlotina, M. A., V. C. Mastro, J. S. Elkinton, and D. E. Leonard. 1999. Dispersal tendencies of neonate larvae of *Lymantria mathura* and the Asian form of *Lymantria dispar* (Lepidoptera : Lymantriidae). *Environmental Entomology* **28**:240-245.

## **CHAPTER 3. Development of a Spatial Analytic Framework for Advanced Risk Information Systems**

### **3.1 Introduction**

In my dissertation I advance a hypothesis that a modeling framework can result in more consistent, tractable, and comprehensive weather-driven pest forecasts that would be relevant to large-scale regulatory pest and disease management. I argue that my approach to setting up a framework has the potential to make uncertainty more explicit, tractable and simpler to communicate. The expected impact is that numerous advantages will result, from being able to make more valid comparisons between different models to improving the role and value of models for informing policy.

A framework is defined as a structure intended to support things, such as buildings, ideas, and facts (Framework n.d.). A framework for an agricultural pest forecast system is a collection of key information streams to support forecasting of pest events, behaviors, and attributes. Biotic and abiotic factors influence agricultural pests; therefore, a collection of essential biotic and abiotic driving factors within the framework is key to make the framework robust.

In this chapter, I describe the process and approaches I used to develop a framework to support seamless integration of various databases, pest biological information, and pest predictive models. This framework is primarily built for decision makers in the United States although the system will be open and can be adapted by any organization.

Such efforts have been attempted by others and some key examples are highlighted below. The discussion below focuses on selected assemblies of models and frameworks that

have been used in area-wide regulatory applications or that have seen use in large-scale, area-wide management programs.

### **3.1.1 Review of frameworks for pest forecasting**

Frameworks for pest management has been developed to model social science and human behavior (McLeod et al. 2015), to model areas suitable for pest establishment and growth (Morissette et al. 2013), to model population dynamics (Gutierrez et al. 1984), and to develop pest risk assessments (Pratique) (Pluess et al. 2012, Holt et al. 2014).

The Software for Assisted Habitat Modeling (SAHM) was developed by United States Geological Survey (USGS) to manage habitat models and input data drivers through visualization workflow software called VisTrails (Morissette et al. 2013) . VisTrails is an open source scientific workflow system that assists in understanding the data and interactions between data and models. The SAHM system has been frequently used by analysts to identify pest suitability and suitable areas for better management practices (Crall et al. 2013, Jarnevich et al. 2014a, Jarnevich et al. 2014b). SAHM is built on open-source software and is available from the USGS Fort Collins Science Center.

The Pratique (Enhancements of Pest Risk Analysis Techniques) is a framework developed to improve management of phytosanitary issues. This project was funded by the European Union (EU) under the 7<sup>th</sup> Research Framework Programme (FP7) in 2008 to create databases required for pest risk assessments (PRAs) for the entire EU; specifically, it was designed to improve methods and techniques used in PRAs and to create a web interface to aid the PRA decision support scheme (European Commission 2016). Although this design is EU-centric, the framework and approaches have been adapted by others

globally. Currently, Pratique has 15 partners from 8 EU countries, EPPO (European Plant Protection Organization), CABI, Australia, New Zealand, and subcontractors in China and Russia (European Commission 2016).

The Pratique framework covers comprehensive information applicable to a variety of phytosanitary questions. Here, I list only key features: 1) pest biological information, 2) host information, 3) inventory and description of key databases in EU, 4) methods and models to assess pest risk assessment, 5) mapping techniques and procedures, 6) estimation of uncertainty, 7) evaluation of systems approach, 8) eradication protocols, and 9) outreach and training methods.

The methods and models provided within Pratique include, but are not limited to, pest spread models, economic/environmental/social impact models, pest pathway models, and impact assessment methods for a pest risk analysis. All these models and information are packaged as a software called “CAPRA” (Computer Assisted Pest Risk Analysis). The models, techniques, data drivers, and CAPRA are available at the Pratique website (<https://secure.fera.defra.gov.uk/pratique/index.cfm>).

### **3.1.2 Pest forecasting and information management in the regulatory agency, USDA APHIS PPQ**

Currently, the United States Department of Agriculture, Animal and Plant Health Inspection Service (USDA APHIS) does not have a formal framework for comprehensive pest risk assessments. Manuals and protocols are available for pest risk assessments, eradication programs, pest control programs (e.g. emergency programs and domestic programs), and quarantine treatments (USDA APHIS 2010, 2012, 2014, 2015a, b).

Information management is one of the key elements in successful pest management. USDA APHIS risk analysts experience challenges with data management, especially for spatial data and spatial models. Geographic information systems (GIS) have become popular in recent years as computing technologies have improved. They are often used as a way to integrate data layers and sometimes constitute frameworks for modeling.

For example, in ecological science, GIS has become an essential modeling technique. In ecology, species responses can be experimented with digitally at the individual level. However, species responses as a group or as a part of ecological landscapes are completely different from individual responses. In reality, no species acts alone. It has been suggested that cumulative effects in ecological processes can only be researched by using GIS technologies and simulation experiments (Graham et al. 1991). Some interactions between species and landscapes can only be understood at a regional level, especially processes like migration or new habitat adaptations that occur only at large scales. Species behave differently under different landscape types, and GIS technologies help understand species interactions at large spatial scales (Manel et al. 2000, Osborne et al. 2001).

The use of GIS was very limited from the late 1980s to the beginning of the 1990s. It was used by a limited number of people in relatively few fields (e.g., forestry, transportation, utility, agriculture and environment, civil engineering) (Tomlinson 1987). Even though spatial analyses were considered and performed in some cases, the majority of GIS use was associated with cartography (Tomlinson 1987). At that time, the inhibiting

factors were availability of digital data, the price of software, the price of hardware, and software capabilities (e.g., no system available for large scale analysis) (Tomlinson 1987).

The current situation is that many GIS limitations from 25 years ago have been overcome. However, there is still limited progress in connecting global scale data to model species behaviors and to forecast critical events. For example, global climatic datasets continue to be sparse and error prone (Hijmans et al. 2005). Solar cycle patterns are relatively well understood by climatologists and effects on landscape level dynamics have been measured. However, solar cycles are not often utilized to address large scale patterns, especially in terms of management of invasive species (Auclair 2015).

Another current limitation is digital data overload. One of the biggest limitations in 1980s was availability of digital data. Now, there are so many digital data in existence that it is often difficult to understand what is available, how it is developed, and how to correctly utilize the data (Walters et al. 2003).

### **3.2 Concept of a Spatial Analytic Framework for Advanced Risk Information Systems**

In order to establish improvements over previous systems and current tools, we wanted to clearly articulate formal criteria that would constitute an ideal, yet practical framework for area-wide regulatory applications. The approach to establishing the criteria was to conduct extensive literature reviews and importantly to establish a dialogue with senior regulators at USDA APHIS and selected staff at large organizations in Europe, Africa, South America and Asia. The dialogue has taken advantage of fora like the International Pest Risk Research Group workshops (<http://www.pestrisk.org/>) and specific

cooperative projects with regulatory bodies in Africa and Asia (e.g., East African Phytosanitary Information Committee).

The Spatial Analytic Framework for Advanced Risk Information Systems (SAFARIS) was designed to provide a seamless environment for pest predictive models. I proposed this idea to address known deficiencies and requirements that USDA APHIS Plant Protection and Quarantine (PPQ) was experiencing and also to accommodate more general applications with a system that facilitated the work of an analyst. Literature reviews and consultations confirmed that these challenges were not unique to USDA (Morissette et al. 2013, Holcombe et al. 2014). The USDA APHIS PPQ previously used a pest forecast system that provided basic pest phenology forecasts and predictions of suitable areas based on total heat unit accumulations; however, due to limited functionality (as per the criteria listed earlier) and lack of external funding, the system is no longer available. The development of SAFARIS intends to provide essential functions for PPQ as part of a continuously available system to provide pest and disease forecasts.

The key goals for developing a framework for pest predictive models for decision makers are 1) to develop and provide commonly used pest predictive models in a web-based environment, 2) to update and manage data and databases that are relevant to pest predictive modeling, 3) to provide timely pest information such as pest event timing, potential habitats and spread, and economic impacts, 4) quantify uncertainty associated with outputs, 5) interpret outputs with uncertainty information for decision makers, and 6) archive analysis results and provide web-interactive mapping to display and share results.

The critical criteria that established guidance to build the framework include:

1. The framework must accommodate epidemiological approaches to modeling pest and disease behavior. In its simplest form, an epidemiological approach implies the ability to capture essential elements of the behavior of the pest, the plant host, the biotic and abiotic environment, and human-based elements (management practices as well as factors that affect pest behavior such as transportation networks and trades) such that pest dynamics models are able to capture essential elements that inform likelihoods and potential impacts.
2. There has to be consideration of driving variables at a global scale. Specifically, climatological characterization, weather forecasts and host distribution must be captured for most countries that engage in trade. The reason for this is that for most countries, a key concern is the introduction of new exotic species. Understanding impacts in new areas depends on understanding dynamics in areas of origin. Furthermore, trade, not just its volume but also its diversity, has made it such that the distances between countries or vast oceans that separate countries have become irrelevant. In real terms, the borders between many countries have become shared beyond our traditional understanding of shared borders. For exotic pests and diseases, trade largely dictates which borders are biologically adjacent to a given country. Biological adjacency refers to the fact that trade pushes borders much closer together in a manner proportional to the total volume of trade of agricultural or environmental significance. Biological adjacency is used to reflect the fact that distance between some borders (e.g., between China and the US) are functionally erased by trade.

3. The effects of climate vs. climate change must be considered by enabling forecasts to be driven by both climatology (historical records) as well as climate forecasts.
4. There has to be an explicit accounting of sources of uncertainty and variability. The differences that might be observed in the behavior of different models can be due to the differences inherent in the model form and formulation or differences in the kind and nature of the driving variables, notably climate-related drivers. Tractability and explicit accounting of uncertainty in practice is supported by having access to raw data as opposed to derived data. A practical example is having ground-based weather data available to users as well as interpolated, gridded data. Explicit consideration of uncertainty also implies establishing mechanisms for verification and validation of specific models. Importantly, the uncertainty associated with forecasts based on historical climate data cannot be treated in the same way as the uncertainty associated with forecasts based on general circulation model (GCM) output. Nevertheless, it is essential that uncertainty needs to be described and communicated no matter what modeling approach or data driver is used.
5. The spatial resolution of the driving variables, especially abiotic variables, must be such that eventual regulatory action is possible at the smallest spatial scale that is still biologically relevant. The reason is that regulations are costly, and there is an interest in minimizing the impacts of regulatory actions. Generally, in the United States, spatial resolution must be at the scale of a “county” or finer. Since counties vary in size, a more generalized approach is to have cell sizes that are 50-kilometer

or less. Spatial relationships (explicit geo-location of modelled features) need to be made explicit using GIS approaches.

6. The temporal resolution of the driving variables, especially abiotic variables, must be such that modeling of key pest behavior is possible in a way that facilitates regulatory actions. For example, if a pest completes a key life cycle (e.g., emergence from egg stage to adult flight) in days, then regulatory actions may be needed before the adult stage is completed. In this case, driving variables that have a 30-day time step (e.g., monthly average or “normals”) would not allow for sufficient temporal resolution that would be meaningful for pest management. Daily is considered a minimum resolution, but I acknowledge that in some cases (some plant diseases), hourly data may be ideal.
7. There are a number of administrative “desiderata”. A) Any framework must be nimble in terms of computer language implementation. There is no long-term stability in languages or development environments so considerations need to be made for a “living” approach to maintaining a framework. B) Frameworks should accommodate server-based as well as desktop-based operation to maximize usability. C) Systems that are used by regulatory agencies should encounter security requirements when operated from administrative servers. Understanding security requirements facilitates adoption of systems by both academia and government. D) Development of a framework should acknowledge that availability of source code facilitates the evaluation of security issues, adoption by regulatory bodies, and use by academic partners as well as developing countries. This means that open source

- approaches are of interest. E) Paradoxically, governance of an open source, freely distributed system necessitates guidelines and protection of intellectual rights to ensure the integrity of the product. In simplest terms, lack of protection exposes an open source system to removal of availability by copyright claims by third parties.
8. Since impacts of invasive exotic species occur over time (often in the future), it is necessary that models include both a short- and long-term modeling capability. Within-year dynamics are key to inform management and to plan management practices. Decades-long forecasts are essential to understand present-day impact assessments (that is, our current understanding of what future losses mean in today's terms).
  9. A very important feature of the framework is that it does not focus on a single model or even a single modeling approach. However, testing the framework does require that specific models be used to verify its functionality. Further, it is assumed that a few selected modeling approaches can be used to address most area-wide, regulatory forecasting needs for exotic pests.
  10. Models that are considered compatible with the current framework include those that are driven by weather variables to describe the abiotic environment and by life table and host distribution data to describe the biotic components.

### **3.2.1 Description of the components and integration of SAFARIS**

As noted, the development of SAFARIS aims to provide essential pest forecasting functions for PPQ as part of a continuously available system. The plan is to take advantage of a limited open-source approach for the framework and key models to foster continued

cooperation, innovation, and creativity. Whereas the code is planned to be shared freely within academic environments and with USDA, I will seek intellectual property rights protection and others as appropriate, to ensure that commercial applications by third parties that obtain copies of the system do not inhibit future access or stifle innovation, creativity and development.

### **3.2.2 Targeted Users and Audience for SAFARIS**

SAFARIS is a framework for pest forecast models that is designed for researchers, risk analysts, decision/policy makers, rapid-responders, and land managers in need of streamlined and tractable (in the sense that the data drivers and models are not black boxes) fast forecasts. The main customers/audience for SAFARIS is currently the regulatory community exemplified by USDA APHIS PPQ. The main system is intended to be an open-source product. Interested cooperators or users are envisioned to access and use the system as a web-enabled and server-based system. Alternatively, a user could download the source code to modify and use on a desktop (or mirror site) and contribute and share ideas as part of a cooperative modeling community. Early case studies (as part of a project I led and which was funded by FAO in Jamaica) demonstrated that SAFARIS may play an important role in providing forecasting functions to partners in developing nations (Takeuchi 2015).

### **3.2.3 Determination of the SAFARIS Components**

A pest forecast system is important not only for regulatory agencies but also for farmers and scientists. Weather information is crucial for forecasting insect and plant pathogen growth. Linking essential climate variables, such as air temperature, precipitation,

relative humidity, and wind speed enables prediction of key events like pest emergence, reproduction, and migration periods. I note that weather is used here to refer to present-time conditions; climate is used to refer to historical data or average trends. Climate change forecasts are assumed to be derived as a product of running GCM models. That kind of information is always referred to as ‘climate change’ data. Unless otherwise noted, “weather forecasts” are assumed to be based on historical climatology, not on GCMs. To develop an appropriate framework it is useful to first review the kinds of models that would be suitable for such a framework.

A pest forecast model that is plugged into the SAFARIS framework can be simple or complex. A simple approach such as the approach associated with phenology or degree-days (also known as “heat units”) model can provide great insight (Dennis et al. 1986). However, similar environmental drivers can also be used to support much more complex approaches like population dynamics models and other models that have increasingly explicit routines to capture more behavioral and epidemiological dynamics (Gutierrez et al. 2010). Clearly, the kind of modeling approach that is most suitable depends on the questions being asked.

Simple phenology models (the so-called “degree-days” or heat units model) estimate the timing of arthropod development and can also be used to estimate the number of potential generations per year if one assumes that other factors are close to ideal (i.e., availability of host material/suitable food and suitability of other abiotic and biotic factors). While a phenology or degree-day model can easily approximate the timing of arthropod development stages based on the current data for immediate forecasts, such models have

several disadvantages. One important disadvantage is that the basic degree day model assumes that arthropod development is a linear function of temperature. Under extreme temperatures (near maximum and minimum threshold temperatures), insects do not develop in the same way they do under optimal and near-optimal temperatures. One approach to address some of the limitations of basic phenology models is to ensure that temperature is modeled in such a way to capture the nonlinear relationship to development. Given the importance of these phenology systems to regulatory decision making, I have investigated practical approaches to overcome the limitations associated with assumptions of linearity.

Other factors beyond climate data are also important in pest forecast systems. Factors such as host type, host density, host distribution, pest population density, natural enemy population density, human movement, and human-assisted activities also influence pest behavior. Data in addition to climatic data that are commonly used for pest forecast models also need to be processed and captured in formats that can be input for the models and archived within the framework.

### **3.3 Components of SAFARIS**

SAFARIS was designed using a multi-compartment model design focused on supporting five specific categories of models or applications. Those are 1) climate pattern analysis (e.g., phenology, suitability), 2) population dynamics models, 3) spread analysis models, 4) economic analysis models, and 5) uncertainty analysis (Figure 3-1). There are many variations around these topics but those categories capture key regulatory and area-wide management applications. Other variations of such applications include population

dynamics under eradication programs and population dynamics under sterile insect treatment (SIT).

In addition to integrating relevant and adaptable pest models, associated data required to run the models are also stored within the SAFARIS using standardized formats that directly feed into the models. Thus, the SAFARIS can be considered to be constituted by three main components: a pest biology database, input data drivers (climate data and non-climate data), and pest forecast models (Figure 3-1).

SAFARIS (Figure 3-1) allows its users to select amongst climate databases. Climate databases include 1) historical climate data since 1950, 2) short-term weather forecasts (7-day forecast) and 3) downscaled GCM outputs. The temporal resolution varies with datasets. Each model requires different temporal resolution climate data; therefore, appropriate climate data sources for the models are already linked within SAFARIS. For example, when a user selects a suitable/unsuitable mapping tool to determine areas that are suitable for a pest's survival during winter months, he/she may choose a model that is connected to monthly climate datasets. However, a phenology model typically uses daily minimum and maximum temperatures to calculate degree-days for a particular pest, so a phenology model must be linked to daily climate datasets instead of hourly or monthly datasets. In addition, a user can run models using historical climate data, short-term weather forecasts, and downscaled GCM climate data to understand pest predictions for near-real time forecast, short-term forecast, and long-term forecast. The spatial resolution also varies with climate datasets. The next section provides detailed description of the climate datasets and information streams.

SAFARIS (Figure 3-1) also contains databases for pest information. Pest information includes pest life cycle data such as growth rate, mortality rate, fecundity rate, lethal and optimal temperatures for growth, host information, dispersal capability, and enemy information. A user can select pre-parameterized pest information to run a phenology model or population dynamics model within SAFARIS or can simply parameterize a pest based on his/her own knowledge.

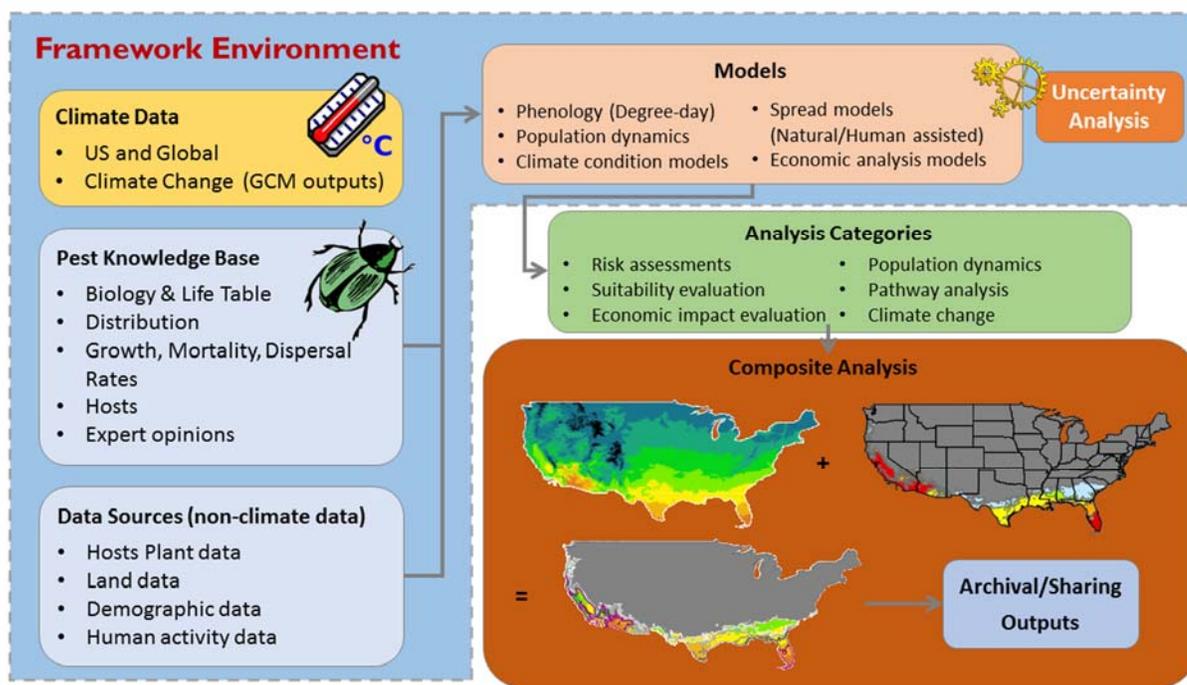


Figure 3-1. The framework environment

### 3.3.1 Information Streams and Data Drivers for Pest Forecasts

A key challenge for any weather and biology driven process like most pest forecasting models is how to identify appropriate data drivers and integrate the information in a streamlined, consistent manner. I have discussed the different kinds of abiotic and biological drivers with meteorological institutions, regulatory bodies and researchers in the US, Africa and Asia. I have conducted in-depth reviews of the literature. I have also

explored the actual data by downloading the information and asking the following questions: Are the data credible? Is it possible to understand how the observations are gathered? Is it possible to calculate variability and other measures of uncertainty? (This is not easy with derived data.) Are the data likely to be available over the long term? Are the data open to public access and can I link to data in a distributed system? Can I use the data streams within the SAFARIS framework? Do they support the criteria established for the system? The section below identifies those data streams that match pest forecasting needs and that have been evaluated for compatibility with the objectives of this initiative.

### **3.3.1.1 Pest information and pest biology data**

Model parameterization as well as parameter fitting require pest biology information. The availability of such information varies with species. However, key parameters need to be collected and stored in a database. The key parameters include pest distribution, host species and distribution, lethal and optimal temperature for pest growth, physiological heat requirements, growth rate, mortality rate, fecundity rate, and information on enemy species and predator species. The key parameters are mainly obtained by literature search, expert opinion, and databases that contain pest information. In addition to literature databases (e.g., AGRICOLA, CAB abstracts, Web of Science), the list below identifies commonly used datasets that contain key biological parameters.

#### **Global Pest and Disease Database**

The Global Pest and Disease Database (GPDD) is developed to archive plant pest information to support USDA APHIS PPQ programs (<http://www.gpdd.info>). The GPDD is

maintained by USDA APHIS PPQ and North Carolina State University, Center for Integrated Pest Management.

The GPDD focuses on quarantine pests to the United States and stores information on plant pests that are not currently in the United States or have limited distributions. The database contains pest taxonomy, identification, biology, distribution, hosts, known impact, detection, and control methods from published references and literature. All references are documented as pdf files and stored in the database so that users are able to obtain the original data source.

The GPDD currently stores more than 5,300 unique pests, 800,000 data records, and 59,000 pdf documents (Suiter 2015). The GPDD has limited access policy that is only available to employees and contractors for USDA APHIS.

### **Crop Protection Compendium**

The Crop Protection Compendium (CPC) is developed by CAB International and contains biological information on crop pests, diseases, and weeds (CABI 2016). The database contains taxonomy, nomenclature, morphology, distribution, history of introduction and spread, habitat conditions, host species, biology, natural enemy, pathway and means of dispersal, environmental impact, social impact, economic impact, and prevention and control. Currently, CPC includes over 26,000 datasheets with 250,000 references.

The database is accessed via the Internet (<http://www.cabi.org/cpc>). This is a licensed database, and the users must purchase a license to access data on the CPC database.

## **ScaleNet**

The ScaleNet contains information about scale and mealybug insects (e.g., Coccoidea, Pseudococcidae). The database includes information on taxonomy, nomenclatural history, biogeography, ecological associations, and economic importance (Garcia Morales et al. 2016). Currently, it contains data from over 23,000 references pertaining to 8,195 species (Garcia Morales et al. 2016).

The original ScaleNet was developed by Yair Ben-Dov, Douglass R Miller and Gary A P Gibson with major support from the United States-Israel Binational Agricultural Research and Development Fund (Garcia Morales et al. 2016). The original ScaleNet ended on July 8, 2015. The current version of ScaleNet is managed by USDA Agricultural Research Service, Systematic Entomology Laboratory.

This is a publically available database. Scale information can be retrieved at <http://scalenet.info/>.

## **Systematic Botany and Mycology Laboratory Fungal Database**

The Systematic Botany and Mycology Laboratory (SBML), which is part of USDA Agricultural Research Service (ARS) and located in Beltsville, MD, updates and maintains SBML Fungal Databases (<https://nt.ars-grin.gov/fungaldatabases/index.cfm>).

The SBML currently maintains 5 fungal databases: 1) Specimens in the U.S. National Fungus Collections, 2) Fungus-Host Distribution, 3) Literature, 4) Index to Saccardo's *Sylloge Fungorum*, and 5) International Mycological Institute's Index of Fungi. All databases are accessible to the public from the ARS SMLB website.

The databases contain fungal taxonomy, specimen collection information, host species, and fungal distribution with associated references from all over the world.

Currently, the databases contains more than 560,000 specimen records describing 80,000 fungal taxa on 56,000 vascular plant hosts (representing 324,000 unique host-fungus combinations), and 33,000 references.

### **PestID**

The USDA Animal and Plant Health Inspection Service (APHIS) maintains a database called PestID, which contains data on pest interceptions at U.S. port-of-entry. The database includes data on date of interception, pathway, detection location, origin and destination of the pest, host material, host part, method of importation (e.g., passenger, permitted cargo), pest type, pest species, pest determination date, and ship/airline information.

#### **3.3.1.2 Climate data**

Information about climate and weather is one of the fundamental drivers of pest forecast. Availability and management of this information is essential to establishing, running and maintaining a framework that supports pest forecasts. One of the unfortunate features of most pest forecasting models is that they use unique data streams, different formats and often unclear methods of applying the information. Some systems and models (e.g., USPest.org, iAIMS, CLIMEX) use point data for climate (from ground-based weather stations) and run forecasts at a point level (Sutherst et al. 2007, Yang et al. 2010, Coop 2015, UCIPM 2016). Other systems and models (e.g., USPest, CLIMEX, ) first interpolate ground data and develop standardized grids using interpolation systems and run forecasts

based on the interpolated grid values; typically, the interpolated values are associated with the grid cell centroid (Sutherst et al. 2007, Coop 2015). SAFARIS will use interpolated values and establish grids to support forecasts. However, it will also make available the un-interpolated data so that different approaches can be supported.

It is not possible to establish a single climate data stream and assume that it is appropriate for all applications. Some data streams have a high resolution (high density of observations) for a small area (sometimes county-level only); sometimes the database describes the continental United States only; some datasets are global but are also very coarse (few data points), which increases uncertainty. In many cases, the only way to establish a global-scale system is to integrate different databases that have different resolutions (spatial and time) which necessitates the use of interpolation.

Given this reality, it is essential that we understand data streams and later how they can be integrated. There is an abundance of relevant data streams and institutional sources. This section reviews those that are considered reliable, available sources of information to support the SAFARIS framework.

### **National Climatic Data Center**

The National Climatic Data Center (NCDC) is one of the three data centers that are managed and operated by the National Centers for Environmental Information (NCEI) (Department of Commerce, National Oceanic and Atmospheric Administration). The NCEI is responsible for preserving, monitoring, assessing, and providing public access to the Nation's climate and historical weather data and information (NCEI 2016). The NCDC provides numerous weather data at global and national scales. The weather station data at

global level is available from the NCDC Climate Data online website (<http://www.ncdc.noaa.gov/cdo-web/>). The information is updated daily. Users are able to specify the station locations as well as the weather data period. The daily temperature data as well as precipitation, wind speed, and wind direction, are provided in comma-separated value (CSV) text file format. The raw data include some invalid data.

### **Climate Research Unit**

The Climate Research Unit (CRU) is part of the School of Environmental Sciences at the University of East Anglia, UK. The CRU provides monthly data of precipitation, average temperature, maximum temperature, and minimum temperature as 0.5 degree (approximately 50 kilometer) raster datasets at global scale since January, 1901 (CRU 2015). The data can be accessed and downloaded at the CRU Data website (<http://www.cru.uea.ac.uk/data>).

### **PRISM**

The PRISM climate group was established by the Northwest Alliance for Computational Science & Engineering at Oregon State University. The PRISM climate group provides daily and monthly averaged precipitation, average temperature, minimum temperature, and maximum temperature for contiguous United States at 4-kilometer resolution raster datasets since 1981 (<http://www.prism.oregonstate.edu/>) (PRISM 2015). The PRISM climate group also provides daily climate data at 800-meter resolution since 1981; however, those data are not publically available.

The contiguous U.S. monthly and 30-year climate normals (averaged monthly and annual climate conditions from 1981 to 2010) (for precipitation, mean temperature,

minimum temperature, maximum temperature, mean dew-point temperature, minimum vapor pressure deficit, and maximum vapor pressure deficit) are also available based on the climatology from 1981 to 2010. The climate normals come in at 4-kilometer and 800-meter resolutions.

The PRISM climate group also modeled time series datasets containing monthly precipitation, monthly mean temperature, monthly minimum temperature, and maximum temperature from 1895 to 1980 by using climatologically-aided interpolation, which uses the long-term average pattern as first estimate of the spatial pattern of climatic conditions for a given month (PRISM 2015).

### **Daily Surface Weather and Climatological Summaries**

The Daily Surface Weather and Climatological Summaries (DAYMET) datasets were originally developed by the Numerical Terradynamic Simulation Group at the School of Forestry, University of Montana (Thornton et al. 1997). The DAYMET is currently archived and distributed through the Oak Ridge National Laboratory, Distributed Active Archive Center for Biogeochemical Dynamics. DAYMET datasets include seven daily variables since 1980, covering the conterminous United States, Mexico, and Southern Canada at 1-kilometer resolution grid. Those seven variables are 1) maximum temperature, 2) minimum temperature, 3) shortwave radiation, 4) vapor pressure, 5) snow-water equivalent, 6) precipitation, and 7) day length.

### **TopoWX**

The Topography Weather (TopoWX) dataset contains gridded minimum and maximum topoclimatic air temperature for the conterminous United States at 30-arc second

resolution (800 meter) (Oyler et al. 2015). This dataset was also developed by the Numerical Terradynamic Simulation Group at the Department of Ecosystem and Conservation Sciences, University of Montana. The gridded daily and monthly minimum temperature and maximum temperature from 1948 to 2014 are accessible through an FTP server or THREDDS server at the TopoWX website (<http://www.ntsug.umt.edu/project/TopoWX>).

### **Gridded Surface Meteorological Data**

The Gridded Surface Meteorological Data (METDATA) dataset was developed at the Department of Geography, University of Idaho (Abatzoglou 2013). The METDATA dataset covers the contiguous United States at 4-kilometer spatial resolution. The current dataset includes daily weather data on precipitation, maximum temperature, minimum temperature, humidity, surface downward shortwave radiation, and 10-meter wind velocity from 1979 to 2014.

The METDATA climatic data were derived by combining meteorological attributes from two databases: the North American Land Data Assimilation System Phase2 (NLDAS-2) (daily weather dataset) and the Parameter-elevation Regression on Independent Slopes Model (PRISM) (monthly data). The NLDAS-2 is temporally rich data, while PRISM enhanced spatial resolution of the climatic data.

### **North American Land Data Assimilation System (NLDAS-2)**

The North American Land Data Assimilation System Phase 2 (NLDAS-2) is a joint project between NOAA NCEP EMC (National Oceanic and Atmospheric Administration, National Centers for Environmental Prediction, Environmental Modeling Center), NASA

GSFC (National Aeronautics and Space Administration, Goddard Space Flight Center), Princeton University, the University of Washington, NOAA NWS OHD (NOAA National Weather Service, Office of Hydrologic Development), and NOAA NCEP CPC (NOAA/NCEP Climate Prediction Center) (NASA 2016). The main objective of NLDAS-2 is to generate high quality land-surface model datasets that are spatially and temporally consistent to support modeling activities (NASA 2016).

The NLDAS-2 provides hourly climate data on air temperature at 2 meters above the surface, surface pressure, specific humidity at 2 meters above the surface, surface downward longwave radiation, surface downward shortwave radiation, precipitation hourly total, potential evaporation, u-wind component at 10 meters above the surface, and v-wind component at 10 meters above the surface since January 1, 1979.

### **Real-Time Mesoscale Analysis (RTMA)**

The Real-Time Mesoscale Analysis (RTMA) is a high spatial and temporal resolution analysis product developed by the National Oceanic and Atmospheric Administration, National Centers for Environmental Prediction (NOAA NCEP) (NCEP 2016). The RTMA covers the contiguous United States, Alaska, Hawaii, Puerto Rico, and Guam at hourly temporal scale. The spatial resolution of the product depends on the locations. The climatic data for Alaska are expressed at 3-kilometer resolution, while the data for the contiguous United States, Puerto Rico, Hawaii, and Guam are represented at 2.5-kilometer resolution.

The RTMA provides climatic data on temperature at 2-meter above ground, dew point temperature at 2-meter above ground, u-component of wind, v-component of wind,

specific humidity, wind direction, wind speed, visibility, and total cloud cover in addition to non-climatic data such as elevation and surface pressure. It is an hourly dataset and provides 24 readings on each variable per day. However, the readings are every hour on the hour; therefore, this datasets may not provide accurate daily minimum and maximum temperatures.

### **European Centre for Medium-Range Weather Forecasts (ECMWF)**

The European Centre for Medium-Range Weather Forecasts (ECMWF) is an independent intergovernmental organization that is constructed assembled by 34 European countries (ECMWF 2016). The ECMWF provides a number of climate datasets, such as short-term climate forecasts (e.g., 10-day, 15-day), seasonal climate forecasts (e.g., 7-month, seasonal), and historical climatology since 1979.

The ERA-Interim dataset is a reanalyzed climate data product covering the period from 1979 to present. The dataset provides raster format data on numerous parameters (e.g., temperature, precipitation, soil temperature, cloud cover, sunshine duration) at a 6-hour interval. Depending on the parameters, 3-hour interval climate data are also available.

This database includes both analysis and forecast data. The analysis data are generated based on observations at weather stations for 4 times a day (0, 6, 12, and 18 UTC). The forecast data are generated from analysis data at 0 and 12 UTC at 3-hour intervals for a continuous period of 10 days. Although a 10-day forecast is conducted twice a day, ECMWF only provides up to a range of 12-hour forecasts to the public.

The climatic data since January 1, 1979 to present (December 31, 2015 as of January 25, 2016) at 3-hour or 6-hour interval can be obtained through the ECMWF server;

however, the data may not represent the true daily values. For example, 2-meter temperature data can be obtained for 0, 3, 6, 9, 12, 15, 18, 21 hour UTC for every day from this dataset. First, only the temperature data at 0, 6, 12, and 18 are based on the actual observations. Second, the true maximum and minimum temperatures most likely do not happen at those four times. The database does provide maximum temperature and minimum temperature at 2-meter at 3-hour interval, but these data are based on the forecasts, not actual observations.

### **General Circulation Models (GCMs)**

General Circulation Models (GCMs) are models developed to represent physical processes in the atmosphere, ocean, cryosphere, and land surface (IPCC 2016b). The GCMs simulate the circulation of atmosphere responding to the increase of greenhouse gas concentrations to forecast long-term global climate conditions (IPCC 2016b).

GCMs are complex models representing three-dimensional analysis over the globe, typically having a horizontal resolution of 250 to 600 kilometers, 10 to 20 vertical layers in the atmosphere, and up to 30 layers in the oceans (IPCC 2016b).

The Intergovernmental Panel on Climate Change (IPCC) of the United Nations is the leading international organization that reviews and assesses climate change on a global scale. The IPCC reports the most recent scientific, technical, and socioeconomic information relevant to the understanding of climate change and provides currently available information and data (IPCC 2016a). There are numerous GCMs developed by organizations worldwide. The information and data of all GCMs are available at the IPCC website.

All GCMs are developed based on greenhouse gas emission scenarios. The emission scenarios describe future greenhouse gas levels in the atmosphere. The emission scenarios are constantly re-evaluated based on the latest greenhouse gas findings and are currently described as “Representative Concentration Pathways” (RCPs). There are 4 RCPs commonly used to describe future emission levels (RCP2.6, RCP4.5, RCP6, and RCP8).

### **North American Regional Climate Change Assessment Program (NARCCAP)**

The North American Regional Climate Change Assessment Program (NARCCAP) was initiated in 2006 to investigate uncertainties in simulation of future climate at a regional scale (Mearns et al. 2013). It is an international program to downscale well-known general circulation model (GCM) outputs spatially and temporally in order to understand future climate, climate change, and uncertainties associated with climate change at a regional scale (NARCCAP 2016). The NARCCAP products cover the United States, Canada, and northern Mexico (NARCCAP 2016).

The regional climate model is similar to global climate models or GCMs, but it covers smaller areas so that it can be run at finer spatial and temporal resolutions (Giorgi and Mearns 1999). GCMs are capable of forecasting long-term climate change effects; however, both the temporal and spatial resolution of GCMs are too coarse for short-term climate forecasts, landscape-scale ecological studies, forecasts of forest fires, forecasts for crop production, water resources, and forecasts for pests and diseases (Giorgi and Mearns 1999, Abatzoglou and Brown 2012). The regional climate models (RCMs) have become popular in recent years due to advances in technologies and understanding of the physical processes of the climate system (Christensen and Hewitson 2007).

The NARCCAP uses a dynamic downscaling method to run six RCMs driven by four atmosphere-ocean general circulation models (AOGCMs) over North America (Mearns et al. 2012). The NARCCAP selected AOGCMs with only one greenhouse gas emission scenario (SRES A2 emission scenario) to understand the uncertainty associated with downscaling from AOGCMs to RCMs (Mearns et al. 2012).

### **Multivariate Adaptive Constructed Analogs**

The Multivariate Adaptive Constructed Analogs (MACA) method is a statistical downscaling technique applied to GCM outputs to produce spatially and temporally sufficient future climate data for biological studies (Abatzoglou and Brown 2012). The MACA method utilizes a training dataset (e.g., weather station observation data) by eliminating historical bias and improving spatial pattern match for forecasted outputs (MACA 2016). This method was applied to 20 well-known GCM outputs derived from two pathway scenarios (RCP4.5 and RCP8.5). The MACA downscaled GCM outputs are daily climatic data on maximum temperature, minimum temperature, maximum daily relative humidity, minimum daily relative humidity, average daily specific humidity near surface, average daily precipitation, average daily downward shortwave radiation, average daily wind speed, average daily eastward component of wind (u-component), and average daily northward component of wind (v-component) (MACA 2016).

There are currently 3 datasets for MACA downscaled GCMs: MACAv1-METDATA, MACAv2-METDATA, and MACAv2-LIVNEH. The MACAv1-METDATA only covers the western United States, while MACAv2-METDATA and MACAv2-LIVNEH use a second version of the MECA method and cover the entire conterminous

United States (MACA 2016). The MACAv1-METDATA and MACAv2-METDATA are 4-kilometer resolution raster datasets that used METDATA as a training dataset, and the MACAv2-LIVNEH is a 6-kilometer resolution raster dataset with Livneh training data. The original observed data that were used as training data in the MACA project are available to the public and can be obtained at MACA project website (<http://maca.northwestknowledge.net/index.php>).

### **National Digital Forecast Database**

The National Digital Forecast Database (NDFD) contains seamless mosaics of digital weather forecasts from National Weather Service field offices and the National Centers for Environmental Prediction. The forecast data are generated and maintained by National Oceanic and Atmospheric Administration (NOAA), National Weather Service (NWS) (NWS 2016).

The NDFD covers the contiguous United States, Hawaii, Guam, Puerto Rico, the Virgin Islands, Alaska, and the North Pacific Ocean islands (NWS 2016). The resolution of the data depends on the coverage regions. The spatial resolution for the contiguous United States is at 5-kilometer, while the spatial resolution is 2.5 kilometers for Hawaii and Guam, 1.25 kilometers for Puerto Rico and the Virgin Islands, 6 kilometers for Alaska, and 10 kilometers for the North Pacific Ocean Domain. Although the forecasted weather information are available for Alaska, the data for Alaska are still considered to be at the experimental stage.

The database contains forecasted weather information on precipitation, temperature, dew point, hazards, relative humidity, wave height, sky cover, snow amount, wind direction,

wind gust, and wind speed. The projection interval and duration vary with parameters. The temperature, dew point, relative humidity, sky cover, wind direction, wind gust, and wind speed are forecasted for next 72 hours at 3-hour interval and for next 168 hours at 6-hour interval. The minimum and maximum temperatures are forecasted every 24 hours for a continuous period of 168 hours. The precipitation amount is forecasted for every 6 hours for a period of 72 hours.

### **Hurricane and Tropical Storm Data**

The National Hurricane Center (NHC) of the National Oceanic and Atmospheric Administration (NOAA) provides hurricane and tropical cyclone data in both pdf report format and GIS format dating back to 1958. Those data can be accessed at the NHC data archival website (<http://www.nhc.noaa.gov/data/>).

Unisys is a global information technology company (UNISYS 2016). Unisys collects weather data, mainly from NOAA, and offers weather information to the public. Global storm data are archived at the Unisys data center and can be downloaded from the data center. Unisys provides hurricane data since 1851 and typhoon data since 1945.

#### **3.3.1.3 Host Species (Plant) Data**

The main sources of information to support pest forecasting models as described here include four categories: data on the distribution of forest species, data on the distribution of row crops (mainly grains), data on the distribution of fruit crops and vegetables, and data on the distribution of plants and weeds. The list below describes commonly used databases that provide information about these four categories.

### **Forest Inventory and Analysis (FIA)**

The Forest Inventory and Analysis (FIA) program is operated and managed by the United States Department of Agriculture, Forest Service (USFS). The FIA program was established some 80 years ago to understand extent, condition, volume, growth and timber use on the forest land in the United States (O'Connell et al. 2015). The program covers both private and public lands such as National Parks, National Forests, and reserved areas. Currently, the FIA program is managed by four regional units (Northern, Southern, Interior West and Pacific Northwest) that are coordinated by a National Office in Washington DC.

The FIA program collects forest data through three phases: Phase 1, remote sensing; Phase 2, ground sampling; and Phase 3, Forest Health Monitoring (FHM) plots. The main purpose of Phase 1 data collection is to classify lands by using remote sensing for land stratification. Through the phase 1 process, a number of points are generated across the United States with land classification information. A subset of these points are selected for field data collection (Phase 2) (O'Connell et al. 2015).

The Phase 2 ground sampling plots represent approximately 1 sample location per 6,000 acres. During Phase 2, forest attributes such as forest type, site attributes, tree species, tree size, and tree conditions are collected (Burkman 2005).

A subset of Phase 2 sampling plots are selected for the purpose of investigating forest health conditions (Phase 3). The Phase 3 plots represent roughly 1 sample location per 90,000 acres. During Phase 3, information about ozone, lichen, soil conditions, and vegetation structure are collected (USFS 2014). The ozone bioindicator data are used to estimate air quality, tree growth effects caused by air pollution, ozone tolerant/intolerant species, and percentage of forests damaged by air pollutants (USFS 2016b).

The FIA program also collects information on epiphytic lichens as one of the forest health indicators. Lichens are classified as fungi and are symbiotic with algae. Lichens are sensitive to air pollutants; therefore, the information about lichens in forests are used to understand changes in forest structure, air quality, and climate (USFS 2016a).

Soil information is collected to evaluate soil physical and chemical properties and the extent of erosion and compaction (USFS 2016c). Those studies are used to understand forest productivity, impact on soils caused by pollution, carbon storage, and erosion (USFS 2016c). Another key category of data that the FIA program collects during Phase 3 are attributes of vegetation structure and biodiversity. The program obtains information on type, abundance, and spatial arrangement of all trees, shrubs, herbs, grass, and fern allies in a sampling plot (USFS 2016d).

The FIA database provides essential U.S. forest information to understand diversity of tree species, types and abundance of tree species, annual tree growth and mortality by species, tree condition, and land ownership. The FIA program also provides forest species information in urban areas (Urban Forest Inventory and Analysis Database: UFIADB) although this is still a pilot program.

### **Forest Biomass**

The United States Department of Agriculture, Forest Service (USFS) has estimated forest biomass across the contiguous United States, Alaska, and Puerto Rico. A forest biomass raster surface was modeled from biomass information collected at each FIA sample plot as functions of more than sixty geospatially predictive data, such as digital elevation models (DEMs), DEM derivatives, Moderate Resolution Spectroradiometer (MODIS)

multi-date composite, vegetation indices, National Land Cover Dataset (NLCD), ecological zones, PRISM climate data, and so on

(<http://data.fs.usda.gov/geodata/rastergateway/biomass/>).

The USFS provides forest biomass information as mg/ha unit at 250-meter resolution. In addition, information on errors associated with forest biomass are available as raster layers.

### **Forest Types**

The United States Department of Agriculture, Forest Service (USFS) provides forest type coverage information as a raster dataset. The national forest type raster dataset was developed by the USFS Forest Inventory and Analysis (FIA) program and the Remote Sensing Applications Center (RSAC) (USFS 2016e). This dataset includes forest coverage for all 50 states, Puerto Rico, and U.S. Virgin Islands at 250-meter resolution in Lambert Azimuthal equal area projection (USFS 2016e). The national forest type dataset was created by combining information from multiple data sources, such as FIA plot data, digital elevation models, Moderate Resolution Spectroradiometer (MODIS) multi-date composites, vegetation indices and vegetation continuous fields, National Land Cover Dataset, ecological zones, and climate data (USFS 2016e).

### **PLANTS Database**

The PLANTS Database, developed and managed by USDA Natural Resources Conservation Service (USDA NRCS), provides information about the vascular plants, mosses, liverworts, lichens of the United States and territories (<http://plants.usda.gov/java/>). The database includes common name, scientific name, taxonomic information, distribution

at state or county level, photos, legal status (e.g., threatened and endangered), indigenous status (e.g., indigenous or introduced), and literature information. The information about plants occurring in the United States and its territories can be obtained by searching with common name, scientific name, or PLANTS designated codes.

### **National Agricultural Statistics Service (NASS) and Cropland Database**

The USDA National Agricultural Statistics Service provides timely, accurate, and useful statistics (e.g., productions, supplies, and prices) about U.S. agriculture by conducting surveys throughout the United States (<http://www.nass.usda.gov/>). NASS datasets provide monthly and annual data on prices and production (acres and volumes) regarding animal products, livestock, fruits, vegetables, and nursery stock at national and state levels. NASS also conducts an AgCensus every 4 years to report detailed information about U.S. agricultural statistics at the county level. These tabular data are available from the NASS website.

NASS also provides a Cropland Data Layer (CDL) for every year dating back to 1997. The CDL covers the contiguous United States and is raster formatted dataset that lists types of cropland at 30-meter resolution. The CDL is generated from satellite imagery and updated each year. The data from 1997 to 2014 are available through the CropScape website (<http://nassgeodata.gmu.edu/CropScape/>). Users are able to download the seamless CDL data for a specified area as well as statewide CDL data. The projection for the seamless CDL dataset is an Albers conic equal area. Users are also able to download the data in KML format, which is compatible with Google Earth, Google Maps, and Google Maps for Mobile (NASS 2016).

## **Normalized Difference Vegetation Index (NDVI)**

The Normalized Difference Vegetation Index (NDVI) is generated to understand vegetation density on land. The density of the plant is determined by the reflections of near-infrared light by remote sensing measurements (Anonymous 2016a). This index is the most widely used index to understand live vegetation density and availability on earth (Gao 1996).

The Advanced Very High Resolution Radiometer (AVHRR) is a radiation-detection imager that is used to remotely determine cloud cover and surface temperature (Robel and Graumann 2014). The AVHRR instrument is maintained and operated by NOAA and currently has 5 detectors, ranging in wavelength (micrometers) from 0.58 to 12.5 (Robel and Graumann 2014). Among those 5 detectors, two are used to develop NDVI (wavelengths of 0.55-0.70 micrometers and 0.73-1 micrometers).

The NDVI is calculated as the difference between near-infrared light and visible light divided by the sum of near-infrared light and visible light ( $NDVI = (NIR - VIS)/(NIR + VIS)$ ) (Anonymous 2016a). A zero means no vegetation, while value of 1 indicates the highest possible density of green leaves (Anonymous 2016a).

The AVHRR NDVI data are available at the USGS website (<https://lta.cr.usgs.gov/ndvi>). The AVHRR NDVI data for conterminous United States and Alaska are available from January, 1989 to present at week and bi-week temporal resolutions and at a 1-kilometer spatial resolution. The AVHRR NDVI dataset is also available at global scale. The global data are available from April, 1992 to May, 1996 as 10-day composited AVHRR data at a 1-kilometer spatial resolution.

The Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI products provide information on vegetation greenness, a composite property of leaf area, chlorophyll, and canopy structure at a global scale (Didan 2016). The temporal and spatial resolutions of the data depends on the type of datasets. The global MODIS NDVI products are available since February, 2000. The temporal resolution ranges from 16-day to monthly. The available spatial resolutions are 250-meter, 500-meter, and 1-kilometer. The MODIS NDVI data are available at NASA website (<http://modis-land.gsfc.nasa.gov/vi.html>).

#### **3.3.1.4 Land data**

The land feature data support pest forecasting models by providing land types and terrain information. These information are critical and useful because some pests are known to inhabit certain features of land. The list below describes commonly used databases that are useful for understanding land data for pest forecasting models.

##### **Digital Elevation Model (DEM)**

A digital elevation model (DEM) is a numeric representation of terrain surface (Ocallaghan and Mark 1984). United States Geological Survey (USGS) has developed several DEMs. The DEM SRTM (Shuttle Radar Topography Mission) was developed by the National Geospatial-Intelligence Agency (NGA) (Department of Defense) and the National Aeronautics and Space Administration (NASA). The SRTM was developed through the use of radar interferometry. The SRTM does not cover the entire earth (80% of the earth surface: between 60 degree north and 56 degree south latitude). It is raster gridded dataset at a 30-meter resolution. The data are now available through the USGS EROS Data Center (<https://lta.cr.usgs.gov/SRTM1Arc>).

Another notable DEM dataset is the USGS 3D Elevation Program (3DEP) (USGS 2015). The 3DEP covers the conterminous United States, Hawaii, and the U.S. territories. 3DEP provides multiple spatial resolutions (0.5-meter, 1-meter, 3-meter, 5-meter, 10-meter, and 30-meter), depending on location.

The main difference between SRTM and 3DEP is the source data used to develop these datasets. The 3DEP DEMs are processed to represent bare earth ground by using historical photogrammetry, aerial data, digitized contours, Lidar, and interferometric synthetic aperture radar (IfSAR) data since 1924, while SRTM was developed exclusively from IfSAR data. The USGS has also stated that the quality of the 3DEP dataset is better than the SRTM dataset. The SRTM vertical errors can be up to 10-meter root mean square error (RMSE), whereas the vertical errors of 3DEP dataset are 1.55-meter RMSE.

### **Digital Surface Model (DSM)**

Digital surface model (DSM) data are digital representations of the Earth's surface, including buildings, trees, and any objects above the ground. DSMs are often generated from Lidar point cloud data. Lidar point cloud data are available from USGS, Center for LIDAR information Coordination and Knowledge (CLICK) website (<http://lidar.cr.usgs.gov/>).

The advantage of DSM data includes object height estimation (the difference between DEM and DSM). The disadvantage is that DSM data are usually not available in raster format. Users have different objectives (e.g., obtaining information for the highest objects, or surface information for understory vegetation); therefore, it is usually available in Lidar point cloud format.

Users are required to convert from las format (Lidar original data format) to text file, text file to point data, select the elevation values based on the objective, then convert the point data to a raster dataset using spatial interpolation methods.

### **National Land Cover Database (NLCD)**

The USGS EROS Center (Earth Resources Observation and Science) provides U.S. land cover data from 1992, 2001, 2006, and 2011. The National Land Cover Database 2011 (NLCD 2011) developed by the Multi-Resolution Land Characteristics (MRLC) Consortium is the latest dataset that describes land cover types across the contiguous United States.

Land cover is categorized into 16 classes and represented at 30-meter resolutions (Homer et al. 2015). The National Land Cover Database 2006 (NLCD 2006) is a 16-class land cover classification scheme that has been applied across the conterminous United States (Fry et al. 2011). The spatial resolution of the NLCD 2006 is 30 meters. The National Land Cover Database 2001 (NLCD 2001) is also a 16-class land cover classification, but it includes all 50 United States and Puerto Rico (Homer et al. 2007). The classification for Alaska is 20-class instead of 16-class. The resolution of the datasets is 30 meters. The National Land Cover Dataset 1992 (NLCD 1992) is a 21-class land cover classification scheme that has been applied consistently across the lower 48 United States at 30-meter resolution (Vogelman et al. 2001).

### **Global Land Cover Characteristics (GLCC) (USGS)**

The Global Land Cover Characteristics (GLCC) database was developed by U.S. Geological Survey (USGS), University of Nebraska-Lincoln, and the Joint Research Centre of the European Commission (Loveland et al. 2000). The 1-kilometer resolution GLCC

data were generated based on the monthly Advanced Very High Resolution Radiometer (AVHRR) Normalized Difference Vegetation Index (NDVI) data at 1-kilometer resolution from April, 1992 to March, 1993.

The GLCC program classified the global landscape into 17 general land cover classes (IGBP-DIS: International Geosphere-Biosphere Programme, Data and Information Systems). This information was used to further characterize the seasonal land cover regions into 20 general land cover types. The GLCC provides IGBP DIS cover and all digital resources that were used in the classification at the USGS GLCC website (<http://edc2.usgs.gov/glcc/glcc.php>).

### **Soil Survey Geographic Database (SSURGO)**

The Soil Survey Geographic (SSURGO) Database, produced and maintained by USDA Natural Resources Conservation Service (NRCS), contains soil information collected by the National Cooperative Soil Survey since 1899 (NRCS 2016).

The National Cooperative Soil Survey is a nationwide survey effort and collects information on soil types, soil classifications, slopes, and soil erosions. Soil data covering the contiguous United States, Alaska, Hawaii, Pacific Islands (American Samoa, Federal States of Micronesia, Guam, Marshall Island, Palau, and Northern Mariana Islands), Puerto Rico, and U.S. Virgin Islands are available at the Web Soil Survey website (<http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx>).

#### **3.3.1.5 Human activity**

Human activities often facilitate new pest introductions and spread into new environments. There are several databases that are useful to evaluate likelihood,

establishment, and spread of alien species due to human activities. The main sources that provide human activity information include four categories: databases for trade, databases that describe transportation, databases that document international travelers' data, and databases that contain U.S. demographic information. This information is often used in pest risk assessments in order to identify pathways of alien species from infested areas to non-infested areas by evaluating potential movements of particular commodities based on trade patterns, transportation patterns, and consumer demands (e.g., demographic information).

### **Databases for Trade Volume**

Trade volumes from foreign countries to the United States and from United States to foreign countries can be obtained from several databases. I list key trade databases to understand overall trade patterns and agricultural commodity trade trends.

#### **World Trade Organization Time Series**

The World Trade Organization (WTO) provides annual import and export value information from all countries in the world. The WTO Time Series database contains trade information (total of all merchandise) since 1948, but it also contains trade information by commodity types (e.g., agricultural products, fuels and mining products, textiles, chemicals, and machinery and transport equipment) since 1980.

#### **FAOSTAT**

The FAOSTAT by the Food and Agriculture Organization of the United Nations provides import and export volumes and values of agricultural commodities from each country in the world since 1961 (<http://faostat3.fao.org/home/E>). Unlike the WTO Time Series database that categorized agricultural products into two categories (food and non-

food), the FAOSTAT database has trade information by almost all major products (e.g., cotton, wheat, banana, cherries, almond, milk, cheese).

### **Global Agricultural Trade System (GATS)**

The Foreign Agricultural Service (FAS) of United States Department of Agriculture provides the Global Agricultural Trade System that contains U.S. agricultural commodity trade information since 1967. This system contains both values and volumes of agricultural commodities and contains detailed information about trading partners. Users are able to obtain import and export information on a specific commodity to a specific country from the United States and from a specific country to the United States.

These three databases offer valuable information and are useful in understanding world and U.S. trade patterns. However, none of the databases specify the export and import ports; in other words, these databases only list trade information between countries.

### **PPQ280**

The USDA Animal and Plant Health Inspection Service (APHIS) maintains several databases associated with trade. The database called PPQ280 contains agricultural commodity imports data from foreign countries. This database contains only agricultural commodities; however, the advantage is that it provides trade volume by country, date, and entry ports.

### **Databases for Transportation**

#### **Freight Analysis Framework (FAF)**

The Freight Analysis Framework (FAF) database is managed by the Center for Transportation Analysis of Oak Ridge National Laboratory through a partnership between

the Bureau of Transportation Statistics (BTS) (United States Department of Transportation: DOT) and the Federal Highway Administration (FHWA) (DOT) ([http://www.ops.fhwa.dot.gov/freight/freight\\_analysis/faf/](http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/)). FAF provides freight movement information among states and major metropolitan areas by all modes of transportation (FHWA 2015). This database provides somewhat limited origin and destination information and does not specify an exact route between locations. However, it does provide the volume flows by commodity category and transportation mode between two locations. The newest version of the FAF database (FAF4) contains the Commodity Flow Survey (CFS) and international trade information and enables users to specify commodity types (e.g., agricultural commodity).

### **Lloyd's List Intelligence**

The transportation database known as Lloyd's List Intelligence (Informa Plc) provides vessel route information. It provides port locations where each vessel has stopped and future anticipated ports. Although this database does not have any information about the types of commodities being transported, it is useful in understanding the number of vessels going to specific ports from specific locations. This is not open source data and requires a subscription to access the database (PLC 2016).

### **Databases for Traveler Information**

#### **World Development Indicators (The World Bank)**

The World Bank generates development indicators at national, regional, and global levels to present the most current and accurate global development data on an annual basis since 1960 (The World Bank 2016). These indicators describe conditions of agriculture and

rural development, aid effectiveness, climate change, economy and growth, education, energy and mining, environment, finance, demographic, health, infrastructure, labor and social protection, poverty, private and public sectors, science and technology, social development, trade, and urban development (The World Bank 2016).

There are indicators that describe annual international travelers (inbound and outbound travelers) by each country since 1995 in the World Development Indicators report. This information is particularly useful in understanding the number of travelers who are coming into and going out of the United States, the growth of international travel, and international travel patterns and trends for each country.

### **U.S. Travel and Tourism Statistics**

The Office of Travel and Tourism Industries (OTTI) is part of the International Trade Administration of the United States Department of Commerce. Their mission is to enhance the international competitiveness of the U.S. travel and tourism industry and increase its exports, thereby creating U.S. employment and economic growth (OTTI 2016). The OTTI provides travel and tourism datasets that contain statistics on international travel to and from the United States (OTTI 2016).

The datasets provide travelers' information (inbound and outbound) by month, year, selected countries, and regions (Africa, Asia, Caribbean, Central America, Europe, Middle East, Oceania, and South America). The selected countries change each year depending on the number of passengers. For example, the selected countries for 2014 data are Argentina, Australia, Brazil, Canada, Chile, China, Colombia, France, Germany, India, Ireland, Italy,

Japan, Mexico (as air passenger), Netherlands, South Korea, Singapore, Spain, Sweden, Switzerland, Taiwan, and United Kingdom for inbound travelers.

## **Demographic Information**

### **U.S. Census**

The United States Department of Commerce, United States Census Bureau (USCB) serves as the leading source of data about U.S. demography and demography-associated and economic characteristics (USCB 2016). The USCB conducts several surveys including a decennial census. The decennial census information is summarized at nation, state, county, zip-code, and census block levels and is available at USCB American FactFinder website (<http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>).

#### **3.3.2 How SAFARIS supports key analysis categories and post processing for pest forecast models**

The pest risk assessment mainly consists of four types of analyses; 1) suitability analysis, 2) pathway (spread) analysis, 3) population dynamics, 4) economic impact analysis, and 5) uncertainty analysis. This is not considered an exhaustive listing of all possible kinds of analyses supported by pest forecasting. However, I use these categories because they relate to some very common applications in large scale management and because they can easily be expanded to accommodate related analyses.

Pest biology is key in each analysis and the input drivers to model each component share the core drivers, but also require unique, different drivers. This section illustrates the shared and unique needs to support analytical types.

### **3.3.2.1 Environmental Suitability Analysis Approach**

Environmental suitability models are used to determine suitable geographic areas for pest establishment by evaluating climatically suitable conditions for given pest's growth and adding supplemental information such as host species distribution, host species availability, and land features. Overall, a basic epidemiological model underlies this approach. That is, the approach focuses on climate (environment), suitable host availability, and pest presence. Those are the three most basic elements of epidemiology: suitable host, suitable climate, and viable pest introduction. Figure 3-2 below describes environmental suitability analysis approaches.

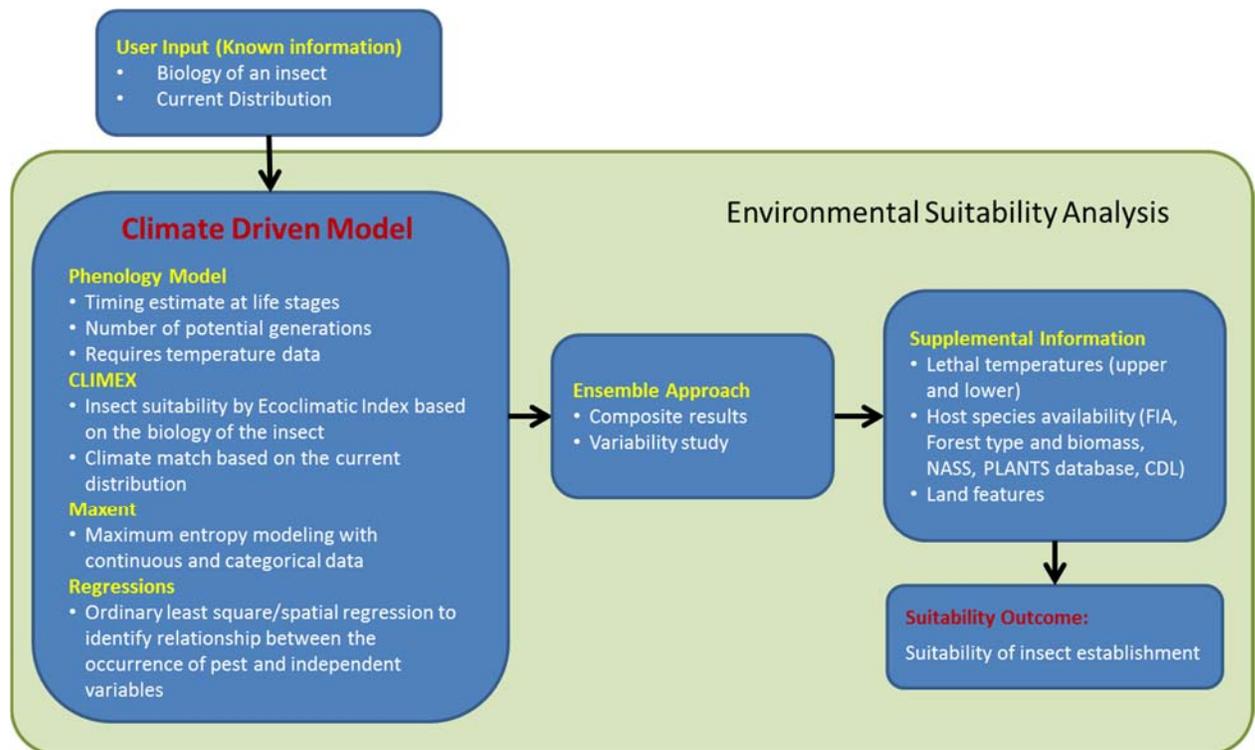
Species distribution models and habitat suitability models are commonly used to determine potential distribution and habitat niche of a pest based on suitability derived from pest biology and/or current distribution (Venette and Cohen 2006, Jarnevich and Reynolds 2011, Senay et al. 2013, Springer et al. 2015). Commonly used approaches to determine environmental suitability are mathematical relationships between physiological conditions and current distributions (e.g., MaxEnt, Generalized Linear Model, and Boosted Regression Trees) and biological simulations based on the required biological conditions for the pest (e.g., CLIMEX).

To use the environmental suitability approach in SAFARIS, the process begins with the user entering known biological information such as data from life table studies and current distribution to estimate pest suitable areas for establishment. Commonly used models in suitability analysis include phenology models, match climate models (e.g., CLIMEX), and pattern matching and fitness driven models (e.g., MaxEnt, Generalized

Linear Model, Random Forest). In addition, spatial tools to identify the areas for specific conditions are often used to eliminate areas unsuitable for long-term pest establishment. In particular, those tools are commonly used to identify unsuitable areas for winter months based on pest lethal temperatures, or for summer months based on pest upper lethal temperatures, and to determine suitable areas based on certain precipitation accumulation levels for pest growth.

Additional pest environmental requirements/conditions for growth such as host availability, host distribution, soil types, and elevation requirements can be integrated with climatologically suitable areas to improve the estimate of environmental suitability for a given location for pest establishment.

The data required to analyze the suitability of an environment for pest establishment depends on the nature of the pests. Climate data appropriate for suitability analysis include weather station data (National Climate Data Center), Climate Research Unit, PRISM, DAYMET, TopoWX, METDATA, and NLDAS-2. Additional information often used in suitability analysis are Forest Inventory and Analysis, Forest Biomass, Forest Types, PLANTS Database, Cropland Data Layer, National Land Cover Database, Digital Elevation Model, and Normalized Difference Vegetation Index.



**Figure 3-2. Environmental suitability analysis approaches**

### 3.3.2.2 Pathway and Spread Analysis Approaches

Pest pathways can be categorized into two groups, natural spread and human-assisted spread. The natural spread pathway is driven by a species' own ability to spread or by abiotic factors not controlled by humans (e.g., wind, rain, ocean currents, rivers, animal-assisted spread). Pests can often be dispersed with assistance of weather events (e.g., hurricane, rain, wind). In addition to natural dispersal, pests can be transported with human activities or combinations of the two factors.

Pathway analysis approaches include 1) evaluating the potential pathways from foreign countries to the United States via human assisted activities or climate events like hurricanes, 2) examining potential long distance movements within the United States, and 3) investigating natural dispersal means based on pest biology and climate conditions.

The potential for transportation of pests from foreign countries varies with pest type and biology (Hulme 2009). Some pests are known to be transported over long distances with infested fruits and vegetables. For example, fruit flies are commonly moved with infested crops. The Mediterranean fruit fly (*Ceratitidis capitata* (Wiedemann)) and Oriental fruit fly (*Bactrocera dorsalis* (Hendel)) are polyphagous and known to attack more than 300 host species (mainly fruits and vegetables) (White and Elson-Harris 1992). Those pests pose a threat if they are transported to uninfested areas in the United States through agricultural commodity trade.

Previously, USDA APHIS had focused on fruit, vegetable, cut flowers, live plants, and log trade for pest long distance movements (Liebhold et al. 2012); however, introductions of wood boring insects (*Anoplophora glabripennis* (Motschulsky) (Asian longhorned beetle), *Tomicus piniperda* (Linnaeus) (pine shoot beetle), and *Agrilus planipennis* Fairmaire (emerald ash bore)) proved that the use of solid wood packing materials as part of international commerce was also a risk factor (Poland and McCullough 2006, Haack et al. 2010).

In addition, there is a so-called “hitch-hiker” pathway constituted by pests that are not associated with a commodity but are present in cargo. Some examples are hitchhiking beetles that fall on cargo when crates are loaded at night. When lights attract some beetles (e.g. Latridiid beetles) as well as wind-blown insects, those species contaminate cargos and vessels at docks. *Lymantria dispar* Linnaeus (Asian gypsy moth (AGM)) is one of the hitch-hikers that poses a great threat to the U.S. forests. Usually egg masses are transported when gravid females oviposit resilient egg masses onto the surface of vessels or

containerized cargo from infested areas overseas (hitch-hiking on vessels) to U.S. port environs or inland areas upon arrival.

Some pests can also travel long distances by their own abilities to spread. In addition, some insects have abilities to fly with the assistance of natural events, such as rain, wind, and hurricane. Grasshoppers are highly mobile insects, and swarms can fly more than 50 kilometers in one night (Anonymous 1994). African locusts (*Schistocerca gregaria*) have been reported to travel long distances (~5,000 kilometers) with the aid of hurricanes (USAID 1989, Richardson and Nemeth 1991). Another example is the Old World bollworm (*Helicoverpa armigera* (Hübner)), which has been reported to migrate 250 km per night (Feng et al. 2010).

Once alien pests arrive in the United States, some pests still continue to travel with the aid of human assisted activities or with their dispersal capacity. Domestic movements of exotic pests are often evaluated based on freight transportation (Colunga-Garcia et al. 2009).

Climatic events and conditions also play important roles for pest dispersal at both regional and local scales. Depending on the population density and food availability, the first instar larvae of AGM can balloon several kilometers with wind (Zlotina et al. 1999, Fowler et al. 2009). The female moth of AGM is also known to fly and is reported to fly up to 3.5 kilometers (Baranchikov 1988).

SAFARIS provides mainly three modules for pathway and spread analysis approaches (Figure 3-3). Those modules are a module to evaluate pathways from foreign countries, a module to evaluate domestic pathways due to human activities, and a module to examine pathways and dispersal means with climatic events and conditions. A user begins

with known biological information such as global distributions, dispersal means, and climate conditions that assist pest movements. A user can choose the appropriate module(s) based on pest biology and data availability.

Pathways from foreign countries are evaluated mainly based on vessel movements, trade data, and international travelers' information. The Lloyd's List Intelligence dataset provides vessel movement information on a global scale and useful information to understand how many vessels are destined for U.S. ports from certain areas. The trade information provides values and volumes at the country level, as well as origins and destinations, types of commodity traded, and the U.S. ports receiving particular commodities. These information can be obtained from databases such as FAOSTAT, WTO Time Series, PPQ 280, and GATS.

Domestic movements due to human assisted activities are evaluated based on freight movements. The FAF datasets provide traffic flow at states and metropolitan levels.

Pathways and dispersal due to climate events and conditions are evaluated by using climatic datasets. As for hurricane or typhoon-borne pests, the potential pathways can be tracked and modeled based on the hurricane paths that are available from Unisys Weather or NHC databases. Other climatic data such as temperature, precipitation, and relative humidity that are obtained from NCDC, CRU, PRISM, DAYMET, TopoWX, METDATA, and NLDAS-2, are commonly used to parameterize pest dispersal aided by climate conditions.

A user can combine results from each module to understand the overall pathway and spread risks. For instance, if a user is examining potential detections of AGM in the

continental United States, he/she can select the module to evaluate movement from foreign countries and the module to evaluate domestic movement, and then combine the results to understand where AGM could be transported from China into the continental United States due to human activities.

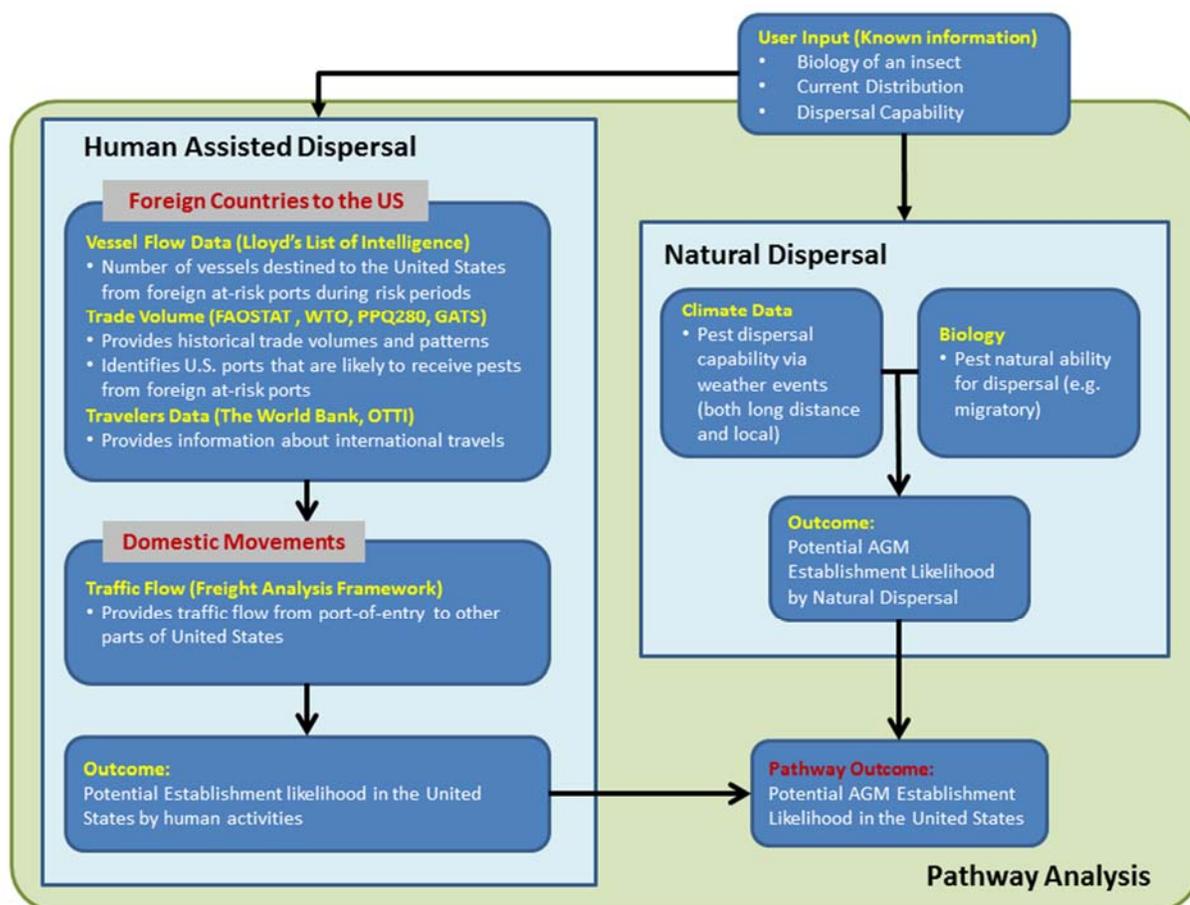


Figure 3-3. Pathways and dispersal means of alien species from infested areas to new colonization sites

### 3.3.2.3 Population Dynamics Analytical Approach

Population dynamics model analyses examine changes in population dynamics over time (Liebhold and Bascombe 2003). Alien pests may be introduced via human-assisted activities, natural events, or pests' ability to disperse and establish in new environments;

however, populations of a pest may not grow due to several reasons such as natural enemies, host conditions, and seasonal effects.

There have been many insect population dynamics studies (Gutierrez et al. 1984, Liebhold and Bascompte 2003, Liebhold et al. 2004, Gutierrez et al. 2010). Insect populations are influenced by many factors such as Allee effects, predator and enemy populations, host species, and weather (Wallner 1987, Elkinton and Liebhold 1990). Yet, many insect models focus on births, deaths, immigration, and emigration and do not consider secondary ecological factors, such as weather conditions. This may be because most population dynamics models are location-specific; therefore, mortality related to climate conditions is already factored into the death totals, assuming climate conditions stay relatively similar during the study or prediction period.

USDA APHIS evaluates pest forecasts on a nationwide scale as well as global scale to protect U.S. agriculture and natural resources. Population fluctuations need to be evaluated and understood for large areas; therefore, the population dynamics model needs to be a function of climate.

Insect population growth for one generation in a closed system (no immigration and emigration) is described in Figure 3-4. In a closed system, population grows when a female produces offspring, and the population decreases through mortality associated with each life stage. Those deaths can be due to natural mortality, natural enemies, host conditions, human activities, and climate conditions. In addition, reproduction is affected by host species, host conditions, climate conditions, natural enemies, and genetics.

The general population dynamics model for the  $i^{\text{th}}$  age class of a population is expressed as a following equation (Gutierrez et al. 2010).

$$\frac{dN_i}{dt} = \frac{k\Delta x}{\Delta} [N_{i-1}(t) - N_i(t)] - \mu_i(t)N_i(t)$$

Where

$N_i$  = abundance of the  $i^{\text{th}}$  class age,

$dt$  = change in time

$k$  = number of age classes

$\Delta$  = expected mean developmental time

$\Delta x$  = daily increment of age

$\mu_i(t)$  = mortality of the  $i^{\text{th}}$  class age

The developmental time, daily increment of age, and mortality are influenced by climate, so those rates are calculated as a function of pest biology and climate data at each insect stage.

Population dynamics analyses evaluate the reproduction rates and the factors affecting mortality rates to examine if particular areas are suitable for population growth. The reproduction rates, growth rates, and mortality rates are modeled with data on climate, soil, host, natural enemies, and land features (Figure 3-4). This analysis is important because particular areas may be suitable for pest establishment, but the environment may not be sufficient, in a practical sense, to support viable pest populations over the long term.

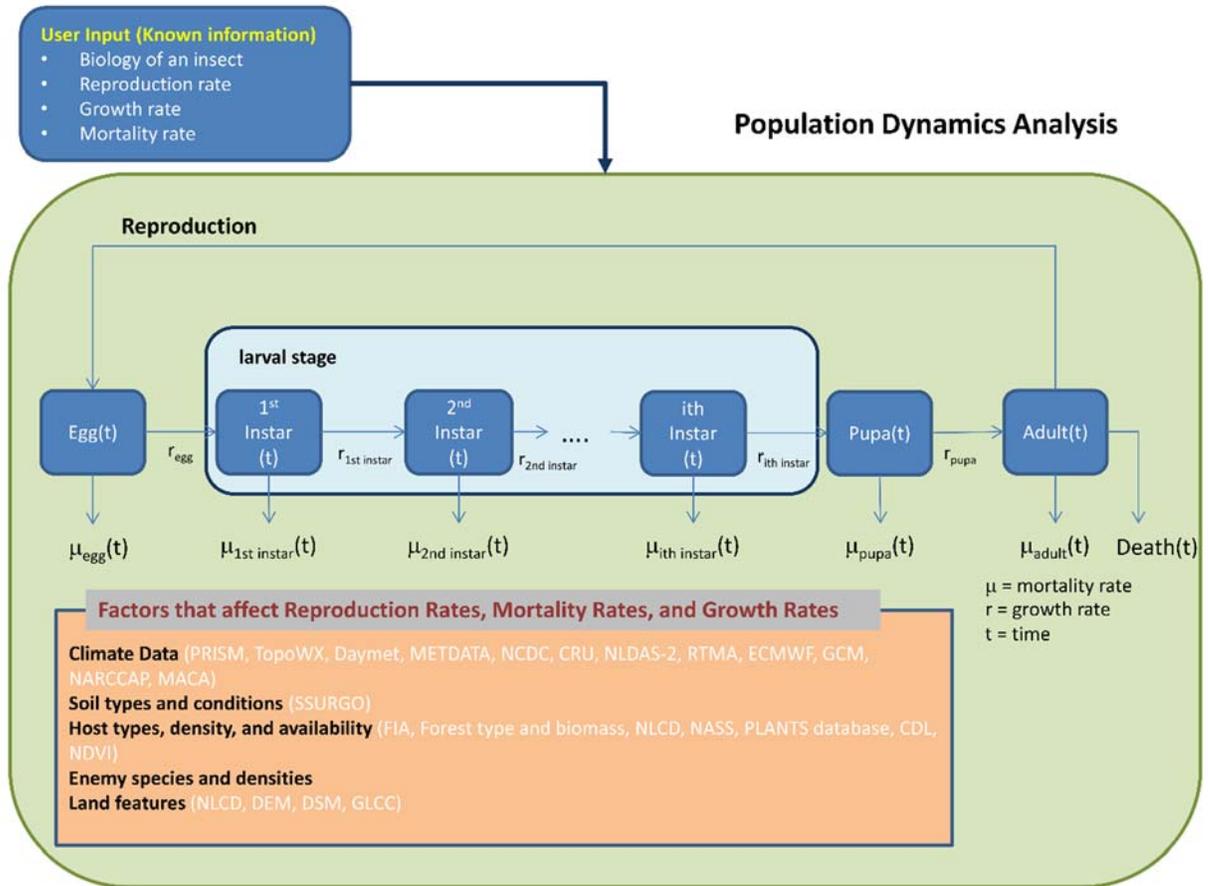


Figure 3-4. Population dynamics analysis in a closed system (no immigrants and emigrants)

### 3.3.2.4 Economic Analysis Approaches

Economic impact analysis evaluates the potential economic consequences over time once a pest is established in a new environment (International Standards for Phytosanitary Measures: ISPM 11) (IPPC 2004). The ISPM 11, which describes pest risk analysis for quarantine pests, lists three techniques for economic assessments: the partial budgeting technique, the partial equilibrium analysis technique, and the computable general equilibrium analysis technique (IPPC 2004, Soliman et al. 2010). These three techniques address three different levels of economic assessments (Figure 3-5) and focus on three different scales. Partial budgeting mainly focuses on farm or the small region scale, while

the partial equilibrium and general equilibrium approaches operates at regional and national/global levels. Also, complexity varies among the three techniques.

The partial budgeting technique evaluates net change in profit by identifying additional costs and reduced revenues as well as additional revenues and avoided costs due to a new pest introduction at the farm level (Soliman et al. 2010). This is the simplest, most commonly used, and transparent economic assessment in pest risk analysis (Trumble and Morse 1993, James and Anderson 1998, MacLeod et al. 2004, Soliman et al. 2010). The USDA APHIS used a partial budgeting approach to evaluate emerald ash borer (*Agilus planipennis* Faimaire) and pine shoot beetle (*Tomicus piniperda* (Linnaeus)) domestic pest control programs (Fowler et al. 2014, Takeuchi et al. 2015). The ISPM 11 reports that the partial budgeting technique is adequate if the economic impacts are minimal and generally limited to the producer level (IPPC 2004).

The partial equilibrium technique examines economic impacts at a broader level than partial budgeting, and IPPC (2004) recommends this technique if there are significant impacts on producer profits, or consumer demand, or both of them. In addition, the partial equilibrium technique evaluates economic effects caused by an introduction of a new pest or new policy implementations at the market level (Soliman et al. 2010).

The partial equilibrium model calculates social welfare by summing consumer surplus and producer surplus derived from supply and demand curves (Soliman et al. 2010). The shifts in supply and demand equilibrium caused by a pest introduction changes consumer surplus and producer surplus. The partial equilibrium assessment determines the difference in social welfare before and after a pest introduction (Soliman et al. 2010).

A number of studies have evaluated economic impacts caused by quarantine pests by using partial equilibrium models (Breukers et al. 2008, Surkov et al. 2009, Cook et al. 2011). However, parameterization and data collection needed for partial equilibrium models can be challenging because parameterization requires data on market impacts such as price elasticity and quantity changes for both supply and demand (Baker et al. 2009, Soliman et al. 2010).

The general equilibrium technique examines economic impacts at a national level that could alter factors such as wages, interest rates, or exchange rates (IPPC 2004). Computable general equilibrium models examine the economic impacts and changes caused by a new pest introduction and policy implementations for multiple markets. Computable general equilibrium models are highly complex because they incorporate factors such as impacts on host species as well as non-host species, impacts for multiple regions, loss of jobs and declining investment in particular industries due to quarantine pests, price changes, and supply and demand changes (Wittwer et al. 2005, Soliman et al. 2010).

I have used SAFARIS to address economic analyses and summarize the approach here. The first phase is to provide the economic model for a yearly impact function based on the annual spread predicted for a pest. USDA APHIS estimated the annual dispersal rate of the emerald ash borer (EAB) (*Agilus planipennis* Faimaire) based on 13 years of field detection records and applied the spread rate to non-quarantine or non-established areas to examine how many years EAB would take to spread across the entire contiguous United States (Takeuchi et al. 2015). Then, the potential annual damage to host species (i.e., *Fraxinus* species) was evaluated based on host species availability and dispersal rate

(Takeuchi et al. 2015). These biology and host information in SAFARIS and economic analysis products can support all three approaches (partial budgeting, partial equilibrium, and general equilibrium). The actual example is not discussed in this section but is used to assess the long-term impact of emerald ash borer over the entire United States.

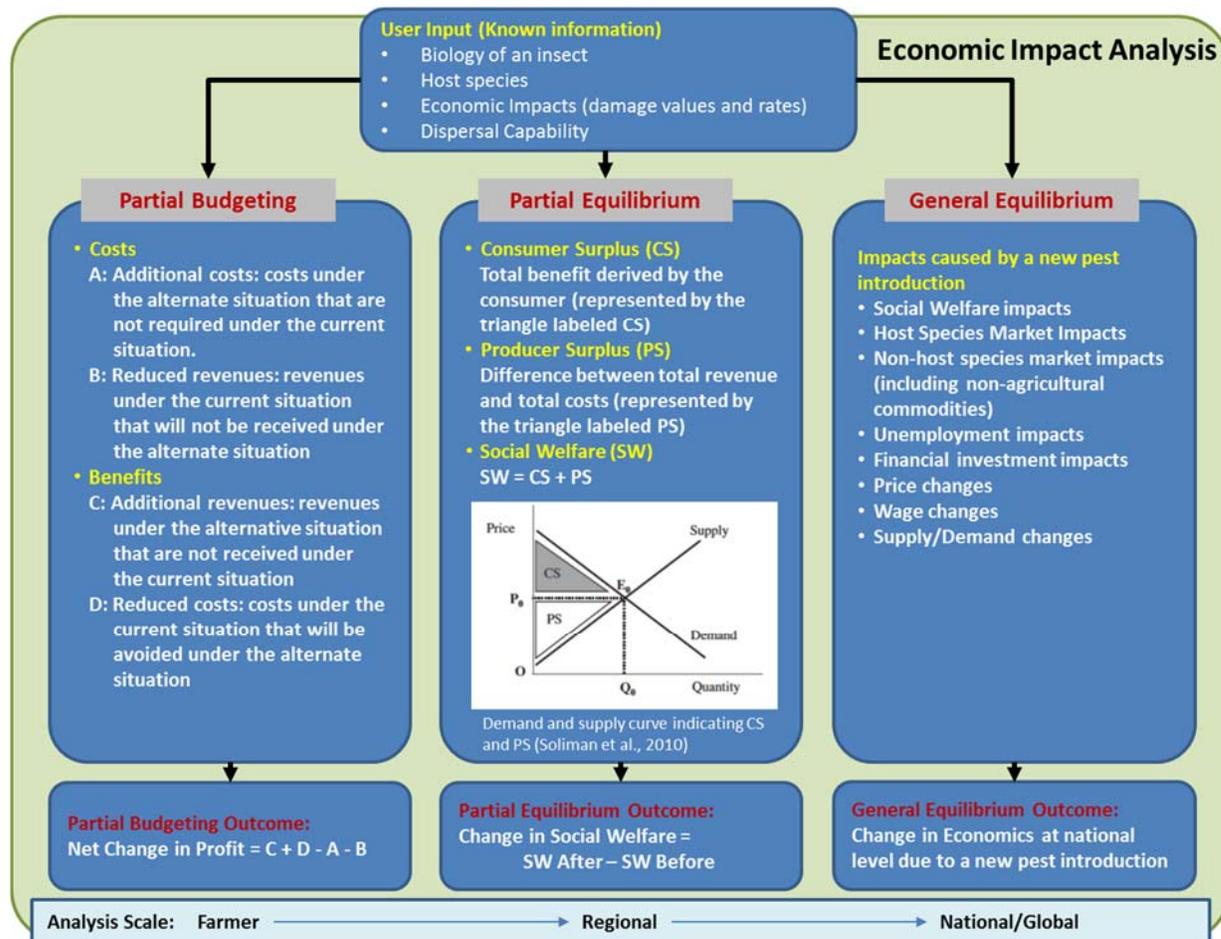


Figure 3-5. Three types of economic impact evaluation analyses (Soliman et al. 2010)

### 3.3.2.5 Uncertainty Analysis Approaches

SAFARIS includes an uncertainty module which is a collection of routines to describe uncertainty. The topic is extremely broad, and the framework does not attempt to be comprehensive. Rather, I focused on approaches that could be consistently used on the data that are known to be available to address the scope of the applications I considered.

The uncertainty module evaluates the uncertainties associated with the pest forecast outputs. Uncertainties within pest forecast outputs are derived from many sources. Elith et al. (2002) and Regan et al. (2002) identified types and sources of uncertainty and ways to deal with the uncertainties and reported that there were 10 types of uncertainties associated with pest forecast systems.

However, there are certain uncertainties that are recognized but cannot be reduced in a pest forecast system. For example, uncertainty may be derived from measurement error. The climate data that are often used in pest forecast systems are developed from climate data collected at weather stations. The machines may have not been calibrated or may have recording errors. Also, linguistic vagueness can contribute to uncertainties in pest forecast outputs.

Those errors are recognized as possible sources of uncertainty associated with pest forecast outputs; however, quantifying and reducing those errors are outside of SAFARIS scope. Here, I focus on three main sources of uncertainties in pest forecast systems: uncertainties in pest biology data, uncertainties in pest models, and uncertainties in climate interpolation methods.

The main goal of estimating uncertainties within the SAFARIS environment is to provide error information to decision makers so that they have a better understanding of pest forecasts and to reduce uncertainties as much as possible by identifying the proportion of errors by each type of uncertainty. The figure below (Figure 3-6) shows the processes of estimating uncertainties within SAFARIS.

In SAFARIS, three main sources of uncertainties (pest biology data, pest forecast models, and climate data) are combined to represent overall uncertainty in pest forecast outputs. The biological data are often derived from field and laboratory experiments. Even when the pests are studied under robust experiments, all biological organisms have variations among individuals. Those errors are often represented as standard errors in the experiments.

Pest forecast models are built with mathematical algorithms based on the relationships between pest biology and input drivers. Those equations also have errors, and those are reported as standard deviations of the model fitness.

The climate data that SAFARIS currently uses are processed (interpolated) products based on information such as weather station data, remote sensing, and digital elevation model results. The interpolation methods are mathematical functions that generate standard deviations. In addition, there are natural variations in weather. It might be hotter one year on a particular day, but it could be cold on the same day next year. These variations also generate uncertainties in pest predictions.

The uncertainty models in SAFARIS identify the sources of uncertainties, combine and integrate those uncertainties, and create an overall uncertainty map associated with the pest prediction.

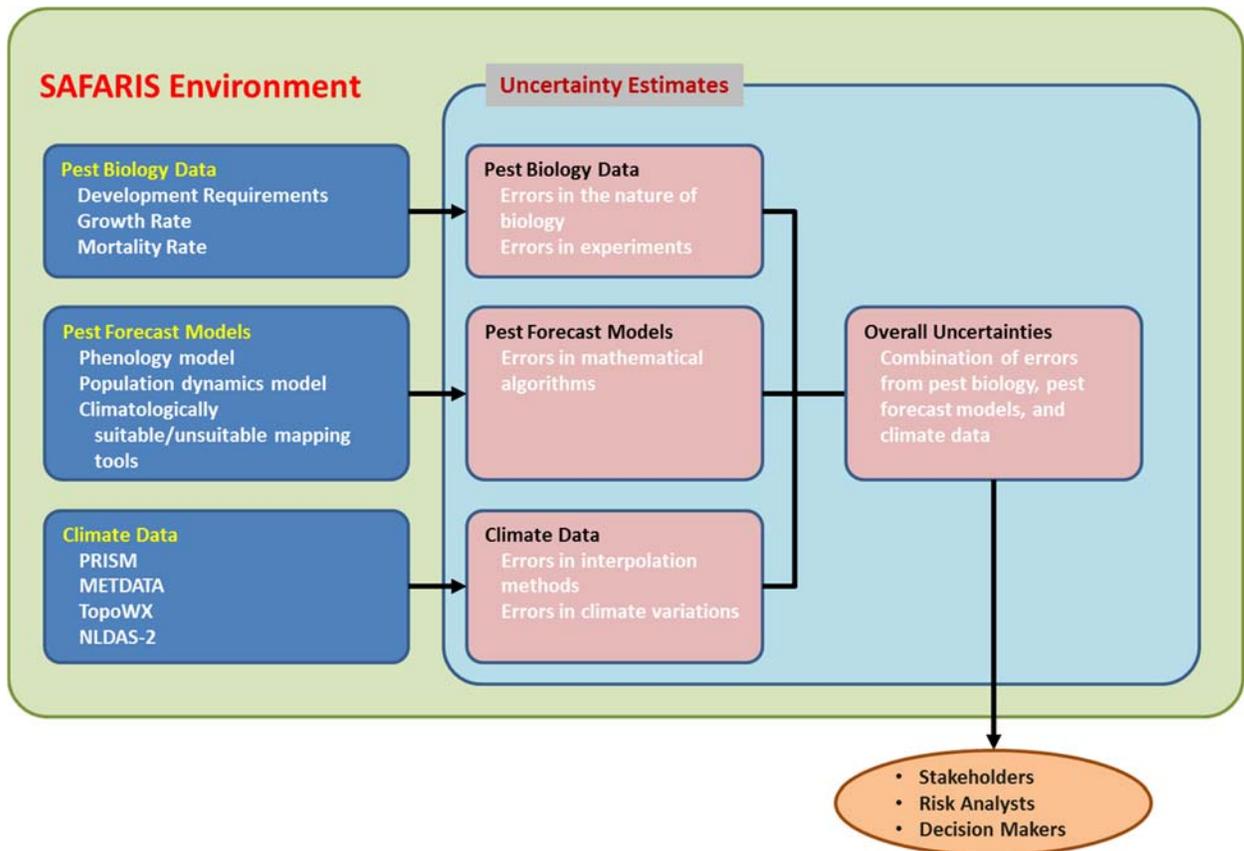


Figure 3-6. Diagram showing the processes of estimating uncertainties associated with pest forecast outputs within SAFARIS

### 3.3.3 Integration within the Spatial Analytic Framework for Advanced Risk Information Systems

The development of the Spatial Analytic Framework for Advanced Risk Information Systems (SAFARIS) began with collecting, processing, and storing commonly used data drivers (key pest predictive modeling data drivers listed in 3.3.1) into databases. The framework was first developed in a desktop computer environment. The desktop version of SAFARIS was programmed with Python® by using ArcGIS® geospatial tools linking to the biological data (e.g., life table information, growth rate, and mortality rate) stored in Excel®.

After the framework in the desktop was evaluated, I developed web-based pest models with Java® scripting and stored web-based pest models and life table biological information for APHIS key arthropod species in a server located at North Carolina State University, Center for Integrated Pest Management (NCSU CIPM). Currently, there are three climate data sources for web-based pest models in SAFARIS: PRISM, MetDATA, and TopoWX. I selected the Java programming language for its practicality and because other USDA APHIS applications used Java in the USDA secure server environment. It is acknowledged that no development language is likely to remain unchanged over time. Thus, abundant source documentation and the publication of this document will help ensure that transitions and migrations to new computer environments and languages is part of an expected and streamlined process. A “wiki” model of cooperation is expected to be one of the first follow up efforts upon the completion of the research described herein.

The pest predictive models chosen were specifically identified as those that could directly and immediately support phytosanitary large-area management programs (including regulatory programs such as those led by APHIS) such as support for first-response to invasive species (New Pest Advisory Group), support for agricultural commodity trade (pest risk analyses), and the Cooperative Agriculture Pest Survey (CAPS) program (environmental suitability analyses).

#### **3.3.3.1 Biology information collection**

The CAPS program currently targets 92 arthropod species (CAPS 2016). The key life table parameters for the CAPS targeted species were collected through literature searches and listed required heat accumulations for each stage and species. I aimed to

simplify parameterization needs by focusing on the minimum set of data that were both commonly available, easy to derive, and yet support informative for useful forecasts. The key parameters obtained from life table studies include minimum threshold temperature for growth, maximum threshold temperature for growth, and required heat accumulations for insect stages. The life table parameters were stored in Excel spreadsheets so that users could select specific species and examine phenology through web-based degree day models. In addition to the life table information stored in Microsoft Excel® files, SAFARIS stores original publications and documents so that users are able to download the source data. Users are able to select a pre-defined life table information for a pest or create a new pest phenology parameters if a particular pest information does not exist within SAFARIS. Users are not able to save the life table information directly into SAFARIS at this time until model parameters are validated in a citable scientific publication.

### **3.3.3.2 Daily weather data collection**

A phenology model requires daily maximum temperatures ( $T_{max}$ ) and daily minimum temperatures ( $T_{min}$ ) in addition to pest biological information (lower threshold temperature) to estimate the heat accumulations required for growth. Three climate data sources for the United States were identified that provide daily temperature data and stored or linked to the web-based SAFARIS. Those data sources are PRISM developed by Oregon State University, METDATA developed by University of Idaho, and TopoWX developed by University of Montana. All three datasets utilize weather station data from NOAA and a digital elevation model (DEM) (Abatzoglou 2013, Oyler et al. 2015, PRISM 2015).

TopoWX also incorporates remotely sensed observations of land temperature (Oyler et al. 2015).

Thematic Real-time Environmental Distributed Data Services (THREDDS) was originally created to find and access scientific data and associated metadata easily and efficiently (Anonymous 2016b). Currently, the focus has shifted to develop a THREDDS Data Server (TDS) that serves contents of the datasets and associated catalogs and metadata (Anonymous 2016b). The datasets stored in a TDS can be accessed through common data access products, such as OPeNDAP, OGC, Web Coverage Service, NetCDF subset, and bulk HTTP file transfer services (Anonymous 2016b). The TDS datasets act as “virtual datasets” because users can access, process, and analyze data on the TDS without downloading to local computers (Anonymous 2016b). The TDS is an open source application that is programmed with Java (Anonymous 2016b).

The State Climate Office of North Carolina in Raleigh, North Carolina maintains, updates, and provides PRISM and METDATA in a TDS. In order to maximize the server space and maintenance efforts at NCSU CIPM, SAFARIS accesses into the State Climate Office for PRISM and METDATA datasets. The TopoWX dataset is currently stored within the NCSU CIPM server. All three daily weather datasets (PRISM, METDATA, and TopoWX) are stored as a NetCDF format. SAFARIS is able to process daily weather data from January 1, 1981 to February 28, 2015 for the PRISM dataset, from January 1, 1981 to December 31, 2014 for METDATA, and from January 1, 1948 to December 31, 2014 for TopoWX.

Accessing data on TDS with SAFARIS has two main advantages. The SAFARIS is not responsible for updating the data on TDS but has access to the most recent, updated data. In addition, analyses can be done without downloading data; therefore, it saves server and local computer resources.

In order to forecast pest patterns and activities under climate change, I downloaded spatially and temporally downscaled General Circulation Models (GCMs) by using the Multivariate Adaptive Constructed Analogs (MACA) method. I selected 5 GCMs for Representative Concentration Pathways (RCP) 4.5 and 8.5 with the MACAv2-LIVNEH downscaling method. Those five GCMs include CanESM2 (Canadian Centre for Climate Modeling and Analysis, Canada), GFDL-ESM2G (NOAA Geophysical Fluid Dynamics Laboratory, USA), HadGEM2-CC (Met Office Hadley Center, UK), IPSL-CM5A-MR (Institut Pierre Simon Laplace, France), and MIROC5 (Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan). The downscaled MACA products were downloaded for January 1, 1950 to December 31, 2099.

The downscaled GCMs were then processed as average daily minimum and maximum temperatures from 1990 to 2005, from 2016 to 2040, and from 2041 to 2065 to represent three time ranges (historical, 25 year-forecast, and 50-year forecast). The downscaled GCMs were averaged for 25 years to represent daily climate data for the next 25 years and 50 years in order to accommodate annual fluctuations and uncertainties associated with the downscaled outputs (Boyles 2015). The downscaled daily GCMs and 25-year averaged GCMs are currently available on the desktop version of SAFARIS.

The original spatial resolutions of those 5 GCMs vary with the products. The original spatial resolution for each GCM is listed in the table below (Table 3-1). All GCMs downscaled by the MACA method have a spatial resolution of 0.0625 degrees (approximately 6.9 kilometers).

**Table 3-1. Five GCMs that were selected for current SAFARIS system (MACA 2016)**

Model Name	Model Country	Model Agency	Original Spatial Resolution (Lon x Lat)
<b>CanESM2</b>	Canada	Canadian Centre for Climate Modeling and Analysis	2.8 deg x 2.8 deg
<b>GFDL-ESM2G</b>	USA	NOAA Geophysical Fluid Dynamics Laboratory	2.5 deg x 2.0 deg
<b>HadGEM2-CC</b>	United Kingdom	Met Office Hadley Center	1.88 deg x 1.25 deg
<b>IPSL-CM5A-MR</b>	France	Institut Pierre Simon Laplace.	2.5 deg x 1.25 deg
<b>MIROC5</b>	Japan	Atmosphere and Ocean Research Institute, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	1.4 deg x 1.4 deg

### 3.3.3.3 Degree-day calculation

Degree-day (heat unit) is a function of daily minimum temperature, maximum temperature and required temperature for pest growth. It is simple yet common in pest management because it is fairly easy to calculate degree days to estimate phenology, i.e., the timing of pest events.

There are several methods to calculate degree-days, and SAFARIS provides the following methods: 1) average method, 2) modified average method, 3) single sine, and 4) single triangulation. Four types of upper cut-off are also available with single sine and single triangulation. Those four upper cut-offs are 1) no cut-off, 2) horizontal cut-off, 3)

vertical cut-off, and 4) intermediate cut-off. Whereas the concept is very simple, the actual application and derivation of methods that cover realistic situations in the field make it such that a final algorithm expression is complex. I developed an integrated approach that captures relevant variations around the topic of estimating degree days. The equations that detail how calculations are made to estimate degree day accumulation are listed in Appendix A.

#### **3.3.3.4 Climatologically Unsuitable/Climatologically Suitable Mapping Tools**

The climatologically unsuitable and climatologically suitable tools identify the areas suitable (or unsuitable) for pest establishment and growth based on the pest biological requirements. The pest biology information is often limited to run a phenology model, spread models or a population dynamic model. Nevertheless, even when the biological information is limited, the potential establishment and spread areas can be estimated with minimum biological information and still produce useful projections.

The climatologically unsuitable and climatologically suitable tools use temperature and precipitation data. Two web-based tools that use temperature data were developed within SAFARIS by linking the PRISM and METDATA datasets at the State Climate Office of NC and the TopoWX dataset at NCSU CIPM. The first tool was designed to find the areas where the specific temperature condition(s) continue for a specific number of consecutive days. Users define those conditions based on pest biology. This tool generates a binary result, meaning that the value of 1 indicates true condition and 0 indicates false condition. The second tool tabulates the number of days that specified temperature condition(s) are true during the particular time period. Users are able to combine daily

minimum temperature, daily minimum temperature, and daily mean temperature conditions for the two web-based tools if using PRISM datasets, but can only combine daily minimum temperature and daily minimum temperature conditions for METDATA and TopoWX because those two datasets do not provide daily mean temperature data.

Two web-based tools that utilized precipitation data from PRISM and METDATA were developed in SAFARIS. One identifies the area based on the precipitation conditions (range of precipitations in millimeters) required for pest growth during particular time period. The results from this tool are displayed as binary values. If the precipitation condition is met, then the result is 1. If not, then the result is 0. Yet another tool provides accumulated precipitation amount in millimeters for a time period specified by the user.

These tools are not simple routines in that they provide categories of useful information that are relevant to phytosanitary applications. The tools are useful to identify suitable or unsuitable areas for pest establishment and growth with basic growth requirement information. The outputs generated from web-based climatologically suitable and climatologically unsuitable models are in NetCDF files and can be easily downloaded from the SAFARIS website for further analyses on desktop computer environments or from other web-based nodes.

#### **3.3.3.5 Web-interface development**

Web-interface and scripts to calculate degree days and to identify areas based on temperature and precipitation suitable/unsuitable areas were developed. The calculations for pest forecasts are managed by scripts written in Java. One script initializes independent variables (drivers) by accessing NetCDF formatted daily weather data for those days

selected by users at the contiguous United States level. I designed and wrote all routines first on a desktop and then mapped the routines to the exact same functions but using web-enabling environments. For the web-based version of the codes and routines, I received programming support from the NCSU Help desk as well as a part time programmer hired as part of the proposals I wrote and for which I obtained grants. I was however directly involved in all code development and I conducted all verification and validation phases for all products I describe.

The web-embedded phenology model was created within SAFARIS. This phenology model was developed for users to easily select climate data source, start date, end date, method of degree-day calculation, and pre-defined pest phenology model. It also allowed users to enter phenology model requirements within SAFARIS. In addition, I created a link to download the original documents and publications if user selected a pre-defined phenology models.

#### **3.3.3.6 Model and data integration**

Connections among databases that contain selected pest biology information, climate data sources, and web-based models were established within the SAFARIS environment (Figure 3-7). Currently, SAFARIS includes three types of pest models, phenology models, climatologically suitable and unsuitable mapping models, and population dynamics models. Each model is connected to required data drivers and accessed through a streamlined, easy-to-use Graphical User Interface (GUI). From the GUI, a user can select pest biology information that is already stored in the SAFARIS pest biology database or the user can supply new pest biology information.

At this time, the web-based models and web-interface for those models reside on the NCSU CIPM server. In addition, the pest biology database and host database also are stored on the NCSU CIPM server.

As for climate data, I tried to link to climate data sources that were stored in a THREDDS Data Server (TDS). SAFARIS accesses PRISM and METDATA that reside in the State Climate Office of North Carolina. A TDS was available for TopoWX. However, I did not connect to the existing TDS, but rather stored TopoWX on a local NCSU CIPM server, since the available TDS service for TopoWX was slow, and it was not reasonable to connect to a remote server. A TDS for NLDAS-2 was not identified, and therefore the data were stored on the local NCSU CIPM server. Downscaled GCMs (Multivariate Adaptive Constructed Analogs) were stored on the NCSU CIPM server since there was no TDS available.

The output generated from any of the web-based models in SAFARIS can be archived and downloaded for further analyses in desktop environments or other web-based nodes. Users are expected to often combine outputs from multiple models (e.g., phenology model and climatologically suitable/unsuitable mapping model). Also, users are able to add supplementary information to the outputs from the SAFARIS models to create a report.

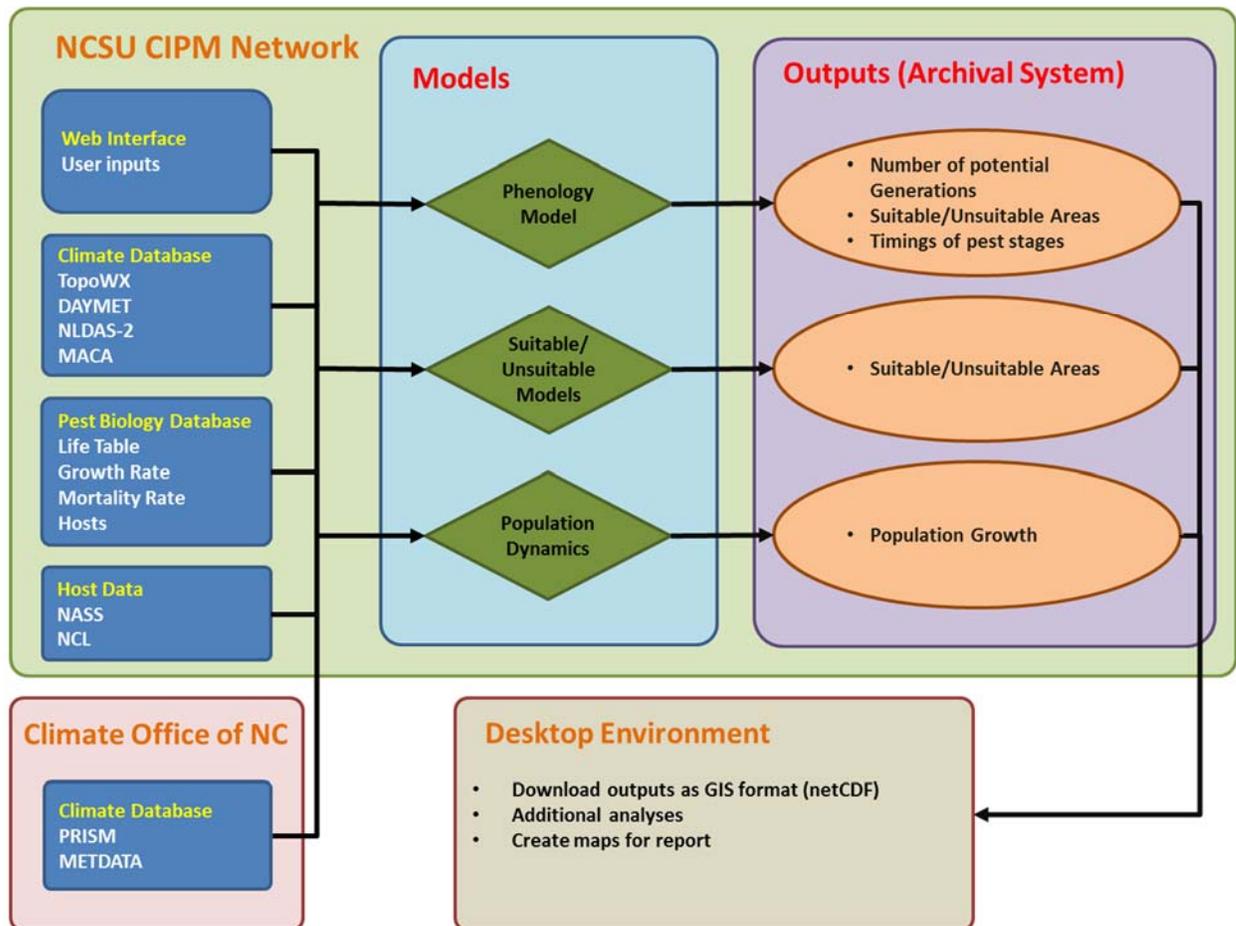


Figure 3-7. Models and associated data that are available at web-based SAFARIS

### 3.3.3.7 Validation of the linkage between data and model within SAFARIS

The links between input data drivers and web based models were tested by running web-based phenology models and climatologically suitable and climatologically unsuitable models within SAFARIS. I tested the web-based phenology modeling component by using life table information for the Oriental fruit fly, *Bactrocera dorsalis* Hendel, to estimate the potential number of Oriental fruit fly generations for one year in the contiguous United States (Rwomushana et al. 2008, Kalia and Yadav 2015). For this validation, I used PRISM, METDATA, and TopoWX datasets to run the phenology model. The sine method

with the horizontal cut-off was used for degree-day accumulations. The parameters used for Oriental fruit fly phenology model are summarized in Appendix C.

In order to test climatologically suitable/unsuitable mapping tools, areas that were suitable for Oriental fruit fly during the winter season were identified using PRISM datasets. The Oriental fruit fly is not known to survive conditions where the temperature falls below 0°C. Therefore, areas where the daily average temperature dropped below 0°C at least for a day were assumed not suitable for Oriental fruit fly survival. Using climatologically suitable/unsuitable mapping tools, the suitable areas were estimated based on the 2010 to 2014 PRISM data.

I also tested the climate change module within SAFARIS by calculating annual degree-day accumulations for Oriental fruit fly using processed daily the downscaled CanESM2 datasets. The same Oriental fruit fly parameters were used for this validation (Appendix C). Again, I used the sine method with the horizontal cut-off to accumulate degree-days to estimate oriental fruit fly growth the under climate change scenario.

### **3.4 Results**

The Spatial Analytic Framework for Advanced Risk Information Systems (SAFARIS) is a framework system that enhances pest forecast abilities by providing a seamless integration of various databases and biological information for a pest. It enables users to provide efficient and timely forecast products and other analyses to phytosanitary decision makers, based on easily available pest information and a series of archived data streams that are integrated with the SAFARIS system.

In the past, APHIS PPQ analysts prepared many pest risk forecasts as ad-hoc projects. There was no mechanism to run forecast models, store associated data drivers, and share models and outputs easily within PPQ. The development of SAFARIS enables APHIS PPQ to manage and store appropriate pest forecast models and associated data drivers for risk analysts in a scientifically defensible form. It also enables risk analysts to run pest forecast models in a timely manner when preparing risk assessments, planning pest surveys, and supporting decision making processes. Importantly, the association with institutions such as NCSU and USDA helps ensure that data streams can be kept up-to-date since the system will be used by large institutional customers that have a stake in the data being maintained for multiple projects and purposes.

The advantages of developing SAFARIS include:

- 1) Commonly used spatial data (e.g., elevation, climate data, and vegetation data) are processed and ready to be plugged into forecast models. The spatial data are maintained up to date.
- 2) Pest forecast models developed by PPQ or cooperators reside within the SAFARIS framework.
- 3) Module design allows users not only to add new information and models but also to select essential models and data based on the objectives of the study.
- 4) Uncertainty characterization becomes possible within SAFARIS because multiple statistical models can be run with exactly the same input data.
- 5) The climate change module is available to add into any of the pest forecast models to incorporate potential climate change effects.

6) It is an archival system for regulatory forecast products.

### **3.4.1 Development of the web-based phenology model**

A web-based phenology model was developed to predict pest events to test the framework and to support several PPQ activities such as New Pest Advisory Group (NPAG) reports, agricultural commodity trade supports, and Cooperative Agriculture Pest Survey (CAPS) activities (Figure 3-8). The web-based phenology models were linked to PRISM and METDATA on TDS at the State Climate Office of NC and TopoWX stored at NCSU CIPM.

Once the users selected a phenology model for a pest of concern, the phenology model parameters were displayed on the web page (Figure 3-9). The users tested and were able to modify the model parameters before submission. SAFARIS includes the scientific reference for the phenology model parameters. All pre-defined phenology models were parameterized based on valid scientific experiments, research or expert estimates (heuristics). In order to make the system transparent, SAFARIS provides original source information for the phenology model that the user selects. In addition, users tested and were able to create new parameter settings by using phenology model templates (Figure 3-10).



## Spatial Analytic Framework for Advanced Risk Information Systems

HOME DEGREE DAY MODEL TEMPERATURE MODEL INSECT MODEL PRECIPITATION MODEL

Climate Data Source:

Begin Date:

End Date:

Upper Cut-Off:

DD Calculation Type/Analysis Model:

Pest Model:

Submit

**Figure 3-8. Web interface for degree-day model within SAFARIS**



## Spatial Analytic Framework for Advanced Risk Information Systems

HOME DEGREE DAY MODEL TEMPERATURE MODEL INSECT MODEL PRECIPITATION MODEL

Climate Data Source:

Begin Date:

End Date:

Upper Cut-Off:

DD Calculation Type/Analysis Model:

Pest Model:

Curious where we got the below data : [Download Reference](#)

Lymantria dispar Female Sheehan Lower Threshold (°C) Upper Threshold (°C) Degree Days Stages

Stages	Lower Threshold (°C)	Upper Threshold (°C)	Degree Days
Egg	3	38	282
Immature Stage 1	8.1	41	376
Immature Stage 2	8.1	41	441
Immature Stage 3	8.1	41	513
Immature Stage 4	8.1	41	607
Immature Stage 5	8.1	41	707
Immature Stage 6	8.1	41	933
Pupae	5.1	41	1167

Submit

Figure 3-9. Web interface for degree-day model parameter settings in SAFARIS

Enter pest name:

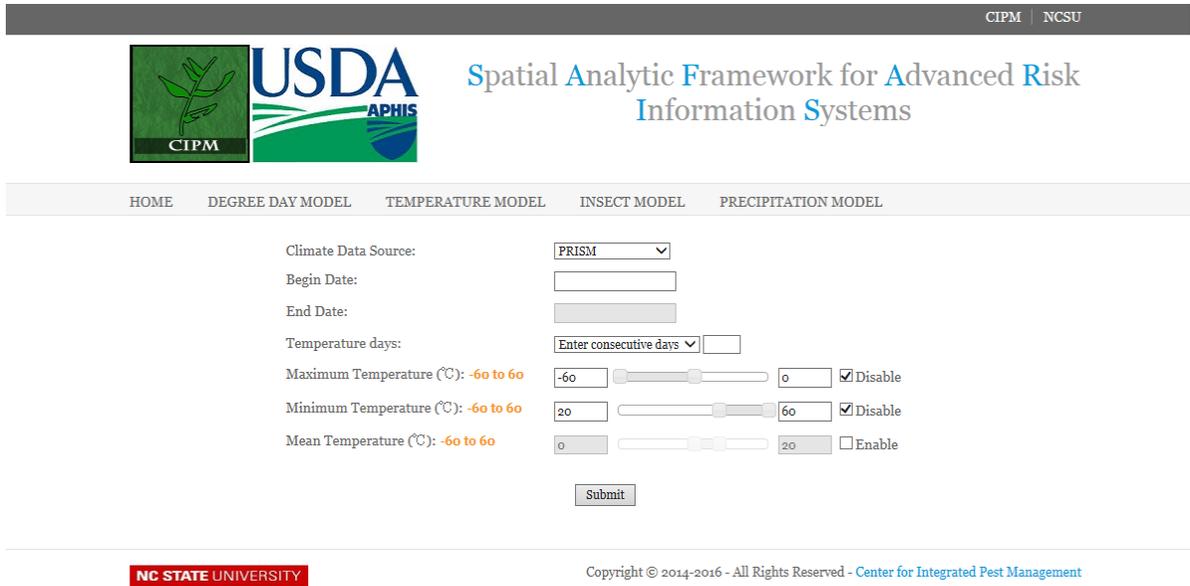
Overwintering Egg                       Overwintering Larvae/Nymph  
 Overwintering Pupae                       Overwintering Adult  
 Egg  
 First Instar/Immature Stage 1                       Second Instar/Immature Stage 2  
 Third Instar/Immature Stage 3                       Fourth Instar/Immature Stage 4  
 Fifth Instar/Immature Stage 5                       Sixth Instar/Immature Stage 6  
 Seventh Instar/Immature Stage 7                       Eighth Instar/Immature Stage 8  
 Ninth Instar/Immature Stage 9                       Tenth Instar/Immature Stage 10  
 All Immature Stages                       Pupae  
 Pre-oviposition Adult                       Adult

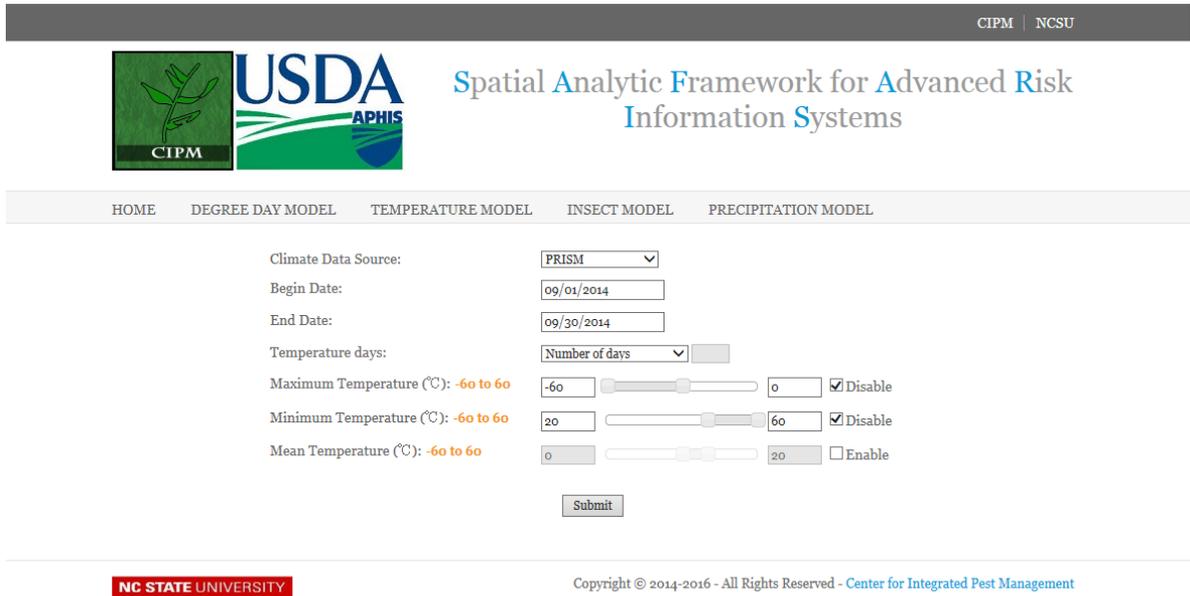
**Figure 3-10. Template for setting phenology model parameters**

### **3.4.2 Development of the web-based climatologically suitable and unsuitable mapping models**

The web-based climatologically suitable/unsuitable mapping models were created within SAFARIS to test the connectivity between climate data and models and to support PPQ activities. I developed four tools as climatologically suitable/unsuitable mapping models; 1) a tool that indicates true or false based on the pest temperature requirement(s) for specific consecutive days (Figure 3-11), 2) a tool that indicates number of days that temperature condition(s) were ideal (or not ideal) based on the pest temperature requirements for growth (Figure 3-12), 3) a tool that indicates true or false based on the precipitation requirement for a pest growth (Figure 3-13), and 4) precipitation tool that indicates accumulated precipitation in millimeters (Figure 3-14). Users tested and were able to select a daily climate data source, define the time period, and set temperature or precipitation criteria conditions in the web interface.



**Figure 3-11. Temperature condition tool that indicates true or false based on the temperature criteria(s) for user defined number of consecutive days**



**Figure 3-12. Temperature tool that indicates number of days that user defined criteria met**

CIPM | NCSU



## Spatial Analytic Framework for Advanced Risk Information Systems

HOME   DEGREE DAY MODEL   TEMPERATURE MODEL   INSECT MODEL   PRECIPITATION MODEL

Climate Data Source:

Begin Date:

End Date:

Select an option:

Total Precipitation (mm): 0 to 1000

---

NC STATE UNIVERSITYCopyright © 2014-2016 - All Rights Reserved - Center for Integrated Pest Management

**Figure 3-13. Precipitation condition tool that indicates true or false based on the criteria users specify**

CIPM | NCSU



## Spatial Analytic Framework for Advanced Risk Information Systems

HOME   DEGREE DAY MODEL   TEMPERATURE MODEL   INSECT MODEL   PRECIPITATION MODEL

Climate Data Source:

Begin Date:

End Date:

Select an option:

---

NC STATE UNIVERSITYCopyright © 2014-2016 - All Rights Reserved - Center for Integrated Pest Management

**Figure 3-14. Precipitation tool that indicates accumulated precipitations in millimeters**

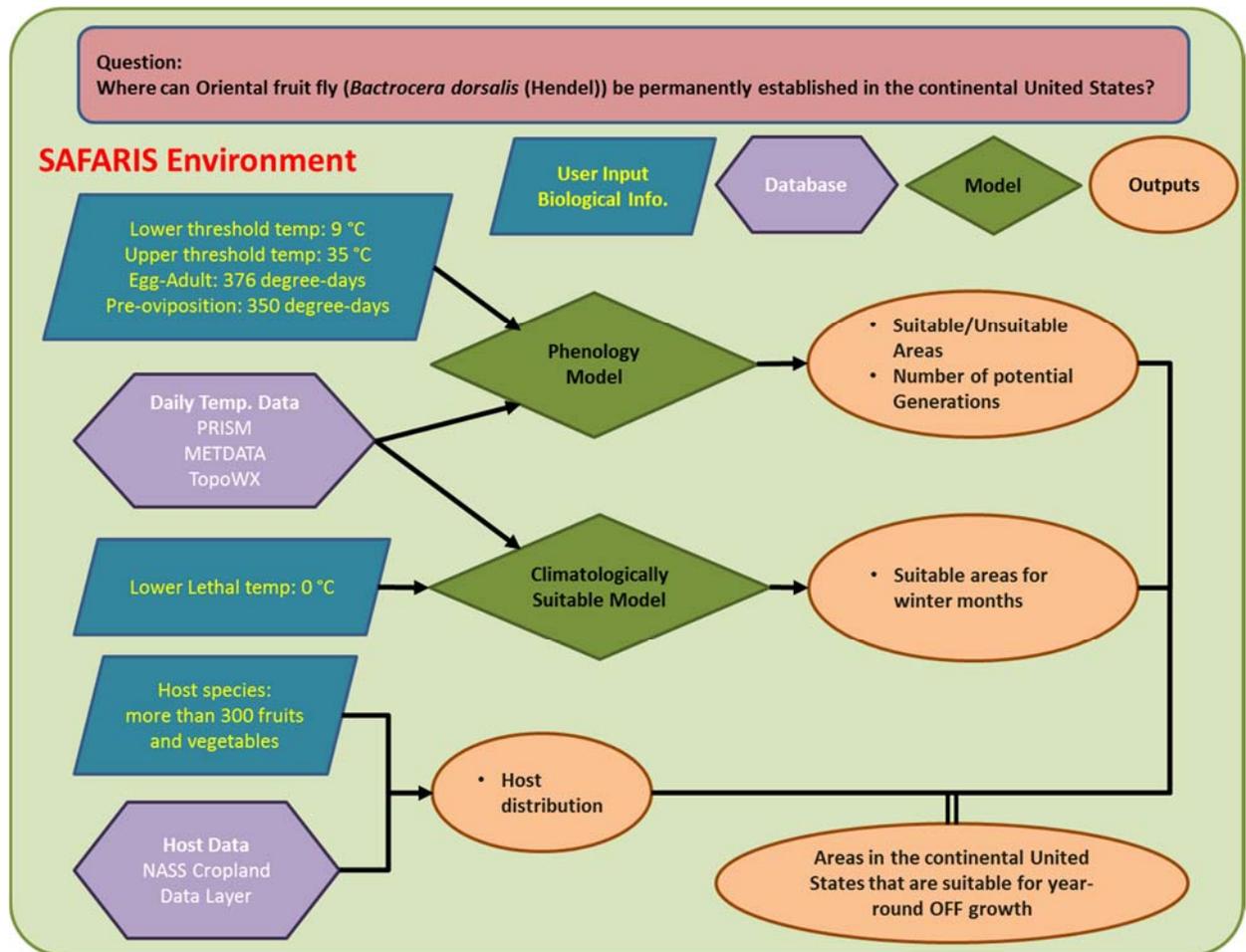
### 3.4.3 Testing of SAFARIS

Since its completion, the SAFARIS phenology model has already supported a number of real-world applications related to PPQ phytosanitary and regulatory activities. In 2015, there were outbreaks of Oriental fruit fly (*Bactrocera dorsalis* Hendel) in Florida (El-Lissy, 2015). Subsequently, USDA APHIS established a new Oriental fruit fly regulated area in the Redland area of Miami-Dade County, Florida.

USDA APHIS uses phenology models for variety of reasons. For instance, the agency places survey traps and monitors according to forecasted pest stages and events. USDA APHIS declares successful eradications if APHIS confirms the area pest free for the forecasted next two generation period estimated by phenology models. USDA APHIS also uses phenology model and climate data (historical and forecasted climatology) to estimate when the pest may experience the completion of two generations. USDA APHIS continues to survey and monitor during those periods and “ground-truths” whether the eradication is successful or not.

In order to test the framework and the connectivity among user input data, databases, and spatial models, areas in the continental United States that are suitable for permanent establishment of Oriental fruit fly (*Bactrocera dorsalis* Hendel)) were estimated with the SAFARIS (Figure 3-15). The flow chart below (Figure 3-15) shows the processes, data, and models used in the SAFARIS to estimate the suitable areas.

First, biological information for Oriental fruit fly was gathered through a literature search. The biology and life table information for *Bactrocera dorsalis* were obtained from experiments by Rwomushana (2008) and Kalia & Yadav (2015) (Appendix C).



**Figure 3-15. Flow chart diagram explaining analysis processes within SAFARIS**

Second, the phenology model was tested within SAFARIS. I used three climate data sources (PRISM, METDATA and TopoWX) to run a phenology model to accumulate degree-days for Oriental fruit fly from January 1 to December 31, 2014. Based on the accumulated degree-days for 2014, the number of potential generations for one year in the continental United States were estimated. The maps below indicating potential numbers of Oriental fruit flies in the contiguous United States with daily climate data from 2014 were generated within SAFARIS (Figure 3-16; Figure 3-17; Figure 3-18). I used the horizontal cut-off single sine method to calculate accumulated degree days for 2014 and estimated potential number of oriental fruit fly generations. The three results from three climate data

sources were shown in Figure 3-16 (PRISM), Figure 3-17 (METDATA), and Figure 3-18 (TopoWX).

There were slight differences among the three phenology outputs derived from three climate data sources, but overall, the results were very similar. The majority of the areas within the contiguous United States was able to at least complete one generation of Oriental fruit fly in a year. The Redland Florida area could have as many as 7 generations a year based on the climatology from 2010 to 2014.

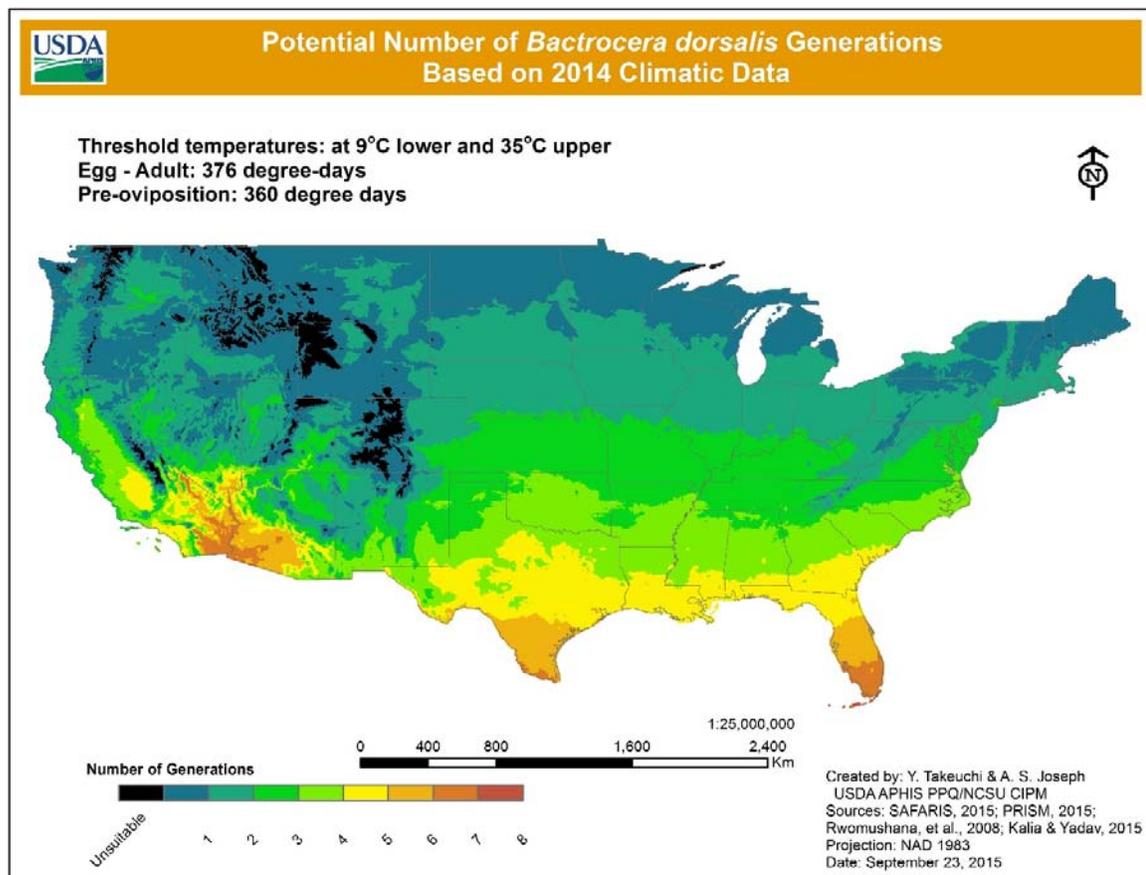
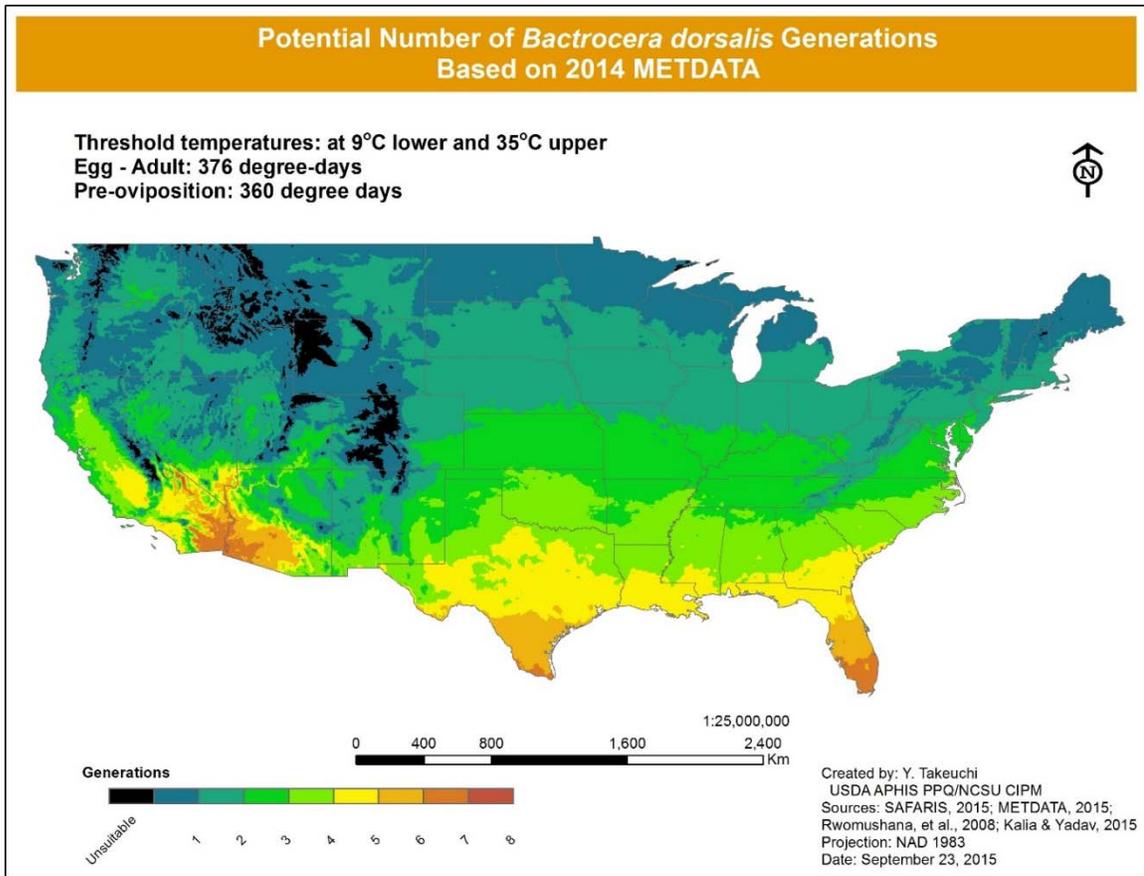
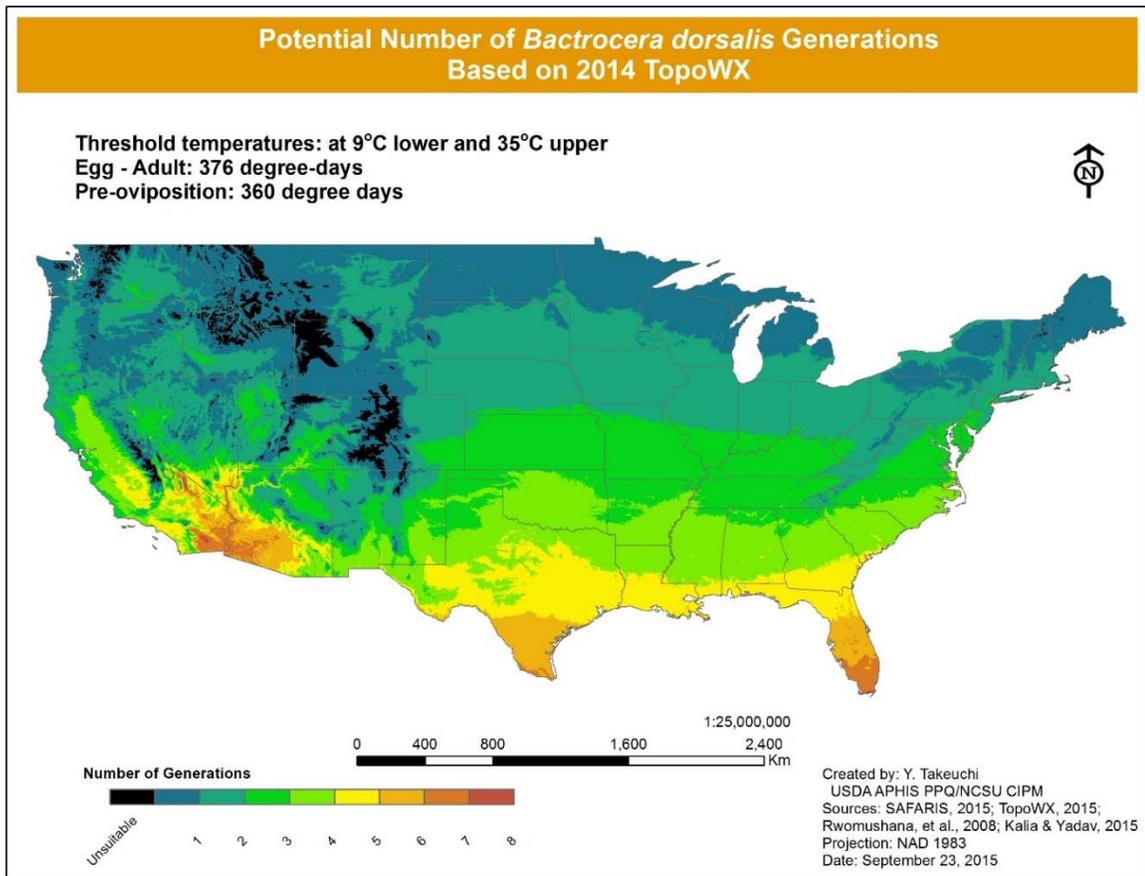


Figure 3-16. Potential number of *Bactrocera dorsalis* generations based on 2014 PRISM climatology



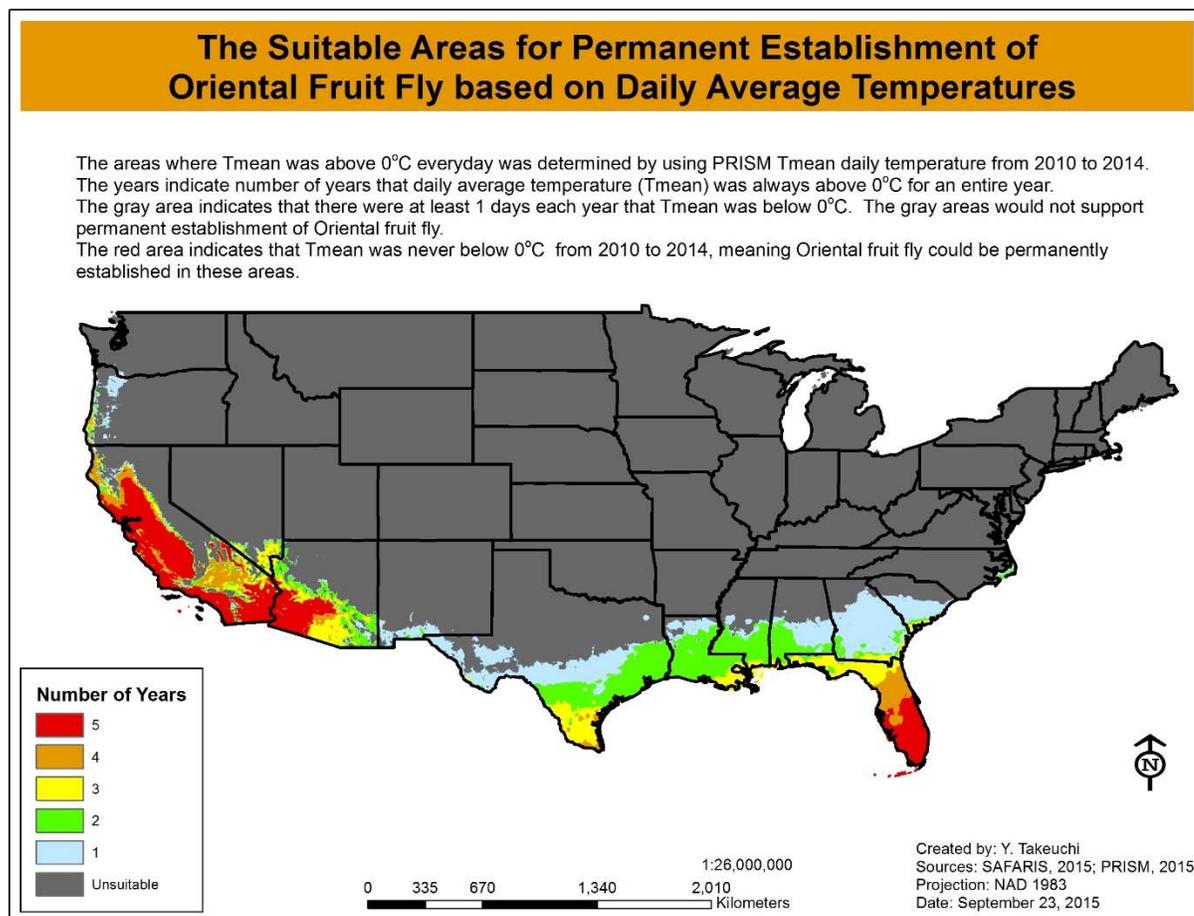
**Figure 3-17. Potential number of *Bactrocera dorsalis* generations per year based on 2014 METDATA climate data**



**Figure 3-18. Potential number of *Bactrocera dorsalis* generations per year based on 2014 TopoWX climate data**

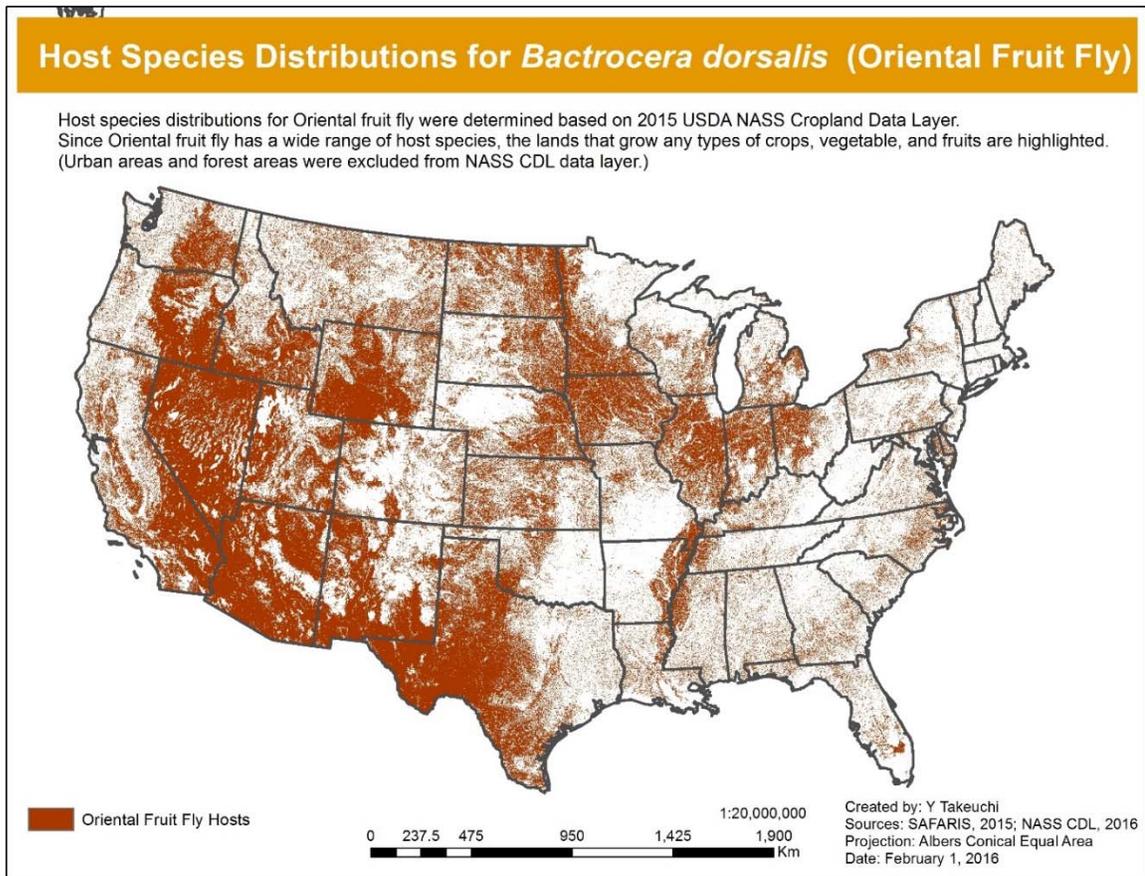
Third, I examined the climatological suitable/unsuitable mapping tools developed in SAFARIS. The Oriental fruit fly could not survive at temperature below 0°C; therefore, the areas that always had daily average temperatures (Tmean) above 0°C were identified with the climatologically suitable/unsuitable mapping tools by using PRISM climate data from 2009 to 2014. The map below (Figure 3-19) indicates unsuitable areas (areas in gray) and potentially suitable areas (colored areas) for Oriental fruit fly survival. The number of years on the map indicates the number of years having a Tmean above 0°C during the five years (2009-2014). The red areas indicate 5 years, meaning Tmean being always above 0°C from

2009 to 2014. The light blue areas indicate 1 year; therefore, the light blue areas had only one year that had a Tmean above 0°C every day from 2009 to 2014.



**Figure 3-19. The suitable areas for permanent establishment of Oriental fruit fly (*Bactrocera dorsalis*) based on the PRISM daily average temperatures**

Oriental fruit fly, *Bactrocera dorsalis* (Hendel), has a wide range of host species. Since *B. dorsalis* is known to attack more than 300 species and most fruits and vegetables, only the locations indicated as urban and forest areas were excluded from host species distribution. The Figure 3-20 displays locations that host species of *B. dorsalis* are grown. The host species are distributed throughout the contiguous United States.



**Figure 3-20. Host species distributions for *Bactrocera dorsalis* generated from 2015 NASS Cropland Data Layer**

The outputs from a phenology model and a climatologically suitable mapping tool and distributions of host species were combined to estimate the areas within the contiguous United States that support *Bactrocera dorsalis* growth for an entire year. Results show that Florida, southern Arizona, and southern California had suitable conditions for an entire year for *B. dorsalis* growth.

In order to test climate change module within SAFARIS, I calculated annual degree-day accumulations for *B. dorsalis* by using downscaled MACA CanESM2 datasets. The single sine with horizontal cut-off method was used to calculate degree days.

The Figure 3-21 and Figure 3-22 display estimates on annual degree-day accumulations for *B. dorsalis* based on downscaled CanESM2 datasets from 2016 to 2040 and from 2041 to 2065, respectively. The calculations were done by using averaged downscaled daily climate data from 2016 to 2014 and from 2041 to 2065.

The forecasted annual degree-day accumulations for 2016-2040 (Figure 3-21) indicated that southern Florida and southern Texas would accumulate 5,800 to 6,000 degree days per year. It is enough physiological heat for *B. dorsalis* to complete 8 generations.

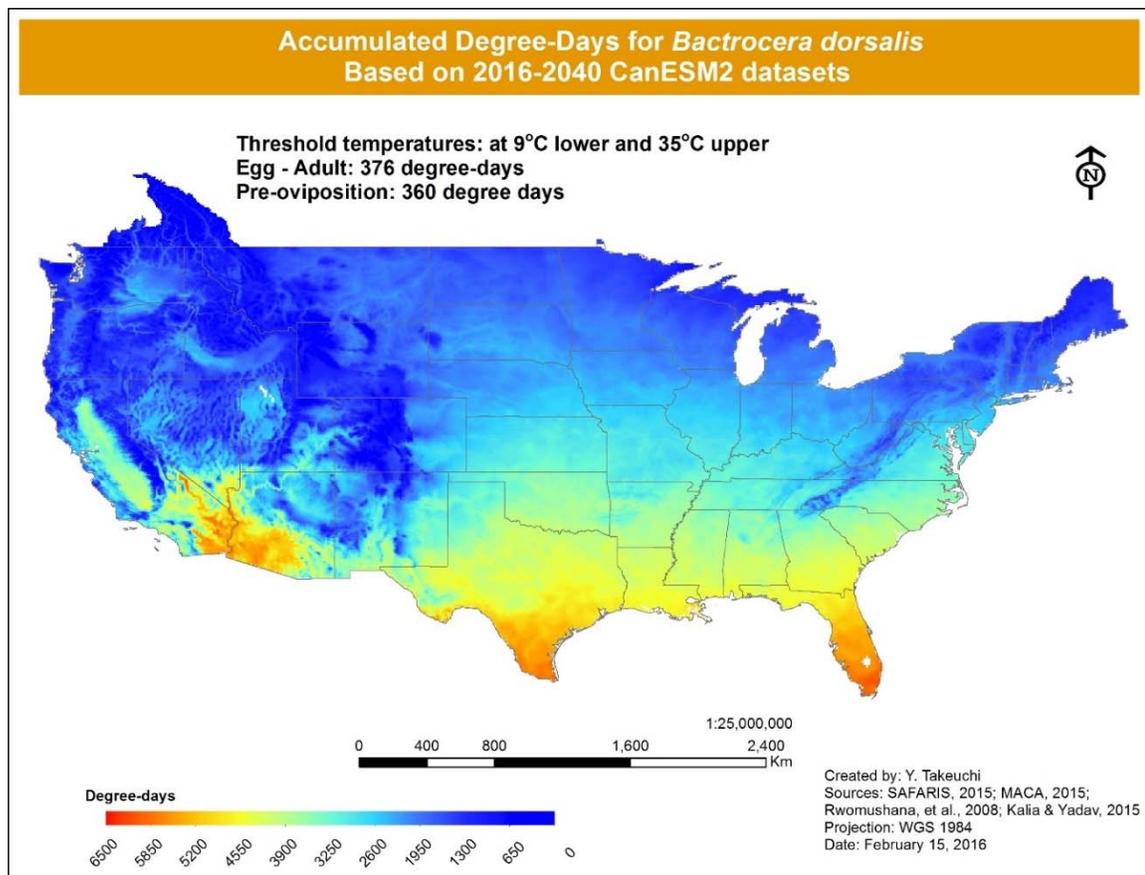
Overall, the forecasted annual degree-day accumulations for 2041-2065 (Figure 3-22) were slightly higher than the ones for 2016 and 2040. In Miami area, the forecasted annual degree-day accumulations based on 2016-2040 data were approximately 5,850 degree-days, while the forecasted accumulation based on 2041-2065 data were 6,025 degree-days.

The differences in degree-days accumulation derived from the historical data and forecasted data were visualized in Figure 3-23. The differences in annual degree-day accumulations between historical climate data (1990-2005) and forecasted 2016-2040 climate data were shown on an upper map, and the differences between historical climate data and forecasted 2041-2065 climate data were shown on a lower map (Figure 3-23).

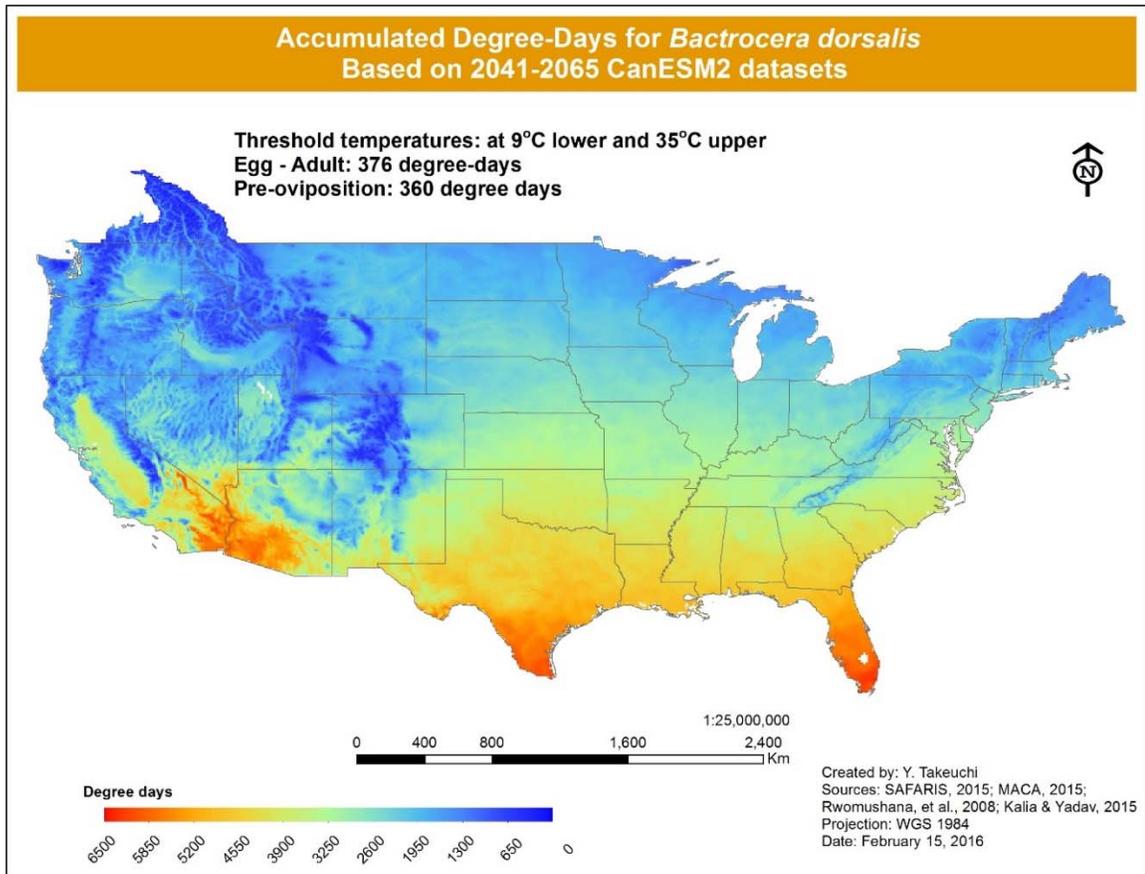
The largest difference estimated was western Texas, southern New Mexico, and southern Arizona. Those areas were forecasted to increase as much as 900 degree-days in a year. Since *B. dorsalis* requires approximately 736 degree-days to complete a generation, those areas could have one additional generation per year.

In the next 50 years, western Texas, southern New Mexico, and southern Arizona were forecasted to be more suitable for *B. dorsalis* growth. Those areas were expected to increase more than 800 degree-days per year based on the phenology model run with forecasted 2041-2065 climate data.

None of the areas indicated negative value when comparing with historical data. This means that the contiguous United States is expected to have warmer climate in general and more suitable conditions for *B. dorsalis*.



**Figure 3-21. Forecast on accumulated degree-days for *B. dorsalis* based on downscaled 2016-2040 CanESM2 data**



**Figure 3-22. Forecast on accumulated degree-days for *B. dorsalis* based on downscaled 2041-2065 CanESM2 data**

## Differences in Accumulated Degree-Days for *Bactrocera dorsalis* Estimated with Historical and Forecasted Climate Data

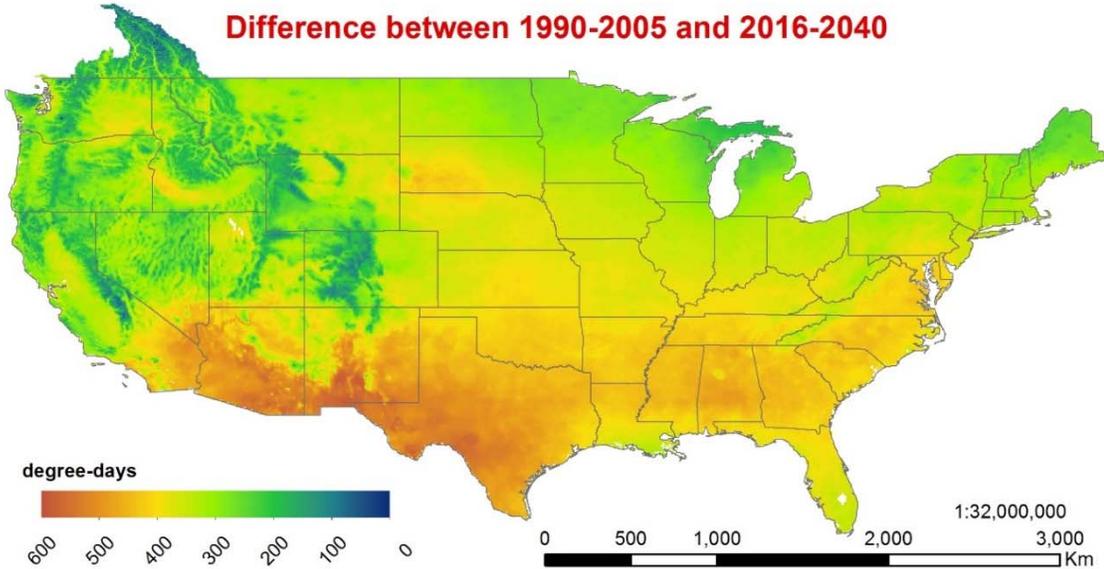
Threshold temperatures: at 9°C lower and 35°C upper

Egg - Adult: 376 degree-days

Pre-oviposition: 360 degree days



### Difference between 1990-2005 and 2016-2040



### Difference between 1990-2005 and 2041-2065

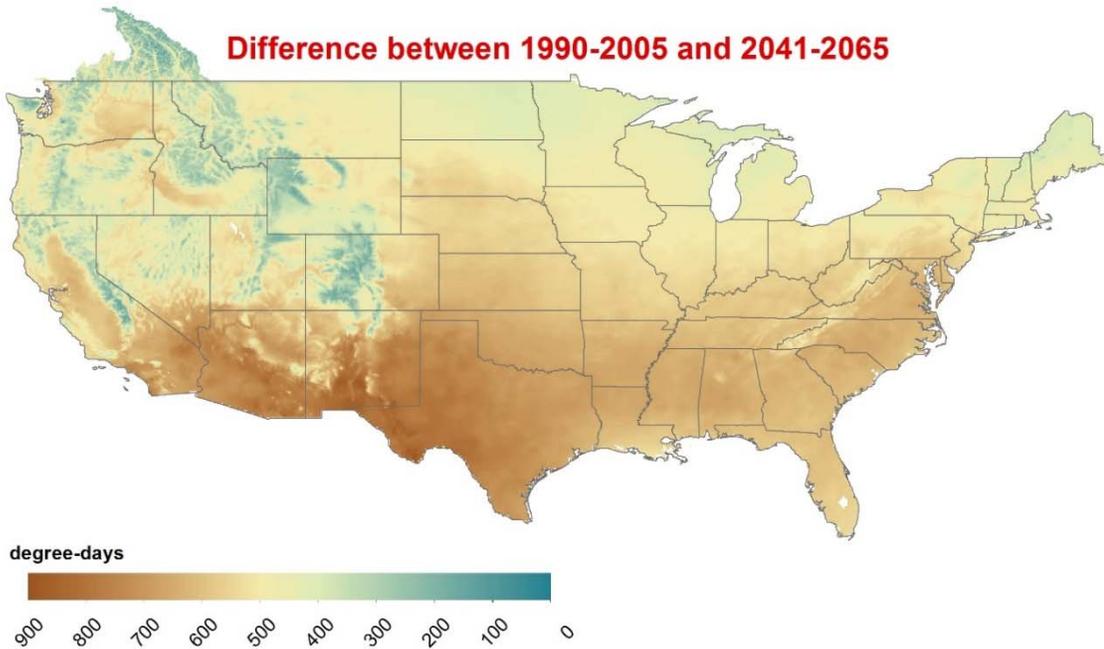


Figure 3-23Figure 3-24. Differences in annual accumulated degree-days for *B. dorsalis* estimated with historical climate data and forecasted climate data

### 3.5 Discussion

Agencies that safeguard natural systems and agriculture like USDA APHIS and international national plant protection organizations like the Administration of Quality Supervision, Inspection and Quarantine (AQSIQ) in China, the Kenya Plant Health Inspectorate Service (KEPHIS) in Kenya, the Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria (SENASICA) in Mexico use climate based forecast models every day to meet their mission to protect agriculture and natural resources. These organizations often conduct spatial analysis at ad-hoc bases with methods that are unique and results that are not easy to compare. One highly inefficient characteristic is that when these modeling projects end, the models proper, but also the data collected during the project, are almost never shared because there is no repository or framework to enable this sharing.

This requires phytosanitary safeguarding organizations like APHIS to re-process climate and other data and re-create a system repeatedly. Now, with the SAFARIS framework, a consistent, transparent, and ready-to-use climate approach is always ready and accessible to the phytosanitary community.

There are already existing frameworks and tools for pest forecast systems (e.g., CLIMEX, PRATIQUE, and USGS SAHM). Although some systems are open source models and tools, almost all tools are mainly developed to run on a desktop environment. Although I envision SAFARIS to be an open source system where everyone has access to the models and associated (non-sensitive data), SAFARIS is currently developed mainly to

support phytosanitary risk analysts. An open-source approach will foster continued cooperation, innovation, and creativity.

In addition, prior frameworks and systems do not have a straightforward ability to deal with historical climate data, near-real time weather data, short-term forecasted climate data, and climate data derived from GCMs. Climate plays an important role in pest developments as well as pest forecasts; therefore, the system should be able to quickly update or connect to the different sources of climate data easily.

Another shortcoming of previously existing framework was not being able to manage uncertainties. For use in a regulatory arena, any system should be able to separate uncertainty associated with input data from the inherent in model proper.

Initially, SAFARIS was built in a desktop computer environment. The desktop computer environment would not fully work for SAFARIS because it has sharing and accessing capabilities to a larger community than an individual or few people accessing a single desktop computer. However, the concept was tested and fully verified and validated.

SAFARIS was tested in several ways. First, I interviewed and established an extended dialogue with phytosanitary managers and researchers from not only PPQ but also several regulatory organizations in other countries. The intent was to ensure that the components of the framework would address PPQ needs as well as their needs.

Second, once the framework was established, communication between components and physical environments needed to be tested. This was done by supplying biological data into pest forecast models and running those models within the SAFARIS environment.

Then, the pest forecast models were verified to see if the calculations were logical in SAFARIS by using simple test data with known outcomes.

The most important test was the third test. I used SAFARIS to validate my hypotheses that the models and tools within SAFARIS could support pest risk assessments. I worked with risk analysts to establish that basic risk analysis work could be supported by providing the likelihood of establishment information. For example, I worked with policy makers to ensure that long term forecasts were useful to inform policy, worked with land managers to confirm that the use of climate change effect was significant to long term resilience of managed system, and worked with economists to test the link between forecast models and economic partial budgeting analysis. These tests confirmed the fundamental value of the framework.

A final test, which was very important for me, was whether I could take the concept of the framework and move to a completely new environment without code or models. I wanted to test this because developing countries often did not have the same resources we had in the United States but would be interested in assisting and supporting pest control and prevention at a global level.

Thus, the concept of SAFARIS was tested with a project to develop a forecast system to manage beet armyworm (*Spodoptera exigua* (Hübner)) in Jamaica (Takeuchi 2015). This project was a part of “Strengthening national Beet Armyworm (*Spodoptera exigua*) Programme” funded by the UN Food and Agricultural Organization to aid the Ministry of Agriculture and Fisheries. Although biological data as well as climatological

data are limited in Jamaica, the system has been developed to forecast beet armyworm behavior and to guide appropriate control measures to farmers by Ministry extension agents.

In order to establish an expanded, even more functional, I am currently extending the basic functionality (as described here) of SAFARIS through a cooperative agreement which I proposed and competitively awarded late in FY2014 with supplemental grants proposed and approved March 2016. At this time, the funds obtained by USDA are being managed through a cooperative agreement between the National Science Foundation, Center for Integrated Pest Management at North Carolina State University (NCSU CIPM) and USDA APHIS PPQ. SAFARIS contains models developed as part of this research but is planned to expand and incorporate models developed by scientists from NCSU, APHIS, and other research counterparts (e.g., USDA Agricultural Research Service, University of Cambridge, Lincoln University). That is, specific, discrete model products (simulation systems) are expected to be a dynamic feature of the framework.

The development of SAFARIS is expected to provide essential functions for phytosanitary management as part of a continuously available system to provide pest and disease forecasts. Establishing all the desired features for the framework and populating the framework with key simulation systems will be an ongoing project. Beyond establishing fundamental components and key case studies as part of my dissertation, the project is expected to continue to expand as part of an existing cooperative agreement for at least 3 to 5-years beyond the terminus of my dissertation research. The additional development will focus mainly on the incorporation and testing of new models and expanding functionality

beyond the scope of regulatory operations to include support of strategic planning by farmers and ranchers as part of USDA climate change sustainability challenges.

### 3.6 References

- Abatzoglou, J. T. 2013. Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology* **33**:121-131.
- Abatzoglou, J. T., and T. J. Brown. 2012. A comparison of statistical downscaling methods suited for wildfire applications. *International Journal of Climatology* **32**:772-780.
- Anonymous. 1994. Migratory grasshopper: *Melanoplus sanguinipes* (Fabricius). Wyoming Agricultural Experiment Station. Wyoming Agricultural Experiment Station Bulletin 912,
- Anonymous. 2016a. Measuring vegetation (NDVI & EVI): Normalized Difference Vegetation Index (NDVI). Earth Observatory, Earth Observing System Project Science Office, NASA Goddard Space Flight Center, Greenbelt, MD. Last accessed January, 2016, [http://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring\\_vegetation\\_2.php](http://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring_vegetation_2.php).
- Anonymous. 2016b. THREDDS fact sheet. University Corporation for Atmospheric Research, Boulder, CO. <https://www.unidata.ucar.edu/publications/factsheets/2007sheets/threddsFactSheet-1.pdf>.
- Auclair, A. (2015). Prediction and forecasting of approach rates of US quarantine pests on commodity imports and in passenger baggage. Personal communication to on September, 2015, from meeting to discuss a development of statistical port targeting tools.
- Baker, R. H. A., A. Battisti, J. Bremmer, M. Kenis, J. Mumford, F. Petter, G. Schrader, S. Bacher, P. De Barro, P. E. Hulme, O. Karadjova, A. O. Lansink, O. Pruvost, P. Pysek, A. Roques, Y. Baranchikov, and J. H. Sun. 2009. PRATIQUE: a research project to enhance pest risk analysis techniques in the European Union. *EPPO Bulletin* **39**:87-93.
- Baranchikov, Y. N. 1988. Ecological basis of the evolution of host relationships in Euroasian gypsy moth populations. Pages 319-338 *in* W. E. Wallner and K. C.

- McManus, editors. *Lymantriidae: a comparison of features of new and old world tussock moths*.
- Boyles, R. (2015). Downscaled GCMs. Personal communication to on from meeting to discuss available downscaled GCMs for the contiguous United States.
- Breukers, A., M. Mourits, W. van der Werf, and A. O. Lansink. 2008. Costs and benefits of controlling quarantine diseases: a bio-economic modeling approach. *Agricultural Economics* **38**:137-149.
- Burkman, B. 2005. *Forest Inventory and Analysis: sampling and plot design*. United States Department of Agriculture, Forest Service, Washington, D.C.
- CABI. 2016. *Crop Protection Compendium*. CAB International. [www.cabi.org/cpc](http://www.cabi.org/cpc).
- CAPS. 2016. The CAPS program. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD. Last accessed February, 2016, <https://caps.ceris.purdue.edu/home>.
- Christensen, J. H., and B. Hewitson. 2007. *Regional climate projections*. Cambridge Univ Press, New York.
- Colunga-Garcia, M., R. A. Haack, and A. O. Adelaja. 2009. Freight transportation and the potential for invasions of exotic insects in urban and periurban forests of the United States. *Journal of Economic Entomology* **102**:237-246.
- Cook, D. C., L. R. Carrasco, D. R. Paini, and R. W. Fraser. 2011. Estimating the social welfare effects of New Zealand apple imports. *Australian Journal of Agricultural and Resource Economics* **55**:599-620.
- Coop, L. 2015. *USPest*. Oregon State University, Integrated Plant Protection Center, Corvallis, Oregon. Last accessed December, 2015, <http://uspest.org/>.
- Crall, A. W., C. S. Jarnevich, B. Panke, N. Young, M. Renz, and J. Morissette. 2013. Using habitat suitability models to target invasive plant species surveys. *Ecological Applications* **23**:60-72.
- CRU. 2015. *Climate Research Unit*. University of East Anglia, School of Environmental Sciences, Climate Research Unit, Norwich, UK. Last accessed <http://www.cru.uea.ac.uk/>.
- Dennis, B., W. P. Kemp, and R. C. Beckwith. 1986. Stochastic-model of insect phenology - estimation and testing. *Environmental Entomology* **15**:540-546.

- Didan, K. 2016. MODIS Land: vegetation index. National Aeronautics and Space Administration (NASA), Washington, DC. Last accessed May, 2016, <http://modis-land.gsfc.nasa.gov/vi.html>.
- ECMWF. 2016. European Centre for Medium-Range Weather Forecasts. European Centre for Medium-Range Weather Forecasts, Reading, UK. Last accessed <http://www.ecmwf.int/en/about>.
- Elith, J., M. A. Burgman, and H. M. Regan. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling* **157**:313-329.
- Elkinton, J. S., and A. M. Liebhold. 1990. Population-dynamics of gypsy moth in North America. *Annual Review of Entomology* **35**:571-596.
- European Commission. 2016. PRATIQUÉ: enhancements of pest risk analysis techniques. European Commission's 7th Framework Programme for Research and Technological Development. Last accessed January, 2016, <https://secure.fera.defra.gov.uk/pratique/index.cfm>.
- Feng, H. Q., F. Gould, Y. X. Huang, Y. Y. Jiang, and K. M. Wu. 2010. Modeling the population dynamics of cotton bollworm *Helicoverpa armigera* (Hubner) (Lepidoptera: Noctuidae) over a wide area in northern China. *Ecological Modelling* **221**:1819-1830.
- FHWA. 2015. Freight Analysis Framework. United States Department of Transportation, Federal Highway Administration, Washington, D.C. Last accessed [http://www.ops.fhwa.dot.gov/freight/freight\\_analysis/faf/](http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/).
- Fowler, G., Y. Takeuchi, R. Sequeira, G. Lougee, W. Fussell, M. Simon, A. Sato, and X. Yan. 2009. Pathway-Initiated Pest Risk Assessment: Asian Gypsy Moth (Lepidoptera: Lymantriidae: *Lymantria dispar* (Linnaeus)) from China into the United States on Maritime Shipments. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Raleigh, NC.
- Fowler, G., Y. Takeuchi, T. Vo, and L. Garrett. 2014. Pine shoot beetle, *Tomicus piniperda* (Linnaeus): analysis of regulatory opinions. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Center for Plant Health Science and Technology, Plant Epidemiology and Risk Analysis Laboratory, Raleigh, NC.
- Framework. n.d. In Merriam-Webster.com. Last accessed <http://www.merriam-webster.com/dictionary/framework>.

- Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold, and J. Wickham. 2011. Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering & Remote Sensing* **77**:858-864.
- Gao, B. C. 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* **58**:257-266.
- Garcia Morales, M., B. D. Denno, D. R. Miller, G. L. Miller, Y. Ben-Dov, and N. B. Hardy. 2016. ScaleNet: A literature-based model of scale insect biology and systematics. <http://scalenet.info>.
- Giorgi, F., and L. O. Mearns. 1999. Introduction to special section: Regional climate modeling revisited. *Journal of Geophysical Research-Atmospheres* **104**:6335-6352.
- Graham, R. L., C. T. Hunsaker, R. V. O'Neill, and B. L. Jackson. 1991. Ecological risk assessment at the regional scale. *Ecological Applications* **1**:196-206.
- Gutierrez, A. P., J. U. Baumgaertner, and C. G. Summers. 1984. Multitrophic models of predator-prey energetics: I. age-specific energetics models - pea aphid *Acrythosiphon pisum* (Homoptera: Aphididae) as an example. *The Canadian Entomologist* **116**:923-932.
- Gutierrez, A. P., N. J. Mills, and L. Ponti. 2010. Limits to the potential distribution of light brown apple moth in Arizona-California based on climate suitability and host plant availability. *Biological Invasions* **12**:3319-3331.
- Haack, R. A., F. Herard, J. H. Sun, and J. J. Turgeon. 2010. Managing invasive populations of Asian longhorned beetle and Citrus longhorned beetle: a worldwide perspective. Pages 521-546 *Annual Review of Entomology*. Annual Reviews, Palo Alto.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**:1965-1978.
- Holcombe, T., C. Jarnevich, C. Talbert, and M. Talbert. 2014. SAHM: VisTrails (Software for Assisted Habitat Modeling for VisTrails). United States Geological Survey, Fort Collins Science Center, Ft. Collins, CO.
- Holt, J., A. W. Leach, G. Schrader, F. Petter, A. MacLeod, D. J. van der Gaag, R. H. A. Baker, and J. D. Mumford. 2014. Eliciting and combining decision criteria using a limited palette of utility functions and uncertainty distributions: illustrated by application to pest risk analysis. *Risk Analysis* **34**:4-16.
- Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J. N. VanDriel, and J. Wickham. 2007. Completion of the 2001 National Land Cover

- Database for the conterminous United States. *Photogrammetric Engineering & Remote Sensing* **73**:337-341.
- Homer, C. G., J. A. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. D. Herold, J. D. Wickham, and K. Megown. 2015. Completion of the 2011 national land cover database for the conterminous United States: representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing* **81**:345-354.
- Hulme, P. E. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology* **46**:10-18.
- IPCC. 2016a. Organization. Intergovernmental Panel on Climate Change. Last accessed Jauary, 2016, <https://www.ipcc.ch/organization/organization.shtml>.
- IPCC. 2016b. What is a GCM? Intergovernmental Panel on Climate Change (IPCC), United Nation. Last accessed January, 2016, [http://www.ipcc-data.org/guidelines/pages/gcm\\_guide.html](http://www.ipcc-data.org/guidelines/pages/gcm_guide.html).
- IPPC. 2004. ISPM No. 11: Pest risk analysis for quarantine pests including analysis of environmental risk and living modified organisms. International Plant Protection Convention, Food and Agriculture Organization, Rome, Italy.
- James, S., and K. Anderson. 1998. On the need for more economic assessment of quarantine policies. *Australian Journal of Agricultural and Resource Economics* **42**:425-444.
- Jarnevich, C. S., W. E. Esaias, P. L. A. Ma, J. T. Morisette, J. E. Nickeson, T. J. Stohlgren, T. R. Holcombe, J. M. Nightingale, R. E. Wolfe, and B. Tan. 2014a. Regional distribution models with lack of proximate predictors: Africanized honeybees expanding north. *Diversity and Distributions* **20**:193-201.
- Jarnevich, C. S., T. R. Holcombe, E. M. Bella, M. L. Carlson, G. Graziano, M. Lamb, S. S. Seefeldt, and J. Morisette. 2014b. Cross-scale assessment of potential habitat shifts in a rapidly changing climate. *Invasive Plant Science and Management* **7**:491-502.
- Jarnevich, C. S., and L. V. Reynolds. 2011. Challenges of predicting the potential distribution of a slow-spreading invader: a habitat suitability map for an invasive riparian tree. *Biological Invasions* **13**:153-163.
- Kalia, V. K., and B. Yadav. 2015. Cost-effective mass rearing of the Oriental fruit fly, *Bactrocera dorsalis* (Hendel) round the year. *International Research Journal of Natural and Applied Sciences* **2**:143-155.
- Liebhold, A., and J. Bascompte. 2003. The Allee effect, stochastic dynamics and the eradication of alien species. *Ecology Letters* **6**:133-140.

- Liebhold, A., W. D. Koenig, and O. N. Bjornstad. 2004. Spatial synchrony in population dynamics. *Annual Review of Ecology Evolution and Systematics* **35**:467-490.
- Liebhold, A. M., E. G. Brockerhoff, L. J. Garrett, J. L. Parke, and K. O. Britton. 2012. Live plant imports: the major pathway for forest insect and pathogen invasions of the US. *Frontiers in Ecology and the Environment* **10**:135-143.
- Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, and J. W. Merchant. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing* **21**:1303-1330.
- MACA. 2016. Multivariate Adaptive Constructed Analogs (MACA) statistical downscaling method. University of Idaho. Last accessed January, 2016, <http://maca.northwestknowledge.net/>.
- MacLeod, A., J. Head, and A. Gaunt. 2004. An assessment of the potential economic impact of Thrips palmi on horticulture in England and the significance of a successful eradication campaign. *Crop Protection* **23**:601-610.
- Manel, S., S. T. Buckton, and S. J. Ormerod. 2000. Testing large-scale hypotheses using surveys: the effects of land use on the habitats, invertebrates and birds of Himalayan rivers. *Journal of Applied Ecology* **37**:756-770.
- McLeod, L. J., D. W. Hine, P. M. Please, and A. B. Driver. 2015. Applying behavioral theories to invasive animal management: towards an integrated framework. *Journal of Environmental Management* **161**:63-71.
- Mearns, L. O., R. Arritt, S. Biner, M. S. Bukovsky, S. McGinnis, S. Sain, D. Caya, J. Correia, D. Flory, W. Gutowski, E. S. Takle, R. Jones, R. Leung, W. Moufouma-Okia, L. McDaniel, A. M. B. Nunes, Y. Qian, J. Roads, L. Sloan, and M. Snyder. 2012. The north American regional climate change assessment program: overview of phase I results. *Bulletin of the American Meteorological Society* **93**:1337-1362.
- Mearns, L. O., S. Sain, L. R. Leung, M. S. Bukovsky, S. McGinnis, S. Biner, D. Caya, R. W. Arritt, W. Gutowski, E. Takle, M. Snyder, R. G. Jones, A. M. B. Nunes, S. Tucker, D. Herzmann, L. McDaniel, and L. Sloan. 2013. Climate change projections of the North American Regional Climate Change Assessment Program (NARCCAP). *Climatic Change* **120**:965-975.
- Morisette, J. T., C. S. Jarnevich, T. R. Holcombe, C. B. Talbert, D. Ignizio, M. K. Talbert, C. Silva, D. Koop, A. Swanson, and N. E. Young. 2013. VisTrails SAHM: visualization and workflow management for species habitat modeling. *Ecography* **36**:129-135.

- NARCCAP. 2016. North American Regional Climate Change Assessment Program. North American Regional Climate Change Assessment Program, National Center for Atmospheric Research (NCAR), Boulder, CO. Last accessed January, 2016, <http://www.narccap.ucar.edu/>.
- NASA. 2016. LDAS Land Data Assimilation Systems. National Aeronautics and Space Administration, Washington, D.C. Last accessed January, 2016, <http://ldas.gsfc.nasa.gov/nldas/>.
- NASS. 2016. CropScape and Cropland Data Layers: FAQ's. USDA National Agricultural Statistics Service, Washington, DC. Last accessed May, 2016, [https://www.nass.usda.gov/Research\\_and\\_Science/Cropland/sarsfaqs2.php](https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php).
- NCEI. 2016. National Centers for Environmental Information. Department of Commerce, National Oceanic and Atmospheric Administration, Asheville, NC. Last accessed January, 2016, <http://www.ncdc.noaa.gov/>.
- NCEP, N. 2016. RTMA documentation. National Oceanic and Atmospheric Administration, National Weather Service, National Centers for Environmental Prediction, College Park, MD. Last accessed [http://nomads.ncep.noaa.gov/txt\\_descriptions/RTMA\\_doc.shtml](http://nomads.ncep.noaa.gov/txt_descriptions/RTMA_doc.shtml).
- NRCS. 2016. Description of SSURGO database. United States Department of Agriculture, Natural Resources Conservation Service, Washington, DC. Last accessed January, 2016, [http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=nrcs142p2\\_053627](http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=nrcs142p2_053627).
- NWS. 2016. NDFD, NWS National Digital Forecast Database. United States Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, MD. Last accessed January, 2016, <http://www.nws.noaa.gov/ndfd/>.
- O'Connell, B. O., E. B. LaPoint, J. A. Turner, T. Ridley, S. A. Pugh, A. M. Wilson, K. L. Waddell, and B. L. Conkling. 2015. The Forest Inventory and Analysis database: database description and user guide for phase 2 (version 6.0.2). United States Forest Service, Washington, D.C.
- Ocallaghan, J. F., and D. M. Mark. 1984. The extraction of drainage networks from digital elevation data. *Computer Vision Graphics and Image Processing* **28**:323-344.
- Osborne, P. E., J. C. Alonso, and R. G. Bryant. 2001. Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *Journal of Applied Ecology* **38**:458-471.

- OTTI. 2016. About OTTI. United States Department of Commerce, International Trade Administration, Office of Travel and Tourism Industries, Washington, D.C. Last accessed January, 2016, <http://travel.trade.gov/about/index.html>.
- Oyler, J. W., A. Ballantyne, K. Jencso, M. Sweet, and S. W. Running. 2015. Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. *International Journal of Climatology* **35**:2258-2279.
- PLC, I. 2016. Lloyd's List Intelligence. Informa PLC. <http://www.lloydslistintelligence.com/llint/index.htm>.
- Pluess, T., V. Jarosik, P. Pysek, R. Cannon, J. Pergl, A. Breukers, and S. Bacher. 2012. Which factors affect the success or failure of eradication campaigns against alien species? *Plos One* **7**:11.
- Poland, T. M., and D. G. McCullough. 2006. Emerald ash borer: invasion of the urban forest and the threat to North America's ash resource. *Journal of Forestry* **104**:118-124.
- PRISM. 2015. PRISM climate data. Oregon State University, Northwest Alliance for Computational Science and Engineering, PRISM Climate Group, Corvallis, OR. Last accessed December, 2015, <http://www.prism.oregonstate.edu/>.
- Regan, H. M., M. Colyvan, and M. A. Burgman. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* **12**:618-628.
- Richardson, C. H., and D. J. Nemeth. 1991. Hurricane-borne African Locusts (*Schistocerca gregaria*) on the Windward Islands. *GeoJournal* **23**:349-357.
- Robel, J., and A. Graumann. 2014. NOAA KLM user's guide with NOAA-N, N Prime, and MetOp supplements. National Oceanic and Atmospheric Administration; National Environmental Satellite, Data, and Information Service; National Climatic Data Center; Remote Sensing and Application Division, Asheville, NC.
- Rwomushana, I., S. Ekesi, C. Ogol, and I. Gordon. 2008. Effect of temperature on development and survival of immature stages of *Bactrocera invadens* (Diptera: Tephritidae). *Journal of Applied Entomology* **132**:832-839.
- Senay, S. D., S. P. Worner, and T. Ikeda. 2013. Novel three-step pseudo-absence selection technique for improved species distribution modeling. *Plos One* **8**:16.
- Soliman, T., M. C. M. Mourits, A. Lansink, and W. van der Werf. 2010. Economic impact assessment in pest risk analysis. *Crop Protection* **29**:517-524.

- Springer, Y. P., C. S. Jarnevich, D. T. Barnett, A. J. Monaghan, and R. J. Eisen. 2015. Modeling the present and future geographic distribution of the lone star tick, *Amblyomma americanum* (Ixodida: Ixodidae), in the continental United States. *American Journal of Tropical Medicine and Hygiene* **93**:875-890.
- Suiter, K. (2015). CIPM projects. Personal communication to on from presentation for Dean Linton visit to CIPM on November 12, 2015.
- Surkov, I. V., A. Lansink, and W. van der Werf. 2009. The optimal amount and allocation of sampling effort for plant health inspection. *European Review of Agricultural Economics* **36**:295-320.
- Sutherst, R. W., G. F. Maywald, and D. Kriticos. 2007. CLIMEX version 3, user's guide. Hearne Scientific Software Pty Ltd, Melbourne, Australia.
- Takeuchi, Y. 2015. Strengthening national beet armyworm (*Spodoptera exigua*) programme: pest forecast system (Mission 3). Food and Agriculture Organization of the United Nations, Kingston, Jamaica.
- Takeuchi, Y., T. vo, and L. Garrett. 2015. Emerald ash borer, *Agrilus planipennis* (Fairmaire): an analysis of the regulatory program., United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Center for Plant Health Science and Technology, Plant Epidemiology and Risk Analysis Laboratory, Raleigh, NC.
- The World Bank. 2016. World Development Indicators. The World Bank, Washington, D.C. Last accessed January, 2016, <http://data.worldbank.org/data-catalog/world-development-indicators>.
- Thornton, P. E., S. W. Running, and M. A. White. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology* **190**:214-251.
- Tomlinson, R. F. 1987. Current and potential uses of geographical information systems: the North American experience. *International Journal of Geographical Information Systems* **1**:203-218.
- Trumble, J. T., and J. P. Morse. 1993. Economics of integrating the predaceous mite *Phytoseiulus persimilis* (Acari: Phytoseiidae) with pesticides in strawberries. *Journal of Economic Entomology* **86**:879-885.
- UCIPM. 2016. Weather, models, & degree-days. University of California, Agriculture & Natural Resources, Statewide Integrated Pest Management Program, Davis, California. Last accessed January, 2016, <http://www.ipm.ucdavis.edu/WEATHER/index.html>.

- UNISYS. 2016. Business description. UNISYS, Blue Bell, PA. Last accessed February, 2016, <http://www.unisys.com/>.
- USAID. 1989. Locust and grasshopper control in Africa/Asia: a programmatic environmental assessment. United States Agency for International Development.
- USCB. 2016. About the Bureau. United States Department of Commerce, United States Census Bureau, Suitland, MD. Last accessed January, 2016, <http://www.census.gov/about.html>.
- USDA APHIS. 2010. Emergency response manual. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD.
- USDA APHIS. 2012. Fresh fruits and vegetables manual. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD.
- USDA APHIS. 2014. Cut flowers and greenery import manual. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD.
- USDA APHIS. 2015a. Emerald ash borer program manual. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD.
- USDA APHIS. 2015b. Treatment manual. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD.
- USFS. 2014. The Forest Inventory and Analysis database: database description and user guide for phase 3 (version 6.0.1). United States Department of Agriculture, Forest Service, Washington, D.C. <http://www.fia.fs.fed.us/>.
- USFS. 2016a. Forest Inventory and Analysis: lichen communities indicators. United States Department of Agriculture, Forest Service, Washington, D.C. Last accessed <http://www.fia.fs.fed.us/>.
- USFS. 2016b. Forest Inventory and Analysis: ozone indicator. United States Department of Agriculture, Forest Service, Washington, D.C. <http://www.fia.fs.fed.us/>.
- USFS. 2016c. Forest Inventory and Analysis: soil quality indicator. United States Department of Agriculture, Forest Service, Washington, D.C. <http://www.fia.fs.fed.us/>.

- USFS. 2016d. Forest Inventory and Analysis: vegetation indicator. United States Department of Agriculture, Forest Service, Washington, D.C.  
<http://www.fia.fs.fed.us/>.
- USFS. 2016e. National forest type dataset. United States Department of Agriculture, Forest Service (USFS), Washington, D.C. Last accessed January, 2016,  
<http://www.fia.fs.fed.us/library/maps/>;  
[http://data.fs.usda.gov/geodata/rastergateway/forest\\_type/](http://data.fs.usda.gov/geodata/rastergateway/forest_type/).
- USGS. 2015. 3D Elevation Program (3DEP). United States Department of Interior, United States Geological Survey, Reston, VA. Last accessed  
<http://nationalmap.gov/3DEP/index.html>.
- Venette, R. C., and S. D. Cohen. 2006. Potential climatic suitability for establishment of *Phytophthora ramorum* within the contiguous United States. *Forest Ecology and Management* **231**:18-26.
- Vogelman, J., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and J. N. Van Driel. 2001. Completion of the 1990's National Land Cover Data set for the conterminous United States. *Photogrammetric Engineering & Remote Sensing* **67**:650-662.
- Wallner, W. E. 1987. Factors affecting insect population dynamics: differences between outbreak and non-outbreak species. *Annual Review of Entomology* **32**:317-340.
- Walters, B. A., J. J. Jiang, and G. Klein. 2003. Strategic information and strategic decision making: the EIS/CEO interface in smaller manufacturing companies. *Information & Management* **40**:487-495.
- White, I. M., and M. M. Elson-Harris. 1992. Fruit flies of economic significance: their identification and bionomics. C.A.B. International, Wallingford, UK.
- Wittwer, G., S. McKirdy, and R. Wilson. 2005. Regional economic impacts of a plant disease incursion using a general equilibrium approach. *Australian Journal of Agricultural and Resource Economics* **49**:75-89.
- Yang, Y., L. T. Wilson, and J. Wang. 2010. Development of an automated climatic data scraping, filtering and display system. *Computers and Electronics in Agriculture* **71**:77-87.
- Zlotina, M. A., V. C. Mastro, J. S. Elkinton, and D. E. Leonard. 1999. Dispersal tendencies of neonate larvae of *Lymantria mathura* and the Asian form of *Lymantria dispar* (Lepidoptera : Lymantriidae). *Environmental Entomology* **28**:240-245.

## CHAPTER 4. Uncertainty Associated with Pest Forecasts

### 4.1 Introduction

Pest forecast models such as phenology models, probabilistic models, species distribution models (SDMs), and population dynamics models, are widely used quantitative tools to evaluate pest behaviors and potential impacts caused by pests (Stinner et al. 1975, Elith and Leathwick 2009, Koch et al. 2009, Gutierrez et al. 2010, Yemshanov et al. 2015); Such prediction and forecast outputs contain errors because these models assume certain relationships between pest biology and input drivers (e.g. climate data, host species availability). Those errors from statistical evaluations lead to uncertainty in predictions.

There are three main sources of uncertainty for insect predictive models. The first source is pest biology. Even if a given species (e.g., an insect pest or disease causal agent) is well studied elsewhere, it is still uncertain how it will react to a new environment. Insect pests such as *Agrilus planipennis* (emerald ash borer) and *Anoplophora glabripennis* (Asian longhorned beetle) are considered non-invasive and non-damaging pests in native habitats (China). However, when those two species were introduced to new environments (continental United States), they have caused significant economic and environmental damage to U.S. forests (Poland and McCullough 2006, Haack et al. 2010). Therefore, predicting economic impacts based on the biology obtained in different regions involves different degrees of uncertainty.

The second source of uncertainty results from input data drivers. Almost all input data drivers used in insect predictive models are interpolated from space and time-based observations. For example, some commonly used climate datasets from NOAA, PRISM

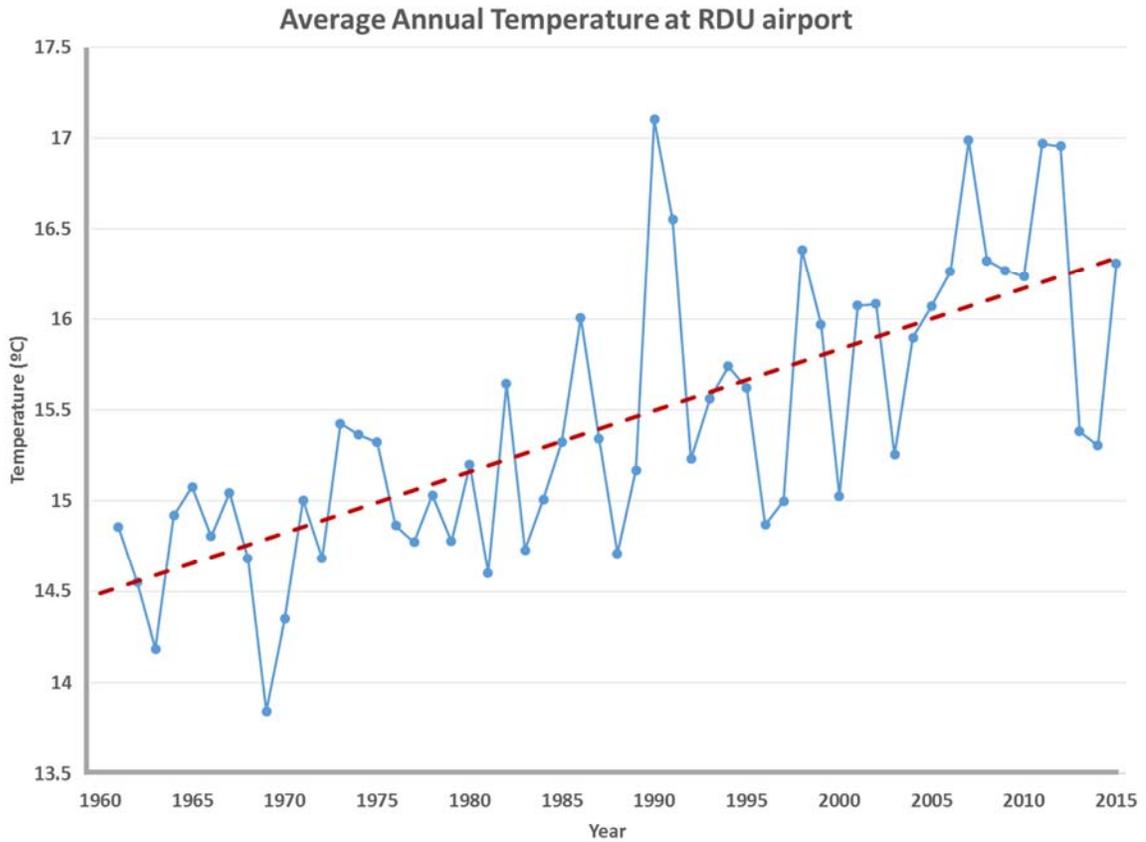
and METDATA are interpolated across the space between ground weather stations based on the mathematical relationships between the land features and observed daily weather data at weather stations (Abatzoglou 2013, Daly and Bryant 2013). Host species distribution and host availability are often also derived products. If a researcher needs uniform data (e.g., a uniform grid), most often these gridded data points are estimated based on climate data, survey data, and satellite images (O'Connell et al. 2015, USDA NASS 2016). Even without interpolation when using point data alone (i.e., calculations based on a single point corresponding to a ground-based weather station or observation point), there is uncertainty due to measurement and human errors (Elith et al. 2002).

The third source is linked to errors introduced by the predictive models themselves. Pest predictive models evaluate pests based on the mathematical relationship derived from known pest information. Even when the predictive model is parameterized well and has a high goodness of fit, the result contains variance across the space and uncertainty across time.

Since most damaging organisms (insect pests, diseases) exhibit a strongly poikilotherm behavior, temperature is one of the major variables that influences growth and development. Overall climate becomes a primary driver to affect development and population dynamics. The plant host is a critical component as well. Since I am considering highly perturbed agricultural and commercial forestry conditions with abundant and productive host material, I will assume that this critical component can be addressed outside the scope of this dissertation. Understanding historical weather patterns, especially temperature, is important and critical for insect predictive modeling. However,

understanding historical weather patterns is very challenging because weather fluctuates significantly day to day and year to year (Figure 4-1). Figure 4-1 is an illustration of these variable patterns. The figure shows average annual temperature recorded at RDU (Morrisville, NC) airport since 1961 (NOAA 2016a). The average annual temperature varies each year; the coldest average annual temperature was in 1969 at 13.8°C, while the hottest average annual temperature was in 1990 at 17.1°C.

In addition to intrinsic variability, temperature for the foreseeable future is expected to have a generally increasing pattern due to global warming (Figure 4-1). The figure below shows consistent increases in annual temperatures measured at a local airport (RDU) since 1961. According to this trend line, the average annual temperature has increased approximately 1.8°C since 1961 at RDU airport. The patterns differ at any given location; however, global average patterns show that the same trend observed locally characterizes global scale observations and expectation (IPCC 2013).



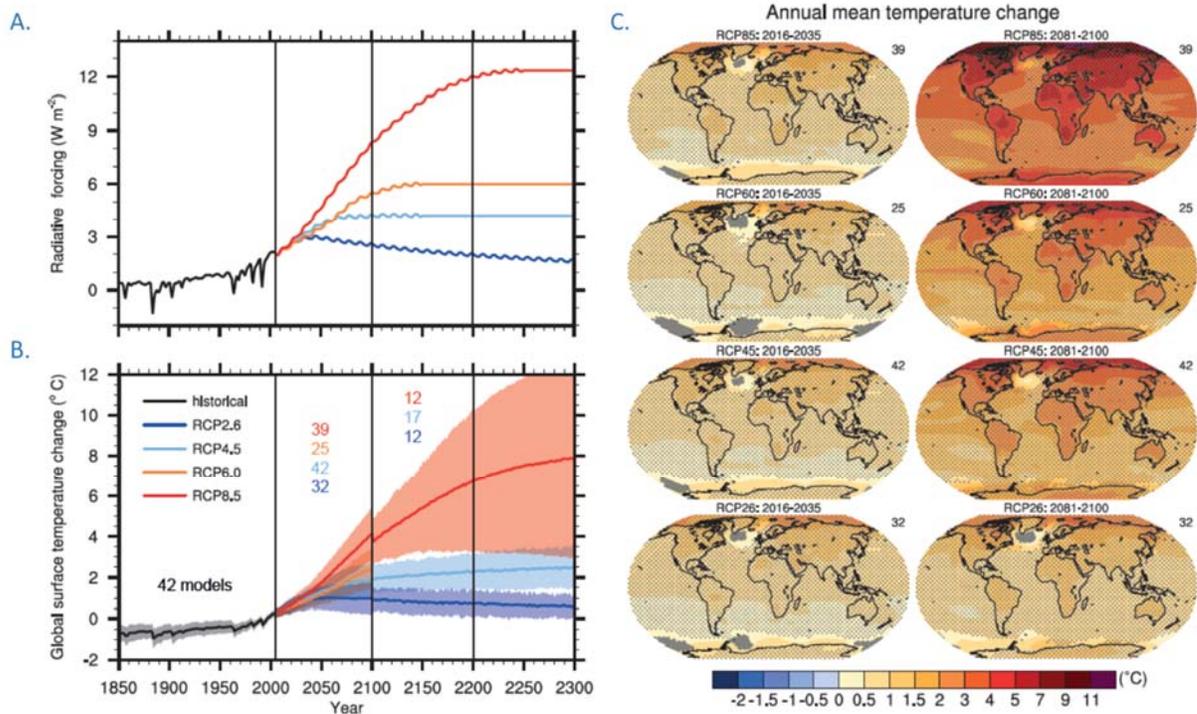
**Figure 4-1.** Average annual temperature at RDU airport (35.633°N, -077.383°W) (NOAA 2016a)

The global warming is caused by increases in greenhouse gasses (H<sub>2</sub>O, N<sub>2</sub>O, CO<sub>2</sub>, CH<sub>4</sub>, and CFCs) in the atmosphere. In average, CO<sub>2</sub> concentrations measured in 2005 were 35% higher than the ones observed during the Industrial Revolution period (1760 - 1840) (Azar and Rodhe 1997, UCAR 2016). The concentration of CO<sub>2</sub> varies by season or month; however, mean 2016 January concentration readings in Mauna Loa, Hawaii was 402.5ppm, while mean 1996 and 2006 January readings were 381.2ppm and 362.0ppm, respectively (NOAA 2016b).

Increased temperature causes significant impacts on insect growth (Porter et al. 1991, Cannon 1998, Bale et al. 2002). In general, warmer condition increases insect growth rate, population density, and suitable habitat (Cannon 1998, Bale et al. 2002). However,

successful life-cycle completion also involves host availability and phenology (Bale et al. 2002). The host species may not react at the same speed with the feeding insect, and insect-host synchrony may be unbalanced under climate change (Bale et al. 2002). In addition, each species (both insect and plant) reacts differently under climate change. Species, whose growth rates respond greatly with higher temperature, tend to benefit under climate change by reducing developmental time, while species, with a slow growth rate or who do not react as much to higher temperature, may not be affected under climate change (Bale et al. 2002). Species may also respond to higher temperature and elevated greenhouse gas concentration differently on different host species (Lindroth et al. 1993). Lindroth et al. (1993) stated that the performance of *Lymantria dispar* larvae was different between aspen and oak that were grown under elevated CO<sub>2</sub> environment.

Estimating insect impact under climate change is a complex process and a challenge (Cannon 1998, Bale et al. 2002). There are a quite number of imprecise facts about how insects and hosts react under climate change. It is not only uncertain what the insect/host reaction to climate change is but also unclear how much greenhouse gas concentration will actually increase over time. Current IPCC (Intergovernmental Panel on Climate Change) assessment report (AR5) states four greenhouse gas emission scenarios (Representative Concentration Pathways: RCP2.6, RCP 4.5, RCP6, and RCP 8.5) (van Vuuren et al. 2011, IPCC 2013) (Figure 4-2). Those four scenarios forecast similar emission levels until 2020; however, those models forecast a significantly different level of greenhouse gas concentrations beyond that date continuing to show great variation and departure from current conditions through 2100.



**Figure 4-2. (A) Total global mean radiative forcing for the four RCP scenarios based on the Model for the Assessment of Greenhouse-gas Induced Climate Change (MAGICC) energy balance model (Note that the actual forcing simulated by the CMIP5 models differs slightly between models). (B) Time series of global annual mean surface air temperature anomalies (relative to 1986–2005) from CMIP5 concentration-driven experiments. (C) Multi-model ensemble average of annual mean surface air temperature change (compared to 1986–2005 base period) for 2016–2035 and 2081–2100, for RCP 2.6, 4.5, 6.0 and 8.5. (IPCC 2013)**

This chapter tests the hypothesis that methods can be developed to incorporate explicit uncertainty into a pest forecast framework. I use the Spatial Analytic Framework for Advanced Information System (SAFARIS) to explore a forecast model and use the framework with this model to evaluate uncertainty associated with *Lymantria dispar* Linnaeus (gypsy moth) suitability estimations under climate change within the continental United States. For simplicity, I do not consider changes related to changes in the behavior, abundance or distribution of hosts (e.g., oak) in this study.

Characterizing, evaluating and communicating uncertainty associated with forecast uses a multistage process. The process begins by identifying an appropriate forecast model;

testing, verifying, validating the behavior of the model; linking the model to the existing data structures provided by SAFARIS; using an adaptation of the mean-variance frontier concept (e.g., as discussed by Yemshanov et al. (2015)) and then exploring how to best characterize and communicate uncertainty. The uncertainty associated with gypsy moth suitability forecasts using the outputs of general circulation models (GCMs) was calculated and visualized using a straightforward mapping of error terms.

## **4.2 Methods**

### **4.2.1 Climate Data**

Downscaling refers to the process to derive data at a given resolution that is more highly resolved than the original data set. For example, monthly data can be used to obtain estimates of daily weather patterns. Alternatively, data collected at a given spatial resolution (e.g., 30-km grid) can be downscaled to a more highly resolved grid (e.g., 15-km grid). The Multivariate Adaptive Constructed Analogs (MACA) is a statistical downscaling method applied to original GCMs to create daily climatology at 4-km to 6-km resolution for the conterminous United States (details presented in Chapter 3; also terminology and acronyms used here are described in Chapter 3).

The monthly downscaled MACA GFDL-ESM2G (NOAA Geophysical Fluid Dynamics Laboratory, USA) for RCP 4.5 and 8.5 were used to generate 19 Bioclim variables that Hijmans et al. (2005) originally established as a core set for representing future climate (Appendix D). The Bioclim variables were first generated from both RCP 4.5 and 8.5 models at each year from 2016 to 2090 and averaged for three time periods: 2016-

2040, 2041-2065, and 2066-2090 using the equations illustrated by O'Donnel and Ignizio (2012). The equations used to calculate 19 Bioclim variables are listed in Appendix D.

Conversion tools from monthly data to bioclimatic variables were developed and available in SAFARIS. The original MACA GCM datasets were stored in SAFARIS. In this study, I only used GFDL-ESM2G GCM datasets; however MACA provides downscaled GCM datasets from 20 different models. SAFARIS can assist calculating bioclimatic variables for specific time periods and specific GCMs based on the users' needs.

#### **4.2.2 Gypsy moth distribution**

The United States Department of Agriculture, Animal and Plant Health Inspection Service (USDA APHIS) has conducted surveys for gypsy moth (Cooperative Agricultural Pest Survey program (CAPS)) for the past four decades. There are over 1.2 million survey records for gypsy moth in the continental United States (1974 through 2015).

The CAPS survey records (e.g., trap collection date, number of pest trapped, and trap locations) are available through APHIS' domestic survey databases. However, some records are missing coordinate information and were therefore removed from the analysis.

The main objective of the CAPS program is to detect new introductions of harmful plant pests and diseases as soon as possible before there is resulting significant damage (<https://www.aphis.usda.gov/aphis/ourfocus/planthealth>). Therefore, the areas where gypsy moth is widely distributed are not surveyed. As soon as an area is generally acknowledged to be infested, surveys are no longer pursued. Thus, there are no positive trap records in the areas where gypsy moth is known to be occurring (definite or confirmed suitable areas). Since CAPS program did not survey where APHIS had confirmed gypsy moth occurrence,

there were few survey records available for the areas that were suitable for gypsy moth growth in most of the regulated area. Therefore, 2,000 points to represent derived-presence (i.e., assumed to be present based on known distribution) points were created based on the current known distribution of gypsy moth and density of the host species. Those points were included as presence points in addition to CAPS survey positive points for MaxEnt analysis. This process was cross-checked with USDA analysts and forest pest experts (Fowler, personal communication 2015) .

In addition to presence points not being well represented though CAPS survey points, having many survey positive points in marginally suitable or unsuitable areas was also problematic. Some states may place traps in only certain years, while borderline states where gypsy moths are likely to be detected (e.g., North Carolina, Virginia, West Virginia, Ohio, and Indiana) are surveyed every year. The reality is that since efforts aim to delimit the infestation, many traps are placed on the edge of suitable areas where conditions may be suboptimal for long term establishment. To minimize noise, I only used positive survey points from 2015 because I was able to confirm the validity of most points with USDA program managers.

#### **4.2.3 Species Distribution Model**

A forecast model using maximum entropy called “MaxEnt” was used to predict suitable areas for gypsy moth establishment based on known distribution locations and 19 bioclimatic variables. This maximum entropy approach is one of many kinds of machine learning algorithms used to navigate a defined search space and establish patterns or find optima. Maximum entropy approaches were first used in applications linked to natural

language interpretation (Berger et al. 1996) and later for the characterization of suitable habitats for many species (Phillips et al. 2005). Phillips et al. (2005) provide a detailed description of the approach. MaxEnt Version 3.3.3k was used. That specific set of algorithms is available as part of the Software for Assisted Habitat Modeling (USGS SAHM) code which in turn is part of a suite of algorithms that constitutes a collection of open-source procedures known as VisTrails (Morissette et al. 2013) based on the R language and other extensions.

MaxEnt can be used to predict the probability of species presence from presence-only data based (Phillips et al. 2005). It generates values ranging from 0 to 1 to rank habitat suitability. MaxEnt also allows users to set aside a “random test percentage” to perform statistical analysis on model performance. It provides the area under the ROC (receiver operating characteristic) Curve (AUC) indicating the fit of the model (Phillips 2010). In addition, MaxEnt facilitates analysis of variable contributions (i.e., percent contribution and permutation importance for each variable used in MaxEnt) and jackknife tests to evaluate the importance of the variables (Phillips 2010).

The Bioclim 19 bioclimatic variables are known to represent annual trends, seasonality, and extremes of temperature and precipitation (Hijmans et al. 2005, Lozier and Mills 2011). Some studies showed that the 19 bioclimatic variables did not predict pest suitability well in MaxEnt, mainly because there were many correlated variables within the model (Fitzpatrick et al. 2008, Peterson and Nakazawa 2008, Lozier and Mills 2011). However, bioclimatic variables were nevertheless commonly used in MaxEnt to predict pest suitability (Pearson et al. 2007, Stohlgren et al. 2010, Webber et al. 2011) and appeared to

perform adequately. In this study, I decided to address correlations. Thus, I examined the correlation among the 19 variables before supplying the data into MaxEnt. As appropriate, one or more of the highly correlated variables based on Pearson's correlation coefficient ( $r > 0.8$ ) were removed to reduce problems associated with multicollinearity. In addition, variable selection was based on a given variable's contribution and jackknife tests.

I used 10-fold cross validation for testing the model performance for each climate period. Seventy (70) percent of the presence data points were used to train the model, and the remaining of 30% of the data were used for testing the model. Since the gypsy moth was not surveyed randomly, Elith et al. (2011) suggested generating random background points (10,000 random points) in MaxEnt. The background points were generated based on the kernel density of the presence data points and the USGS SAHM code was used for the basic model algorithm.

The parameterization of MaxEnt (e.g., selection of climatic variables as independent variables) was done based on the correlation evaluation among 19 variables and relevance to biology of gypsy moth by using bioclimatic variables generated from historical data. The same types of variables generated from GCMs were used to estimate the likelihood of gypsy moth establishment for the future.

Mean climatic variables for each time period under each climate change scenario were first supplied using SAFARIS and linked to MaxEnt code available as part of the code sets in USGS SAHM. In order to incorporate yearly climate variability into the analysis, 5 different sets of bioclimatic values were randomly generated at one standard deviation of the mean in SAFARIS and supplied to MaxEnt in addition to mean climatic variables.

In this study, I did not use information on forest species types and volume to run MaxEnt. Current forest condition information is available; however, those conditions under climate change are uncertain and not readily available. Therefore, only climatic data were used to estimate likelihood of gypsy moth establishment by MaxEnt. However, the topography information was included. I used Shuttle Radar Topography Mission (SRTM) elevation data (at 1km resolution) as one of the independent variables to estimate the habitat suitability for gypsy moth.

#### **4.2.4 Uncertainty Evaluation**

The uncertainty associated with model output was evaluated by exploring uncertainty associated with 1) suitability for gypsy moth growth, 2) standard deviation derived from MaxEnt model, and 3) standard deviation derived from averaging bioclimatic variables (Bioclim) from downscaled GCMs.

Yemshanov et al. (2015) described the mean-variance frontier concept (MVF) that could be applied to understand uncertainty associated with plant pests. They evaluated uncertainty associated with pest prediction output by plotting model output against standard deviation. The location can be prioritized by combining model output and standard deviation information. However, the ranking depends on managers and decision makers. One may focus on highly suitable areas regardless of uncertainty. Or one may focus on highly suitability areas with higher variance, while one may focus on highly suitability areas with lower variance (risk-averse).

The MVF approach was applied to incorporate mean probability and uncertainty to prioritize locations according to management needs. The mean probability from MaxEnt

model for 2066-2090 for each climate change scenario was examined against standard deviation to prioritize locations for decision making processes.

Each gridded cell has its mean probability and associated standard deviation results from species distribution model (MaxEnt). Instead of evaluating at each gridded cell, uncertainty evaluation was conducted at the county level because regulatory actions often take place at that level. The mean probability (y-axis) and standard deviation (x-axis) from each U.S. county in the contiguous United States were plotted to visualize the pattern of associated uncertainty.

In this study, I wanted to identify the areas that have high suitability with high variance. The locations with high suitability and high variance are placed at the right-upper quadrant of the scatter plot. Based on the scatter of the county values, the mean-standard deviation ratio at each county was calculated. The algorithm used to calculate relative risk is based on the value of probability and standard deviation. Then, each county is ranked based on the relative risk to determine the areas where decision makers should focus on pest management actions.

### **4.3 Results**

The results first use the SAFARIS environment to capture and examine the nature of climate data. The biological aspects and epidemiology of the case study organism, gypsy moth, are reviewed and also captured in the SAFARIS framework. The application of the MaxEnt model is examined to predict suitability of gypsy moth under climate change. Finally uncertainty associated with suitability prediction is evaluated. The uncertainty

algorithms presented were added as a feature (user defined option) within the SAFARIS modeling environment.

### **Climate Data**

The statistically downscaled monthly GFDL-ESM2G (NOAA Geophysical Fluid Dynamics Laboratory, USA) for RCP 4.5 and 8.5 were obtained from <http://maca.northwestknowledge.net/>. Multivariate Adaptive Constructed Analogs (MACA) provides daily and monthly climate data from 1990 to 2099. The monthly maximum temperature (Tmax), minimum temperature (Tmin), and precipitation datasets in netCDF format were similarly obtained.

Nineteen Bioclim variables were generated from monthly downscaled MACA GFDL-ESM2G RCP 4.5 and 8.5 models by the method originally proposed by (Hijmans et al. 2005, O'Donnell and Ignizio 2012). In order to understand the expected changes in the future, 19 Bioclim variables were also created based on GFDL-ESM2G base period (1990-2005) monthly data in SAFARIS.

In order to understand the climate change scenarios, annual temperature (Bioclim 1) and annual precipitation (Bioclim 12) were first examined and compared against historical climate data. Figure shows forecast annual mean temperature (Bioclim 1) for 2066-2090 using GFDL-ESM2G RCP 4.5 and 8.5 models and expected changes in 2066-2090 compared to current climate (base period: 1990-2005). Models for both scenarios (RCP 4.5 and 8.5) resulted in all locations within contiguous United States expected to have a higher annual mean temperature by 2066-2090. The model forecasts for the RCP 4.5 scenario

predicted that the annual mean temperature for 2066-2090 would increase 1.0 to 2.3°C, while the model for RCP 8.5 predicted 2.2 to 4.4°C increase by 2066-2090.

The expected annual precipitation (Bioclim 12) for 2066-2090 by GFDL-ESM2G RCP 4.5 and 8.5 models are shown in Figure 4-4. The expected changes were evaluated against the base period (1990-2005) under two scenarios (RCP 4.5 and 8.5) and are also shown in Figure 4-4. Results from those two models were slightly different; however, trends were similar. The southern United States, especially Texas, Louisiana, Mississippi, and southern Florida are expected to be drier than current conditions, while the northeastern states are expected to have significantly more precipitation by 2066-2090. Also, coastal California and the San Joaquin Valley are expected to have drier conditions by 2066-2090 from both models. The differences between RCP 4.5 and RCP 8.5 models were related to intensity. The RCP 8.5 model forecasted drier conditions than the RCP 4.5 model in the southern United States, especially for Florida, Texas, Louisiana, Mississippi, and California (Figure 4-4). On the other hand, the RCP 8.5 model predicted wetter conditions than the RCP 4.5 model in northern United States (Figure 4-4).

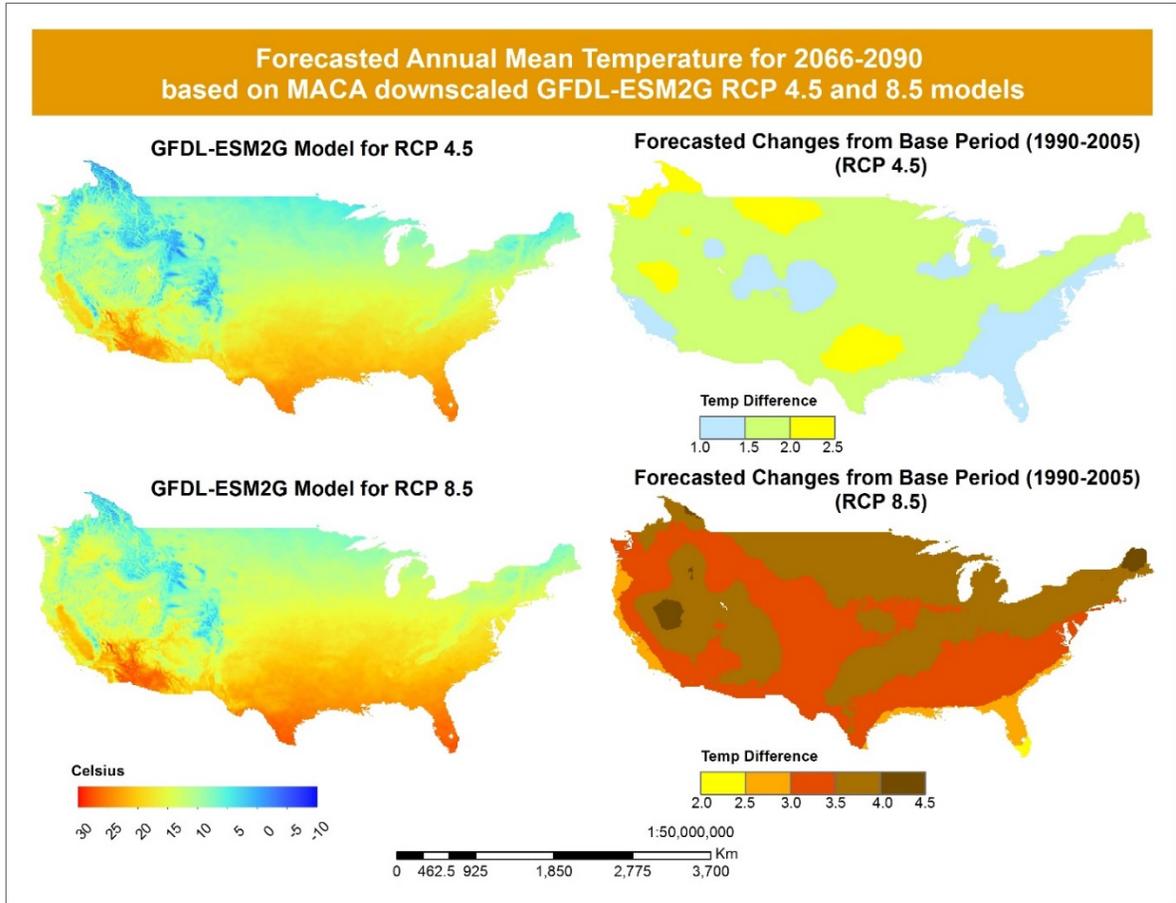
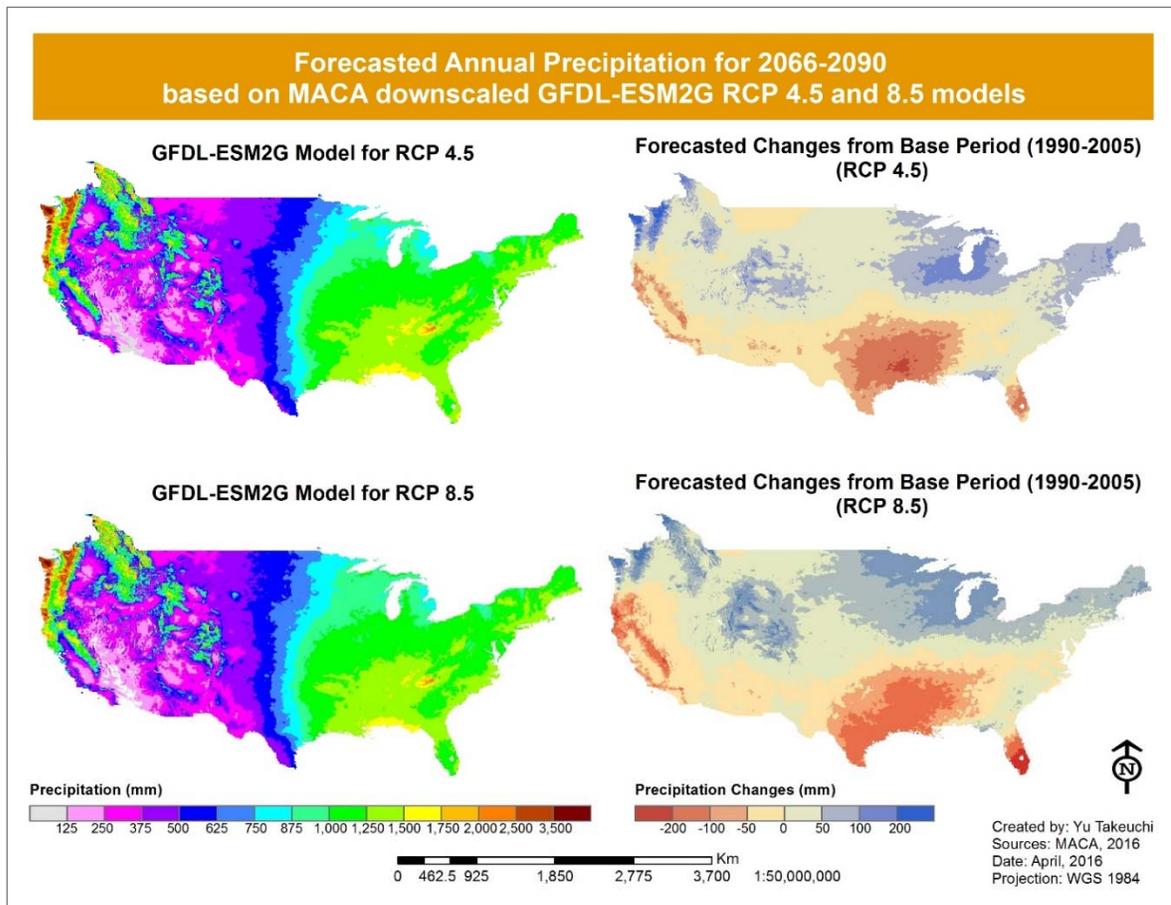


Figure 4-3. Forecasted annual mean temperature for 2066-2090 based on GFDL-ESM2G RCP 4.5 and 8.5 models and expected changes in annual mean temperature from base period of 1990-2005.



**Figure 4-4. Forecasted annual precipitation for 2066-2090 based on GFDL-ESM2G RCP 4.5 and 8.5 models and expected changes in annual precipitation from base period of 1990-2005.**

Although errors associated with downscaling (both spatial and temporal) processes in MACA products were not available, the standard deviation at each location was calculated by averaging each Bioclim variable over three time periods (2016-2040, 2041-2065, and 2066-2090). The maximum temperatures of the warmest month (Bioclim 5) were determined for each year from 2016 to 2090 and averaged by three time periods (Figure 4-5).

Overall, the GFDL-ESM2G model with RCP 8.5 scenario predicted higher temperatures than the model with RCP 4.5 scenario. The standard deviation maps indicated that the results from the RCP 4.5 model forecasted greater fluctuations in maximum

temperature of the warmest month among years. Both models resulted in smaller standard deviations in Southern states and western states. However, Southern Texas and Florida had smaller standard deviations from the RCP 4.5 model than the RCP 8.5 model. This indicates that climate conditions forecasted in the RCP 8.5 model are more uncertain. As Figure 4-2 indicates, the RCP 8.5 model assumes a significant increase in greenhouse gas in the future; however, the uncertainty on the amount of greenhouse gas increase becomes larger compare to the assumption for the RCP 4.5 model. North Central states had higher standard deviations within the contiguous United States. This could indicate that the maximum temperature of the warmest month in the North Central states is a value for which we have less confidence when using the GFDL-ESM2G models. The uncertainty associated with the maximum temperature of the warmest month estimated by the RCP 4.5 model was high in North Central states.

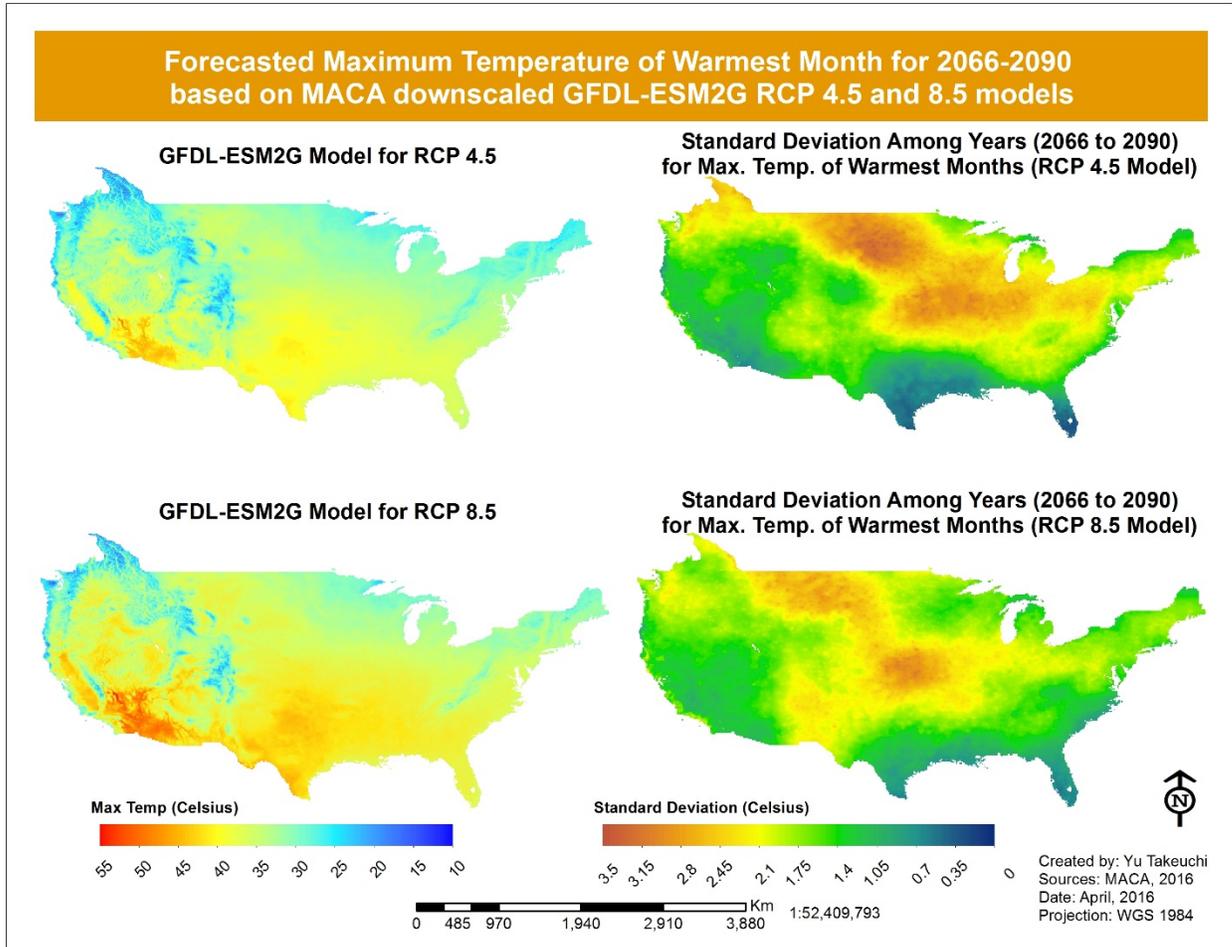
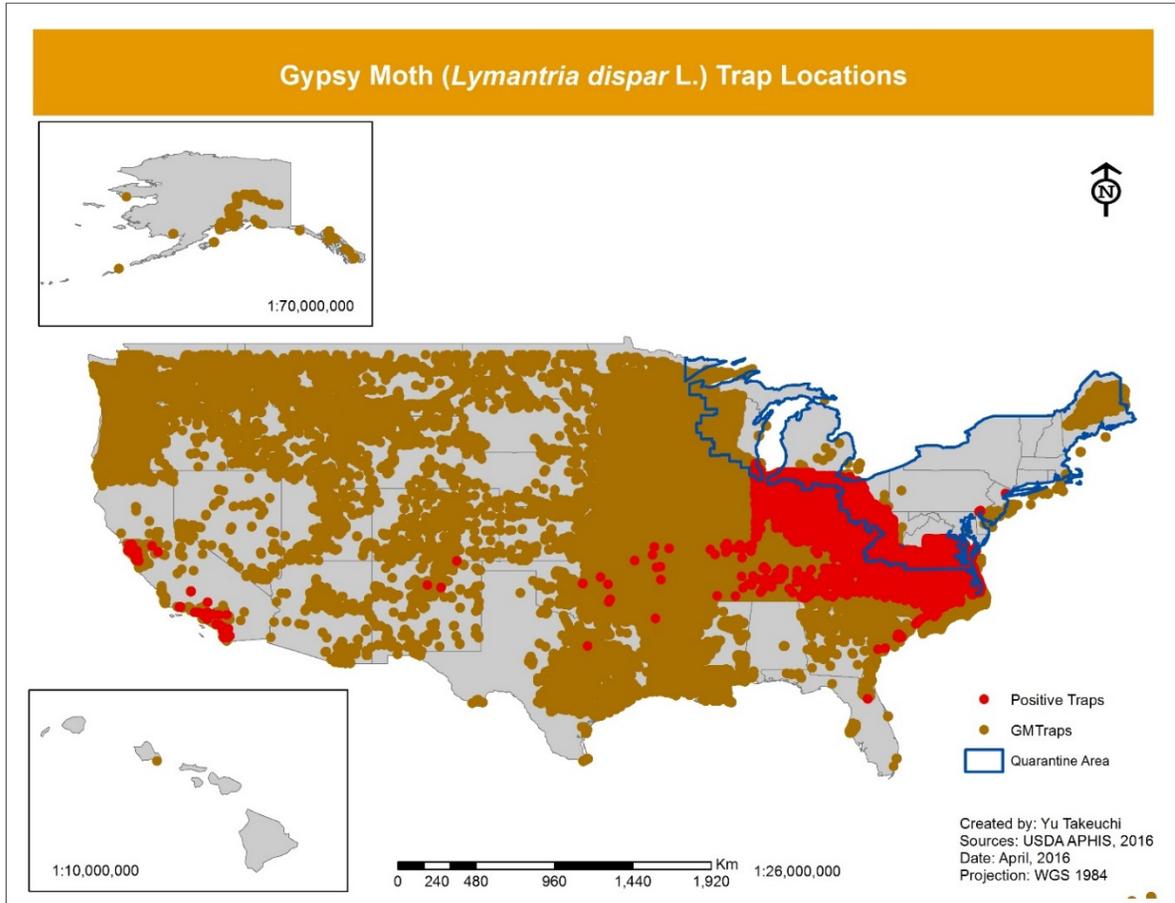


Figure 4-5. Forecasted maximum temperature of the warmest month for 2066-2090 based on GFDL-ESM2G RCP 4.5 and 8.5 models and standard deviations among years

### Gypsy moth distribution

The European gypsy moth (EGM), *Lymantria dispar* (Linnaeus), is currently distributed in nineteen (19) northeastern states and the District of Columbia in the United States (Figure 5-1). USDA APHIS conducts surveys for pests of elevated concern throughout the United States (Cooperative Agricultural Pest Survey (CAPS)). EGM is one of the pests that are currently surveyed, with surveys for EGM conducted on a national scale. Figure 4-6 shows the locations where traps for EGM have been placed since 1974.



**Figure 4-6. Survey trap locations for European gypsy moth**

The initial results revealed that the MaxEnt model with 19 Bioclim variables did not perform well for indicating the likelihood of gypsy moth establishment. The main reason was that positive findings of gypsy moth were found at large geographic areas (e.g., California, Florida, Texas, Arkansas, Georgia); therefore, the model predicted most of the United States as suitable areas (map not shown, it includes most of the continental United States).

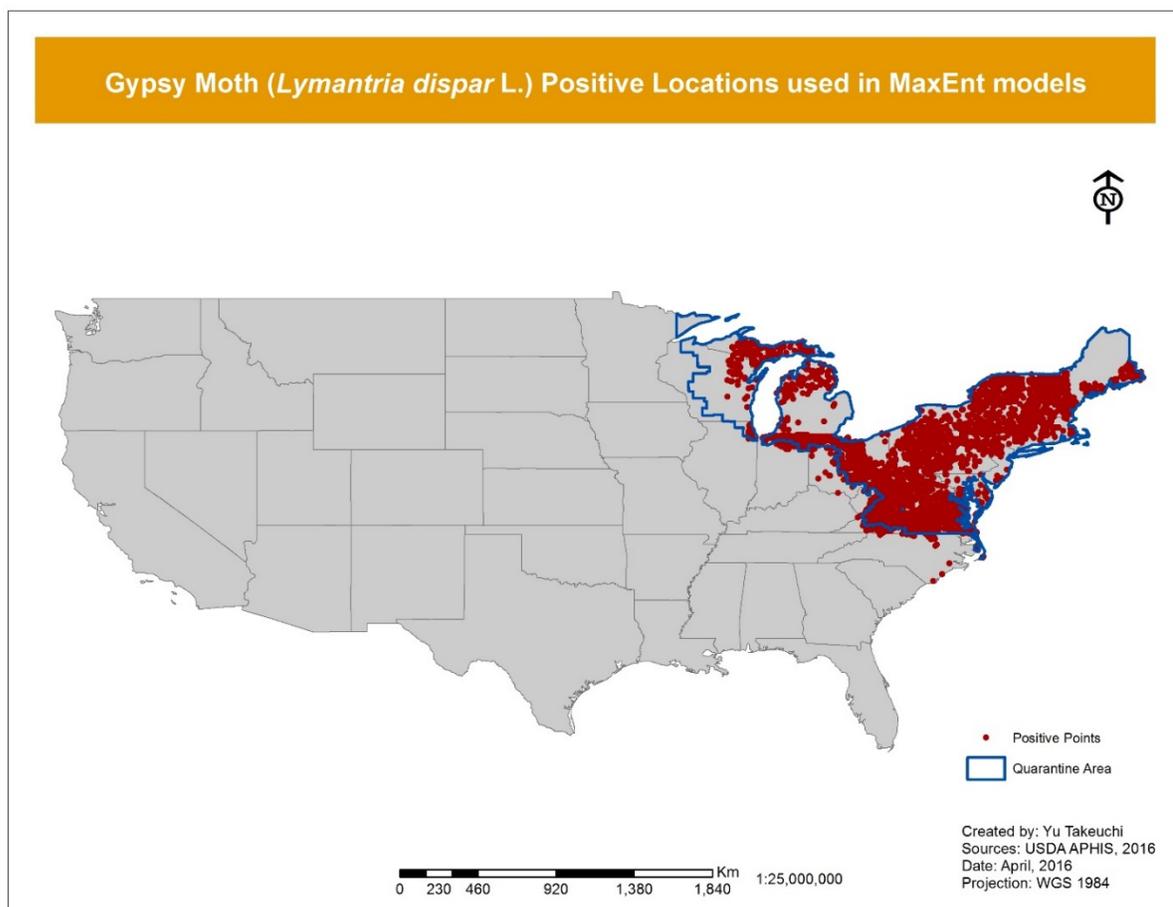
An epidemiological investigation of the nature of the CAPS data and the nature of USDA surveys was needed to better ascertain what the data records meant. The problem with CAPS survey data is that they record all gypsy moth findings, and there was no method

to differentiate positive records between detection records and establishment records. For example, large numbers of gypsy moth (Asian gypsy moth) had been detected at ports and near ports; however, these records did not represent gypsy moth establishment or suitable conditions for gypsy moth growth. A gypsy moth egg mass found in Florida is well known to be far beyond its climatic suitability range. But the database shows Florida interceptions are positive and does not distinguish such a record from those from areas where the moth has become endemic (e.g., in Pennsylvania or Michigan). Based on consultations with APHIS experts, some of the positive records that seemed to be detection points were removed as positive locations for MaxEnt simulations. The APHIS PPQ experts consulted included Glenn Fowler (risk analyst at Plant Epidemiology and Risk Analysis Laboratory), Anthony Man-Son Hing (National Operations Manager at Field Operations, Pest Management Programs), and David Lance (Assistant Director at Otis Laboratory).

The verified positive findings (not detections) included data from states that bordered the main infested area (area assumed to be largely infested and where the moth had become endemic). These borderline states are surveyed every year. In contrast, areas where the moth is known to occur are not surveyed (they do not show presence but rather no data).

I used 2015 survey records that I was able to correct using expert input and that could be confirmed to contain positive records to use as input MaxEnt models. A positive record was considered a record that was not transient and was likely located in an area where the moth could conceivably complete its life cycle.

The map below (Figure 4-7) shows the 2015 positive traps and 2,000 points generated based on current distribution of gypsy moth and forest density to better represent the distribution of gypsy moth in northeastern United States.



**Figure 4-7. Gypsy moth positive locations used in MaxEnt to estimate habitat suitability**

There is clearly a lot of variability in the observed data. As noted, some points do not have any data although the points are well-known to be infested for many years. For those, we derived presence using historical records, expert input, and tree distribution and abundance. There are also pseudo-positives. These false positives are due to detection records in areas where the pest is very unlikely to exist given its known biology. In the case of gypsy moth, high temperatures limit the establishment and it is well known that

subtropical and tropical areas are not suitable. Yet, CAPS database includes positive locations in Florida. Updated positive detections for EGM is displayed in Figure 4-7.

One approach is to use MaxEnt. MaxEnt uses presence data as a driver. Given the location of the CAPS data points (without correction) it was noted before that such an approach would yield results showing nearly all of the US as suitable. This is clearly unreasonable. A second approach is to set up a training set for MaxEnt based on the corrected data set. That approach is investigated below. A third approach is to use a phenology model to determine suitability. This approach was investigated in Chapter 5 and illustrated in Figure 5-12 (Chapter 5). A straightforward estimation of areas where the pest can develop shows that the majority of the contiguous United States is suitable to complete a generation of gypsy moth. That is also unrealistic. The problem is that extreme temperatures will result in high temperature induced inactivation. High temperature inactivation is not part of most common phenology models that are based on degree day accumulation. Additionally, there is a key problem with conventional phenology (accumulated degree days) in that some species like gypsy moth require a period of chilling conditions to complete diapauses, a necessary part of the life cycle. Absence of those conditions make it impossible to develop.

I extended the phenology model to accommodate chilling conditions, to accommodate extreme (high temperature) conditions and to accommodate population growth based on temperature. The results were shown in Figure 5-12 (Chapter 5). The results are much more consistent with Figure 4-7. There is an extended area of suitability

(extends West beyond the current known distribution) and is likely due to the fact that “slow the spread” efforts have kept this species from attaining its full range.

Consideration of the pest epidemiology suggests that the approach summarized in Figure 5-12 (Chapter 5) is a more appropriate mechanism to manage the uncertainty associated with the interplay between the variability inherent in the data for pest biology and our ability to establish patterns from these data and develop forecast models. The use of MaxEnt with the modified data set in Figure 4-7 is explored below.

### **MaxEnt**

The MaxEnt model was used to estimate the potential distribution of gypsy moth based on the current distribution plus climatic variables. Among the 19 bioclimatic variables, 6 variables were selected for the model based on the model performance, relevance to gypsy moth, and correlation evaluations among climatic variables with historical bioclimatic variables (1990-2005). The variables that had higher than 0.8 Pearson’s correlation coefficient was removed from the model in order to limit collinearity. Those selected 6 variables were 1) mean diurnal range (Bioclim 2), 2) minimum temperature of coldest month (Bioclim 6), 3) mean temperature of wettest quarter (Bioclim 8), 4) precipitation seasonality (Bioclim 15), 5) precipitation of warmest quarter (Bioclim 18), and 6) precipitation of coldest quarter (Bioclim 19). In addition to bioclimatic values, elevation data were included in the MaxEnt analysis. The mean values and associated standard deviation for these selected Bioclim variables and elevation map are shown in Appendix E. The relative influence of the variable contributions to MaxEnt model for historical bioclimatic variables were summarized in Table 4-1. The most significant

environmental contributor was precipitation seasonality (Bio15), accounting for an 85.9% contribution to habitat suitability probability. The permutation importance indicates the contribution for each variables by randomly permuting the values of the variable among the training points and measuring decrease in training AUC; therefore a large decrease implies that the model depends heavily on the particular variable (Phillips 2010).

**Table 4-1. Relative influence of the variable contribution to MaxEnt for historical climatic data**

Variable	Percent contribution	Permutation importance
Mean diurnal range (Bio2)	0.3	3.1
Minimum temperature of coldest month (Bio6)	3.2	5.3
Mean temperature of wettest quarter (Bio8)	0.3	3.8
Precipitation seasonality (Bio15)	85.9	62.5
Precipitation of warmest quarter (Bio18)	5.0	11.6
Precipitation of coldest quarter (Bio19)	4.5	10.0
SRTM Elevation	0.8	3.7

The first averaged selected bioclimatic variables for each time period were used to run MaxEnt. The model performance for MaxEnt runs for all time periods and RCPs were very similar. The test AUC values ranged from 0.864 to 0.867. The precipitation seasonality influenced the probability of presence the most compared to other variables for all time periods for both RCPs. Figure 4-8 shows the probability of gypsy moth presence estimated with MaxEnt model by using MACA GFDL-ESM2G 1990-2005 data. In order to understand the threshold probability that was suitable for gypsy moth growth, the lowest presence threshold (LPT) was estimated. The LPT represents the lowest output value for observed presence record (Pearson et al. 2007, Webber et al. 2011). The LPT was estimated with the MaxEnt results derived from historical bioclimatic data by rejecting the lowest 10%

of possible predicted values (Pearson et al. 2007). The LPT value was 0.461 for gypsy moth.

The MaxEnt model predicted the probability of the presence of gypsy moth based on the 1990-2005 dataset fairly well. The suitable areas were predicted to be within gypsy moth quarantine areas, except for some areas in Indiana, Ohio, Kentucky, Tennessee, and North Carolina. Although those areas are not designated as quarantine areas as of today, there have been detections of gypsy moth from CAPS traps. On the other hand, there were some areas that the MaxEnt results indicated as unsuitable areas that are within current quarantine areas (e.g., northern Maine, Wisconsin, and northeast Minnesota). I note that those areas have limited gypsy moth populations. I concluded that the MaxEnt model results with LPT of 0.461 represent current distribution of gypsy moth well.

The standard deviations derived from running 10 replicates of MaxEnt on historical bioclimatic variables are shown in Figure 4-8. The higher standard deviations were reported in northern Maine, Wisconsin, eastern half of quarantine areas in Minnesota, eastern Missouri, northern Arkansas, Illinois, Indiana, Kentucky, Tennessee, North Carolina, and South Carolina. Although the mean probability of presence was low for gypsy moth growth in Wisconsin, Illinois, Kentucky, and North Carolina, those areas received high standard deviation, meaning those areas should be considered uncertain regarding gypsy moth suitability. Specifically, those areas may be suitable or may not be suitable, and the model does not clearly differentiate them.

The same bioclimatic variable were used with forecasted climatic values in MaxEnt to predict the probability of gypsy moth presence for the future. I forecasted the probability

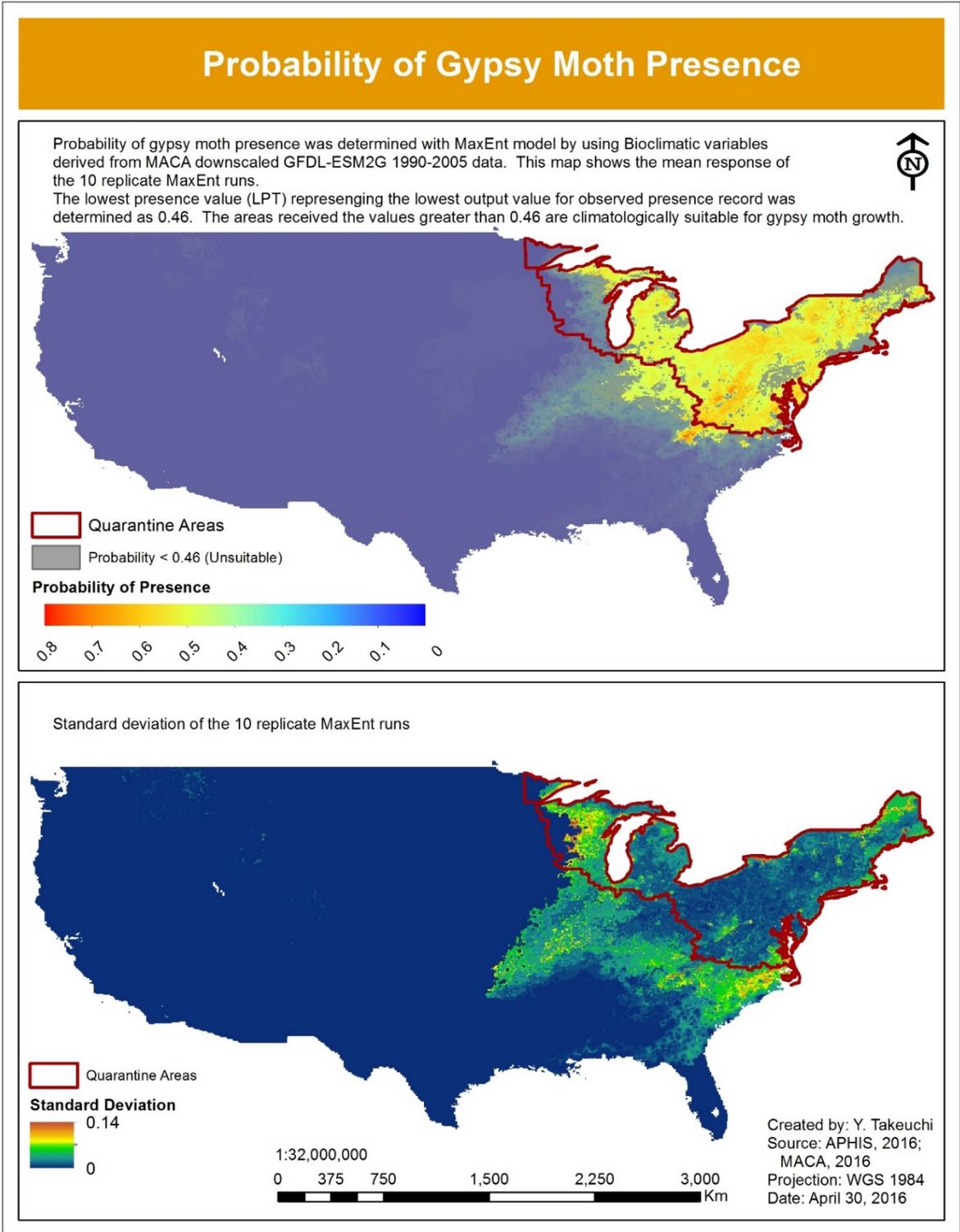
of gypsy moth presence for the 2016-2040, 2041-2065, and 2066-2090 time frames under two climate change scenarios (RCP 4.5 and RCP 8.5). The probabilities of gypsy moth presence based on MACA downscaled GFDL-ESM2G RCP 4.5 model data are shown in Figure 4-9. The results from three time periods (2016-2040, 2041-2065, and 2066-2090) were very similar. All three time periods models indicated that the northeastern United States was suitable for gypsy moth. The standard deviations of 10 replicate MaxEnt runs from three time periods are also displayed in Figure 4-9. The standard deviation derived from RCP 4.5 model data (Figure 4-9) was also similar to the standard deviations derived from historical data (Figure 4-8). The higher standard deviations were observed in a similar areas compared to the ones from historical data; however, in general, the standard deviations observed in RCP 4.5 (except 2016-2040) were higher than the ones from historical data.

The probability of gypsy moth presence estimated by using MACA downscaled GFDL-ESM2G RCP 8.5 model are displayed in Figure 4-10. MaxEnt results from RCP 8.5 models were also similar to the results from RCP 4.5 and historical data. All results from the three time periods indicated that northeastern United States were suitable for this species. The standard deviations associated with those MaxEnt runs also showed similar trends. The standard deviation derived from 2066-2090 models were higher than the models from other time periods for RCP 8.5.

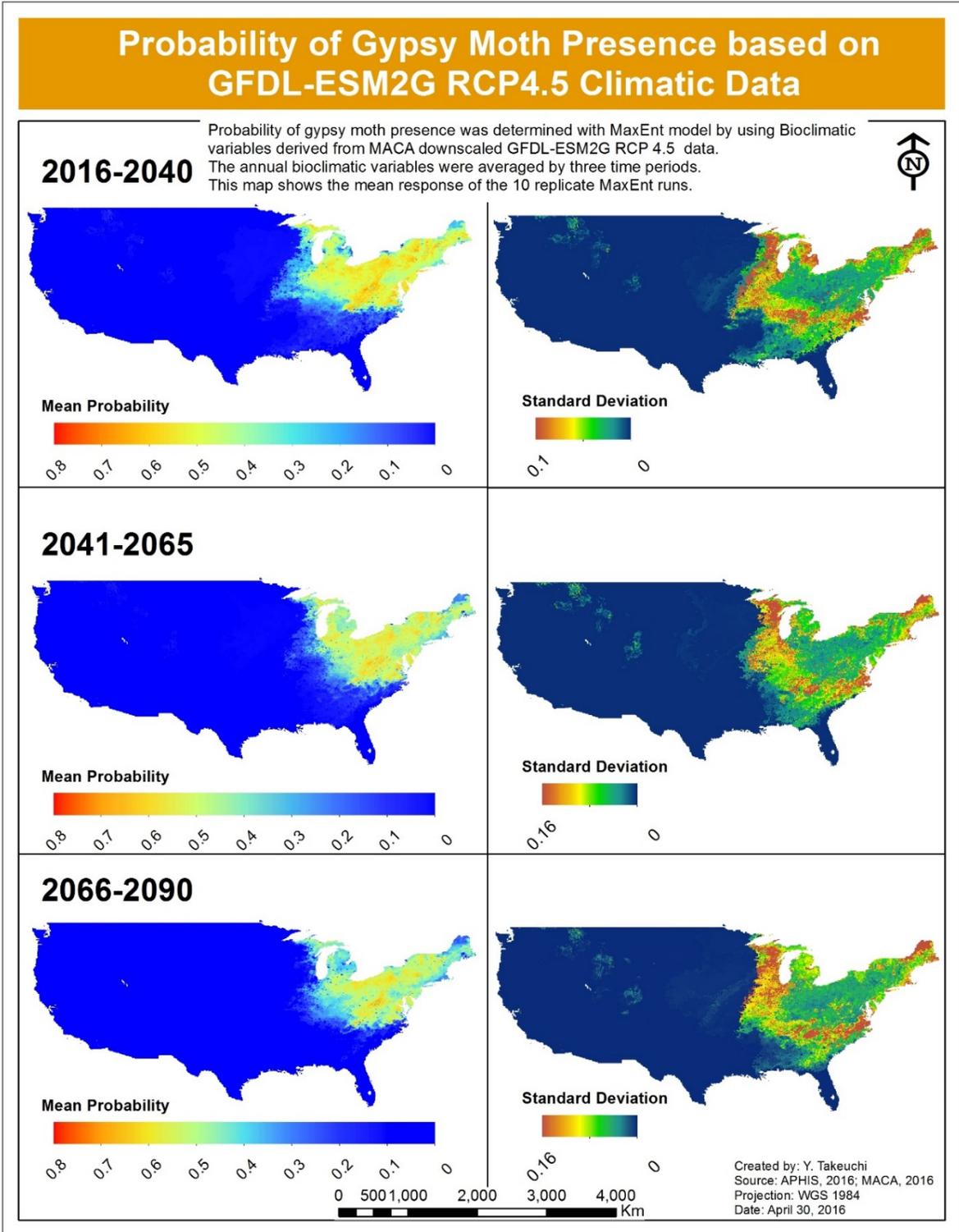
In general, all MaxEnt models run on averaged climate data (e.g., averaged minimum temperature of coldest month from 2016 to 2040) generated relatively small standard deviations (~0.16). I examined the averaged probability (Figure 4-8; Figure 4-9; Figure 4-10), maximum probability, and minimum probability generated from 10 replicates

of each time period (not shown in figure). Since the standard deviations were small, there were only tiny differences between minimum and maximum probability estimated at each location.

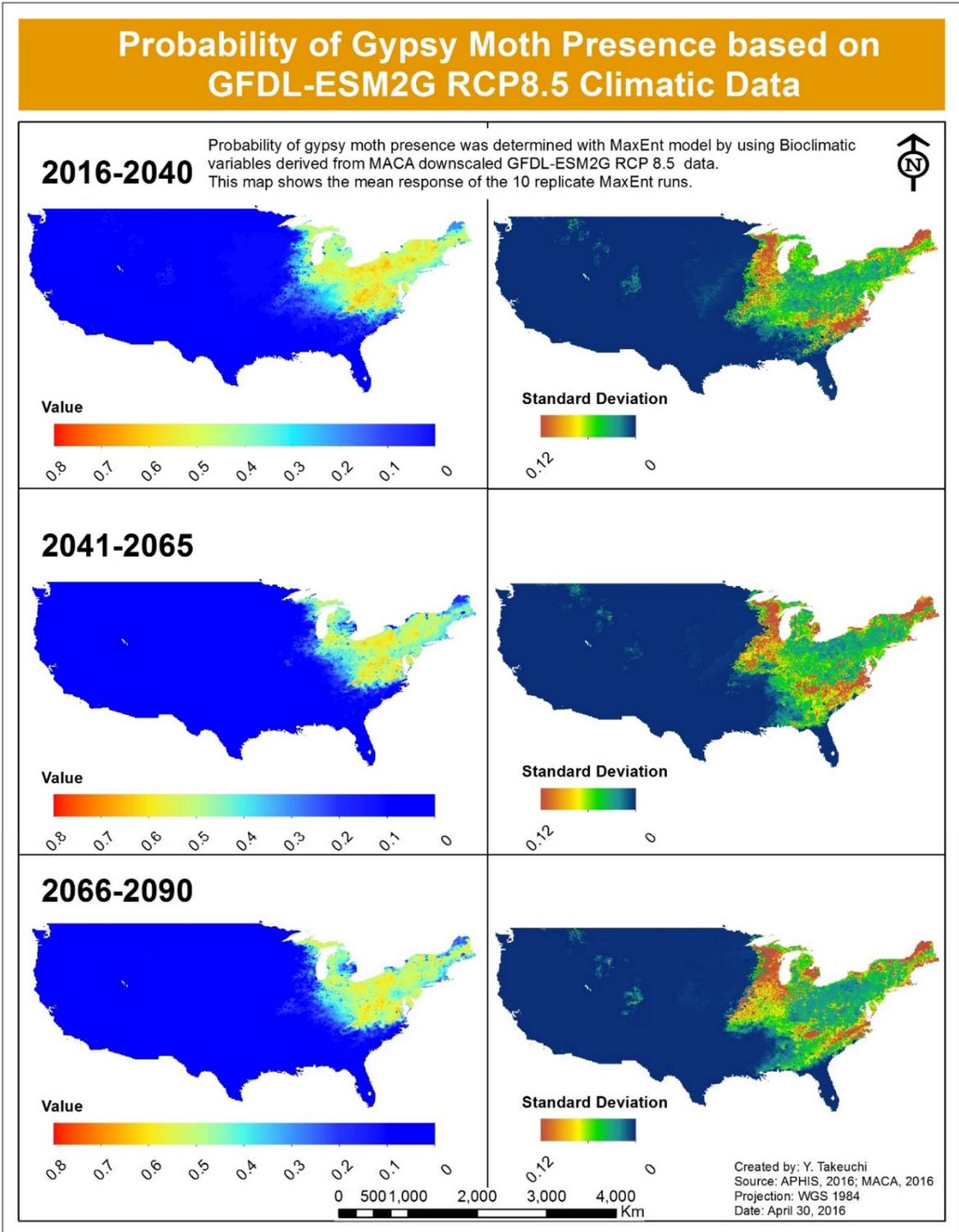
The model results suggested that suitable geographic areas for gypsy moth were likely to shrink under climate change. The models from all time periods for both climate change pathways indicated that the majority of suitable areas from the current condition (based on historical data) would remain suitable. There were some areas that were estimated to become unsuitable or suitable in the future; however, the areas that are forecast to become unsuitable were larger than the areas expected to become suitable. Based on the 2066-2090 climatic data from RCP 8.5 data resulted in the areas shown in blue (e.g., eastern Michigan, eastern Indiana, and eastern New Hampshire) becoming unsuitable, while the areas shown in red (central Maine) likely becoming suitable for gypsy moth (Figure 4-11).



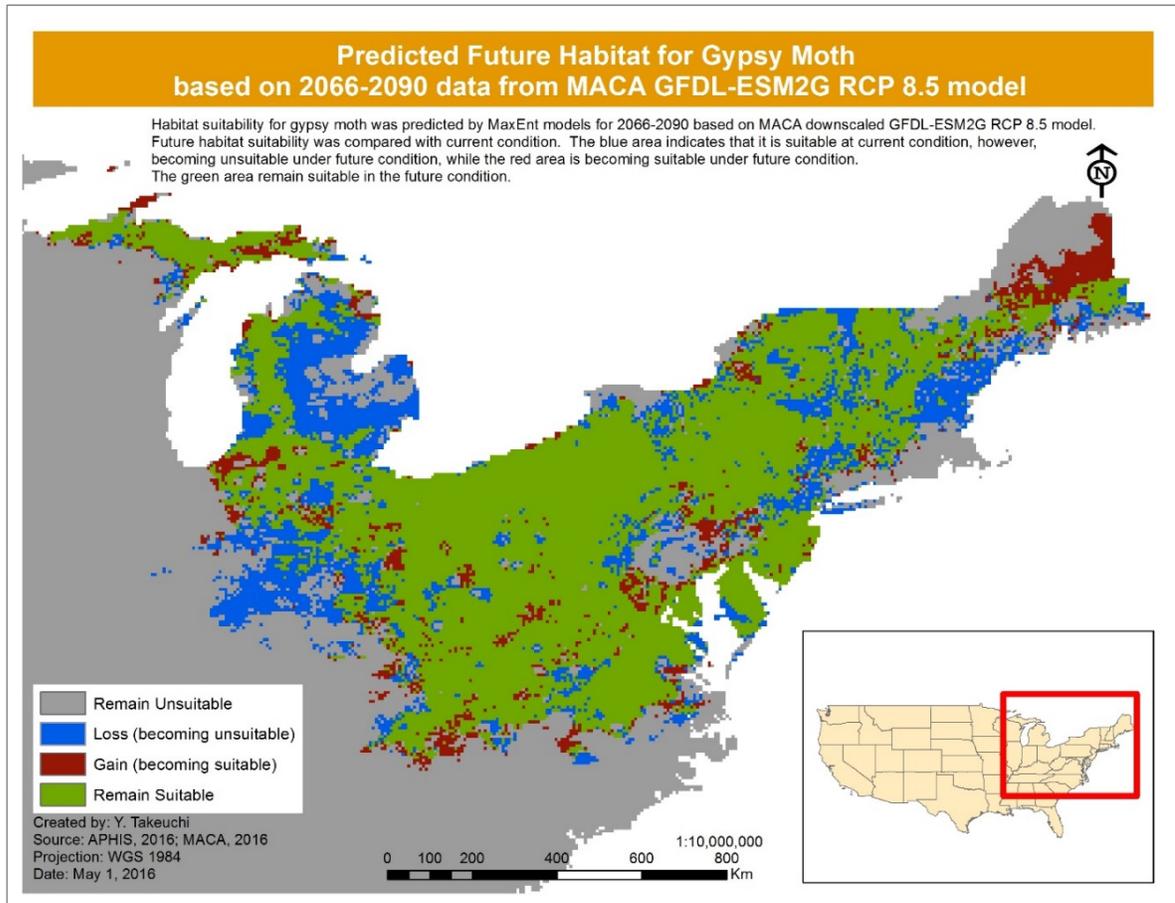
**Figure 4-8. Probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G 1990-2005 data**



**Figure 4-9. Probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP4.5 averaged data by three time periods (2016-2040, 2041-2065, and 2066-2090)**



**Figure 4-10. Probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP8.5 averaged data by three time periods (2016-2040, 2041-2065, and 2066-2090)**



**Figure 4-11. Predicted future habitat for gypsy moth based on 2066-2090 data from MACA downscaled GFDL-ESM2G RCP 8.5**

MaxEnt was set up to use different seeds for each run to select training and testing randomly. However, there was no method to incorporate stochasticity in data drivers in MaxEnt. Assuming bioclimatic variables were normal random variables, 5 different sets of selected bioclimatic variables were randomly generated based on the standard deviations and means. Those 5 sets were then supplied to MaxEnt to predict the probability of gypsy moth presence. I did not generate 5 sets for historical bioclimatic variables because probability density functions evaluation study proved that baseline data from GFDL-ESM2G were well matched with observed climate data (Boyles and Aldridge 2016);

therefore, I assumed that historical data from GFDL-ESM2G were well represented with observed data.

There were a total of 60 MaxEnt runs for each time period and each climate change scenario (10 replicates for 5 sets of randomly generated bioclimatic variables and averaged bioclimatic variables). That is, 60 MaxEnt runs were performed for 2016-2040, 2041-2065, and 2066-2090 climatic variables derived from the RCP 4.5 model and RCP 8.5 model. Model performance (AUC) was evaluated and summarized for each time period and each climate change scenario (Table 4-2).

Models for all time periods and for both climate change scenarios performed better than random (0.5). An average test AUC values from all models were above 0.84 with relatively small standard deviations.

**Table 4-2. Average test AUC for 60 MaxEnt runs and associated standard deviation**

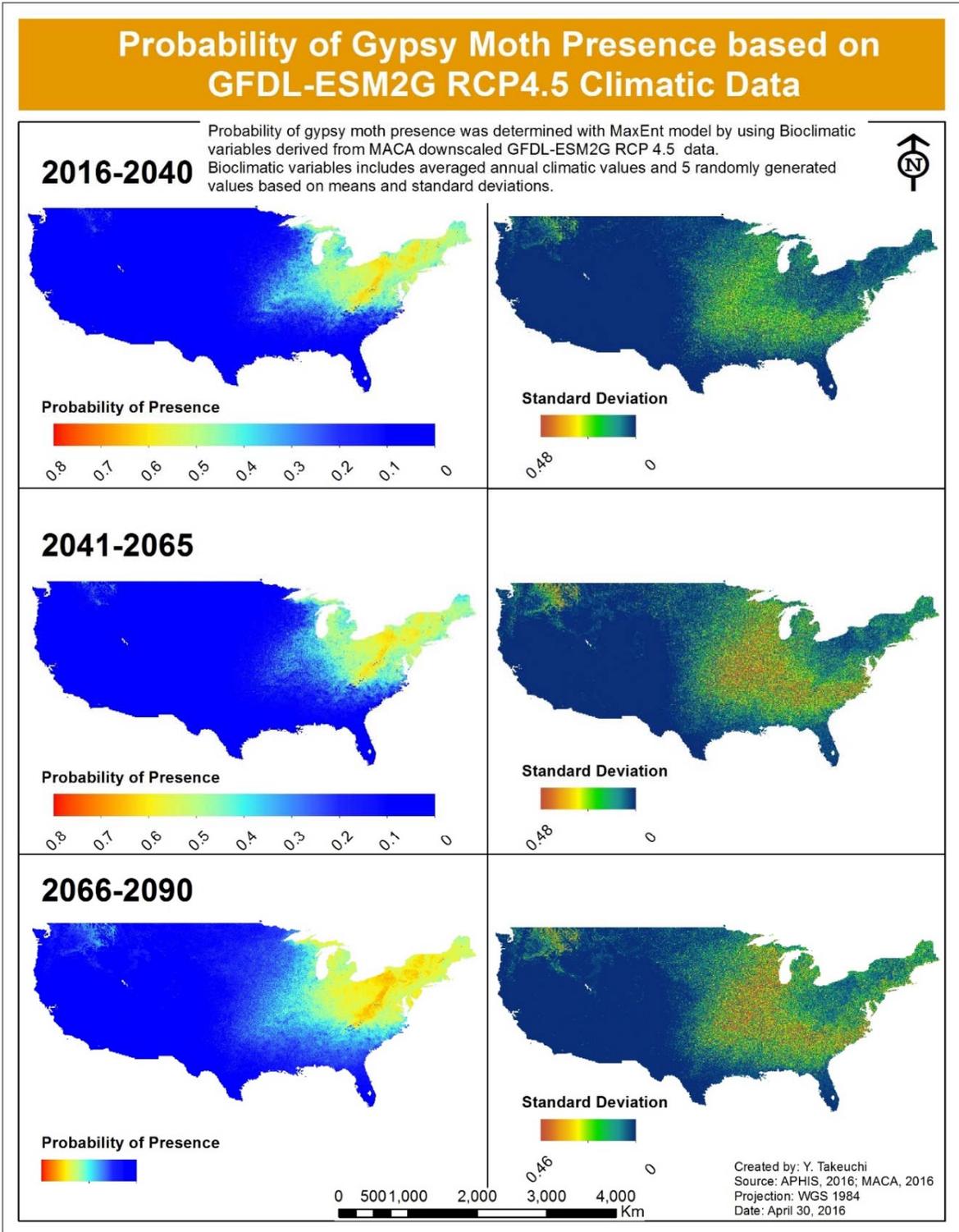
Model/Time Period	Average test AUC	Standard Deviation
RCP 4.5 / 2016 – 2040	0.845	0.0029
RCP 4.5 / 2041 – 2065	0.844	0.0031
RCP 4.5 / 2066 – 2090	0.846	0.0037
RCP 8.5 / 2016 – 2040	0.845	0.0034
RCP 8.5 / 2041 – 2065	0.847	0.0027
RCP 8.5 / 2066 – 2090	0.844	0.0031

The model outputs from using averaged and randomly generated bioclimatic variables suggested that northeastern United States had suitable conditions (Figure 4-12; Figure 4-13) for pest establishment. This trend was observed for both climate change scenarios (RCP 4.5 and 8.5) and all three time periods. It also matched with the MaxEnt results derived by using only mean bioclimatic variables. However, the main difference was that the model run with both mean and randomly generated variables predicted larger areas

to be suitable for gypsy moth. For example, northern Maine was unlikely to be suitable based on the model with mean bioclimatic variables. But when randomly generated bioclimatic variables were included in the model, the probability of the presence estimated by MaxEnt model increased. In addition, the standard deviations associated with MaxEnt outputs (model performance) were much larger when randomly generated bioclimatic variables were used in the MaxEnt compared to the one derived from the MaxEnt model run with mean bioclimatic variables.

There are currently four well-known climate change scenarios (Representative Concentration Pathways: RCPs) (IPCC 2013) (Figure 4-2). The RCP scenarios are developed based on the forecasted greenhouse emission at global scale. The greenhouse emission level influences ambient temperature; however, global surface temperature is expected to stay relatively stable among all 4 RCPs until for the next fifteen years, until 2030 (Figure 4-2). After that and through 2100, global surface temperature is expected to increase approximately 4°C and 2°C under RCP 8.5 and 4.5 scenarios, respectively (IPCC 2013). The outputs derived from 2041-2065 and 2066-2090 were significantly different compared to the outputs derived from historical data. The differences between RCP 4.5 and 8.5 were small or not noticeable (Figure 4-14). The models with RCP 4.5 and 8.5 data resulted similar gypsy moth habitat suitability for 2066-2090. Compared to habitat suitability predicted with historical climatology (1990-2005), the model with RCP8.5 data indicated that larger areas were to become increasingly suitable in 2066-2090 than the model with RCP 4.5 data. For example, although the model with RCP 4.5 data indicated that central Maine was becoming suitable, the model with RCP 8.5 data showed larger areas

in central Maine to be suitable. However, the model with RCP 4.5 data resulted in larger areas remaining suitable compared with currently suitable areas, while the model with RCP 8.5 data indicated fewer areas were to remain suitable. Therefore, the suitable habitat areas for gypsy moth in 2066-2090 predicted with RCP 4.5 and 8.5 data resulted in very similar forecasts.



**Figure 4-12. Probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP4.5 data (mean and randomly generated climatic variables based on mean and standard deviation by time periods)**

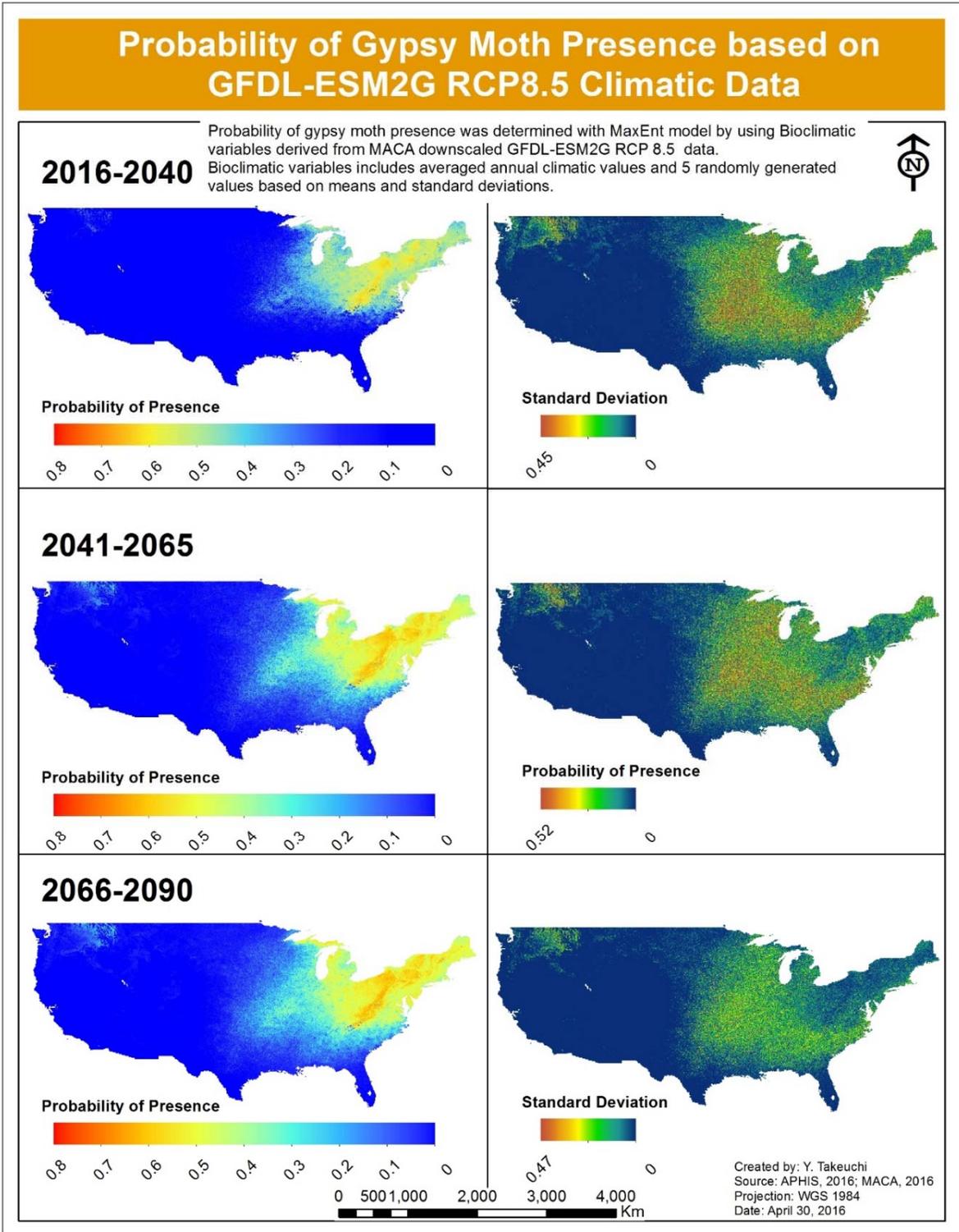
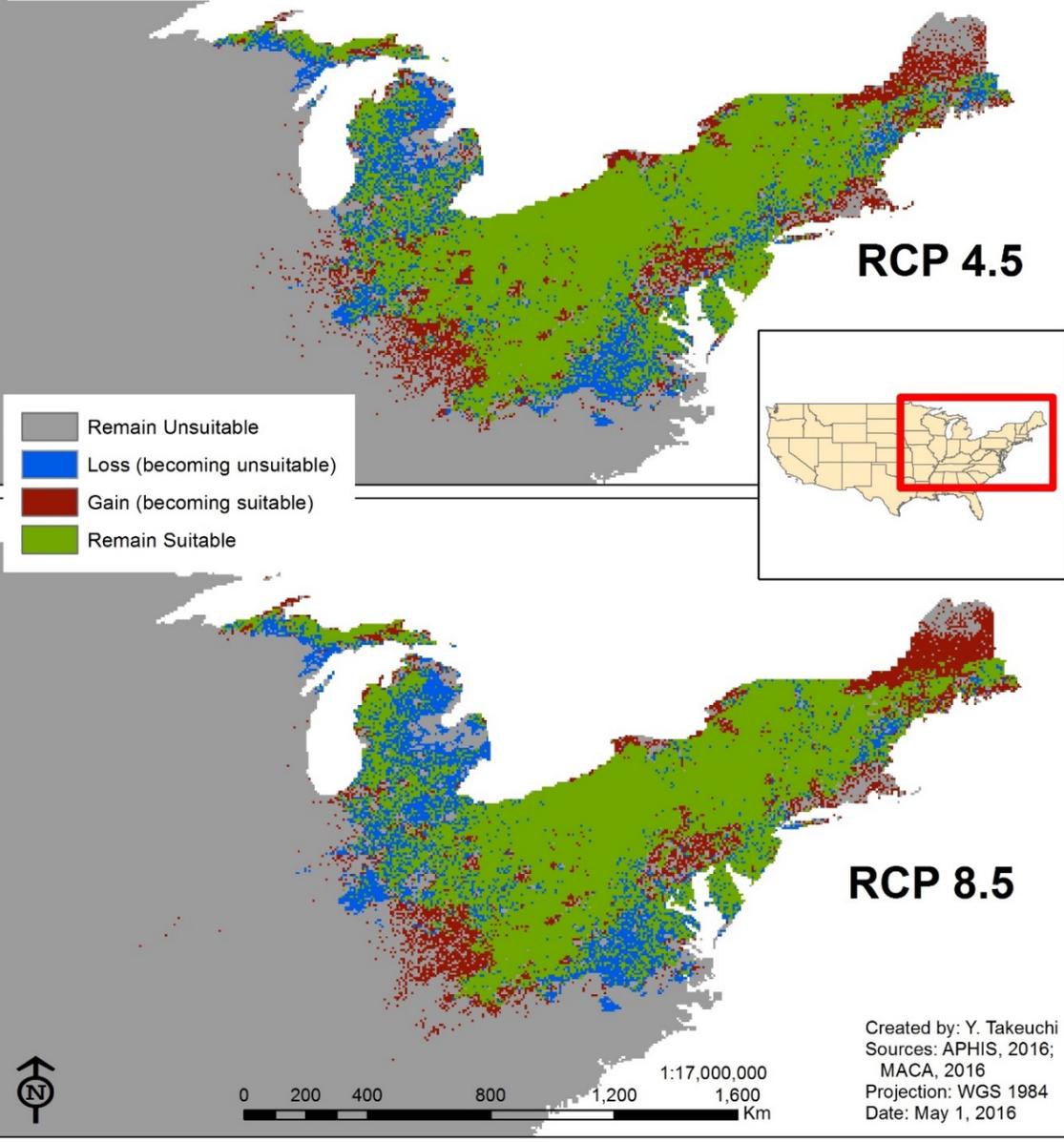


Figure 4-13. Probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP8.5 data (mean and randomly generated climatic variables based on mean and standard deviation by time periods)

**Predicted Future Habitat for Gypsy Moth**  
**based on 2066-2090 data from MACA DGDL-ESM2F RCP 4.5 and 8.5 Models**

Habitat suitability for gypsy moth was predicted by MaxEnt models for 2066-2090 based on MACA downscaled GFDL-ESM2G RCP 4.5 and 8.5 models. Climate data used in this model were mean bioclimatic variables for 2066-2090 and randomly generated climatic values based on the mean and standard deviation from 2066 to 2090. Future habitat suitability was compared with current condition. The blue area indicates that it is suitable at current condition, however, becoming unsuitable under future condition, while the red area is becoming suitable under future condition. The green area remain suitable in the future condition.



**Figure 4-14. Predicted future habitat for gypsy moth for 2066-2090 data estimated with mean and randomly generated climate data from MACA downscaled GFDL-ESM2G RCP 4.5 and RCP 8.5**

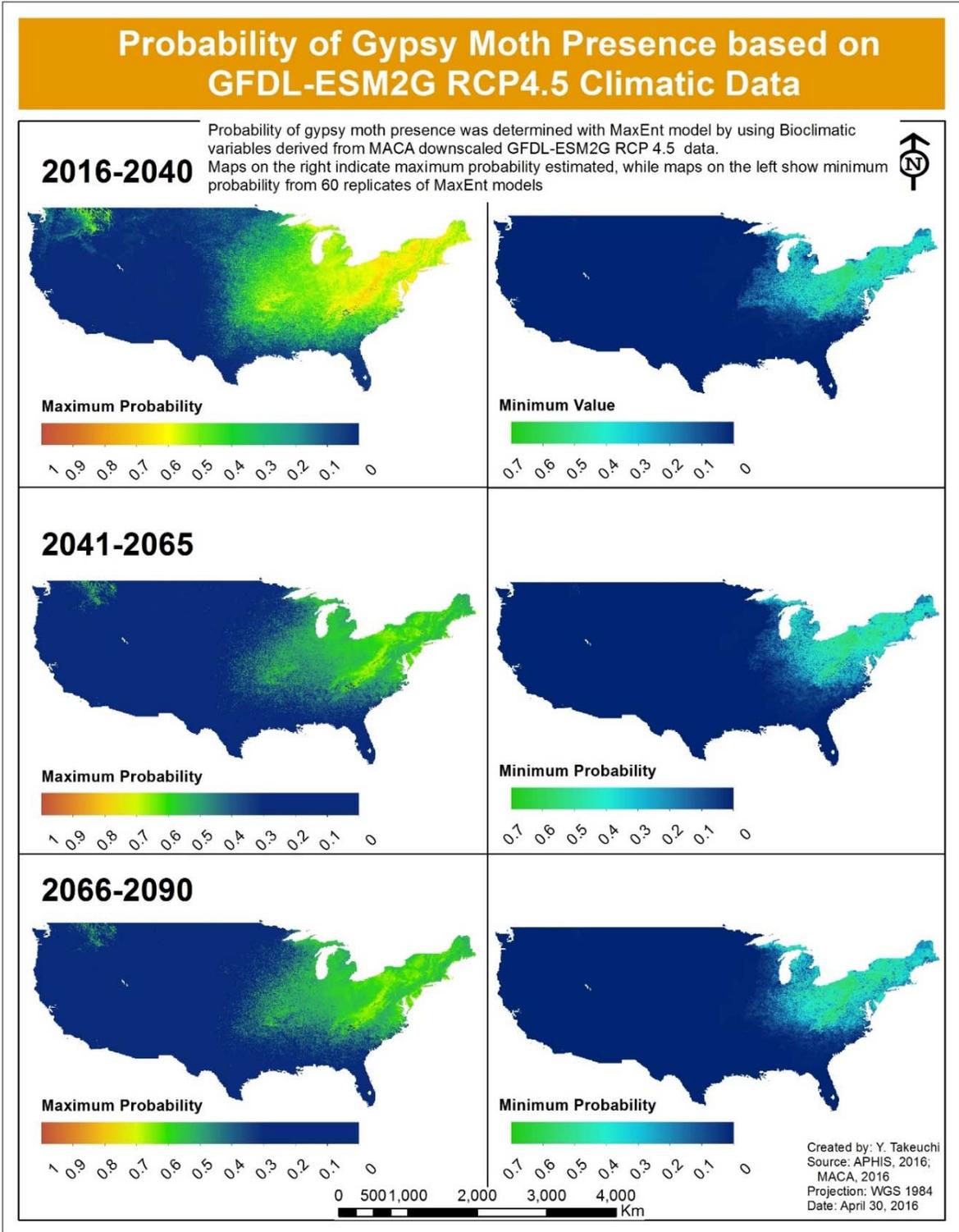
Since the models that were supplied with mean and randomly generated climatic values resulted in high standard deviations, the model results were less reliable (i.e., lower confidence) than the models supplied with only mean climatic values. This is expected since running the MaxEnt model with mean climatic values would not incorporate the annual climate variability. Climate conditions vary greatly each year as shown in Figure 4-1. Simulating climate data in MaxEnt captures the climate variability, resulting in higher standard deviation (higher uncertainty) associated with MaxEnt results.

To illustrate uncertainty distinctly, I displayed maximum and minimum values resulted from 60 replicates of MaxEnt instead of confidence interval around the mean for each time period and each climate scenario (Figure 4-15). There were significant differences between minimum and maximum probabilities. The MaxEnt models with RCP 4.5 data revealed that eastern United States, except Florida, Gulf coastal area, and southern Texas, could be suitable for gypsy moth habitat. In addition, northeastern Washington, northern Idaho, and northwestern Montana also resulted in high enough probability to conclude this would support gypsy moth growth. Those additional areas where gypsy moth habitat became suitable were uncertain because those areas could be suitable depending on climatic conditions. Another important result was that some areas (e.g., northeastern Washington, northern Idaho, and northwestern Montana) had a very low probability with low standard deviation when the models were run with mean climatic values. Those areas were only observed as high probability when climatic conditions were simulated based on the mean and standard deviation associated with the time periods. Similar results were observed from the MaxEnt models with RCP 8.5 data (Figure 4-16), indicating that the

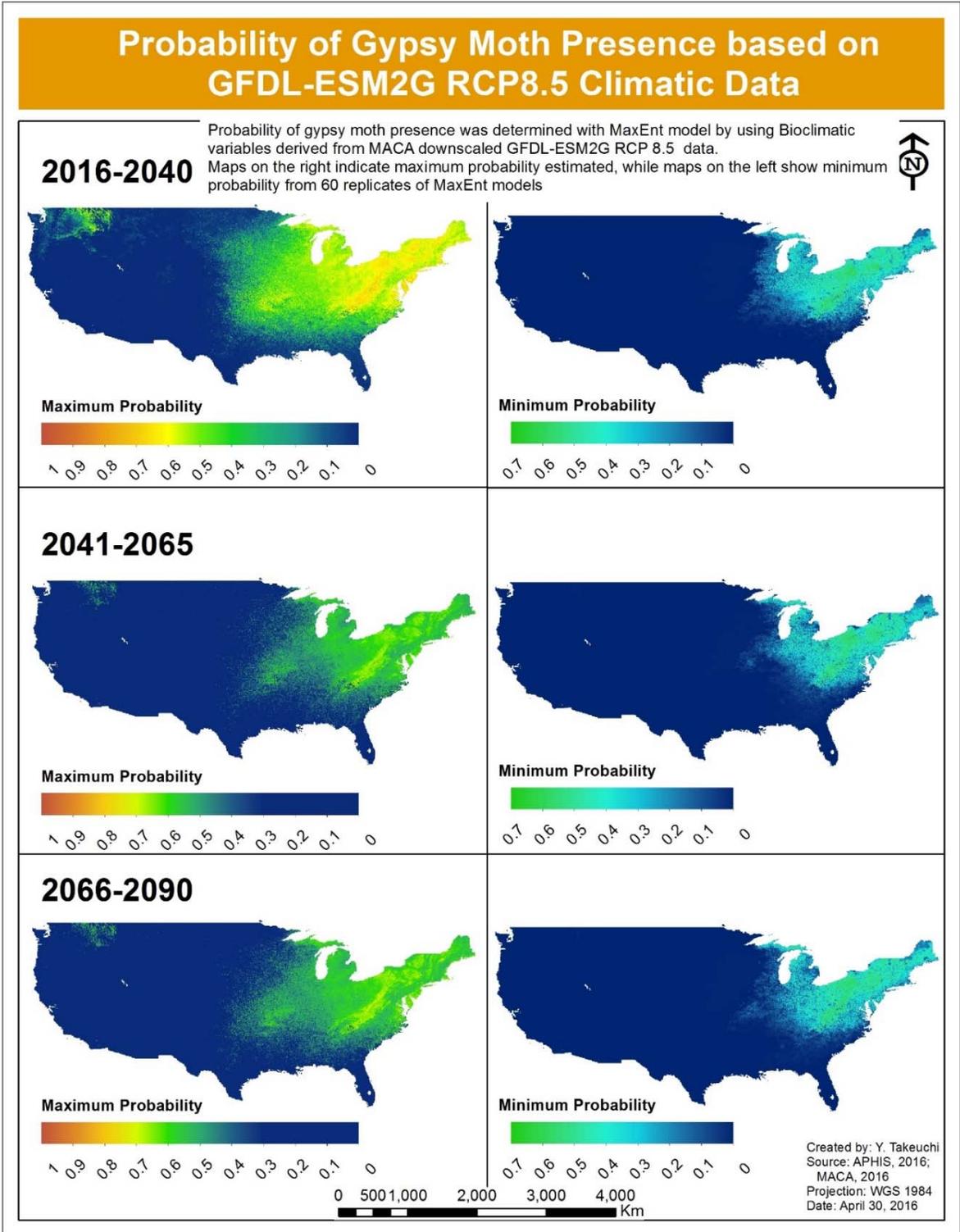
climate change prediction differences between the RCP 4.5 and 8.5 did not affect the suitability for gypsy moth in the contiguous United States.

The minimum probability (maps on the left in Figure 4-15 and Figure 4-16) indicates the minimum values produced by MaxEnt replicates. I can conclude with high certainty that the areas receiving greater than 0.461 minimum probability are suitable for gypsy moth.

The northeastern United States (light blue to green areas in maps on the left in Figure 4-15 and Figure 4-16) continues to be a suitable habitat for gypsy moth at least for the next 75 years.



**Figure 4-15. Maximum and minimum probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP4.5 data (mean and randomly generated climatic variables based on mean and standard deviation by time periods)**



**Figure 4-16. Maximum and minimum probability of gypsy moth presence estimated with MaxEnt by using MACA downscaled GFDL-ESM2G RCP8.5 data (mean and randomly generated climatic variables based on mean and standard deviation by time periods)**

## **Evaluating Uncertainty for SAFARIS Forecasts**

In this section I examine uncertainty associated with forecasts. The overall uncertainty is examined and then the sources of uncertainty are separated and examined. An important aspect of communicating the importance of the uncertainty to decision makers is shown by examining the geographic-dependent aspects of uncertainty. Importantly, this sections aims to establish not only methods to explore uncertainty but also methods to clearly communicate why this is important to decision makers.

The first evaluation of uncertainty explores the uncertainty associated with the input drivers (focusing on climate and geospatial characteristics) and model goodness of fit. The second section examines current practice. This current practice (suitable locations shown on a map) is used as a baseline to explore what is added by adding uncertainty associated with goodness of fit indicators. The third section examines uncertainty associated with different GCM models.

The MaxEnt modeling study using the SAFARIS framework indicated that the majority of the eastern United States could be suitable habitat for gypsy moth. Examining maximum probability, minimum probability, and standard deviation using MaxEnt suggested that uncertainty information needed to be included to gain critical information for decision making.

In order to incorporate uncertainty into the decision making process, the mean-variance frontier (MVF) concept was applied to MaxEnt outputs. Instead of analyzing uncertainty at each cell level, uncertainty was evaluated at the county level. APHIS often makes decisions at the county level; therefore, understanding uncertainty and prioritizing

locations by county was more practical and applicable than at cell level. The mean probability and standard deviation of gridded cells within each county were summarized. I used maximum probability and standard deviation observed within each county for uncertainty evaluation to understand maximum potential likelihood. I note that this process could be easily repeated at the cell level if a different application is being pursued.

### **Evaluation of uncertainty associated with the input drivers and model goodness of fit**

The mean probability of gypsy moth presence was plotted against the standard deviation of the probability derived from GDFL-ESM2G RCP 8.5 2066-2090 data for each county in contiguous United States (Figure 4-17). In this study, the areas with higher standard deviation were given higher priority because I want to gain more information where the uncertainty is high and also want to inform decision makers that the areas with high uncertainty could result far away from the expected mean value. Therefore, a priority ranking was assigned to each point from the upper right outer boundary.

The mean-standard deviation ratio was estimated to incorporate uncertainty into mean outputs from MaxEnt. As for the forecast output derived from GDFL-ESM2G RCP 8.5 2066-2090, the equation used to adjust mean outputs to incorporate uncertainty (standard deviation) at the county level was:

Mean Standard Deviation Ratio

$$= \text{Max Probability Observed} + 3 * \text{Max Standard Deviation Observed}$$

The mean-standard deviation ratio was then converted to ranks by sorting U.S. counties based on the mean-standard deviation ratio. The county that received the highest mean-standard deviation ratio was assigned 1 followed by assigning 2 for the second highest

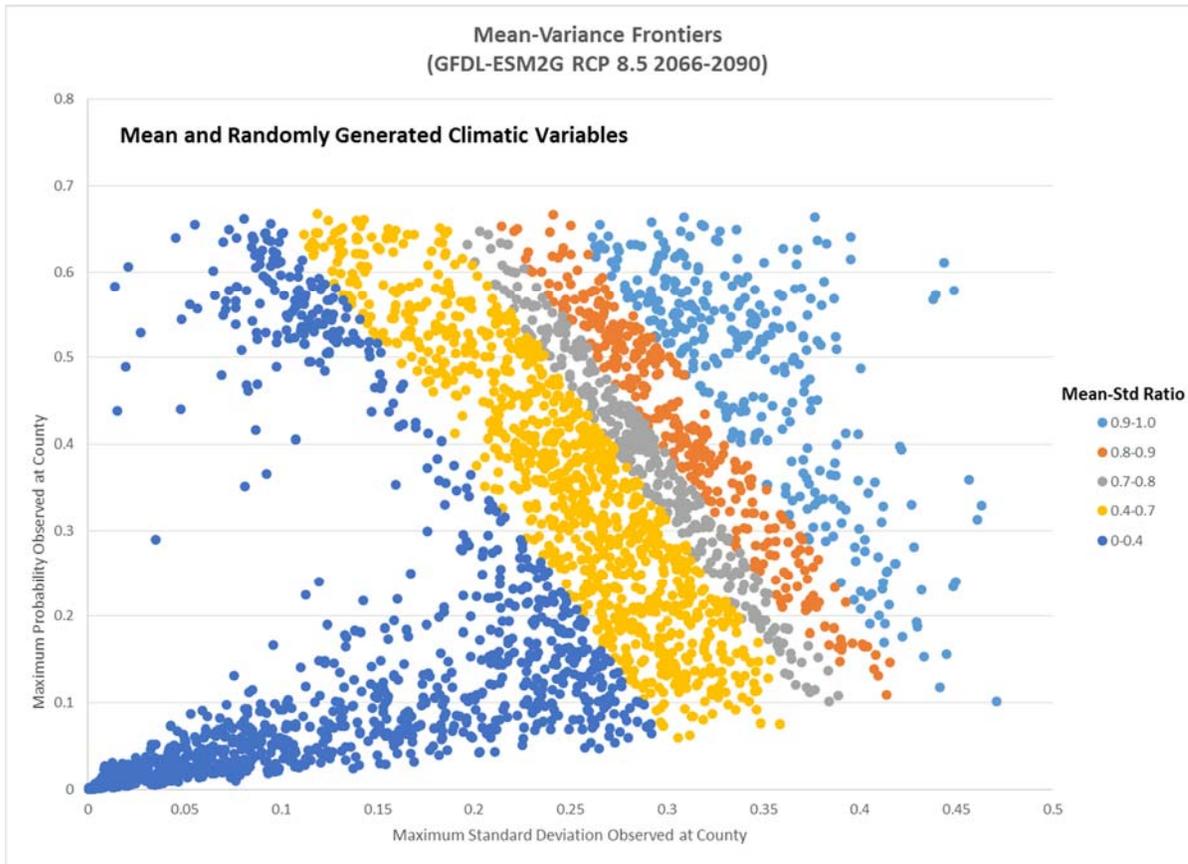
mean-standard deviation ratio. Then, the ranks were converted to relative rankings that were ranged from 0 to 1. I used the equation provided by Yemshanov et al. (2015). The relative ranking of jth location can be expressed as:

$$r_j = 1 - \{(l_j - 1)/\max[l_j]\}$$

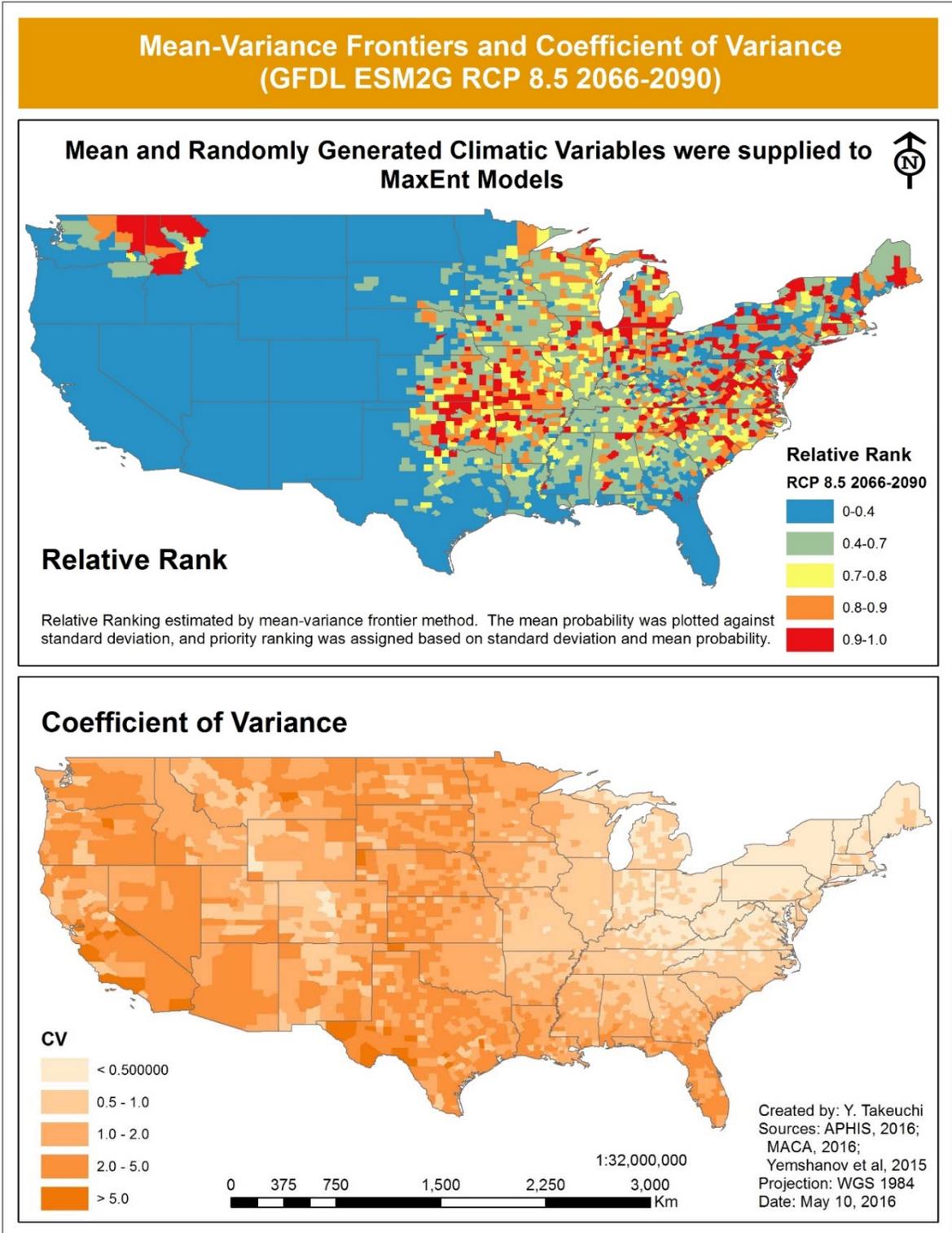
where  $\max[l_j]$  is a maximum rank value in the classification,  $l_j$  is priority ranking assigned by plotting mean probability and standard deviation. I evaluated 3,094 counties; therefore  $\max[l_j] = 3,094$  in this study.

A county map indicating the relative ranking is shown in Figure 4-18. The map above in Figure 4-18 displays relative ranking estimated by MVF. The MVF evaluation resulted in areas of Arkansas, Missouri, Kansas, Oklahoma, southern Appalachian mountain regions, southern Michigan, Indiana and Ohio having higher priority in addition to some counties in Washington, Idaho, and Montana. The map below indicates that the coefficient of variation ( $CV = \sigma / \mu$ ). Western states had high relative variability; however, those counties still did not rank as high priority because the mean probabilities were very low. However, the CV has one major disadvantage. The CV becomes very high and is sensitive when the predicted probability is close to zero. Since many of Western United States resulted in very small mean probability, the CV results showed extremely high values.

However, the CV map also provides additional information and could be used to inform the number of traps required for gypsy moth as an illustration of its application. If we were to detect gypsy moth in high VC areas, those areas must have a higher number of traps placed in order to confirm gypsy moth presence or absence.



**Figure 4-17. Scatter plot between probability and standard deviation at county level derived from MaxEnt model with GFDL ESM2G RCP 8.5 2066-2090 data (mean and randomly generated climatic variables)**

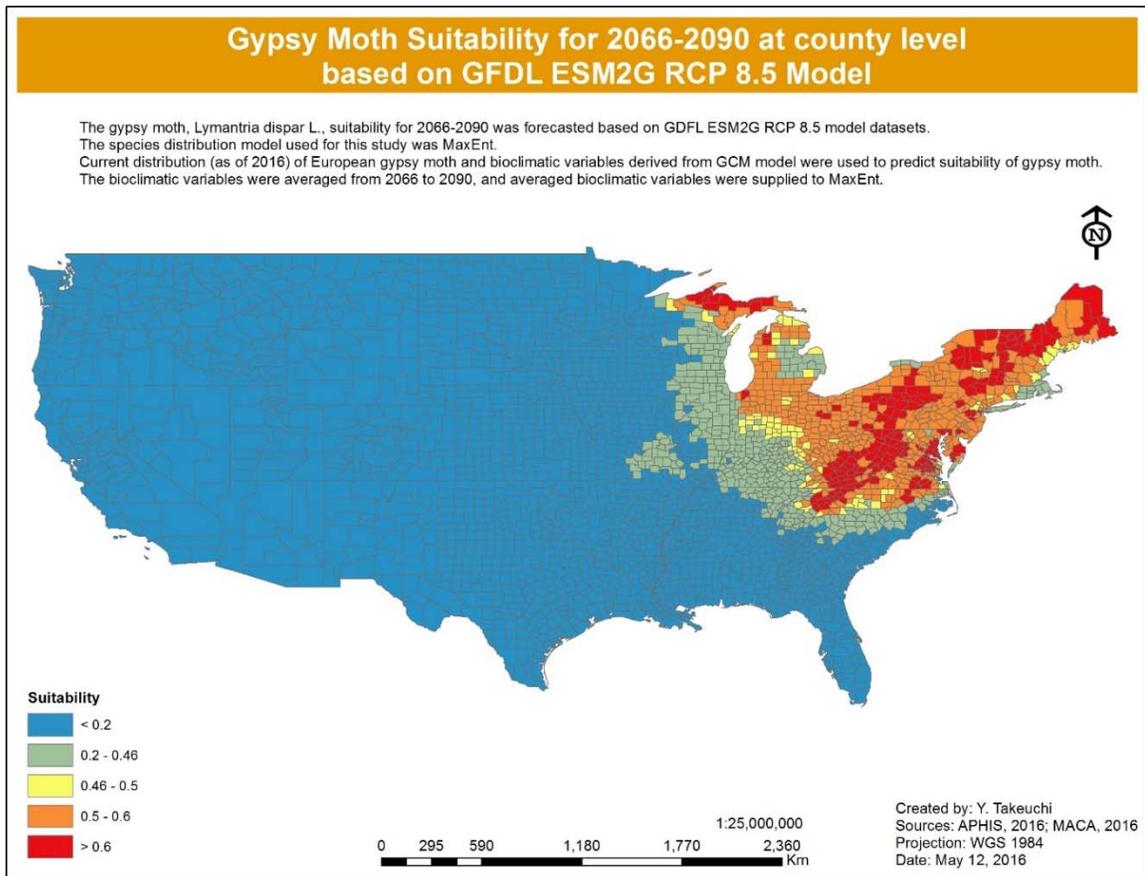


**Figure 4-18. Relative rank estimated by mean-variance frontier approach and coefficient of variation for GFDL ESM2G RCP 8.5 2066-2090 data (mean and randomly generated climatic variables)**

### **Evaluation of uncertainty associated with model goodness of fit**

Historically, APHIS has not considered uncertainty associated with pest forecasts. For example, APHIS usually presents a county level suitability map such as for gypsy moth in 2066-2090 shown in Figure 4-19. This map was developed based on the MaxEnt model with mean forecasted climatic values for 2066-2090 from the RCP 8.5 model and selected for the maximum probability predicted within the county. The areas indicated with yellow, orange, and red were above LTP (0.461); therefore, the interpretation is that the northeastern United States is highly suitable as gypsy moth habitat.

The risk of making decisions based on the prediction in Figure 4-19 is that we do consider uncertainty associated with how climate will change in the future. Considering the areas with high uncertainty and high or medium likelihood of establishment should be part of the decision making processes for better safeguarding protocols for unwanted pests.



**Figure 4-19. Suitability of gypsy moth for 2066-2090 based on GFDL ESM2G RCP 8.5 model**

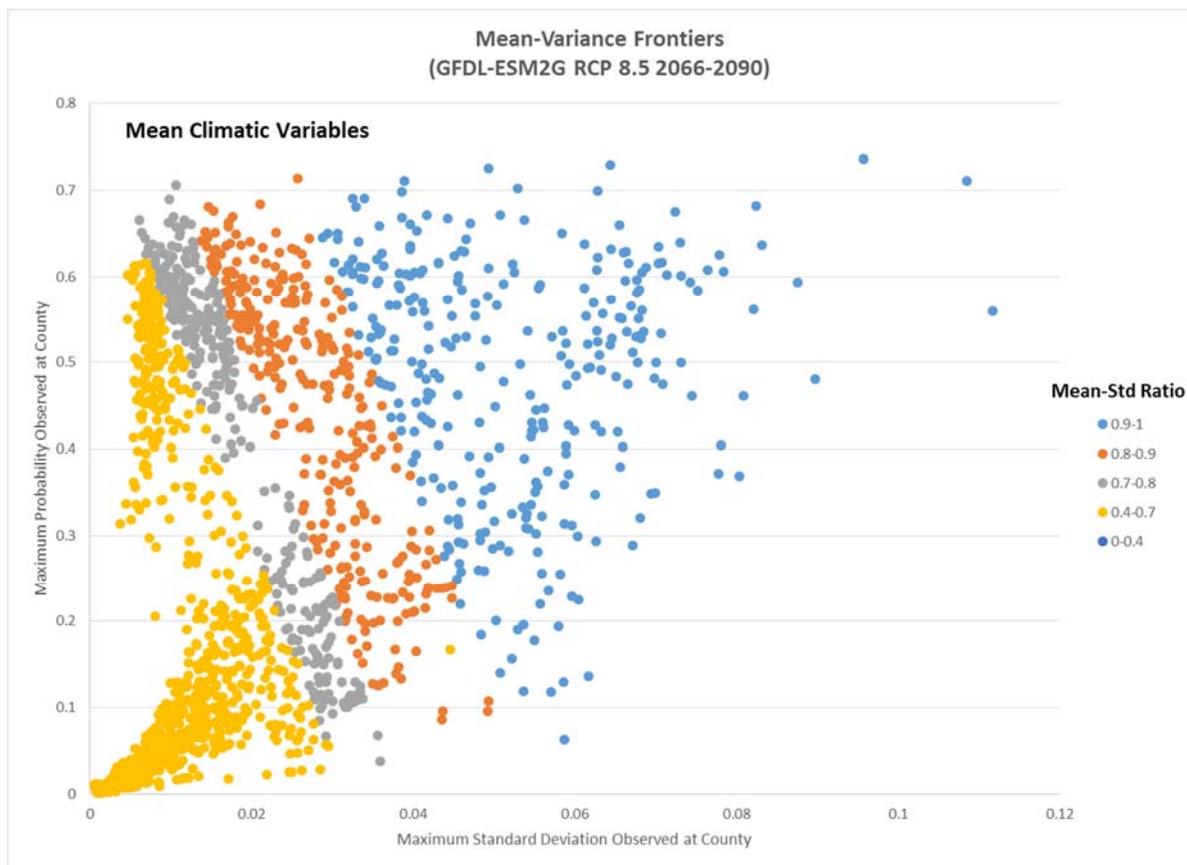
The relative rank at the county level was estimated for the suitability output for 2066-2090 derived from only the mean climate values of the RCP 8.5 datasets. The process I used is to add an uncertainty analysis layer onto the APHIS current pest prediction protocols to explore if there are meaningful differences. Since MaxEnt was run on the mean climatic values, this output did not inherit annual climate variability. The standard deviations represented the model performance, meaning how confident the model outputs were based on the goodness of fit for 10 replicates at each location (cell).

Equation for adjusting mean probability to incorporate standard deviation for the MaxEnt model run with mean climatic values was expressed as:

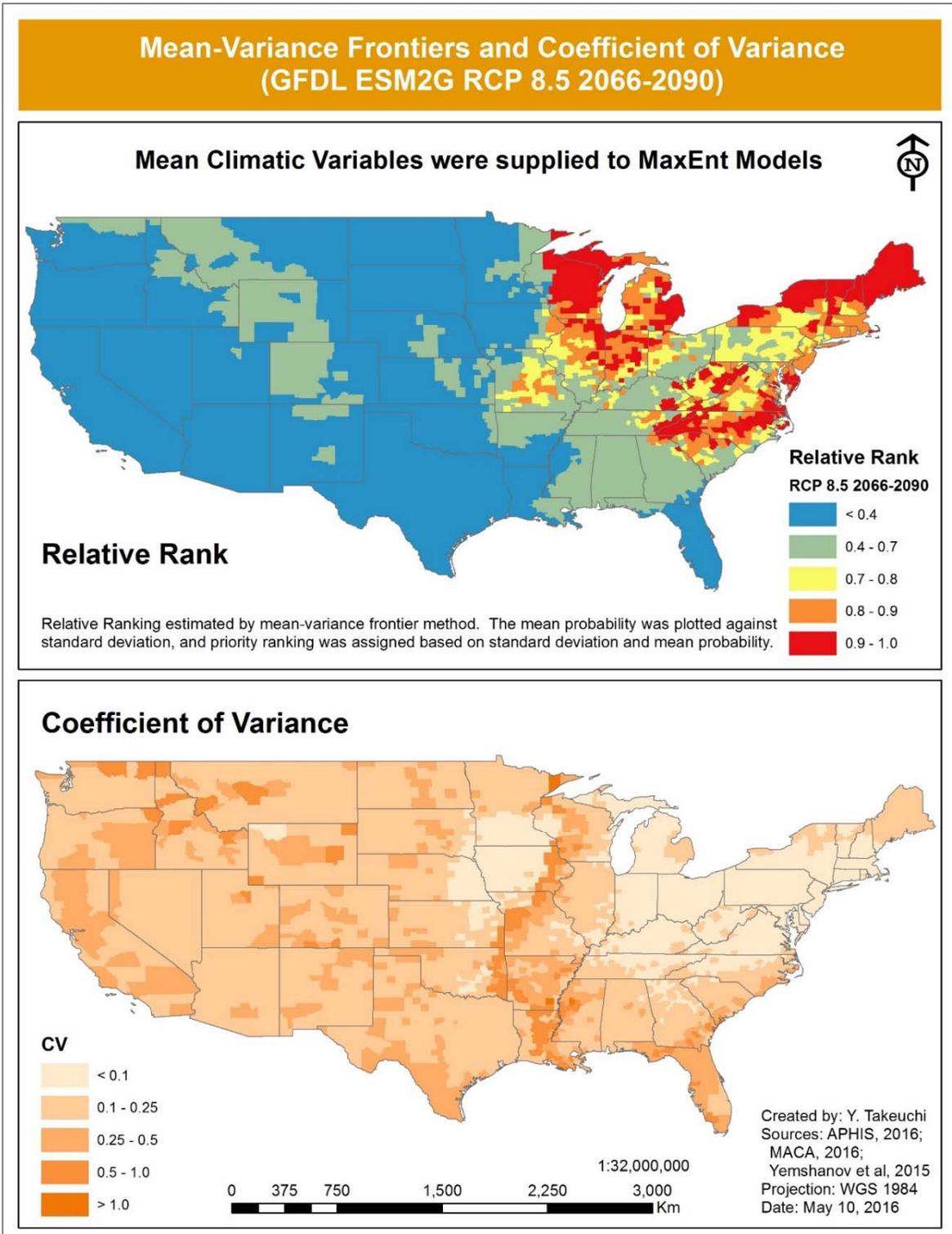
## Mean Standard Deviation Ratio

$$= \text{Max Probability Observed} + 25 * \text{Max Standard Deviation Observed}$$

The standard deviations generated from the MaxEnt model with mean climatic values were much smaller because climatic variability uncertainty was not included in this model (Figure 4-20). Therefore, a different adjustment was needed to calculate the mean-standard deviation ratio. The mean-standard deviation ratio was examined for the model outputs with only mean climatic variables (Figure 4-20). Then, each county was ranked based on mean-standard deviation ratio and displayed in the map below (Figure 4-21).



**Figure 4-20. Scatter plot between probability and standard deviation at county level derived from MaxEnt model with GFDL ESM2G RCP 8.5 2066-2090 data (mean climatic variables)**



**Figure 4-21. Relative rank estimated by mean-variance frontier approach and coefficient of variation for GFDL ESM2G RCP 8.5 2066-2090 data (mean climatic variables)**

The relative rank (Figure 4-21) for 2066-2990 from MaxEnt outputs with only mean climatic variables indicated that higher priority should be placed in northern New England region, North Carolina to northern South Carolina, and eastern Midwest region. The coefficient of variation calculated with maximum probability and maximum standard deviation observed at each county indicated that western United States were more uncertain about the suitability predicted by MaxEnt. However, comparing the CV values between the MaxEnt supplied with mean and randomly generated climate variables (Figure 4-18) and with only mean climate variables (Figure 4-21), shows that the uncertainty was much greater for the models with mean and randomly generated climatic values. This means that the main sources of uncertainty for gypsy moth suitability model for the future was due to climate uncertainty rather than fitness of the model based on the relationship between biology and climatic conditions through distribution of the gypsy moth.

#### **Evaluation of uncertainty associated with GCM RCP models**

To evaluate uncertainty associated with climate projections, I compared relative ranks for 2066-2090 derived from mean and randomly generated climate values from GFDL ESM2G RCP 4.5 and 8.5 models. The relative ranks for the RCP 8.5 model are displayed in Figure 4-18. The relative ranks for the RCP 4.5 model were estimated based on the same procedures used with the RCP 8.5.

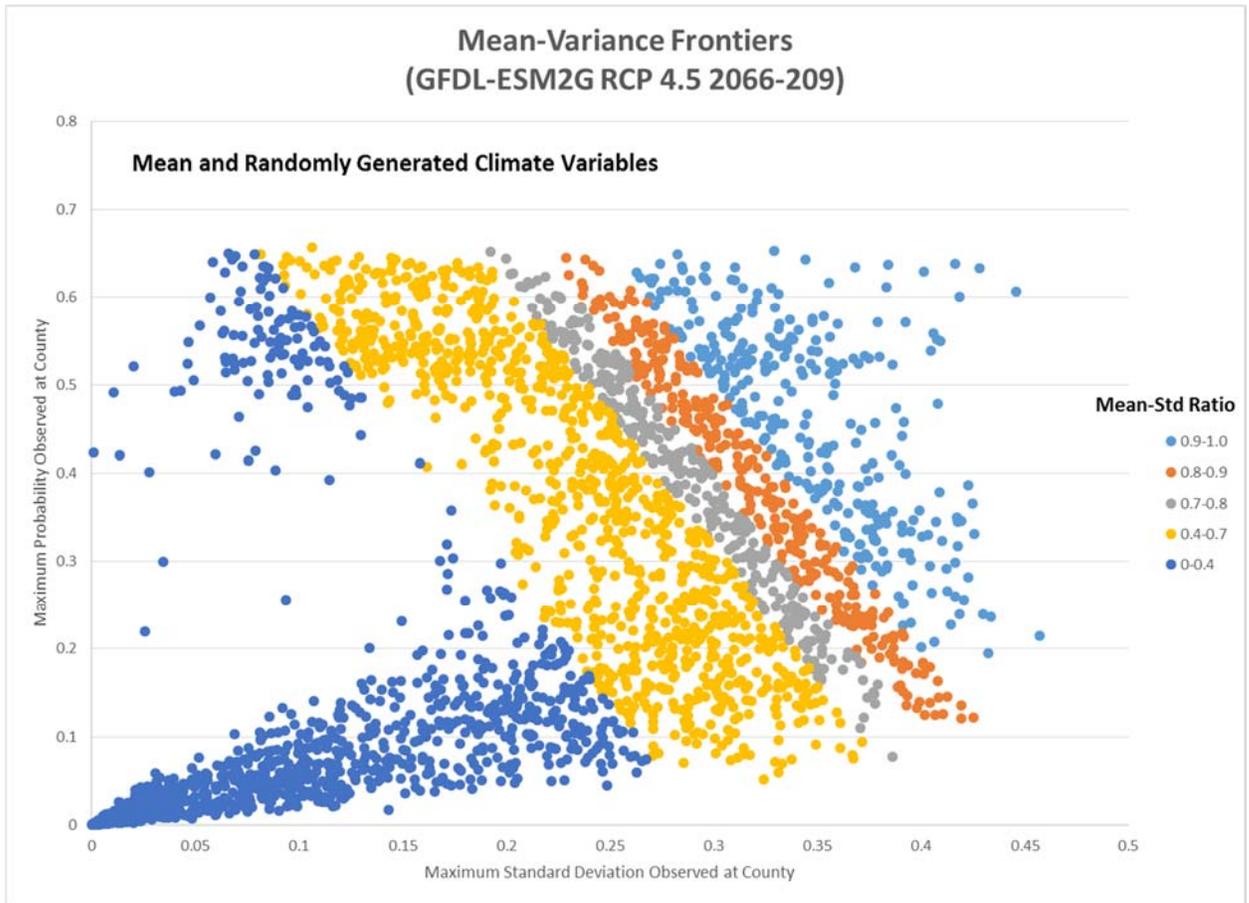
The adjusting equation to incorporate standard deviation into prediction for the MaxEnt model run with the RCP 4.5 model (both mean and randomly generated climate variables) was expressed as:

Mean Standard Deviation Ratio

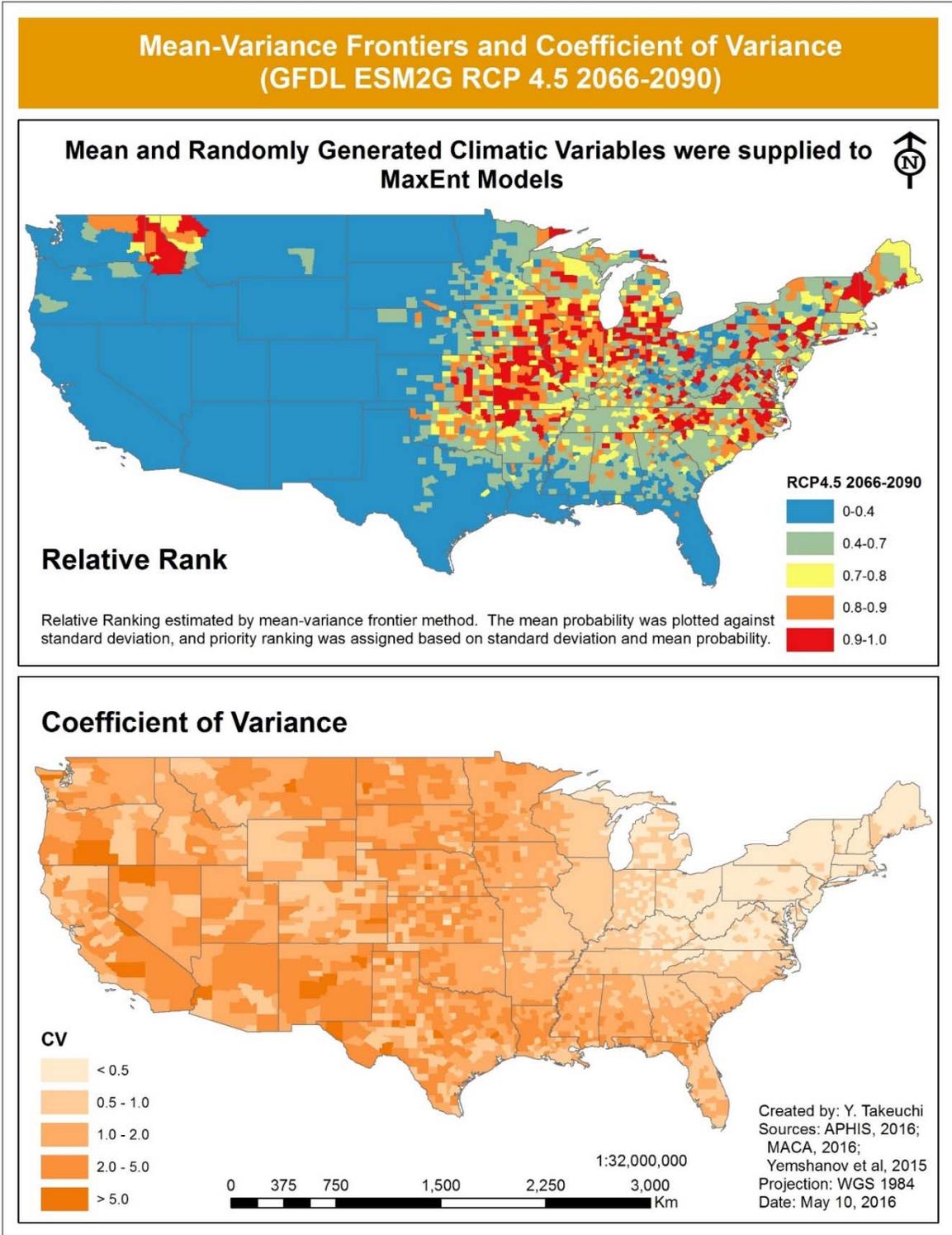
$$= \text{Max Probability Observed} + 3 * \text{Max Standard Deviation Observed}$$

The maximum probability and maximum standard deviation observed at the county level were plotted as y-axis and x-axis, respectively (Figure 4-22). The model results from the RCP 4.5 had similar outputs derived from the RCP 8.5, I was able to use the same adjusting equation from the previous example.

The relative rank at the county level for 2066-2990 MaxEnt prediction with mean and randomly generated climatic variables showed that Indiana, Illinois, Iowa, Missouri, Oklahoma, eastern Kansas, Arkansas, and North Carolina should be a higher priority for gypsy moth management (Figure 4-23). This priority ranking was very similar to the one estimated from the model supplied with mean and randomly generated climate data from the RCP 8.5 model. Although the assumption for the greenhouse gas emission level differs between the RCP 4.5 and 8.5 models, the gypsy moth suitability and level of uncertainty associated with the prediction for 2066-2090 did not differ in terms of prioritizing counties for pest management within the contiguous United States.



**Figure 4-22. Scatter plot between probability and standard deviation at county level derived from MaxEnt model with GFDL ESM2G RCP 4.5 2066-2090 data (mean and randomly generated climatic variables)**



**Figure 4-23. Relative rank estimated by mean-variance frontier approach and coefficient of variation for GFDL ESM2G RCP 4.5 2066-2090 data (mean and randomly generated climatic variables)**

In summary, three levels of uncertainty were evaluated in this study. I examined the uncertainty associated with model fitness and climate variability, 2) uncertainty associated with model fitness, and 3) uncertainty associated with GCM RCP models. Through these evaluations, I also wanted to point out how important it is to incorporate uncertainty information into the pest predictions.

Incorporating uncertainty provided insights in how to prioritize areas for pest management. This study focused on gaining additional information to reduce uncertainty; therefore, the higher weight was placed on the areas that had high suitability with high uncertainty. Conducting a study or surveying in the areas received high relative rankings (e.g., Illinois, Indiana, and Missouri) would likely to gain more information about gypsy moth that are useful for gypsy moth prediction for the future.

Another key element with observing uncertainty associated with pest forecast was to understand variations around the mean. I examined the coefficient of variance to understand the areas with high uncertainty. All three uncertainty evaluations resulted in outcomes where the western United States had a high CV, indicating that the uncertainty in the western United States was quite high. Gypsy moth is currently surveyed nationwide (including Hawaii and Alaska). If the CAPS program is surveying pests to confirm absence or presence of the pests, the counties that have high uncertainty, especially for the counties with low probability and high standard deviation, would be required to have more number of traps to detect gypsy moth or to confirm the pest free areas (IPPC 1995). These considerations should be part of the decision making processes, and resources should be allocated according to the pest predictions and associated uncertainties.

#### **4.4 Discussion**

The future habitat suitability for gypsy moth was evaluated by using forecasted climatic variables under different climate change scenarios. The MaxEnt models performed better than random for all time periods for gypsy moth, with relatively good AUC values. The models for all time periods under both climate change scenarios predicted that current distribution areas (northeastern United States) would mainly remain suitable for at least the next 75 years.

USDA APHIS frequently produces pest forecast maps to support agricultural trade, develop pest risk assessments, assist pest surveys, and manage quarantine pests. Climate data is one of the key elements (data drivers) for pest forecast models, and APHIS has been normally using averaged climatic products as data drivers and has ignored errors associated with model outputs.

This study revealed that using averaged climatic variables is problematic. When MaxEnt models were run by using averaged climatic variables for the specific time periods (e.g., 2016-2040), the mean probability resulting from the model indicated that only the northeastern United States were climatologically suitable for gypsy moth growth (Figure 4-9; Figure 4-10). However, when climatic variables were also simulated in MaxEnt models, other parts of the United States (e.g., northeastern Washington, northern Idaho, and northwestern Montana) resulted in higher probability of establishment (as indicated by larger areas suitable), meaning those areas could support gypsy moth growth if the climatic conditions become suitable during a specific year.

Those areas still remain with high uncertainty because I do not know precisely how climate would play out every day. Permanent establishment may not be supported but for some pests, seasonal incursions can still have massive impacts. MaxEnt outputs derived from mean climatic variables did not rate those areas as suitable areas, but the output derived from mean and simulated climatic values around the mean based on standard deviation suggests that there may be some years that would support gypsy moth development. It is critical to use the additional information gained by simulating climatic variables to incorporate seasonality into the analysis.

I randomly generated climatic values around the means to incorporate climate variability into the pest forecasts. Instead of generating random values around the mean for climate variables, possible and more meaningful combinations of 6 climatic variables such as those I used in this study should be supplied to MaxEnt to incorporate climatic variability (Arellano, personal communication, 2016). In addition, standard deviations from several simulations of MaxEnt (each simulation can consist of 10 replicates of each model run) should be combined instead of taking outputs from all replicates and simulations to calculate one mean and standard deviation (Arellano, personal communication, 2016). This approach allows me to combine standard errors from both replicate and combination (combination of meaningful climatic values) effects.

Generating potential climate condition combinations is still a challenge. It is almost impossible to have all possible combination of several climatic variables. Determining meaningful combinations of data driver values within a certain distance of standard deviation from the mean is still not a systematic process and I have used empirical

approaches based on my own understanding of the processes. Importantly, the pest framework, SAFARIS, helps to generate some combinations of the climatic values by automatically processing climate values based on specific criteria.

The mean probability and the maximum and minimum probabilities generated from MaxEnt replicates were examined in this study. Maximum and minimum probabilities can be evaluated by examining standard deviation at each location; however, I used maximum and minimum probabilities at each location to illustrate the differences clearly. The areas that had wider differences between maximum and minimum had higher standard deviation, indicating that those areas had high uncertainties. Based on the mean and standard deviation, those areas could become suitable for gypsy moth establishment. For example, the entire eastern United States, except Florida, Gulf coastal area, and Texas, could become suitable for gypsy moth in the near future based on the climatic values estimated under GDFL-ESM2G RCP 8.5 model.

In order to incorporate uncertainty into a decision making process, a mean-variance frontier method was used to evaluate pest forecast outputs. Yemshanov et al. (2015) pointed out that outputs should be ranked based on the objective of the study. I wanted to identify areas that could gain more information about gypsy moth; therefore, areas with high uncertainty received the highest priority in this study.

The mean-variance frontier (MVF) approach was applied to the results from MaxEnt to rank locations based on the mean probability and associated standard deviation. Historically, APHIS has not evaluated errors or uncertainty associated with the pest forecasts. This case study demonstrated that incorporating prediction errors was important

because APHIS could better prioritize resources efficiently to gain more information about pests like gypsy moth.

The MVF approach was a simple but powerful method. Since the MVF only requires a mean prediction and associated standard deviation (or variance), this approach can be adapted to many different types of pest forecast models. The MVF analysis tool is now part of SAFARIS tools and can be used on any pest forecast models that SAFARIS currently provides (e.g., phenology model, population dynamics, temperature/precipitation condition tools). Additionally, the MVF can be applied to multiple sources of uncertainty. For example, degree-day requirements for the insect growth are used to estimate the phenology of the insects. APHIS risk analysts normally use degree-day requirement at a point where 50% of individuals complete a certain stage. It can be a richer, more representative approach to run degree-day requirements at different individual completion (e.g., 25%, 90%) and estimate standard errors for the biology of the pest to incorporate uncertainty into the pest forecast outputs. Also, phenology models can be run at multiple years to estimate annual climate variability. The MVF evaluation can be performed on both biology and climate aspects to incorporate uncertainty into the pest forecast outputs.

Three additional observations are made here to clarify where future research is indicated: 1) host information is not included in this climate change study; 2) there are several more GCMs available, and 3) this study focuses on the United States, I plan to expand similar analyses to a global scale. The effects on host species (e.g., distribution, growth) need to be evaluated in order to improve current understanding of how gypsy moth might change its habitat under climate change. It is clearly uncertain how climate might

change into future in specific terms that would be needed to increase our certainty about pest establishment or geographic suitability. I only used climatic values from one GCM because the focus of this study was to quantify uncertainty associated with pest prediction. But multiple GCMs should be used to minimize uncertainty associated with climate change.

#### 4.5 References

- Abatzoglou, J. T. 2013. Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology* **33**:121-131.
- Arellano, C. (2016). Conversation on probabilistic models. Personal communication to on May 12, 2016, from meeting discussion.
- Azar, C., and H. Rodhe. 1997. Targets for stabilization of atmospheric CO<sub>2</sub>. *Science* **276**:1818-1819.
- Bale, J. S., A. Buse, J. Butterfield, J. Farrar, J. C. Coulson, V. K. Brown, I. D. Hodkinson, G. J. Masters, T. M. Bezemer, and C. Awmack. 2002. Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. *Global change biology* **8**:1-16.
- Berger, A. L., V. J. D. Pietra, and S. A. D. Pietra. 1996. A maximum entropy approach to natural language processing. *Computational linguistics* **22**:39-71.
- Boyles, R., and H. R. Aldridge. 2016. Probability density functions evaluations on 20 climate models North Carolina State University, The Climate Office of North Carolina.
- Cannon, R. J. C. 1998. The implications of predicted climate change for insect pests in the UK, with emphasis on non-indigenous species. *Global change biology* **4**:785-796.
- Daly, C., and K. Bryant. 2013. The PRISM climate and weather System - an introduction. Oregon State University, Northwest Alliance for Computational Science and Engineering.
- Elith, J., M. A. Burgman, and H. M. Regan. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling* **157**:313-329.

- Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* **40**:677-697.
- Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17**:43-57.
- Fitzpatrick, M. C., A. D. Gove, N. J. Sanders, and R. R. Dunn. 2008. Climate change, plant migration, and range collapse in a global biodiversity hotspot: the *Banksia* (Proteaceae) of Western Australia. *Global change biology* **14**:1337-1352.
- Fowler, G. (2015). Distribution of European gypsy moth. Personal communication to on October 1, 2015, from conversation.
- Gutierrez, A. P., N. J. Mills, and L. Ponti. 2010. Limits to the potential distribution of light brown apple moth in Arizona-California based on climate suitability and host plant availability. *Biological Invasions* **12**:3319-3331.
- Haack, R. A., F. Herard, J. H. Sun, and J. J. Turgeon. 2010. Managing invasive populations of Asian longhorned beetle and Citrus longhorned beetle: a worldwide perspective. Pages 521-546 *Annual Review of Entomology*. Annual Reviews, Palo Alto.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**:1965-1978.
- IPCC. 2013. Climate change 2013: the physical science basis. Intergovernmental Panel on Climate Change.
- IPPC. 1995. ISPM No. 4: Requirements for the establishment of pest free areas. International Plant Protection Convention, Rome, Italy.
- Koch, F. H., D. Yemshanov, D. W. McKenney, and W. D. Smith. 2009. Evaluating critical uncertainty thresholds in a spatial model of forest pest invasion risk. *Risk Analysis* **29**:1227-1241.
- Lindroth, R. L., K. K. Kinney, and C. L. Platz. 1993. Responses of deciduous trees to elevated atmospheric CO<sub>2</sub>: productivity, phytochemistry, and insect performance *Ecology* **74**:763-777.
- Lozier, J. D., and N. J. Mills. 2011. Predicting the potential invasive range of light brown apple moth (*Epiphyas postvittana*) using biologically informed and correlative species distribution models. *Biological Invasions* **13**:2409-2421.
- Morisette, J. T., C. S. Jarnevich, T. R. Holcombe, C. B. Talbert, D. Ignizio, M. K. Talbert, C. Silva, D. Koop, A. Swanson, and N. E. Young. 2013. VisTrails SAHM:

- visualization and workflow management for species habitat modeling. *Ecography* **36**:129-135.
- NOAA. 2016a. Global Surface Summary of the Day - GSOD. National Oceanic and Atmospheric Administration, National Centers for Environmental Information. <https://data.noaa.gov/dataset/global-surface-summary-of-the-day-gsod>.
- NOAA. 2016b. Trends in atmospheric carbon dioxide. U.S. Department of Commerce, National Oceanic & Atmospheric Administration, Earth System Research Laboratory, Boulder, CO. Last accessed January, 2016, <http://www.esrl.noaa.gov/gmd/ccgg/trends/>.
- O'Connell, B. O., E. B. LaPoint, J. A. Turner, T. Ridley, S. A. Pugh, A. M. Wilson, K. L. Waddell, and B. L. Conkling. 2015. The Forest Inventory and Analysis database: database description and user guide for phase 2 (version 6.0.2). United States Forest Service, Washington, D.C.
- O'Donnell, M. S., and D. A. Ignizio. 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States. U.S. Geological Survey Data Series 691:10p.
- Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* **34**:102-117.
- Peterson, A. T., and Y. Nakazawa. 2008. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Global Ecology and Biogeography* **17**:135-144.
- Phillips, S. 2010. A brief tutorial on MaxEnt. *Lessons in Conservation* **3**:107-134.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2005. Maximum entropy modeling of species geographic distributions. *Ecological Modeling* **190**:231-259.
- Poland, T. M., and D. G. McCullough. 2006. Emerald ash borer: invasion of the urban forest and the threat to North America's ash resource. *Journal of Forestry* **104**:118-124.
- Porter, J. H., M. L. Parry, and T. R. Carter. 1991. The potential effects of climatic change on agricultural insect pests. *Agricultural and Forest Meteorology* **57**:221-240.
- Stinner, R. E., Butler G. D, Jr., J. S. Bacheler, and C. Tuttle. 1975. Simulation of temperature-dependent development in population-dynamics models. *Canadian Entomologist* **107**:1167-1174.

- Stohlgren, T. J., P. Ma, S. Kumar, M. Rocca, J. T. Morisette, C. S. Jarnevich, and N. Benson. 2010. Ensemble habitat mapping of invasive plant species. *Risk Analysis* **30**:224-234.
- UCAR. 2016. How much carbon dioxide (and other kinds of greenhouse gas) is already in the atmosphere? University Corporation for Atmospheric Research, National Center for Atmospheric Research. Last accessed January, 2016, <https://www2.ucar.edu/climate/faq/how-much-carbon-dioxide-and-other-kinds-greenhouse-gas-already-atmosphere>.
- USDA NASS. 2016. CropScape and Cropland Data Layer. United States Department of Agriculture, National Agricultural Statistics Service, Washington, DC. Last accessed March, 2016, [http://www.nass.usda.gov/Research\\_and\\_Science/Cropland/SARS1a.php](http://www.nass.usda.gov/Research_and_Science/Cropland/SARS1a.php).
- van Vuuren, D. P., J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G. C. Hurtt, T. Kram, V. Krey, J. F. Lamarque, T. Masui, M. Meinshausen, N. Nakicenovic, S. J. Smith, and S. K. Rose. 2011. The representative concentration pathways: an overview. *Climatic Change* **109**:5-31.
- Webber, B. L., C. J. Yates, D. C. Le Maitre, J. K. Scott, D. J. Kriticos, N. Ota, A. McNeill, J. J. Le Roux, and G. F. Midgley. 2011. Modelling horses for novel climate courses: insights from projecting potential distributions of native and alien Australian acacias with correlative and mechanistic models. *Diversity and Distributions* **17**:978-1000.
- Yemshanov, D., F. H. Koch, and M. Ducey. 2015. Making invasion models useful for decision makers; incorporating uncertainty, knowledge gaps, and decision-making preference. Pages 206-222 *in* R. C. Venette, editor. *Pest Risk Modeling and Mapping for Invasive Alien Species*. CAB International, Wallingford, UK.

## **CHAPTER 5. Risk Assessment for Asian Gypsy Moth (*Lymantria dispar*) using the Spatial Analytic Framework for Advanced Risk Information System**

### **5.1 Introduction**

The gypsy moth, *Lymantria dispar* (Linnaeus), exhibits several biotypes, all of which are considered damaging and are regulated by many countries as quarantine pests of concern. The North American infestation of the European biotype was first introduced into the northeastern United States from Europe in 1869 (Montgomery and Wallner 1988). This European biotype has become established and spread to 19 states and the District of Columbia in the eastern United States and is referred to as the North American gypsy moth (NAGM) or European Gypsy Moth (EGM) (USDA-APHIS-PPQ 2015a) (Figure 5-1). EGM is considered a major forest pest in the United States and Canada due to its ability to cause economic and environmental damage (Tobin et al. 2012). Another distinct biotype, the Asian biotype (AGM) has traits, described below, which if it became introduced, would likely increase the damage, expand overall spread and require the development of new management tools and programs.

AGM and closely related taxonomic forms are pests of forests in China, eastern Russia, Korea and Japan. The AGM biotype is considered to be a more threatening pest than EGM because: 1) AGM females are capable of long distance flight, unlike EGM females; 2) its host range is broader than that of EGM (Humble and Stewart 1994, Zlotina et al. 1999); and 3) practical pest management alternatives for North American conditions may not be available.

Pogue and Schaefer (2007) reclassified AGM into five discrete taxonomic groups depending on biology, morphology, and distribution. These are: *Lymantria dispar asiatica*,

*L. dispar japonica*, *L. albescens*, *L. umbrosa* and *L. postalba*. *Lymantria dispar asiatica* is the species found in China, Korea and Russia. As is the case for most *Lymantria*, females of this species are attracted to lights and can exhibit mass migrations.

Over the past 20 years AGM has been intercepted at U.S. ports and locations in: California, Idaho, South Carolina, North Carolina, Oregon, Oklahoma, Texas and Washington State (USDA-APHIS-PPQ 2014, 2015b) (Figure 5-2). The most obvious pattern is the concentration of interceptions near the U.S. coasts and ports of entry. The second most obvious pattern is that most interceptions occur along the West Coast of the United States and Canada. This second observation is likely linked to the fact that less than 25% of shipments (see maritime pathway study: (Mastro et al. 2007)) that originate in China call at Eastern U.S. ports. This pattern is likely to experience shifts with the recently initiated project to enlarge the Panama Canal to accommodate vessels of any size. Of all these interceptions, the inland interception in Texas in 2006 was likely due to the movement of containers offloaded from infested shipments or containers which were exposed at a staging area where gypsy moth had an opportunity to deposit eggs.

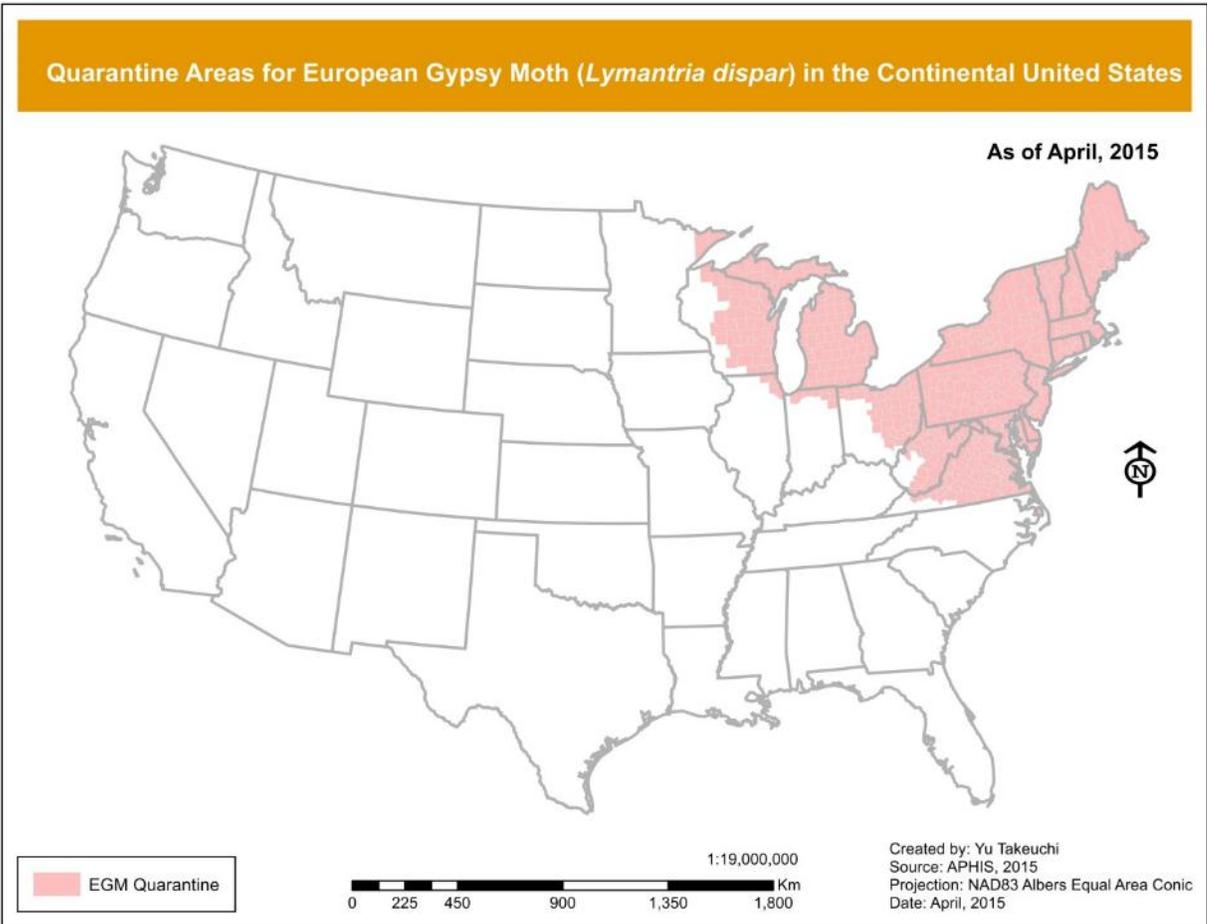
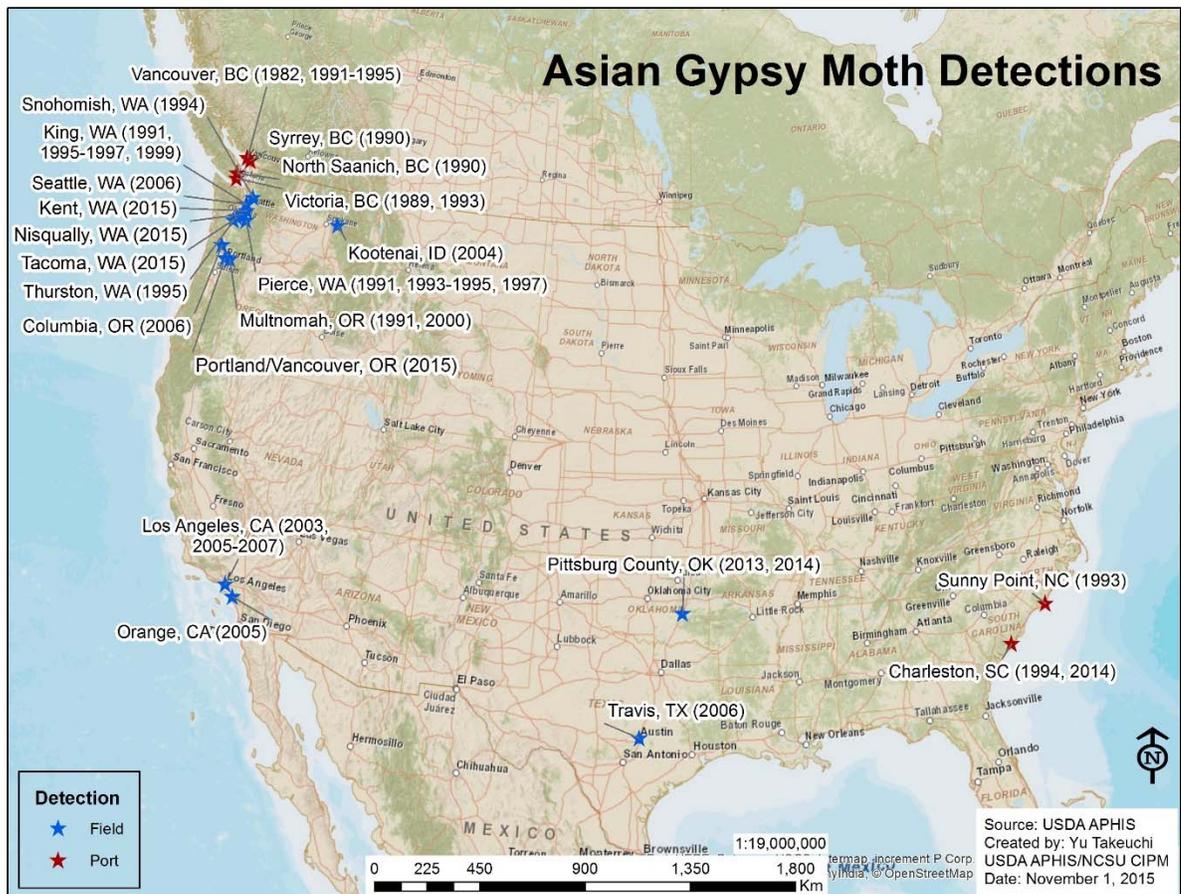


Figure 5-1. Quarantine Areas for European Gypsy Moth (*Lymantria dispar*) in the continental United States as of April, 2015



**Figure 5-2. Asian gypsy moth detections in the United States**

The objectives of this study are to understand the likelihood of Asian Gypsy Moth (*Lymantria dispar asiatica*) introduction within the contiguous United States using three models: phenology model, temperature-driven demographic model, and human assisted movement model. Successful introductions occur when AGM spread via natural means or human assisted movements and find hosts in climatologically suitable areas. In this study, I analyze the potential introduction of AGM characterizing GM biology, host species, and truck movements from at-risk ports.

## **5.2 Biological Information**

The biology of gypsy moth has been described by several authors including Humble & Stewart (1994), Andresen et al., (2001) and Pogue & Shafer (2007). This section highlights aspects that may help understand the forecast model and the use of the SAFARIS (Spatial Analytic Framework for Advanced Risk Information) framework; the section is not intended as a comprehensive review of all aspects of AGM biology.

### **5.2.1 Morphology**

**Eggs:** Asian gypsy moth egg masses are beige in color (Humble and Stewart 1994). One egg mass can contain from 500 to over 1,200 eggs, approximately five millimeters long by two millimeters wide (Humble and Stewart 1994, Andresen et al. 2001). Egg masses are typically laid on tree boles and larger branches but can be laid on any surface (Humble and Stewart 1994). Additionally, USDA APHIS has repeatedly recorded observations of moths attracted to lights and ovipositing on nearly any structure including vessel superstructure, containers, buildings, automobiles and even along roadways and sidewalks (Humble and Stewart 1994).

**Larvae:** Newly hatched larvae are typically hairy and are approximately 2 to 3 mm long. The number of instars varies with host species; however, typically male larvae go through five instars, while female larvae go through six (Barbosa and Greenblatt 1979, Wilson 2015). Mature male larvae are typically 40-50mm and matured female larvae are about 60-70mm (Humble and Stewart 1994, CABI 2015). First-instar larvae are grey-black. Later instars have distinctive colors, with two rows of large spots along the back usually

arranged in five pairs of blue and six pairs of red from head to rear (Humble and Stewart 1994).

**Pupae:** The pupae are dark red-brown color and matted with reddish hair (Humble and Stewart 1994). The female pupae (15-35mm) are usually larger than male (15-20mm) (Humble and Stewart 1994). Pupae can be attached to trees, stones or other objects by silken threads (CABI 2015).

**Adults:** There is variation in adult forewing color and pattern in adult moths (Pogue & Schaefer, 2007). Male moths from Korea and China have dark brown forewing, while males from central Russia have light brown forewing (Pogue & Schaefer, 2007). Female moths typically have white forewings, however, may also have slight brownish forewings, especially for some moths from China (Pogue & Schaefer, 2007). Wingspan in males is about 38mm, and females are slightly larger than males (Dixon and Foltz 1985).

### **5.2.2 Life Cycle**

Gypsy moth (*Lymantria dispar* L.) is a univoltine species. European gypsy moths (EGM) hatch in early spring in the continental United States when the foliage of hosts like oak starts expanding (Wilson 2015). The growth rate varies with temperature, host availability, and other environmental factors; however, the larval stage normally lasts approximately 40 days. Matured larvae go into pupal stage in early summer (June to early July). The pupal stage lasts about three weeks. Male moths emerge a few days before female moths (July to August). The female EGM mates shortly after emergence and lays a single egg mass within 24 hours (Giebultowicz et al. 1991). The female EGM is flightless; however, the female AGM is capable of flying. The female AGM is known to fly up to 30

kilometers (Humble and Stewart 1994). The egg mass remains at the same place until next spring, approximately 8 to 9 months later (Andresen et al. 2001).

### **5.2.3 Host species**

Gypsy moth (GM) (*Lymantria dispar* L.) is a polyphagous species. It has been reported that European gypsy moth (EGM) attacks more than 300 species of trees and shrubs (Doane and McManus 1981). Although gypsy moth attacks many trees and shrubs, the genus *Quercus* species are the most favored hosts for EGM (Liebhold et al. 1995). The most preferred host species for Asian gypsy moth (AGM) are same as EGM; however, AGM has a wider host range than EGM and has been reported to feed on more than 600 species in more than 100 plant families (Humble and Stewart 1994).

Host tree species are mainly distributed and most common throughout eastern United States., Pacific coastal areas, and Rocky Mountains in the continental United States, covering approximately 33% of total land areas of the United States (Smith et al. 2004) (Figure 5-3). Although GM prefers *Quercus* species and other deciduous species over conifers, GM will feed on most deciduous tree species. Especially when AGM is abundant, most tree and bushes are at risk.

Although the distribution of tree species is helpful to understand the GM potential habitats, the distribution does not indicate the volume (mass) of the trees occurring at given areas. Figure 5-4 indicates the mass (or weight) of trees in mega-grams per hectare (Blackard et al. 2008). The northern Pacific coastal area and Sierra Nevada had the highest tree biomass (high tree density) within the contiguous United States that accounts for more than 200 tons of biomass above the ground per hectare. Combining the preferred host

distribution and biomass information can help us better understand the potential damage caused by GM.

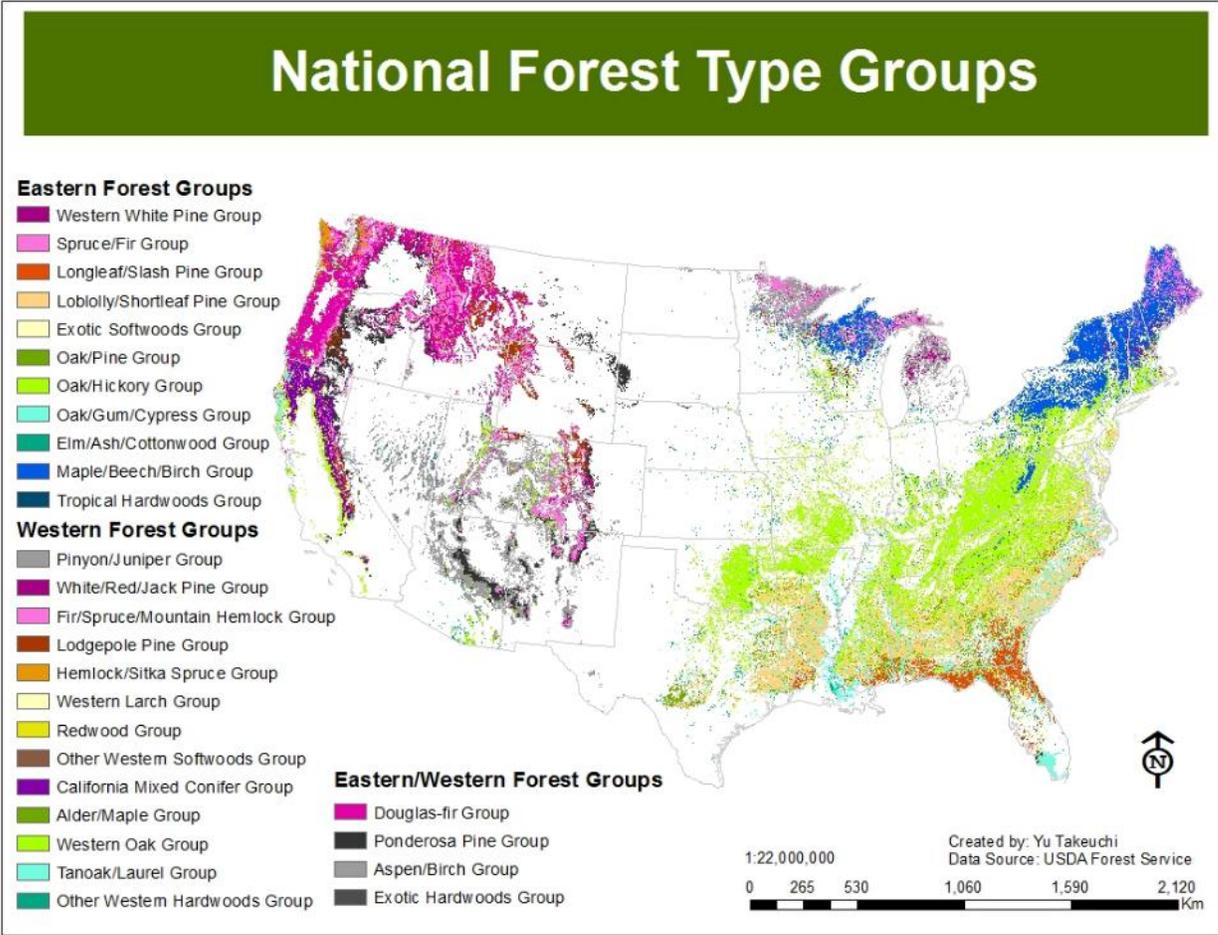
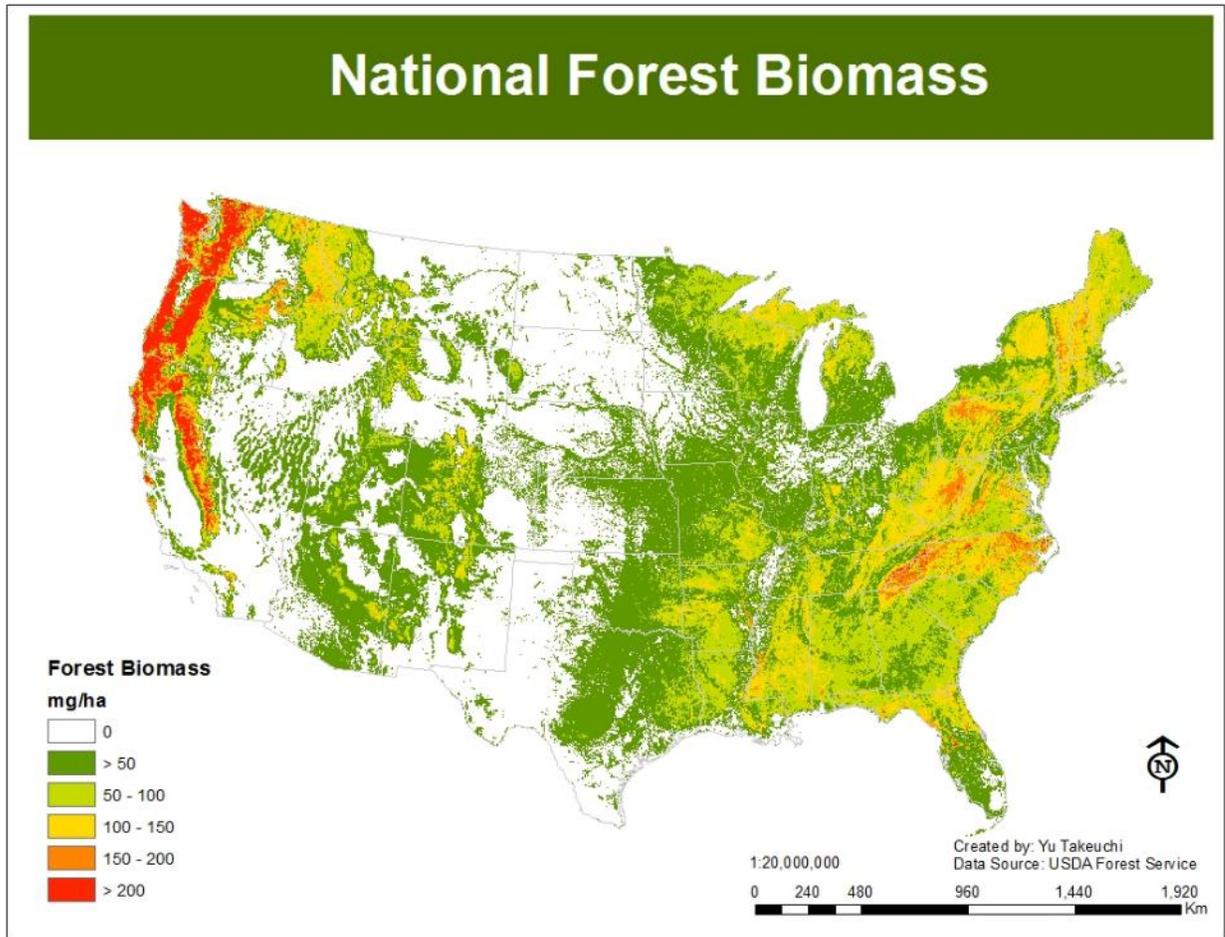


Figure 5-3. National forest type groups in the continental United States



**Figure 5-4. Forest biomass (mg per hectare) in the continental United States**

### **5.3 Analytic Methods**

#### **5.3.1 Spatial Analytic Framework**

Prevention, eradication, containment, and mitigation of exotic agricultural pests are affected by various biotic and abiotic factors. These factors include weather and climate patterns, host availability, human and wind-dispersal, and effectiveness and economics of management practices. Each of these factors has to be assessed from the port of entry to the field or from the infested area to un-infested area. Climate plays a key role in determining

pest spread and distribution because agricultural pests are poikilothermic organisms. In addition, future climate change scenarios have to be included in the assessment and analysis.

The Spatial Analytic Framework for Advanced Risk Information Systems (SAFARIS) is an integrated pest forecasting framework composed of 3 parts: database containing model drivers, knowledge base that stores pest information and expert opinions, and pest forecast models and analytic tools. The SAFARIS assists predicting exotic species behavior for assisting pest survey, risk assessments, pest emergency responses, and economic assessments.

### **5.3.2 Phenology Model**

The initial approach to forecasting suitable areas for AGM uses models that focus on the phenology of the species. Such an approach provides a relationship between climatic conditions, notably temperature patterns, and rate of development of the pest such that the timing of events such as oviposition, larval instars, pupal and adult stages can be determined. The phenology model, however, is integrated with spatial layers that contain appropriate host material as well as other environmental factors known to affect development of the pest. Thus, when I refer to a phenology model here, I refer to an approach where I integrate biologically relevant data along with development rate estimates. Specifically, I integrate temperature dependent development, high resolution space/time climatic data, host presence, known extreme topographical or prevailing conditions to refine estimates of development over large continental and global scales. The SAFARIS is used to provide a computing environment that facilitates the integration of appropriate biological and abiotic drivers.

The phenology of GM (*L. dispar* L.) was determined by using the degree-day requirements estimated by Johnson et al. (1983), Sheehan (1992), Carter et al. (1992), and Russo et al. (1993). The phenology model was used to predict AGM initiation times for larval stage, pupa stage, and adult emergence within the continental United States. The degree-day parameters for male and female GM are summarized in Table 5-1 and

Table 5-2. The lower threshold temperature for GM development was 3°C (Johnson et al. 1983, Sheehan 1992). The upper threshold temperatures for GM development varied with the insect stages. The upper threshold temperature for egg stage was 40°C (Russo et al. 1993), while the upper threshold temperatures for larval and pupal stages were 38°C (Sheehan 1992).

The weather data used for the phenology model were described by the PRISM Climate Group (2015). PRISM daily temperature data cover the contiguous United States with 4-kilometer resolution developed by PRISM Climate Group at Oregon State University. Daily maximum temperatures and minimum temperatures from 2009 to 2013 were averaged by calendar day and used as drivers for the phenology model. Although the lower threshold temperature for egg stage was 2°C higher than the one for larval and pupal stages, the upper threshold temperature was set to 38°C for all stages in this model. This can introduce a slight underestimate of degree-day accumulations but was not considered a significant factor. Setting the same threshold temperature was done for computational convenience since it was not considered to introduce significant biological error. In addition, GM biology varied with locations, and different lower and upper threshold temperatures were determined by different researchers (Johnson et al. 1983, Carter et al. 1992, Russo et al. 1993). Phenology model parameters indicated in Table 5-1 and

Table 5-2 represent the best for overall conditions for the continental United States.

Whereas there are many ways to approach the estimation of degree-days, the double sine with horizontal cut-off method was selected for degree-day calculations (Johnson et al. 1983, Carter et al. 1992, Russo et al. 1993) because it provides a more accurate integration of actual temperature patterns compared to alternative methods, like triangulation. The initiation times for larval stage, pupa stage, and adult stage were estimated at weekly intervals, for reporting convenience (daily reports are easily produced but are not needed to support most management decisions or for the purposes of this evaluation).

In this study, only female parameters were used to determine the AGM suitability because females takes longer time to develop than males, and I am interested in the reproductive timing of the AGM. The double-sine method was used for degree day accumulation to estimate the timing of the AGM life stages.

**Table 5-1. Male gypsy moth (*Lymantria dispar* L.) degree-day parameters for each stage of development**

<b>GM Stage</b>	<b>Lower Thres</b>	<b>Upper Thres</b>	<b>Required DD</b>	<b>Accumulated DD</b>
<b>Egg Eclosion</b>	3°C	36°C	300 DD	300 DD
<b>First Instar</b>	3°C	38°C	100 DD	400 DD
<b>Second Instar</b>	3°C	38°C	69 DD	469 DD
<b>Third Instar</b>	3°C	38°C	76 DD	545 DD
<b>Fourth Instar</b>	3°C	38°C	98 DD	643 DD
<b>Fifth Instar</b>	3°C	38°C	240 DD	883 DD
<b>Pupa</b>	3°C	38°C	277 DD	1160 DD

**Table 5-2. Female gypsy moth (*Lymantria dispar* L.) degree-day parameters for each stage of development**

GM Stage	Lower Thres	Upper Thres	Required DD	Accumulated DD
<b>Egg Eclosion</b>	3°C	36°C	300 DD	300 DD
<b>First Instar</b>	3°C	38°C	94 DD	394 DD
<b>Second Instar</b>	3°C	38°C	65 DD	459 DD
<b>Third Instar</b>	3°C	38°C	72 DD	531 DD
<b>Fourth Instar</b>	3°C	38°C	93 DD	624 DD
<b>Fifth Instar</b>	3°C	38°C	100 DD	724 DD
<b>Sixth Instar</b>	3°C	38°C	226 DD	950 DD
<b>Pupa</b>	3°C	38°C	234 DD	1184 DD

### 5.3.3 Population Trends Model

Whereas a model based on phenology provides powerful insights as to suitable areas and the potential impact (albeit indirectly, by estimating the potential number of generations), a population model provides additional insights which are especially useful when assessing the impact of a given species.

A population trends model was built based on the effect of temperature on mortality and development rate. Population trends have been modeled based on the experiments by Casagrande, et al. (1987) and Gutierrez (2010). The developmental rate of *Lymantria dispar* (GM: Gypsy Moth) was established based on research reported by Casagrande, et al. (1987) and Andersen et al. (2001); additionally, their results were used as the basis for the temperature-dependent population models for each AGM life stage.

The effects of temperature on developmental rates of GM at larval and pupal stages are summarized in the figures below (Figure 5-6; Figure 5-7) and expressed by following equations (Equation 1 and 2). These equations were developed based on the experimental results that Casagrande, et al. (1987) reported.

The lower threshold temperatures of 11.1°C and 13.3°C were used in these equations, instead of 3°C used in the phenology model. Based on the experiments conducted by Casagrande, et al. (1987), larval and pupal developments were not completed at 11.1°C and 13.3°C, respectively. The upper threshold temperatures of 32.0°C and 33.0°C were also determined based on the studies by Casagrande, et al. (1987).

The mean duration of the larval stage

Equation 1 (Figure 5-6):

$$\text{Male larval growth rate} = \frac{0.002456 * (T_{avg} - \text{lower thres. temp})}{(1 + 2.35426^{(T_{avg} - \text{upper thres. temp})})}$$

$$\text{Female larval growth rate} = \frac{0.002141 * (T_{avg} - \text{lower thres. temp})}{(1 + 2.088343^{(T_{avg} - \text{upper thres. temp})})}$$

Where, lower threshold temperature = 11.1°C

upper threshold temperature = 32.0°C

Equation 2 (Figure 5-7):

$$\text{Male pupal growth rate} = \frac{0.005593 * (T_{avg} - \text{lower thres. temp})}{(1 + 3.433153^{(T_{avg} - \text{upper thres. temp})})}$$

$$\text{Female pupal growth rate} = \frac{0.007406 * (T_{avg} - \text{lower thres. temp})}{(1 + 2.058586^{(T_{avg} - \text{upper thres. temp})})}$$

Where lower threshold temperature = 13.3°C

upper threshold temperature = 33.0°C

In insect studies, the mortality rates also are influenced by temperature patterns. The relationship between temperature and mortality rate for larval stage was parameterized

based on GM biological data collected by Casagrande et al. (1987) (Equation 3) (Figure 5-8). The daily mortality rate was the lowest at 25.6°C (Casagrande et al. 1987); however, reanalysis of the data resulted that the lowest mortality rate was at 24.5°C. The mortality at pupal stage was sensitive to temperature but was difficult to parameterize. Casagrande et al. (1987) reported that there were no successful pupal development at temperatures below 13.3°C and there were some mortality observed at 17.8 °C and 28 °C. But there were no pupal mortality at 20°C, 25°C, 25.6°C, 30°C, 31.1°C, and 32°C. Therefore, I simplified the pupal mortality rate to be 1 if average mean temperature during pupal stage was below 13.3°C (Equation 4). There was no mortality if average mean temperature during pupal stage was above 13.3°C.

Equation 3 (Figure 5-8):

$$\begin{aligned} \text{Larval mortality rate} \\ &= 0.0000395 * T_{avg}^2 - 0.0019300345 * T_{avg} \\ &+ 0.0240450957 \end{aligned}$$

Equation 4:

If mean  $T_{avg}$  during pupal stage  $\leq 13.3$ ,  
*Pupal mortality rate* = 1

If mean  $T_{avg}$  during pupal stage  $> 13.3$ ,  
*Pupal mortality rate* = 0

In addition to the mortality caused by climate conditions, there are other factors affecting the survival of GM. Several parasite species of GM have been identified. In North America, *Ooencyrtus kuvanae* and *Anastatus disparis* are known to parasitize GM eggs; *Apanteles melanoscelus*, *Phobocampe disparis*, *Blepharipa pratensis*, *Parasetigena silverstris*, *Exorista larvarum*, and *Compsilura concinnata* parasitize GM larvae; *Brachymeria intermedia* is known to attach GM pupae (Weseloh 1976, Campbell 1981,

Brown and Cameron 1982). More importantly, an entomopathogenic fungus, *Entomophaga maimaiga*, is known to kill GM larvae (Andreadis and Weseloh 1990), and has caused significant reductions of GM populations (and impacts) in most of its current US distribution (Hajek et al. 1995). Birds and mammals (e.g. blue jays, white-footed mouse, and short tail shrew) also affect GM populations.

Although there are many factors reportedly affecting the population of GM in the field, only mortality caused by temperature was included in this model. First, it is difficult to estimate the population levels of natural enemies in the fields. There are some studies done in the fields to understand the interactions between natural enemies and GM population levels (Morse and Simmons 1979, Gould et al. 1990). However, the environments and conditions vary with locations, and it is difficult to accurately estimate the microclimate throughout the continental United States. Second, natural enemies are not known to be a critical factor in GM introduction potential which is a key phytosanitary focus. Finally, we do have a good understanding of the effects of the *Entomophaga maimaiga* on GM, but we have yet to characterize the impact of natural enemies on AGM.

The temperature effect on egg masses has not been well understood. The egg mass is cold tolerant, and its lethal low temperature has been reported as -29°C (Madrid and Stewart 1981) under laboratory condition and as -80 °C ambient temperature in the field, depending on the locations where egg masses were laid (Sullivan and Wallace 1972). Several studies indicated that presence of snow cover affected the survival of egg mass (Sullivan and Wallace 1972, Higashiura 1989, Andresen et al. 2001). Several studies reported that snow cover protected from air temperature and increased survival of egg

masses (Sullivan and Wallace 1972, Higashiura 1989), while egg mass survival was greater when the egg masses were laid above snow line (Leonard 1972). In addition, Andresen et al. (2001) reported that angle and exposure to the sun (aspects) influenced egg mass temperature and mortality more than air temperature.

Whereas it was a straightforward calculation when the parameters are available, the cold temperature effect on egg mass mortality was not modeled in this study for two main reasons. First, it has been shown that egg mass can survive at extreme cold temperature and prevailing conditions are such that adding these effects does not change the outcomes for the continental United States as shown below (Figure 5-5). Figure 5-5 shows the monthly average temperature for winter months (December, January, and February) and mean winter months air temperature. There was no area that indicated that mean temperature was below  $-29^{\circ}\text{C}$  in the contiguous United States.

The second reason that low temperature mortality is not considered as key feature is that egg mass mortality is affected by factors such as snow line and aspect to the sun where the egg mass is laid. These factors are difficult to estimate because these parameters are not usually available, especially for the entire contiguous United States level. Therefore, egg mortality was not factored in this model because it was not considered to add biologically significant information and it was not considered practical for the kinds of phytosanitary applications envisioned for the framework and associated models.

GM is known to require certain level of cold environment during winter for the eggs to hatch next spring. Although cold temperature mortality was not factored into this model,

the mortality was modeled if GM did not accumulate enough cold temperature during winter period.

GM diapause is experienced by the eggs only; the temperature during the diapause stage affects hatching rate and days to first hatch (Keena 1996). GM egg masses require exposure to cold temperature for certain periods of time to successfully eclose in the spring. The experiments by Keena (1996) revealed that there were significant variation among GM strains; however, none of the egg masses hatched if egg masses were kept at 5°C for 30 days. The egg hatching effect influenced by exposure to cold temperature was parameterized by using experiments on Russian strain conducted by Keena (1996) (Figure 5-9; Figure 5-10). The hatching rate (range of 0 to 1) was computed using Equation 5.

Equation 5:

If number of days that  $T_{min}$  is below 5 °C less than 30 days,

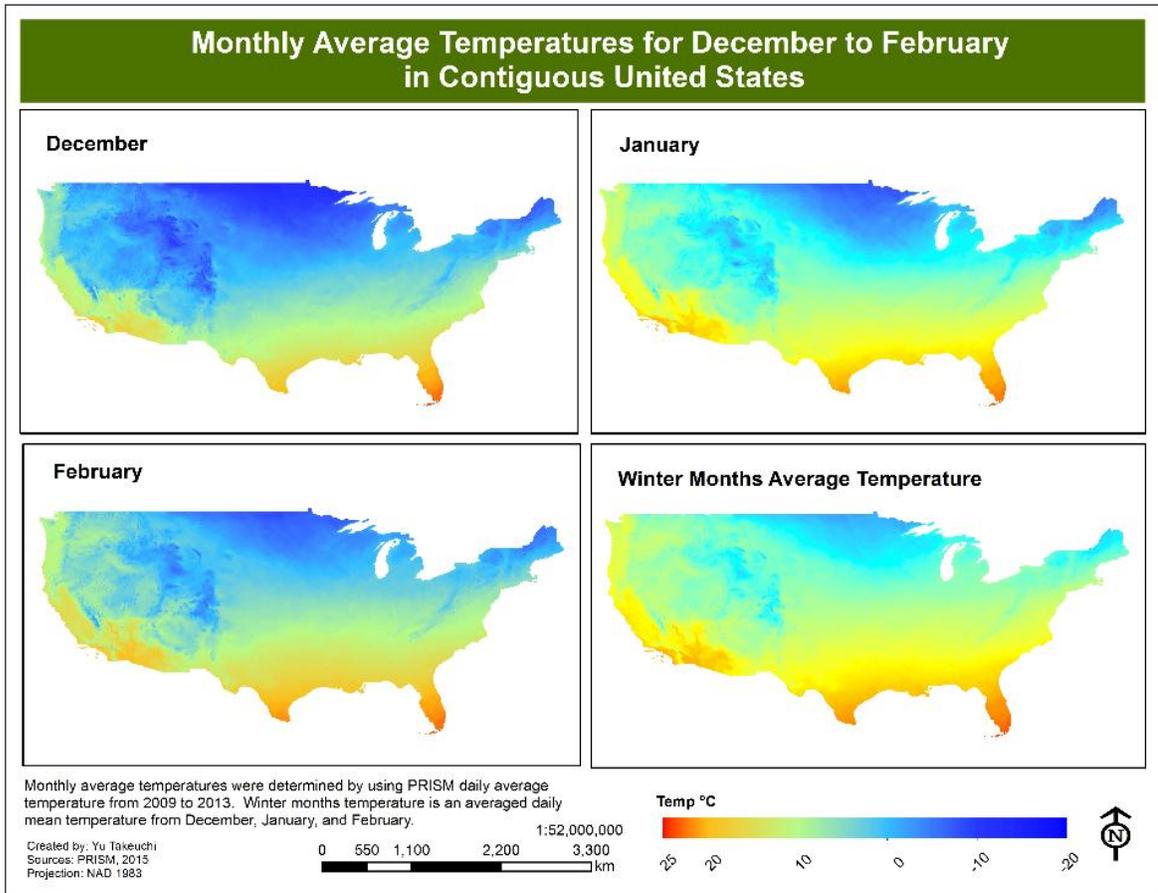
$$Hatching\ rate = 0$$

If number of days that  $T_{min}$  is below 5 greater than 30 days,

$$Hatching\ rate = -0.00009 * (Number\ of\ days\ T_{min} < 5^{\circ}C)^2 + 0.0267 * (Number\ of\ days\ T_{min} < 5^{\circ}C) - 0.75$$

If hatching rate > 1, hatching rate = 1

If hatching rate < 0, hatching rate = 0



**Figure 5-5. Monthly average temperature for winter months developed by using PRISM daily Tmean data**

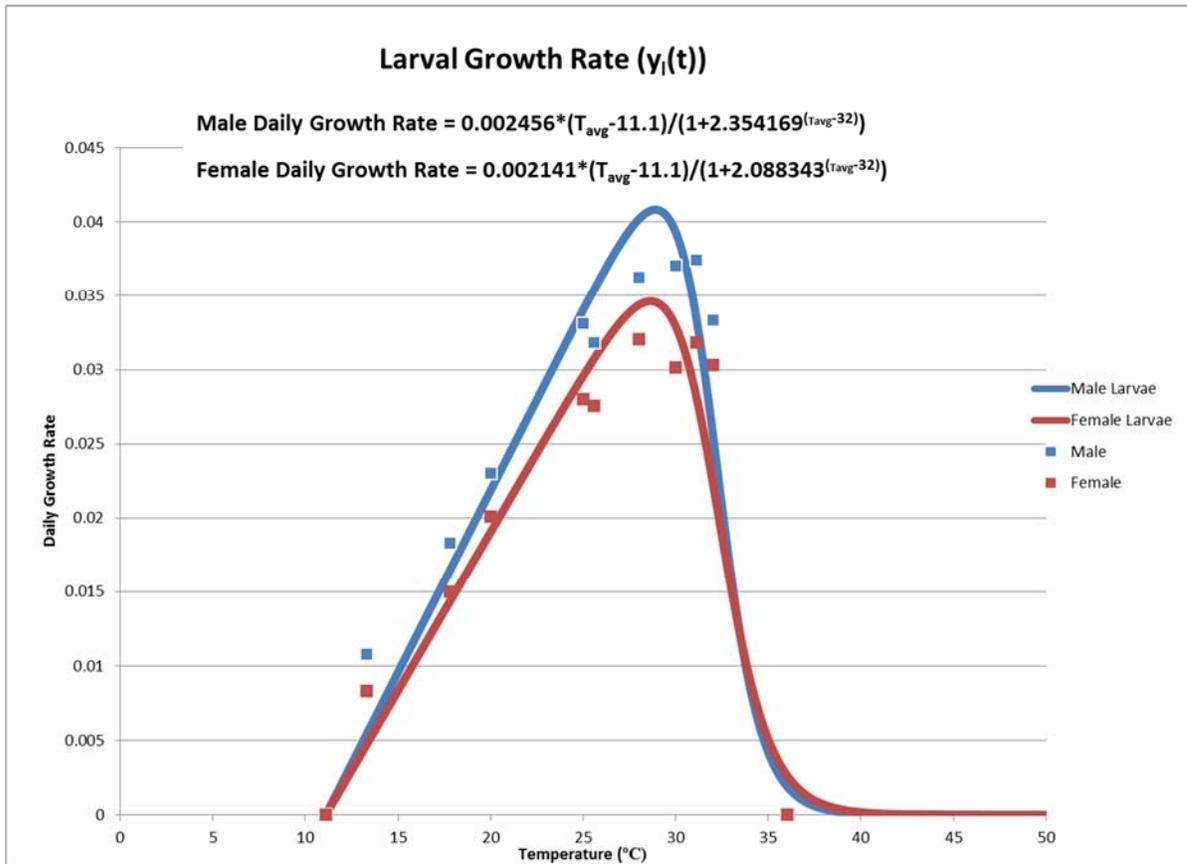
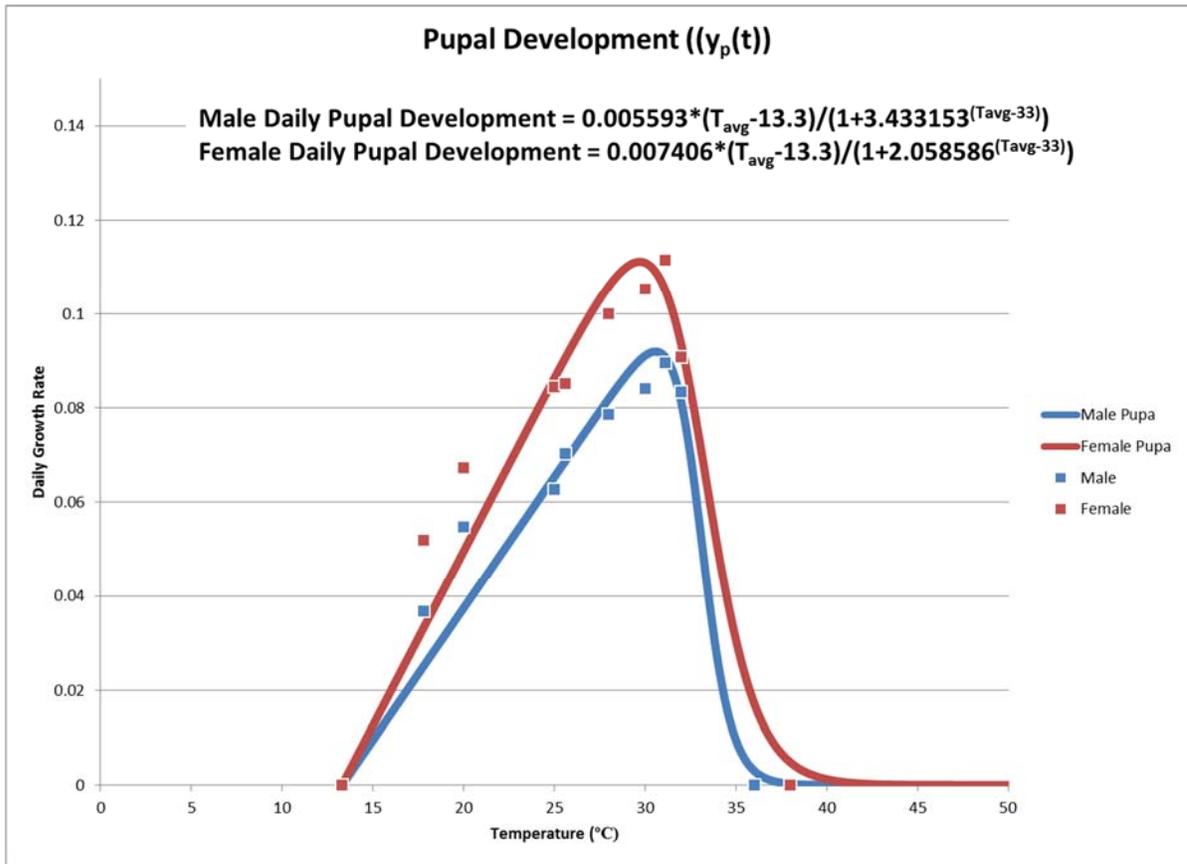


Figure 5-6. The effect of temperature on the developmental rates of GM larvae (data from Casagrande et al. (1987))



**Figure 5-7. The effect of temperature on the developmental rates of GM pupae (data from Casagrande et al. (1987))**

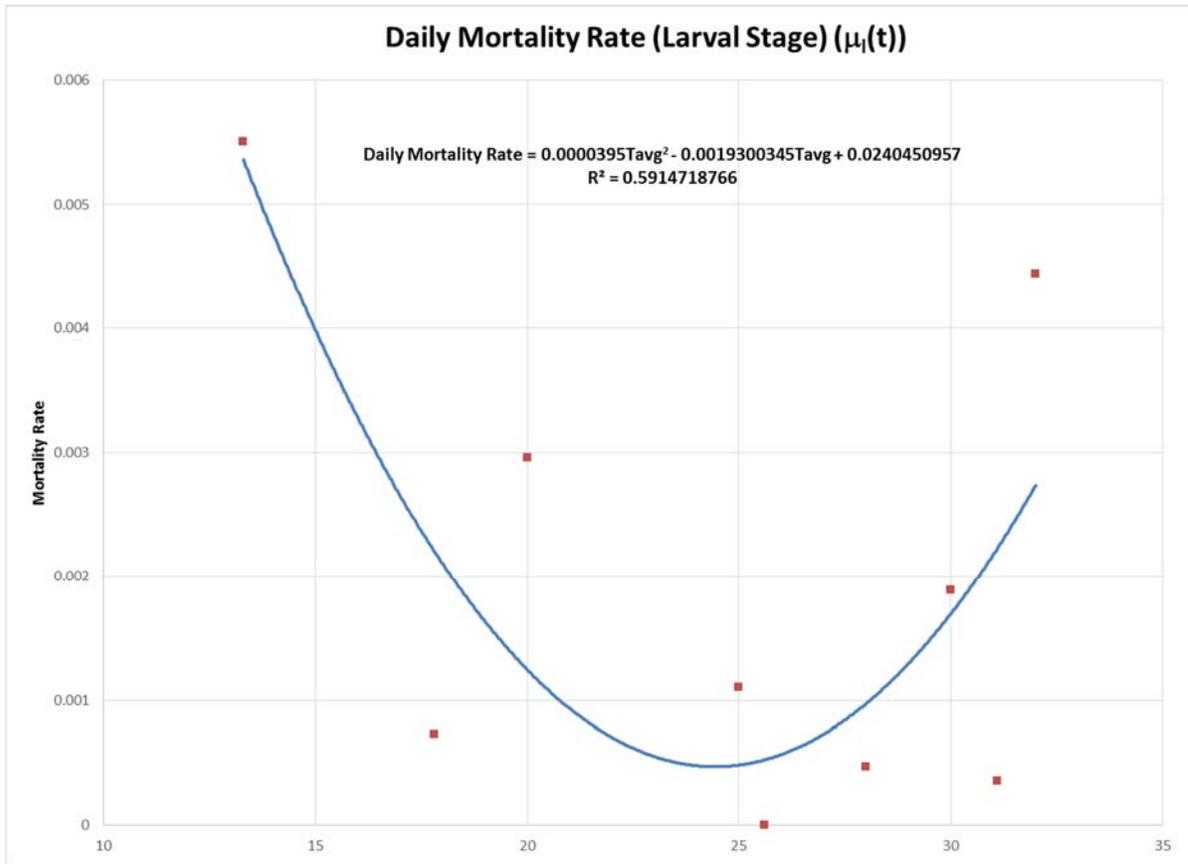
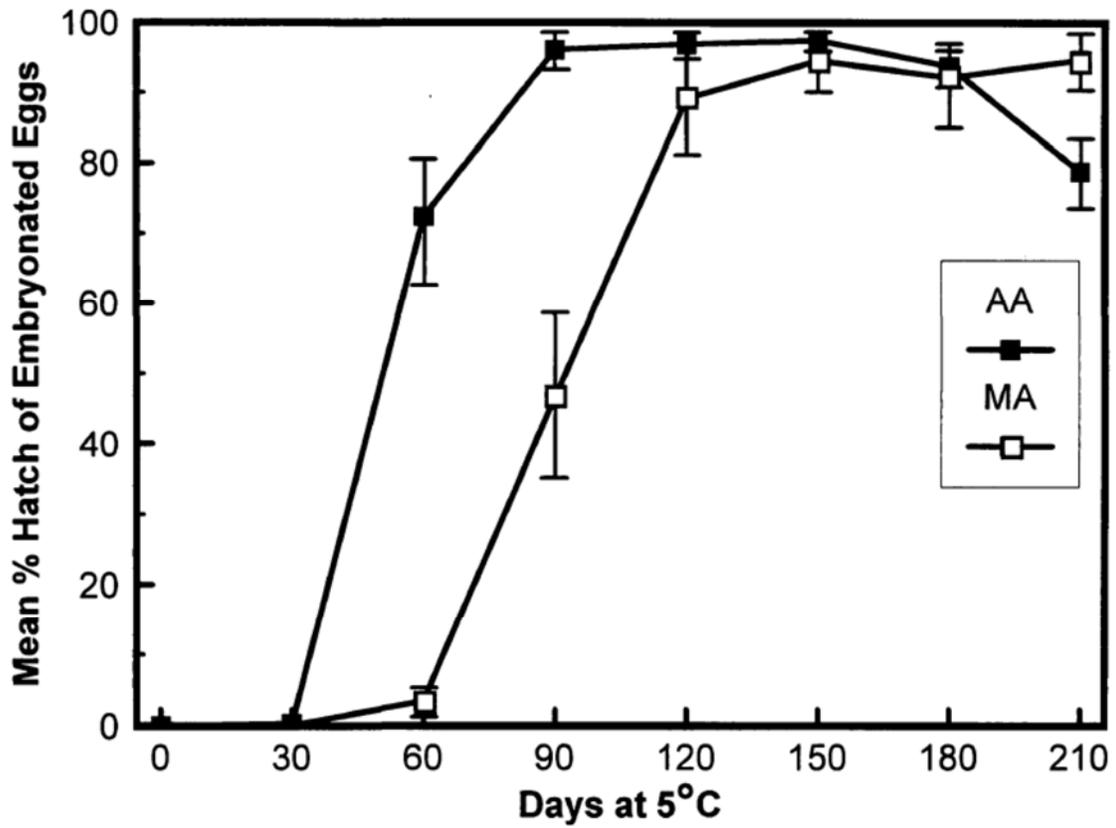


Figure 5-8. The effect of temperature on daily mortality rate at larval stage (data from Casagrande et al. (1987))



**Fig. 2.** Comparison of the effects of different durations at 5°C, followed by incubation at 25°C, on mean percentage of hatch of embryonated eggs for gypsy moth strains from far eastern Russia (AA) and Massachusetts (MA). Bars = 95% CI.

**Figure 5-9.** Mean percentage hatch at different number of days that egg mass had been exposed to temperature at 5°C (from Keena (1996))

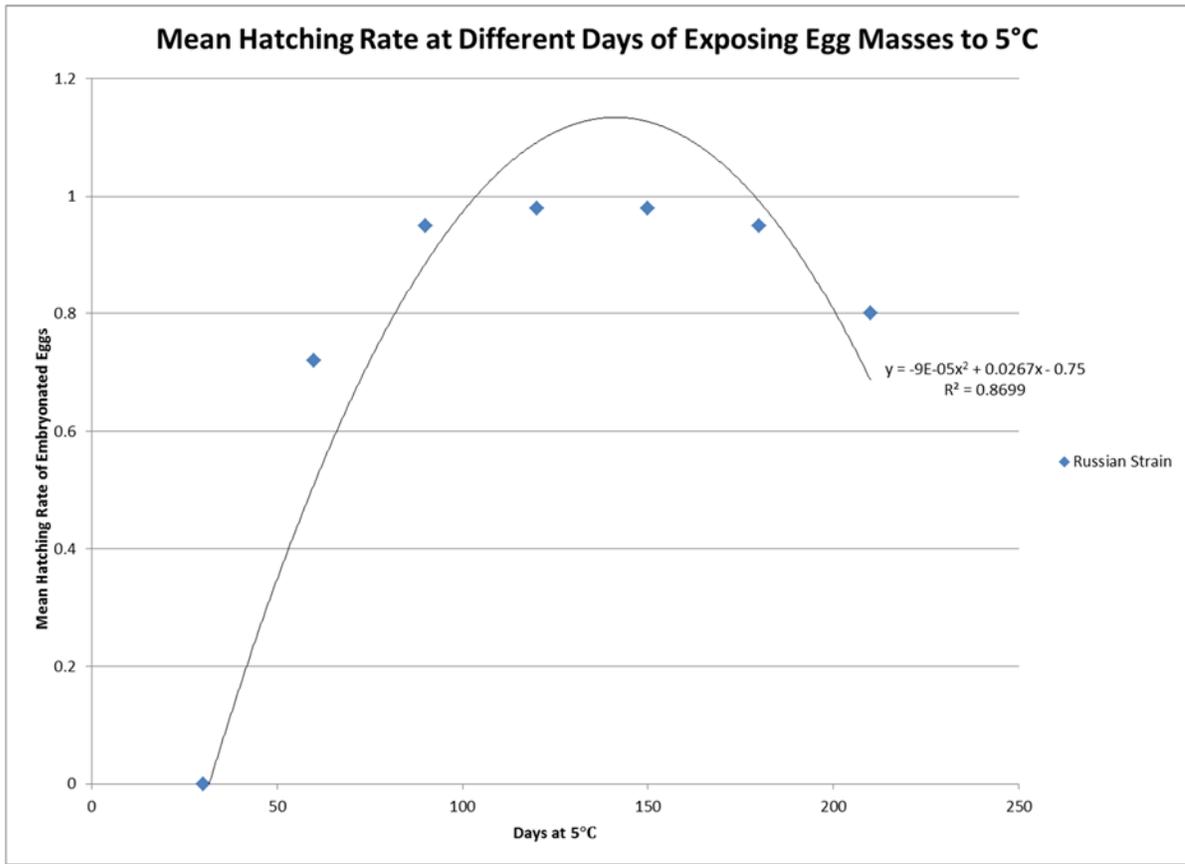


Figure 5-10. Mean hatching rate estimates based on the experiments by Keena (1996)

Using historical records of daily temperature data from 2009 to 2013, GM population fluctuations were estimated for continental United States by using a temperature-driven demographic model (Figure 5-11). For each cell of a 4-km raster dataset the initial population level was set to 10 egg masses. The number of eggs within an egg mass varies with host species and environments; however, in this study, it was set that each egg mass contained 700 eggs. The ratio of female to male within an egg masses was reported to be 65:35 (Campbell 1967, 1981); therefore, there were 455 females within an egg mass in this study. Suitable areas are indicated by increases in population, while unsuitable areas are defined by decreasing populations. The purpose of the simulation was to provide insights into population fluctuations and potential pest impact within the continental United States.

The timing of GM egg eclosion was estimated by using the phenology model described earlier. The stages of GM were estimated based on the growth rate and daily temperature. The number of GM surviving each generation was estimated based on the stage of GM, the mortality rate of GM at a specific stage and observed daily temperature at each grid cell.

I note that I am modeling theoretical populations (AGM is not in the US as of this writing) and the interest is in understanding relative differences and relative impacts. For example, the initial population level is set to 10 egg masses at each cell level throughout the contiguous United States. This does not represent the true population levels in fields, but this model will thus assist us to understand the population fluctuations due to daily temperature conditions.

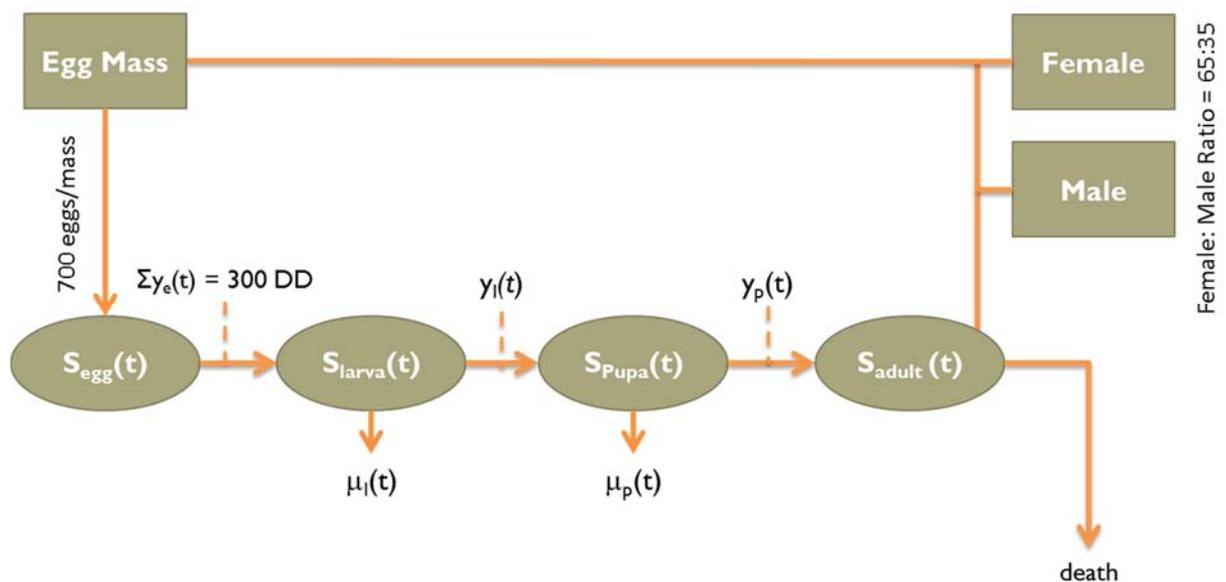


Figure 5-11. Population dynamic model at each life stage of gypsy moth

### 5.3.4 Modeling Human Assisted Movement and spread

The Asian gypsy moth (*Lymantria dispar* L.) has been intercepted in the continental United States via maritime vessels and passengers. Natural transoceanic spread was not

considered a biologically significant likelihood. The entry of AGM into the continental United States is likely through sea ports or airports. This model evaluates AGM at-risk areas based on at-risk ports, traffic volumes, and environmentally suitable areas.

In this study, three at-risk ports were selected based on the AGM detections within the continental United States. Those three at-risk ports were Los Angeles/Long Beach (California), Seattle (Washington), and Charleston (South Carolina). This section of analysis evaluated the traffic volumes and flows from those at-risk ports to identify risk areas in the continental United States.

The destinations were identified by the cities that received more than 1.5 million tons of cargo annually by trucks from those three potentially at-risk ports (considered initial point sources). Since AGM can be transferred for very long distance attached to the surface of a material (e.g., cargo, truck), all commodity types were selected. The traffic volumes at the destinations from three at-risk ports were obtained by using the Freight Analysis Framework (FAF) Data Tabulation Tool at FAF Version 3 website (<http://faf.ornl.gov/fafweb/Extraction1.aspx>).

The Freight Analysis Framework combines multiple data sources to develop a comprehensive dataset of freight movement throughout United States by all modes of transportation. The FAF version 3 provides estimates for tonnage, value, and domestic ton-miles by region of origin and destination (Southworth et al. 2010). The FAF dataset was funded by U.S. Department of Transportation, Federal Highway Administration and developed by the Center for Transportation Analysis, Oak Ridge National Laboratory. This dataset contains traffic information on US Interstates and major highways.

The FAF dataset was processed to record direction specific traffic volumes at major street-level as a stand-alone table for 2001-2012 time periods. This process was done by joint effort of USDA APHIS Plant Protection and Quarantine (PPQ), Center for Plant Health Science and Technology (CPHST), USDA Forest Health Technology Enterprise team, USDA Forest Service, Southern Research Station, and the Canadian Forest Service. The traffic volume table can be joined to ESRI Streetmap v10 dataset to incorporate spatial information.

Once origins and destinations were identified, the most likely routes between origins and destinations by truckers were determined by the distance, the travel time to destination, and the types of highway. The most likely routes were identified by using ESRI Streetmap v10 dataset and ArcGIS Find Route tool. It was considered reasonable and likely that truckers select major roads (e.g., Interstates, US highway); therefore major roads were selected even when the distances were shorter on minor roads. This assumption was confirmed later with survey and transportation data.

The female AGM is known to fly long distances (Reineke and Zebitz 1998). Since AGM can dislodge from cargos while being transported by trucks, I created a 40-kilometer buffer areas along the most likely routes identified earlier to represent potential areas that AGM could fly. The distance that AGM female can travel is still unknown; however, USDA APHIS once found a female AGM approximately 40 kilometers away from Los Angeles/Long Beach port area. The AGM was likely transported via a maritime vessel destined to Los Angeles/Long Beach port; therefore, I assumed that female AGM is capable of flying for 40 kilometers.

The availability of host species within the buffer area was examined. The AGM is a polyphagous species and is known to attack any deciduous trees. The location of the deciduous trees throughout US were identified and multiplied by the forest density to represent as the availability of the host species for AGM. The forest tree species and forest biomass raster datasets were downloaded from USDA Forest Service FSGeodata Clearinghouse website (<http://data.fs.usda.gov/geodata/>) (Figure 5-3; Figure 5-4). Both forest type and forest biomass datasets are at 250-meter resolution with Albers conical equal area projection. The forest biomass data were divided by the biomass value at 96% quantile to represent as forest density. The reason for using a biomass value at the 96% quantile was due to large standard deviation in the data, and the maximum value in the dataset was considered an outlier. The tree species type data were categorized by the GM preference. The deciduous forests received 1 as the GM preference, while hardwood-softwood forests and hardwood forests received 0.5 and 0, respectively. The availability of the GM host species was then determined by multiplying GM preference and density information.

The availability of host species was incorporated with the traffic volumes along the most likely routes to identify the areas that were highly likely to be have introductions and establishment.

### **5.3.5 Asian Gypsy Moth Likelihood of Introduction**

The likelihood of Asian gypsy moth (AGM) introduction into the contiguous United States was determined based on the phenology model, population dynamics trends, and human assisted movements. The phenology model determines the climatologically suitable areas for AGM, while population dynamic trend model with phenology model estimates the

areas that populations of AGM are likely to increase (or decrease) based on climates. The combination of those two models can differentiate between the optimal environment and the borderline environment for AGM growth.

Since the major long distance movement of AGM is by hitchhiking onto vessels, vehicles, and trains, identifying the human assisted movements and trends from the international ports that receive high number of cargos from infested areas/countries (e.g., China, Russia, and Japan) can further identify where AGM could be transported and established. In this study, the at-risk ports were identified based on number/volume of cargo units arriving/received as well as historical AGM detections in the continental United States. Then, traffic patterns from at-risk ports were characterized based on traffic volume and likely routes between at-risk ports and the major cities.

## **5.4 Results**

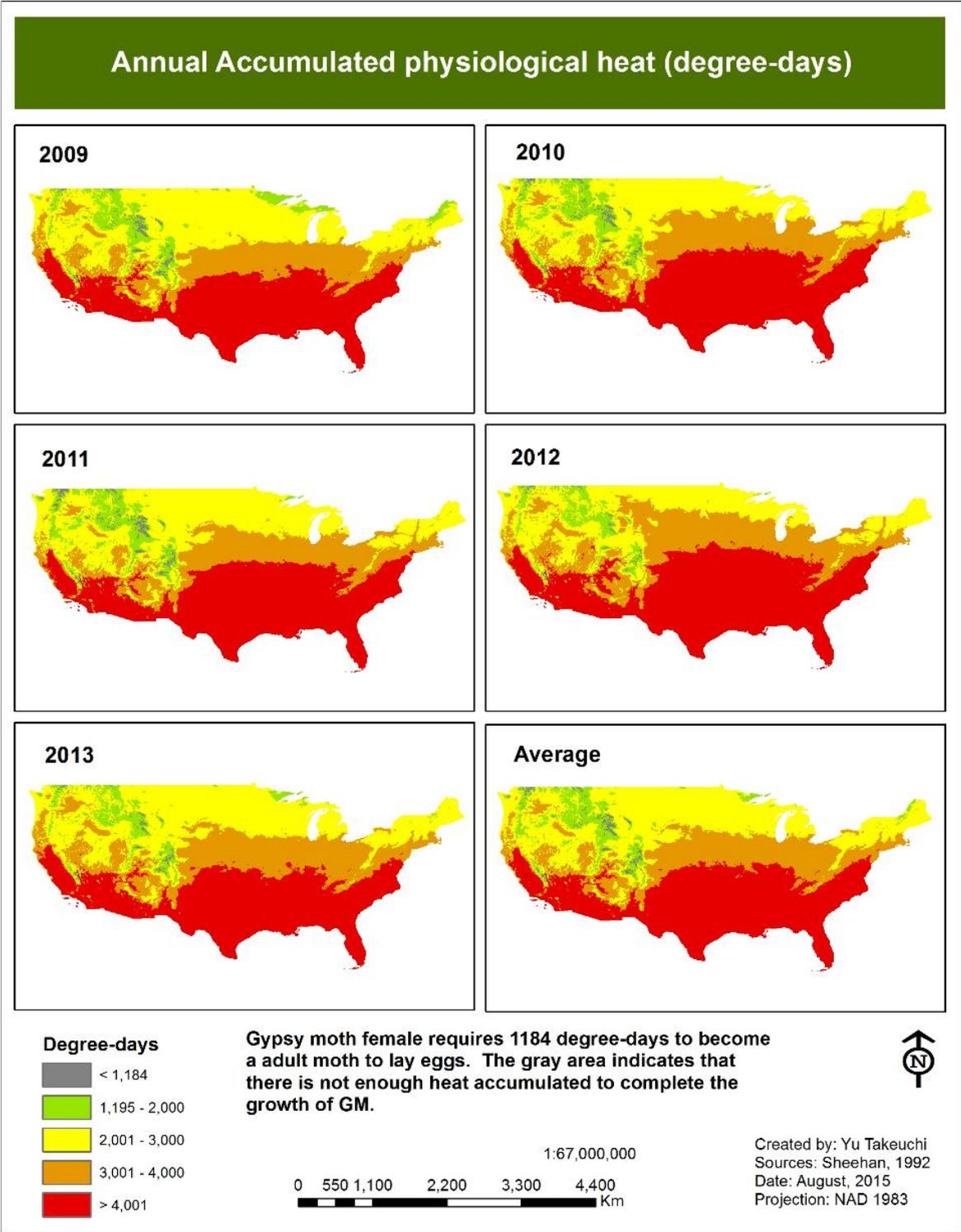
### **5.4.1 Phenology Model**

The phenology of gypsy moth (GM) in the continental United States was estimated based on the daily climatic conditions (Tmin and Tmax) at week interval by using PRISM dataset (PRISM Climate Group 2015). Figure 5-12 shows the accumulated degree-days since January 1 to August 26 to evaluate the climatically suitable areas within the United States where GM is capable of completing a generation. The accumulated degree-days slightly varied with years. In 2012, more degree-days were accumulated in the southeast United States compared to 2010 and 2013. Despite this variation, most areas in the United

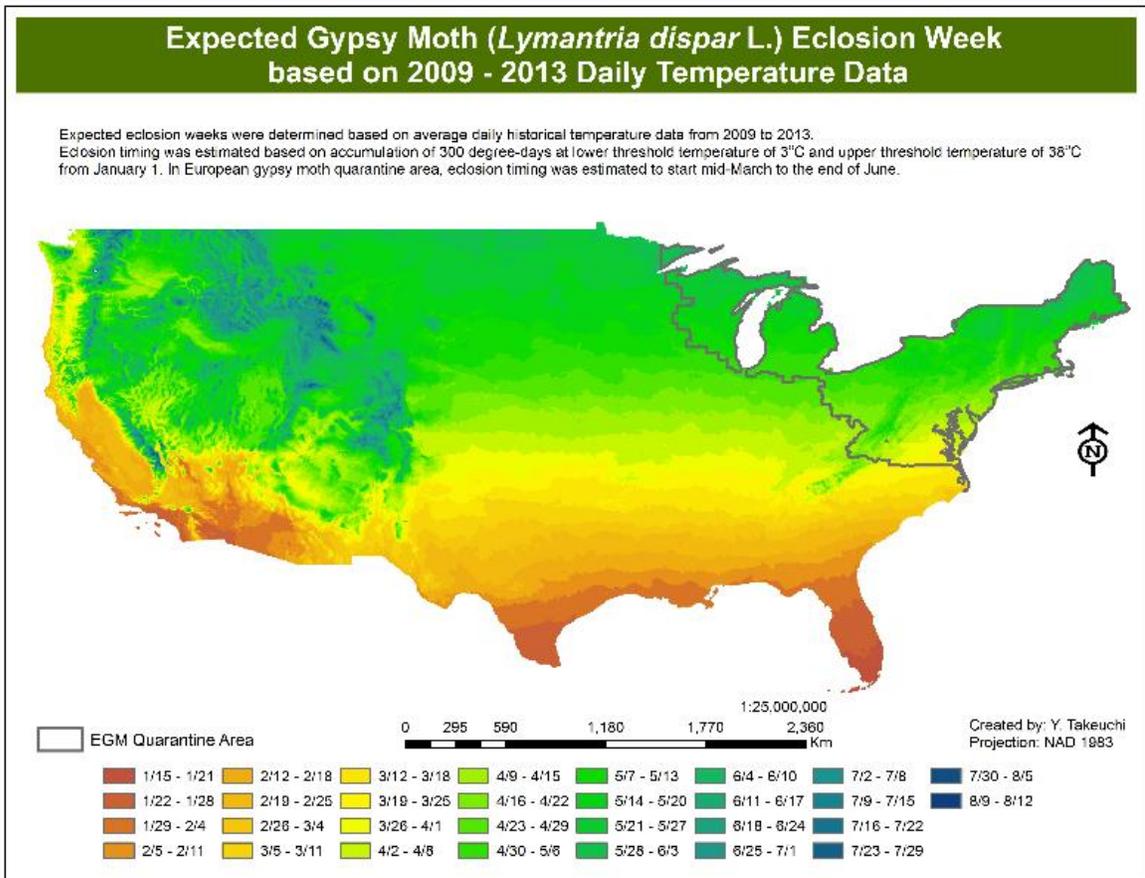
States were climatically suitable for GM and AGM growth. The few areas in the northern Rocky Mountain range were unsuitable every year.

Below maps show the estimated timing of eclosion (Figure 5-13), pupation (Figure 5-14), and oviposition (Figure 5-15) for female GM based on the average daily temperature data from 2009 to 2013. In Pennsylvania, GM start eclosing in late April to early May, start pupation in late June to early July, and start ovipositing in mid-July (Hoover 2000). The phenology model predicted fairly well. Based on the historical daily climate data, the phenology model predicted eclosion in mid-April to mid-May, pupation in late-May to early-July, and oviposition in mid-June to late July for Pennsylvania.

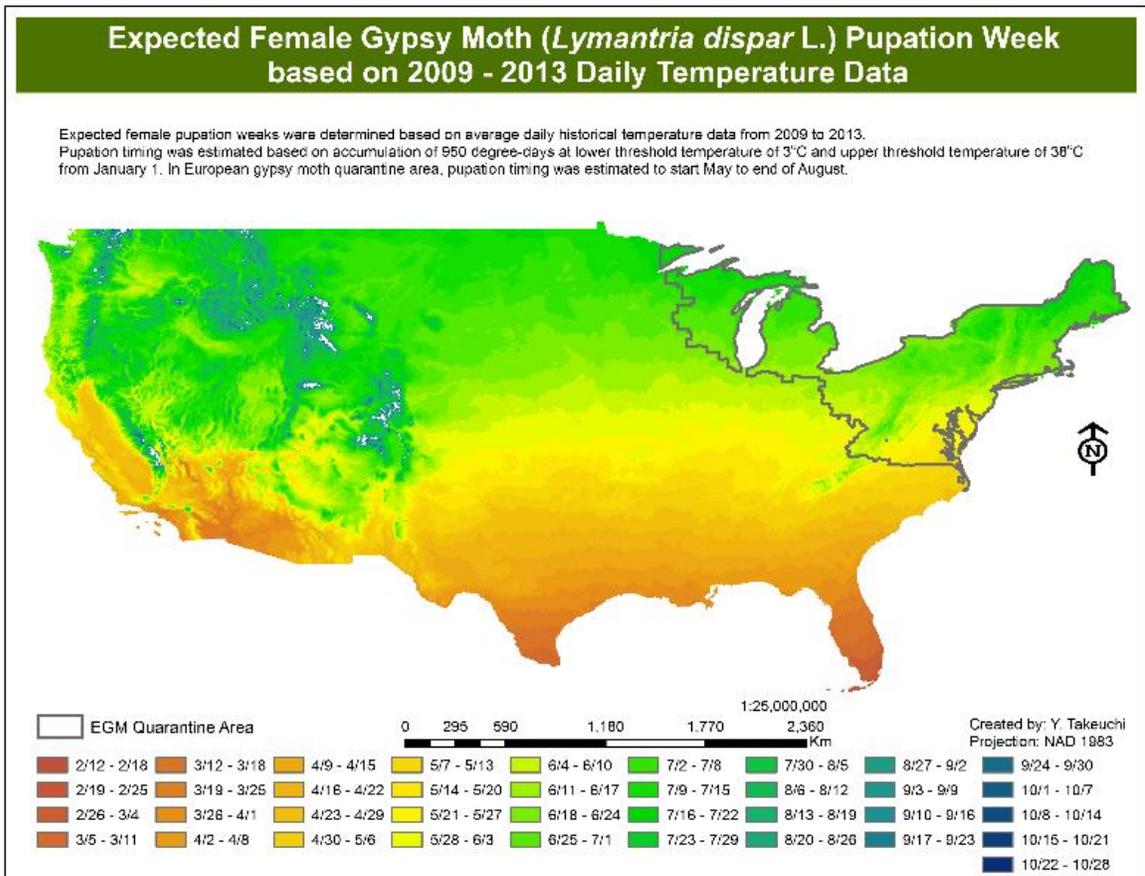
The phenology model also indicated that the most of the areas in the continental United States, except small areas in Rocky Mountain, were climatologically suitable to complete a generation in a year. The host species for GM are distributed in the continental United States, except Great Plains and Great Basin regions (Figure 5-3), indicating that the majority of the continental United States is at risk of GM.



**Figure 5-12. Annual accumulated degree-days at lower and upper threshold temperatures of 3°C and 38°C**



**Figure 5-13. Expected gypsy moth eclosion week based on 2009-2013 historical temperature data**



**Figure 5-14. Expected gypsy moth pupation week based on 2009-2013 historical temperature data**

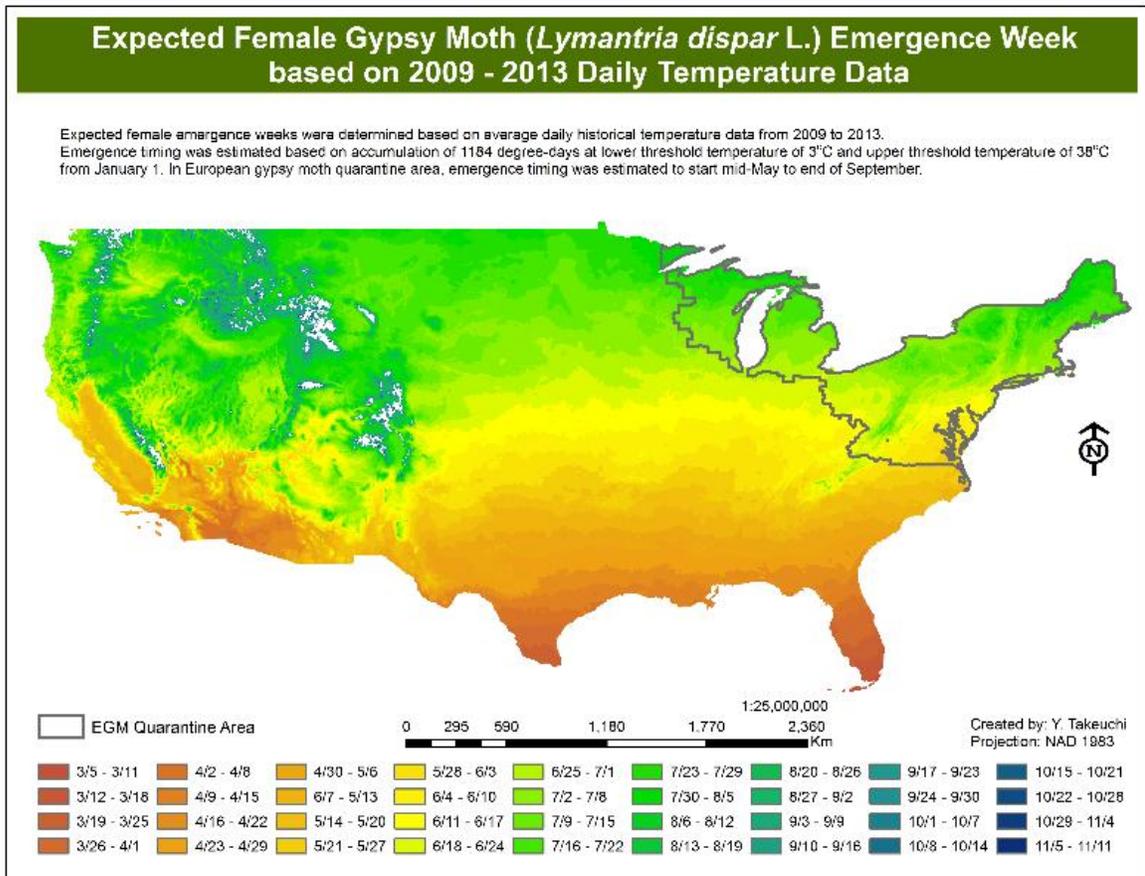


Figure 5-15. Expected gypsy moth emergence week based on 2009-2013 historical temperature data

#### 5.4.2 Population Trends Model

The gypsy moth (GM) population fluctuations within the contiguous United States were examined by using a temperature-driven demographic model. The timing of the GM stages were determined based on the phenology model described in

Table 5-2 for eclosion and the daily growth rates derived from Casagrande et al., (1987) study (Equation 1 and 2) for pupation and adult emergence. The mortality rates for larval stage and pupal stage were calculated based on the daily average temperature according to the each insect stage estimated from the daily growth rates.

The GM stage estimates based on the growth rates were similar to the results from the phenology model; however, the phenology model (degree-day) estimated slightly earlier timing (1 to 4 weeks) than the temperature driven demographic model for majority of the areas in the contiguous United States (Figure 5-15; Figure 5-16; Figure 5-17).

The Year 2009 was colder than usual in northern Michigan and Minnesota; it resulted that there was not enough physiological heat to complete a generation in 2009 (Figure 5-16). However, Year 2010, 2011, and 2012 had enough physiological heat to complete a generation by the end of September (2010 emergence week is shown in Figure 5-17).

The mortality rate during larval stage varied with years; however, it indicated that GM would not survive consistently in western coastal areas (Figure 5-18). The Rocky Mountain region had mortality rate of 100% because those areas did not have enough physiological heat to complete larval stage.

The mortality rate for pupal stage was also factored into this population model. The mortality in this phase was expressed either 100% mortality or 0% mortality based on the average mean temperature during the pupation period either below or above 13.3°C. In addition, if enough physiological heat was not accumulated to complete the pupal stage, then mortality rate was converted to 100% so that there would be no remaining GM

populations in those areas. Each year's pupal mortality rate was combined to indicate the areas that experienced no mortality for the examined years and the areas that had 100% mortality every year. The Figure 5-19 indicates the number of years that pupal mortality was estimated based on climate data from 2009 to 2013. The light gray area indicates that there was no mortality during pupal stage estimated for those 5 years. The red area indicates that pupal stage mortality was resulted every year during those 5 years. The mortality rate induced by temperature during pupation stage indicated that Rocky Mountain region, Sierra Nevada Mountain region and northwestern coastal area were not suitable for GM growth. In some years, northern Minnesota, northern Wisconsin, and northern Michigan were not suitable for pupal growth. Except north east of Minnesota, those experienced only one or two years of unsuitable climate. Another area that was potentially unsuitable for pupal growth was in northern Vermont, northern New Hampshire, and northern Maine. Northern Maine results indicated that it was not suitable for pupal growth almost every year.

Egg hatching rate was estimated based on egg masses being exposed to enough cold temperature during winter (Figure 5-20). The hatching rate was indicated as 0 if a GM generation was not completed in the previous year or if enough coldness (at least 30 days at or below 5°C) was not accumulated. The hatching rate varies each year; however, most of Northeast, northern South, Central states consistently had good winter conditions for GM egg masses. Florida and southern Texas showed unsuitable conditions due to warm climate during winter months. The Rocky Mountain region was also unsuitable for every year. However, this was due to GM not being able to complete a generation in previous year. The

same reason applies to the results in northern East North Central for 2010. In 2009, those areas did not have enough physiological heat to complete a generation.

The temperature driven demographic model resulted that most Eastern United States were suitable for GM growth (Figure 5-21). The unsuitable areas were Rocky Mountain region, northwest coastal area, northern East North Central, northern New England, Florida, Gulf coast area, and south Texas. I did not express the results with GM population level because this model only considered the mortalities due to temperature. In reality, the majorities of the GM mortality come from host species, parasites, predators, natural conditions (other than temperature: e.g. wind), and diseases (Campbell 1976, Campbell et al. 1977, Hamilton and Lechowicz 1991). Campbell et al. (1977) reported that there was more than 99% mortality in Glenville, NY due to parasites, predators, diseases, and some other unknown reasons. Therefore, the population levels resulted from this model are quite overestimated due to not considering major mortality sources.

The areas that support fast growth were identified to be south of New England, Mid Atlantic, Appalachian Highlands, Midwest, east of Heartland area, and Utah. Utah also had unsuitable areas due to having cold climate at higher elevations. The Figure 5-21 indicates the relative growth. Lower relative growth rates do not mean that the areas were not suitable. Any value other than 0 supports GM growth, but it was relatively low compared to the maximum population level simulated from this model.

The temperature driven demographic model was still useful to identify the suitable areas for GM growth. It has identified the additional unfavorable areas than the phenology model and also indicated sources of GM mortality. The phenology model only identified

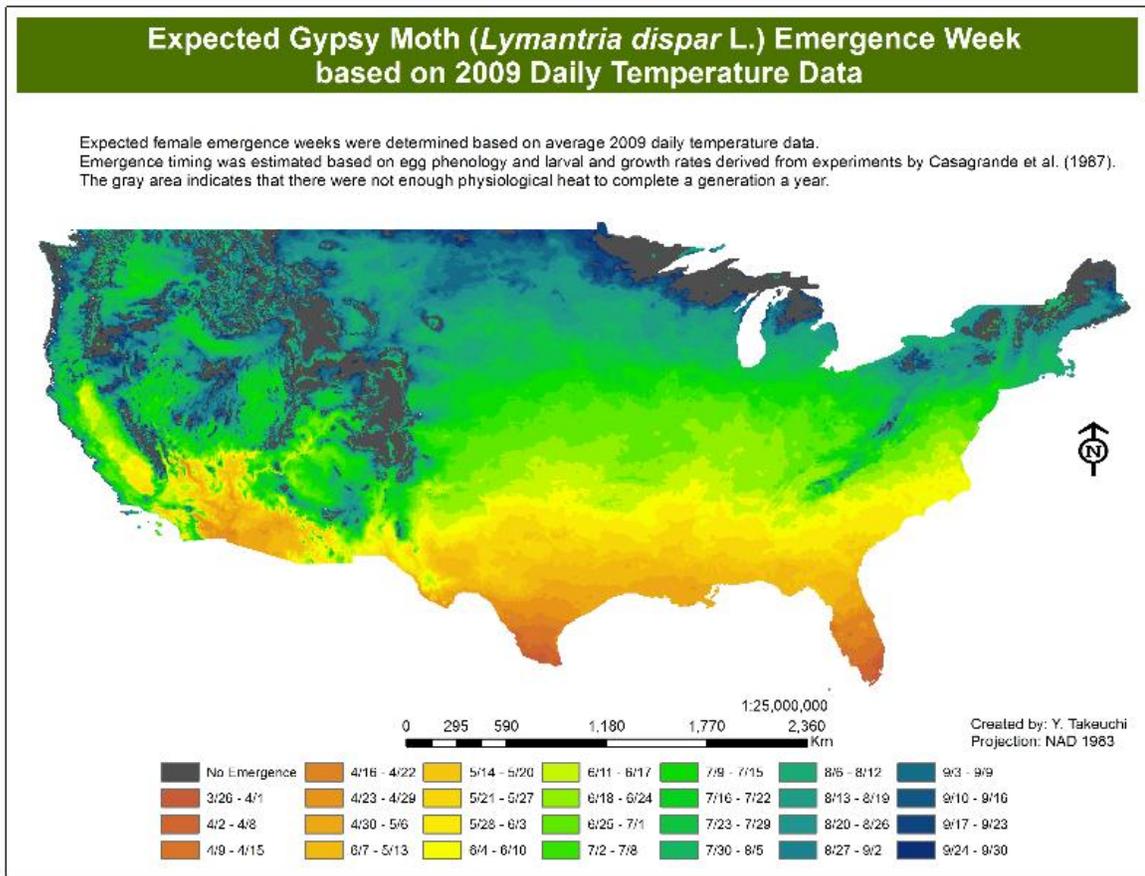
that small area in Rocky Mountain was unsuitable due to cold temperature. On the other hand, this demographic model showed that Florida, Gulf coast, and south of Texas were not suitable for GM growth because it was too hot for egg masses to successfully hatch in following spring. This result coincided with the results by Allen et al. (1993). Allen et al. (1993) reported that GM was unlikely to be established in areas south of Gainesville (29°41'North). The demographic model also showed that northwest coastal area was unsuitable. This was mainly because the mortality rate at larval stage was high. In northwest coastal area, the mild summer temperature resulted in longer days to reach to pupal stage. In Seattle area, the eclosion was estimated at the beginning of April, while the pupation was forecasted in mid to late August. On the other hand, in Harrisburg, PA, eclosion and pupation were forecasted in mid to late April and mid-June, respectively. Having longer time to develop larvae and pupae resulted in accumulation of higher mortality rates in northwest coastal area.

The northern New England was not suitable for GM growth due to colder temperature. But the suitability for GM growth in this region varied with years. It had suitable conditions for some years; however, it was unsuitable as an overall result. This was because the GM population had already died out during an unsuitable year; therefore, there was no population to increase when the weather was more suitable in later years. GM is not widely distributed in north Maine (Figure 5-22). This may be because the environment is not suitable for GM every year in northern Maine area. Even if the population of GM increases under favorable condition for some years, population decreases due to some

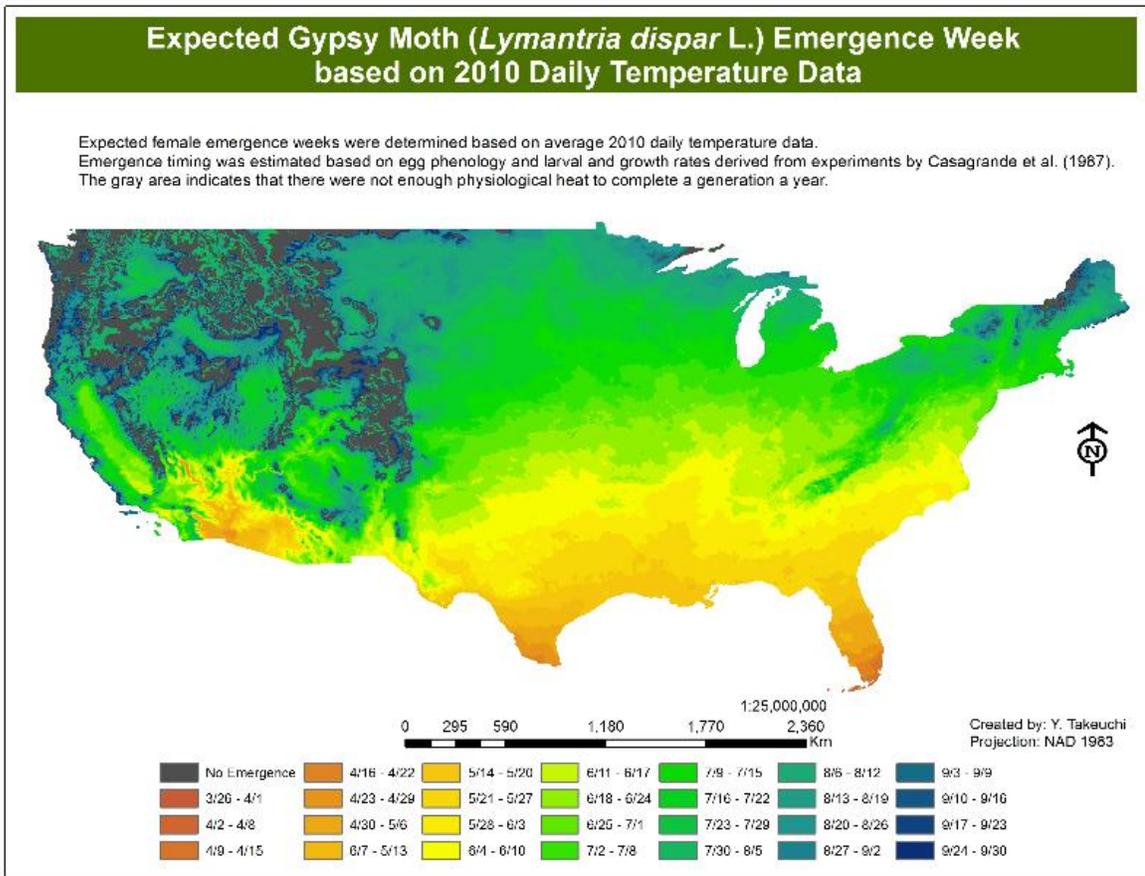
extreme cold temperatures for some years. Therefore, GM has not been permanently established in northern Maine.

The northern Michigan, northern Wisconsin, and northern Minnesota regions also showed similar results. The entire state of Michigan is currently under quarantine.

Although the quarantine areas in Wisconsin has been expanding in last 10 years, GM has a limited distribution in State of Wisconsin.



**Figure 5-16. Expected emergence week based on 2009 historical climate data and GM biology parameters derived from experiments by Casagrande et al. (1987)**



**Figure 5-17. Expected emergence week based on 2010 historical climate data and GM biology parameters derived from experiments by Casagrande et al. (1987)**

## Temperature Induced Mortality during Larval Stage of Gypsy Moth

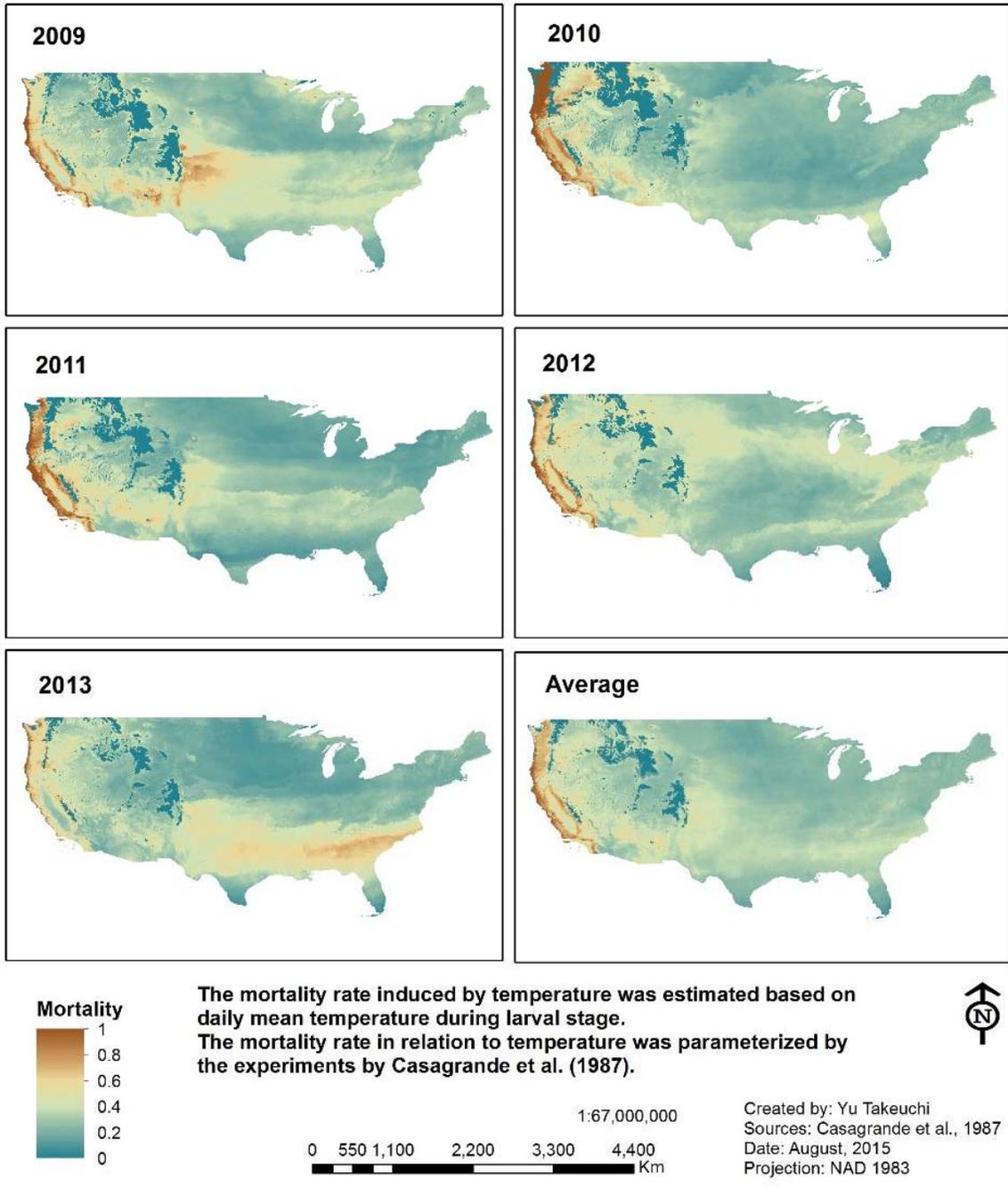
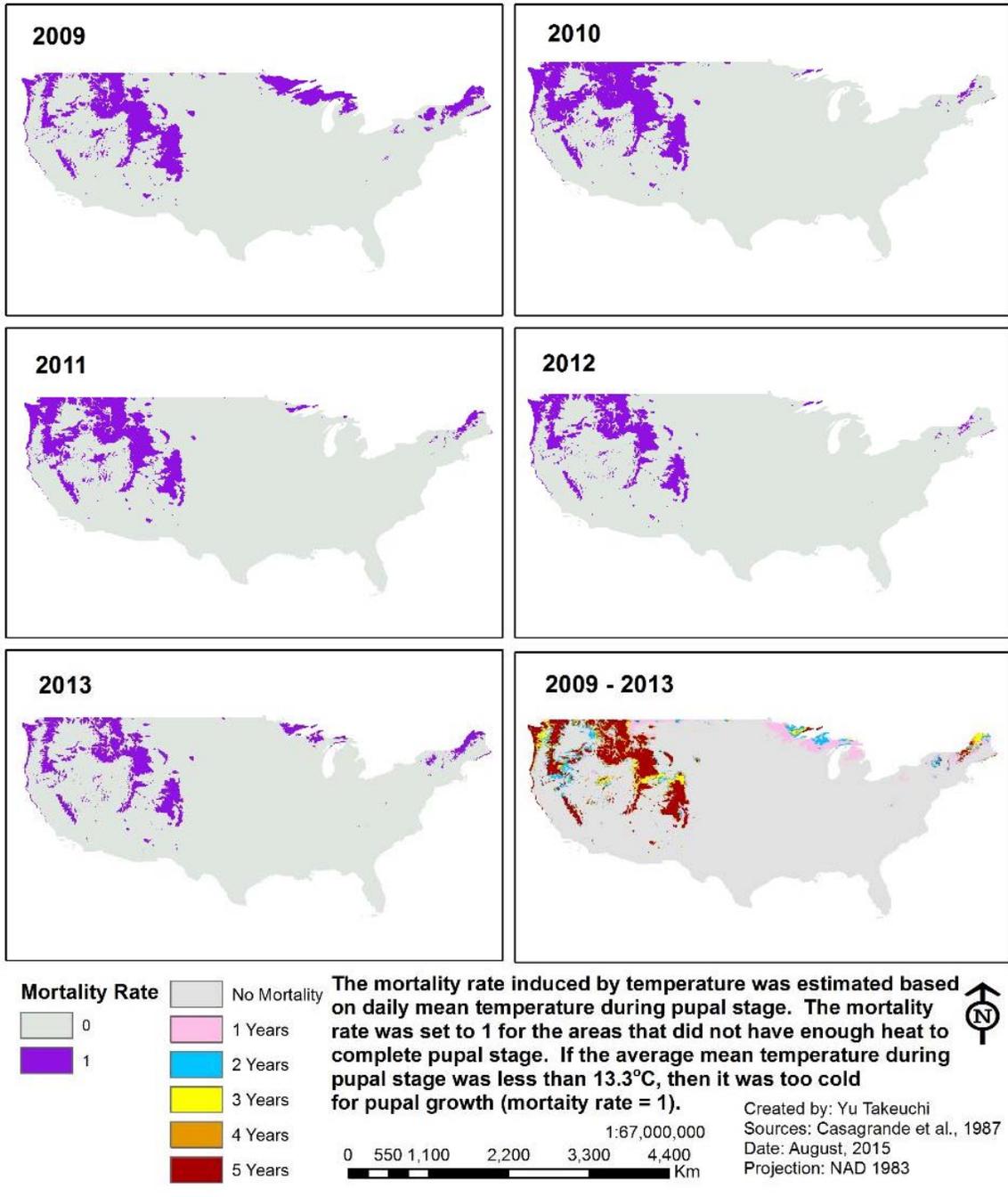


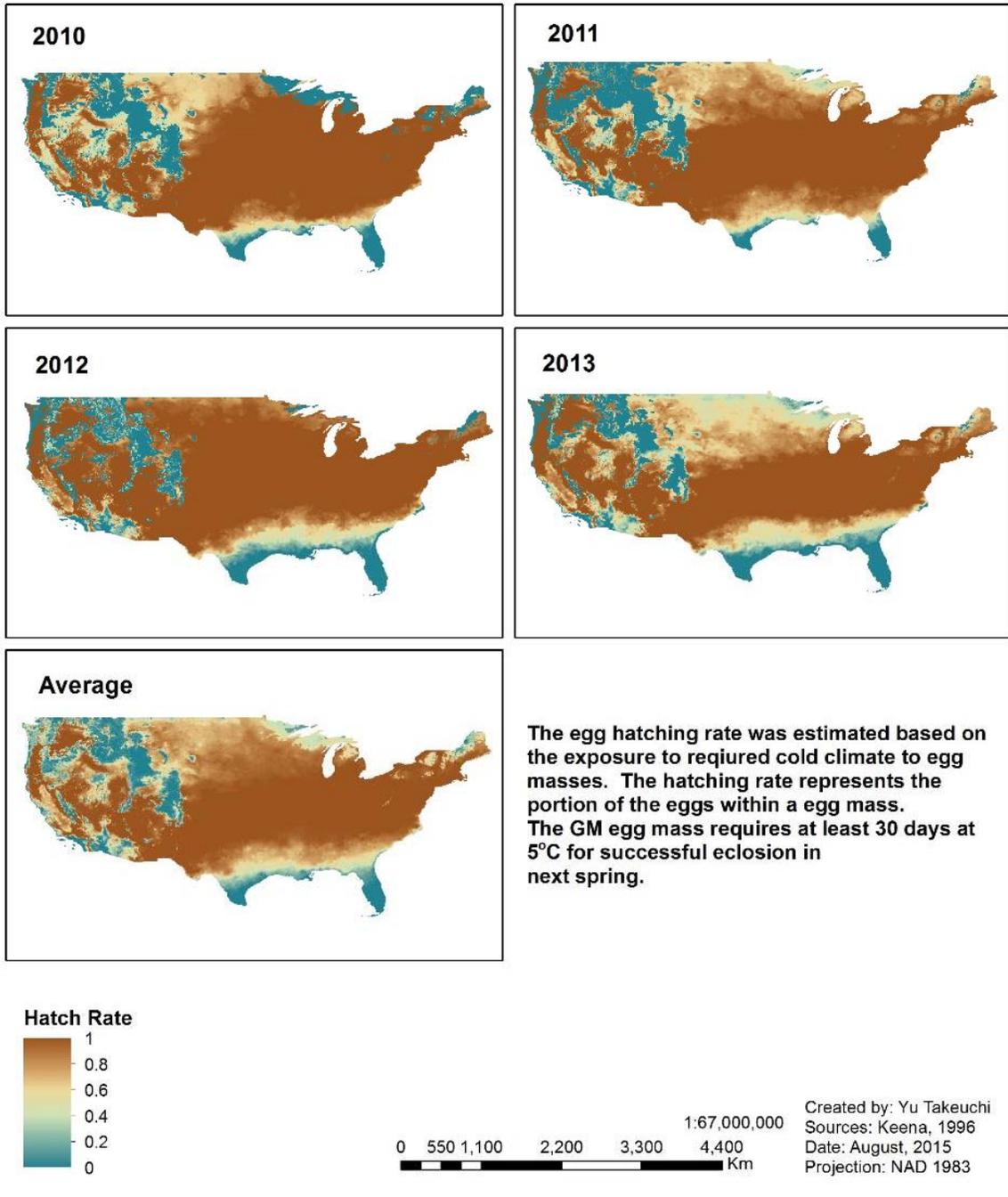
Figure 5-18. Temperature induced mortality rate during larval stage

## Temperature Induced Mortality during Pupal Stage of Gypsy Moth



**Figure 5-19. Temperature induced mortality rate during pupal stage**

## Egg Hatching Rate Estimates by Exposure to Cold Climate



**Figure 5-20. Egg hatching rate predicted by exposure to cold climate conditions**

## Gypsy Moth Suitability Expressed in the Relative Population Growth

The population growth rates expressed with relative growth. The relative growth was determined by population levels divided by the largest population level resulted from this analysis. The rate was determined by simulating GM populations with larval and pupal growth rates, larval and pupal mortality rates, and hatching rate from 2009 to 2013. The initial population was set to 10 egg masses at 4-kilometer resolution cells throughout contiguous United States.

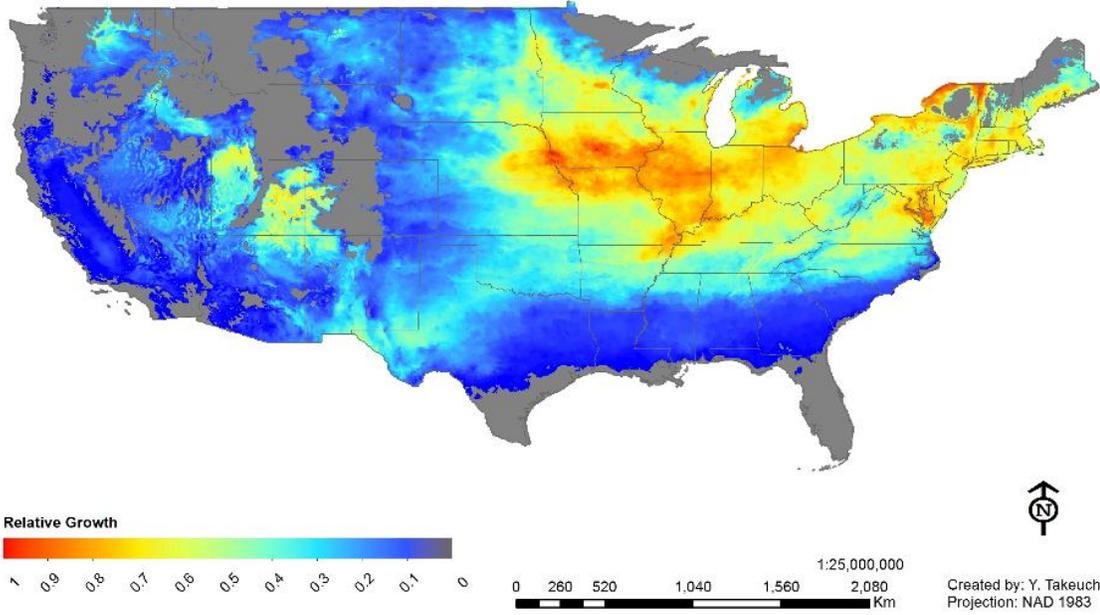
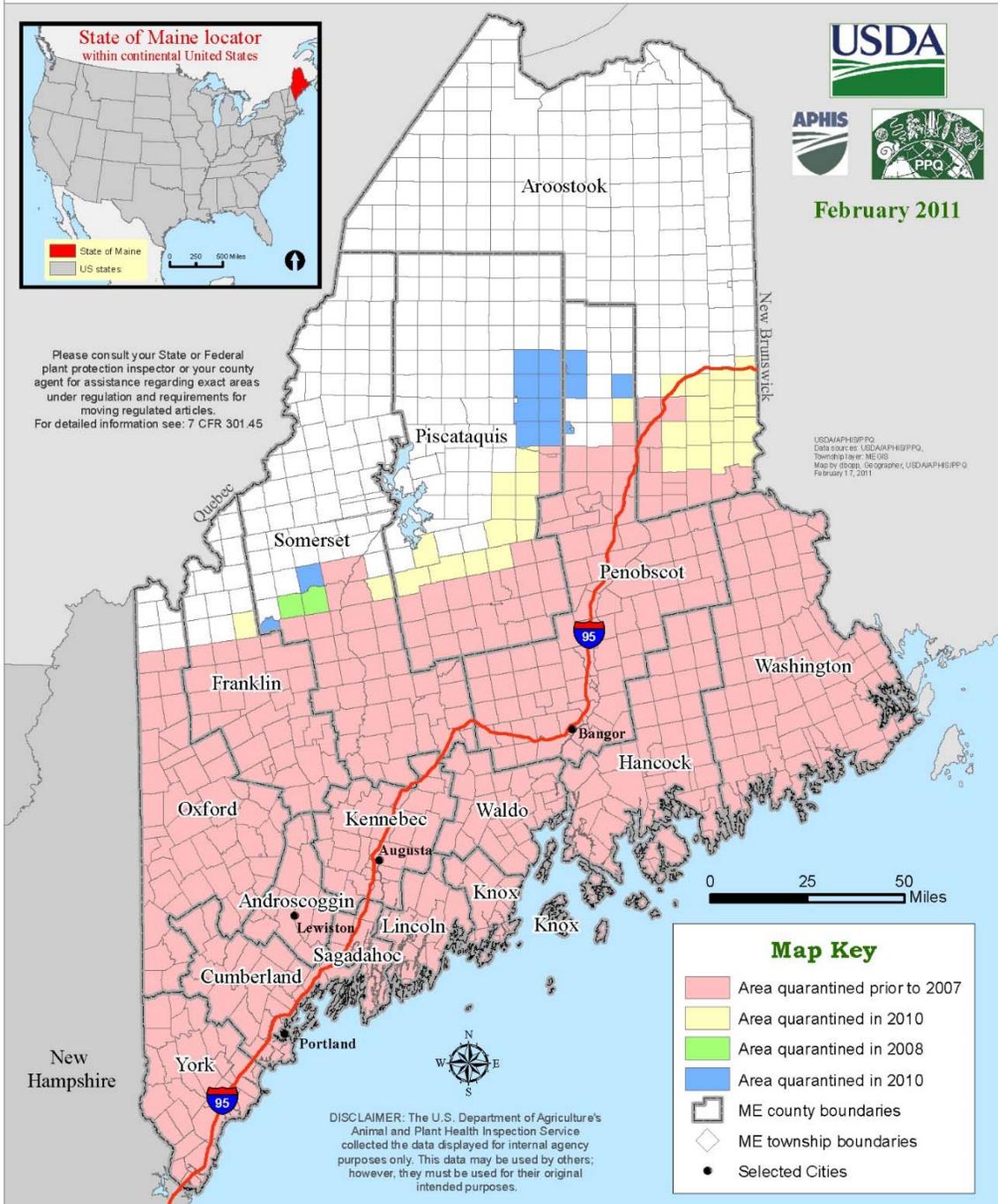


Figure 5-21. Gypsy moth suitability expressed with relative population growth

# European Gypsy Moth (*Lymantria dispar*) quarantine area State of Maine



**Figure 5-22. European gypsy moth quarantine area in Maine (map created by USDA APHIS PPQ (2011))**

### 5.4.3 Human Assisted Movement Model

The three at-risk ports (Los Angeles/Long Beach, Seattle, and Charleston) were selected based on the Asian Gypsy Moth (AGM: *Lymantria dispar asiatica* Vnukovskij) historical detections in the continental United States (Fowler et al. 2009). The most frequent detection area was Seattle area although the last detection in Seattle area was in 2006. Los Angeles area also had multiple detections since 2003, and the last detection was in 2007. Charleston area had two detections. One was in 1994, and the other was in 2014.

The 2007 U.S. Commodity Flow Surveys (CFS) tables are based on FAF analysis zones, that contiguous United States are divided into 123 regions (FAF flow analysis regions) (Southworth et al. 2010). The commodity volume flows for 2007 by trucks from at-risk ports to FAF flow analysis regions are summarized in Figure 5-23.

A large quantity of commodities shipped out from Los Angeles area remained in the state of California; however, large quantity of goods from Los Angeles was also transshipped throughout the continental United States. The cities/areas that received more than 1.5 million tons of commodities by truck movement from the Los Angeles area were other locations in California, Washington State, Arizona except the Tucson/Nogales area, Portland, New Jersey, Las Vegas, Salt Lake City, Houston, Dallas, Chicago, Detroit, Memphis, Atlanta, and New York cities.

The commodities shipped from Seattle were mainly sent to the West coast although a majority remained in the state of Washington. In addition to West coast states, some commodities were sent to Idaho, Minneapolis/St. Paul and Chicago from Seattle; however, the annual commodity volumes transported to those areas were less than 1.5 million tons.

The cities and areas received greater than 1.5 million tons of commodity annually from Seattle by truck movement were state of Washington, Portland, and Los Angeles.

Compared to commodity movement from Los Angeles area and Seattle, the commodity movement from Charleston, South Carolina was localized. The commodities received and sourced from Charleston were shipped out to other locations in South Carolina, plus the states of Georgia and North Carolina. The cities and areas that received greater than 1.5 million tons of commodities annually from Charleston by trucks were mainly within South Carolina, locations in Georgia but excluding Atlanta and Savannah areas, and Charlotte, North Carolina.

The most likely routes from origins to destinations were identified (Figure 5-24). When the entire states were identified as destinations, major cities within the states were selected as destinations to determine the most likely routes. The major cities selected for Washington State were Seattle, Spokane, Tacoma, Vancouver, and Bellingham. The major cities selected for state of California were San Diego, Fresno, Sacramento, San Francisco, and Los Angeles. The large cities selected for South Carolina were Columbia, Charleston, Greenville, Spartanburg, and Hilton Head Island. The large city selected for Georgia to represent receiving large amount of commodities from Charleston, South Carolina was Athens. The routes between Los Angeles and Seattle were selected in both directions since large volumes of commodities were shipped from Los Angeles to Seattle and from Seattle to Los Angeles. Associated traffic volumes with selected roads were scaled to between 0 and 1 by dividing the traffic volumes by the largest traffic volume in one road segment in the

dataset (Figure 5-25). The large volumes of traffic were found around cities; however, there were still large amount of traffic found along highways in the United States.

Along the most likely routes between at-risk ports and destinations, 40-kilometer buffer areas were created to examine the availability of host species within the areas that AGM could easily fly (Figure 5-26). The availability of the host species were calculated by forest type and forest biomass. Although the large amount of forest biomass was found in western coastal states (Washington, Oregon, and California) (Figure 5-4), the AGM host species were mainly available throughout eastern states, especially Appalachian mountain range. The forest species that were commonly found in western coastal states are hardwood species; therefore, it was not preferred species for AGM.

The relative likelihood of AGM introduction within the contiguous United States from Los Angeles/Long Beach, Seattle, and Charleston by truck movements were identified based on the availability of AGM host species and traffic volume along the most likely routes from those three cities to common destination areas (Figure 5-27). The heavy traffic volumes were found both directions between Seattle and Los Angeles; however, very few areas resulted as high risk area between Seattle and Los Angeles due to AGM preferred host availability. The high risk cities were Oklahoma City (OK), Pittsburg (PA), Salt Lake City (UT), Jackson (MS), Birmingham (AL), Atlanta (GA), Greenville (SC), Charlotte (NC), St. Louis (MO), and Columbus (OH). The risk areas around Oklahoma City were quite large, extending far out of Oklahoma City areas along I-40. Some risk areas were found between Abilene, TX to Jackson, MS along I-20, while majority of the areas between Jackson, MS to

Atlanta, GA along I-20 was at high risk. Also, the areas between Columbus, OH to New York City, NY were at high risk.

This result coincides with recent detections of AGM in Oklahoma City, OK. USDA APHIS detected AGMs in near McAlester, OK and near Edmond, OK in 2013 and again in near McAlester, OK in 2014. Since Oklahoma City does not have a maritime port, AGM must have traveled via trucks or trains by remaining on the cargo from Asian countries. Incorporating human activities assisted to precisely identify the likelihood of AGM introductions within the contiguous United States.



The cities that received high volume of commodities shipped by trucks were identified. In this example, cities receiving more than 1.5 million tons of cargo annually were selected.

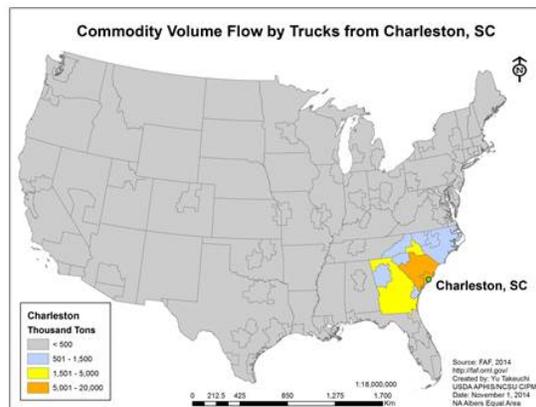
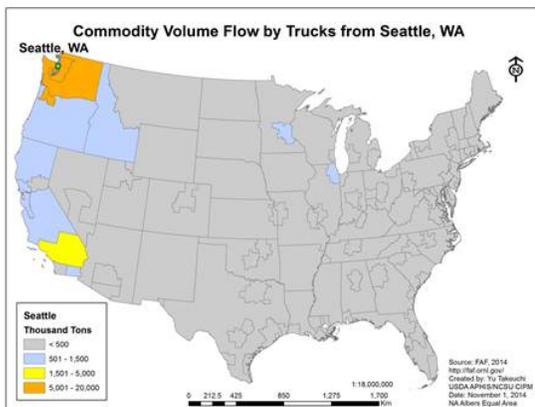


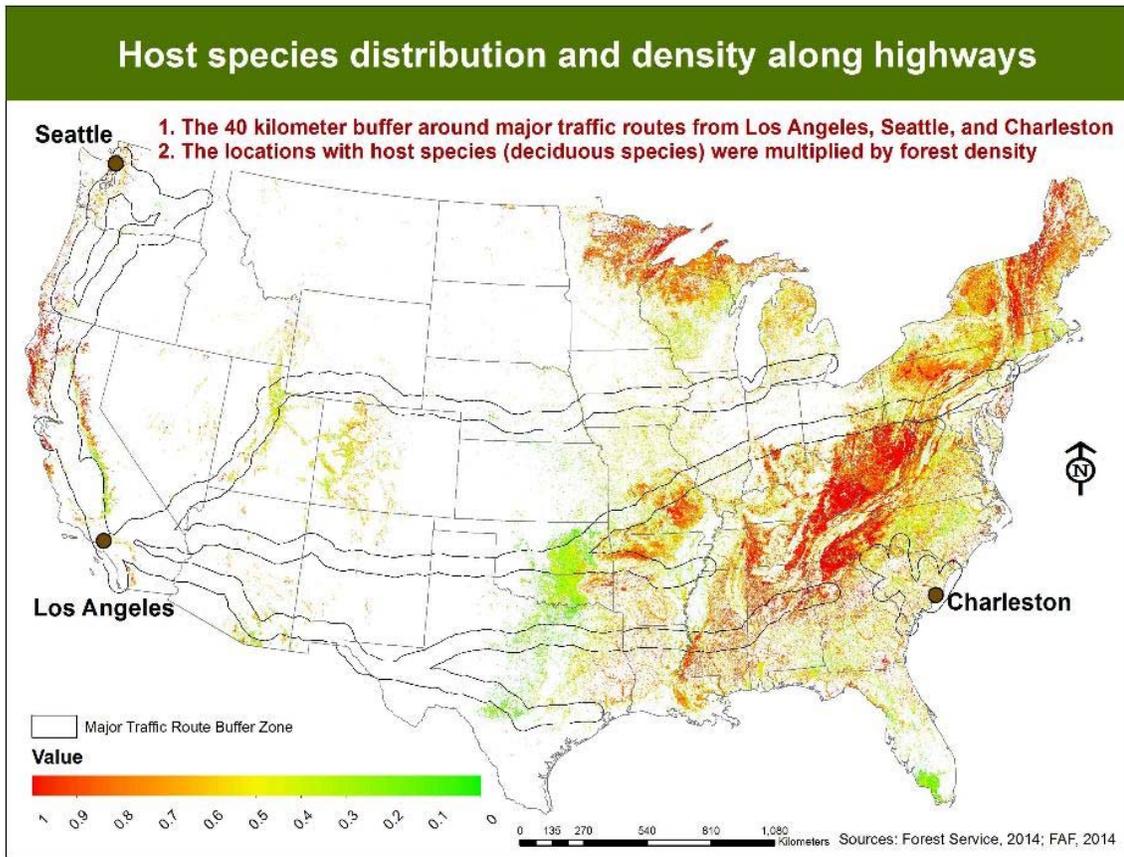
Figure 5-23. The commodity volume flows by trucks from three AGM at-risk ports (Los Angeles, Seattle, and Charleston)



Figure 5-24. Most likely routes from Seattle, Los Angeles, and Charleston to common destinations



Figure 5-25. The relative traffic volume expressed between 0 and 1 along the most likely routes from at-risk ports



**Figure 5-26. The host species availability determined by tree species types and density along most likely routes**

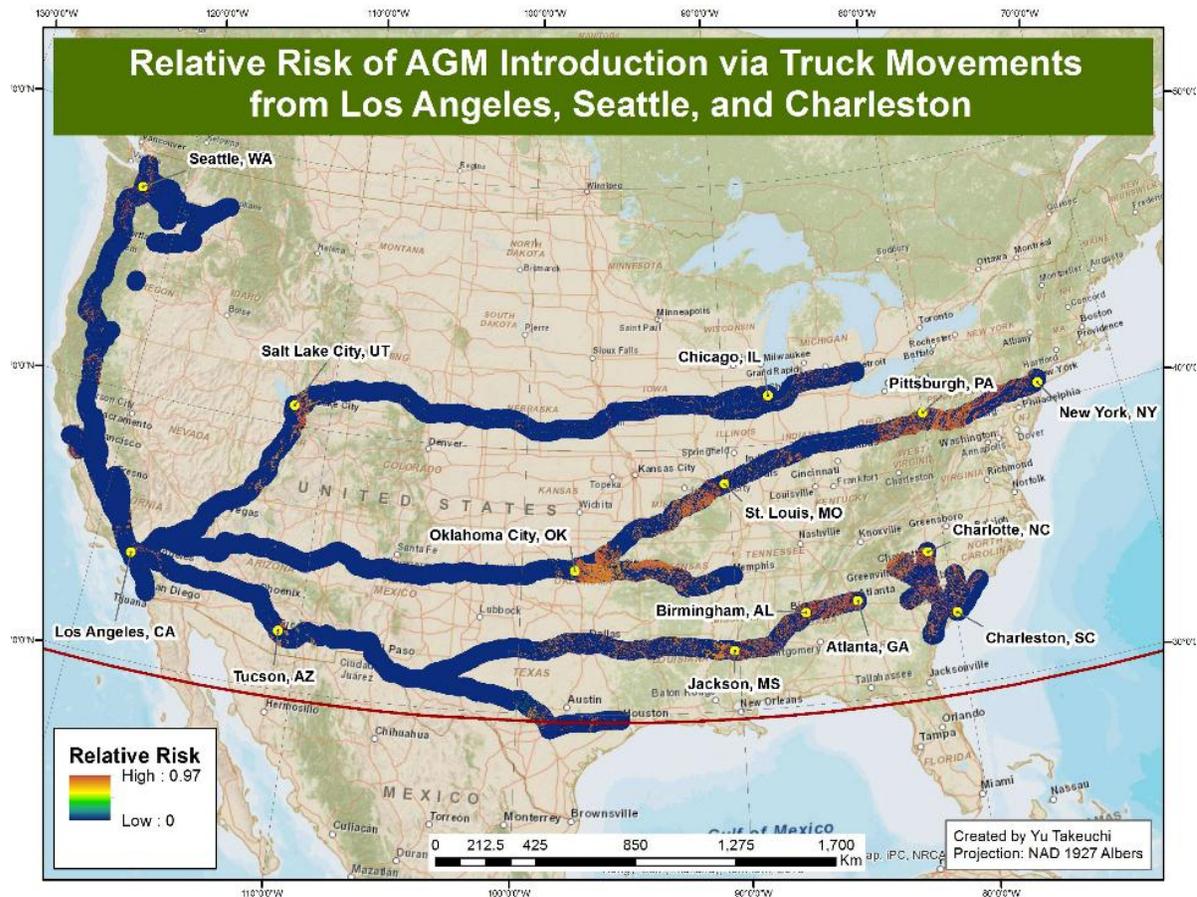


Figure 5-27. The relative likelihood of AGM introduction within the contiguous United States from Los Angeles, Seattle, and Charleston by truck movements

#### 5.4.4 Asian Gypsy Moth Likelihood of Introduction

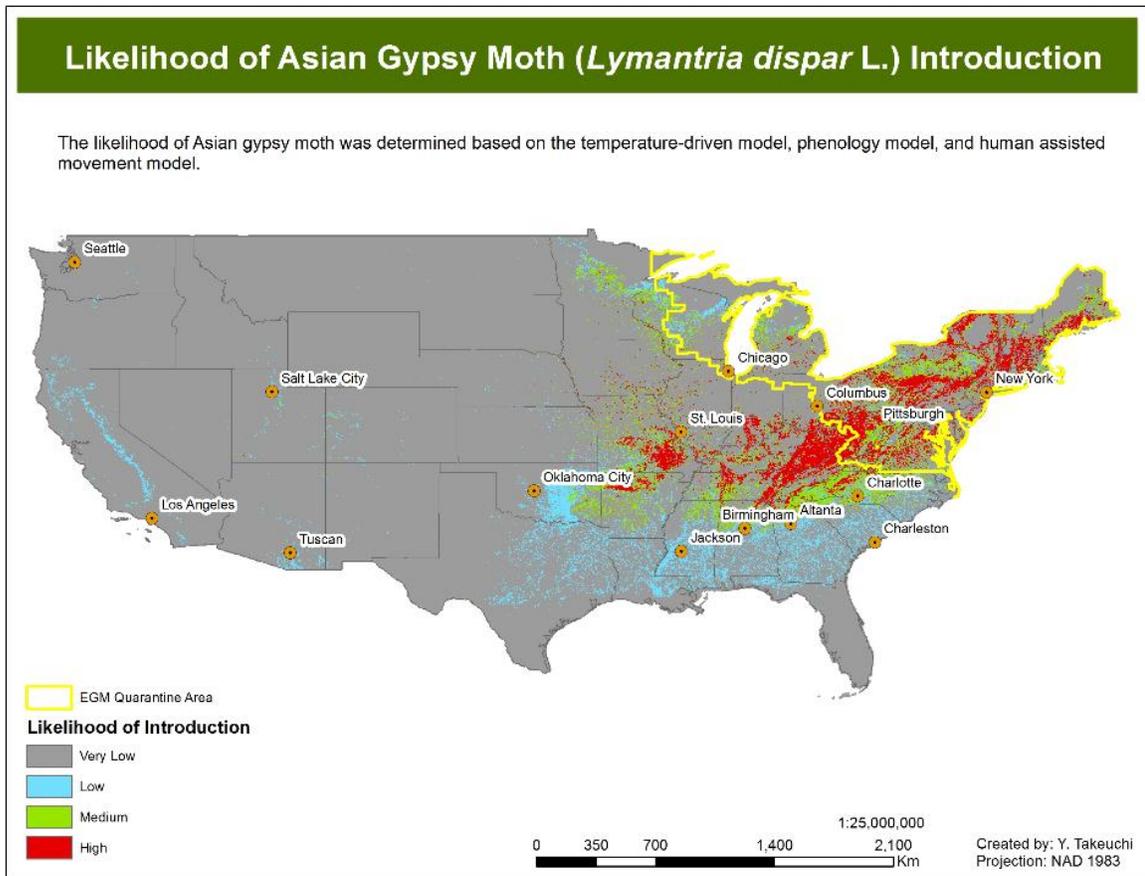
The likelihood of Asian gypsy moth (AGM) introduction was determined by combining results from the phenology model, temperature-driven demographic model, and human assisted movement model (Figure 5-28).

The majority of the contiguous United States could support GM growth climatically. However, the preferred host species are mainly found in Eastern United States. There are some preferred hosts in California, Colorado, Utah, Arizona, and western coastal areas. There are almost no host species available in Great Plain region.

The overall results showed that northeastern United States had high likelihood of AGM introduction. The northeastern cities like New York, Columbus (OH), Pittsburgh, St. Louise, and Chicago receive large quantity of commodities from Los Angeles/Long Beach area, indicating that AGM could be easily transported with human assisted movements. In addition, the host distributions and climatic conditions support AGM growth around those cities.

The Northeast and northern part of Upper Midwest regions are currently under quarantine areas for EGM. Although there are some differences between AGM and EGM (Humble and Stewart 1994, Keena 1996), they have similar life cycles and biology requirements. AGM is likely to successfully be established in those areas.

The areas that were currently out of quarantine areas but resulted high likelihood of introduction were Kentucky, Tennessee, and Missouri. The southeastern United States did not result as high as northeastern United States, however, still could support AGM introduction and growth in those regions. The detections in Oklahoma City are the good examples. The Oklahoma City surrounding area had relatively low likelihood of introduction compared to other areas. However, AGM were successfully transported alive and able to reside on roadside.



**Figure 5-28. Likelihood of Asian gypsy moth introduction based on phenology model, temperature driven demographic model, and human assisted movement model.**

## 5.5 Discussion

The phenology model is often used to estimate the timing of the insect stages and potential number of generations per year. It is also used to determine the potential insect establishment areas based on the accumulated physiological heat required for the insect development (Fowler et al. 2009). The phenology model in this study resulted that majority of the contiguous United States were suitable for GM establishment.

Conceptually, the phenology model can be used to identify suitable areas for insect growth. If the area has accumulated enough physiological heat for one complete generation

in certain time period, the area is climatically suitable for the insect. However, the phenology model assumes the linear growth rate over temperature between lower and upper threshold temperatures when the growth rates are not at constant rate at extreme temperatures for the insect. Also, the mortality rates are high at extreme temperatures, and the phenology model does not consider anything about the mortality at specific temperature range. It is still useful to estimate the timings of the insect stages, however, should not be used to identify suitable areas for growth.

The temperature driven demographic model developed from the biological data constructed a phenology model indicated some additional unsuitable areas within the contiguous United States and also showed the favorable areas by increasing population levels based on 5-year historical climatic conditions. Although the temperature driven demographic model was constructed with the same biological data for the phenology model, it provided more information and realistic estimates where GM could be established in the contiguous United States. It still requires some fine tuning. This model is very sensitive to one year of unusual climatic condition. If one particular year has unsuitable condition, the population becomes zero. There is no population to be increased even when the climate condition is more suitable in following years. In reality, the population might become very low when GM experiences unsuitable conditions; however, it is unlikely to completely eliminate the population with only one-year of unsuitable condition.

Another major function needs to be added into the demographic model is host species effects. This model was parameterized based on the average data to represent GM behavior the same throughout the contiguous United States. The growth rate, mortality rate,

fecundity rate were different at different host species (Hamilton and Lechowicz 1991, Sheehan 1992, Hajek et al. 1995). In the future, the temperature driven demographic model should be able to incorporate host species effects.

The models used in this study did not include impacts of natural enemies, including entomopathogens and parasitoids. It is an adaptation of the multi-trophic modeling approach developed by several researchers (Gutierrez 1992, Hunter and Price 1992). Thus, a module to capture the effects of species like *E. maimaga* as well as parasitoids is left as an important follow up to the present study.

The human assisted movement model identified some of locations that were at high risk based on the amount of commodities shipped, the traffic volume, routes between high-risk ports and common destinations, and preferred host species availability. The results coincided with recent detections of AGM, indicating that incorporating human activities into the AGM analysis was critical. However, this model still needs some improvements, such as examining train movements, number of vessels destined to US ports from AGM occurring countries during AGM flight period, applying weight on movement distance, and incorporating seasonality; however, it proves that this type of analysis is important and helpful to identify the likelihood of AGM introductions.

## **5.6 References**

Allen, J. C., J. L. Foltz, W. N. Dixon, A. M. Liebhold, J. J. Colbert, J. Regniere, D. R. Gray, J. W. Wilder, and I. Christie. 1993. Will the gypsy moth become a pest in Florida. Florida Entomologist **76**:102-113.

- Andreadis, T. G., and R. M. Weseloh. 1990. Discovery of *Entomophaga maimaiga* in North American gypsy moth, *Lymantria dispar*. Proceedings of the National Academy of Sciences of the United States of America **87**:2461-2465.
- Andresen, J. A., L. S. Bauer, D. P. Lusch, C. W. Ramm, D. G. McCullough, B. E. Potter, and C. N. Koller. 2001. Effects of winter temperatures on gypsy moth egg masses in the Great Lakes region of the United States. Agricultural and forest meteorology **110**:85-100.
- Barbosa, P., and J. Greenblatt. 1979. Suitability, digestibility and assimilation of various host plants of the gypsy moth *Lymantria dispar* L (Lepidoptera, Lymantriidae). Oecologia **43**:111-119.
- Blackard, J. A., M. V. Finco, E. H. Helmer, G. R. Holden, M. L. Hoppus, D. M. Jacobs, A. J. Lister, G. G. Moisen, M. D. Nelson, R. Riemann, B. Ruefenacht, D. Salajanu, D. L. Weyermann, K. C. Winterberger, T. J. Brandeis, R. L. Czaplowski, R. E. McRoberts, P. L. Patterson, and R. P. Tymcio. 2008. Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. *in* R. S. A. C. USDA Forest Service Forest Inventory and Analysis, editor. Elsevier.
- Brown, M. W., and E. A. Cameron. 1982. Spatial distribution of adult of *Ooencyrtus kuvanae* (Hymenoptera, Encyrtidae), an egg parasite of *Lymantria dispar* (Lepidoptera, Lymantriidae). Canadian Entomologist **114**:1109-1120.
- CABI. 2015. Crop Protection Compendium. CAB International. <http://www.cabi.org/cpc/>.
- Campbell, R. W. 1967. Studies on sex ratio of gypsy moth. Forest Science **13**:19-&.
- Campbell, R. W. 1976. Comparative analysis of numerically stable and violently fluctuating gypsy moth populations. Environmental Entomology **5**:1218-1224.
- Campbell, R. W. 1981. Population dynamics. Pages 65-214 *in* C. C. Doane and M. L. McManus, editors. The Gypsy Moth: Research Toward Integrated Pest Management. U.S. Department of Agriculture, Forest Service, Washington, D.C.
- Campbell, R. W., R. J. Sloan, and C. E. Biazak. 1977. Sources of mortality among late instar gypsy moth (Lepidoptera-Lymantriidae) larvae in sparse populations. Environmental Entomology **6**:865-871.
- Carter, M. R., F. W. Ravlin, and M. L. McManus. 1992. Effect of defoliation on gypsy moth phenology and capture of male moths in pheromone-baited traps. Environmental Entomology **21**:1308-1318.
- Casagrande, R. A., P. A. Logan, and W. E. Wallner. 1987. Phenological model for Gypsy Moth, *Lymantria dispar* (Lepidoptera, Lymantriidae), larvae and pupae. Environmental Entomology **16**:556-562.

- Dixon, W. N., and J. L. Foltz. 1985. The gypsy moth, *Lymantria dispar* (L.) (Lepidoptera: Lymantriidae).
- Doane, C. C., and M. L. McManus. 1981. The gypsy moth: research toward integrated pest management. U.S. Department of Agriculture, Forest Service, Washington, D.C. Technical Bulletin 1584,
- Fowler, G., Y. Takeuchi, R. Sequeira, G. Lougee, W. Fussell, M. Simon, A. Sato, and X. Yan. 2009. Pathway-Initiated Pest Risk Assessment: Asian Gypsy Moth (Lepidoptera: Lymantriidae: *Lymantria dispar* (Linnaeus)) from China into the United States on Maritime Shipments. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Raleigh, NC.
- Giebultowicz, J. M., A. K. Raina, E. C. Uebel, and R. L. Ridgway. 1991. Two-step regulation of sex-pheromone decline in mated gypsy moth females. Archives of Insect Biochemistry and Physiology **16**:95-105.
- Gould, J. R., J. S. Elkinton, and W. E. Wallner. 1990. Density-dependent suppression of experimentally created gypsy moth, *Lymantria dispar* (Lepidoptera: Lymantriidae), populations by natural enemies. Journal of Animal Ecology **59**:213-233.
- Gutierrez, A. P. 1992. Physiological basis of ratio dependent predatory prey theory - the metabolic pool model as a paradigm. Ecology **73**:1552-1563.
- Gutierrez, A. P., N. J. Mills, and L. Ponti. 2010. Limits to the potential distribution of light brown apple moth in Arizona-California based on climate suitability and host plant availability. Biological Invasions **12**:3319-3331.
- Hajek, A. E., J. A. A. Renwick, and D. W. Roberts. 1995. Effects of larval host plant on the gypsy moth (Lepidoptera, Lymantriidae) fungal pathogen, *Entomophaga maimaiga* (Zygomycetes: Entomophthorales). Environmental Entomology **24**:1307-1314.
- Hamilton, D. J., and M. J. Lechowicz. 1991. Host effects on the development and fecundity of gypsy moth, *Lymantria dispar*, reared under field conditions. Canadian Journal of Zoology-Revue Canadienne De Zoologie **69**:2217-2224.
- Higashiura, Y. 1989. Survival of eggs in the gypsy moth *Lymantria dispar*. II. Oviposition site selection in changing environments. Journal of Animal Ecology **58**:413-426.
- Hoover, G. A. 2000. Gypsy moth *Lymantria dispar* (Linnaeus). Department of Entomology, College of Agricultural Sciences, Pennsylvania State University, University Park, PA.
- Humble, L., and A. J. Stewart. 1994. Forest Pest Leaflet: gypsy moth. Canadian Forest Service, Natural Resources Canada, Burnaby, BC.

- Hunter, M. D., and P. W. Price. 1992. Playing chutes and ladders - heterogeneity and the relative roles of bottom up and top down forces in natural communities. *Ecology* **73**:724-732.
- Johnson, P. C., D. P. Mason, S. L. Radke, and K. T. Tracewski. 1983. Gypsy moth, *Lymantria dispar* (L.) (Lepidoptera: Lymantriidae), egg eclosion: degree-day accumulation. *Environmental Entomology* **12**:929-932.
- Keena, M. A. 1996. Comparison of the hatch of *Lymantria dispar* (Lepidoptera: Lymantriidae) eggs from Russia and the United States after exposure to different temperatures and durations of low temperature. *Annals of the Entomological Society of America* **89**:564-572.
- Leonard, D. E. 1972. Survival in a gypsy moth population exposed to low winter temperatures.
- Liebhold, A. M., K. W. Gottschalk, R. M. Muzika, M. E. Montgomery, R. Young, K. O'Day, and B. Kelley. 1995. Suitability of North American tree species to gypsy moth: a summary of field and laboratory tests., U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station.
- Madrid, F. J., and R. K. Stewart. 1981. Ecological significance of cold hardiness and winter mortality of eggs of the gypsy moth *Lymantria dispar* in Quebec. *Environmental Entomology* **10**:586-589.
- Mastro, V., S. Weihman, B. Chambers, E. LaGassa, S. Spinella, E. Johnson, R. Sequeira, Y. Takeuchi, G. Fowler, T. Kalaris, H. Meissner, B. Wang, and N. Jones. 2007. Use of HACCP-type methodology for mitigating risk of AGM introductions., United State Department of Agriculture, Animal and Plant Health Inspection Service, Riverdale, MD.
- Montgomery, M. E., and W. E. Wallner. 1988. The gypsy moth, a west-ward migrant. Pages 353-375 in A. A. Berryman, editor. *Dynamics of Forest Insect Populations*. Plenum Press, New York.
- Morse, J. G., and G. A. Simmons. 1979. Simulation model of gypsy moth introduced into Michigan forests. *Environmental Entomology* **8**:293-299.
- Pogue, M. G., and P. W. Schaefer. 2007. A review of selected species of *Lymantria* Hübner [1819] (Lepidoptera: Noctuidae: Lymnatriinae) from subtropical and temperate regions of Asia, including the descriptions of three new species, some potentially invasive to North America. USDA Forest Health Technology Enterprise Team.
- PRISM Climate Group. 2015. PRISM gridded climate data. Oregon State University. <http://prism.oregonstate.edu>.

- Reineke, A., and C. P. W. Zebitz. 1998. Flight ability of gypsy moth females (*Lymantria dispar* L.) (Lep., Lymantriidae): a behavioural feature characterizing moths from Asia? *Journal of Applied Entomology* **122**:307-310.
- Russo, J. M., A. M. Liebhold, and J. G. W. Kelley. 1993. Mesoscale weather data as input to a gypsy moth (Lepidoptera, Lymantriidae) phenology model. *Journal of Economic Entomology* **86**:838-844.
- Sheehan, K. A. 1992. User's guide for GMPHEN: gypsy moth phenology model. United States Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. General Technical Report NE-158,
- Smith, W. B., P. D. Miles, J. S. Vissage, and S. A. Pugh. 2004. Forest resources of the United States, 2002. U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, Minnesota.
- Southworth, F., D. Davidson, H. L. Hwang, B. E. Peterson, and S. M. Chin. 2010. The Freight Analysis Framework, Version 3: overview of the FAF national freight flow tables. Oak Ridge National Laboratory, Oak Ridge, TN.
- Sullivan, C. R., and D. R. Wallace. 1972. The potential northern dispersal of the gypsy moth, *Porthetria dispar* (Lepidoptera: Lymantriidae). *The Canadian Entomologist* **104**:1349-1355.
- Tobin, P. C., B. B. Bai, D. A. Eggen, and D. S. Leonard. 2012. The ecology, geopolitics, and economics of managing *Lymantria dispar* in the United States (vol 58, pg 195, 2012). *International Journal of Pest Management* **58**:377-377.
- USDA-APHIS-PPQ. 2011. European gypsy moth (*Lymantria dispar*) quarantine area: State of Maine. USDA Animal and Plant Health Inspection Service, Riverdale, MD.
- USDA-APHIS-PPQ. 2014. USDA-APHIS PPQ Asian gypsy moth survey and response guidelines. U.S. Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD.
- USDA-APHIS-PPQ. 2015a. Gypsy moth program manual. U.S. Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD.
- USDA-APHIS-PPQ. 2015b. USDA continues Asian gypsy moth survey in Oklahoma. U.S. Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Riverdale, MD. Last accessed October, 2015, [https://www.aphis.usda.gov/aphis/newsroom/news/!ut/p/z0/hY7LTsMwEEW\\_hUW\\_XaNzQsG8riqj6QEKg4I01SU3sxpIJ7eGRvyd4D-zu0VydO6ChAk344VsUz4Rh4ld9a3bHxXq-elK7-83jnVo-H7erh4Mq1KGELeg\\_Ci83P4Yi7tf7FvSA4q49vTFUjZgk2FnH4WSjQSJ-](https://www.aphis.usda.gov/aphis/newsroom/news/!ut/p/z0/hY7LTsMwEEW_hUW_XaNzQsG8riqj6QEKg4I01SU3sxpIJ7eGRvyd4D-zu0VydO6ChAk344VsUz4Rh4ld9a3bHxXq-elK7-83jnVo-H7erh4Mq1KGELeg_Ci83P4Yi7tf7FvSA4q49vTFUjZgk2FnH4WSjQSJ-)

p8b2liT945xGJ6c\_Xy56CbphEvsIUOHgfDIZSUzwdcQ4zlRCQ\_YzReY-w6-r-VqP5oRicy7UvMyhx9i4mZpe5i6g4x5NOw5pND2Lg6HTdRnaq2\_q1L5l/.

Weseloh, R. M. 1976. Behavioral responses of parasite, *Apanteles melanoscelus*, to gypsy moth silk. *Environmental Entomology* **5**:1128-1132.

Wilson, P. L. 2015. Gypsy moth introduction and life cycle. NCDA&CS Plant Industry Division - Plant Protection Section, Raleigh, NC. Last accessed <http://www.ncagr.gov/plantindustry/plant/entomology/gm/introduction.htm>.

Zlotina, M. A., V. C. Mastro, J. S. Elkinton, and D. E. Leonard. 1999. Dispersal tendencies of neonate larvae of *Lymantria mathura* and the Asian form of *Lymantria dispar* (Lepidoptera : Lymantriidae). *Environmental Entomology* **28**:240-245.

## APPENDICES

## **Appendix A. Equations to estimate heat units (daily degree days)**

There are number of methods to calculate daily degree-days. SAFARIS currently provides 4 upper cut-off methods and four degree-day calculation methods to calculate daily accumulated degree-days.

To calculate degree-days, daily cycle of temperature is expressed in three different ways – constant, sine wave, or triangulation (Baskervi and Emin 1969, Allen 1976, Wilson and Barnett 1983, Zalom et al. 1983). The constant means that 24-hour air temperature is expressed as one value, which is normally calculated as an average product of daily maximum and daily minimum temperatures for the day. The sine wave method expresses daily cycle of air temperature as a shape of sine wave, whereas the triangulation method expresses daily cycle as a shape of triangle.

In addition to how diurnal temperature shape is expressed, there are four cutoff methods to calculate degree-days. Those cutoff methods are 1) no cutoff, 2) horizontal cutoff, 3) vertical cutoff, 4) intermediate cutoff. There are 6 possible relationships exist between daily temperature cycle and upper and lower thresholds of organisms (Figure A-1). In this appendix, I list a degree-day equation for each case.

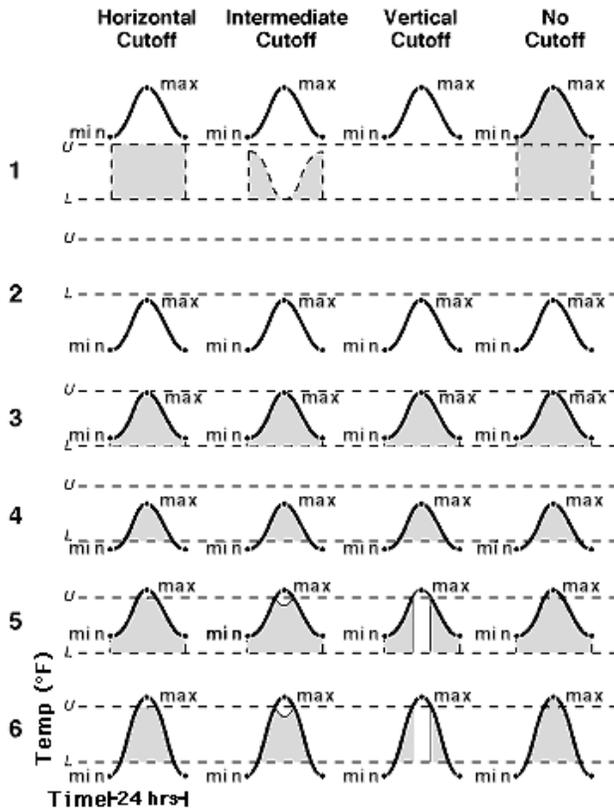


Figure A-1. Six possible relationships can exist between the daily temperature cycle and the upper and lower thresholds (UCIPM 2016)

Tmax = daily maximum temperature

Tmin = daily minimum temperature

LTT = lower threshold temperature

UTT = upper threshold temperature

Constant Diurnal Temperature Cycle:

1. Average method

$$daily\ DD = \frac{(Tmax + Tmin)}{2} - LTT$$

2. Modified average method:

If Tmin ≥ LTT,

$$\text{daily DD} = \frac{(T_{\max} + T_{\min})}{2} - LTT$$

If  $T_{\min} < LTT$ ,

$$\text{daily DD} = \frac{(T_{\max} + LTT)}{2} - LTT$$

If  $T_{\min} \geq LTT$  and  $T_{\max} \leq UTT$ ,

$$\text{daily DD} = \frac{(T_{\max} + T_{\min})}{2} - LTT$$

If  $T_{\max} > UTT$  and  $T_{\min} \geq LTT$ ,

$$\text{daily DD} = \frac{(UTT + T_{\min})}{2} - LTT$$

If  $T_{\max} > UTT$  and  $T_{\min} < LTT$ ,

$$\text{daily DD} = \frac{(UTT + LTT)}{2} - LTT$$

Single Sine Wave Temperature Cycle:

$$\alpha = \frac{(T_{\max} - T_{\min})}{2}$$

$$\theta_1 = \sin^{-1} \left[ \left( LTT - \frac{(T_{\max} + T_{\min})}{2} \right) * \frac{1}{\alpha} \right]$$

$$\theta_2 = \sin^{-1} \left[ \left( UTT - \frac{(T_{\max} + T_{\min})}{2} \right) * \frac{1}{\alpha} \right]$$

1. No Cut Off

If  $T_{\max} \leq LTT$ ,

$$\text{daily DD} = 0$$

If  $T_{\min} \geq LTT$ ,

$$\text{daily DD} = \frac{(T_{\max} + T_{\min})}{2} - LTT$$

If  $T_{\min} < LTT$  and  $T_{\max} > LTT$ ,

$$\text{daily DD} = \frac{1}{\pi} \left[ \left( \frac{T_{\max} + T_{\min}}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_1 \right) + \alpha \cos \theta_1 \right]$$

2. Horizontal Cut Off

Case 1: If  $T_{\min} \geq UTT$ ,

$$\text{daily DD} = UTT - LTT$$

Case2: If  $T_{max} \leq LTT$ ,

$$daily\ DD = 0$$

Case3: If  $T_{min} \geq LTT$  and  $T_{max} \leq UTT$ ,

$$daily\ DD = \frac{(T_{max} + T_{min})}{2} - LTT$$

Case4: If  $T_{min} < LTT$  and  $T_{max} > LTT$  and  $T_{max} \leq UTT$ ,

$$daily\ DD = \frac{1}{\pi} \left[ \left( \frac{T_{max} + T_{min}}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_1 \right) + \alpha \cos \theta_1 \right]$$

Case 5: If  $T_{min} \geq LTT$  and  $T_{max} > UTT$  and  $T_{min} < UTT$ ,

$$daily\ DD = \frac{1}{\pi} \left[ \left( \frac{T_{max} + T_{min}}{2} - LTT \right) \left( \theta_2 + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_2 \right) - \alpha \cos \theta_2 \right]$$

Case6: If  $T_{min} < LTT$  and  $T_{max} > UTT$ ,

$$daily\ DD = \frac{1}{\pi} \left[ \left( \frac{T_{max} + T_{min}}{2} - LTT \right) (\theta_2 - \theta_1) + \alpha (\cos \theta_1 - \cos \theta_2) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_2 \right) \right]$$

### 3. Vertical Cut Off

Case 1: If  $T_{min} > UTT$ ,

$$daily\ DD = 0$$

Case2: If  $T_{max} < LTT$ ,

$$daily\ DD = 0$$

Case3: If  $T_{min} > LTT$  and  $T_{max} < UTT$ ,

$$daily\ DD = \frac{(T_{max} + T_{min})}{2} - LTT$$

Case4: If  $T_{min} < LTT$  and  $T_{max} > LTT$  and  $T_{max} \leq UTT$ ,

$$daily\ DD = \frac{1}{\pi} \left[ \left( \frac{T_{max} + T_{min}}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_1 \right) + \alpha \cos \theta_1 \right]$$

Case 5: If  $T_{min} > LTT$  and  $T_{max} > UTT$ ,

$$Remove = (\pi - 2\theta_2)(UTT - LTT) \frac{1}{(2\pi)}$$

$$daily\ DD = \frac{1}{\pi} \left[ \left( \frac{Tmax + Tmin}{2} - LTT \right) \left( \theta_2 + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_2 \right) - \alpha \cos \theta_2 \right] - Remove$$

Case6: If Tmin < LTT and Tmax > UTT,

$$Remove = (\pi - 2\theta_2)(UTT - LTT) \frac{1}{(2\pi)}$$

$$daily\ DD = \frac{1}{\pi} \left[ \left( \frac{Tmax + Tmin}{2} - LTT \right) (\theta_2 - \theta_1) + \alpha (\cos \theta_1 - \cos \theta_2) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_2 \right) \right] - Remove$$

#### 4. Intermediate Cut Off

Case 1: If Tmin > UTT,

$$daily\ DD = 2UTT - LTT - \frac{(Tmax + Tmin)}{2}$$

Case2: If Tmax < LTT,

$$daily\ DD = 0$$

Case3: If Tmin > LTT and Tmax < UTT,

$$daily\ DD = \frac{(Tmax + Tmin)}{2} - LTT$$

Case4: If Tmin < LTT and Tmax > LTT and Tmax =< UTT,

$$daily\ DD = \frac{1}{\pi} \left[ \left( \frac{Tmax + Tmin}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_1 \right) + \alpha \cos \theta_1 \right]$$

Case 5: If Tmin > LTT and Tmax > UTT,

$$Remove = 2 * \left\{ \left[ \frac{(Tmax + Tmin)}{2} - LTT \right] - \frac{1}{\pi} \left[ \left( \frac{Tmax + Tmin}{2} - LTT \right) \left( \theta_2 + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_2 \right) - \alpha \cos \theta_2 \right] \right\}$$

$$daily\ DD = \left[ \frac{(Tmax + Tmin)}{2} - LTT \right] - Remove$$

Case6: If Tmin < LTT and Tmax > UTT,

$$\begin{aligned}
Remove &= 2 * \left\{ \left[ \frac{(Tmax + Tmin)}{2} - LTT \right] \right. \\
&\quad \left. - \frac{1}{\pi} \left[ \left( \frac{(Tmax + Tmin)}{2} - LTT \right) \left( \theta_2 + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_2 \right) \right] \right. \\
&\quad \left. - \alpha \cos \theta_2 \right\} \\
daily\ DD &= \frac{1}{\pi} \left[ \left( \frac{(Tmax + Tmin)}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_1 \right) + \alpha \cos \theta_1 \right] - Remove
\end{aligned}$$

Double Sine Wave Temperature Cycle:

$$Tmin_1 = Tmin\ of\ Current\ Day$$

$$Tmin_2 = Tmin\ of\ Next\ Day$$

$$Tmin = Tmin_1\ or\ Tmin_2$$

$$\alpha_1 = \frac{(Tmax - Tmin_1)}{2}$$

$$\alpha_2 = \frac{(Tmax - Tmin_2)}{2}$$

$$\theta_{11} = \sin^{-1} \left[ \left( LTT - \frac{(Tmax + Tmin_1)}{2} \right) * \frac{1}{\alpha_1} \right]$$

$$\theta_{12} = \sin^{-1} \left[ \left( LTT - \frac{(Tmax + Tmin_2)}{2} \right) * \frac{1}{\alpha_2} \right]$$

$$\theta_{21} = \sin^{-1} \left[ \left( UTT - \frac{(Tmax + Tmin_1)}{2} \right) * \frac{1}{\alpha_1} \right]$$

$$\theta_{22} = \sin^{-1} \left[ \left( UTT - \frac{(Tmax + Tmin_2)}{2} \right) * \frac{1}{\alpha_2} \right]$$

$$daily\ DD = daily\ DD_1 + daily\ DD_2$$

1. No Cut Off

If  $Tmax \leq LTT$ ,

$$daily\ DD_1 = 0$$

$$daily DD_2 = 0$$

If  $T_{min} \geq LTT$ ,

$$daily DD_1 = \frac{1}{2} \left[ \frac{(T_{max} + T_{min_1})}{2} - LTT \right]$$

$$daily DD_2 = \frac{1}{2} \left[ \frac{(T_{max} + T_{min_2})}{2} - LTT \right]$$

If  $T_{min} < LTT$  and  $T_{max} > LTT$ ,

$$daily DD_1 = \frac{1}{2\pi} \left[ \left( \frac{(T_{max} + T_{min_1})}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_{11} \right) + \alpha_1 \cos \theta_{11} \right]$$

$$daily DD_2 = \frac{1}{2\pi} \left[ \left( \frac{(T_{max} + T_{min_2})}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_{12} \right) + \alpha_2 \cos \theta_{12} \right]$$

## 2. Horizontal Cut Off

Case 1: If  $T_{min} \geq UTT$ ,

$$daily DD_1 = UTT - LTT$$

$$daily DD_2 = UTT - LTT$$

Case2: If  $T_{max} \leq LTT$ ,

$$daily DD_1 = 0$$

$$daily DD_2 = 0$$

Case3: If  $T_{min} \geq LTT$  and  $T_{max} \leq UTT$ ,

$$daily DD_1 = \frac{1}{2} \left[ \frac{(T_{max} + T_{min_1})}{2} - LTT \right]$$

$$daily DD_2 = \frac{1}{2} \left[ \frac{(T_{max} + T_{min_2})}{2} - LTT \right]$$

Case4: If  $T_{min} < LTT$  and  $T_{max} > LTT$  and  $T_{max} \leq UTT$ ,

$$daily DD_1 = \frac{1}{2\pi} \left[ \left( \frac{(T_{max} + T_{min_1})}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_{11} \right) + \alpha_1 \cos \theta_{11} \right]$$

$$daily DD_2 = \frac{1}{2\pi} \left[ \left( \frac{(T_{max} + T_{min_2})}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_{12} \right) + \alpha_2 \cos \theta_{12} \right]$$

Case 5: If  $T_{min} \geq LTT$  and  $T_{max} > UTT$  and  $T_{min} < UTT$ ,

$$daily DD_1 = \frac{1}{2\pi} \left[ \left( \frac{(T_{max} + T_{min_1})}{2} - LTT \right) \left( \theta_{21} + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{21} \right) - \alpha_1 \cos \theta_{21} \right]$$

$$daily DD_2 = \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_2}{2} - LTT \right) \left( \theta_{22} + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{22} \right) - \alpha_2 \cos \theta_{22} \right]$$

Case6: If  $Tmin < LTT$  and  $Tmax > UTT$ ,

$$daily DD_1 = \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_1}{2} - LTT \right) (\theta_{21} - \theta_{11}) + \alpha_1 (\cos \theta_{11} - \cos \theta_{21}) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{21} \right) \right]$$

$$daily DD_2 = \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_2}{2} - LTT \right) (\theta_{22} - \theta_{12}) + \alpha_2 (\cos \theta_{12} - \cos \theta_{22}) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{22} \right) \right]$$

### 3. Vertical Cut Off

Case 1: If  $Tmin \geq UTT$ ,

$$\begin{aligned} daily DD_1 &= 0 \\ daily DD_2 &= 0 \end{aligned}$$

Case2: If  $Tmax \leq LTT$ ,

$$\begin{aligned} daily DD_1 &= 0 \\ daily DD_2 &= 0 \end{aligned}$$

Case3: If  $Tmin \geq LTT$  and  $Tmax \leq UTT$ ,

$$\begin{aligned} daily DD_1 &= \frac{1}{2} \left[ \frac{(Tmax + Tmin_1)}{2} - LTT \right] \\ daily DD_2 &= \frac{1}{2} \left[ \frac{(Tmax + Tmin_2)}{2} - LTT \right] \end{aligned}$$

Case4: If  $Tmin < LTT$  and  $Tmax > LTT$  and  $Tmax \leq UTT$ ,

$$\begin{aligned} daily DD_1 &= \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_1}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_{11} \right) + \alpha_1 \cos \theta_{11} \right] \\ daily DD_2 &= \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_2}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_{12} \right) + \alpha_2 \cos \theta_{12} \right] \end{aligned}$$

Case 5: If  $Tmin \geq LTT$  and  $Tmax > UTT$  and  $Tmin < UTT$ ,

$$\begin{aligned} Remove_1 &= (\pi - 2\theta_{21})(UTT - LTT) \frac{1}{(4\pi)} \\ Remove_2 &= (\pi - 2\theta_{22})(UTT - LTT) \frac{1}{(4\pi)} \end{aligned}$$

$$\begin{aligned}
daily\ DD_1 &= \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_1}{2} - LTT \right) \left( \theta_{21} + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{21} \right) \right. \\
&\quad \left. - \alpha_1 \cos \theta_{21} \right] - Remove_1 \\
daily\ DD_2 &= \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_2}{2} - LTT \right) \left( \theta_{22} + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{22} \right) \right. \\
&\quad \left. - \alpha_2 \cos \theta_{22} \right] - Remove_2
\end{aligned}$$

Case6: If  $Tmin < LTT$  and  $Tmax > UTT$ ,

$$\begin{aligned}
Remove_1 &= (\pi - 2\theta_{21})(UTT - LTT) \frac{1}{(4\pi)} \\
Remove_2 &= (\pi - 2\theta_{22})(UTT - LTT) \frac{1}{(4\pi)} \\
daily\ DD_1 &= \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_1}{2} - LTT \right) (\theta_{21} - \theta_{11}) + \alpha_1 (\cos \theta_{11} - \cos \theta_{21}) \right. \\
&\quad \left. + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{21} \right) \right] - Remove_1 \\
daily\ DD_2 &= \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_2}{2} - LTT \right) (\theta_{22} - \theta_{12}) + \alpha_2 (\cos \theta_{12} - \cos \theta_{22}) \right. \\
&\quad \left. + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{22} \right) \right] - Remove_2
\end{aligned}$$

#### 4. Intermediate Cut Off

Case 1: If  $Tmin \geq UTT$ ,

$$\begin{aligned}
daily\ DD_1 &= \frac{1}{2} \left[ 2UTT - LTT - \frac{(Tmax + Tmin_1)}{2} \right] \\
daily\ DD_2 &= \frac{1}{2} \left[ 2UTT - LTT - \frac{(Tmax + Tmin_2)}{2} \right]
\end{aligned}$$

Case2: If  $Tmax \leq LTT$ ,

$$\begin{aligned}
daily\ DD_1 &= 0 \\
daily\ DD_2 &= 0
\end{aligned}$$

Case3: If  $Tmin \geq LTT$  and  $Tmax \leq UTT$ ,

$$daily\ DD_1 = \frac{1}{2} \left[ \frac{(Tmax + Tmin_1)}{2} - LTT \right]$$

$$daily DD_2 = \frac{1}{2} \left[ \frac{(Tmax + Tmin_2)}{2} - LTT \right]$$

Case4: If  $Tmin < LTT$  and  $Tmax > LTT$  and  $Tmax \leq UTT$ ,

$$daily DD_1 = \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_1}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_{11} \right) + \alpha_1 \cos \theta_{11} \right]$$

$$daily DD_2 = \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_2}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_{12} \right) + \alpha_2 \cos \theta_{12} \right]$$

Case 5: If  $Tmin \geq LTT$  and  $Tmax > UTT$  and  $Tmin < UTT$ ,

$$Remove_1 = \left\{ \left[ \frac{(Tmax + Tmin_1)}{2} - LTT \right] \right. \\ \left. - \frac{1}{\pi} \left[ \left( \frac{Tmax + Tmin_1}{2} - LTT \right) \left( \theta_{21} + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{21} \right) \right] \right. \\ \left. - \alpha_1 \cos \theta_{21} \right\}$$

$$Remove_2 = \left\{ \left[ \frac{(Tmax + Tmin_2)}{2} - LTT \right] \right. \\ \left. - \frac{1}{\pi} \left[ \left( \frac{Tmax + Tmin_2}{2} - LTT \right) \left( \theta_{22} + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{22} \right) \right] \right. \\ \left. - \alpha_2 \cos \theta_{22} \right\}$$

$$daily DD_1 = \frac{1}{2} \left[ \left( \frac{Tmax + Tmin_1}{2} - LTT \right) \right] - Remove_1$$

$$daily DD_2 = \frac{1}{2} \left[ \left( \frac{Tmax + Tmin_2}{2} - LTT \right) \right] - Remove_2$$

Case6: If  $Tmin < LTT$  and  $Tmax > UTT$ ,

$$Remove_1 = \left\{ \left[ \frac{(Tmax + Tmin_1)}{2} - LTT \right] - \frac{1}{\pi} \left[ \left( \frac{Tmax + Tmin_1}{2} - LTT \right) \left( \theta_{21} + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{21} \right) - \alpha_1 \cos \theta_{21} \right] \right\}$$

$$Remove_2 = \left\{ \left[ \frac{(Tmax + Tmin_2)}{2} - LTT \right] - \frac{1}{\pi} \left[ \left( \frac{Tmax + Tmin_2}{2} - LTT \right) \left( \theta_{22} + \frac{\pi}{2} \right) + (UTT - LTT) \left( \frac{\pi}{2} - \theta_{22} \right) - \alpha_2 \cos \theta_{22} \right] \right\}$$

$$daily DD_1 = \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_1}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_{11} \right) + \alpha_1 \cos \theta_{11} \right] - Remove_1$$

$$daily DD_2 = \frac{1}{2\pi} \left[ \left( \frac{Tmax + Tmin_2}{2} - LTT \right) \left( \frac{\pi}{2} - \theta_{12} \right) + \alpha_2 \cos \theta_{12} \right] - Remove_2$$

Single Triangulation Temperature Cycle:

1. No Cut Off

If Tmax ≤ LTT,

$$daily DD = 0$$

If Tmin ≥ LTT,

$$daily DD = \frac{(Tmax + Tmin)}{2} - LTT$$

If Tmin < LTT and Tmax > LTT,

$$daily DD = \frac{(Tmax - LTT)^2}{2(Tmax - Tmin)}$$

2. Horizontal Cut Off

Case 1: If Tmin ≥ UTT,

$$daily DD = UTT - LTT$$

Case2: If Tmax ≤ LTT,

$$\text{daily DD} = 0$$

Case3: If  $T_{min} \geq LTT$  and  $T_{max} \leq UTT$ ,

$$\text{daily DD} = \frac{(T_{max} + T_{min})}{2} - LTT$$

Case4: If  $T_{min} < LTT$  and  $T_{max} > LTT$  and  $T_{max} \leq UTT$ ,

$$\text{daily DD} = \frac{(T_{max} - LTT)^2}{2(T_{max} - T_{min})}$$

Case 5: If  $T_{min} \geq LTT$  and  $T_{max} > UTT$  and  $T_{min} < UTT$ ,

$$\text{daily DD} = \frac{(T_{max} + T_{min})}{2} - \frac{(T_{max} - UTT)^2}{2(T_{max} - T_{min})} - LTT$$

Case6: If  $T_{min} < LTT$  and  $T_{max} > UTT$ ,

$$\text{daily DD} = \frac{(T_{max} - LTT)^2 - (T_{max} - UTT)^2}{2(T_{max} - T_{min})}$$

### 3. Vertical Cut Off

Case 1: If  $T_{min} \geq UTT$ ,

$$\text{daily DD} = 0$$

Case2: If  $T_{max} \leq LTT$ ,

$$\text{daily DD} = 0$$

Case3: If  $T_{min} \geq LTT$  and  $T_{max} \leq UTT$ ,

$$\text{daily DD} = \frac{(T_{max} + T_{min})}{2} - LTT$$

Case4: If  $T_{min} < LTT$  and  $T_{max} > LTT$  and  $T_{max} \leq UTT$ ,

$$\text{daily DD} = \frac{(T_{max} - LTT)^2}{2(T_{max} - T_{min})}$$

Case 5: If  $T_{min} \geq LTT$  and  $T_{max} > UTT$  and  $T_{min} < UTT$ ,

$$\text{daily DD} = \frac{(T_{max} + T_{min})}{2} - \frac{(T_{max} - UTT)^2}{2(T_{max} - T_{min})} - \frac{(T_{max} - UTT)(UTT - LTT)}{(T_{max} - T_{min})} - LTT$$

Case6: If  $T_{min} < LTT$  and  $T_{max} > UTT$ ,

$$\text{daily DD} = \frac{(T_{max} - LTT)^2 - (T_{max} - UTT)^2}{2(T_{max} - T_{min})} - \frac{(T_{max} - UTT)(UTT - LTT)}{(T_{max} - T_{min})}$$

4. Intermediate Cut Off

Case 1: If  $T_{min} \geq UTT$ ,

$$daily\ DD = 2UTT - \frac{(T_{max} + T_{min})}{2} - LTT$$

Case2: If  $T_{max} \leq LTT$ ,

$$daily\ DD = 0$$

Case3: If  $T_{min} \geq LTT$  and  $T_{max} \leq UTT$ ,

$$daily\ DD = \frac{(T_{max} + T_{min})}{2} - LTT$$

Case4: If  $T_{min} < LTT$  and  $T_{max} > LTT$  and  $T_{max} \leq UTT$ ,

$$daily\ DD = \frac{(T_{max} - LTT)^2}{2(T_{max} - T_{min})}$$

Case 5: If  $T_{min} \geq LTT$  and  $T_{max} > UTT$  and  $T_{min} < UTT$ ,

$$daily\ DD = \frac{(T_{max} + T_{min})}{2} - \frac{(T_{max} - UTT)^2}{(T_{max} - T_{min})} - LTT$$

Case6: If  $T_{min} < LTT$  and  $T_{max} > UTT$ ,

$$daily\ DD = \frac{(T_{max} - LTT)^2 - 2(T_{max} - UTT)^2}{2(T_{max} - T_{min})}$$

Double Triangulation Temperature Cycle:

$$T_{min_1} = T_{min\ of\ Current\ Day}$$

$$T_{min_2} = T_{min\ of\ Next\ Day}$$

$$T_{min} = T_{min_1}\ or\ T_{min_2}$$

$$daily\ DD = daily\ DD_1 + daily\ DD_2$$

1. No Cut Off

If  $T_{max} \leq LTT$ ,

$$daily\ DD_1 = 0$$

$$daily\ DD_2 = 0$$

If  $T_{min} \geq LTT$ ,

$$daily DD_1 = \frac{1}{2} \left[ \frac{(Tmax + Tmin_1)}{2} - LTT \right]$$

$$daily DD_2 = \frac{1}{2} \left[ \frac{(Tmax + Tmin_2)}{2} - LTT \right]$$

If  $Tmin < LTT$  and  $Tmax > LTT$ ,

$$daily DD_1 = \frac{(Tmax - LTT)^2}{4(Tmax - Tmin_1)}$$

$$daily DD_2 = \frac{(Tmax - LTT)^2}{4(Tmax - Tmin_2)}$$

## 2. Horizontal Cut Off

Case 1: If  $Tmin \geq UTT$ ,

$$daily DD_1 = UTT - LTT$$

$$daily DD_2 = UTT - LTT$$

Case2: If  $Tmin \geq UTT$ ,

$$daily DD_1 = 0$$

$$daily DD_2 = 0$$

Case3: If  $Tmin \geq LTT$  and  $Tmax \leq UTT$ ,

$$daily DD_1 = \frac{1}{2} \left[ \frac{(Tmax + Tmin_1)}{2} - LTT \right]$$

$$daily DD_2 = \frac{1}{2} \left[ \frac{(Tmax + Tmin_2)}{2} - LTT \right]$$

Case4: If  $Tmin < LTT$  and  $Tmax > LTT$  and  $Tmax \leq UTT$ ,

$$daily DD_1 = \frac{(Tmax - LTT)^2}{4(Tmax - Tmin_1)}$$

$$daily DD_2 = \frac{(Tmax - LTT)^2}{4(Tmax - Tmin_2)}$$

Case 5: If  $Tmin \geq LTT$  and  $Tmax > UTT$  and  $Tmin < UTT$ ,

$$daily DD_1 = \frac{1}{2} \left[ \frac{(Tmax + Tmin_1)}{2} - \frac{(Tmax - UTT)^2}{2(Tmax - Tmin_1)} - LTT \right]$$

$$daily DD_2 = \frac{1}{2} \left[ \frac{(Tmax + Tmin_2)}{2} - \frac{(Tmax - UTT)^2}{2(Tmax - Tmin_2)} - LTT \right]$$

Case6: If  $Tmin < LTT$  and  $Tmax > UTT$ ,

$$daily DD_1 = \frac{(Tmax - LTT)^2 - (Tmax - UTT)^2}{2(Tmax - Tmin_1)}$$

$$\text{daily } DD_2 = \frac{(Tmax - LTT)^2 - (Tmax - UTT)^2}{2(Tmax - Tmin_2)}$$

### 3. Vertical Cut Off

Case 1: If  $Tmin \geq UTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= 0 \\ \text{daily } DD_2 &= 0 \end{aligned}$$

Case2: If  $Tmax \leq LTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= 0 \\ \text{daily } DD_2 &= 0 \end{aligned}$$

Case3: If  $Tmin \geq LTT$  and  $Tmax \leq UTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= \frac{1}{2} \left[ \frac{(Tmax + Tmin_1)}{2} - LTT \right] \\ \text{daily } DD_2 &= \frac{1}{2} \left[ \frac{(Tmax + Tmin_2)}{2} - LTT \right] \end{aligned}$$

Case4: If  $Tmin < LTT$  and  $Tmax > LTT$  and  $Tmax \leq UTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= \frac{(Tmax - LTT)^2}{4(Tmax - Tmin_1)} \\ \text{daily } DD_2 &= \frac{(Tmax - LTT)^2}{4(Tmax - Tmin_2)} \end{aligned}$$

Case 5: If  $Tmin \geq LTT$  and  $Tmax > UTT$  and  $Tmin < UTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= \frac{1}{2} \left[ \frac{(Tmax + Tmin_1)}{2} - \frac{(Tmax - UTT)^2}{2(Tmax - Tmin_1)} \right. \\ &\quad \left. - \frac{(Tmax - UTT)(UTT - LTT)}{(Tmax - Tmin_1)} - LTT \right] \\ \text{daily } DD_2 &= \frac{1}{2} \left[ \frac{(Tmax + Tmin_2)}{2} - \frac{(Tmax - UTT)^2}{2(Tmax - Tmin_2)} \right. \\ &\quad \left. - \frac{(Tmax - UTT)(UTT - LTT)}{(Tmax - Tmin_2)} - LTT \right] \end{aligned}$$

Case6: If  $Tmin < LTT$  and  $Tmax > UTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= \frac{1}{2} \left[ \frac{(Tmax - LTT)^2 - (Tmax - UTT)^2}{2(Tmax - Tmin_1)} - \frac{(Tmax - UTT)(UTT - LTT)}{(Tmax - Tmin_1)} \right] \\ \text{daily } DD_2 &= \frac{1}{2} \left[ \frac{(Tmax - LTT)^2 - (Tmax - UTT)^2}{2(Tmax - Tmin_2)} - \frac{(Tmax - UTT)(UTT - LTT)}{(Tmax - Tmin_2)} \right] \end{aligned}$$

#### 4. Intermediate Cut Off

Case 1: If  $T_{min} \geq UTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= \frac{1}{2} \left[ 2UTT - \frac{(T_{max} + T_{min_1})}{2} - LTT \right] \\ \text{daily } DD_2 &= \frac{1}{2} \left[ 2UTT - \frac{(T_{max} + T_{min_2})}{2} - LTT \right] \end{aligned}$$

Case2: If  $T_{max} \leq LTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= 0 \\ \text{daily } DD_2 &= 0 \end{aligned}$$

Case3: If  $T_{min} \geq LTT$  and  $T_{max} \leq UTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= \frac{1}{2} \left[ \frac{(T_{max} + T_{min_1})}{2} - LTT \right] \\ \text{daily } DD_2 &= \frac{1}{2} \left[ \frac{(T_{max} + T_{min_2})}{2} - LTT \right] \end{aligned}$$

Case4: If  $T_{min} < LTT$  and  $T_{max} > LTT$  and  $T_{max} \leq UTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= \frac{(T_{max} - LTT)^2}{4(T_{max} - T_{min_1})} \\ \text{daily } DD_2 &= \frac{(T_{max} - LTT)^2}{4(T_{max} - T_{min_2})} \end{aligned}$$

Case 5: If  $T_{min} \geq LTT$  and  $T_{max} > UTT$  and  $T_{min} < UTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= \frac{1}{2} \left[ \frac{(T_{max} + T_{min_1})}{2} - \frac{(T_{max} - UTT)^2}{(T_{max} - T_{min_1})} - LTT \right] \\ \text{daily } DD_2 &= \frac{1}{2} \left[ \frac{(T_{max} + T_{min_2})}{2} - \frac{(T_{max} - UTT)^2}{(T_{max} - T_{min_2})} - LTT \right] \end{aligned}$$

Case6: If  $T_{min} < LTT$  and  $T_{max} > UTT$ ,

$$\begin{aligned} \text{daily } DD_1 &= \frac{1}{2} \left[ \frac{(T_{max} - LTT)^2 - 2(T_{max} - UTT)^2}{2(T_{max} - T_{min_1})} \right] \\ \text{daily } DD_2 &= \frac{1}{2} \left[ \frac{(T_{max} - LTT)^2 - 2(T_{max} - UTT)^2}{2(T_{max} - T_{min_2})} \right] \end{aligned}$$

## References:

- Allen, J. C. 1976. A modified sine wave method for calculating degree-days. *Environmental Entomology* **5**:388-396.
- Baskervi, G. I., and P. Emin. 1969. Rapid estimation of heat accumulation from maximum and minimum temperatures. *Ecology* **50**:514-&.
- UCIPM. 2016. How to manage pests: degree-days. University of California, Agriculture & Natural Resources, Statewide Integrated Pest Management Program, Davis, California. Last accessed January, 2016, <http://ipm.ucdavis.edu/WEATHER/ddconcepts.html#7>.
- Wilson, L. T., and W. W. Barnett. 1983. Degree-days: an aid in crop and pest management. *California Agriculture* **37**:4-7.
- Zalom, F. G., P. B. Goodell, L. T. Wilson, W. W. Barnett, and W. J. Bentley. 1983. Degree-days: the calculation and use of heat units in pest management., University of California, Division of Agriculture and Natural Resources. Leaflet 21373,

**Appendix B. CLIMEX parameter values for *Epiphyas postvittana* developed by Lozier & Mills (2011) and He et al. (2012)**

Parameters by Lozier & Mills (2011)

CLIMEX parameter	Value
Temperature (TI)	
DV0 Limiting low temperature	7.5
DV1 Lower optimal temperature	15
DV2 Upper optimal temperature	25
DV3 Limiting high temperature	31
PDD Degree-days per generation	673.6
Moisture (MI)	
SM0 Limiting low moisture	0.15
SM1 Lower optimal moisture	0.5
SM2 Upper optimal moisture	0.8
SM3 Limiting high moisture	1.4
Cold Stress (CS)	
TTCS Cold stress temperature threshold	5
THCS Cold stress temperature rate	-0.005
Heat Stress (HS)	
TTHS Heat stress temperature threshold	31
THHS Heat stress temperature rate	0.0045
Dry Stress (DS)	
SMDS Dry stress threshold	0.15
HDS Dry stress rate	-0.005
Wet Stress (WS)	
SMWS Wet stress threshold	1.4
HWS Wet stress rate	0.001

Parameters by He et al. (2012)

CLIMEX parameter	Value
Temperature (TI)	
DV0 Limiting low temperature	7
DV1 Lower optimal temperature	13
DV2 Upper optimal temperature	23
DV3 Limiting high temperature	30
PDD Degree-days per generation	673.6
Moisture (MI)	
SM0 Limiting low moisture	0.25
SM1 Lower optimal moisture	0.8
SM2 Upper optimal moisture	1.5
SM3 Limiting high moisture	2.5
Cold Stress (CS)	
TTCS Cold stress temperature threshold	5
THCS Cold stress temperature rate	-0.0005
DTCS Cold stress degree-day threshold	0
DHCS Cold stress degree-day rate	0
TTCSA Cold stress temp. threshold (avg)	0
THCSA Cold stress temp. rate (avg)	0
Heat Stress (HS)	
TTHS Heat stress temperature threshold	31
THHS Heat stress temperature rate	0.01
DTHS Heat stress degree-day threshold	0
DHHS Heat stress degree-day rate	0
Dry Stress (DS)	
SMDS Dry stress threshold	0.2
HDS Dry stress rate	-0.01
Wet Stress (WS)	
SMWS Wet stress threshold	2.5
HWS Wet stress rate	0.002

**References:**

- He, S., S. P. Worner, and T. Ikeda. 2012. Modeling the potential global distribution of light brown apple moth *Epiphyas postvittana* (Lepidoptera: Tortricidae) using CLIMEX. *Journal of Asia-Pacific Entomology* **15**:479-485.
- Lozier, J. D., and N. J. Mills. 2011. Predicting the potential invasive range of light brown apple moth (*Epiphyas postvittana*) using biologically informed and correlative species distribution models. *Biological Invasions* **13**:2409-2421.

### **Appendix C. A life table of Oriental fruit fly (*Bactrocera dorsalis*)**

I used biological information from Rwomushana (2008) and Kalia & Yadav (2015) to parameterize a phenology model for Oriental fruit fly. The phenology model was used to estimate potential number of generations within the contiguous United States. The model parameters were:

Lower threshold temperature: 9°C

Upper threshold temperature: 35°C

Egg – Pupa: 376 degree-days

Pre-oviposition: 360 degree-days

The web-based phenology model within SAFARIS was used to accumulate degree-days for each year to estimate potential number of generations per year. Based on the experiments by Rwomushana (2008) and Kalia & Yadav (2015), Oriental fruit fly required 376 degree-days to complete pupal stage from egg stage in addition to 360 degree-days to become mature enough to lay eggs. Therefore, it required 736 degree-days from egg stage to next generation egg stage.

#### **References:**

- Kalia, V. K., and B. Yadav. 2015. Cost-effective mass rearing of the Oriental fruit fly, *Bactrocera dorsalis* (Hendel) round the year. *International Research Journal of Natural and Applied Sciences* **2**:143-155.
- Rwomushana, I., S. Ekesi, C. Ogol, and I. Gordon. 2008. Effect of temperature on development and survival of immature stages of *Bactrocera invadens* (Diptera: Tephritidae). *Journal of Applied Entomology* **132**:832-839.

## Appendix D. Bioclim variables

Bioclimatic variables are commonly used climatic variables derived from monthly temperature and rainfall values in order to generate more biologically meaningful variables (Hijmans et al. 2005, O'Donnell and Ignizio 2012). The temperature values used to generate Bioclimatic values were measured in degrees Kelvin. The temperature for Bioclimatic values (e.g. Bio1, Bio2, Bio5....) were converted to Celsius by simply subtracting 273.15 from each original value.

### Notation for Equation

$i$  = month

$T_{max}$  = monthly mean of daily maximum temperature

$T_{min}$  = monthly mean of daily minimum temperature

$$T_{avg} = \frac{(T_{max_i} + T_{min_i})}{2}$$

PPT = total monthly precipitation

### BIO1 = Annual Mean Temperature

$$Bio1 = \frac{\sum_{i=1}^{i=12} T_{avg_i}}{12}$$

### BIO2 = Mean Diurnal Range (Mean of monthly (max temp – min temp))

$$Bio2 = \frac{\sum_{i=1}^{i=12} (T_{max_i} - T_{min_i})}{12}$$

### BIO3 = Isothermality ((BIO2/BIO7)\*100)

$$Bio3 = \frac{Bio2}{Bio7} * 100$$

**BIO4 = Temperature Seasonality (standard deviation \* 100)**

$$Bio4 = \frac{SD\{Tavg_1 + \dots + Tavg_{12}\}}{(Bio1 + 273.15)} * 100$$

**BIO5 = Max Temperature of Warmest Month**

$$Bio5 = \max(\{Tmax_1 + \dots + Tmax_{12}\})$$

**BIO6 = Min Temperature of Coldest Month**

$$Bio6 = \min(\{Tmin_1 + \dots + Tmin_{12}\})$$

**BIO7 = Temperature Annual Range (BIO5 – BIO6)**

$$Bio7 = Bio5 - Bio6$$

**BIO8 = Mean Temperature of Wettest Quarter**

$$Qppt_{max} = \max \left[ \begin{array}{l} \sum_{i=1}^{i=3} PPT_i \\ \sum_{i=2}^{i=4} PPT_i \\ \sum_{i=3}^{i=5} PPT_i \\ \sum_{i=4}^{i=6} PPT_i \\ \sum_{i=5}^{i=7} PPT_i \\ \sum_{i=6}^{i=8} PPT_i \\ \sum_{i=7}^{i=9} PPT_i \\ \sum_{i=8}^{i=10} PPT_i \\ \sum_{i=9}^{i=11} PPT_i \\ \sum_{i=10}^{i=12} PPT_i \\ \sum_{i=11}^{i=1} PPT_i \\ \sum_{i=12}^{i=2} PPT_i \end{array} \right]$$

Where precipitation is evaluated for 12 consecutive sets of 3 months.

$$Bio8 = \frac{\sum_{i=1}^{i=3} Tavg_i}{3} \text{ at } Qppt_{max}$$

**BIO9 = Mean Temperature of Driest Quarter**

$$Qppt_{min} = \min \left[ \begin{array}{l} \sum_{i=1}^{i=3} PPT_i \\ \sum_{i=2}^{i=4} PPT_i \\ \sum_{i=3}^{i=5} PPT_i \\ \sum_{i=4}^{i=6} PPT_i \\ \sum_{i=5}^{i=7} PPT_i \\ \sum_{i=6}^{i=8} PPT_i \\ \sum_{i=7}^{i=9} PPT_i \\ \sum_{i=8}^{i=10} PPT_i \\ \sum_{i=9}^{i=11} PPT_i \\ \sum_{i=10}^{i=12} PPT_i \\ \sum_{i=11}^{i=1} PPT_i \\ \sum_{i=12}^{i=2} PPT_i \end{array} \right]$$

Where precipitation is evaluated for 12 consecutive sets of 3 months.

$$Bio9 = \frac{\sum_{i=1}^{i=3} Tavg_i}{3} \text{ at } Qppt_{min}$$

**BIO10 = Mean Temperature of Warmest Quarter**

$$QT_{max} = \max \left[ \begin{array}{l} \sum_{i=1}^{i=3} Tavg_i \\ \sum_{i=2}^{i=4} Tavg_i \\ \sum_{i=3}^{i=5} Tavg_i \\ \sum_{i=4}^{i=6} Tavg_i \\ \sum_{i=5}^{i=7} Tavg_i \\ \sum_{i=6}^{i=8} Tavg_i \\ \sum_{i=7}^{i=9} Tavg_i \\ \sum_{i=8}^{i=10} Tavg_i \\ \sum_{i=9}^{i=11} Tavg_i \\ \sum_{i=10}^{i=12} Tavg_i \\ \sum_{i=11}^{i=1} Tavg_i \\ \sum_{i=12}^{i=2} Tavg_i \end{array} \right]$$

Where temperatures are evaluated for 12 consecutive sets of 3 months.

$$Bio10 = \frac{\sum_{i=1}^{i=3} T_{avg_i}}{3} \text{ at } QT_{max}$$

**BIO11 = Mean Temperature of Coldest Quarter**

$$QT_{min} = \min \left[ \begin{array}{l} \sum_{i=1}^{i=3} T_{avg_i} \\ \sum_{i=2}^{i=4} T_{avg_i} \\ \sum_{i=3}^{i=5} T_{avg_i} \\ \sum_{i=4}^{i=6} T_{avg_i} \\ \sum_{i=5}^{i=7} T_{avg_i} \\ \sum_{i=6}^{i=8} T_{avg_i} \\ \sum_{i=7}^{i=9} T_{avg_i} \\ \sum_{i=8}^{i=10} T_{avg_i} \\ \sum_{i=9}^{i=11} T_{avg_i} \\ \sum_{i=10}^{i=12} T_{avg_i} \\ \sum_{i=11}^{i=1} T_{avg_i} \\ \sum_{i=12}^{i=2} T_{avg_i} \end{array} \right]$$

Where temperatures are evaluated for 12 consecutive sets of 3 months.

$$Bio11 = \frac{\sum_{i=1}^{i=3} T_{avg_i}}{3} \text{ at } QT_{min}$$

**BIO12 = Annual Precipitation**

$$Bio12 = \sum_{i=1}^{i=12} PPT_i$$

**BIO13 = Precipitation of Wettest Month**

$$Bio13 = \max(\{PPT_1 + \dots + PPT_{12}\})$$

**BIO14 = Precipitation of Driest Month**

$$Bio14 = \min(\{PPT_1 + \dots + PPT_{12}\})$$

**BIO15 = Precipitation Seasonality (Coefficient of Variation)**

$$Bio15 = \frac{SD\{PPT_1 + \dots + PPT_{12}\}}{(Bio12/12)} * 100$$

**BIO16 = Precipitation of Wettest Quarter**

$$Bio16 = \max \left[ \begin{array}{l} \sum_{i=1}^{i=3} PPT_i \\ \sum_{i=2}^{i=4} PPT_i \\ \sum_{i=3}^{i=5} PPT_i \\ \sum_{i=4}^{i=6} PPT_i \\ \sum_{i=5}^{i=7} PPT_i \\ \sum_{i=6}^{i=8} PPT_i \\ \sum_{i=7}^{i=9} PPT_i \\ \sum_{i=8}^{i=10} PPT_i \\ \sum_{i=9}^{i=11} PPT_i \\ \sum_{i=10}^{i=12} PPT_i \\ \sum_{i=11}^{i=1} PPT_i \\ \sum_{i=12}^{i=2} PPT_i \end{array} \right]$$

**BIO17 = Precipitation of Driest Quarter**

$$Bio17 = \min \left[ \begin{array}{l} \sum_{i=1}^{i=3} PPT_i \\ \sum_{i=2}^{i=4} PPT_i \\ \sum_{i=3}^{i=5} PPT_i \\ \sum_{i=4}^{i=6} PPT_i \\ \sum_{i=5}^{i=7} PPT_i \\ \sum_{i=6}^{i=8} PPT_i \\ \sum_{i=7}^{i=9} PPT_i \\ \sum_{i=8}^{i=10} PPT_i \\ \sum_{i=9}^{i=11} PPT_i \\ \sum_{i=10}^{i=12} PPT_i \\ \sum_{i=11}^{i=1} PPT_i \\ \sum_{i=12}^{i=2} PPT_i \end{array} \right]$$

**BIO18 = Precipitation of Warmest Quarter**

$$QT_{max} = \max \begin{bmatrix} \sum_{i=1}^{i=3} Tavg_i \\ \sum_{i=2}^{i=4} Tavg_i \\ \sum_{i=3}^{i=5} Tavg_i \\ \sum_{i=4}^{i=6} Tavg_i \\ \sum_{i=5}^{i=7} Tavg_i \\ \sum_{i=6}^{i=8} Tavg_i \\ \sum_{i=7}^{i=9} Tavg_i \\ \sum_{i=8}^{i=10} Tavg_i \\ \sum_{i=9}^{i=11} Tavg_i \\ \sum_{i=10}^{i=12} Tavg_i \\ \sum_{i=11}^{i=1} Tavg_i \\ \sum_{i=12}^{i=2} Tavg_i \end{bmatrix}$$

Where temperatures are evaluated for 12 consecutive sets of 3 months.

$$Bio18 = \sum_{i=1}^{i=3} PPT_i \text{ at } QT_{max}$$

**BIO19 = Precipitation of Coldest Quarter**

$$QT_{min} = \min \begin{bmatrix} \sum_{i=1}^{i=3} Tavg_i \\ \sum_{i=2}^{i=4} Tavg_i \\ \sum_{i=3}^{i=5} Tavg_i \\ \sum_{i=4}^{i=6} Tavg_i \\ \sum_{i=5}^{i=7} Tavg_i \\ \sum_{i=6}^{i=8} Tavg_i \\ \sum_{i=7}^{i=9} Tavg_i \\ \sum_{i=8}^{i=10} Tavg_i \\ \sum_{i=9}^{i=11} Tavg_i \\ \sum_{i=10}^{i=12} Tavg_i \\ \sum_{i=11}^{i=1} Tavg_i \\ \sum_{i=12}^{i=2} Tavg_i \end{bmatrix}$$

Where temperatures are evaluated for 12 consecutive sets of 3 months.

$$Bio19 = \sum_{i=1}^{i=3} PPT_i \text{ at } QT_{min}$$

**References:**

- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**:1965-1978.
- O'Donnell, M. S., and D. A. Ignizio. 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States. U.S. Geological Survey Data Series 691:10p.

Appendix E. Bioclimatic variables and standard deviations used for MaxEnt models

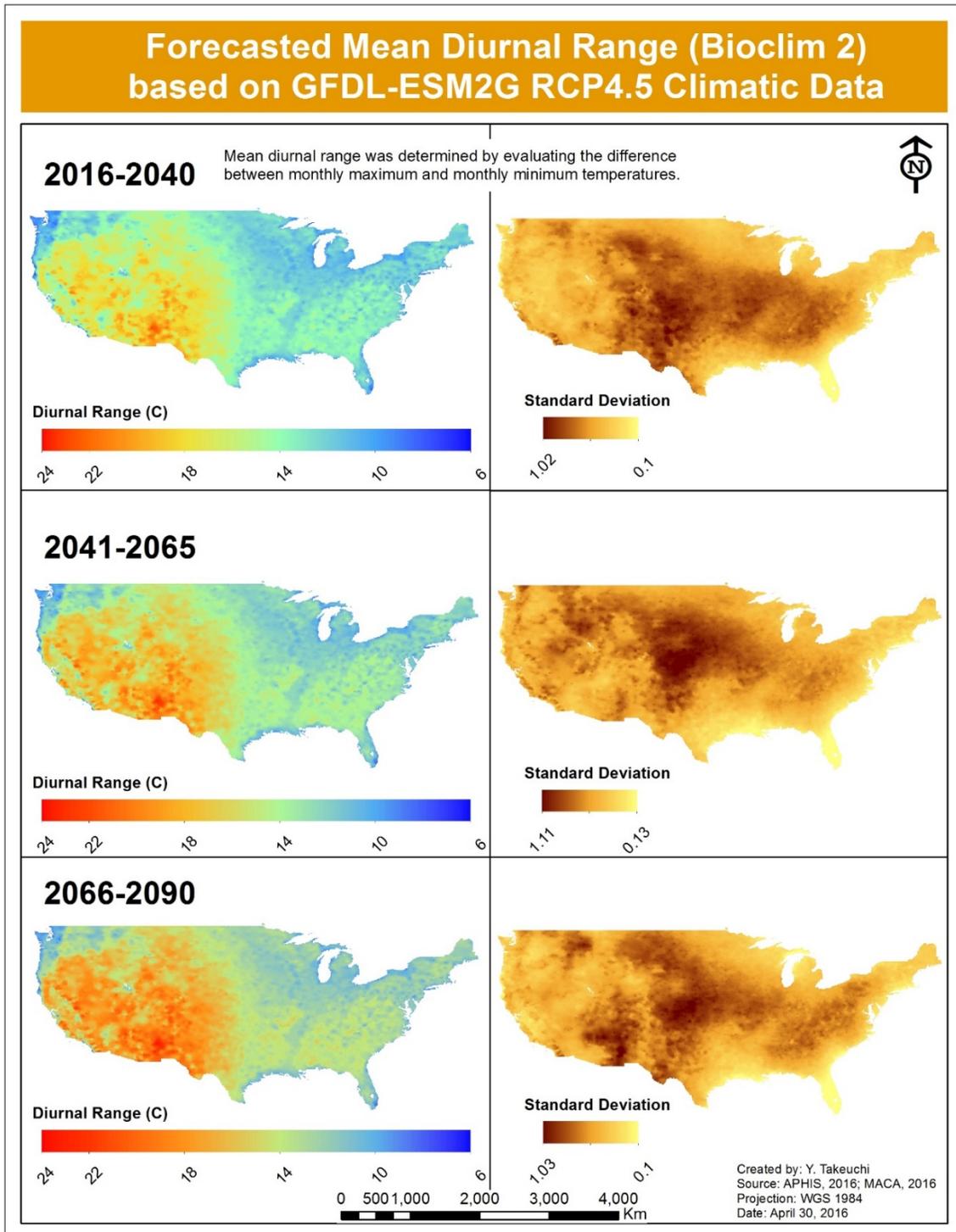
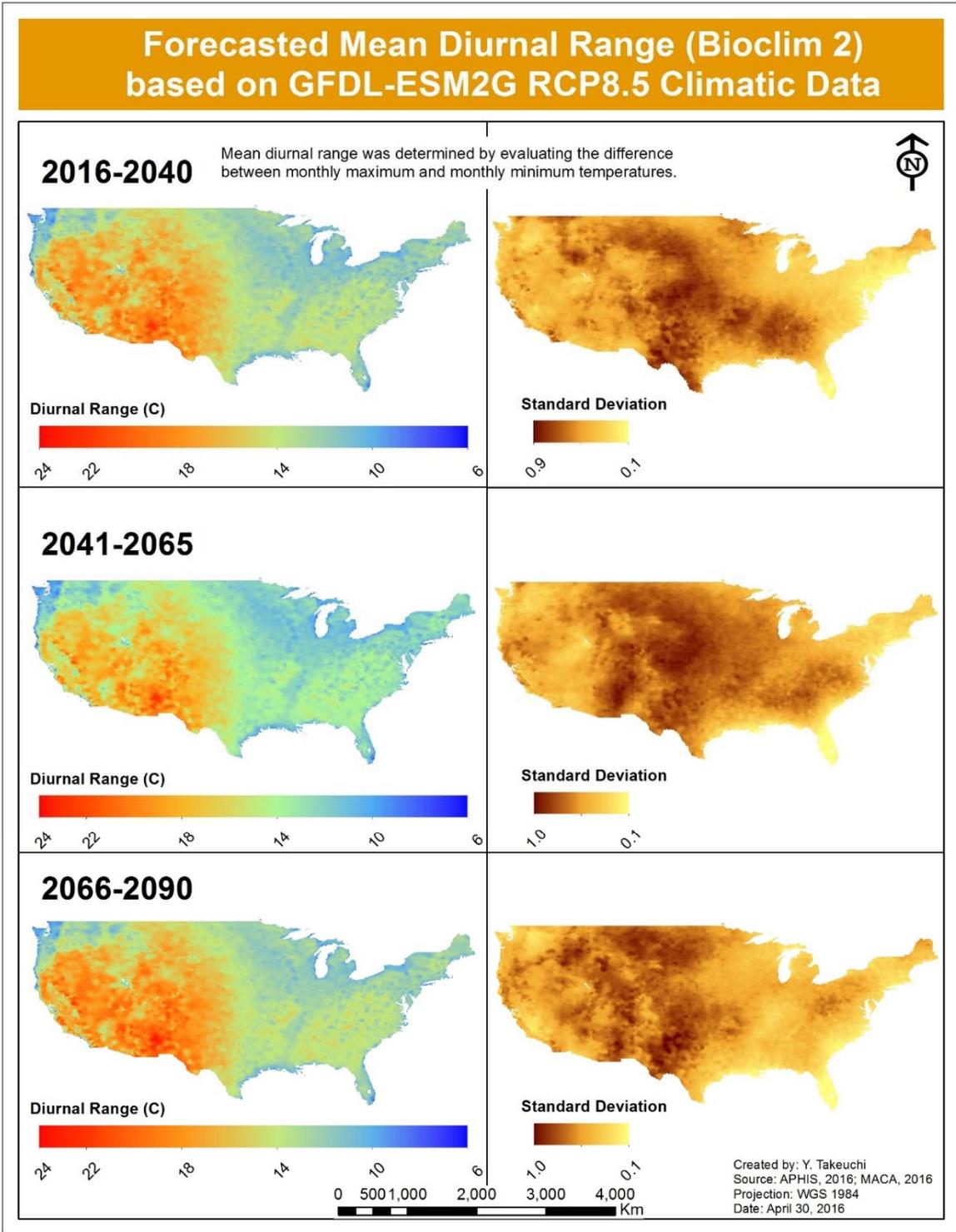
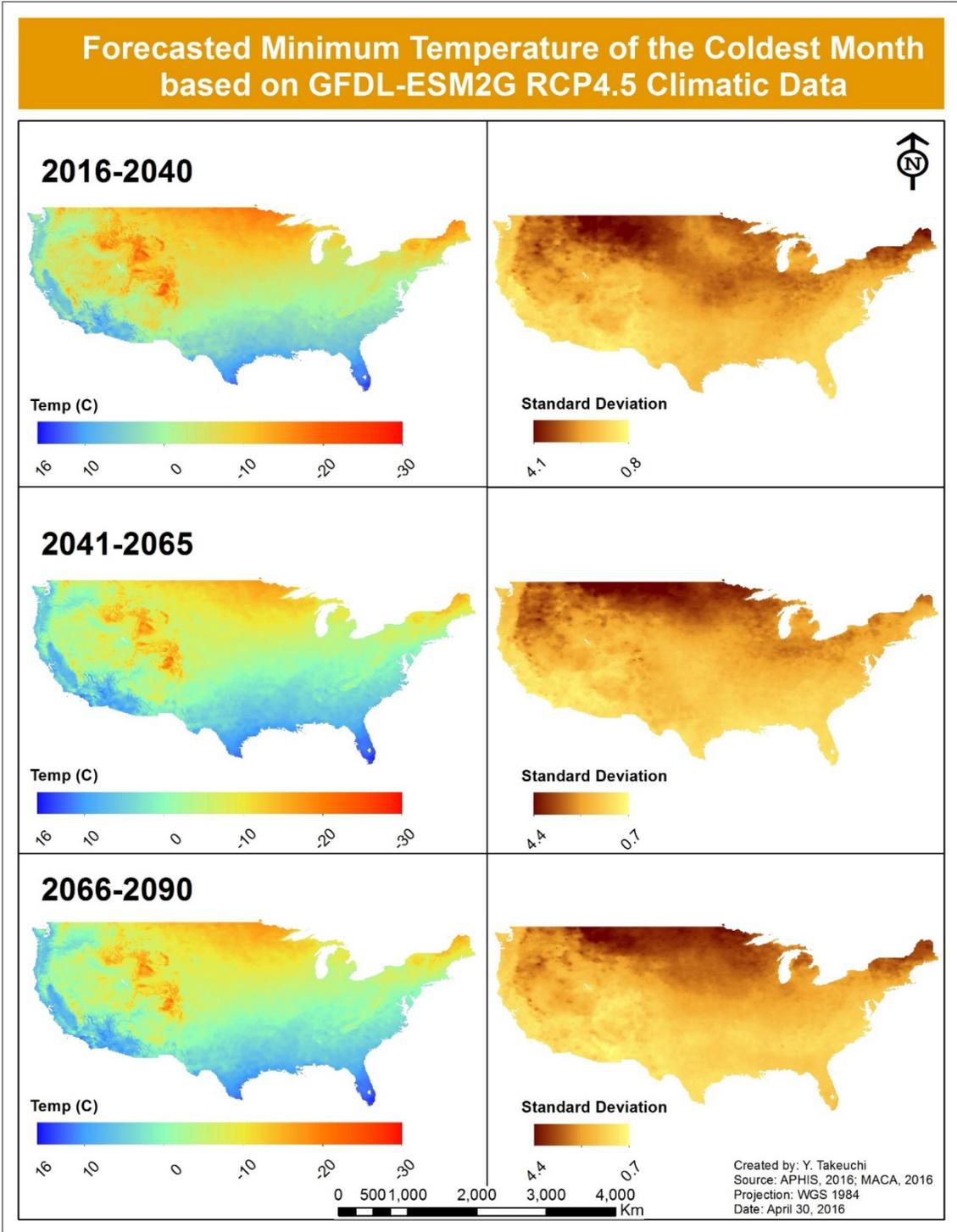


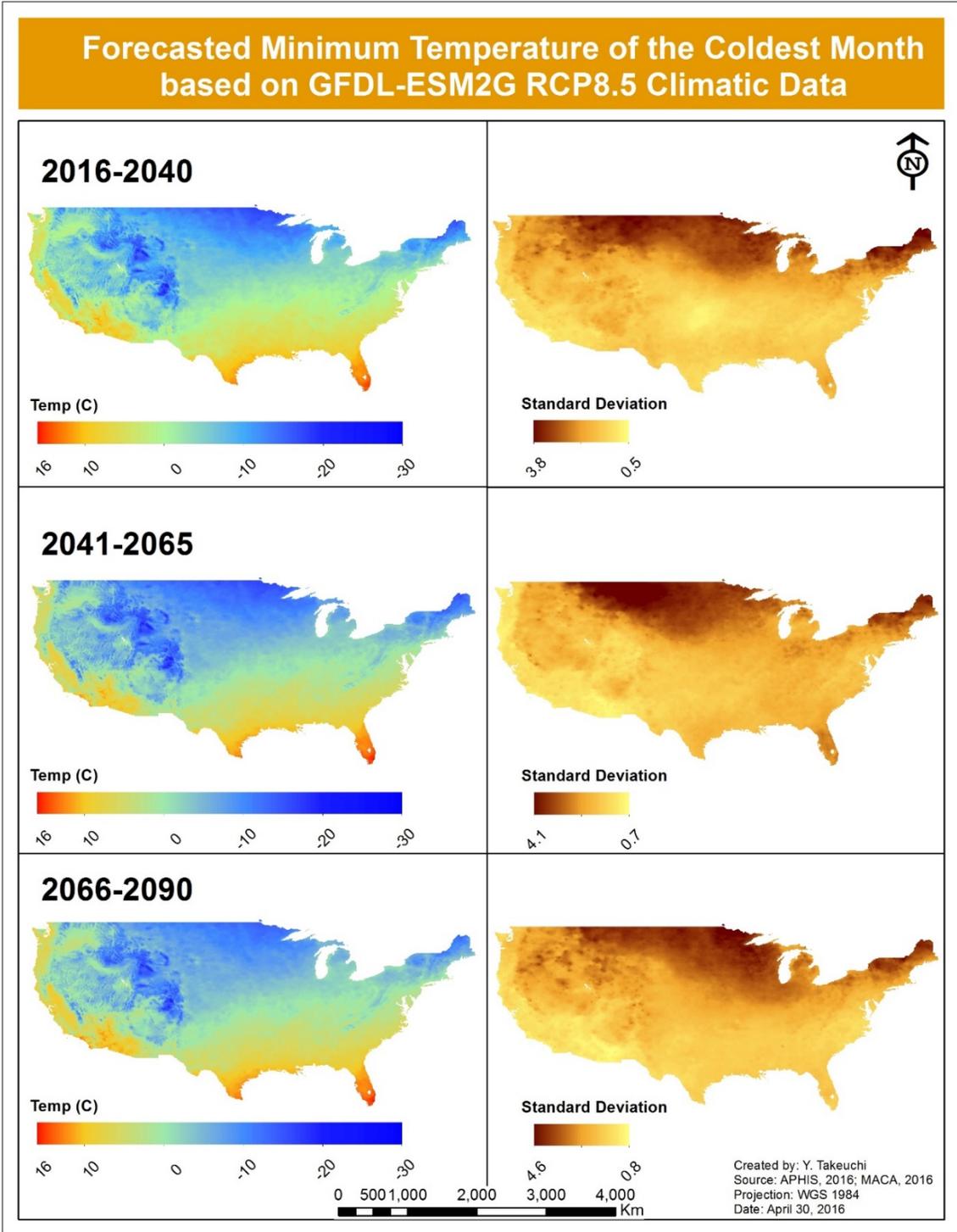
Figure A-2. Mean diurnal range (Bioclim 2) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model



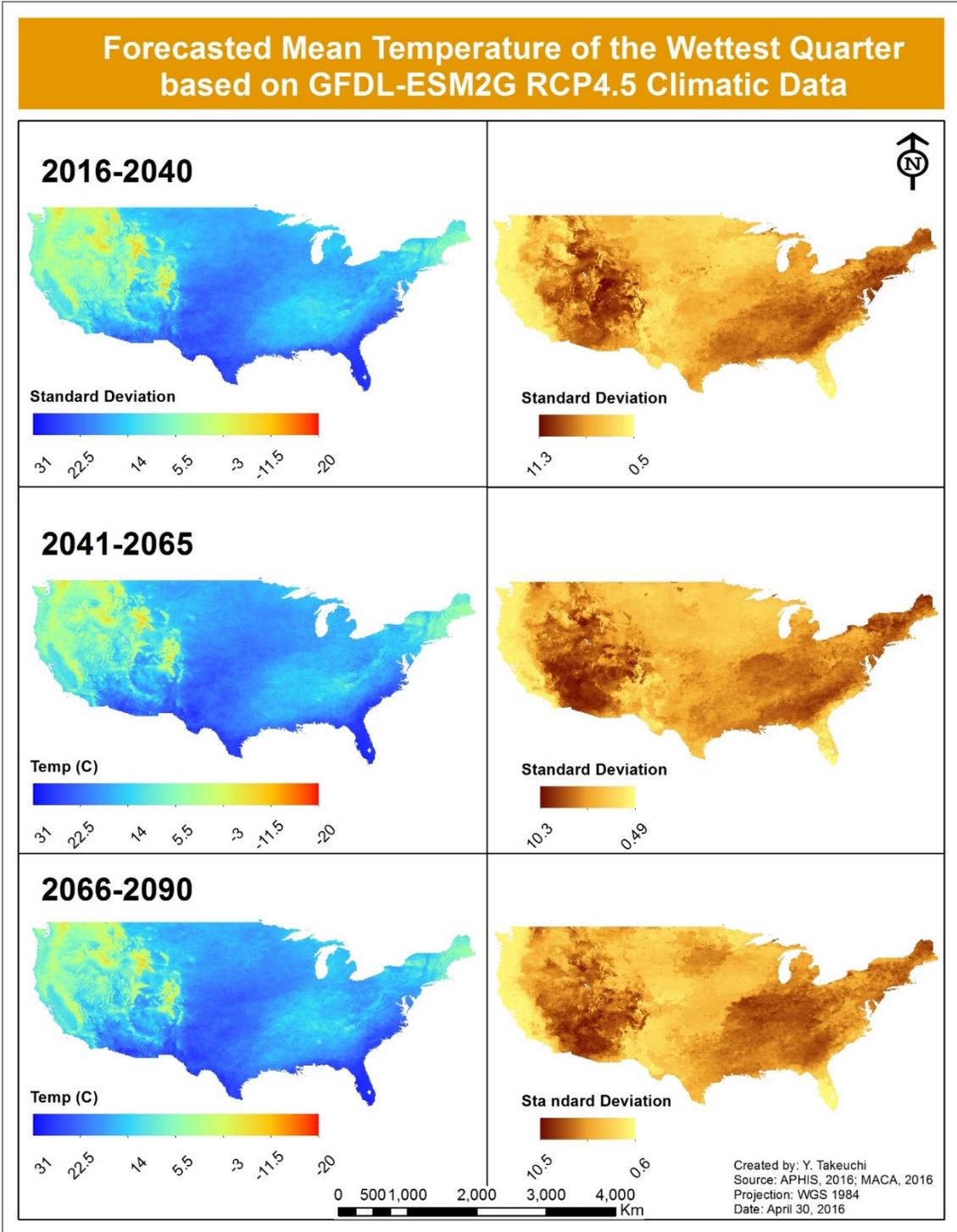
**Figure A-3. Mean diurnal range (Bioclim 2) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model**



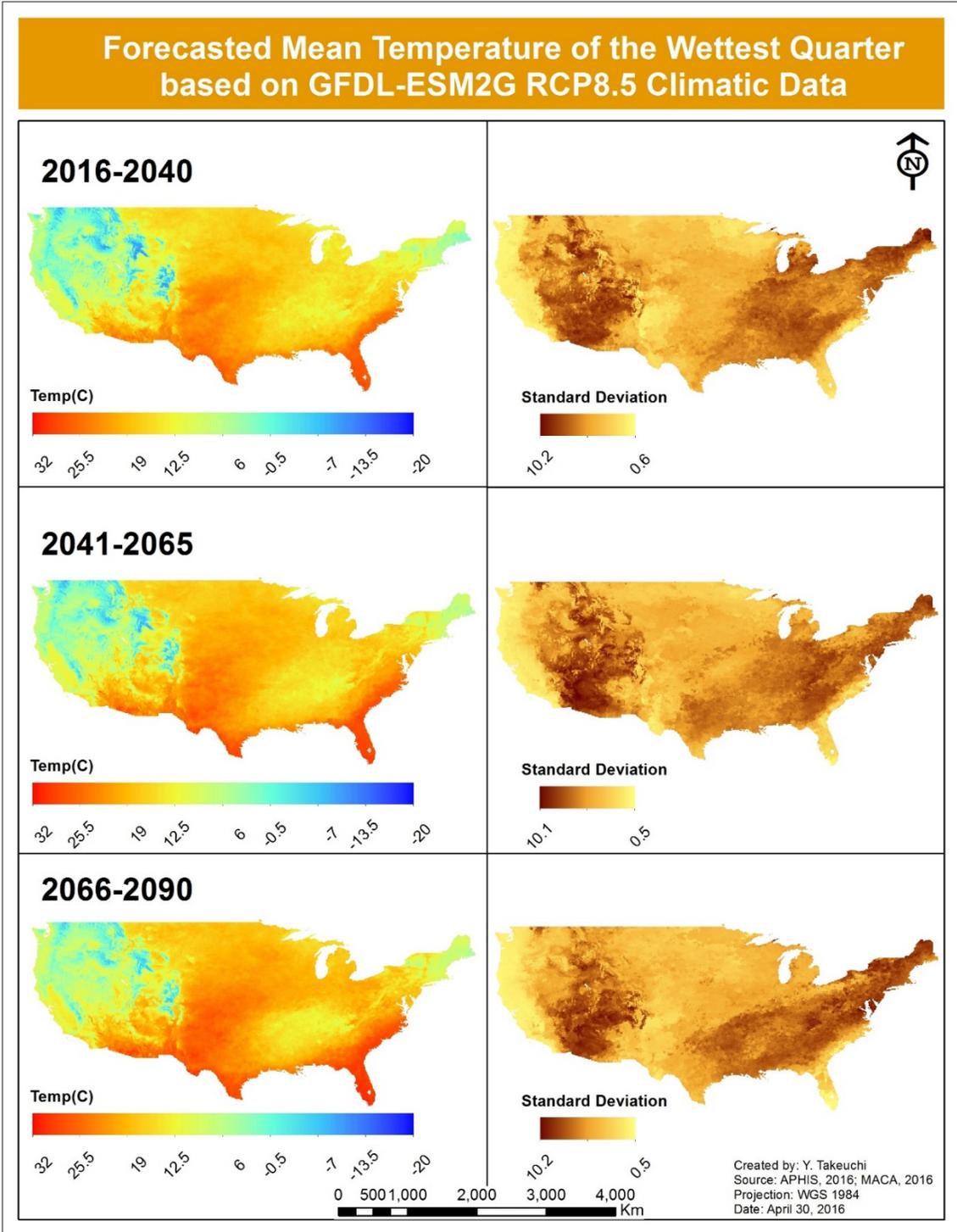
**Figure A-4. Minimum temperature of the coldest month (Bioclim 6) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model**



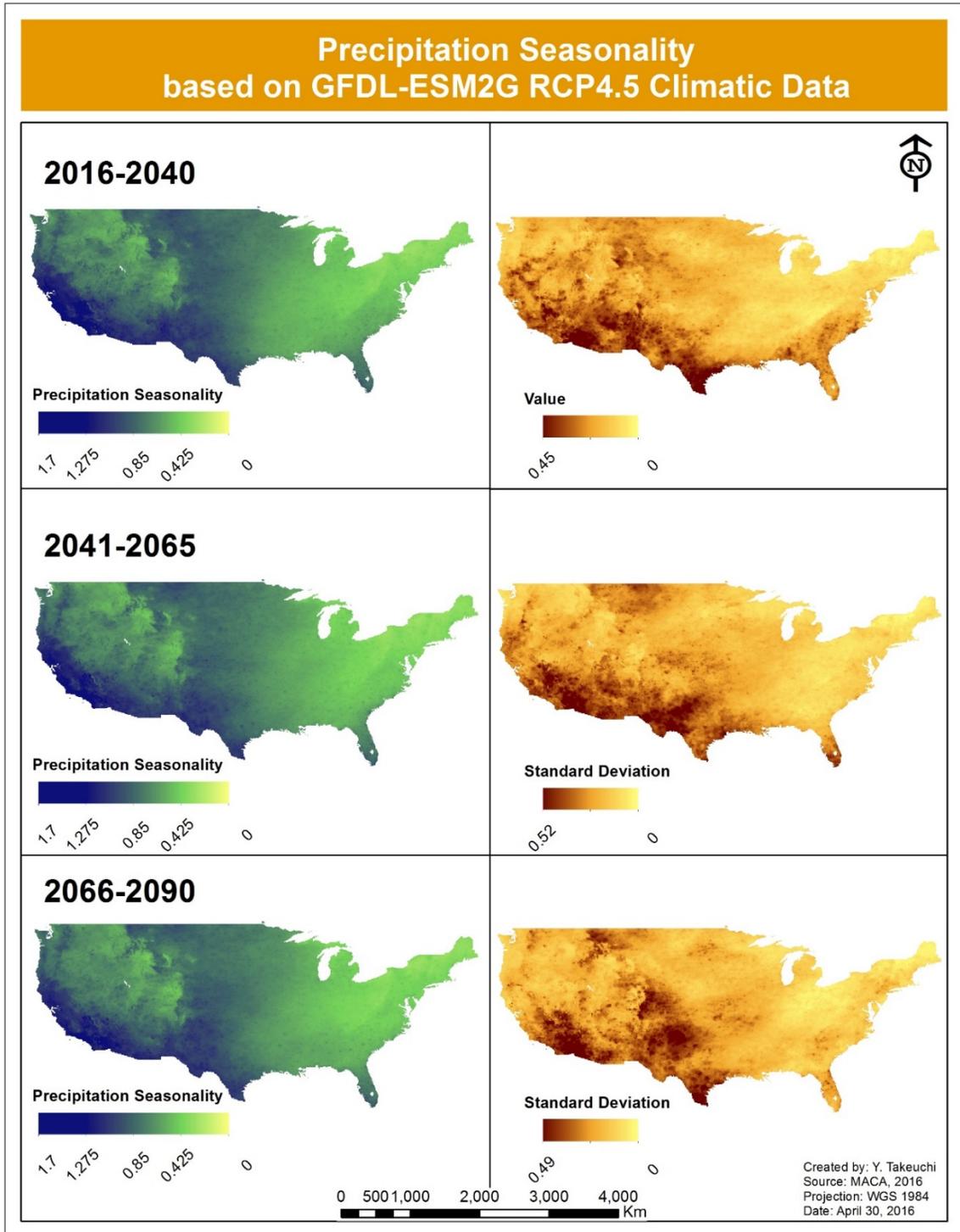
**Figure A-5. Minimum temperature of the coldest month (Bioclim 6) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model**



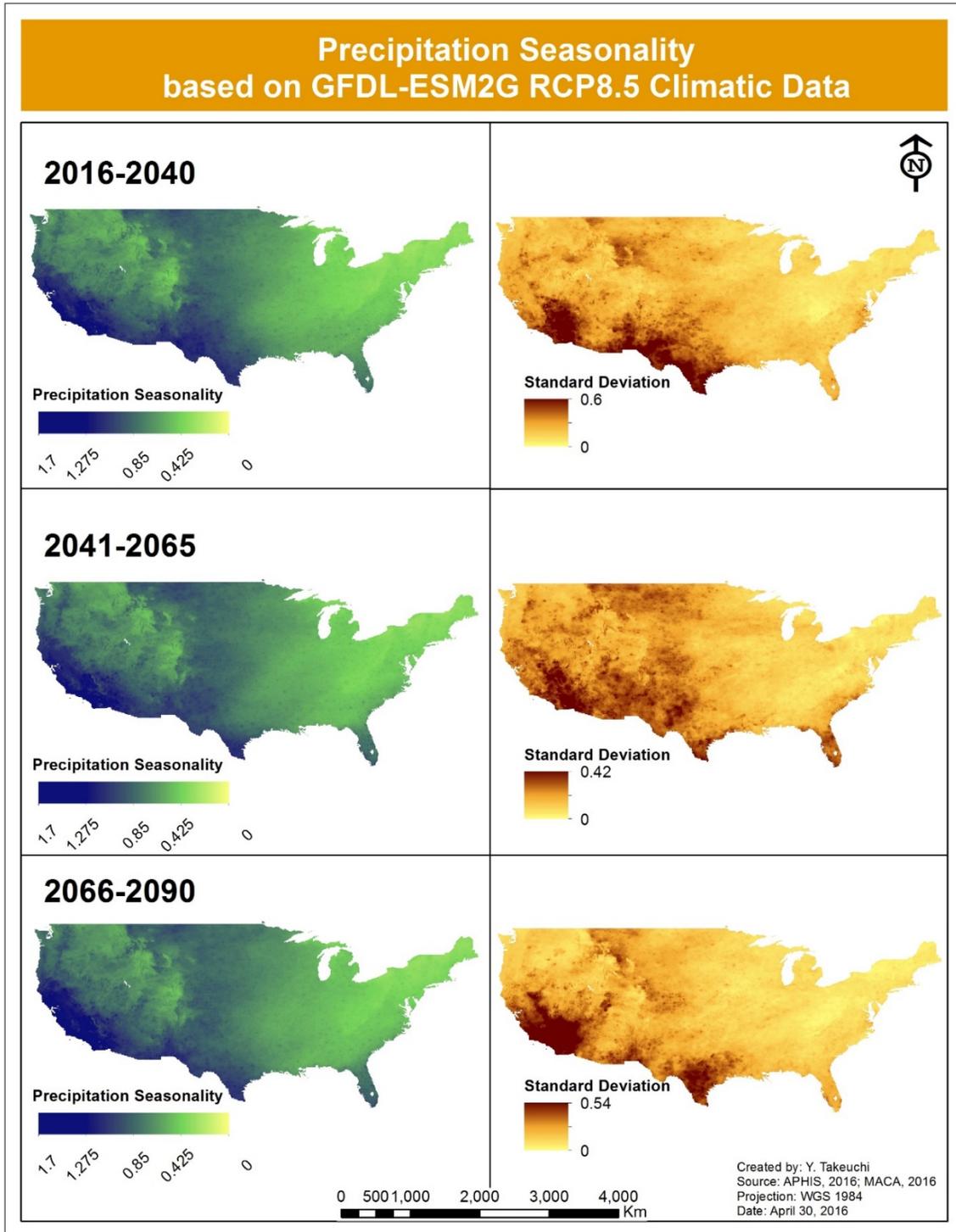
**Figure A-6. Mean temperature of the wettest quarter (Bioclim 8) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model**



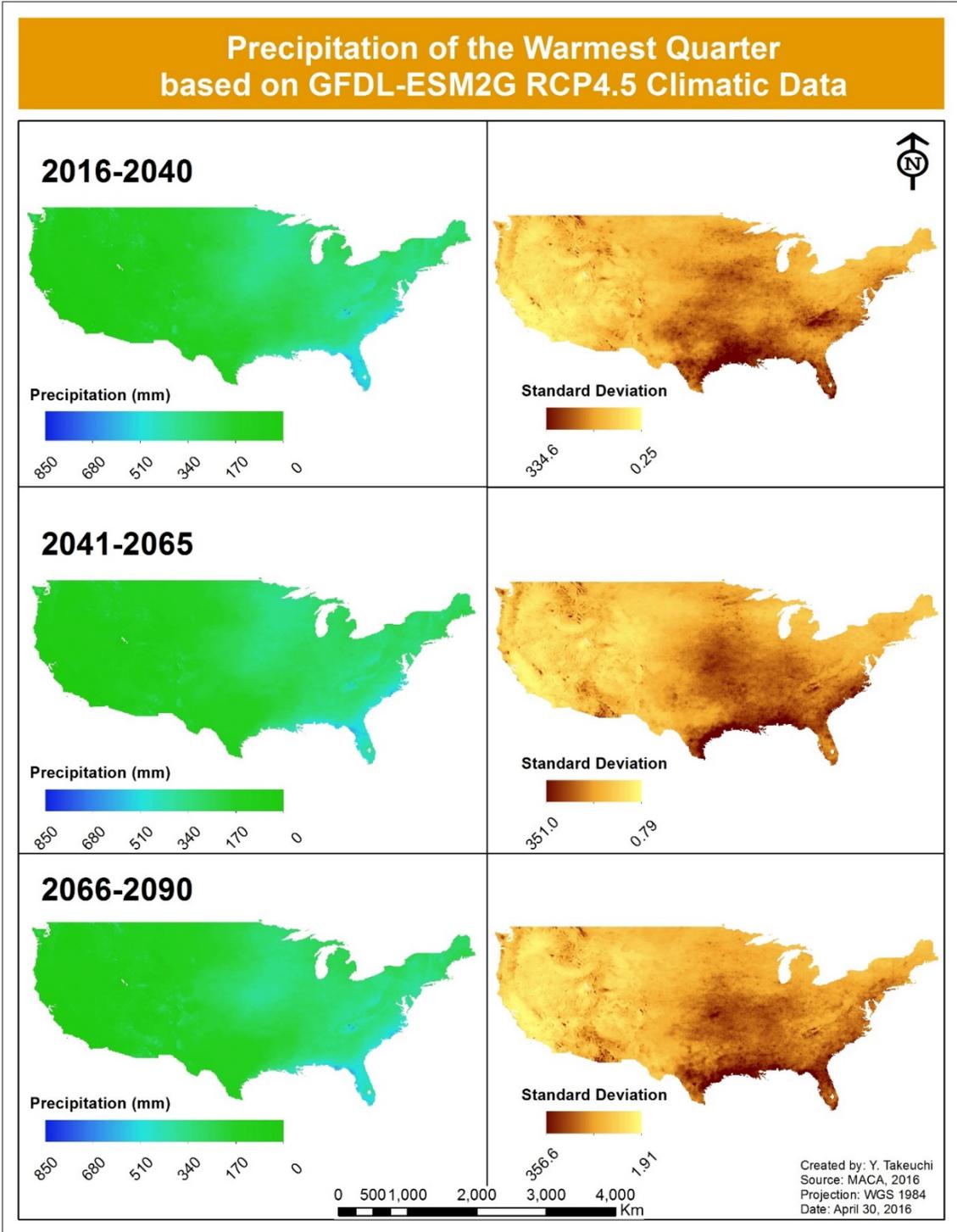
**Figure A-7. Mean temperature of the wettest quarter (Bioclim 8) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model**



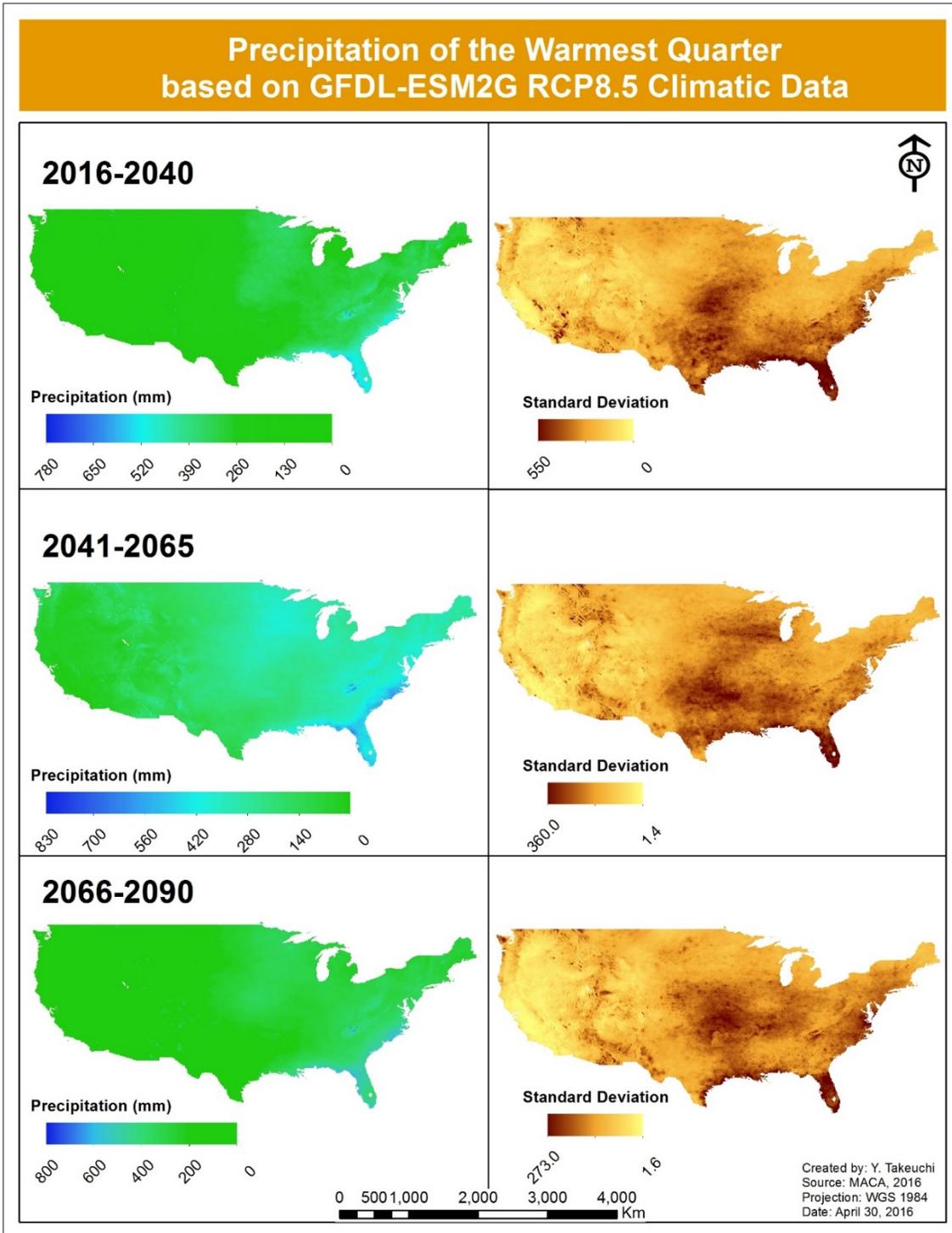
**Figure A-8. Precipitation seasonality (Bioclim 15) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model**



**Figure A-9. Precipitation seasonality (Bioclim 15) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model**



**Figure A-10. Precipitation of the warmest quarter (Bioclim 18) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model**



**Figure A-11. Precipitation of the warmest quarter (Bioclim 18) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model**

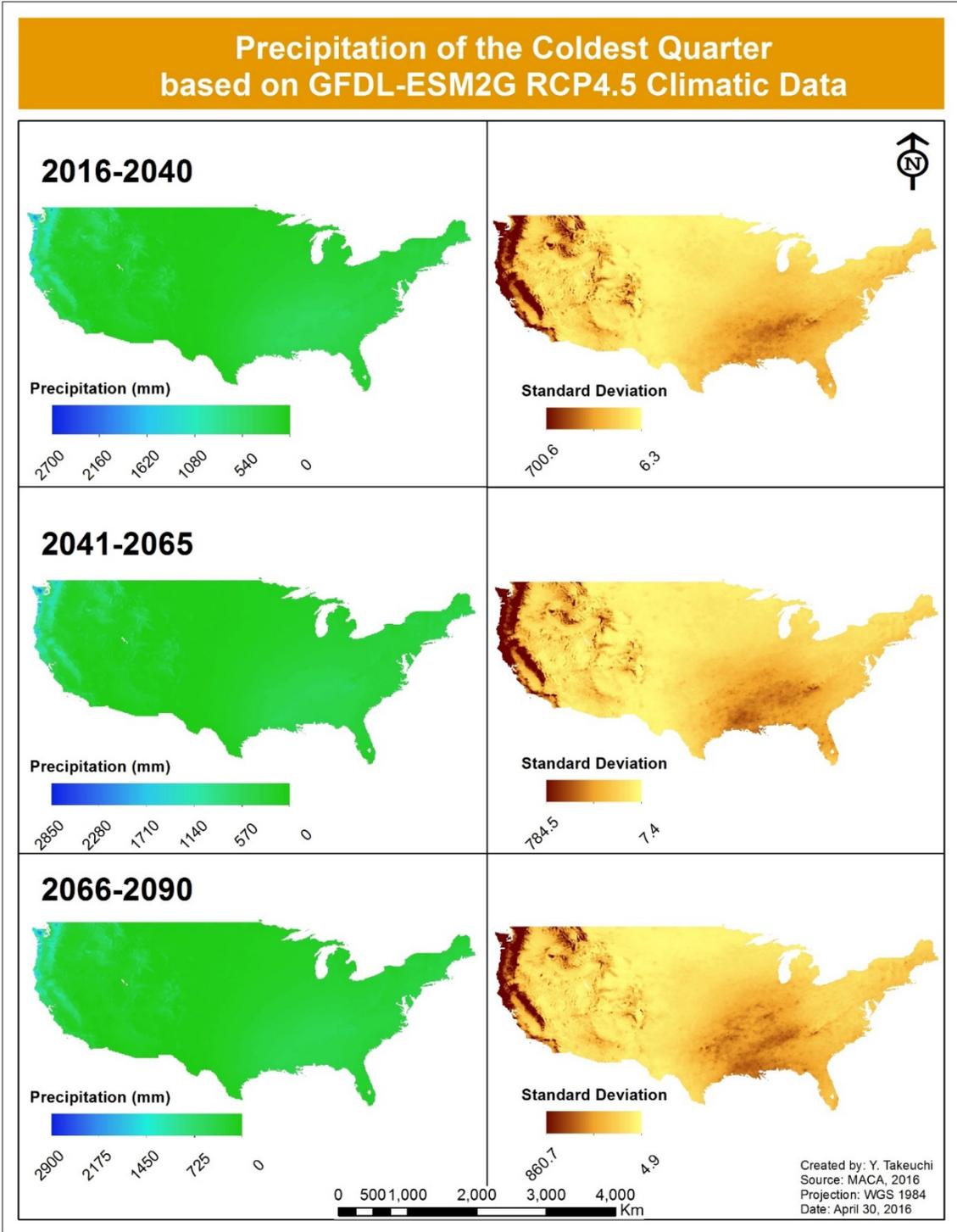
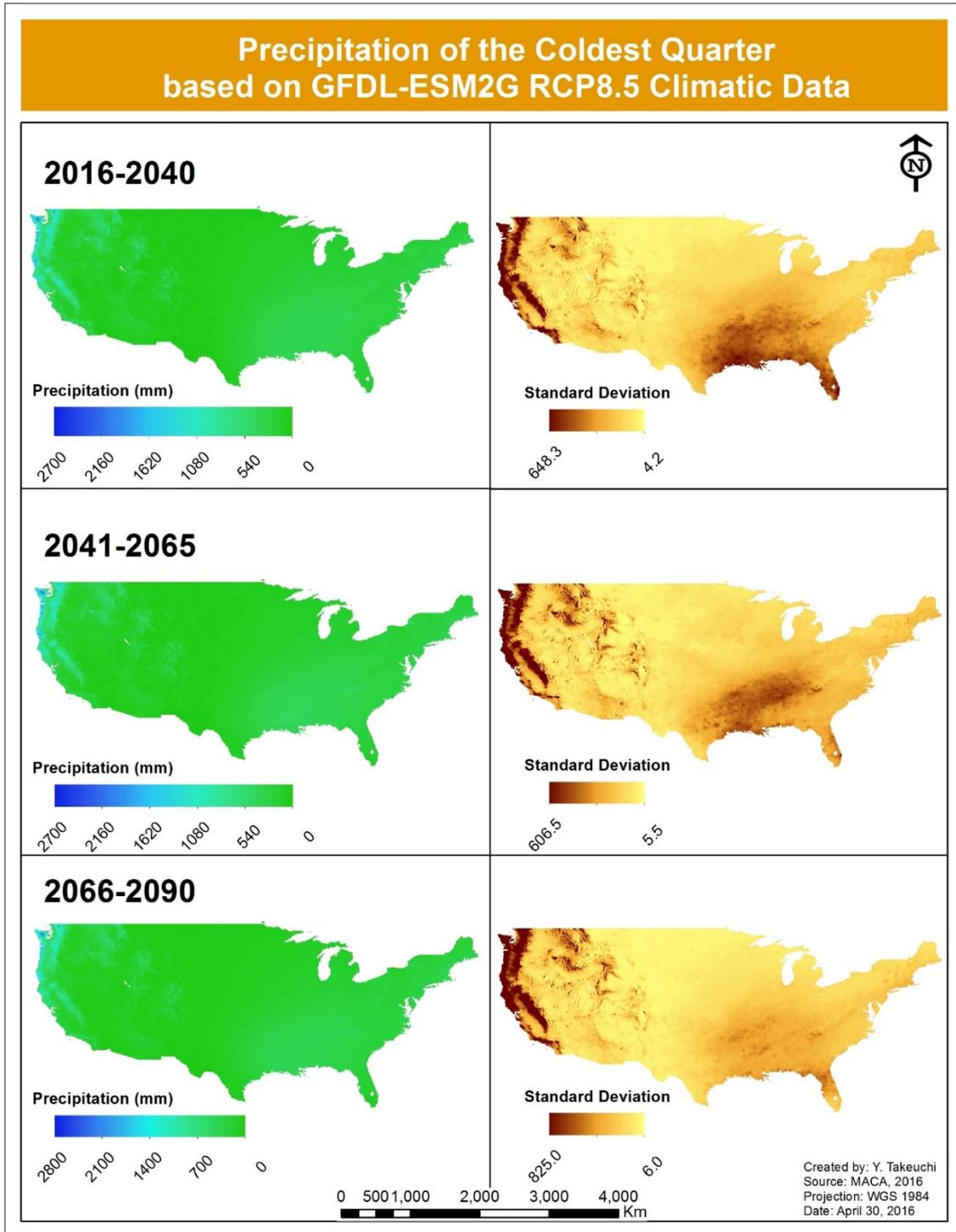
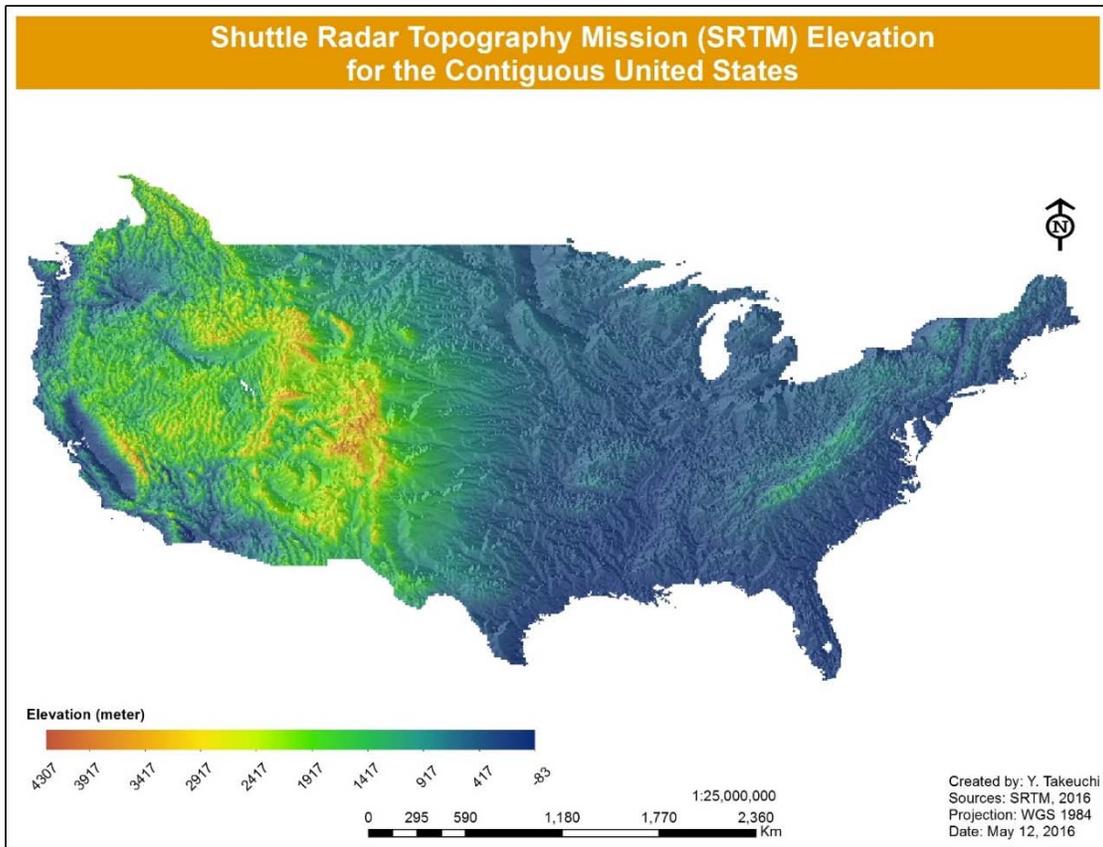


Figure A-12. Precipitation of the coldest quarter (Bioclim 19) and associated standard deviation derived from GFDL-ESM2G RCP 4.5 model



**Figure A-13. Precipitation of the coldest quarter (Bioclim 19) and associated standard deviation derived from GFDL-ESM2G RCP 8.5 model**



**Figure A-14. Shuttle Radar Topography Mission Elevation data for the contiguous United States**