CLIMATE CHANGE AND AIRBORNE ALLERGENS

By

YONG ZHANG

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Chemical and Biochemical Engineering

Written under the direction of

Panos G. Georgopoulos, Ph.D.

And approved by

New Brunswick, New Jersey

January, 2015

ABSTRACT OF THE DISSERTATION

Climate Change and Airborne Allergens

By Yong Zhang

Dissertation Director:

Panos G. Georgopoulos, Ph.D.

Climate change is expected to alter the spatiotemporal dynamics of airborne allergens and potentially increase occurrence of allergic airway disease. Climate change impact on allergenic pollen was investigated through statistical analysis and modeling of observed airborne pollen counts and climatic factors, and through simulation using a deterministic modeling system. A probabilistic exposure model was developed to study exposures to allergenic pollen during the 1990s (1994-2000) and the 2000s (2001-2010) in nine climate regions in the contiguous United States (CONUS).

The allergenic pollen seasons of representative trees, weeds and grass during the 2000s across the CONUS have been observed to start 3.0 days earlier on average than in the 1990s. The average peak value and annual total of daily counted airborne pollen have increased by 42.4% and 46.0%, respectively.

The deterministic modeling system consists of modules of emission, meteorology and air quality. It correctly predicted the observed pollen season start date and duration, and airborne level at the majority of monitor stations for oak and ragweed pollen, and performed reasonably well for birch, mugwort and grass pollen. Dry deposition, emission and vertical eddy diffusion were the dominant processes determining the ambient pollen concentrations.

The response of the allergenic pollen season to climate change varies in different climate regions for different taxa in the CONUS. Under scenarios of regionally and

economically oriented future development, the weed and grass pollen concentrations were predicted to decrease from period of 2001-2004 to 2047-2050 in the majority of regions. The number of hours in which birch and oak pollen concentrations exceed the threshold values for triggering allergy has been predicted to increase in the majority of regions.

Inhalation and dermal deposition were the dominant exposure routes for allergenic pollen. The aggregated exposure to allergenic pollen in outdoor environments was more than twice as that in indoor environments during the 2000s in the CONUS. Meantime, inhalation exposure for children of 1-4 years old was two to five times higher than for other age groups. Changes in exposures to allergenic pollen between the 2000s and the 1990s varied in different climate regions for different taxa.

Acknowledgements

I would like to express sincere thanks to my advisor Dr. Panos G. Georgopoulos for his guidance and financial support over the years of my study at Rutgers University. I deeply appreciate the opportunities and facilities he provided at the Computational Chemodynamics Laboratory for me to learn, to grow, and to prosper in my research and life.

I am also grateful to my committee members Dr. Yannis Androulakis, Dr. Marianthi Ierapetritou, Dr. Leonard Bielory and Dr. Alan Robock for their insightful comments and for taking time to guide me through my dissertation. I appreciate Dr. Bielory's help for providing the observed pollen counts from NAB-AAAI stations across the United States. I thank Dr. Robock and Dr. Ruby Leung for their guidance on obtaining and using the NARCCAP meteorology model output. I thank Dr. Androulakis from the bottom of my heart for serving as my committee member in an extremely short notice.

I thank Dr. Sastry Isukapalli and Dr. Christos Efstathiou for their help in starting my dissertation project, Climate Change and Allergic Airway Disease. I also thank Dr. Yee Chiew for his excellent teaching and continuous encouragement. I am thankful to Dr. Clifford Weisel for providing me opportunity to work on the Federal Aviation Administration project, Airliner Cabin Environmental Research.

I would also like to thank my CCL colleagues Ms. Linda Everett, Ms. Teresa Boutillette, Xiaogang Tang, Zhongyuan Mi, Dwaipayan Mukherjee, Ting Cai, Steven Royce, Jocelyn Alexander, Peter Koutsoupias, Shu Xu, Pamela Shade, Allison Patton, Leo Choi and Yanran Wang. It has been a wonderful experience to discuss and work with them at CCL. My special thanks go to Ms. Linda Everett for her help on many things in my studies and life.

My deepest gratitude goes to my wife Fang Yu and my beloved parents Tingcai

Zhang and Yuzhi Chen. I could not have pursued and finished my PhD studies without their selfless and unconditional support and patience. The priority of my life and work has always been to ensure their happiness and health in the past, present and future.

This work has been funded in part by USEPA under STAR Grant EPA-RD-83454701-0 (Climate Change and Allergic Airway Disease) to Rutgers University, by the NIEHS sponsored Center for Environmental Exposures and Disease (CEED) Grant Number: NIEHS P30ES005022, and by the Ozone Research Center which is funded by the State of New Jersey Department of Environmental Protection.

Dedication

To my wife Fang Yu
And to my parents Tingcai Zhang and Yuzhi Chen

Table of Contents

List of Tables

List of Figures

Chapter 1
INTRODUCTION

1.1 Background

Climate change has been shown to cause dramatic changes in natural ecosystems and cultivated agricultural systems, and to increase the occurrence of disease in both[5,6]. Climate trends and variations impact many prevalent human diseases such as malaria, asthma and hay fever. These climate-linked diseases have raised increasing concerns related to public health[7–9].

In particular, climate change is expected to modify the patterns of emission and transport of allergenic pollen from trees, weeds and grasses[10–13]. Like dust mites and cockroaches in indoor environments[14], outdoor allergenic pollen is one of the main triggers of allergic airway disease, affecting up to 30% of the population of industrialized countries[15]. It acts synergistically with common air pollutants, such as ozone and particulate matter, to exacerbate allergic airway disease[16], resulting in related high medical costs[17].

In addition, gene flow through pollen transport is a key determinant of impacts of a variety of plantations (e.g., forest trees) on surrounding plant populations and ecosystems. Long range transport and growing production of pollen have raised concern of cross pollination near genetically modified plants[18].

Understanding the spatiotemporal patterns of changes in pollen season timing and levels is thus important in assessing climate impacts on allergic airway disease and cross pollination[19–21]. Changes in temperature and precipitation have been and will be heterogeneous, and enhanced warming and precipitation are very likely to occur at higher latitudes[22]. Even in the vicinity of a single locality, different taxa are observed to respond differently to climate change[23]. Studies on allergenic pollen from multiple taxa are needed to elucidate climate impacts on allergenic pollen and potential consequences on public health. These studies could be either analyses of observed airborne pollen data or modeling simulations in a large geographic area spanning different climate regions.

1.2 Review of Studies Focused on Climate Change and Allergenic Pollen

The recent US National Climate Assessment looks at all recent studies on how allergenic pollen season has changed with change in climate, and shows studies of some regions for some taxa [24]. Most analyses of observed airborne pollen data on assessment of climate change effects on allergenic pollen season have involved individual or a few taxa at a single or limited number of locations [25,11,13].

Modeling studies on allergenic pollen and climate change can be generally classified into two categories. The models of the first category are basically statistical relationships based on regression between observed phenology, aerobiology and factors of climate and/or meteorology [26]. The second category usually consists empirical or mechanistic models constructed based on existing air quality or meteorology modeling systems [27,28].

Table 1.1 summarizes the statistical or empirical modeling studies focused on linking climate change and pollen season timing and airborne levels. Dahl *et al.* studied the effects of climate change on birch pollen from 1979 to 1994 by regressing annual production on three variables: (1) annual production during the inception year, which is the year prior to flowering season when the male catkins are initiated; (2) hourly temperature accumulation from May 1st to July 20th in the inception year; and (3) hourly temperature accumulation during the main pollen season in the flowering year [29]. They found that hourly temperature sum and annual production in the prior year are important variables to predict annual production in the current year. Based on regression analysis of start dates and mean monthly temperatures from 1997 to 2000, Emberlin *et al.* reported that birch pollen season tends to start earlier in Europe, and that it is closely related to mean monthly temperatures in January, February, March, April and May [30].

Trend analysis on start date, peak date, annual production and peak value performed by Rasmussen and Yli-Panula *et al.* showed that birch pollen season in both Denmark and Finland tend to start earlier with a rising annual production [31,32]. Their studies also showed that monthly temperature and precipitations are key climatic factors influencing pollen season timing and airborne level. Statistical analyses on birch pollen were also performed to examine other climatic factors such as sunshine hours [33] and other pollen indices such as season length.

Similar trend and regression analysis on allergenic pollen of other species, such as oak, Platanus [34], ragweed [35] and grass [36], were also conducted to identify their trends and relationships with climatic factors. Stepwise regressions were used by Garcia-Mozo *et al.* to study relationships among start dates, annual production, daily concentrations of oak pollen, and monthly temperatures and precipitation in multiple regions of Spain [37]. The above relationships were then combined with future meteorology data from the Regional Climate Model developed by the Hadley Center to determine the pollen season timing and airborne levels in future years. They showed that pollination season could start on average one month earlier and airborne pollen concentration could increase by 50% at the end of the 21st century.

Table 1.2 is adapted from Efstathiou *et al.* to summarize modeling studies that have focused on large-scale emissions and long-range transport of pollen [38]. Kawashima *et al.* constructed an emission model by regressing the airborne pollen data with hourly air temperature and wind speed [39]. The emission model was then coupled with a Eulerian-type diffusion model and wind speed extrapolated from meteorology stations to simulate the transport of cedar pollen. Schueler *et al.* developed an emission module by fitting the airborne pollen count and flowering time using a fourth-order polynomial curve [40]. This emission module was then incorporated into the Meteorological Institute Mesoscale Model to simulate the dispersion of oak pollen. Results from their studies showed that mesoscale atmospheric models are applicable to simulate pollen dispersal in a large geographical area.

Without explicit emission modules, Lagrangian algorithms were used by some researchers to track the movement of pollen particles or puffs containing pollen. Starting

Table 1.1: Statistical or empirical studies focused on linking the climate change and pollen timing and quantity.

Study and Location	**Period**	**Plant Species**	**Model Format**	**Climatic Factor**	**Pollen Index**	**Prediction for Future**
Sweden [29]	1979-1994	Birch	Multiple regression	Temperature	Annual Production	No
Switzerland [36]	1969-1996	Hazel, birch, grass	Simple regression	Temperature, Precipitation	Start Date, Peak Date, Annual Production	No
Denmark [31]	1977-2000	Birch	Simple regression	Temperature, Precipitation	Start date, Annual Production, Peak Value	No
Europe [30]	1970-2000	Birch	Simple regression	Temperature	Start Date	No
Spain, Italy [34]	1982-2003	Platanus	Simple regression	Temperature	Start Date	No
Spain [37]	1992-2004	Oak	Stepwise regression	Temperature, Precipitation	Start Date, Annual Production, Daily Concentration	Yes
Switzerland [10]	1969-2006	Birch	Simple regression	Temperature	Start Date, Annual Production	No
Finland [33]	1989-2006	Birch	Stepwise regression	Temperature, Sunshine Hours	Annual Production	No
Finland [32]	1974-2004	Birch	Simple regression	Temperature, Precipitation	Start Date, Season Length, Annual Production, Peak Value	No
North America [35]	1995-2009	Ragweed	Stepwise regression	Frost Free Days	Season Length	No
Europe [43]	1977-2009	Multiple taxa	Trend, correlation analysis	Temperature	Annual Production	No
Australia,New zealand [44]	1988-2011	Multiple taxa	Temporal analysis, comparison	Temperature, Precipitation, Land use	Pollen Count	No
United States [45]	1994-2010	Multiple taxa	Trend, correlation analysis	Frost Free Days,Growing Degree Days	Annual Production, Start Date, Season Length, Peak Value	No

from the measured pollen count, Aylor *et al.* used a Lagrangian stochasic model and measured meteorology data to calculate concentrations and fluxes of maize pollen [41]. Based on area coverage of oak trees from the Biogenic Emissions Landuse Database, version 3.1 (BELD3.1), Pasken *et al.* assumed a uniform diurnal profile of pollen emission [42]. The trajectories of emitted oak pollen particles and puffs containing pollen were then simulated using the National Center for Atmospheric Research/ Penn State Fifth Generation Mesoscale Model (MM5) and the National Oceanographic and Atmospheric Administration (NOAA) Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) Model. It was indicated from their results that a better understanding of pollen release is critical to improving pollen concentration forecasts.

Helbig *et al.* developed an empirical emission model which consists of a characteristic concentration, a lumped meteorology adjustment factor and a characteristic velocity [46]. The characteristic concentration is parameterized using annual total emission flux and a characteristic length. The lumped meteorological adjustment factor is related to humidity, wind speed and temperature. This empirical emission model was then coupled with mesoscale meteorology and air quality modeling system to simulate emission and transport of hazel and alder pollen. Later, the empirical emission model was adjusted by Vogel *et al.* to calculate emission and transport of birch pollen using an operational weather forecast system [47]. It was also modified by Efstathiou *et al.* to predict emission and transport of birch and ragweed pollen using the Community Multiscale Air Quality (CMAQ) model and meteorology model MM5 [38].

The latest studies of modeling pollen emission and dispersion focused on development of a deterministic modeling system. Zhang *et al.* developed a framework incorporating the Weather Research and Forecast (WRF) model and CMAQ model to study distributions of multiple allergenic pollen in southern California under climate change scenarios [48,49]. Sofiev *et al.* simulated birch pollen emission and transport in Europe using the European-scale operational System for Integrated modeLing of Atmospheric coMposition (SILAM) [50,51]. Deterministic modeling systems have also been developed and applied to provide operational forecast of ragweed pollen concentration in Europe [52,53]. These modeling studies provided a very good starting point toward

Table 1.2: Modeling studies focusing on large-scale emissions and long-range transport of pollen

Study and Location	Resolution & vertical layers	Plant Taxon	Emissions Model	Vegetation Data / Model	Meteorology Data / Model	Transport Model
Japan [39]	(10x10 km), 2-D	Cedar	Meteorological parameterization	Forest maps and remote sensing	AMeDAS weather stations	Gaussian model
Germany [46]	(4x4 km), 35	Hazel, Alder	Empirical model	Forest maps	KAMM 3-D	DRAIS CTM
USA [42]	(12x12 km), 8	Oak	Uniform diurnal profile	BELD3.1 database	MM5, Eta 3-D	HYSPLIT
Finland [54]	(1x1 km), single layer	Birch	Aerobiology observations	CORINE, PELCOM, Forest surveys	SILAM	SILAM
Germany [40]	(500x500 m), 32	Oak	Meteorological parameterization	Forest maps	METRAS	METRAS
Switzerland [47]	(7x7 km), 40	Birch	Empirical model	National Forest Inventory	COSMO-ART	COSMO-ART
NorthEastern USA [38]	(12x12 km), 22	Birch, Ragweed	Empirical model	BELD3.1, PLANTS, MODIS LAI	MM5	CMAQ
California, USA [48,49]	(12x12 km, 4x4 km), 29	Multiple Taxa	STaMPS	NASS of USDA, tree inventory	WRF	CMAQ
Europe [51,50]	($0.25° \times 0.25°$), 40, 74	Birch	Probabilistic model	CORINE, PELCOM, Forest surveys	ECMWF, HIRLAM	SILAM
Central Europe [52]	(7x7 km), 40	Ragweed	Empirical model	Empirical model	COSMO-ART	COSMO-ART
Europe [53]	($0.25° \times 0.25°$), 8	Ragweed	Empirical model	Ecological model	ECMWF	SILAM
This study, USA	(50x50 km, 12x12 km), 34	Multiple Taxa	Mechanistic model	BELD3.1, Empirical model	WRF	CMAQ

operational forecasts of daily pollen levels. Further improvements on emission modeling can be made to incorporate information such as detailed vegetation maps, the spatial distribution of start and length of pollen season, and diurnal flowering likelihood.

In terms of modeling climate change effects of spatiotemporal dynamics of aeroallergens, there are major deficiencies in existing modeling systems. First, there are no links among climate change, emission and transport of allergenic pollen in existing modeling systems. Most of the emission parameters, such as start dates, season lengths and annual total emission flux, are static without considering long term influences of multiple climatic factors. Simulated future meteorology profiles from regional meteorology models and global climate models are not yet utilized to drive the pollen emission and transport models.

Second, the existing emission models are either based on regressed relationships or empirical formulations. They were not constructed mechanistically based on first-principle physics. The existing emission models were generally parameterized using only part of the information from physics, aerobiology, phenology and meteorology. For example, diurnal emission patterns cannot be captured because hourly flowering likelihood is not incorporated. Values of some emission parameters are based on assumptions due to scarcity of measurement. As a result, pollen concentration estimations generated by most of existing modeling systems are qualitative or semi-quantitative [54].

Third, pollen removal processes are not fully studied. In existing modeling systems, pollen particles or puffs which contain pollen are usually treated as inert substances, whose concentrations are usually determined by only considering processes of deposition and advection. The contributions of each physical processes on ambient pollen concentration have not been fully instigated.

Last, but not least, statistical analysis of observed airborne pollen and climate data is usually limited to simple trend analysis and regressions. On one hand, these trend and regression analyses have usually targeted a specific monitoring station or region. On the other hand, these analyses are generally for a single pollen index and a single climatic factor. A generalized statistical relationship needs to be derived based on multiple pollen indices and multiple climatic factors from multiple regions.

1.3 Main Hypotheses

Recorded changes in pollen season timing, such as start date, peak date, and season length, and airborne pollen levels, such as annual production, mean and maximum daily concentrations, are most probably due to the observed changes of multiple climatic factors, such as increasing temperatures and rising CO_2 levels. Future pollen season timing and airborne levels will be altered by the expected climate change in the coming decades reported by the Intergovernmental Panel on Climate Change (IPCC).

Increasing temperature will trigger an earlier onset of pollination of allergenic trees, weeds and grass; and increasing warming period will lead to longer pollen seasons. Rising CO_2 levels will favor evolutional and productive plant growth, and potentially increase pollen production. Emission and transport of pollen grains are impacted by multiple climatic factors such as temperature, precipitation, humidity, friction velocity and wind speed.

Diversified responses of pollen season timing and airborne level to climate change are expected for different species. For the same species, patterns of pollen season timing shifting and pollen level variation will be different for different regions due to natural variability existing in both climate and plant growth.

Pollen indices are normally distributed variables which fluctuate around mean trends depending on the combination of multiple random climate/meteorology factors. Pollen of the same genus has similar responses to climate/meteorology changes, such as variations in CO_2 levels, temperature, precipitation, humidity, friction velocity and wind speed.

1.4 Objectives of the Thesis

The hypotheses mentioned above were examined by the following studies. Statistical analyses were carried out to analyze the trends of pollen season timing and levels, and their relationships with multiple climatic factors based on observed airborne pollen count and climatic data. Patterns of pollen season timing shifting and pollen level variation were identified by analyses of airborne pollen data, geographical data and climate data in different periods and regions. A mechanistic emission model was developed and combined with an adapted air quality model to simulate spatiotemporal dynamics of pollen timing and levels. Statistics based on simulation results were derived to investigate the responses of pollen season timing and airborne levels to expected climate changes.

The fundamental questions this work tries to answer are:

- How have the allergenic pollen seasons responded to the changing climate in the past two decades?
- How will the expected climate change impact the spatiotemporal distributions of aeroallergens and their population exposures?
- When will the allergenic pollen seasons start, and how long will they last for future years under climate change scenarios?
- What will be the spatiotemporal emission and airborne concentration profiles of five representative allergenic pollen, which are birch (*Betula*), oak (*Quercus*), ragweed (*Ambrosia*), mugwort (*Artemisia*) and grass (*Poaceae*)?

The specific objectives of this work are:

- To conduct case studies to determine how climate change will impact the spatiotemporal distributions of aeroallergens and their population exposures by integrating information on (a) emissions of pollen (b) meteorology (c) land cover and land use and (d) population demographics,
- To develop the statistical modeling system using Bayesian statistics and machine learning models for (a) analysis of climate change impact on allergenic pollen season based on observed climate and pollen data, and for (b) identification of the empirical relationships among the observed pollen season start date, duration, ambient level, and the observed meteorological, phenological and geographical factors,
- To develop a regional deterministic modeling system based on existing phenology, emission, meteorology and air quality models for studying production, emission

and dispersion of multiple airborne allergenic pollen of five representative species: birch (*Betula*), oak (*Quercus*), ragweed (*Ambrosia*), mugwort (*Artemisia*) and grass (*Gramineae*), and

- To simulate population exposures to multiple airborne allergenic pollen through multiple routes such as inhalation, dermal contact and ingestion by developing new modules and adapting existing exposure models.

1.5 Outline Of The Thesis

The overall modeling framework was developed through adjustment and incorporation of the WRF model, the Sparse Matrix Operator Kernel Emissions (SMOKE) model, the CMAQ model, and the statistical learning models. These models represent state-of-the-science in fields of meteorology, emission and air quality simulations. They were designed in different modules which could be easily assembled for different configurations and applications. Multiple databases are available to be used as input to drive these models. Figure 1.1 presents schematically the overall WRF-SMOKE-CMAQ-Pollen modeling system. Major databases used in the current study are summarized in Table 1.3. Information about each model step is listed in Table 1.4. The modules and databases in the diagram of Figure 1.1 are explained accordingly in the following chapters.

Observed climate and pollen data were analyzed to identify the effects of historical climate change on pollen season timing and airborne levels, and to provide parameterization for Bayesian analysis, machine learning models and pollen emission module. IPCC scenario A2 was used to drive a General Circulation Model (GCM) and meteorology model WRF, and to provide information on Land Use and Land Coverage (LULC). The emission scenarios in the fourth assessment report of IPCC have been replaced by Representative Concentration Pathways in the fifth assessment report [22]. The WRF meteorology data were processed to provide input for emission and transport modules to simulate the spatial and temporal distributions of allergenic pollen. Time series of airborne pollen levels could be simulated mechanistically from CMAQ or statistically from observed pollen counts during 1994-2010 in nine climate regions. The simulated time series of airborne pollen concentrations were then combined with information of activity pattern and exposure route to simulate the general population exposures to pollen under climate change scenarios.

The selected meteorology dataset covers the whole region of North America for historical years 2001-2004 and future years 2047-2050. These data have been evaluated and archived by the NARCCAP. Its temporal resolution is three hours, and spatial resolution is 50 x 50 km with 34 vertical layers. The meteorology dataset was selected because (1) it has been validated and published by the North American Regional Climate Change Assessment Program (NARCCAP); and (2) it is available for multiple historical and futures years. An emission dataset and a meteorology dataset for 2007 from the US Environmental Protection Agency (USEPA) and the New Jersey Department of Environmental Protection (NJDEP) were also selected for demonstration of simultaneous simulation of allergenic pollen and air pollutants, such as ozone and particulate matter, through the developed modeling system. These two datasets cover the Ozone Transport Commission (OTC) domain in the northeastern US with spatial resolution of 12x12km, temporal resolution of 1 hour, and 34 vertical layers.

Observed daily airborne pollen counts are collected from certified monitoring stations of the National Allergy Bureau (NAB) of the American Academy of Allergy, Asthma & Immunology (AAAAI) across the contiguous United States and parts of Canada. Data before 1994 are not used in this study because of the lower sampling frequency and few available monitoring stations. Most of the certified pollen stations were established after 2000. Before 1994, pollen stations usually only reported pollen data on a weekly or monthly or even yearly basis. Pollen data for 2001 and 2002 are not available to us because of an issue of intellectual property.

BELD3.1 was selected to provide information on the area coverage for different plant species. This selection is mainly based on the fact that it is the only database to provide area coverage information for multiple species (230 vegetation classes in total) and it has a high spatial resolution (1x1km).

Data of observed climatic factors are collected based on two considerations: (1) they are from quality controlled databases archived by different agencies and organizations

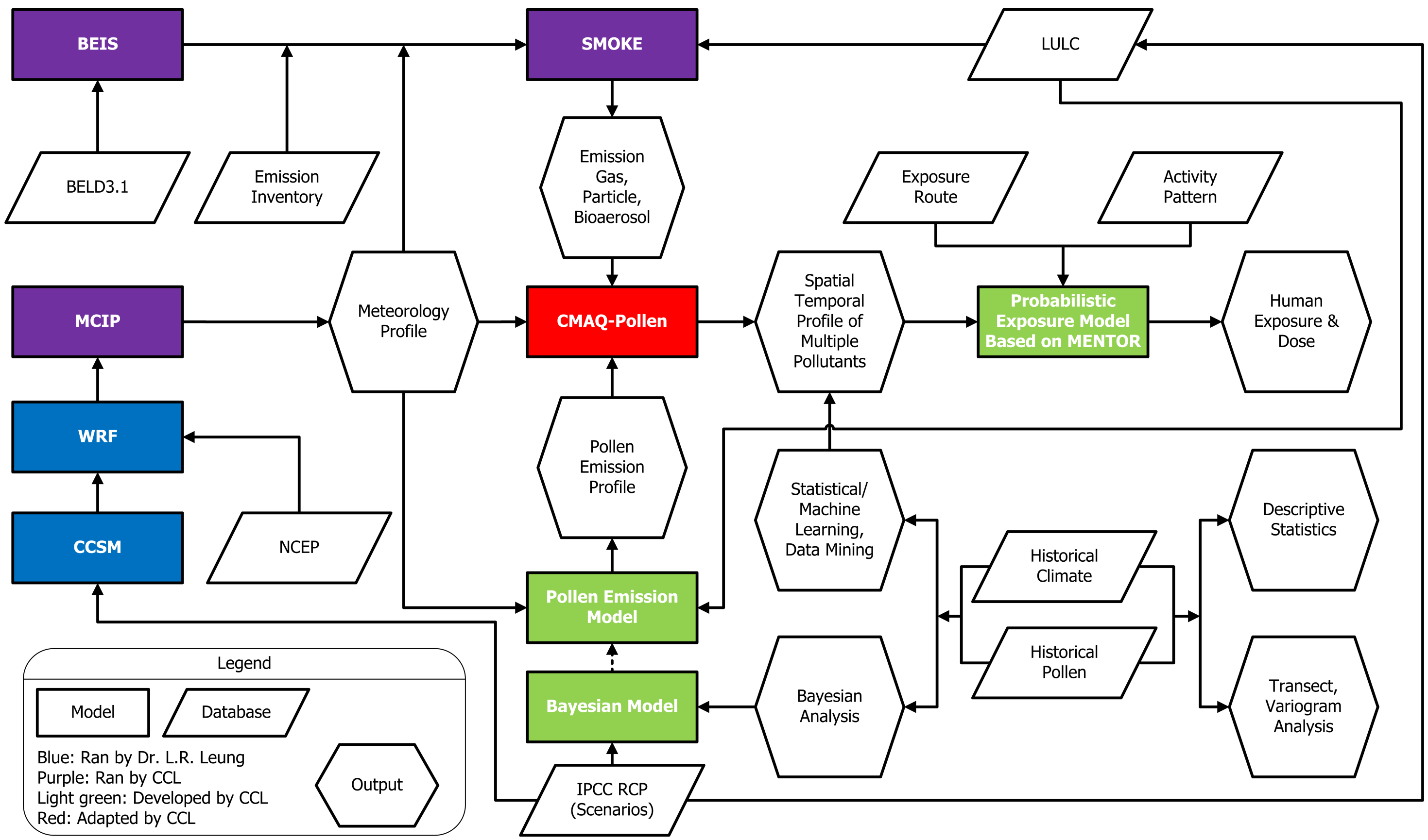

Figure 1.1: Schematic diagram of overall WRF-SMOKE-CMAQ-Pollen modeling system linking climate change and exposure.

Table 1.3: Major databases and observations used in this study

Data	**Period**	**Spatiotemporal Resolution**	**Model / Observation**	**Reference**	**Note**
Historical meteorology	2001-2004	3-hourly; 50x50km; North America	WRF	[55,56]	IC & BC: NCEP reanalysis of observation
Historical meteorology	2007	hourly; 12x12km; Northeastern US	WRF	[57]	USEPA and NJDEP
Future meteorology	2047-2050	3-hourly; 50x50km; North America	WRF	[55,56]	IC & BC: Output of CCSM
Emission	2007	hourly; 12x12km; Northeastern US	SMOKE	[58]	USEPA and NJDEP
Pollen count	2003-2011	Daily; 86 AAAAI stations; North America	Observation	From Dr. Bielory	Tree, weed, grass, spore; Spreadsheet
Pollen count	1994-2000	Daily; 86 AAAAI stations; North America	Observation	AAAAI report; From Dr. Bielory	Tree, weed, grass, spore; Plots in reports
Land Use Land Cover	Established in 1998	Stationary; 1x1km; North America	BELD3.1	[59]	Fraction of vegetation coverage
Observed Climatic Factors (e.g. CO_2 level)	Varied periods for historical and future years	Hourly / Daily / Monthly/ Yearly; Global	ESRL[a]and NCDC[b]of NOAA; DDC[c] of IPCC	Websites of NOAA and IPCC	Data used for historical analysis

[a] Earth System Research Laboratory (http://www.esrl.noaa.gov/gmd/dv/site/map1.html)

[b] National Climatic Data Center (http://www.ncdc.noaa.gov/oa/climate/climatedata.html)

[c] Data Distribution Centre (http://www.ipcc-data.org/ddc_co2.html)

such as NOAA and IPCC; and (2) they are from the meteorology/climate monitoring stations closest to the corresponding AAAAI pollen stations.

The advantage of the WRF-SMOKE-CMAQ-Pollen modeling system comes from four aspects. First, the models for both pollen emission and transport are constructed mechanistically based on basic principles of physics, phenology and meteorology. The emission model is parameterized using observed climate/meteorology and airborne pollen data from multiple years at multiple stations. Second, multiple climatic factors are taken into consideration in both pollen emission and transport models in order to evaluate the long term effects of climate change. Third, because of the utilization of the CMAQ model, multiple common air pollutants such as ozone and particulate matter (PM) can also be simulated simultaneously with allergenic pollen to examine their effects on allergic airway diseases. Lastly, patterns of pollen season timing shifting and airborne level variation in the US are, for the first time investigated using observations of airborne pollen counts of multiple taxa, meteorology and climate factors of multiple years from multiple stations.

Limitations of the proposed modeling system are mainly from two issues: (1) the BELD3.1 does not provide specific area coverage information for ragweed and mugwort. Area coverage of ragweed and mugwort obtained through an empirical algorithm from different vegetation classes may potentially contain uncertainty; (2) since the meteorology model WRF, pollen emission model, and air quality model CMAQ-Pollen are run separately, no feedback is considered between different sub-models in the current study; (3) since only four years simulation were used for each of the periods of 2001-2004 and 2047-2050, the simulation could not capture the internal variability of climate; and (4) since the meteorology model output was from only one regional model and one global climate model, the simulation results may not capture the full atmospheric physics and driving forces of complex climate system.

Table 1.4: Configurations and methodologies for proposed studies

Study	**Model / Method**	**Resolution & layers**	**Period**	**Domain**
Process archived meteorology	MCIP[a], M3TOOL[b]	50x50km or 12x12km,hourly,34	2001-2004, 2047-2050	CONUS[c]
Descriptive statistics of observed pollen	Trend, comparison	Multiple stations, daily, yearly	1994-2010	In US
Machine learning and Bayesian analysis	Regression, support vector machine, neuron network, decision tree	Multiple stations, yearly	1969-2006	In Europe,US
Latitudinal analysis	Regression,GDD[d]	Multiple stations	1994-2010	In US
Synchronization analysis	Correlation, Variogram	Multiple stations	1994-2010	In US
Emission profiles of pollen	Emission module[e]	50x50km or 12x12km,hourly,1	2001-2004, 2007, 2047-2050	CONUS, Northeast US
Anthropogenic emission profiles	SMOKE[h]	12x12km,hourly,1	2007	Northeast US
Concentration profiles of pollen	Adapted CMAQ[f]	50x50km or 12x12km,hourly,34	2001-2004, 2007, 2047-2050	CONUS, Northeast US
Exposures to allergenic pollen	Probabilistic model based on MENTOR[g]	climate region,daily	1994-2000, 2001-2010	CONUS

[a] MCIP: Meteorology-Chemistry Interface Processor [b] M3TOOL: Models-3 Tools [c] CONUS: Contiguous continental US
[d] GDD: Growing Degree Days [e] Developed in the current study [f] CMAQ: Community Multiscale Air Quality
[g] MENTOR: Modeling Environment for Total Risk studies [60] [h] SMOKE: Sparse Matrix Operator Kernel Emissions

1.6 Climate and Meteorology Data

1.6.1 Global climate model

The climate and meteorology data used in this study are from the established dataset of the North American Regional Climate Change Assessment Program [56]. The Atmosphere-Ocean General Circulation Model adopted for generating the selected dateset is the Community Climate System Model (CCSM) developed by the National Center for Atmospheric Research (NCAR).

Future climate simulations in NARCCAP were driven by scenario A2 from the Special Report on Emissions Scenarios (SRES) of the IPCC [55]. A2 is at the higher end of the SRES emission scenarios; it assumes future development will be regionally and economically oriented with projection of the CO_2 level being 850 ppm in 2100 [61]. A2 is preferred because it will generate a relatively larger climate change; and in terms of impacts and adaption, if one can adapt to larger climate change, then the smaller climate changes of the lower end scenarios can also be adapted to.

1.6.2 Regional climate model

In NARCCAP, WRF was used as a Regional Climate Model (RCM) to generate the selected meteorology dataset. Historical meteorology simulations were conducted based on boundary conditions generated using reanalysis of observed meteorology data from the National Centers for Environmental Prediction (NCEP); and future meteorology simulation was performed based on boundary conditions generated using output of the global climate model CCSM [55].

As depicted in Figure 1.2, the established meteorology data (3 hourly, 50 x 50 km) from NARCCAP are first processed using Meteorology-Chemistry Interface Processor Version 3.6 (MCIP3.6) [62], and then interpolated using Models-3 Tools (M3TOOL) to hourly resolved data for the pollen transport model.

For identification of patterns in pollen season timing shifting and airborne level variation across the contiguous US, the spatiotemporal resolution of the processed meteorology data (hourly, 50x50 km) for the pollen transport model is capable of capturing

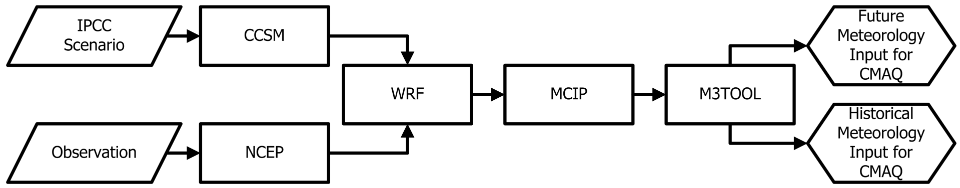

Figure 1.2: Schematic diagram of obtaining input meteorology for pollen emission and transport models from global climate model CCSM and regional meteorology model WRF.

the features of large scale physical transport processes such as horizontal advection, diffusion, dry deposition and cloud process as mentioned in Chapter 4. Because of the large horizontal resolution, the transport model cannot fully characterize sub-grid process such as vertical advection. The transport effects of vertical advection may be therefore weak in pollen transport model driven by the meteorology data with spatial resolution of 50x50 km.

The effect of vertical advection on pollen transport was further investigated using WRF meteorology data with higher spatial resolution. These meteorology data cover the Ozone Transport Commission domain in the Northeastern US with spatial resolution of 12 x 12 km and temporal resolution of one hour[57] (Table 1.3).

Chapter 2

OBSERVED ALLERGENIC POLLEN SEASON VARIATIONS UNDER CHANGING CLIMATE

2.1 Introduction

This chapter investigates allergenic pollen season variations and their relationships with climatic factors under changing climate conditions on the basis of observed airborne pollen count. The observed airborne pollen counts were used to derive six indices related to allergenic pollen season timing and airborne levels. Trends and changes of these derived pollen indices were analyzed for allergenic taxa of representative trees, weeds and grass. Statistical analyses were conducted to correlate the changes in pollen indices with the changes in climatic factors, and to identify the spatiotemporal patterns of changes in pollen indices. Bayesian analyses and machine learning models were used to investigate the relationships among pollen indices, airborne levels and concentrations, and observed climatic and meteorological factors.

2.2 Methods

On the basis of observed airborne pollen counts and climate/meteorology data, Figure 2.1 schematically diagrams assessment of allergenic pollen season variations under changing climate in the contiguous US [45]. Observed daily counted airborne pollen data and corresponding meteorology/climate factors were preprocessed to obtain indices of pollen season timing and airborne level, growing degree days (GDD), frost free days (FFD) and accumulative precipitation during the periods of 1994-2000 and 2001-2010. Mean pollen indices in these two periods were then compared and used to calculate the trends, correlation coefficients and variograms to identify the spatiotemporal patterns of changes in allergenic pollen seasons in the CONUS. The derived pollen indices were

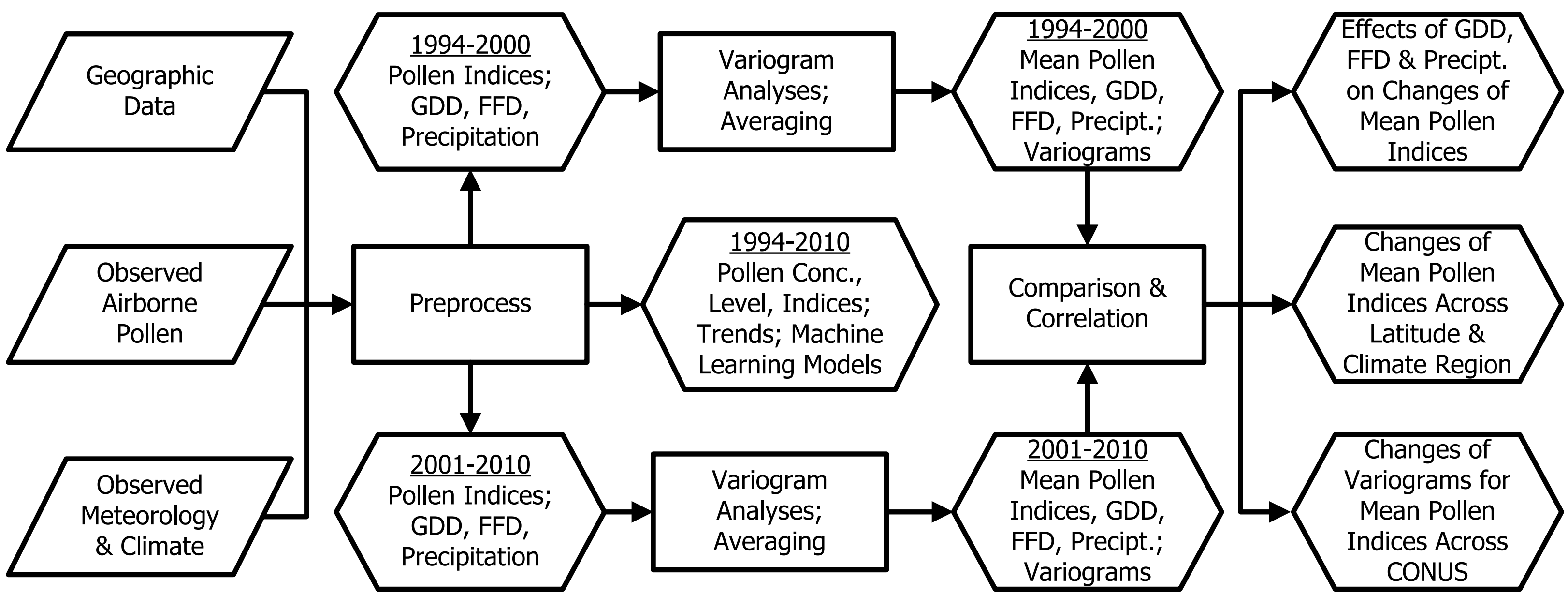

Figure 2.1: Schematic diagram for assessment of allergenic pollen season variations under the changing climate in the contiguous US (CONUS).

also used to train and evaluate the Bayesian and machine learning models for predicting these pollen indices from observed meteorology factors.

2.2.1 Data source

Observed daily airborne pollen counts were obtained from all available monitoring stations of the National Allergy Bureau (NAB) at the American Academy of Allergy, Asthma and Immunology (AAAAI) during the period of 1994-2010 across the CONUS (Figure 2.2). To retrieve more available data, airborne pollen data from two neighboring Canadian monitoring stations were also incorporated into the current study. The reported pollen data were only classified at the level of genus. Fifty-eight NAB-AAAI stations were selected because they recorded valid data for at least four years for performing further analyses and modeling parameterization. The main climate characteristics and geographical locations of the studied stations are listed in Table B.1. Observed daily temperatures, precipitation and other climatic factors were obtained from the National Oceanic and Atmospheric Administration (NOAA) meteorology stations nearest to the corresponding NAB-AAAAI pollen stations.

2.2.2 Pollen indices

Start Date (SD), Peak Date (PD), Season Length (SL), Peak Value (PV), Annual Mean (AM) and Annual total Production (AP) of daily counted airborne pollen were selected as six pollen indices to assess climate change impacts on allergenic pollen season timing and airborne level. With day 1 being January 1st, the start date (days from January 1st) is the day when the cumulative pollen count reaches 5% and end date when it reaches 95% of annual total count. The definitions of the start and end dates were demonstrated using the observed daily pollen count during the allergenic pollen season in 2010 at the monitor station in Springfield, New Jersey (Appendix B.1). This method was used to exclude long-range-transport pollen grains from the local pollen season. These long-range-transport pollen grains from surrounding regions may influence pollen counts at the beginning and end of local pollen seasons [63,64]. Season length (day) is defined as the duration between start and end dates. Peak date is reached when the maximum daily count is registered. Peak value (pollen grains/m^3)

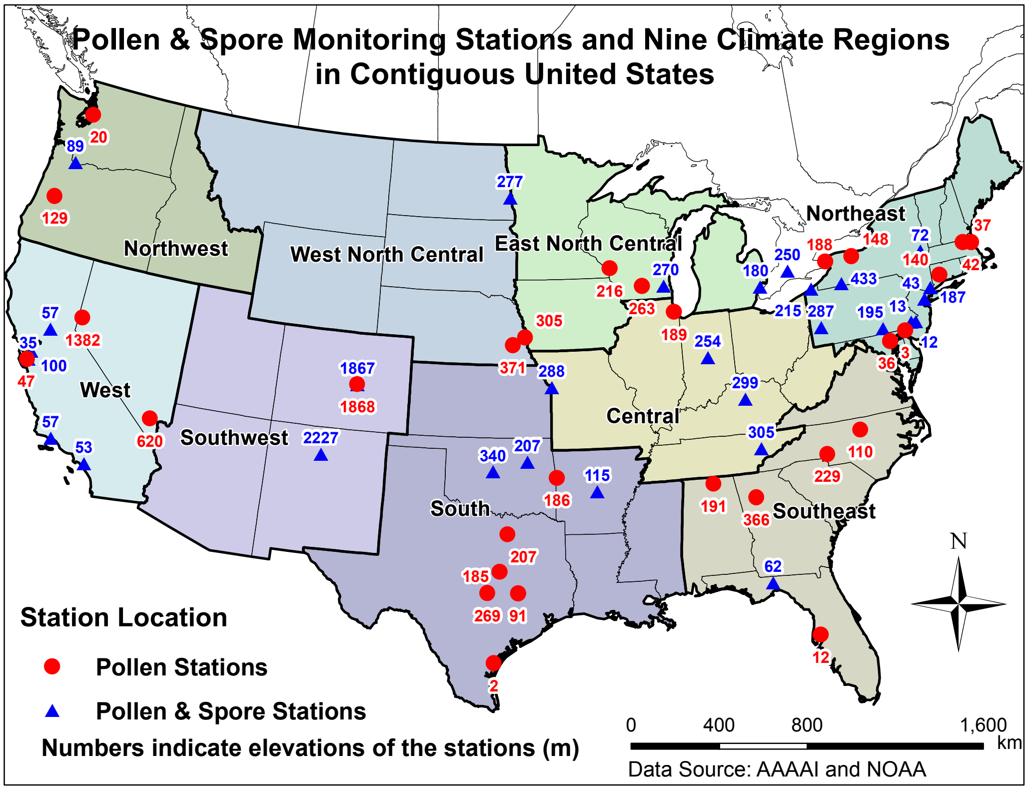

Figure 2.2: Distribution of the studied pollen stations (n = 58) across the nine climate regions in the contiguous US. The climate regions are classified according to the long term observed temperature and precipitation based on the database of National Climatic Data Center of the National Oceanic and Atmospheric Administration [1].

is the maximum daily count recorded on the peak date during a pollen season. Annual production (pollen grains/m^3) is defined as the sum of daily counts during a pollen season. Annual Mean (pollen grains/m^3) is the mean daily pollen count during the pollen season.

Further assumptions about the derived pollen indices are as follows: (1) Start date of summer-flowering ragweed and mugwort should not be earlier than June 21st, because weeds (e.g. ragweed) are generally short-day species which require accumulation of a consolidated period of darkness to flower [65,35]; and start date of spring-flowering birch, oak and grass should not be later than June 21st, because spring-flowering trees (e.g. birch) and grass are generally long-day species. Shortening photoperiod after June 21st will trigger growth cessation, cold acclimation and dormancy development on these plants[66,67]. (2) Pollen data with SL of less than 7 or greater than 80 days are assumed unreasonable and excluded from further analyses[68]; the exception is made for grass, which usually includes many different species and can have a pollen season length longer than 80 days.

As an example, an observed daily pollen curve is illustrated in Figure 2.3 together with the derived pollen indices based on raw airborne pollen counts at a NAB-AAAAI pollen monitoring station located in Springfield, New Jersey. The derived pollen indices (SD, PD, SL, PV, AM and AP) were examined according to the above principles before further analyses.

2.2.3 Climatic factors

Allergenic pollen season timing and levels have been widely reported to be associated with Growing Degree Days (GDD), Frost Free Days (FFD) and accumulated precipitation[69,35,70]. The fixed-period GDD value was calculated for each taxon in each year at each NAB-AAAAI station. GDD in a fixed period was calculated using equation 2.1,

$$GDD = \sum_{i=ID}^{LD} (T_i - T_b), \qquad T_i \geq T_b \tag{2.1}$$

where ID and LD are the Initial Date and Last Date to accumulate the temperature difference between daily temperature T_i and base temperature T_b. The parameters of ID, LD and T_b for each species were obtained through equation 2.2 by maximizing the correlation coefficients between GDD and SD, using observed daily pollen count and temperature from 1994 to 2010 at the studied NAB-AAAAI stations,

$$(ID, LD, T_b) = \arg \max_{ID,LD,T_b} (\mid \rho(SD, GDD) \mid) \tag{2.2}$$

where ρ is the Spearman correlation coefficient between SD and GDD. The Spearman correlation coefficient was used because it can handle small nonlinearity, if any, existing between SD and GDD. ID took the value from January 1st, February. 1st, $\cdots$, and December 1st; LD took the value from January 31st, February. 28th, $\cdots$, and December 31st; T_b assumed a value from -2 to 10 °C with an interval being 0.5 °C.

FFD is defined as the interval between the last frost day during spring and the first frost day (daily minimum temperature below 0 °C) during fall. Pollen levels were

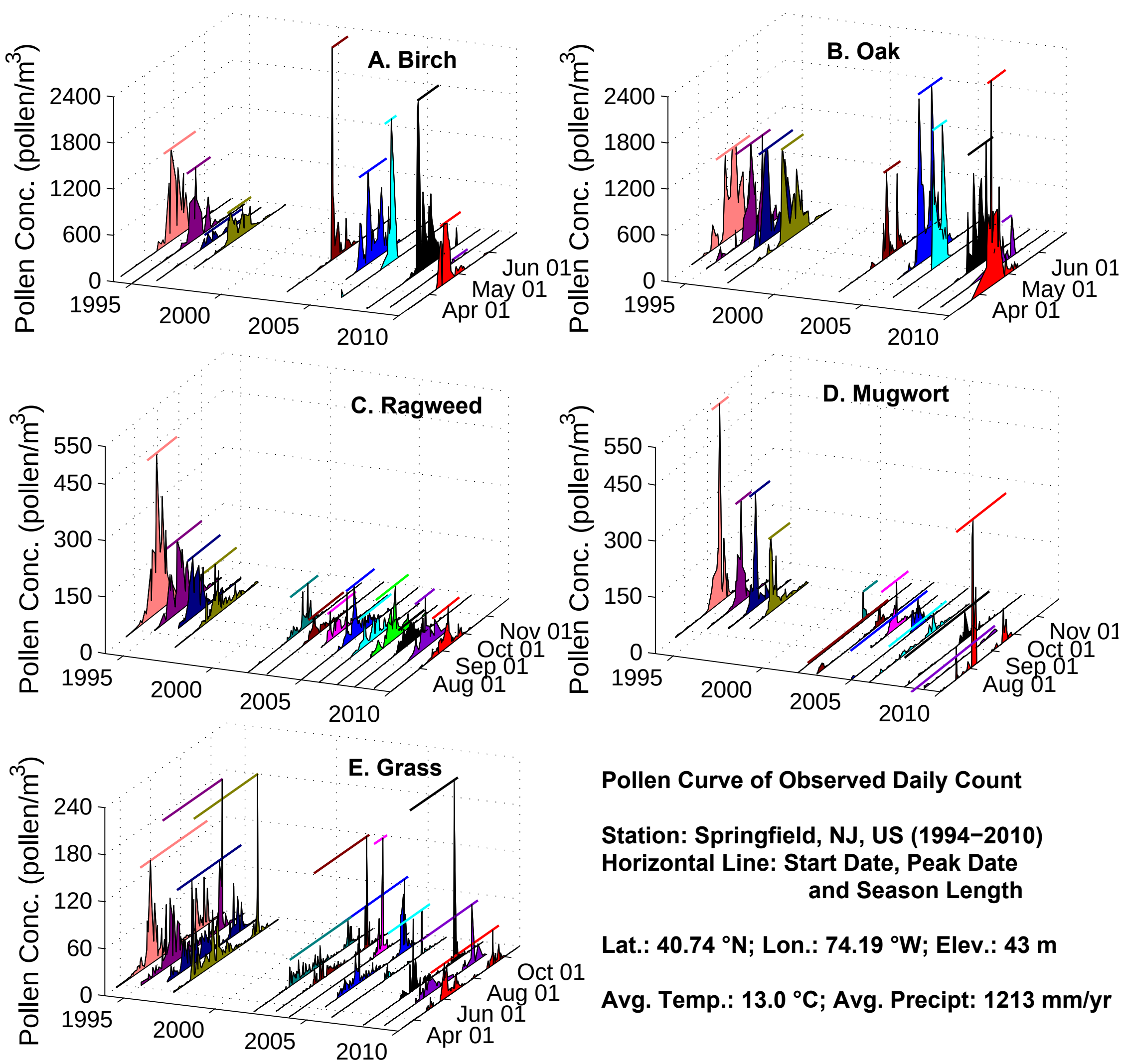

Figure 2.3: Daily pollen curves observed at the Springfield, NJ station, 1994-2010. Top horizontal line represents season length, its start and end points indicate start and end dates, respectively, based on 5% and 95% of the total annual counted airborne pollen. The contacted point on the line represents the peak value and peak date. (A) Birch; (B) Oak; (C) Ragweed; (D) Mugwort; and (E) Grass.

affected by precipitation preceding and during the pollen seasons[71,70]. Accumulated precipitation in fixed periods was used in the current study to investigate the climate change impacts on allergenic pollen levels. These fixed periods were selected to approximately cover the allergenic pollen seasons and the time right before the seasons at the studied NAB-AAAAI stations (Table 2.1).

Table 2.1: Initial Date (*ID*, mmdd), Last Date (*LD*, mmdd), and base temperature (T_b, °C) for calculating fixed-period Growing Degreee Days (GDD), Frost Free Days (FFD) and accumulated Precipitation (mm).

	Brich			Oak			Ragweed			Mugwort			Grass		
	ID	LD	T_b	ID	LD	T_b	ID	LD	T_b	ID	LD	T_b	ID	LD	T_b
GDD	0201	0430	2	0201	0430	5.5	0201	0228	0.5	0201	0228	1.5	0101	0531	6
FFD	0101	1231	-	0101	1231	-	0101	1231	-	0101	1231	-	0101	1231	-
Precip.	0101	0630	-	0101	0630	-	0501	1031	-	0501	1031	-	0101	0630	-

2.2.4 Difference of mean pollen indices between periods of 1994-2000 and 2001-2010

To reduce the effects of the natural climate and plant-growth variability on pollen indices[72,73], mean pollen indices were calculated for the past decade (2001-2010) and the 1990s (1994-2000) at each station. Pollen data before 1994 were scarce and usually reported on a weekly basis, and thus not adequate for deriving start date and duration of pollen season. Because of proprietary issues, airborne pollen data in 2001 and 2002 are not available to us for most of the studied stations. This makes seven years of airborne pollen data available for the period of 1994-2000 and approximately eight years of data available for the period of 2001-2010. For calculating the changes in mean pollen indices between the past decade and the 1990s, at least three years of pollen data in each of the two periods are required. Student's t tests were performed to check the significance of changes in pollen indices during the periods of 1994-2000 and 2001-2010 for each of the five pollen taxa. Since hypotheses tests on four pollen indices based on the same group of observed pollen data may potentially cause spurious significant findings, the Benjamini Hochberg procedure was used to guarantee a false discovery rate of less than 5%[74].

Changes in mean pollen indices at station i in climate regions j were calculated using equation 2.3,

$$\begin{cases} \Delta\overline{SD}_{i,j} = \overline{SD}_{i,j,2} - \overline{SD}_{i,j,1} \\ \Delta\overline{SL}_{i,j} = \overline{SL}_{i,j,2} - \overline{SL}_{i,j,1} \\ \Delta\overline{AP}_{i,j}/\overline{AP}_{i,j,1} = (\overline{AP}_{i,j,2} - \overline{AP}_{i,j,1})/\overline{AP}_{i,j,1} \\ \Delta\overline{PV}_{i,j}/\overline{PV}_{i,j,1} = (\overline{PV}_{i,j,2} - \overline{PV}_{i,j,1})/\overline{PV}_{i,j,1} \end{cases} \tag{2.3}$$

where $\overline{SD}_{i,j,1}$, $\overline{SL}_{i,j,1}$, $\overline{AP}_{i,j,1}$ and $\overline{PV}_{i,j,1}$ are the mean SD, SL, AP and PV, respectively, during the period of 1994-2000 at station i in climate region j; and $\overline{SD}_{i,j,2}$, $\overline{SL}_{i,j,2}$, $\overline{AP}_{i,j,2}$ and $\overline{PV}_{i,j,2}$are the mean SD, SL, AP and PV, respectively, during the period of 2001-2010.

The regional ($\overline{\Delta \overline{PI}_j}$) and nationwide ($\overline{\Delta \overline{PI}}$) average changes in mean pollen index were calculated using equation 2.4,

$$\begin{cases} \overline{\Delta \overline{PI}_j} = \frac{\sum_i \Delta \overline{PI}_{i,j}}{n_j} \\ \overline{\Delta \overline{PI}} = \frac{\sum_j \sum_i \Delta \overline{PI}_{i,j}}{\sum_j n_j} \end{cases} \tag{2.4}$$

where $\Delta \overline{PI}_{i,j}$ is the change in mean pollen index (PI), it could be the change of any of the four pollen indices obtained from equation 2.3; the n_j is the number of available NAB-AAAAI pollen stations in climate region j.

2.2.5 Trend and correlation analysis

Regression analysis was performed to identify trends of start date, season length, peak value and annual production of allergenic pollen during 1994-2010 at each of the NAB-AAAAI stations. At least six years of pollen data are required for conducting trend analyses of pollen indices at a NAB-AAAAI pollen monitoring station. Correlation analyses were conducted to examine the relationships between changes in mean pollen indices and changes in mean climatic factors.

2.2.6 Variogram analysis

Variogram is a geostatistical technique to investigate and quantify the spatial variability of widely distributed spatial phenomena[75]. It can be used to explain the trends, cyclicity, geometric anisotropy and zonal anisotropy of a given spatial variable. Larger variogram means higher variability and lower spatial correlation among spatial variables separated by certain distances. In the current study, fifteen spatial lags were used for calculating the variogram of mean pollen indices. The values of spatial lag varied from 45 km to 2360 km depending on the data availability of different allergenic species. All the spatial lags were calculated isotropically (i.e. equally treated for different directions) among NAB-AAAAI pollen monitoring stations across the CONUS.

The variogram of a mean pollen index of a given period was calculated by equation 2.5,

$$2\gamma(d) = \frac{1}{N(d)} \sum_{N(d)} [NMPI(u) - NMPI(u+d)]^2 \tag{2.5}$$

where $2\gamma(d)$ is the variogram with spatial lag of d; $NMPI(u)$ is the Normalized Mean Pollen Index at location u; $N(d)$ is the number of pairs of pollen stations separated by a spatial distance of d. As shown in equation 2.6.

$$NMPI(u) = MPI(u)/OMPI(u) \tag{2.6}$$

The $NMPI(u)$ was obtained by dividing the Mean Pollen Index $[MPI(u)]$ of a given period by the Overall Mean Pollen Index during the entire observation period of 1994-2010 at the same location $[OMPI(u)]$.

The normalized semi-variogram $\gamma_N(d)$ was derived using equation 2.7,

$$\gamma_N(d) = \gamma(d)/\max_d(\gamma(d)) \tag{2.7}$$

where the $\max_d(\gamma(d))$ is the maximum semi-variogram of all spatial lags d for a given pollen index (either SD, SL, AP or PV) during the periods of 2001-2010 and 1994-2000. The changes of normalized semi-variogram at spatial lag d $(\Delta\gamma_N(d))$ for a given mean pollen index were calculated using equation 2.8,

$$\Delta\gamma_N(d) = \gamma_N^{(2)}(d) - \gamma_N^{(1)}(d) \tag{2.8}$$

where $\gamma_N^{(1)}(d)$ and $\gamma_N^{(2)}(d)$ are the normalized semi-variograms during the periods of 1994-2000 and 2001-2010, respectively.

2.2.7 Bayesian analysis

As shown in Figure 2.4, Bayesian analysis[76,77,70] was carried out to study the relationship between multiple pollen indices and multiple climatic factors using historical birch pollen and climate data in three representative stations in Europe, and five stations in the US. The established relationships were used to relate the future pollen index (mainly annual total and start date) with the future temperature and CO_2 level projected by IPCC. Observed birch pollen indices are assumed normally distributed

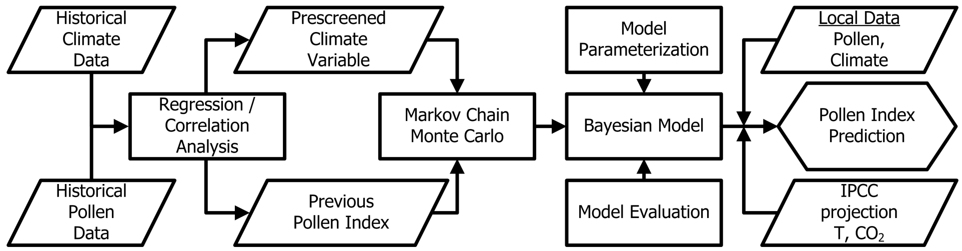

Figure 2.4: Schematic diagram of Bayesian analysis.

variables which fluctuate around mean trends depending on the combination of multiple random climate/meteorology factors, and that pollens of the same genus (Betula) have similar responses to climate/meteorology changes. The ordinary norm linear regression model[76] is presented in equation 2.9,

$$(Y|\beta, \sigma^2, X) \sim N_n(X\beta, \sigma^2 I_n) \tag{2.9}$$

where $Y = (y_1, \cdots, y_n)^T$ is a vector of pollen indices, the five year overlapping mean of either annual production (pollen/m^3) or peak value (pollen/m^3) or start date (day) or peak date (day). With day 1 being January 1st, the start date is defined when the cumulative pollen count reached a certain percentage of the annual production[31] and peak date is reached when the daily maximum count is registered. X is the $n \times k$ matrix of explanatory variables in which each column vector x_i corresponds to values of a climatic factor in n years and k is the number of variables. I_n is the $n \times n$ identity matrix. β and σ^2 are the unknown vector of coefficient and variance, respectively.

Detailed formulations of Bayesian analysis can be found in literature[70] and Appendix B.3.

2.2.8 Machine learning model

For discussion of machine learning models, pollen levels in sections 2.2.8 and 2.3.7 refer to three qualitative levels, which are high, medimum and low. Pollen concentrations refer to the daily pollen counts per cubic meter.

Support Vector Machine (SVM)is a widely used machine learning model to solve both the classification and regression problems[78,79]. SVM classifies data samples through maximization of the distance between decision boundary and data samples based on support vectors. Support Vector Regression (SVR) is a similar machine learning model as SVM, but used for regression problems[79,80]. Both SVM and SVR can take advantage of kernel tricks to transform the low dimensional data samples into high dimensional spaces, so that these data samples can be easily separated or regressed[78,79].

Neural NetWork (NNW)[81] and decision tree[82] are also commonly used machine learning models used for classification and regression. NNW mimics human brains to classify and quantify data samples through input layer, hidden layers and output layer[81]. It can capture the highly nonlinear relationship within the data through addition of hidden layers and complex activation functions. Decision tree classify data samples through maximization of information gain[82]. Regression tree is essentially stage wise regressions, in which different regressions were used for different subgroup of data samples[83]. Detailed descriptions and formulations of SVM, SVR, M5P, decision and regression trees can be found in the literature[78,79,81,82,80,83].

The relationship between the observed airborne pollen levels (i.e., high, medium and low) and meteorological factors were investigated using Support Vector Machine[78,79], Neural NetWork[81] and decision tree[82]. The relationship between the actual concentrations of observed airborne pollen and meteorological factors were studied using Support Vector Regression[79,80], neural network[81] and M5P regression tree[83].

In the current study, machine learning models were used to predict oak pollen levels and concentrations in Springfield, New Jersey. The data for machine learning models are the observed airborne oak pollen counts during the period of 1994-2010 from the monitoring station in Springfield, New Jersey, and daily temperature, precipitation and wind speed from the meteorology station in Newark, New Jersey. As shown in Figure 2.5, the observed airborne pollen counts and meteorology data were first processed into structured input and output variables (Table 2.2); this structured dataset was then used to train and evaluate the machine learning models, so that the models could be used to predict daily airborne pollen levels and concentrations using observed information in the previous days.

For prediction of airborne pollen levels, daily airborne oak pollen counts were transformed into three levels on the basis of two threshold values as shown in equation 2.10,

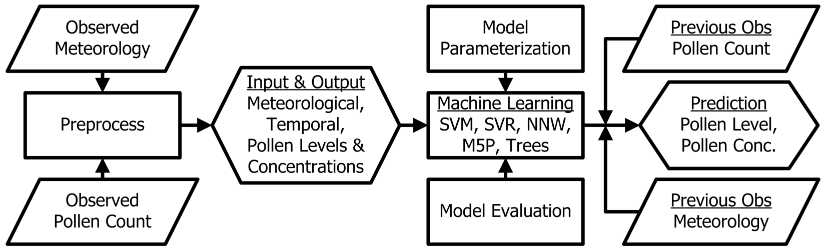

Figure 2.5: Schematic diagram of machine learning model.

$$PL = \begin{cases} 1 \quad \text{(Low Level)} & \text{if} \quad c < c_{Thr1} \\ 2 \quad \text{(Medium Level)} & \text{if} \quad c_{Thr1} \leq c < c_{Thr2} \\ 3 \quad \text{(High Level)} & \text{if} \quad c \geq c_{Thr2} \end{cases} \tag{2.10}$$

where PL and c are the daily pollen level and concentration (pollen grain/m^3), respectively. The threshold concentrations of oak pollen, c_{Thr1} and c_{Thr2}, take values of 13 (pollen grain/m^3) and 227 (pollen grain/m^3), respectively[15]. Three machine learning models, SVM, NNW and decision tree, were then used to solve the classification problem of daily airborne pollen counts.

Table 2.2: Input and output variables for machine learning models. t=0, refers to the time that the forecast is produced; t=-h refers to h days before the forecast, whereas cumulative variables start at 1st January of each year.

	Group	Variable	Time
Input	Temporal	Day of Year (DOY) ={1,…,365}	t = -3, -2, -1
	Meteorological	Max. Daily Temperature (MaxT)	t = -3, -2, -1
		Cumulative MaxT (MaxTSum)	t = -3, -2, -1
		Min. Daily Temperature (MinT)	t = -3, -2, -1
		Cumulative MinT (MinTSum)	t = -3, -2, -1
		Mean Daily Temperature (Temp)	t = -3, -2, -1
		Cumulative Temp (TempSum)	t = -3, -2, -1
		Daily Precipitation (Prcpt)	t = -3, -2, -1
		Cumulative Prcpt (PrcptSum)	t = -3, -2, -1
		Wind speed (Wdspd)	t = -3, -2, -1
	Pollen	Pollen Concentration	t = -3, -2, -1
Output	Regression	Pollen Concentration	t = 0
	Classification	Pollen Levels = {3-High, 2-Medium, 1-Low}	t = 0

For prediction of actual daily airborne concentration of oak pollen, three machine learning models, SVR, NNW and regression tree M5P, were used to solve the regression problem of daily airborne pollen counts. The input and output variables for both classification and regression models are listed in Table 2.2.

The procedure for training, optimizing and evaluating a machine learning model, for example SVM for classification problem, is demonstrated as follows:

(1) Split the data into two halves as training and test dataset, respectively.

(2) Implement SVM using R package (e.g., e1071) based on default parameters.

(3) Tune the model multiple times (e.g., tune.svm) in parameter space to obtain the optimum parameters based on training data.

(4) Check the bias and variance: calculate the training and test errors on data samples of different size based on the optimum parameters.

(5) Evaluate the model performance: calculate the accuracy, precision, recall and F1 score for classification model; and calculate root mean square error, correlation coefficient and index of agreement for regression model.

For the classification problem, the precision and recall are defined according to confusion table. The confusion table for pollen level i is presented in Table 2.3. TP_i, TN_i, FP_i and FN_i represent numbers of cases of True Positive, True Negative, False Positive and False Negative for observed and predicted pollen level i, respectively.

Table 2.3: Confusion table for classification models.

Level i		Observation	
		TRUE	FALSE
Prediction	TRUE	TP_i	FP_i
	FALSE	FN_i	TN_i

The overall accuracy Acc, precision P, recall R and $F1$ score for observed and predicted pollen levels were calculated using equation 2.11,

$$\begin{cases} Acc = \frac{\sum_{i=1}^{|C|} TP_i + TN_i}{\sum_{i=1}^{|C|} TP_i + FP_i + TP_i + FN_i} \\ P = \frac{1}{|C|} \sum_{i=1}^{|C|} \frac{TP_i}{TP_i + FP_i} \\ R = \frac{1}{|C|} \sum_{i=1}^{|C|} \frac{TP_i}{TP_i + FN_i} \\ F1 = \frac{2PR}{P+R} \end{cases} \tag{2.11}$$

where $|C| = 3$ is the number of classes of pollen levels.

For the regression problem, the Index of Agreement IA was calculated using equation 2.12,

$$IA = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(|P_i - \overline{O}| + |O_i - \overline{O}|)^2} \tag{2.12}$$

where the P_i and O_i are the i^{th} predicted and observed pollen concentrations, respectively. The N is the number of total observations. $\overline{O}$ is the average concentration of all observations.

2.3 Results and Discussion

2.3.1 Trends of pollen indices

Trends of oak pollen indices at representative stations

As a demonstration, Figure 2.6 shows the trends of start date, season length, peak value and annual mean for oak pollen season during 1994-2010 at six representative pollen monitoring stations: Fargo (North Dakota), College Station (Texas), Omaha (Nebraska), Pleasanton (California), Cherry Hill and Newark (New Jersey). These stations were selected because (1) they have airborne daily pollen counts available for birch and oak for multiple years between 1994 and 2010; and (2) they are located at representative geographical and climatic regions across the CONUS [68].

As shown in Figure 2.6, oak pollen season was found to start earlier at five out of the six monitoring stations; season lengths at Fargo, Omaha and Pleasanton were observed to be longer, while season lengths at College Station, Cherry Hill and Newark appeared to be shorter. Increases in annual mean and peak value have been identified at most of the studied stations. Detailed description of the trend analyses of pollen indices at these six representative stations are documented in the literature for birch, oak, ragweed and mugwort [68,84].

Trends of pollen indices for five taxa at all studied stations

Similar for the six representative stations, trend analysis was conducted for start date, season length, peak value and annual production of birch, oak, ragweed, mugwort and grass pollen at all monitoring stations, where sufficient observations of pollen count have been recorded. Figure 2.7 summarizes the results of trend analyses of start date, season

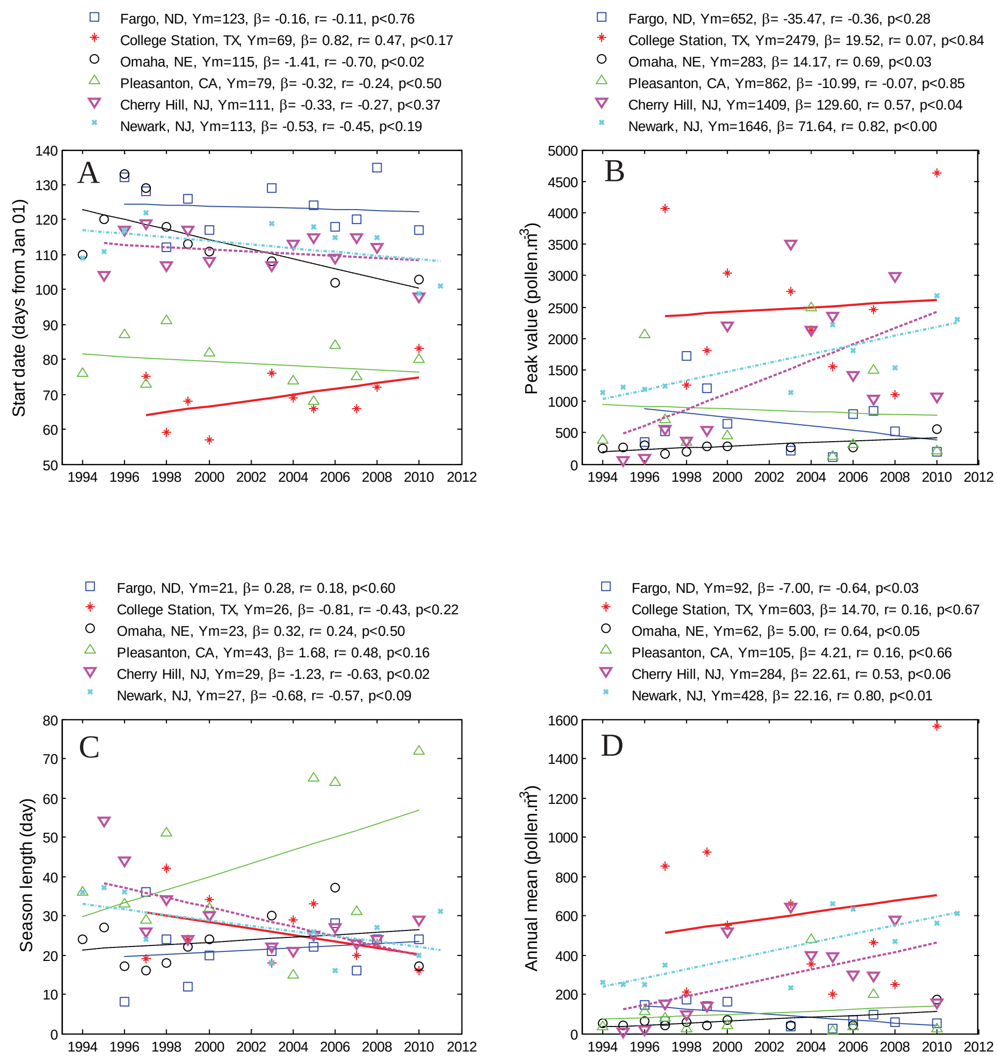

Figure 2.6: Oak pollen indices and calculated trends from 1994 to 2011 for six stations in the US. *Ym* is the mean value, β the annual trend, *r* the correlation coefficient and *p* the significance of the trend.

length, peak value and annual production of allergenic pollen season during 1994-2010 for each taxon at each NAB-AAAAI station. For example, for start date of birch pollen, trend analyses on start date were performed at each station based on available pollen data from 1994 to 2010. The number of stations where decreasing trends (i.e., negative slope) of birch pollen start date have been observed, was plotted as the first bar in the left side of Figure 2.7A; the section of solid bar gives the number of stations where significant decreasing trends have been observed ($p < 0.05$, Student's t test). Likewise, the first bar in the right side of Figure 2.7A indicates the number of stations where increasing trends of birch pollen start date have been observed.

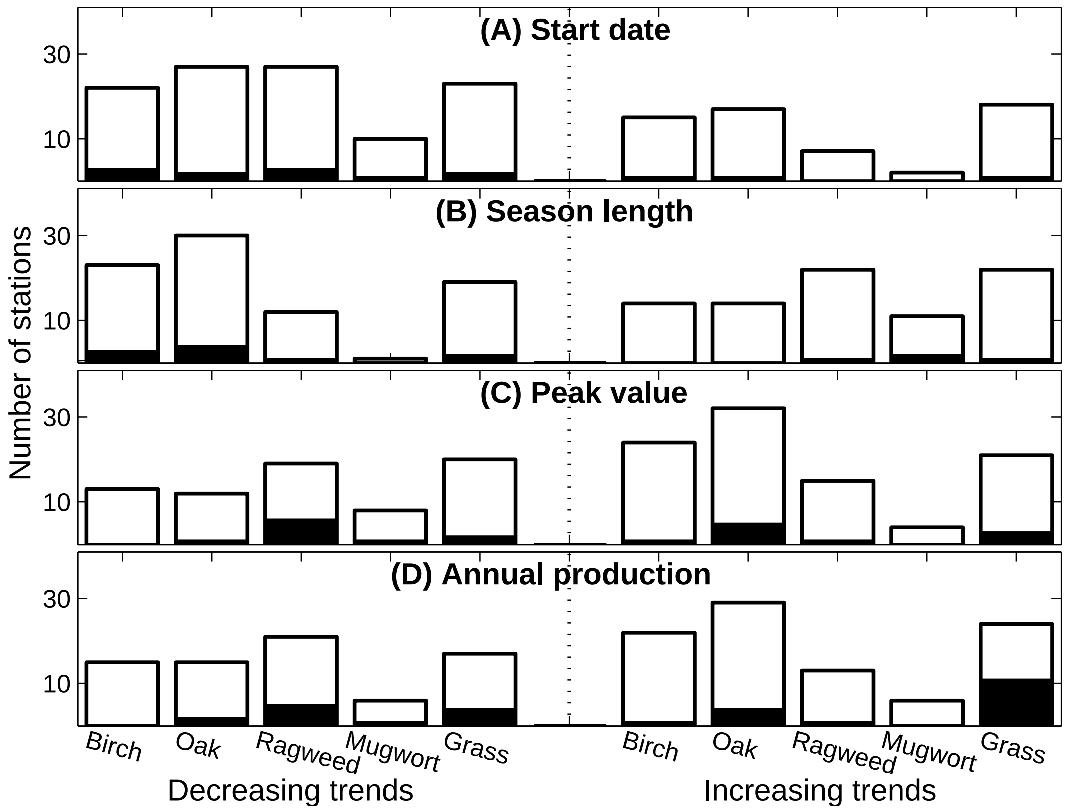

Figure 2.7: Number of stations where decreasing or increasing trends of pollen indices have been observed from 1994 to 2010. (A) Start Date, (B) Season Length, (C) Peak Value, and (D) Annual Production. The black bar indicates the number of stations at which the observed trends are significant at 5% level based on the Student's t test. Decreasing trends indicate that pollen season tends to start earlier, season length tends to be shorter, and peak value and annual production tend to decrease.

Decreasing trends during 1994-2010 indicate that the pollen season tends to start earlier, season length tends to be shorter, and peak value and annual production tend to decrease. The allergenic pollen season during the period of 1994-2010 across the CONUS showed early start trends at 59%, 61%, 79%, 83% and 56% of the 50 studied stations for birch, oak, ragweed, mugwort and grass, respectively. Approximately 7% of the studied stations showed trends of significantly earlier start dates ($p < 0.05$, Student's t test). Season lengths tended to be shorter at 62% and 68% of the studied stations for birch and oak, respectively, but appeared to be longer at 65%, 92% and 54% of the studied stations for ragweed, mugwort and grass, respectively. The number of stations with significantly different start dates and season lengths in general are proportional to the number of stations with increasing or decreasing trends of start date and season length.

The peak value and annual production of daily counted airborne pollen tended to increase for spring-flowering taxa at most of the studied stations. Around 62% of the observations showed increasing trends in peak value and annual production during the period of 1994-2010 (one observation corresponds to airborne pollen count for one taxon at one station during 1994-2010). For the peak value and annual production of the summer-flowering taxa, decreasing trend and significant decreasing trend are more common than increasing trend. The widely increasing trends of peak value and annual production of spring flowering taxa are consistent with a European study focused on the trends of observed annual airborne pollen counts from multiple taxa across Europe[43]. The study reported that 59% of the observed trends of annual airborne pollen counts increased during various periods from 1977 to 2009 at different European pollen monitoring stations.

Trends of mean pollen indices across latitudes

Figure 2.8 displays the overall mean pollen indices (average over 1994-2010) and the corresponding standard deviations across latitude. These overall mean pollen indices and their standard deviations are listed in Tables B.2 and B.3. The spring-flowering species (birch, oak and grass) flowered earlier at the lower latitudes than those at the higher latitudes; while the short-day summer flowering species[35,65] (ragweed and mugwort) started flowering from higher latitudes and gradually shifted to lower latitudes. Pollen season lengths decrease as latitudes go higher.

Except for birch, annual production of allergenic airborne pollen tended to decrease with the increase of latitudes. This exception for birch pollen is mainly caused by spatial distribution of birch forest. Birch trees mainly distribute in the northern, particularly the northeastern CONUS, according to the Biogenic Emission Landuse Database (BELD)[85].

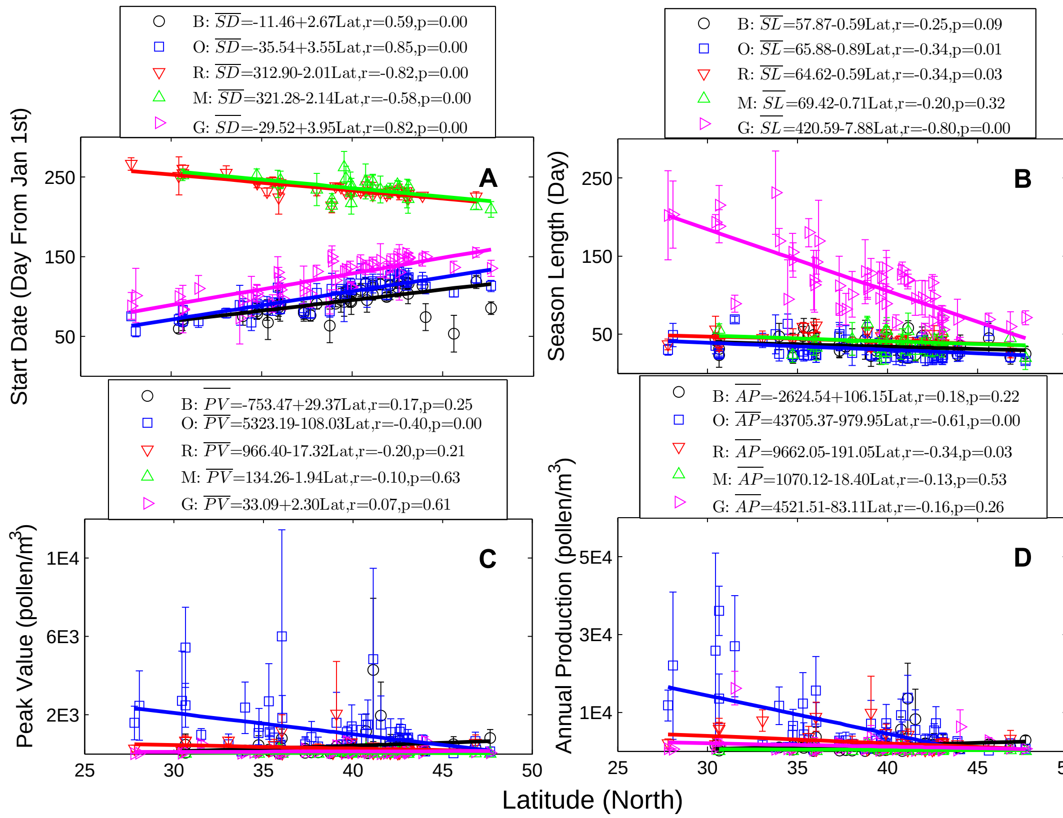

Figure 2.8: Overall mean Start Date (A), Season Length (B), Peak Value (C) and Annual Production (D) and their standard deviations across latitudes. An overall mean pollen index is defined as the average over entire observation time of 1994-2010 at a station. Regression equations are presented in the legends.

2.3.2 Changes of mean pollen indices between periods 2001-2010 and 1994-2000

Figure 2.9 displays the changes of mean pollen indices between the periods of 1994-2000 and 2001-2010 in nine climate regions across the CONUS. The relative change in peak value was calculated by dividing the changes in mean peak value from two periods by the mean peak value in the period of 1994-2000, i.e. $\Delta\overline{PV}/\overline{PV1} = (\overline{PV2} - \overline{PV1})/\overline{PV1}$ (likewise for annual production). The box plot was generated using changes in mean pollen indices at different stations within the same climate region. Tables 2.4 and 2.5 list the summary statistics for the changes of mean pollen indices in nine climate regions and in the CONUS, respectively.

Changes in pollen indices vary by climate region and taxon. The allergenic pollen

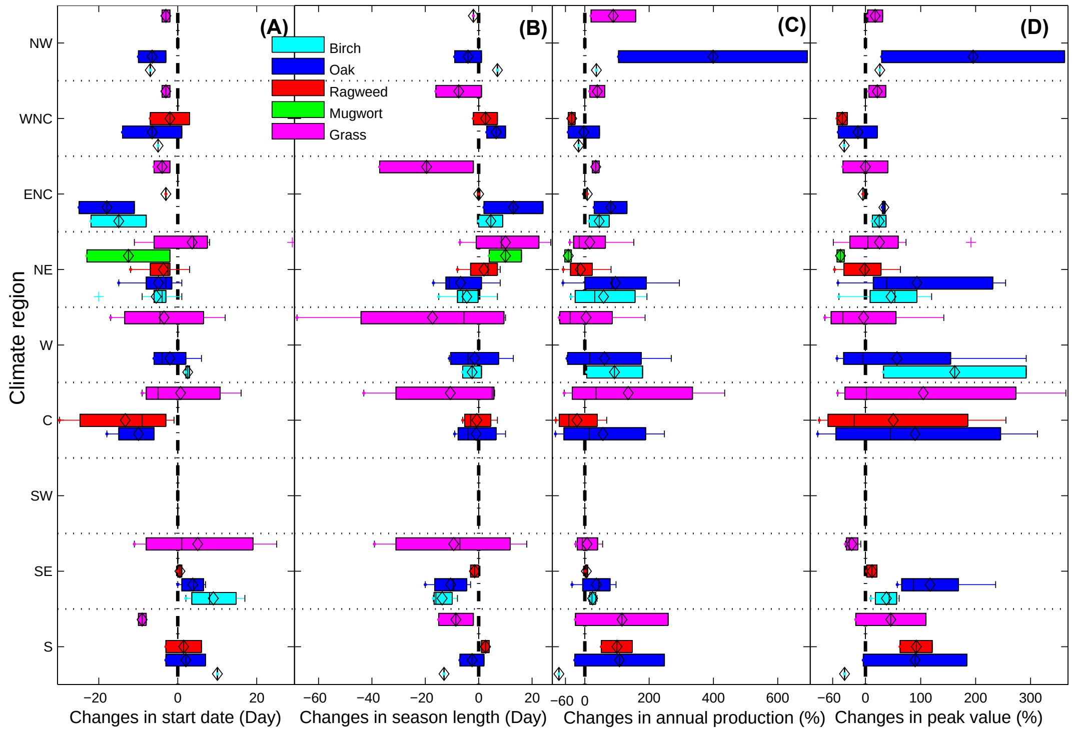

Figure 2.9: Changes in mean pollen indices during the periods of 2001-2010 from the means during 1994-2000 across the contiguous US. The nine climate regions are shown in Figure 2.2 South (S), Southeast (SE), Southwest (SW), Central (C), West (W), Northeast (NE), East North Central (ENC), West North Central (WNC), and Northwest (NW). (A) Start Date, (B) Season Length, (C) Annual Production, and (D) Peak Value. In each box plot the central black line is the median; the black diamond is the mean; two sides are the 25th (q_1) and 75th (q_3) percentiles; the whiskers represent q_3+1.5(q_3-q_1) and q_1-1.5(q_3-q_1), respectively. "Outliers" were plotted as plus ('+'). A negative number indicates earlier pollen season start date, shorter season length, and decreasing pollen levels.

season in most of the climate regions tended to start earlier in the past decade than in the 1990s, but it tended to start later in the South and Southeast climate regions. In general, the allergenic pollen season for the northeastern CONUS (e.g., Northeast and East North Central climate regions) in the past decade appeared to last longer than in the 1990s; while for the southern CONUS (e.g., South and Southeast climate regions) it appeared to be shorter (Figure 2.9, Table 2.4). The later onset and shorter duration of the allergenic pollen season in the South and Southeast regions are consistent with the decreasing trends of temperature in these regions[22]. Allergenic pollen levels across the CONUS were observed to increase substantially across different geographic areas in the past decade compared to the 1990s.

Table 2.4: Changes of allergenic pollen seasons between the periods of 2001-2010 and 1994-2000 in the nine climate regions across the contiguous US. The changes in mean pollen index for a given species during these two periods at all available stations in one region were used to calculate the regional average. The 95% confidence interval of each pollen index in each climate region was also calculated. Asterisk ($*$) indicates statistically significant difference at 5% level based on Student's t test.

Regions		South	South East	South West	Central	West	North East	East North Central	West North Central	North West
Number of Stations		8	6	2	5	7	13	4	2	3
Start Date (Days)	B	10.0*	9.0*			2.5*	-5.4*	-15.0*	-5.0*	-7.0*
	O	2.0	3.8*		-10.0	-2.0	-4.9*	-18.0*	-6.5	-6.5
	R	1.5	0.5		-13.3		-3.7*	-3.0*	-2.0	
	M						-12.5*			
	G	-9.0*	5.0		0.7	-3.5	3.6	-4.0*	-3.0	-3.0*
Average		-0.1	3.2*		-3.6	-0.7	-3.3*	-6.4*	-2.8	-2.0*
95% CI		-2.7	-0.6		-8.0	-2.7	-6.1	-12.0	-6.3	-4.0
		2.6	7.0		0.8	1.4	-0.5	-0.8	0.7	0.0
Season Length (Days)	B	-13.0*	-13.7*			-2.5	-4.4*	4.5	0.0	7.0*
	O	-2.5	-10.5*		-1.0	-1.5	-6.8*	13.0	6.5	-4.0
	R	2.5*	-1.5		-0.7		2.0	0.0	2.5	
	M						10.0*			
	G	-8.5*	-9.3		-10.7	-17.3	10.0*	-19.5	-7.5	-2.0*
Average		-1.9*	-6.3*		-1.9	-3.1	0.5	-0.3	0.3	-0.4
95% CI		-4.7	-12.2		-7.2	-9.0	-3.0	-9.0	-4.6	-2.4
		0.9	-0.4		3.3	2.9	4.0	8.3	5.2	1.7
Peak Value (%)	B	-38.0*	37.8*			162.4*	46.3*	24.8*	-39.0*	25.8*
	O	90.2	117.3*		90.1	57.4	93.2*	33.1*	-13.9	195.6
	R	92.3*	11.5		50.1		-2.1	-4.6*	-42.4*	
	M						-45.5*			
	G	46.0	-25.0*		105.0	-3.5	25.5	-0.2	21.3	17.3
Average		26.2*	29.5*		38.7	20.8	35.5*	9.2	-10.9	34.7
95% CI		-6.3	-1.9		-22.1	-14.8	8.1	-5.7	-32.8	-25.2
		58.6	60.9		99.6	56.3	62.8	24.2	10.9	94.7
Annual Production (%)	B	-80.2*	24.7*			92.9*	58.6*	44.9	-19.3*	36.2*
	O	108.4	36.3		56.9	61.5	95.4*	80.6*	-2.7	398.5
	R	99.9*	5.2		-23.5		-12.9	8.2*	-40.6*	
	M						-51.5*			
	G	115.4	6.8		135.3	4.2	16.3	34.4*	38.8	89.0
Average		35.4*	13.9*		26.6	17.3	33.8*	27.3*	-2.8	77.8
95% CI		-16.2	-2.3		-31.9	-14.7	2.7	1.7	-29.5	-37.7
		87.1	30.1		85.2	49.2	65.0	52.9	23.9	193.3

B: Birch; O: Oak; R: Ragweed; M: Mugwort; G: Grass;

As shown in Table 2.5, the allergenic pollen seasons for birch, oak, ragweed, mugwort, and grass during the decade of 2001-2010 started on average 2.3, 4.4, 4.0, 12.5 and 0.2 days earlier, respectively, than those during the period of 1994-2000. The average pollen seasons for spring-flowering birch, oak and grass during the past decade (2001-2010) were 4.4, 3.1 and 4.8 days shorter, respectively, than in the 1990s (1994-2000); while those for summer-flowering ragweed and mugwort were 1.3 and 10.0 days longer, respectively.

The average annual production of pollen for birch, oak, and grass pollen have increased 42.8%, 92.5% and 43.4%, respectively during the same periods. For ragweed and mugwort, the average annual productions have decreased by 3.1% and 51.5%, respectively. The average peak values for birch, oak, ragweed and grass pollen have increased 44.9%, 86.4%, 12.4% and 23.0%, respectively during the same periods; with the exception of mugwort, its PV decreased by 45.4%. These results are consistent with the results in Figure 2.9 and Table 2.4.

Table 2.5: Differences of mean pollen indices between periods of 2001-2010 and 1994-2000 in the contiguous US. 95% confidence intervals are included in the parentheses. The changes in a mean pollen index for a given taxa during two periods at all available stations were used to calculate the nationwide average and the 95% confidence intervals. Asterisk (*) indicates statistically significant difference at 5% level based on Students t test and Benjamini-Hochber control procedure (false discovery rate $<$ 5%).

	Start Date (Days)	Season Length (Days)	Peak Value (%)	Annual Production (%)	# of stations
Birch	-2.3 (-7.0, 1.9)	-4.4* (-8.8, -0.6)	+44.9* (7.9, 82.0)	+42.8* (4.6, 81.1)	19
Oak	-4.4* (-7.4, -1.5)	-3.1 (-7.0, 0.8)	+86.4* (37.9, 134.8)	+92.5* (29.4, 155.7)	28
Ragweed	-4.0 (-7.6, -0.4)	+1.3 (-1.1, 3.6)	+12.4 (-22.9, 47.7)	-3.1 (-30.0, 23.8)	20
Mugwort	-12.5 (-145.9, 120.9)	+10 (-66.2, 86.2)	-45.4 (-127.4, 36.5)	-51.5 (-179.0, 76.1)	2
Grass	-0.2 (-4.7, 4.3)	-4.8 (-13.7, 4.2)	+23.0 (-15.0, 61.0)	43.4 (-3.4, 90.3)	26
Average	-3.0* (-4.9,-1.1)	-2.6 (-5.4, 0.2)	+42.4* (21.9, 62.9)	+46.0* (21.5, 70.5)	31

Overall, the allergenic pollen seasons for five representative taxa started on average 3.0 (95% Confidence Interval, 1.1-4.9) days earlier during the past decade than during the 1990s across the CONUS (Table 2.5). Significantly earlier start dates (p value < 0.05, Student's t test with Benjamini Hochberg control procedure) are shown for 6.3% of the observations, with an average advancement of 17.0 (95% CI, 8.3-25.7) days in a decade; and 2.1% of the observations showed significantly later start dates than previously. The average advancement of allergenic pollen season onset in the past decade is consistent with the reported decadal advancements of phenology events (e.g., flowering) of trees, weeds and grasses[86–89]. Pollen seasons for spring-flowering allergenic taxa (birch, oak and grass) in the past decade appeared to be on average 3.1-4.8 days shorter than in the 1990s. Pollen seasons of summer-flowering taxa (ragweed and mugwort) appeared to be 1.3-10 days longer than previously.

The average allergenic airborne pollen levels have increased by 42.4% (95% CI, 21.9%-62.9%) and 46.0% (95% CI, 21.5%-70.5%) based on peak values and annual production, respectively (Table 2.5). For allergenic airborne pollen levels, 16.8% of the observations showed significant increase in annual production with an average increase of 179.9% (95% CI, 96.6%-263.2%); and 6.3% of the observations showed significant increase in peak value with an average increase of 283.6% (95% CI, 231.9%-335.4%).

2.3.3 Spatiotemporal patterns of changes of mean pollen indices

Changes of pollen indices across latitudes

Changes in average allergenic pollen season timing and airborne levels between the past decade and the 1990s were identified as functions of latitude (Figure 2.10). Changes in mean start date were found to decrease from later start to earlier start with increasing latitude. Changes in mean season length increased from shorter season to longer season with increasing latitude. Similar latitudinal effects on altered ragweed pollen season length in North America have been reported by Ziska *et al.*[35]. The latitudinal effects on average allergenic airborne pollen levels varied for different taxa. Overall, changes in average annual production appear to be large at higher latitudes and small at lower latitudes; while changes in average peak value appear to be small at higher latitudes and large at lower latitudes.

Allergenic pollen seasons for spring-flowering birch and oak start from the south and shift gradually towards the north, and their season lengths at lower latitudes are generally longer than those at higher latitudes. The enhanced warming at higher latitudes[22] leads to larger increases in GDD and FFD than at lower latitudes, and thus drives the allergenic plants at higher latitudes to flower earlier and last for a longer duration. This makes the start dates from north to south more synchronous and the

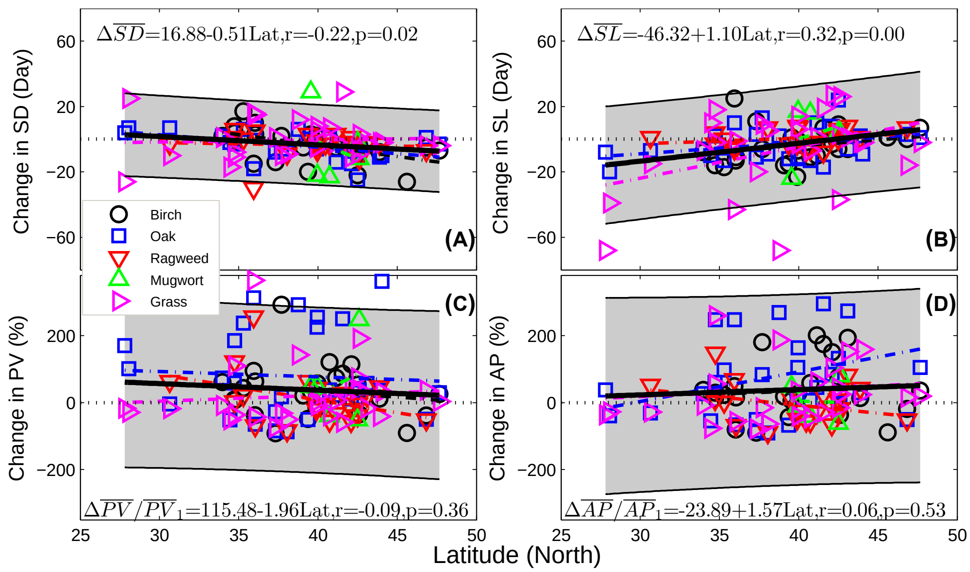

Figure 2.10: Changes in mean pollen indices between the periods of 2001-2010 and 1994-2000 as a function of latitude. (A) Start Date, (B) Season Length, (C) Peak Value, and (D) Annual Production. Heavy black lines represent the overall trends; dashed lines give trends for individual taxa; shaded gray area is the 95% CI of overall trend. Horizontal dotted lines are zero lines.

season length more uniform during the past decade than previously.

Variogram analysis

Variograms were calculated for mean pollen indices during periods of 1994-2000 and 2001-2010 to further investigate the synchronization of start date of spring flowering species, and the homogeneity of season length and airborne pollen production (Figure B.2). Figure 2.11 displays the changes of normalized semi-variograms for mean pollen indices between the periods of 1994-2000 and 2001-2010. The box plot for a given pollen index and species was generated using changes of normalized semi-variograms at different spatial lags. The symbols in Figure 2.11 are the same as those defined in Figure 2.9. The horizontal line is zero line representing no changes. Negative changes in variogram indicate that allergenic pollen seasons in the past decade appear to have more synchronous start date, uniform season length, and homogeneous peak value and annual production than in the 1990s across the CONUS.

The normalized semi-variograms of SD in the last decade for spring-flowering birch

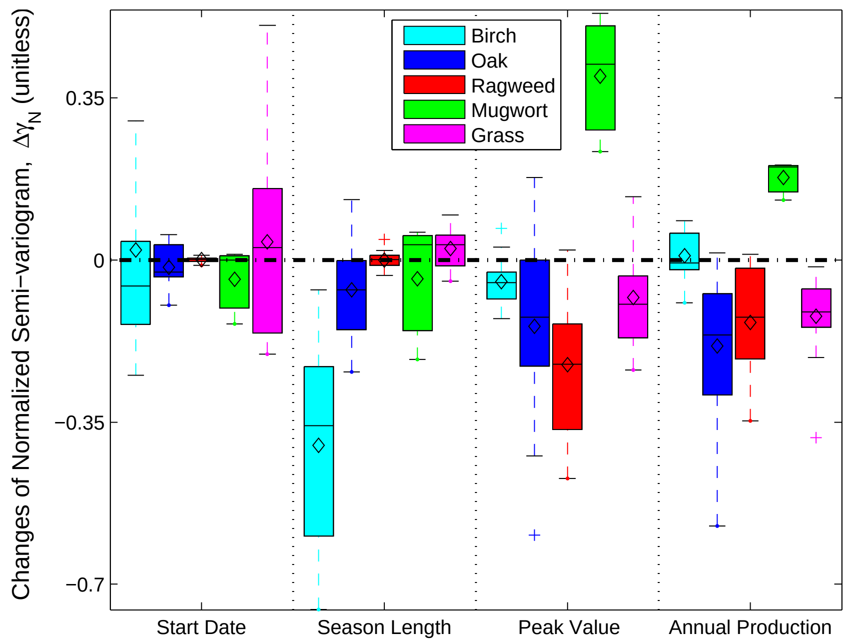

Figure 2.11: Changes in normalized semi-variogram for mean pollen indices during the period of 2001-2010 from those during 1994-2000 across the CONUS. The symbols are the same as defined in Figure 2.9. Negative changes in variogram indicate that allergenic pollen seasons in the past decade appear to have more synchronous start date, uniform season length, and homogeneous peak value and annual production than the 1990s across the CONUS.

and oak were generally lower than those in the 1990s (Figures 2.11 and B.2). This indicates that flowering of spring-flowering species in the past decade were more synchronous among different regions than previously. This is consistent with the reported flowering synchrony of birch trees in Finland during 1989-2006 [33]. The normalized semi-variograms for season length, particularly for birch and oak, were also smaller in the past decade than in the 1990s. This suggests pollen season lengths of birch and oak were more uniform among different regions in the past decade than previously.

Except for birch and mugwort, normalized semi-variograms of annual production and peak value during the past decade were smaller than in the 1990s (Figure 2.11). This indicates the annual production and peak value of allergenic airborne pollen appeared to be more homogeneous among different regions in the past decade than previously. The exception for birch is mainly caused by the area coverage of birch trees in the CONUS. Birch trees mainly distribute in the northern, particularly the northeastern CONUS, according to the Biogenic Emission Landuse Database (BELD) [85,68]. However, the enhanced precipitation and warming at higher latitudes may have favored the growth of birch trees and expanded their habitats in the northern CONUS, and thus increased birch pollen production in these areas.

For mugwort, the large positive changes in variograms for peak value and annual production are most likely caused by the scarcity of observed airborne pollen counts. Some of the variograms for mugwort, particularly for large spatial lags during the period of 1994-2000, could not be calculated due to scarce pollen data (Tables B.1 and 2.5).

2.3.4 Relationship with recent climate variation

Figure 2.12 presents the relationships between changes of mean pollen indices and changes of mean climatic factors during the periods of 2001-2010 and 1994-2000 across the CONUS. The trend lines for changes in peak value (Figure 2.12C) and annual production (Figure 2.12D) are divided into two stages at $\Delta Prc = 100$ mm. This precipitation change of 100 mm was roughly the "valley point" of the curves describing the relationships between change of airborne pollen level and change of precipitation.

The changes in mean start date are negatively related to changes in GDD between the past decade and the 1990s while the changes in season lengths are positively related to changes in FFD (2.12A and B). Accumulated precipitation during pollen season exerts dual effects on airborne pollen levels (2.12C and D). When the change of precipitation is less than 100 mm, increase of precipitation tends to reduce the airborne pollen levels. Conversely, when the change of precipitation is greater than 100 mm, increase of precipitation tends to increase the airborne pollen levels. The reason for dual effects of precipitation is discussed in section 2.3.5.

Figure 2.13 shows changes in mean pollen indices, Growing Degree Days (GDD), Frost Free Days (FFD) and accumulated precipitation between periods of 2001-2010 and 1994-2000 across latitudes. Increase of FFD, GDD and accumulative precipitation at higher latitudes cause allergenic pollen season in the north to start earlier, last longer and have higher pollen levels in the past decade than the 1990s. These results are consistent with those presented in Figure 2.12.

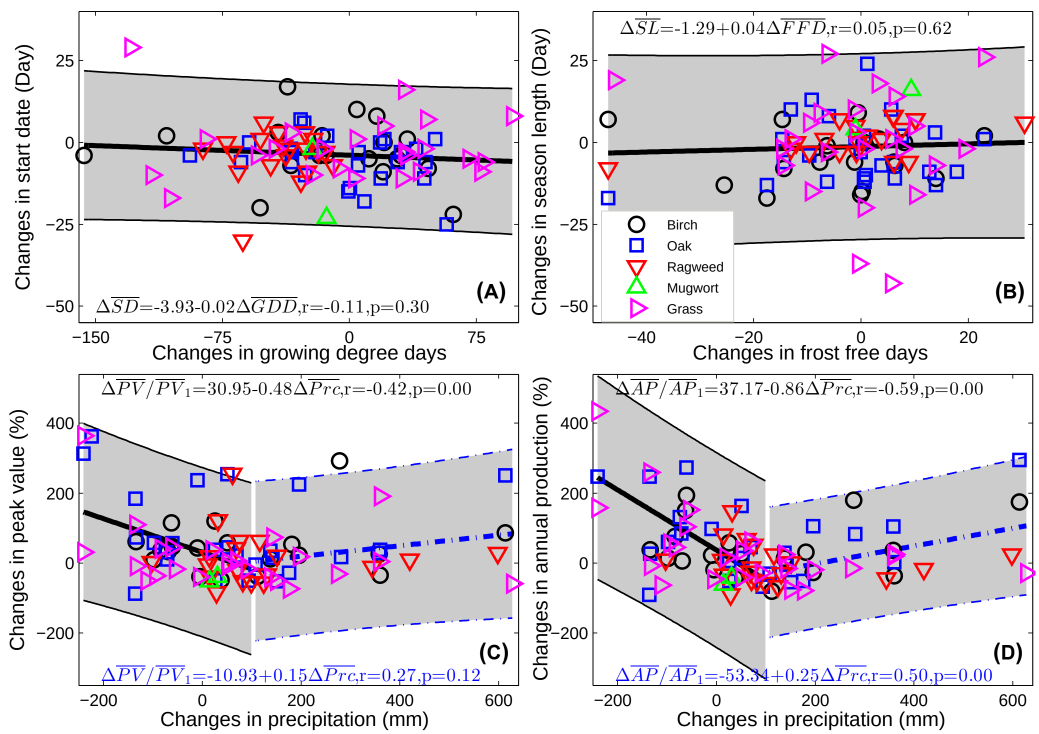

Figure 2.12: Changes in mean pollen indices and changes in mean climatic factors between periods of 2001-2010 and 1994-2000. (A) Start Date and GDD, (B) Season Length and FFD, (C) Peak Value and precipitation, and (D) Annual Production and precipitation. Heavy lines represent the trends; shaded gray areas are the 95% CIs. The initial and last dates and base temperature used to calculate GDD, FFD and accumulated precipitation are listed in Table 2.1. In panels (C) and (D), the trend line is divided into two stages at $\Delta Prc = 100$ mm to show the dual effects of precipitation on airborne pollen levels.

2.3.5 Impacts of temperature and precipitation

Over the past two decades, temperature and precipitation changes over North America have been larger at higher latitudes and altitudes[22]. This enhanced warming and precipitation at higher latitudes and altitudes has caused poleward and upward shifts of distribution ranges of plants and animals across different ecosystems[90,91]. The spatiotemporal patterns of changes in allergenic pollen season timing and airborne levels are likely due to the latitudinal patterns of temperature and precipitation in the Northern Hemisphere. The larger increase of temperature and precipitation at higher latitudes[22] caused larger changes in start date and annual production of allergenic pollen at higher latitudes (Figure 2.10A and D). Change of peak value and season length

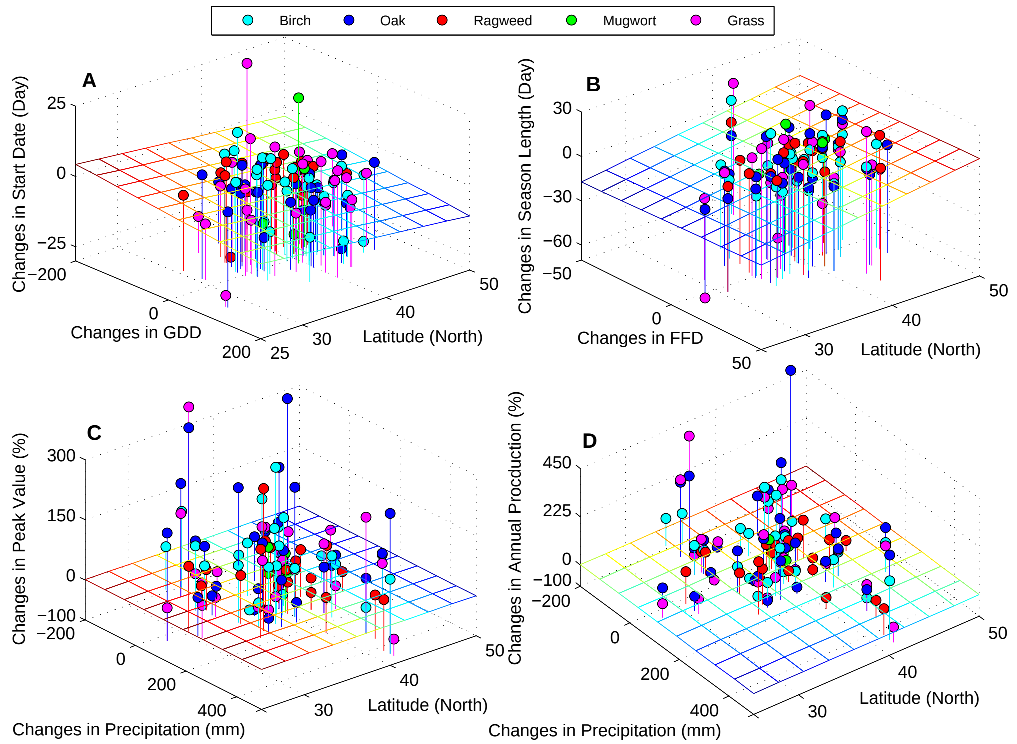

Figure 2.13: Changes in mean pollen indices, Growing Degree Days (GDD), Frost Free Days (FFD) and accumulated precipitation between periods of 2001-2010 and 1994-2000 across latitudes. A trend plane was also plotted in each of the subplots to show the changes of mean pollen indices and climatic factors across latitude. (A) Changes in Start Date and GDD, and Latitude; (B) Changes in Season Length and FFD, and Latitude; (C) Changes in Peak Value and Precipitation, and Latitude; (D) Changes in Annual Production and Precipitation, and Latitude.

may be dominated by changes in precipitation. Larger increase of precipitation and its frequency at higher latitudes washes out more airborne pollen during the pollen season, and thus reduces the peak value of airborne pollen at higher latitudes (Figure 2.10C). The reduced season length of allergenic pollen at lower latitudes is most likely caused by the decreasing temperatures in the South and Southeast regions in the CONUS [22], and those at middle latitudes are likely due to increased precipitation and rainy days.

On one hand, increasing precipitation can directly wash out more airborne pollen, and therefore decrease the peak values and annual total counts of airborne pollen. On the other hand, climate change, even on the scale of years to decades, can change the distributions and abundances of plants and animals[92–94]; a large increase in precipitation may favor the growth and expansion of habitat of allergenic plants at higher latitudes, at locations that have not been favorable for plant growth because of dry and cold conditions, thus increasing the production of airborne pollen.

The dual effect of precipitation on airborne allergenic pollen levels is particularly prominent at higher latitudes. If similar trends of enhanced warming and precipitation at higher latitudes continue, earlier exposure times and higher exposure levels to allergenic pollens may occur with potentially substantial consequences to public health. This will likely increase the prevalence (number of individuals becoming allergic) and the morbidity (severity and duration) of the population suffering from allergies and asthma.

2.3.6 Bayesian analyses on observed and projected birch pollen seasons

Bayesian model evaluation

The details of variable selection and parameterization of a Bayesian model are presented in the literature[70]. Modeling results are compared with corresponding observed values in Figure 2.14 for five different locations. Three diagonal lines have been plotted in each panel: the middle line has a slope of unity, the upper line has a slope of 2 or 1.25, and the lower line has a slope of 0.5 or 0.75. It is illustrated that the phenologically observed values of the four pollen indices can be well matched by the modeling values. Most of the points of annual productions and peak values either from Turku (Finland) or Basel (Switzerland) fall into the range between diagonal lines 0.5 and 2; and those of start and peak dates from five locations fit into the space between diagonal lines 0.75 and 1.25. The estimated pollen indices in the US stations can capture the trends, but the deviations are larger compared with the estimates for European locations because of the non-local parameterizations of the models.

Root mean square error (RMSE) and RMSE relative to mean value of pollen index were calculated to quantify the deviation between the observations and estimations. The relative RMSEs (RRMSE) are approximately 30%, 50% and 20% for estimates of annual productions and peak values in Basel, Turku, and Copenhagen, respectively.

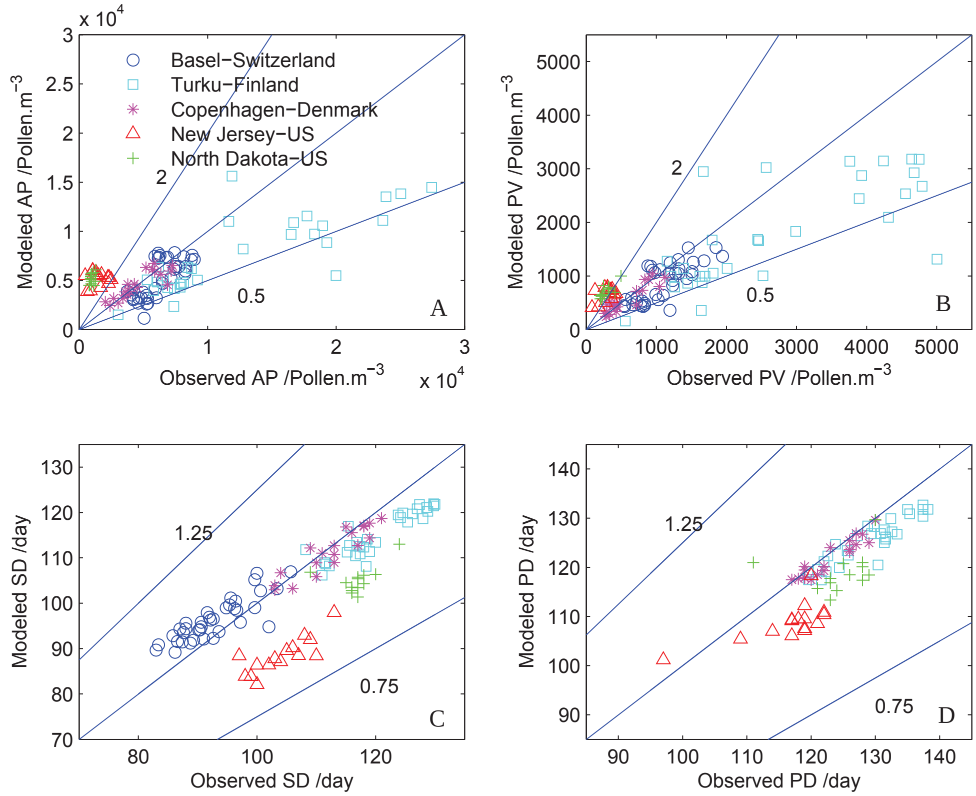

Figure 2.14: Comparison of pollen indices between the phenological observations and mean model estimations for five different locations. (A) annual production, (B) peak value, (C) start date, and (D) peak date. Three diagonal lines have been plotted in each panel: the middle line has a slope of unity, the upper line has a slope of 2 or 1.25, and the lower line has a slope of 0.5 or 0.75.

RRMSEs of estimates of start and peak dates for the three European locations are between 1.5% and 5.1%, which are much lower than those for annual production and peak value. For the US stations, the RRMSE of annual production and peak value range from 123.8% to 370.7%, and RRMSE of start and peak dates vary between 6.1% and 15.2%.

The deviations between estimations and observations are most likely due to the following: (1) For estimates of pollen indices in Basel, Turku, New Jersey and North Dakota, the Bayesian models were not parameterized with the local pollen and climate data; (2) The spring temperature and especially the annual mean CO_2 concentrations used in evaluations were not derived from the exact sites where the four pollen stations are located; (3) Because of the data availability of multiple climate factors, the Bayesian models used were not the optimum ones; and (4) The Bayesian models used to predict mean trends did not incorporate the information on inter-annual variation.

Bayesian model prediction

The historical estimates and future predictions of pollen indices under three representative IPCC scenarios B1, A2 and A1B, are presented on left and right, respectively, of Figure 2.15 using heavy lines. The top 5% HPD regions of future predictions were also calculated and are shown as a shaded area around the mean trends. Vertical dotted lines at 2010 identify the historical data and future predictions. Alternative development pathways are assumed in different IPCC scenarios which cover a wide range of demographic, economic and technological driving forces and resulting GHG emissions[95,22]. Scenario B1 assumes future development will be globally and environmentally oriented with projection of CO_2 level being 600 ppm in year 2100; and A2 assumes regionally and economically oriented development with projection of CO_2 level being 850 ppm; while A1B features with rapid economic growth and a balanced emphasis on all energy sources, and with projection of CO_2 level being 800 ppm. These emission scenarios in the fourth assessment report of IPCC have been replaced by Representative Concentration Pathways in the fifth assessment report[22].

Comparison between estimated pollen indices and historical observations indicate that the variations of pollen indices can be reasonably characterized by the estimates. Overall, the mean trends of historical pollen indices can be reasonably captured by the mean model estimates with the exceptions of pollen indices in two US stations where start and peak dates were systematically underestimated, and annual production and peak value were overestimated. Simple comparisons between global mean pollen indices in future years and the corresponding mean values in 2000 are summarized in the literature[70]. Under scenario B1, the global means of annual production and peak value in 2020 to 2040 will be 1.3-2.2 and 1.1-1.9 times as many as the mean values of 2000, respectively; while the start and peak dates will be 19 days and 23 days earlier, respectively. Under scenario A2, the annual production and peak values will be 1.4-2.5

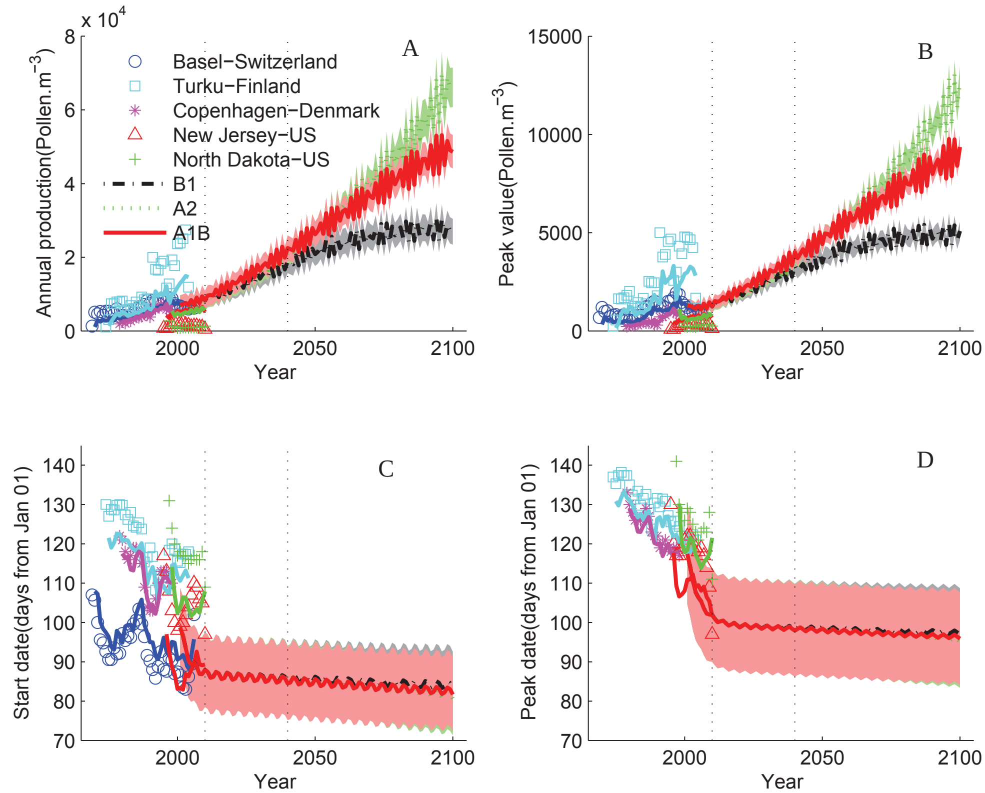

Figure 2.15: Predictions of mean trends of pollen indices based on the global annual mean temperatures and global annual mean CO_2 concentrations projected by the IPCC under three representative scenarios. (A) annual production, (B) peak value, (C) start date, and (D) peak date. Heavy lines are the mean trends and the corresponding shaded areas are top 5% HPD regions. Also shown on the left are time series of historical pollen indices and their corresponding mean trends calculated by the model.

and 1.2-2.2 times higher, respectively; while the start and peak dates will be also 19 days and 23 days earlier, respectively. Pollen indices under scenario A1B are similar to those of A2.

These ratios and differences are within the ranges reported in the literature [96]. The start and peak dates in 2000 were observed 14 days and 17 days earlier, respectively, than in 1977 [31]. Extreme observation has also been reported in Turku by [32] showing that the annual production of birch pollen in 1993 was 70,445 pollen grains/m^3 which was 119.4 times greater than that recorded in 1994. Further discussion and application of the developed Bayesian model are presented in the literature [70].

2.3.7 Predict oak pollen levels and concentrations using machine learning models

Classification of airborne pollen levels

Application of SVM to predict airborne pollen levels is used as an example here to demonstrate the process of parameter optimization and performance evaluation for machine learning model. Figure 2.16 presents the heat map of SVM classifier's cross-validation error rate as a function of regularization parameter and Gaussian kernel parameter. It shows that SVM classifier achieves the lowest cross-validation error rate of 12.3% with regularization parameter of 4 and Gaussian kernel parameter of 0.008.

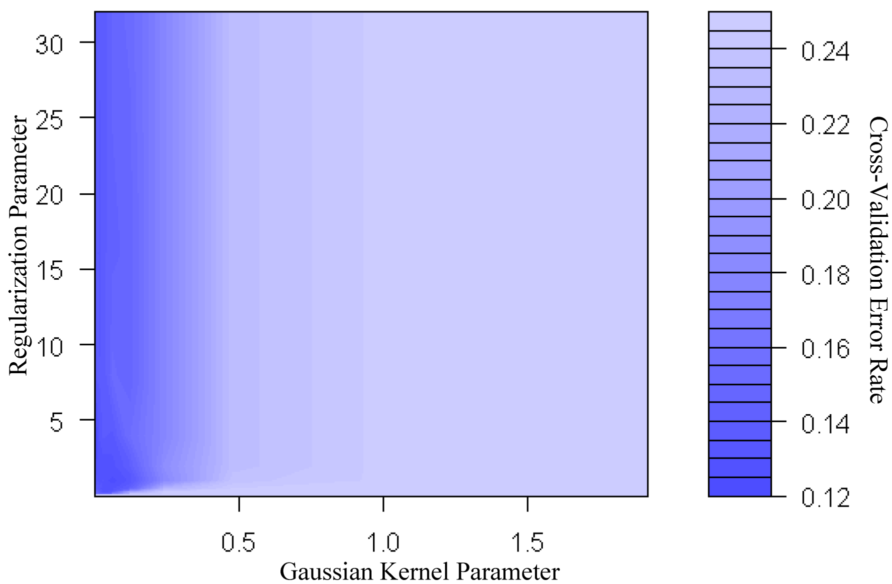

Figure 2.16: Heat map of SVM classifier's cross-validation error rate as a function of regularization parameter and Gaussian kernel parameter for prediction of airborne pollen levels

The learning curve of SVM classifier is presented in Figure 2.17 to check whether the parameterized SVM model has bias (i.e., underfitting) or variance (i.e., overfitting) issues. When the size of the data sample is smaller (< 200), the test error rate is higher than the training error rate. This means the SVM model is not trained sufficiently using training data, and therefore it can not perform equally well on test data. When the size of the data sample increases, the training and error rates converges to a lower value (around 12%). This means the SVM has been trained sufficiently, and can correctly capture the information in both training and test data. Since both the training and test error rates converges to a lower value, the SVM classifier is robust, and does not have bias (underfitting) or variance (overfitting) issues.

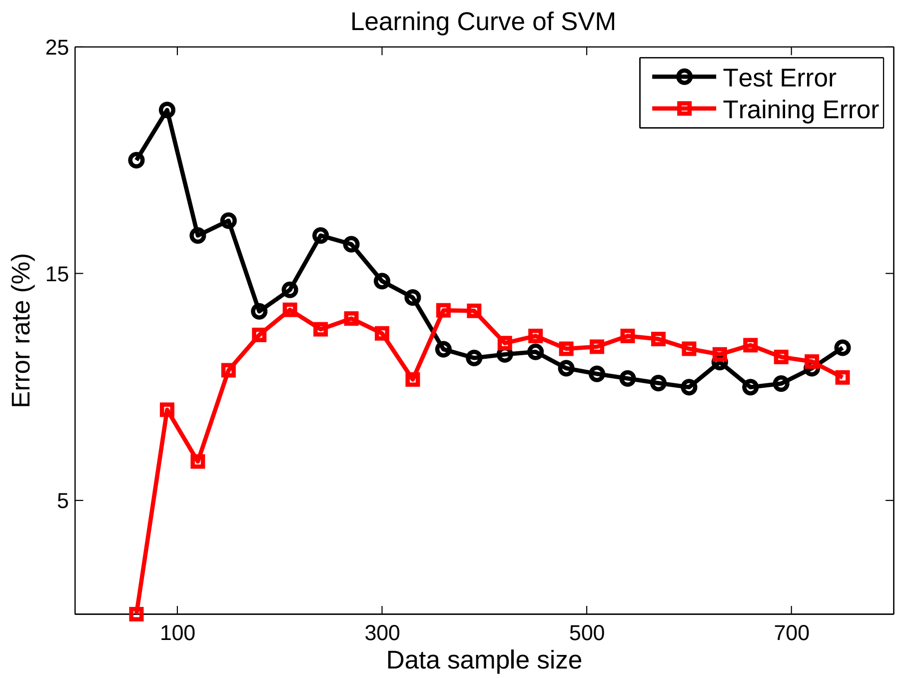

Figure 2.17: Training and test error rates of SVM classifier for observed data samples of different size.

The modeling structure of the decision tree classifier and the neural network classifier are presented in Figures B.3 and B.4, respectively. The optimum parameters and performance evaluation results for SVM, decision tree and neural network are summarized in Table 2.6.

Table 2.6: Machine learning models' configurations and their performance metrics on estimates of daily oak pollen level.

Model	R Package	Parameters	Accuracy	Precision	Recall	F1 Score
SVM	e1071	γ=0.008, C=4,	0.9034	0.8370	0.7273	0.7783
Decision Tree	rpart	Node, Leaf	0.9086	0.7870	0.7835	0.7852
NNW	nnet	1 hidden layer, 22 neurons	0.7859	0.4845	0.3485	0.4054

C, regularization parameter; γ, Gaussian kernel parameter.

Regression of airborne pollen concentrations:

Figure 2.18 presents the comparison between observed and predicted daily oak pollen concentrations during 1994-2010 in Springfield, New Jersey using support vector regression, regression tree and neural network. It shows that the majority of the points predicted by support vector regression and regression tree fall into the range between diagonal lines 0.5 and 2.0. The estimates from neural network can not capture the variations in the observed airborne pollen data. The poor performance of Neural Network could be due to: (1) the simple structure of hidden layer, and (2) linear activation function used for regression problems in R package nnet.

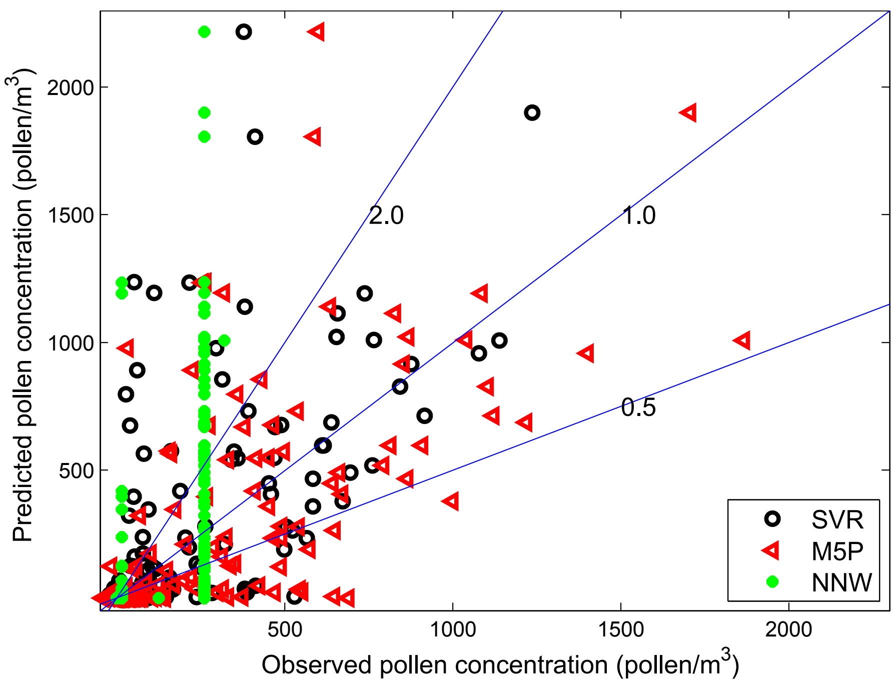

Figure 2.18: Comparison between observed and predicted daily oak pollen concentrations during 1994-2010 in Springfield, New Jersey using different machine learning models.

Figure 2.19 shows the comparison between observed and predicted time series of daily pollen concentrations during oak pollen season in 1996 in Newark, New Jersey using support vector regression, regression tree and neural network. The estimates from support vector regression and regression tree (M5P) match well with the observed pollen curves. As also shown in Figure 2.18, estimates from neural network fail to capture the trend and variation in the observed pollen count.

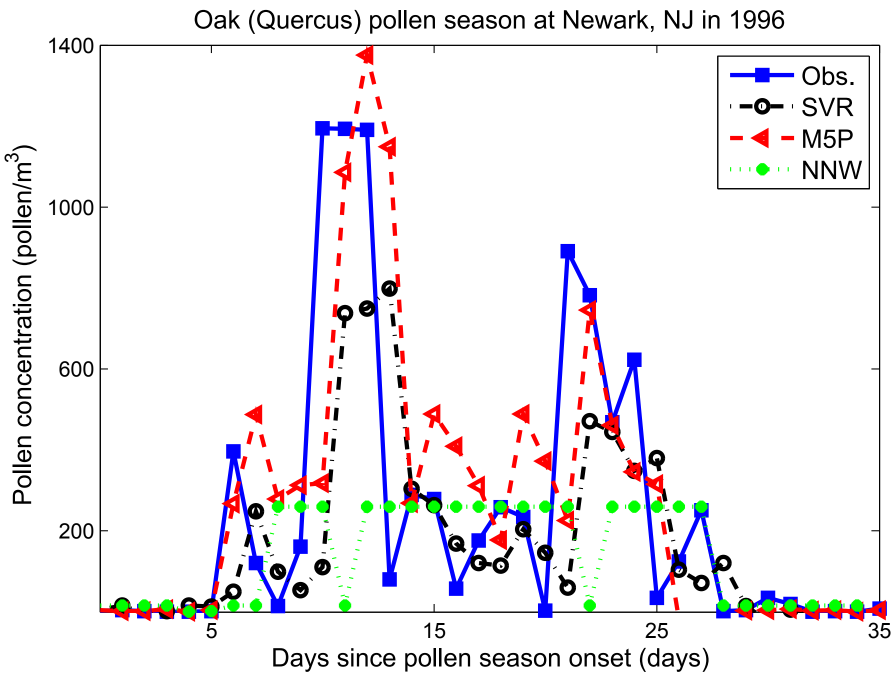

Figure 2.19: Comparison between observed and predicted time series of daily pollen concentrations during oak pollen season in 1996 in Newark, New Jersey using different machine learning models.

The modeling structure of regression tree and neural network classifier are presented in Figures B.5 and B.6, respectively. The optimum parameters and performance evaluation results for support vector regression, M5P regression tree and neural network are summarized in Table 2.7.

As shown in tables 2.6 and 2.7, algorithms based on support vector machine and tree outperformed those based on neural network for both classification and regression problems. For estimates of airborne pollen level, the SVM and decision tree achieved accuracy of around 90%, and an F1 score of around 78%. For estimates of airborne pollen concentration, the SVR and M5P have a correlation coefficient of around 0.7, and the index of agreement between 0.80 and 0.85.

For application of machine learning models to predict airborne pollen levels and concentrations, support vector machine could be the best choice. On one hand, SVM can take advantage of kernel methods to transform the low dimensional features into high dimensional space. This is particularly useful in the case of nonlinear classification based on low dimensional features. On the other hand, tree based algorithms are

Table 2.7: Machine learning models' configurations and their performance metrics on estimates of daily oak pollen concentration.

Model	R Package	Parameters	RMSE	Correlation Coefficient	Index of Agreement
SVR	e1071	γ=0.008, C=4, ε=0.05	202	0.6860	0.8049
M5P	RWeka	Node, Leaf	201	0.6964	0.8510
NNW	nnet	1 hidden layer, 20 neurons	274	0.5335	0.4778

C, regularization parameter; γ, Gaussian kernel parameter; ε, slack variable.

generally prone to overfitting the data for both classification and regression problems. It is hard to extrapolate the tree based models to estimate pollen levels and concentrations at other locations.

2.3.8 Uncertainty in observation-based analyses and statistical models

The variable number of NAB-AAAAI stations in the nine climate regions could potentially cause bias when we compare the allergenic pollen season variations among different climate regions (Figure 2.9). Specifically, since there are only three NAB-AAAAI stations in each of the Northwest, West North Central and Southwest regions there is a scarcity of data for these regions. To reduce this bias, Figure 2.10 was generated to account for allergenic pollen season variations across latitudes without confining the data to climate regions. Figures 2.9 and 2.10 should be considered together for comparison of allergenic pollen season variations among different regions and locations. Furthermore, incorporation of the airborne pollen data during the missing years and more recent years (e.g., 2011-2013) into the analyses could improve the results of the current study.

The causal attribution of changes in allergenic pollen season timing and levels to variation and trend of a single climatic factor in Figure 2.12 is substantially compounded by multiple other factors and their combinations [71,91,70]. The distances between NAB-AAAAI pollen stations and corresponding closest NOAA meteorology stations vary from a few kilometers to tens of kilometers depending on the stations. The mismatch of locations between pollen and meteorology stations may play a role in the weak relationships found in Figure 2.12.

Factors affecting pollen season timing and airborne levels interact in complex ways, and it may not be surprising to find a weak correlation with temperature or precipitation changes[91,43]. Population shifts and changes of land use in the proximity of the NAB-AAAAI counting stations may play an important role in determining the amount of airborne pollen collected at the corresponding stations[44,14,97]. Because of the fertilizer effect of CO_2 in the atmosphere, increase of CO_2 level itself or combined with rising temperature has been reported to substantially influence pollen and spore production[98–101]. Data describing these compounding factors (e.g., CO_2 level and land changes) are generally not available or very limited during the period of 1994-2010 for most of the NAB-AAAAI pollen stations.

The statistical relationship established from Bayesian analysis and machine learning models were derived on the basis of observed pollen and climatic data at discrete monitoring stations in the past years. It is advisable to be cautious to apply these statistical relationships to other locations. If local observations are available, it is always good to reparameterize the statistical models using local observations. For Bayesian analysis, in particular, the predictions after 2040 (second vertical dotted line) are expected to contain substantial uncertainties. Biological limitations and physics should be taken into consideration in terms of interpreting and using the predictions from Bayesian analysis and machine learning models.

2.4 Summary

The allergenic pollen seasons of representative trees, weeds and grass during the past decade (2001-2010) across the contiguous United States have been observed to start 3.0 (95% Confidence Interval (CI), 1.1-4.9) days earlier on average than in the 1990s (1994-2000). The average peak value and annual total of daily counted airborne pollen have increased by 42.4% (95% CI, 21.9%-62.9%) and 46.0% (95% CI, 21.5%-70.5%), respectively. Changes of pollen season timing and airborne levels depend on latitude, and are associated with changes in growing degree days, frost free days, and precipitation[45]. These changes are likely due to recent climate change and particularly the enhanced warming and precipitation at higher latitudes in the contiguous United States.

A Bayesian framework has been presented for modeling the effects of climate change on annual production, peak value, start date, and peak date of birch pollen [70]. Predictions of these models under three representative IPCC scenarios [95] indicate that annual productions and peak values of birch pollen will increase dramatically, while the start and peak dates of the birch pollen season will occur earlier in future years.

Support vector machine, neural network, decision and regression tree have been applied to estimate the daily airborne oak pollen levels and concentrations. For estimates of airborne pollen levels, the SVM and decision tree achieved accuracy of around 90%, and F1 score of around 78%. For estimates of airborne pollen concentration, the SVR and M5P had correlation coefficient of around 0.7, and index of agreement of between 0.80 and 0.85.

Chapter 3
EMISSION OF AIRBORNE ALLERGENS

3.1 Introduction

This chapter deals with the development of a mechanistic pollen emission model based on mass balance. The mechanistic emission model was developed based on the physical processes such as direct emissions and re-suspension of pollen particles, and accounted for meteorological parameters such as surface temperature, friction velocity, humidity, precipitation, etc, and information of land use and land cover. It also incorporates results from the analysis of observed pollen and meteorology data for estimating pollen season onset and duration, flowering likelihood and vegetation coverage. Different components of the emission model and their connections are illustrated in Figure 3.1, and described in the methods section.

3.2 Methods

The mechanistic pollen emission model was constructed based on mass balance of emission flux surrounding the plant crowns. As shown in Figure 3.2, the characteristic pollen concentrations c^* (Pollen/m^3) in the near surrounding of plant crowns depends on upward emission flux F_e (Pollen/(m^{2}h)), resuspension flux q_r (Pollen/(m^{2}h)), direct emission flux from plant crowns q_e (Pollen/(m^{2}h)), deposition flux F_s (Pollen/(m^{2}h)), and lateral emission fluxes F_L, F_R, F_B and F_F (Pollen/(m^{2}h)). This mass balance of pollen grains was formulated through equation 3.1,

$$\begin{aligned} H_C S_{TB} \frac{dC^*}{dt} =& (F_R - F_L) S_{RL} + (F_B - F_F) S_{BF} - F_e S_{TB} \\ & - F_s (S_{TB} + S_C) + K_e q_e S_C + K_r q_r (S_C + S_{TB}) \end{aligned} \tag{3.1}$$

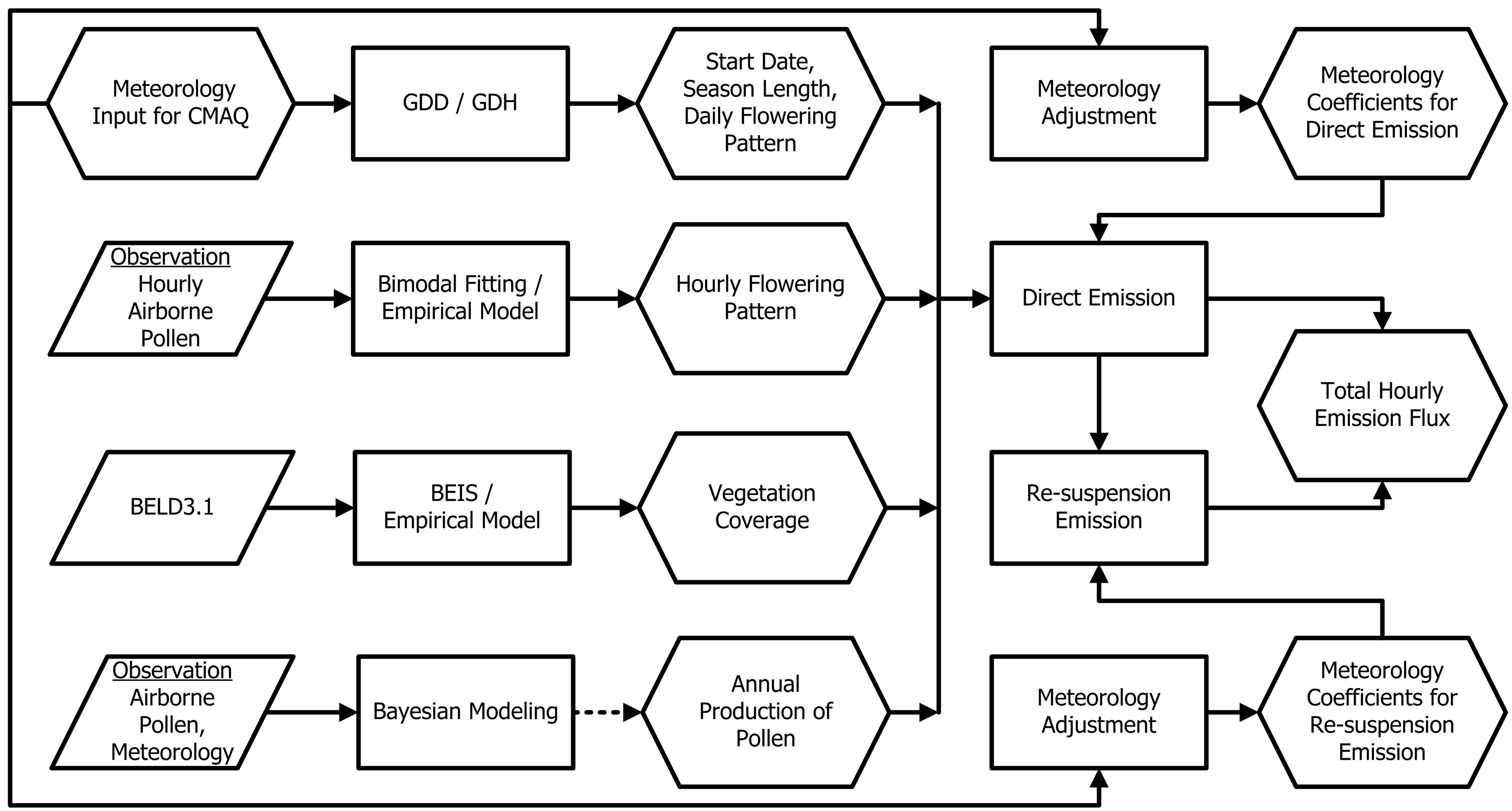

Figure 3.1: Diagram of pollen emission model. The mechanistic emission model consists of modules of direct emission, resuspension, meteorology adjustment, vegetation coverage, flowering likelihood, start date and season length.

where H_C (m) is the plant height. S_{TB}, S_{RL} and S_{BF} (m^2) are the areas of the top, left and back surface of the model box, respectively. S_C (m^2) is the surface area of the plant crowns. K_e and K_r (dimensionless) are the lumped meteorology adjustment factors for direct emission and resuspension fluxes, respectively.

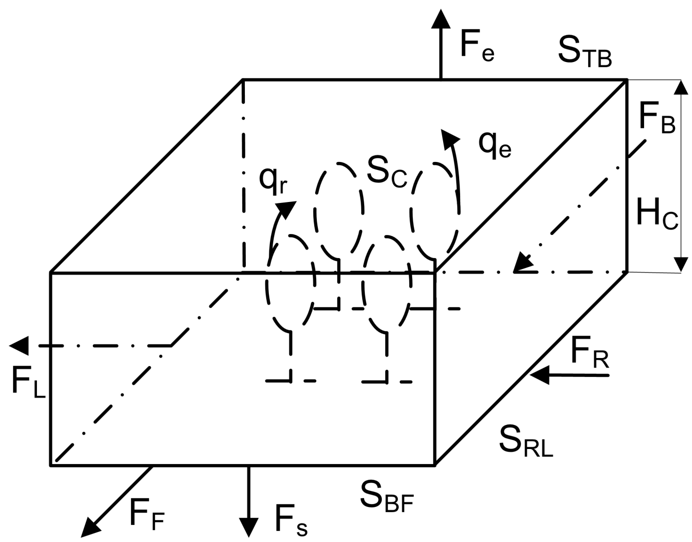

Figure 3.2: Schematic representation of pollen emission model. The emission flux of pollen grains depends on mass transfer of pollen in the near surrounding of plant crowns.

During the pollen season, two assumptions can be made: (1) quasi-steady state for mass transfer of pollen grains in the near surrounding of plant crowns (i.e., $\frac{dC^*}{dt} = 0$); and (2) lateral emission fluxes balance out (i.e., F_L=F_R, F_B=F_F). On the basis of these two assumptions, Equation 3.1 can be reduced to Equation 3.2,

$$F_e + F_s(1 + LAI) = K_e q_e LAI + K_r q_r (1 + LAI) \tag{3.2}$$

where LAI (dimensionless) is the Leaf Area Index, which according to definition can be used to approximate the quantity S_C/S_{TB}.

The upward emission flux F_e and deposition emission flux F_s can also be calculated according to the deposition velocity v_d and a characteristic velocity μ_* in the near surrounding of plant crowns as shown in equation 3.3.

$$\begin{cases} F_e = C^* \mu_* \\ F_s = C^* v_d \end{cases} \tag{3.3}$$

In the current study, μ_* is approximated using friction velocity as reported in the literature[46,38].

The resuspension emission q_r can be associated with direct emission flux q_e through a proportional factor C_r (dimensionless) as shown in equation 3.4,

$$\begin{cases} q_e = q_p L_d L_h \\ q_r = C_r q_e \end{cases} \tag{3.4}$$

where q_p (Pollen/(m^{2}yr)) is total emission flux during a pollen season, which can generally be measured and obtained from aerobiology literature. L_d (%) and L_h (%) are the daily and hourly flowering likelihood, respectively.

The upward emission flux F_e can thus be solved by the combination of equations 3.2-3.4. The resultant upward emission flux F_e is presented in equation 3.5,

$$F_e = \frac{q_p L_d L_h (K_e LAI + C_r K_r (1 + LAI))}{1 + v_d (1 + LAI)/\mu_*} \tag{3.5}$$

where all the terms in the right side can either be measured, or parameterized and approximated through measurable factors.

The pollen emission flux in a modeling grid with area of S_g can thus be calculated through equation 3.6,

$$F_g = F_e S_g P_c \tag{3.6}$$

where P_c (%) is the percentage of area coverage of allergenic plants in the corresponding modeling grid. Each of the terms in equations 3.5 and 3.6 are listed in Table 3.1 and described in the following subsections.

3.2.1 Vegetation coverage of allergenic plants

The area coverage of birch, oak and grass were sourced from the Biogenic Emissions Land use Database[85], version 3.1 (BELD3.1). Area coverage for birch, oak and grass were generated using Spatial Allocator[2] to redistribute the 1x1 km BELD3.1 data into 50x50 km grids of CONUS. The area coverage of ragweed and mugwort were generated using an algorithm on the basis of observed ragweed and mugwort pollen counts and vegetation coverage information from BELD3.1. As an example, the following discussion

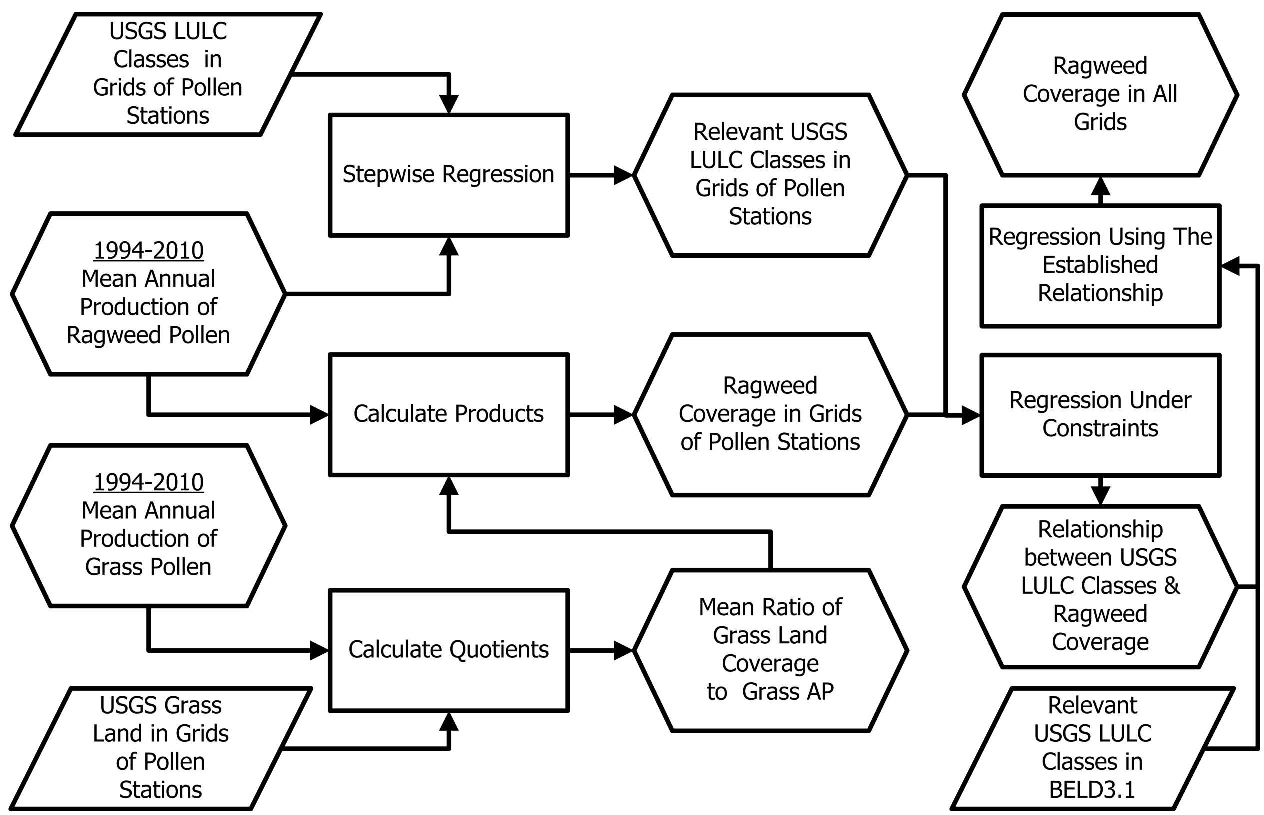

Figure 3.3: Schematic representation of the algorithm used to estimate ragweed coverage. The ragweed plant coverage is estimated based on land use and land cover information from BELD3.1 and ragweed pollen counts data from NAB-AAAAI.

focuses on development of ragweed coverage in each of the 50x50 km grids covering the CONUS. Figure 3.3 is a schematic representation of the algorithm for estimation of ragweed coverage in each grid. This is a two stage algorithm similar to the methods reported in the literature[102–104]. The first stage is to estimate the ragweed plant coverage in the grids which contain a monitoring station collecting the ragweed pollen count, and to identify the relationship between ragweed coverage and relevant land use and land coverage. The second stage is to estimate the ragweed plant coverage in the remaining grids using the relationship established in the first stage.

The pollen counts collected at each monitoring station are mainly from the ragweed plants in the grid that contains the monitoring station. The basic assumption of the algorithm is that the average ratio of grass area coverage to mean annual production of grass pollen is roughly the same as the average ratio for ragweed. This assumption is mathematically presented in equation 3.7,

$$\overline{P_R/\overline{AP_R}} \approx \overline{P_G/\overline{AP_G}} \tag{3.7}$$

where the $\overline{AP_R}$ and $\overline{AP_G}$ (pollen/m^3) are the mean annual production of ragweed and grass pollen, respectively. The P_R and P_G (%) are the area coverage of ragweed and grass plant in the corresponding grids, respectively.

The average ratio for grass, i.e., the right hand side of equation 3.7, can be calculated using grassland coverage and mean annual production of grass pollen in each of the grids containing a monitor station. The grassland coverage data are from BELD3.1. The mean annual production data are from grass pollen counts during 1994-2010 at the NAB-AAAAI monitor stations. The area coverage of ragweed plants in the grids (P_{Rg}) containing monitor stations, can therefore be estimated through equation 3.8.

$$P_{Rg} \approx \overline{AP_{Rg}} \times \overline{P_G/\overline{AP_G}} \tag{3.8}$$

The mean annual production of ragweed pollen from the available monitor stations was also used as a dependent variable of stepwise regression to select the relevant LULC classes in the corresponding grids. The LULC classes fed to stepwise regression include: urban land, dry crop land, crop grass land, crop wood land, grass land, shrub land, shrub grass land, savanna land, mixed forest land, sparse barren land, and wood tundra land. The area coverage of each LULC class in each grid can be obtained from BELD3.1. It was found that the mean AP of ragweed pollen was mainly relevant to area coverage of grass land, shrub land, crop grass land, and savanna land; and that the mean AP of mugwort pollen was mainly relevant to area coverage of grass land and shrub land.

The estimation of ragweed plant coverage in the remaining grids was generated using equation 3.9,

$$P_R = b_G P_G + b_{Sh} P_{Sh} + b_{CG} P_{CG} + b_{Sa} P_{Sa} \tag{3.9}$$

where P_{Sh}, P_{CG} and P_{Sa} (%) are the area coverage of shrub land, crop grass land, and savanna land, respectively. b_G, b_{Sh}, b_{CG} and b_{Sa} (dimensionless) are the corresponding coefficients. The coefficient b_{Sh} represents roughly the fraction of shrub land area occupied by ragweed plants, likewise for other coefficients. These coefficients were obtained by minimizing equation 3.10 under constraints,

$$\begin{aligned} & \min_{b_G, b_{Sh}, b_{CG}, b_{Sa}} (P_{Rg} - \hat{P_{Rg}})^2 \\ \text{Suject to:} \quad & 0 \leq b_G, b_{Sh}, b_{CG}, b_{Sa} \leq 1 \\ & 0 \leq (b_G P_G + b_{Sh} P_{Sh} + b_{CG} P_{CG} + b_{Sa} P_{Sa}) \leq 100 \end{aligned} \tag{3.10}$$

where $\hat{P_{Rg}}$ is the regressed value (equation 3.9) of ragweed plant coverage in grids containing monitor stations.

3.2.2 Flowering likelihood

Daily flowering likelihood

Daily flowering likelihood L_d is estimated based on the assumption that flowering likelihood increases gradually from the first flowering day to a peak in the middle and then decreases gradually to zero at the end of pollen season. The functional form for daily flowering likelihood was adapted from the literature[46]. It can be paramerized using Equation 3.11,

$$L_d(d) = c_b\left(\frac{d}{SL} - \frac{d^2}{SL^2}\right) \tag{3.11}$$

where d is number of days from the start date of pollen season, SL (days) is the season length, c_b is a normalizing constant which makes $\sum L_d = 1$. The start date and season length of allergenic pollen season are described later in section 3.2.3.

Hourly flowering likelihood

The bimodal characteristic has been observed for daily pollen release at flower scale [105]. This bimodal characteristic reflects mainly the diurnal and nocturnal features of pollen release. For birch, oak and grass, the hourly flowering likelihood L_h was constructed using two norm distributions with different mean and standard deviation as shown in equation 3.12,

$$L_h(t) = \alpha \frac{1}{\sqrt{2\pi}\sigma_d} e^{\frac{-(t-\mu_d)^2}{2\sigma_d^2}} + (1-\alpha) \frac{1}{\sqrt{2\pi}\sigma_n} e^{\frac{-(t-\mu_n)^2}{2\sigma_n^2}}, \quad t = 0, 1, ..., 23 \tag{3.12}$$

where t is hour number. α and $1-\alpha$ are the diurnal and nocturnal fractions of daily pollen release, respectively. μ_d and μ_n are the diurnal and nocturnal means of pollen release time, respectively. σ_d and σ_n are the diurnal and nocturnal standard deviations, respectively.

For birch, the fraction, means and standard deviations were obtained by fitting the data reported by Vogel et al. [47]. For oak, these parameters were obtained by fitting the data reported by Pasken et al. [42]. For grass, these parameters were obtained by fitting the data reported by Latalowa et al. [106]. The hourly or bi-hourly pollen count data for each species reported in the references were averaged over multiple days across multiple places depending on the data availability.

For mugwort, the flowering likelihood L_h was parameterized using Laplace distribution based on the features in the observed hourly pollen count. The distribution is presented in equation 3.13,

$$L_h(t) = \frac{1}{2\sigma_{t_0}} e^{-\frac{|t-t_0|}{\sigma_{t_0}}}, \quad t = 0, 1, ..., 23 \tag{3.13}$$

where the t_0 and σ_{t_0} are the location and scale parameters, respectively. These parameters were obtained by fitting the data reported by von Wahl et al. [107].

For ragweed, the flowering likelihood L_h was simulated using the algorithm developed by Martin et al. [105]. This algorithm relates the distributions of pollen emission with the relative humidity and elapsed time after sunrise. For simulation of the hourly flowering likelihood (L_h) for ragweed in a grid, the sunrise time was calculated according to the latitude and longitude of the grid; the hourly relative humidity was derived from the WRF simulation data.

3.2.3 Prediction of allergenic pollen season onset and duration

As shown in Figure 3.4, two observation-based models and one process-based model were developed to predict the SD and SL of allergenic pollen seasons of birch, oak, ragweed, mugwort and grass. The models were parameterized and cross validated using nationwide observations of airborne pollen data, and climate and/or meteorology data during the period of 1994-2010 in the CONUS. These three models were developed to: (1) provide a new scheme to handle the variable selection and parameter optimization in the field of predicting pollen season onset and duration; and (2) identify modeling approaches applicable to large geographic areas (e.g. continent) to simulate the SD and SL for multiple allergenic species at multiple spatiotemporal scales.

The sources of observed start date, season length and meteorology data have been described in sections 2.2.1 and 2.2.2.

Observation-based model (M1)

Allergenic pollen season onset and duration are associated with multiple climatic, meteorological and geographical factors. The associations can be generally described as $Y = f(T, P_r, Lat, H, GDD, Y_p, FFD, \Delta GDD, \Delta Y_p, \Delta FFD, \Delta T, \Delta P_r)$. Y is the vector of pollen season timing index ((SD, SL), days). T is a set which consists of the monthly, seasonal and annual mean temperatures (°C), representing the temperature effect on pollen season onset and duration. P_r is a set which consists of the monthly, seasonal or annual accumulative precipitations (mm), representing the precipitation effect on pollen season onset and duration. Monthly mean temperature and accumulative precipitation from September of the previous year to August of the current year were incorporated into the preliminary correlation analyses. Lat is the latitude (°N), H is the elevation above sea level (m), GDD is the growing degree days in a fixed period (Degree days), Y_p is the corresponding SD and SL in the previous year (days), and FFD is the Frost

Free Days (days). ΔGDD, ΔY_p, ΔFFD, ΔT and ΔP_r are the deviations in GDD, Y_p, FDD, T and P_r for a given year from the corresponding long term averages over the period of 1994-2010 at a given location, respectively. The deviation in independent variable X is defined using equation 3.14,

$$\begin{cases} \Delta X = X - \overline{X}_{Lat} \\ \overline{X}_{Lat} = b_0 + b_1 Lat \end{cases} \tag{3.14}$$

Where $\overline{X}_{Lat}$ and ΔX are the long term mean and deviations at corresponding latitude Lat, respectively; b_0 and b_1 are the coefficients.

For start date of ragweed and mugwort, the relationship among pollen season onset, duration, climatic, meteorological and geographical factors can be further formulated using equation 3.15,

$$\begin{cases} \Delta Y = Y - \overline{Y}_{Lat} \\ \overline{Y}_{Lat} = a_0 + a_1 Lat \\ \Delta Y = g(\Delta GDD, \Delta Y_p, \Delta FFD, \Delta T, \Delta P_r) \end{cases} \tag{3.15}$$

where $\overline{Y}_{Lat}$ are the long term mean SD and SL at latitude Lat, ΔY are the deviations in SD and SL from the corresponding long term mean, a_0 and a_1 are the coefficients.

The fixed-period GDD were calculated using Equation 2.1. The parameters ID, LD and T_b from Equation 2.2 were further optimized using simulated annealing[108,109] on higher resolution grids in the parameter space. For this further optimization using simulated annealing, ID took the value from January 1st, January 15th, February 1st, $\cdots$, December 1st and December 15th. LD took the value from January 15th, January 31st, February. 15th, $\cdots$, December 15th and December 31st; T_b assumed a value from -2 to 10 °C with an interval being 0.25 °C.

Simplified observation-based model (M2)

A simplified observation-based model can be generally represented as $Y = f(T, P_r, Lat, H, GDD, FFD, \Delta GDD, \Delta FFD, \Delta T, \Delta P_r)$. This simplified model does not incorporate the influence of allergenic pollen season in the previous year, namely Y_p and ΔY_p in

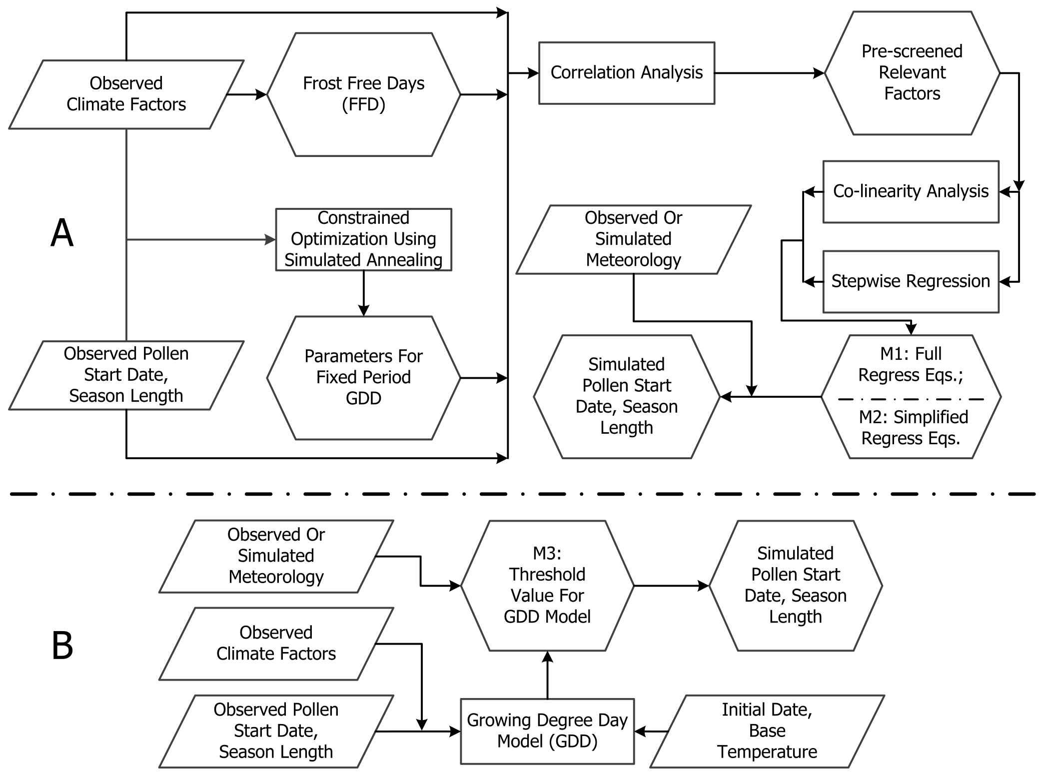

Figure 3.4: Schematic diagrams of the simulation methods. (A) Observation-Based model and its simplified version (M1 and M2 in Table 3.4), and (B) Growing Degree Day (GDD) model (M3 in Table 3.4).

model M1. Because of sparse collection of airborne pollen data in some regions for given years, it is usually difficult to derive precise Y_p and ΔY_p for these regions. In this situation, the simplified model can provide reasonable approximate estimates of SD and SL.

GDD model (M3)

The GDD model was adopted to describe the onset (i.e., SD) and end dates of allergenic pollen season. As shown in Equation 3.16,

$$\begin{cases} GDD_{Thr,SD} = \sum_{ID}(T_i - T_b), \quad T_i \geq T_b \\ GDD_{Thr,ED} = \sum_{ID}(T_i - T_b), \quad T_i \geq T_b \end{cases} \tag{3.16}$$

the start and end dates are simulated as the date when the accumulated temperature difference between daily temperature T_i and base temperature T_b reached threshold values $GDD_{Thr,SD}$ and $GDD_{Thr,ED}$, respectively. The SL is the interval between start and end dates. The ID and T_b are the optimum parameters based on an algorithm of simulated annealing.

The $GDD_{Thr,SD}$ and $GDD_{Thr,ED}$ may change slightly in different years even for the same species at the same location. In the current study, as shown in Equation 3.17,

$$\begin{cases} GDD_{Thr,SD} \approx \overline{GDD}_{Thr,SD} = c_{SD0} + c_{SD1}Lat \\ GDD_{Thr,ED} \approx \overline{GDD}_{Thr,ED} = c_{ED0} + c_{ED1}Lat \end{cases} \tag{3.17}$$

$GDD_{Thr,SD}$ and $GDD_{Thr,ED}$ at a given latitude are approximated using their long term averages ($\overline{GDD}_{Thr,SD}$ and $\overline{GDD}_{Thr,ED}$) over the period of 1994-2010 at that latitude. The c_{SD0}, c_{SD1}, c_{ED0} and c_{ED1} are coefficients.

Model parameterization

For observation-based models, first, correlation analyses were conducted between SD and each of the climatic factors on the basis of nationwide observations of airborne pollen data and climate and/or meteorology data in the US, likewise for SL. The factors considered in the current study include annual mean temperature, annual accumulative precipitation, monthly mean temperature and accumulative precipitation from

September of the previous year to August of the current year, *GDD*, *FFD*, latitude, elevation, and *SD* and *SL* in the previous year. Three or four of the factors with the highest correlation coefficients were prescreened to incorporate the influence of pollen information in the previous year, and temperature and precipitation in both current and previous years. Second, the prescreened factors were further selected through stepwise regression, and collinearity analyses based on Variance Inflation Factor (VIF). VIF between climatic factors was required to be less than 5 to insure the independence of the final selected factors. Finally, the selected regression equation was parameterized using nationwide observed *SD* or *SL* and climatic factors.

For the *GDD* model, first, a *GDD* threshold value at a given station was calculated using observed *SD* or *SL* and meteorology data from that station for each available year during 1994-2010. Second, the long term averaged threshold *GDD* over the period of 1994-2010 (i.e., $\overline{GDD}_{Thr,SD}$ and $\overline{GDD}_{Thr,ED}$) at each station was calculated to parameterize its relationship with latitude (Equation 3.17).

As shown in Table B.1 in the Appendix, airborne pollen data are missing for some of the species in some years at some stations. The variable selection and parameterization are only based on the observed airborne pollen data available for each of the species at each of the NAB-AAAAI stations during 1994-2010. The missing data were not counted towards the variable selection and parameterization of the models. The parameterization processes are depicted using schematic diagrams in Figure 3.4.

Model evaluation

The model performance was evaluated using root mean square errors (RMSE) and deviations between observed and simulated mean SD and SL at each of the NAB-AAAAI stations during 1994-2010. The model performance was also validated using cross-validation by splitting the data into a training set and a validation set [110,111]. I used a leave-one-out cross validation procedure, in which one year's data (e.g., 1994) were held out as a validation set, and the data of remaining years (e.g., 1995-2010) were used as a training set. The model was first trained using the training set and then used to predict the *SD* and *SL* for the validation year. The process was repeated for each of the years during 1994-2010 to check for validation errors.

Model application

The observation-based model M1 was applied to generate the spatially resolved SD and SL of allergenic pollen seasons for 2001-2004 and 2047-2050 in the US with spatial resolution of 50x50 km using the NARCCAP archived meteorology data. For each of the 50x50 km grid cells, the M1 method for SD and SL was executed using the dependent variables in the corresponding cell. The SD_p and SL_p in each cell, required by the observation-based model, were approximated using the long term mean SD and SL through equations in Figure 2.8 A and B. The simulated results were mapped only on cells in which the area coverage of a given allergenic species is greater than zero (section 3.2.1). These spatially resolved maps of SD and SL were used to drive the pollen emission model and transport model for studying spatiotemporal distributions of allergenic pollen in the CONUS under the changing climate.

3.2.4 Meteorology adjustment factors

Meteorological adjustment factor K_e is mainly related to the friction velocity, which is important to entrain the pollen from flower to atmosphere. Threshold friction velocity u_{*t} is required to activate the saltation process leading to dust entrainment. Shao *et al.*[112] introduced a physical parameterisation of u_{*t} for dry and bare soils as shown in equation 3.18,

$$u_{*t} = (\alpha_1[\rho_p g d_p/\rho_a + \alpha_2/(\rho_a d_p)])^{1/2} \tag{3.18}$$

where factors $\alpha_1 = 0.0123$ and $\alpha_2 = 3 \times 10^{-4}$ (kg/s) are defined on the basis of the results obtained from a wind tunnel experiment, ρ_p (kg/m^3) and ρ_a (kg/m^3) are the pollen and air densities, respectively, d_p is the diameter of pollen.

Since pollen is intrinsically different from soil, a modified friction velocity was introduced through equation 3.19 by Helbig *et al.*[46] to account for the meteorological effect on pollen emission,

$$\begin{cases} u_{*te} = A u_{*t} \\ u_{*tr} = B u_{*t} \end{cases} \tag{3.19}$$

where A and B are the meteorological coefficients for direct emission and resuspension emission, respectively. They were further parameterized using equation 3.20,

$$\begin{cases} A = \frac{3}{\alpha_T + \alpha_U + \alpha_V} \\ B = \frac{2}{\beta_U + \beta_V} \end{cases} \tag{3.20}$$

where α_T, α_U and α_V are the resistances due to temperature T (°C), relative humidity U (%) and wind speed at 10 m V (m/h), respectively; β_U and β_V are the corresponding resistances for resuspension emission.

The three resistances were further parameterized using equation 3.21,

$$\begin{cases} \alpha_T = c_{Te} T / T_{te}, \quad \alpha_U = c_{Ue} U_{te} / U, \quad \alpha_V = c_{Ve} V / V_{te} \\ \beta_U = c_{Ur} U_{tr} / U, \quad \beta_V = c_{Vr} V / V_{tr} \end{cases} \tag{3.21}$$

where c_{Te}, c_{Ue} and c_{Ve} are species specific constants, T_{Te}, U_{Ue} and U_{Ve} are the threshold values of temperature, relative humidity and wind speed for direct pollen emission. c_{Ur}, U_{tr}, c_{Vr} and V_{tr} are the corresponding constants and threshold values for resuspension emission.

Finally, the meteorological adjustment factors K_e and K_r were obtained through equation 3.22 by modifying the method from Helbig et al.[46]. Equation 3.22 indicates that higher ground surface temperature, wind speed, and lower humidity favor pollen release.

$$\begin{cases} K_e = 1 - e^{-u_*/u_{*te}} \\ K_r = 1 - e^{-u_*/u_{*tr}} \end{cases} \tag{3.22}$$

3.2.5 Deposition velocity

Deposition velocity was calculated using the resistance model[113] as presented in equation 3.23,

$$v_d = \frac{1}{r_a + r_b + r_a r_b v_s} + v_s \tag{3.23}$$

where the r_a and r_b (hr/m) are the aerodynamic resistance and quasi-laminar resistance, respectively. v_s (m/hr) is the settling velocity, which was calculated using Stokes' equation 3.24,

$$v_s = \frac{\rho_p d_p^2 g C_c}{18\mu} \tag{3.24}$$

where air dynamical viscosity μ can be found to be 1.8×10^{-5}kg/(s m)[54]. C_c is the slip correction factor used to correct the noncontinuum effects of particles in the air according to their diameters[113]. In the current study, it takes the value of 1.008 for pollen grains, which have an average diameter of around 20 μm.

The aerodynamic resistance r_a and quasi-laminar resistance r_b were calculated using equation 3.25[113],

$$\begin{cases} r_a = \frac{\ln(H_C)/z_0}{\kappa\mu_*} \\ r_b = \frac{1}{\mu_*(Sc^{-2/3}+10^{-3/St})} \end{cases} \tag{3.25}$$

where κ=0.41 is the von Karman constant; $Sc = \gamma/D$ is the Schmidt number; $St = v_s\mu_*^2/(g\gamma)$ is the Stokes number; $D = k_BTC_c/(3\pi\mu d_p)$ is the molecular diffusivity; and $\gamma = \mu/\rho_a$ is the kinetic viscosity.

3.2.6 Sensitivity analyses on pollen emission model

Global sensitivity analyses were performed to test the sensitivity of the pollen emission model to multiple inputs and parameters based on Morris's design[114]. This design estimates the main effect of a parameter by computing a number of local sensitivities at random points of the parameter space. The mean of these randomized local sensitivities indicates the overall influence of a given parameter on the output metric, while the corresponding standard deviation indicates the effects of interaction and nonlinearity[115].

For birch and oak, the emission model was run from March 1st to April 30th, 2004 covering the CONUS with spatial resolution of 50x50 km and temporal resolution of one hour. For ragweed and mugwort, the emission model was run from August 1st to September 30th, 2004 covering the same domain with the same resolution. For grass, the emission model was run from March 1st to June 30th, 2004 covering the same domain with the same resolution.

For investigation of the spatiotemporal pattern, the mean ($F_{g,hrMn}$), maximum ($F_{g,hrMx}$), seasonal total ($F_{g,HrSum}$) and standard deviation ($F_{g,hrStd}$) of the simulated

Table 3.1: Parameters for pollen emission model. These parameters were derived from the literature, and also used for global sensitivity analysis.

Parameter and ID	Birch	Oak	Ragweed	Mugwort	Grass	Reference
1 H_c, plant height (m)	15	30	0.69	1.5	0.35	[116–122]
2 C_r, proportional factor (unitless)	0.7	0.7	0.7	0.7	0.7	[46]
3 q_p, annual emission flux (pollen grain /(m^{2}yr))	1.4×10^9	1.0×10^9	2.8×10^9	2.8×10^9	9.4×10^8	[26,40,120,123,124]
4 LAI, leaf area index (m^2/m^2)	1.5	3.4	1.2	1.2	2.0	[125–127,65,128]
5 μ_*, friction velocity (m/h)	WRF Data	WRF Data	WRF Data	WRF Data	WRF Data	[55,56]
6 c_{Te},c_{Ue},c_{Ve}, correction factor for direct emission (unitless)	1	1	1	1	1	[46]
7 c_{Ur},c_{Vr}, correction factor for resuspension (unitless)	1	1	1	1	1	[46]
8 T_{te}, threshold temperature for direct emission (°C)	10	10	0	10	10	[46,52]
9 U_{te}, threshold relative humidity for direct emission (%)	60	60	60	60	60	[46]
10 V_{te}, threshold velocity for direct emission (m/s)	2.65	2.65	2.9	2.65	2.65	[46,52]
11 U_{tr}, threshold relative humidity for resuspension (%)	85	85	85	85	85	[46]
12 V_{tr}, threshold velocity for resuspension (m/s)	0.9	0.9	0.9	0.9	0.9	[46]
13 r_a, aerodynamic resistance (hr/m)	Equation3.25	Equation3.25	Equation3.25	Equation3.25	Equation3.25	[113]
14 r_b, quasi-laminar resistance (hr/m)	Equation3.25	Equation3.25	Equation3.25	Equation3.25	Equation3.25	[113]
15 v_s, settling velocity (m/h)	Equation3.24	Equation3.24	Equation3.24	Equation3.24	Equation3.24	[113]
16 P_c, percentage of area coverage (%)	BELD3.1	BELD3.1	Equation3.9	Equation3.9	BELD3.1	[85]
17 L_d, daily flowering likelihood (%)	Equation3.11	Equation3.11	Equation3.11	Equation3.11	Equation3.11	[46]
18 L_h, hourly flowering likelihood (%)	Equation3.12	Equation3.12	Literature	Equation3.13	Equation3.12	[47,42,106,107,105]
19 u_{*t}, threshold friction velocity (m/s)	Equation3.18	Equation3.18	Equation3.18	Equation3.18	Equation3.18	[46]
20 z_0, surface roughness (m)	10	10	0.1	0.1	0.1	[113]
21 d_p, diameter of pollen grain (μm)	22	28	18	21	35	[54,129,130]
22 ρ_p, density of pollen grain (kg/m^3)	800	1200	1280	1280	968	[54,129,130]
23 C_c, slip correction factor (unitless)	1.008	1.008	1.008	1.008	1.008	[113]

hourly emission in each 50x50 km grid were calculated using equation 3.26,

$$\begin{cases} F_{g,hrMn}(i,j) = \frac{\sum_t F_g(i,j,t)}{N_t} \\ F_{g,hrMx}(i,j) = \max_t F_g(i,j,t) \\ F_{g,hrSum}(i,j) = \sum_t F_g(i,j,t) \\ F_{g,hrStd}(i,j) = \frac{\sum_t (F_g(i,j,t) - F_{g,hrMn}(i,j))^2}{N_t} \end{cases} \tag{3.26}$$

where $F_g(i,j,t)$ is the pollen emission flux in grid (*i*,*j*) at time t; N_t is the number of temporal index t. These four emission metrics $F_{g,hrMn}(i,j)$, $F_{g,hrMx}(i,j)$, $F_{g,hrSum}(i,j)$ and $F_{g,hrStd}(i,j)$, hereafter are also sometimes referred to as hourly mean, hourly maximum, seasonal total and standard deviation of pollen emission.

The regional mean hourly emission (F_{hrMn}), maximum hourly emission (F_{hrMx}), mean seasonal emission (F_{SnMn}), and maximum seasonal emission (F_{SnMx}) were selected as metrics for testing the emission model's sensitivity to multiple inputs and parameters. Definition of these metrics are presented in equation 3.27,

$$\begin{cases} F_{hrMn} = \frac{\sum_{i,j,t} F_g(i,j,t)}{N_i N_j N_t} \\ F_{hrMx} = \max_{i,j,t} F_g(i,j,t) \\ F_{SnMn} = \frac{\sum_{i,j} (\sum_t F_g(i,j,t))}{N_i N_j} \\ F_{SnMx} = \max_{i,j} (\sum_t F_g(i,j,t)) \end{cases} \tag{3.27}$$

where N_i and N_j are the number of spatial indices i and j, respectively. These four pollen emission metrics, F_{hrMn}, F_{hrMx}, F_{SnMn} and F_{SnMx}, hereafter are also sometimes referred to as regional hourly mean, regional hourly maximum, regional seasonal mean and regional seasonal maximum pollen emission, respectively.

In the current study, each of the 23 parameters (Table 3.1) was sampled 12,000 times according to Morris' method from 500 random trajectories (each has 24 steps) in the parameter space[114,115]. Each of the parameters was perturbed between 50% and 150% of its base value or distribution while keeping other parameters unchanged. Equation 3.28 was used to calculate the Normalized Sensitivity Coefficient (NSC) for regional hourly mean emission at a local point:

$$NSC_{hrMn} = \frac{\Delta F_{hrMn}/F_{hrMn}}{\Delta P/P} \tag{3.28}$$

where F_{hrMn} and P are the regional mean hourly emission flux and the input parameter, respectively; and ΔF_{hrMn} and ΔP are the perturbations in the emission flux and input parameters, respectively. The local NSCs for regional maximum hourly emission, regional mean seasonal and regional maximum seasonal emission were calculated in the same way as in equation 3.28.

The global NSC of a parameter, NSCg, is defined as the mean of the corresponding local sensitivities. The average absolute global NSC, $\overline{|NSCg|}$, for each parameter and pollen taxon can be derived based on means of the absolute NSCg. Similarly, the standard deviations averaged over each parameter and pollen taxon ($\overline{STD}$) can be obtained to evaluate the interaction and nonlinearity effect of input parameters on modeling output.

3.3 Results and Discussion

3.3.1 Vegetation coverage map of allergenic plants

Table 3.2 lists the optimum coefficients in Equation 3.9, which were used to calculate the area coverage of ragweed and mugwort.

Table 3.2: Coefficients used in the developed algorithm to calculate the area coverage of ragweed and mugwort plants.

Land Class **Coefficient**	**Grass** b_G **(unitless)**	**Shrub** b_{sh} **(unitless)**	**Crop Grass** b_{CG} **(unitless)**	**Savanna** b_{Sa} **(unitless)**
Ragweed	0.7684	0	0.5000	0.7497
Mugwort	0.2548	0.0598	0	0

Figure 3.5 presents the percentage of the area occupied by birch, oak, ragweed, mugwort and grass in each of the 50x50 km grids covering the CONUS. Birch trees mainly distribute along the east coast of the CONUS, in particular, Maine, Vermont and New Hampshire have the highest area coverage (12.1%-16.0%) of birch trees. Oak trees distribute across eight of nine climate regions in the CONUS, with the highest area coverage (28.1%-51.0%) in the West, Southeast and South climate regions.

Grass distributes in most of the states across the CONUS. The highest area coverage (65.1%-98.4%) are found in the South and West North Central climate regions. Ragweed and mugwort mainly distribute in the western US. Their distribution shows a pattern similar to that of grass. The area coverage of ragweed in the South and West North Central climate regions is between 60.1% and 76.0%. The area coverage of mugwort in the South and West North Central climate regions is between 20.1% and 25.1%.

These vegetation coverage maps are important inputs to the pollen emission model. They were used in Equation 3.6 to calculate the allergenic pollen emission fluxes in each of the 50x50 km grids in the CONUS.

3.3.2 Flowering likelihood

Table 3.3 lists the parameters and references for calculating the hourly flowering likelihood. All parameters were derived by fitting the hourly flowering likelihood function using the hourly pollen counts.

Table 3.3: Parameters for calculating hourly flowering likelihood.

Parameter	Birch	Oak	Ragweed[a]	Mugwort	Grass
α, daytime fraction (unitless)	0.762	0.533			0.847
μ_d, daytime mean (hr)	9.0	4.6			12.0
μ_n, nighttime mean (hr)	19.0	17.7			21.2
σ_d, daytime deviation (hr)	5.4	3.4			6.1
σ_n, nighttime deviation (hr)	3.4	3.3			1.9
t_0, location parameter (hr)				9.6	
σ_{t_0}, scale parameter (hr)				2.1	
Reference	[47]	[42]	[105]	[107]	[106]

[a] Calculated based on the method reported by Martin *et al.*[105].

Figure 3.6 presents the hourly flowering likelihood estimated from our models and those derived from observed hourly pollen counts. The bimodal features of hourly flowering likelihood observed in Figure 3.6 for birch, oak, ragweed and grass reflect the pollen emissions in early morning and late afternoon. The simulated hourly flowering likelihood could capture the main features observed in the hourly pollen counts.

These hourly flowering likelihood functions were used to represent the diurnal emission pattern in the pollen emission model through Equation 3.5. In each of the 50x50

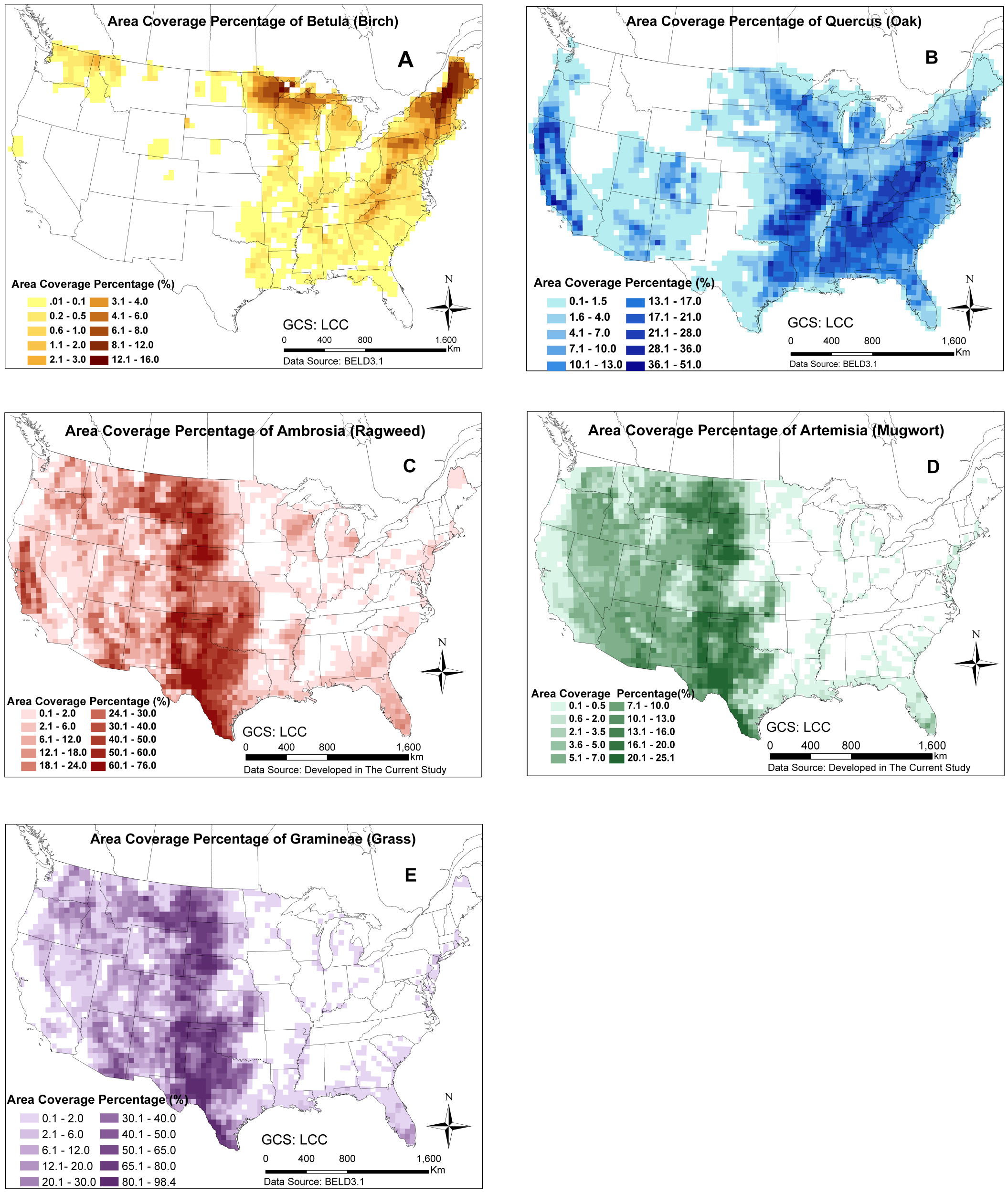

Figure 3.5: Area coverage of allergenic plants in the contiguous United States with spatial resolution of 50x50 km. (A) Birch, (B) Oak, (C) Ragweed, (D) Mugwort, and (E) Grass. Area coverage for birch, oak and grass was generated using Spatial Allocator [2] to redistribute the 1x1 km BELD3.1 data into 50x50 km grid. Area coverage for ragweed and mugwort was generated using the algorithm developed in section 3.2.1.

km grids, daily and hourly flowering likelihoods were calculated on a given day based on the methods presented in section 3.2.2. These hourly flowering likelihoods were then normalized using the sum of L_h on a given day, so that they added up to 1 from hour 0 to hour 23. The normalized L_h were then used to drive the pollen emission model through Equation 3.5.

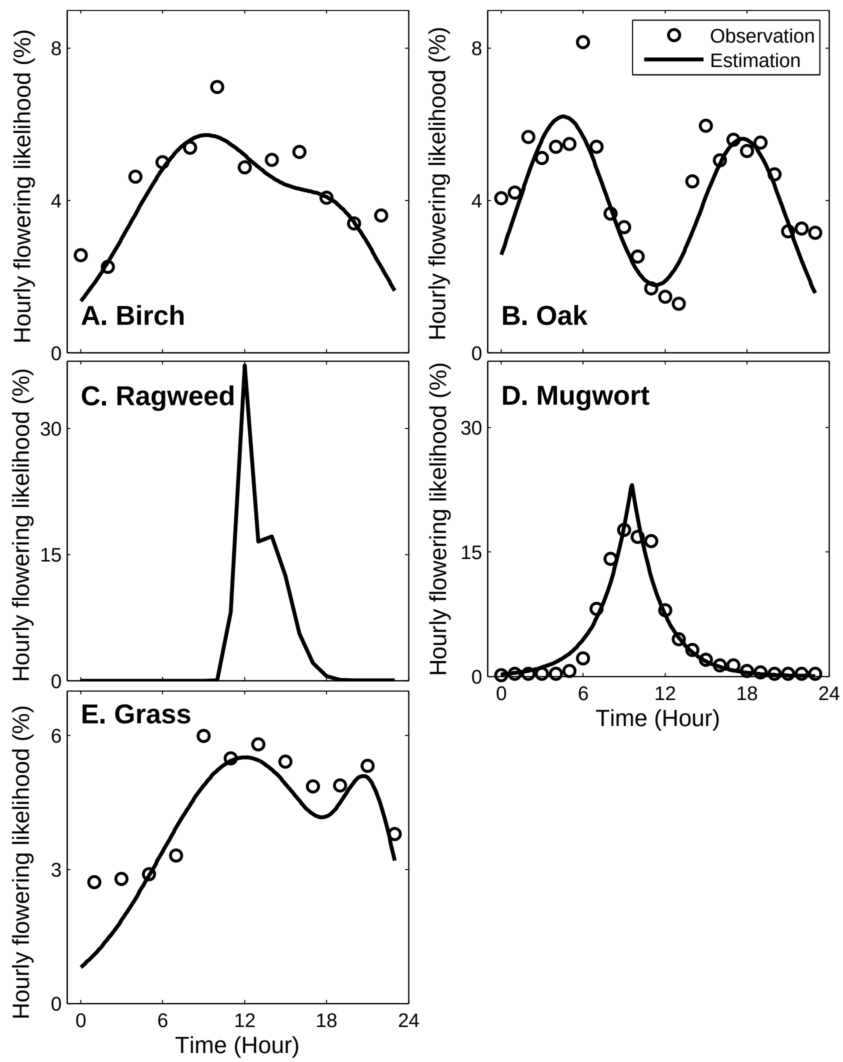

Figure 3.6: Hourly flowering likelihood estimated from models and those derived from observed hourly pollen counts. (A) Birch, (B) Oak, (C) Ragweed, (D) Mugwort, and (E) Grass.

3.3.3 Allergenic pollen season onset and duration

Threshold GDD across latitude

Figure 3.7 presents the average threshold GDD values for start and end dates across latitudes. These average threshold values were calculated based on observed start and end dates of allergenic pollen seasons during 1994-2010 at the studied monitoring stations across the CONUS. For the same species, the GDD threshold values for start and end dates of allergenic pollen season generally decrease as latitudes go higher in the US. The exception is the threshold GDD for start date of oak pollen season. It slightly increases at higher latitudes. Oak has more subspecies than other allergenic plants. The observed SD may represent pollen season timing from many subspecies. This may influence the general features of threshold GDD values for SD of oak pollen season.

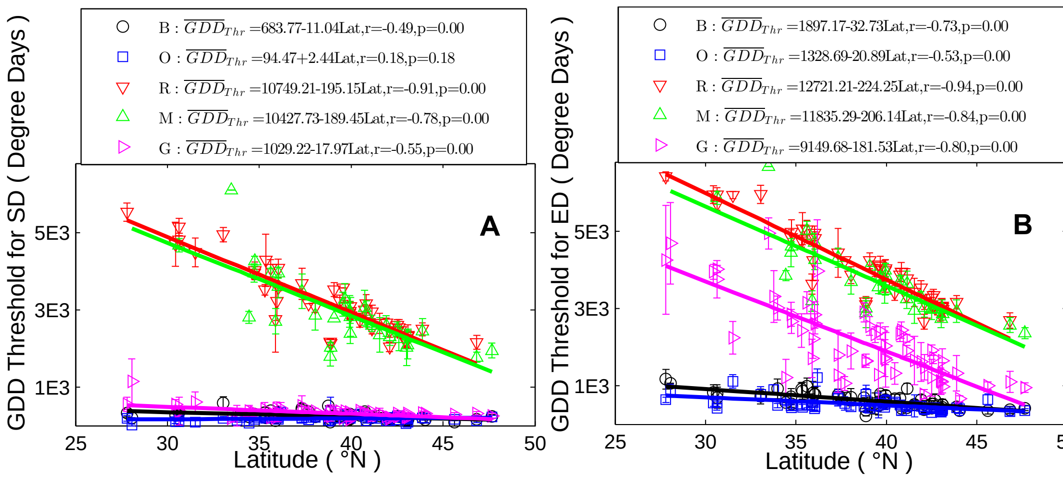

Figure 3.7: Mean threshold GDDs and their corresponding standard deviations for birch (B), oak (O), ragweed(R), mugwort (M) and grass (G) across the latitudes. The mean is defined as the average over the period of 1994-2010 at a station. Regression equations for the trends are presented in the legends. (A) GDD threshold for start date, and (B) GDD threshold for end date.

Although ragweed and mugwort start flowering from the cold north towards the warm south (Figure 2.8A), they flower earlier at a given location if the annual mean temperature is higher at that location (Figure C.1 in Appendix). This can be explained by their threshold GDD for SD. At higher latitude (cold north), the threshold GDD for SD ($\overline{GDD}_{Thr,SD}$), required for ragweed and mugwort to flower, is much lower than that at the low latitude (warm south, Figure 3.7A). At a given location (latitude), the higher the temperature, the earlier the $\overline{GDD}_{Thr,SD}$ can be reached, and therefore the earlier the ragweed or mugwort will bloom.

Model parameterization

Figure 3.8 presents the correlation coefficients between SD, ΔSD or SL and each individual climatic factor. For SD of birch and oak pollen, fixed-period GDD has the largest magnitude of correlation coefficient (-0.75 and -0.84 respectively; Figure 3.8 A1 and B1), and therefore can be used as a factor to account for influence of the current year temperature on SD. Monthly mean temperature from November of the previous year to March of the current year (T_{NDJFM}) also have relatively larger correlation coefficients, and therefore can be used as a factor to account for influence of the preceding year temperature on SD. Similarly, for SD of grass pollen (Figure 3.8 C1), the annual mean temperature (T_c) and monthly mean temperature from September to December of the previous year (T_{SOND}) can be used as factors to incorporate influence of temperature on SD.

Since ragweed and mugwort flower from the cold north towards the warm south, direct correlation analyses between SD and temperature mask the feature that higher temperature at the same location drives ragweed and mugwort to bloom earlier (Figure C.1). Instead, correlation analyses were conducted between the deviations in SD (ΔSD) and the deviations in climatic factors. The ΔSD of ragweed pollen is closely correlated with deviations in annual mean temperature (ΔT_C) and monthly mean temperature from November of the previous year to March of the current year (ΔT_{NDJFM}, Figure 3.8 D1). The ΔSD of mugwort pollen is closely correlated with deviations in FFD (ΔFFD) and annual accumulative precipitation (ΔPrc, Figure 3.8 E1).

Following similar analyses, relevant climatic factors can be prescreened to account for the influence of temperature, precipitation and other factors on SL (Figure 3.8 A2-E2). Also shown in Figure 3.8 is that SD or SL in the previous year exerts important influence on SD or SL in the current year. Thus SD_p or SL_p were also used as prescreened factors for further analyses. Table 3.4 lists the prescreened factors and optimum parameter values for fixed-period GDD.

The prescreened variables were further selected using stepwise regression and collinearity analyses. The final selected variables and regression equations for the observation-based model and its simplified version are also listed in Table 3.4. In addition, Table

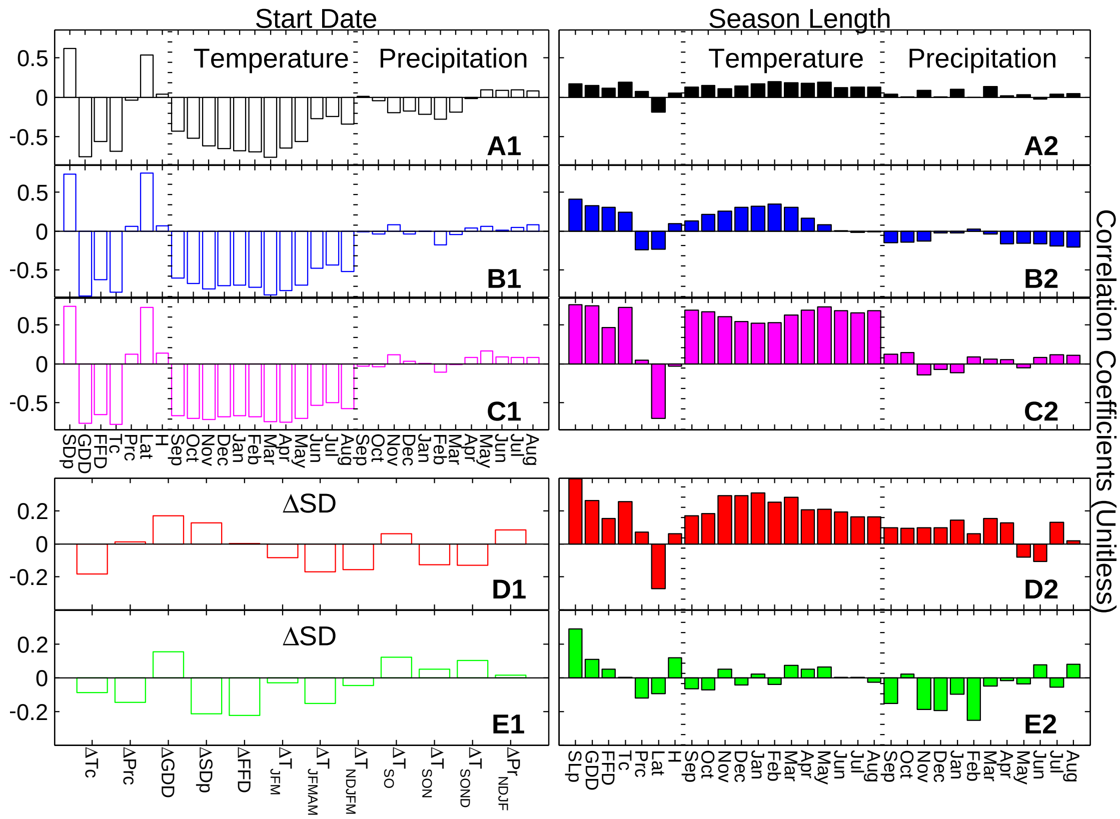

Figure 3.8: Correlation coefficients between Start Date (*SD*), Deviation in *SD* (ΔSD), or Season Length (*SL*) and an individual climatic factor. These factors include *SD* and *SL* in the previous year (SD_P, SL_P), Growing Degree Days (*GDD*), Frost Free Days (*FFD*), annual accumulative precipitation (Prc) and mean temperature (T_C) in the current year, latitude (*Lat*), height (*H*), and monthly accumulative precipitation and mean temperature from September in the preceding year to August in the current year. For start dates of ragweed and mugwort, correlation coefficient is based on correlation between ΔSD and the deviation in a climatic factor. Subscripts in the aggregated climatic factors indicate the consecutive months, in which the mean temperatures or total precipitation are calculated; e.g., T_{NDJFM} is the mean temperature in months from November in the preceding year to March of the current year. (A1-E1) correlation coefficients of *SD* or ΔSD for birch, oak, grass, ragweed and mugwort, respectively. (A2-E2) correlation coefficients of SL for birch, oak, grass, ragweed and mugwort, respectively.

Table 3.4: Summary of three models simulating the Start Date (*SD*) and Season Length (*SL*) of pollen season. Listed in the table are the prescreened variables, optimum parameters for Growing Degree Day (*GDD*), selected regression equation, the corresponding *P* value, adjusted coefficient of determination (R^2), and Root Mean Square Error (RMSE). All model and statistical parameters were calculated based on simulated and observed *SD* and *SL* for each year during 1994-2010 at all the studied pollen monitoring stations. Subscripts in the aggregated climatic factors indicate the consecutive months in which the mean temperatures or total precipitations are calculated; e.g., T_{NDJFM} is the mean temperature in months from November of the preceding year to March of the current year.

			Birch	Oak	Ragweed	Mugwort	Grass
Start Date	Variables		GDD, SD_p, T_{NDJFM}	GDD, SD_p, T_{NDJFM}	ΔT_c, ΔSD_p, ΔT_{NDJFM}	ΔFFD, ΔSD_p, ΔPr_c	T_c, SD_p, T_{SOND}
	ID, LD, T_b (ºC)		Feb. 1st, Apr. 15th, 3	Mar. 1st, Apr. 15th, 5	Feb. 1st, Feb. 28th, 0.5	Feb. 1st, Feb. 28th, 1.8	Feb. 1st, May 15th, 7.3
	M1	Equation	$SD=95.24-0.05GDD+0.20SD_p$, $P = 0.00$, $R^2 = 0.59$	$SD=106.35-0.09GDD+0.18SD_p$, $P = 0.00$, $R^2 = 0.73$	$\Delta SD=7.9\times10^{-3}-1.61\Delta Tc + 0.11\Delta SD_p$, $P = 0.00$, $R^2 = 0.03$	$\Delta SD=-0.25-0.19\Delta FFD-0.17\Delta SD_p$, $P = 0.00$, $R^2 = 0.08$	$SD=134.54-3.69T_c+0.31SD_p$, $P = 0.00$, $R^2 = 0.64$
		RMSE (days)	12.9	10.4	10.0	15.1	15.1
	M2	Equation	$SD=118.66-0.06GDD$, $P = 0.00$, $R^2 = 0.58$	$SD=128.42-0.10GDD$, $P = 0.00$, $R^2 = 0.72$	$\Delta SD=2.9\times10^{-3}-1.53\Delta Tc$, $P = 0.01$, $R^2 = 0.02$	$\Delta SD=-0.20-0.20\Delta FFD$, $P = 0.01$, $R^2 = 0.05$	$SD=193.33-5.23T_c$, $P = 0.00$, $R^2 = 0.60$
		RMSE (days)	13.3	10.7	10.0	15.2	16.0
	M3	Equation	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)
		RMSE (days)	26.1	29.4	12.3	5.3	33.8
Season Length	Variables		T_{JFMAM}, SL_p, T_{NDJFM}	T_{JFM}, SL_p, T_{NDJFM}, Pr_c	T_{JFM}, SL_p, T_{NDJFM}	GDD, SL_p, Pr_{NDJF}	GDD, SL_p, T_{SO}
	ID, LD, T_b (ºC)		Jan. 1st, Apr. 30th, 9.5	Feb. 1st, Mar. 31st, 6.3	Mar. 1st, Mar. 15th, 1.3	Feb. 1st, Apr. 15th, 9.5	Feb. 15th, Jun 30th, 10
	M1	Equation	$SL=24.13+0.47T_{JFMAM}+0.16SL_p$, $P = 0.00$, $R^2 = 0.05$	$SL=23.76+0.60T_{JFM}-0.01Pr_c+0.29SL_p$, $P = 0.00$, $R^2 = 0.23$	$SL=28.14+0.58T_{NDJFM}+0.27SL_p$, $P = 0.00$, $R^2 = 0.18$	$SL=38.42-0.02Pr_{NDJF}+0.21SL_p$, $P = 0.00$, $R^2 = 0.10$	$SL=19.64+0.05GDD+0.46SL_p$, $P = 0.00$, $R^2 = 0.64$
		RMSE (days)	14.2	11.9	9.9	16.9	29.8
	M2	Equation	$SL=28.55+0.60T_{JFMAM}$, $P = 0.00$, $R^2 = 0.03$	$SL=33.93+0.76T_{JFM}-0.01Pr_c$, $P = 0.00$, $R^2 = 0.15$	$SL=38.92+0.72T_{NDJFM}$, $P = 0.00$, $R^2 = 0.11$	$SL=48.48-0.03Pr_{NDJF}$, $P = 0.00$, $R^2 = 0.06$	$SL=37.46+0.10GDD$, $P = 0.00$, $R^2 = 0.54$
		RMSE (days)	14.4	12.5	10.4	17.3	33.8
	M3	Equation	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)
		RMSE (days)	15.5	15.0	29.9	33.4	5.7

3.4 lists the *GDD* model on the basis of threshold values obtained through Figure 3.7 A and B. Additional information regarding simulated annealing, FFD and annual mean temperature across latitudes are depicted in Figures C.3 and C.4.

On the basis of the Root Mean Square Error (RMSE), the observation-based model (M1) performed better than its simplified version (M2) and the GDD model (M3). The simplified observation-based model has the advantage of simulating the *SD* and *SL* without being dependent on continuously monitored aerobiological or phenological data. The *GDD* model performed best at predicting the *SL* for grass pollen, and *SD* for mugwort pollen. However, the *GDD* threshold values for start and end dates (Figure 3.7 A and B) may face difficulties for application at latitudes outside the ranges of observations. Particularly, the $\overline{GDD}_{Thr,ED}$ for grass pollen decreases quickly as latitude goes higher; it is equal to the $\overline{GDD}_{Thr,SD}$ at a latitude of around 49.65 °N. This makes the *GDD* model unfeasible for latitudes higher than 49.65 °N.

Accumulation of chilling force, masting behavior, photoperiod and soil humidity also affect the *SD* and *SL* of allergenic seasons[26,73,15]. In the current study, the initial date for accumulating daily temperature for SD is February 1st for birch, ragweed, mugwort and grass, and March 1st for oak. It is reasonable to assume that the chilling requirement was met before these initial dates in the CONUS. Observed data and practical models of masting behavior, photoperiod and soil humidity are rarely available, and extremely limited for large geographic area like the US. These factors should be taken into consideration in future studies.

Model evaluation

On the basis of the minimum RMSE for *SD* or *SL* for each species in Table 3.4, optimum simulation results were used to calculate the deviation in simulated mean *SD* or *SL* from the corresponding mean observations during 1994-2010 at each studied station (i.e., $SD_{Dev} = \overline{SD}_{Obs} - \overline{SD}_{Sim}$). The results are displayed in Figures 3.9 and 3.10 according to the locations of the studied NAB-AAAAI monitoring stations. To indicate the pollen abundance of a given allergenic species at a station, the size of each circle in Figures 3.9 and 3.10 were plotted proportionally to the average seasonal airborne pollen counts during 1994-2010 at that station. The average of seasonal counts and maximum daily counts are also listed in Table B.3 in the Appendix.

The deviations of simulated mean SDs are within 0-6 days at 63.6%, 72.4%, 68.8%, 94.4% and 56.4% of the studied stations for birch, oak, ragweed, mugwort and grass, respectively. The deviations of simulated mean SLs are within 0-6 days at 60.0%, 69.0%, 66.7%, 25.0% and 74.6% of the studied stations for birch, oak, ragweed, mugwort and grass, respectively. In general, the models could simulate the middle range SD and SL reasonably well, but face difficulty in reproducing the extreme large or small observed SD and SL.

The RMSEs derived from cross-validation of the model are comparable with those listed in Table 3.4, which are based on models trained using the whole dataset from each of the stations during 1994-2010. This indicates that the proposed models are less likely to overfit the data. As an example, Figure C.2 presents a comparison between the observed and predicated SD and SL through the leave-one-out cross validation procedure using the M1 method for all pollen stations during 1994-2010.

Model application

Figures 3.11 and 3.12 demonstrates the spatially resolved (50x50 km) maps of allergenic pollen season onset and duration for 2004 in the CONUS. For birch, oak and grass, the allergenic pollen seasons in 2004 proceeded from the south toward the north. For ragweed and mugwort, the pollen seasons progressed from the north to the south. The durations of allergenic pollen season in 2004 were longer in the south than those in the north. These results are consistent with the long term observations in Figure 2.8A and B.

The spatially resolved maps of allergenic pollen season onset and duration can be used to drive the operational pollen forecasting model at multiple spatiotemporal scales, and to study the production, release, distribution and health effects of allergenic pollen in the US. Based on the availability of airborne pollen data, the methods presented in the current study can potentially be adapted to other countries, regions and species for studying biogenic aeroallergens in large geographic areas.

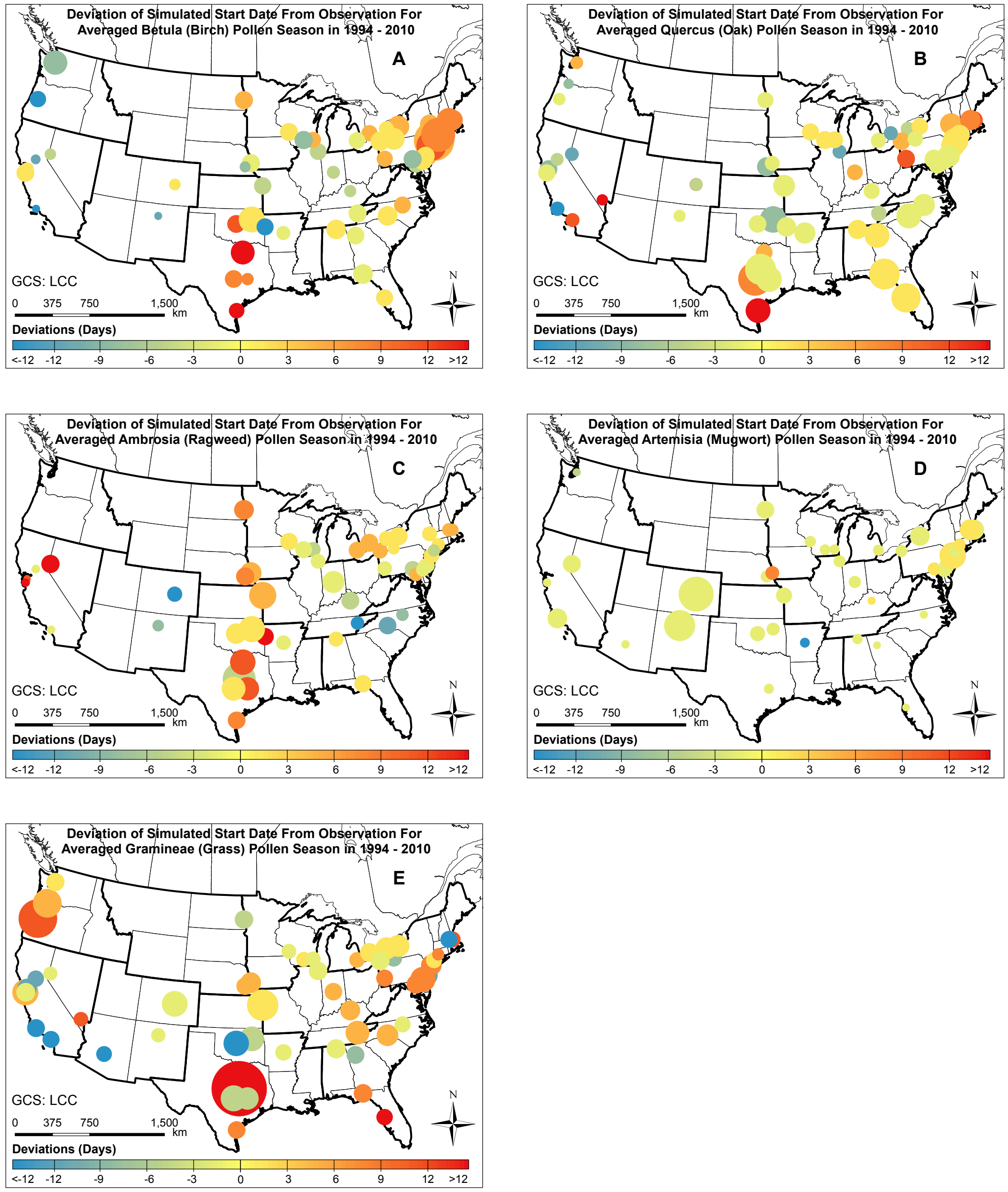

Figure 3.9: Deviation of simulated mean Start Date (SD) from the mean observations in 1994-2010 at each NAB-AAAAI station, i.e., $SD_{Dev} = \overline{SD}_{Obs} - \overline{SD}_{Sim}$. (A) Birch, (B) oak, (C) Ragweed, (D) Mugwort, and (E) Grass. The size of the circle indicates the average pollen abundance for a given species during the same period.

Although there are some spatial correlations of SD and SL among different pollen stations, these spatially resolved maps of allergenic pollen season onset and duration are generally hard to derive from geostatistics (e.g. kriging). On one hand, geostatistics

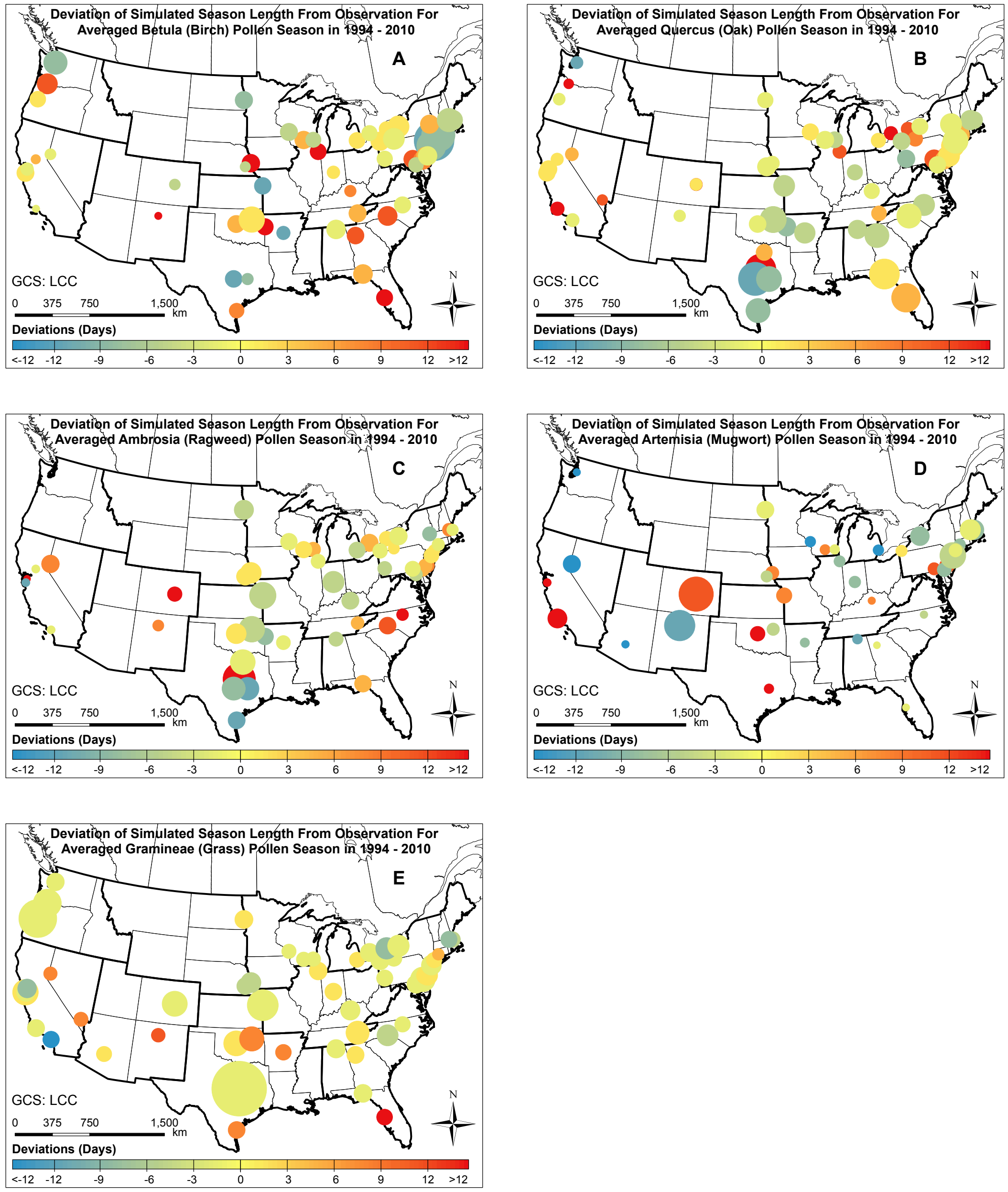

Figure 3.10: Deviation of simulated mean Season Length (SL) from the mean observations in 1994-2010 at each NAB-AAAAI station, i.e., $SL_{Dev} = \overline{SL}_{Obs} - \overline{SL}_{Sim}$. (A) Birch, (B) Oak, (C) Ragweed, (D) Mugwort, and (E) Grass. The size of the circle indicates the average pollen abundance for a given species during the same period.

generally need sufficient spatial sample points to generate reasonable interpolation values at unknown locations. Specifically, it is hardly convincing to interpolate spatially resolved maps of SD and SL using airborne pollen data from 58 stations across the US.

On the other hand, geostatistics generally generate stationary interpolations of a given spatial phenomenon. For the current study, it is challenging to use geostatistics to extrapolate allergenic pollen season onset and duration for years other than the study period of 1994-2010.

3.3.4 Spatiotemporal pattern of allergenic pollen emission

Temporal pattern of pollen emission

Figure 3.13 presents the representative time slices of oak pollen emission for 2004. Figures 3.13A and B show spatial distribution of emissions at morning and afternoon time in an early period (March) of oak pollen season in the CONUS; and Figures 3.13C and D show emission distribution at morning and afternoon time in a late period (April) of pollen season.

As illustrated in Figure 3.13, oak pollen emissions started from Southeastern CONUS in March and then shifted gradually toward Northern CONUS in April. Oak pollen emissions in the Southeast, South and West climate regions are higher than those in other regions in the CONUS. This is consistent with the distributions and density of oak trees as shown in Figure 3.5B.

Figure 3.14 presents the representative time slices of ragweed pollen emission for 2004. Figures 3.14A and B show spatial distribution of emissions at morning and afternoon time in an early period (August) of ragweed pollen season in the CONUS; and Figures 3.14C and D show emission distribution at morning and afternoon time in a late period (September) of pollen season.

As illustrated in Figure 3.14, ragweed pollen emissions started from Northern CONUS in August and then shifted gradually toward Southern CONUS in September. Ragweed pollen emissions in the Southwest and West North Central climate regions are higher than those in other regions in the CONUS. This is consistent with the distributions and density of ragweed vegetation as shown in Figure 3.5C.

Grass and birch pollen emissions follow temporal patterns similar to that of oak; mugwort pollen emissions follows temporal patterns similar to that of ragweed.

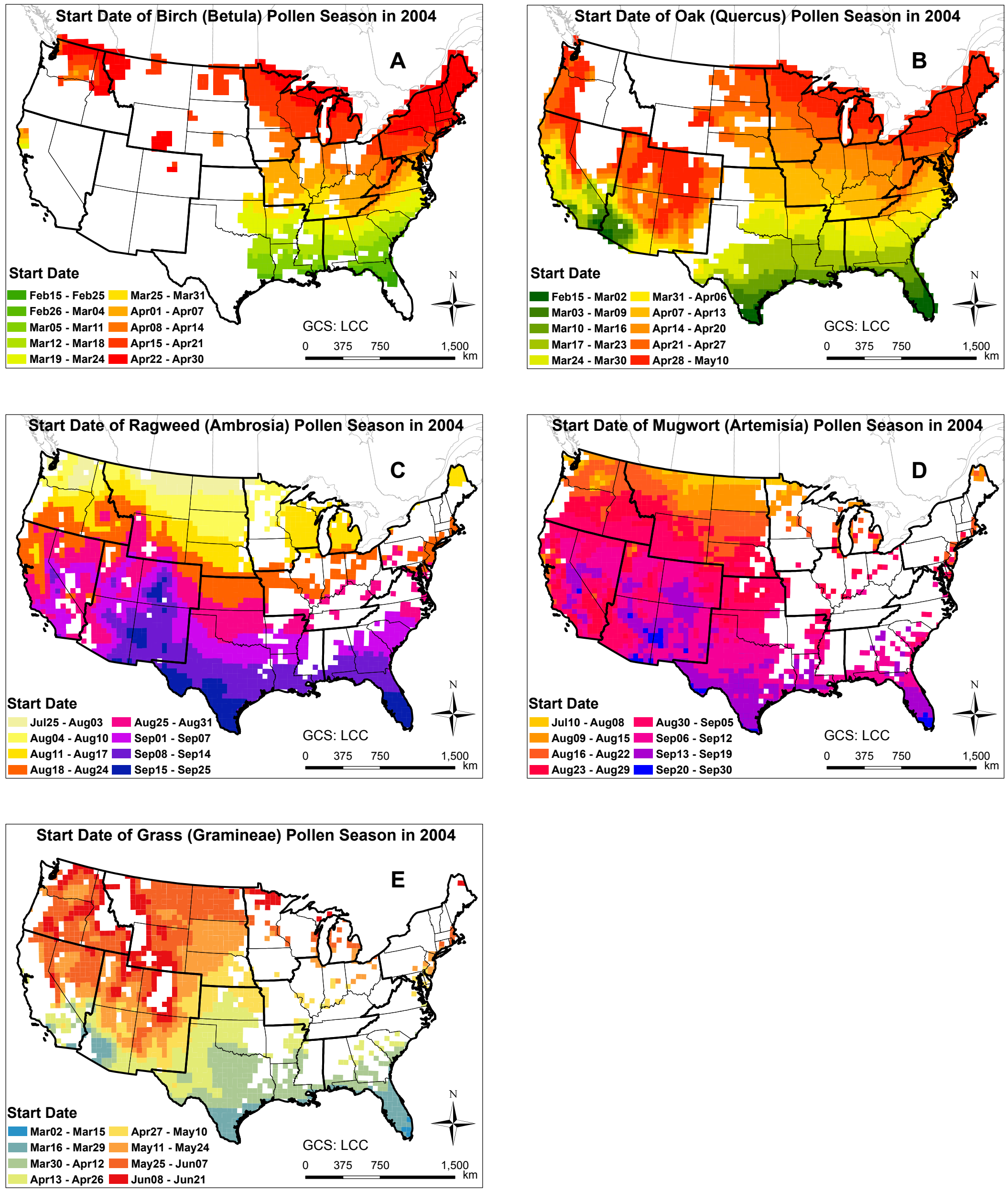

Figure 3.11: Simulated Start Date (SD) of allergenic pollen season using method M1 (Table 1) based on NARCCAP archived meteorology simulation data in 2004. Data were mapped only on cells in which the area coverage of a given allergenic species is greater than zero. (A) Birch, (B) Oak, (C) Ragweed, (D) Mugwort, and (E) Grass.

Spatial patterns of pollen emissions

Figure 3.15 depicts the spatial patterns of oak pollen emissions during the pollen season in 2004. The spatial patterns were examined for four metrics: mean, maximum, seasonal

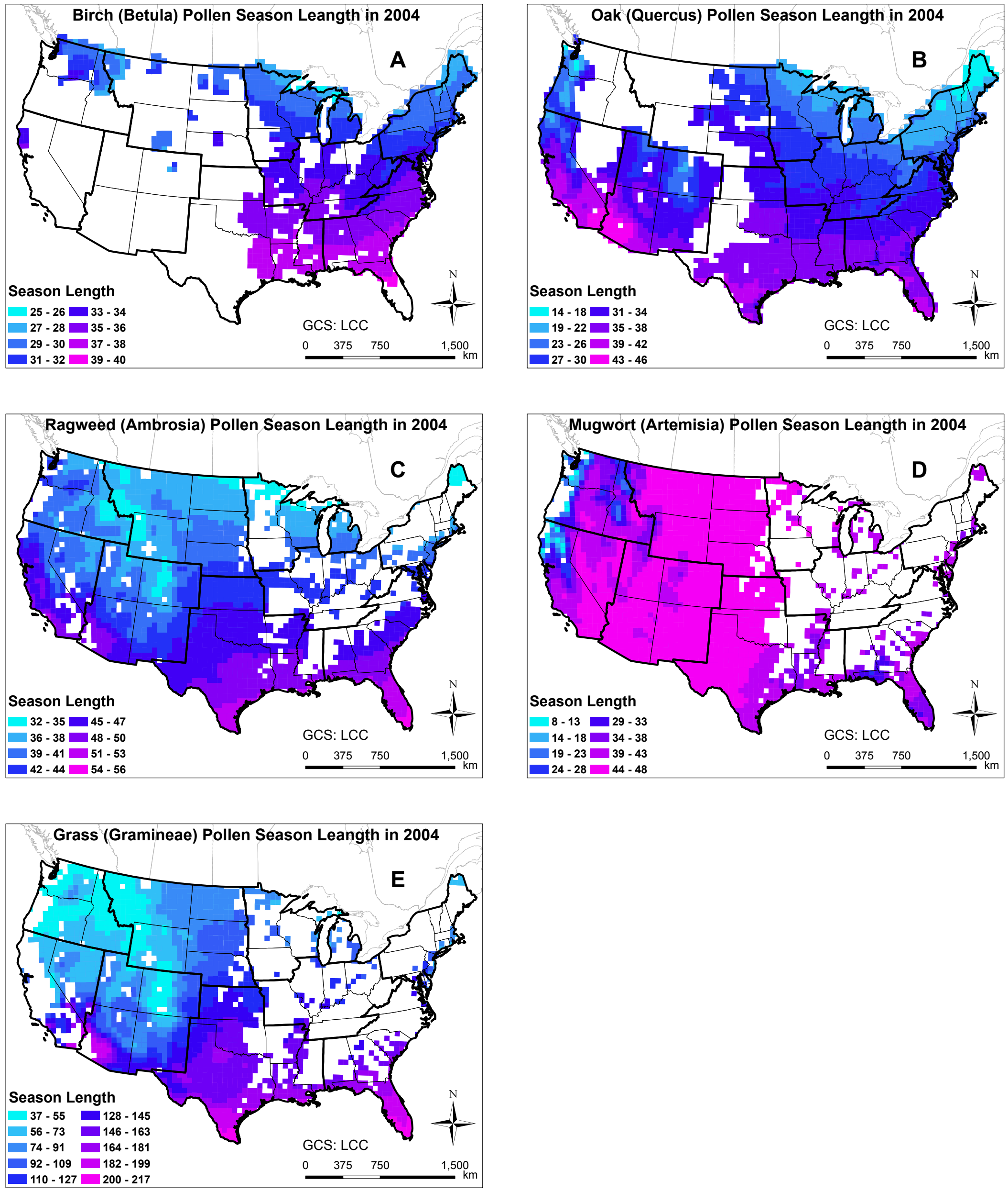

Figure 3.12: Simulated Season Length (SL) of allergenic pollen season using method M1 (Table 1) based on NARCCAP archived meteorology simulation data in 2004. Data were mapped only on cells in which the area coverage of a given allergenic species is greater than zero. (A) Birch, (B) Oak, (C) Ragweed, (D) Mugwort, and (E) Grass.

total, and standard deviation of hourly emissions at each 50x50 km grid covering the CONUS. These four metrics were calculated using Equation 3.26 for each 50x50 km grid based on the simulated hourly emissions of oak pollen between March 1st, 2004

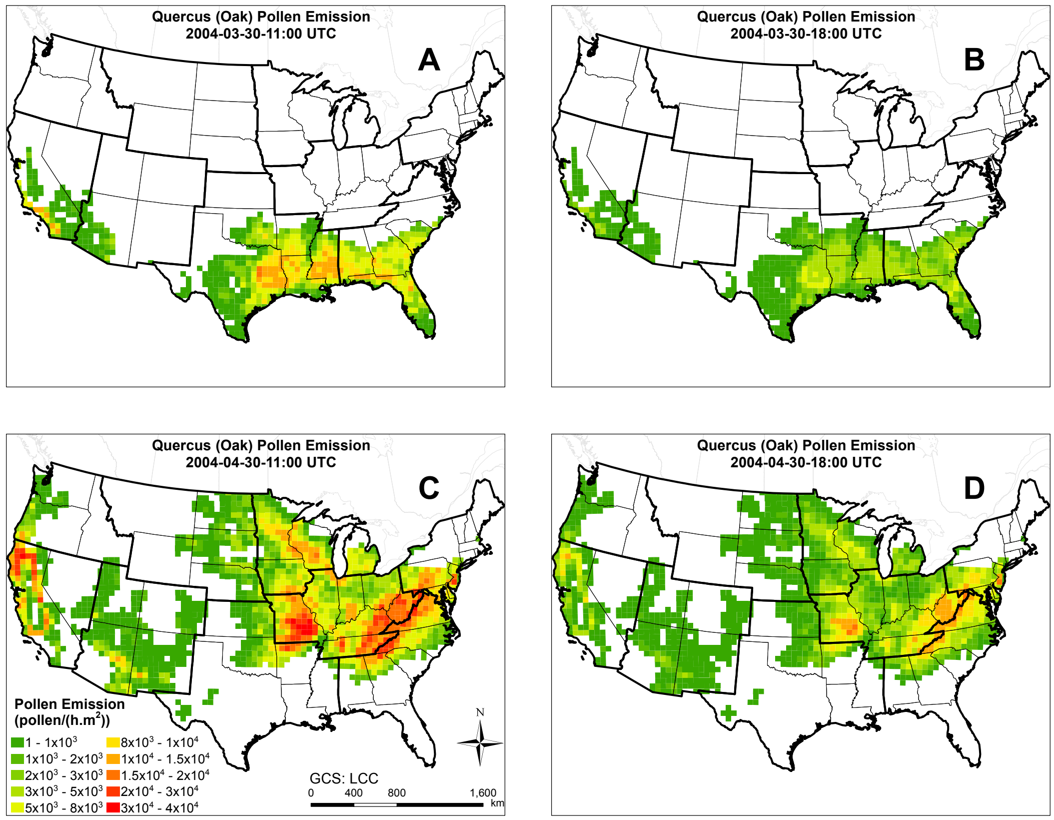

Figure 3.13: Time slices of spatiotemporal emission profiles of oak pollen for 2004. (A) Mar. 30th 11:00 UTC, (B) Mar. 30th 18:00 UTC, (C) Apr. 30th 11:00 UTC, and (D) Apr. 30th 18:00 UTC.

and April 30th, 2004.

The oak pollen emissions varied remarkably in different regions. The mean hourly emission flux can range from 1-200 pollen grain/(m^{2}h) in the West North Central region to 4001-6500 in the Southeast region. The spatial pattern of mean, maximum, seasonal and standard deviation of hourly emission flux all roughly follow the pattern of area coverage of oak trees as shown in Figure 3.5B.

Figure 3.16 depicts the spatial pattern of ragweed pollen emission during the pollen season in 2004. These four emission metrics were calculated for each 50x50 km grid using Equation 3.26 based on the simulated hourly emission of ragweed pollen between August 1st, 2004 and September 30th, 2004.

The ragweed pollen emission varied remarkably in different regions. The mean hourly emission flux can range from 1-6x10^4 pollen grains/(m^{2}h) in the Northeast region

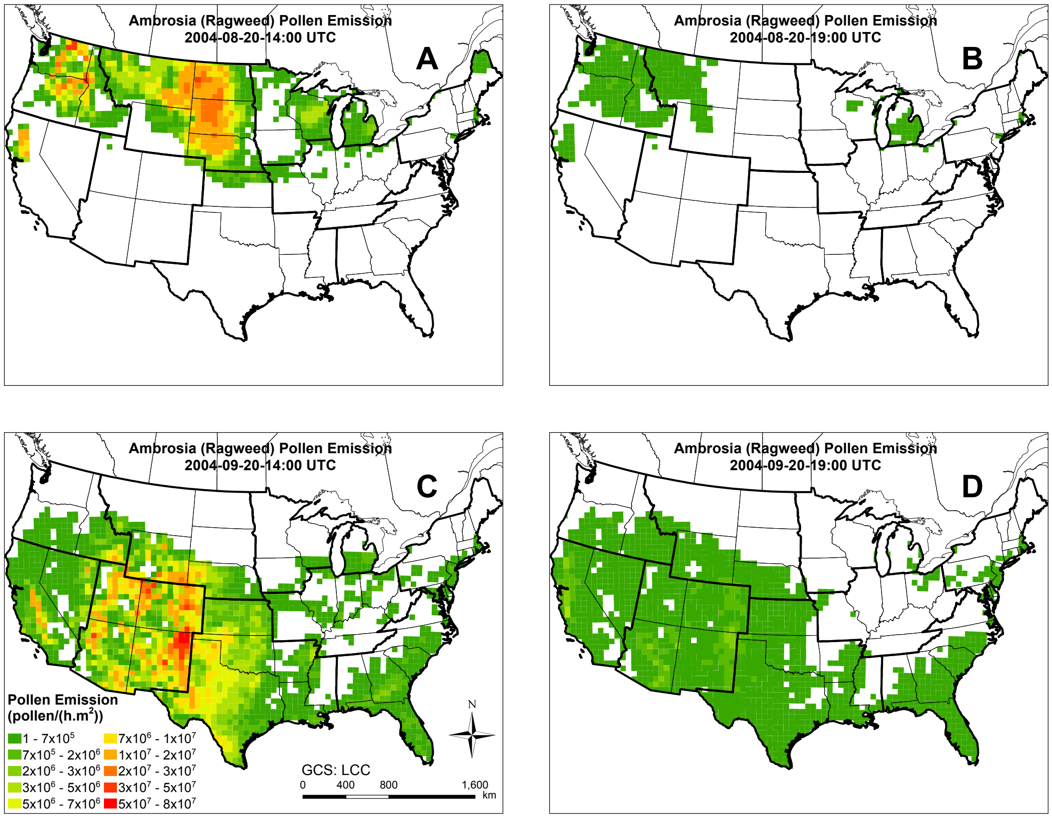

Figure 3.14: Time slices of spatiotemporal emission profiles of ragweed pollen for 2004. (A) Aug. 20th 14:00 UTC, (B) Aug. 20th 19:00 UTC, (C) Sep. 20th 14:00 UTC, and (D) Sep. 20th 19:00 UTC.

to $2\text{x}10^6$-$3\text{x}10^6$ in the West North Central and South regions. The spatial pattern of mean, maximum, seasonal and standard deviation of hourly emission flux all roughly follow the pattern of area coverage of ragweed plants as shown in Figure 3.5C.

Figures C.5, C.6 and C.7 in the Appendix display the mean, maximum, seasonal total and standard deviation of simulated hourly pollen emission for birch, mugwort and grass, respectively. Similar to Figures 3.15 and 3.16, the pollen emission patterns of birch, mugwort and grass also roughly follow their area coverage in the CONUS. The slight difference in the patterns between pollen emission and the area coverage are mainly caused by the meteorology factors in different regions.

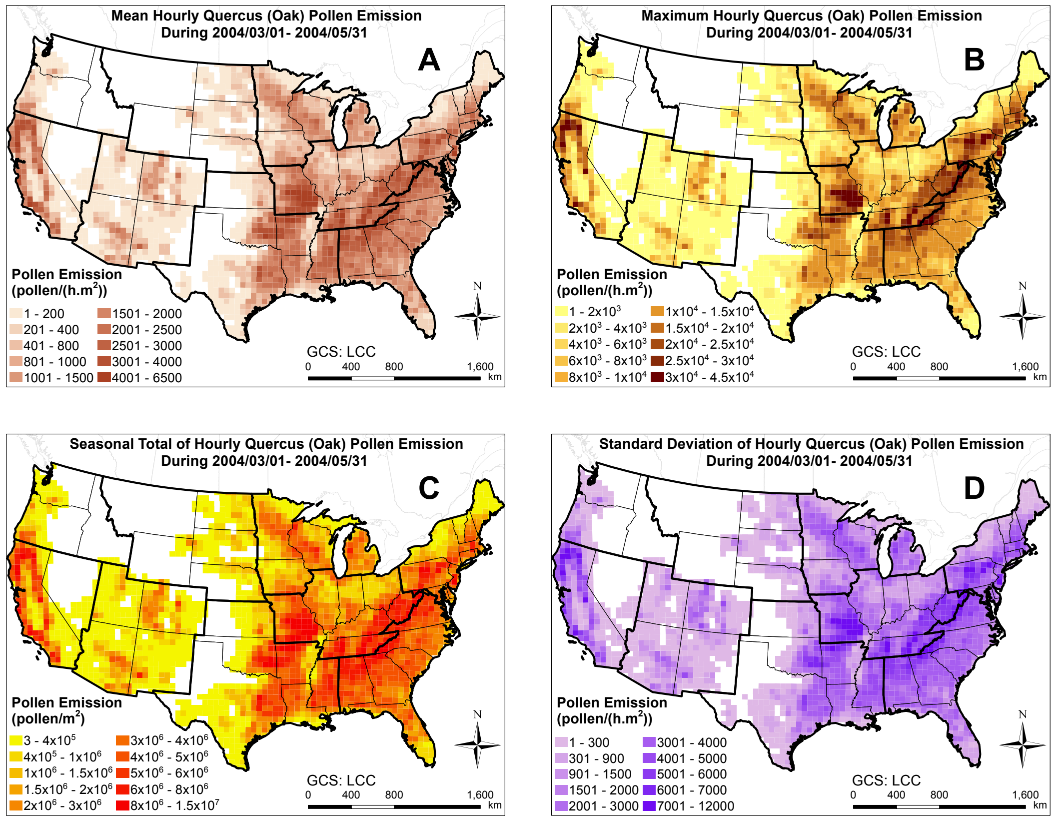

Figure 3.15: Spatial pattern of mean, maximum, seasonal total and standard deviation of hourly emission of oak pollen. (A) Hourly mean, (B) Hourly maximum, (C) Seasonal total, and (D) Standard deviation.

3.3.5 Sensitivity analysis

The global sensitivity coefficients for each input parameter in Table 3.1 are very close for four regional emission metrics calculated using Equation 3.27. Figure C.8 in the Appendix presents the global sensitivity coefficients for each input parameter for four regional emission metrics of oak pollen emissions. The regional hourly mean and maximum, and seasonal mean and maximum pollen emissions responded almost the same to the perturbations in the parameter space. The regional hourly mean emission flux was used as a metric for further discussion on sensitivity analysis.

The global sensitivity of the simulated regional mean hourly pollen emissions to different parameters is presented in Figures 3.17 and 3.18. The global NSC of all parameters varied between -0.08 and 0.08 for pollen emissions from oak, ragweed, mugwort and grass, indicating the robustness of the emission model to these four taxa. Birch

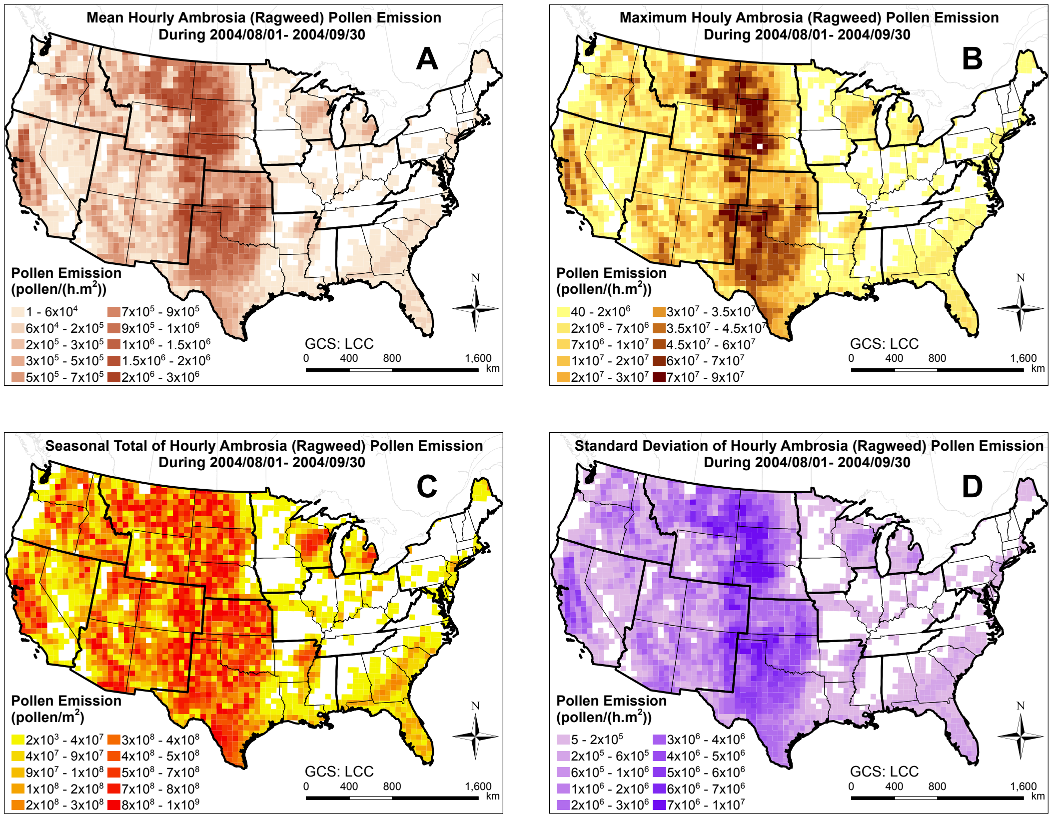

Figure 3.16: Spatial pattern of mean, maximum, seasonal total and standard deviation of hourly emissions of ragweed pollen. (A) Hourly mean, (B) Hourly maximum, (C) Seasonal total, and (D) Standard deviation.

pollen emissions were more sensitive to parameter perturbations (Figure 3.17E). The average absolute global NSC, $\overline{|NSCg|}$, were 0.176 and 0.039 for regional hourly mean and maximum birch pollen emissions, respectively. The sensitive parameters for birch pollen emission included: the threshold temperature for direct emission (T_{te}), the density of birch pollen grain (ρ_p), the daily flowering likelihood (L_d), the threshold wind speed for direct emission (V_{te}), the settling velocity (v_s), the quasi-laminar resistance (r_b), and the hourly flowering likelihood (L_h).

The standard deviations of NSCs for pollen emissions of oak, ragweed, mugwort and grass were between 0.454 and 0.650. This indicated low interaction and nonlinearity effects among parameters for pollen emissions of oak, ragweed, mugwort and grass. High interaction and nonlinearity effects among parameters were found for birch pollen emission. The average interaction effects (i.e.,$\overline{STD}$) were 4.183 and 1.059 for regional hourly mean and maximum birch pollen emissions, respectively. Parameters with high interaction and nonlinearity effects included: the threshold temperature for direct emission (T_{te}), the density of birch pollen grain (ρ_p), the daily flowering likelihood (L_d), the quasi-laminar resistance (r_b), the threshold wind speed for direct emission (V_{te}), the aerodynamic resistance (r_a), and the hourly flowering likelihood (L_h).

The emission model's sensitive and interactive response to perturbations in input parameters for birch pollen could be potentially related to the locations of birch trees and properties of birch pollen grains. As shown in Figure 3.5, birch trees mainly grow in the Northeast and East North Central climate regions. Birch pollen season in these two regions is very short compared with those in other regions because of cold weather through the later winter and earlier spring (Figure 2.8B). Perturbations in meteorology factors, such as threshold temperature and wind speed for direct emission, in the short flowering season in these two regions are thus important, and could possibly dominate the birch pollen emissions. This is consistent with sensitive response of birch pollen season to temperature in these regions observed in the past two decades in the CONUS (Figure 2.9)[45]. The density of birch pollen grains is 800 kg/m^3, which is smaller than water density[54]. Perturbations in the pollen density could easily lead to nonlinear changes in deposition velocity (Equations 3.23, 3.24 and 3.25) and threshold friction velocity (Equation 3.18), and therefore cause nonlinear perturbations in pollen emission flux.

Uncertainties in sensitive and interactive input parameters result in large deviations of model predictions. Further discussion regarding the impact of uncertainties from different sources on distribution of allergenic airborne pollen are presented in Chapter 4.

3.4 Summary

A mechanistic pollen emission model was developed based on mass balance of pollen grain fluxes in the near surroundings of allergenic plants. The emission model consists of direct emission and resuspension, and accounts for influences of temperature, wind

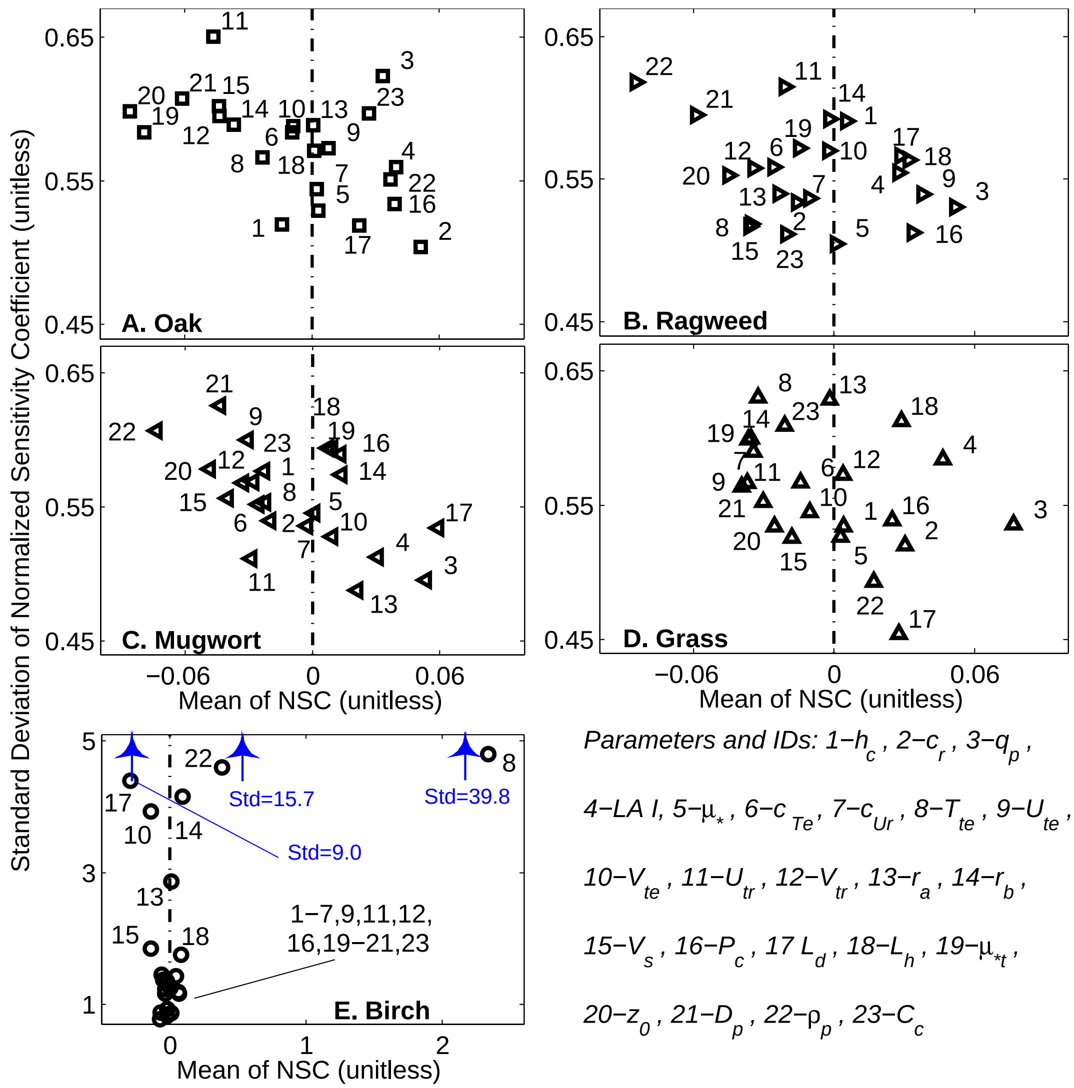

Figure 3.17: Mean and standard deviation of Normalized Sensitivity Coefficient (NSC) for each parameter for the pollen emission model for five allergenic taxa. The vertical dashed lines represent the NSC values of 0. All parameters are described in table 3.1. (A) Oak, (B) Ragweed, (C) Mugwort, (D) Grass, and (E) Birch.

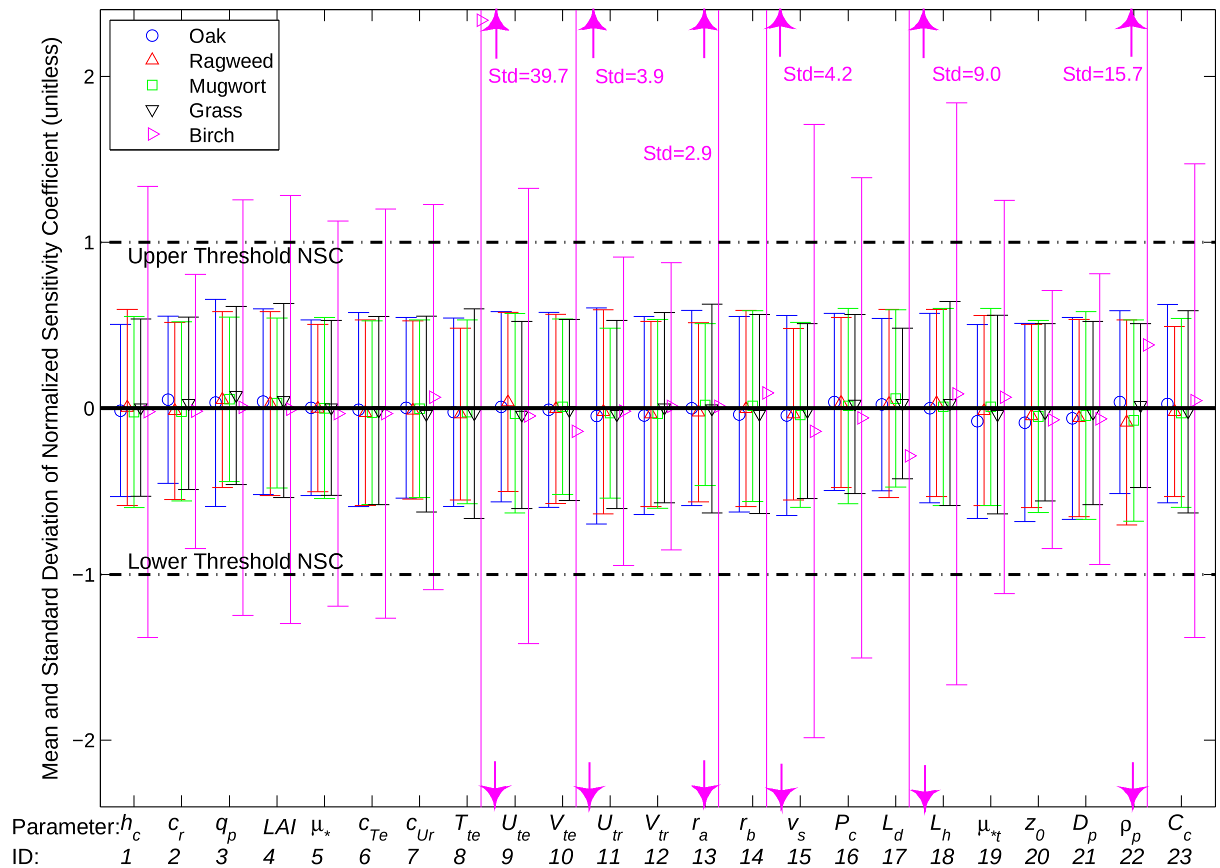

Figure 3.18: Mean and standard deviation of Normalized Sensitivity Coefficient (NSC) for each parameter for the pollen emission model for five allergenic taxa based on global sensitivity analyses using Morris' design. Horizontal dashed lines (NSC=±1) represent threshold NSCs. All parameters are described in table 3.1.

velocity and relative humidity. Modules of this emission model have been developed and parameterized to provide pollen season onset and duration, hourly flowering likelihood, and vegetation coverage for birch, oak, ragweed, mugwort and grass.

The estimated mean start date and season length for birch, oak, ragweed, mugwort and grass pollen season in 1994-2010 are mostly within 0 to 6 days of the corresponding observations for the majority of the NAB-AAAAI monitoring stations across the contiguous US. The simulated spatially resolved maps for onset and duration of allergenic pollen season in the contiguous US are consistent with the long term observations. The spatiotemporal pattern of pollen emissions generally follows the corresponding pattern of area coverage of allergenic plants and diurnal pattern of hourly flowering likelihood.

The emission model is robust with respect to the pollen emission of oak, ragweed, mugwort and grass; but highly sensitive and interactive to perturbations in input parameters for birch pollen emissions. The sensitive and interactive parameters included the threshold temperature and wind speed for direct emissions, the density of pollen grains, the aerodynamic resistance and quasi-laminar resistance, and the flowering likelihood.

Chapter 4

DISTRIBUTION OF AIRBORNE ALLERGENS UNDER CHANGING CLIMATE

4.1 Introduction

In this chapter, the pollen emission model developed in Chapter 3 was used to drive the adapted CMAQ model to simulate the spatiotemporal distributions of airborne pollen concentrations. The simulation results were evaluated using the observed pollen count from the NAB-AAAAI pollen stations across the contiguous United States. Process analyses were conducted to investigate the contribution of each physical process on airborne pollen concentration. The validated model was then run to simulate the spatiotemporal distributions of allergenic pollen during the periods of 2001-2004 and 2047-2050. These simulation results were analyzed to elucidate the climate change impacts on allergenic pollen season timing and airborne levels in nine climate regions across the CONUS (Figure 2.2).

4.2 Methods

4.2.1 Model configuration

The configurations of each component model are listed in Table 4.1. The WRF-SMOKE-CMAQ-Pollen modeling system was first run for 2004 covering the CONUS with a spatial resolution of 50x50 km and temporal resolution of one hour. The simulation time for birch and oak was between March 1st, 2004 and April 30th, 2004; for ragweed and mugwort between August 1st and September 30th; and for grass between March 1st and June 30th. The simulation results for 2004 were evaluated using the observed pollen counts from NAB-AAAAI monitoring stations across the CONUS. The evaluated WRF-SMOKE-CMAQ-Pollen modeling system was then applied to simulate the spatiotemporal distributions of allergenic pollen during the periods of 2001-2004 and 2047-2050 for studying climate change impact on allergenic pollen season.

The WRF-SMOKE-CMAQ-Pollen modeling system was also configured to simultaneously simulate the distributions of anthropogenic allergens (e.g., ozone and particulate matter) and allergenic pollen for 2007, covering the northeastern CONUS with spatial resolution of 12x12 km and temporal resolution of one hour. The simulation time for oak is between March 1st and May 30th, 2007; and for ragweed between July 1st and September 30th, 2007. The chemistry module in CMAQ was turned off for simulation of pollen concentrations in the CONUS domain, but turned on for simultaneous simulation of concentrations of pollen and anthropogenic pollutants.

Table 4.1: Configuration of the meteorology, emission and transport model for studying distributions of airborne allergens under changing climate.

	Model	Resolution, Layers	Period	Domain	Configuration Reference
Meteorology	WRF2.1.1	50x50 km, hourly, 34	2001-2004, 2047-2050	CONUS	[55,56]
	WRF3.1.1	12x12 km, hourly, 34	2007	Northeast US	[57]
Pollen emission	Current Study	50x50 km, hourly, 1	2001-2004, 2047-2050	CONUS	Chapter 3
	Current Study	12x12 km, hourly, 1	2007	Northeast US	Chapter 3
Anthropogenic emission	SMOKE3.0	12x12 km, hourly, 1	2007	Northeast US	[58]
Pollen transport	Adapted CMAQ4.7.1	50x50 km, hourly, 34	2001-2004, 2047-2050	CONUS	Chapter 4
	Adapted CMAQ4.7.1	12x12 km, hourly, 34	2007	Northeast US	Chapter 4

4.2.2 Physical processes governing transport of airborne pollen

Physical processes

The pollen transport model CMAQ-Pollen was adapted from the existing CMAQ modeling system[131]. Pollen grains were treated as coarse mode aerosol. Physical properties such as density, diameter and diameter distributions and other related information (e.g. allowed maximum aerosol diameter) of coarse mode were adapted in relevant CMAQ modules such as aero_depv.F, AERO_EMIS.F, AERO_INFO.f and aero_subs.f *etc.*, so that the adapted CMAQ-Pollen model could handle the simulation of spatial and temporal distributions of pollen.

Figure 4.1 presents different modules of the transport model and their connections. The physical processes governing the transport and removal of pollen grains from air include cloud process, dry deposition, horizontal and vertical advection, and horizontal and vertical diffusion. Dry deposition process is treated in vertical diffusion process as a flux boundary condition at the bottom of the model layer, while wet deposition is simulated in cloud process, and depends on the precipitation rate and pollen concentration in cloud water. Effects of convection on pollen transport are treated separately through modules of horizontal and vertical advection.

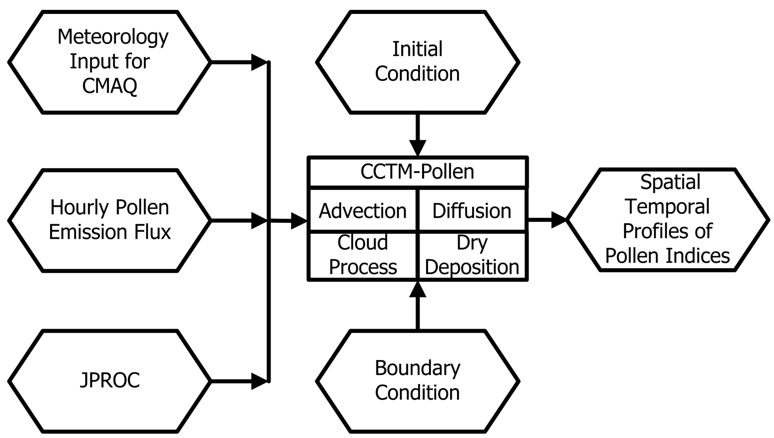

Figure 4.1: Schematic presentation of pollen transport model. The model was adapted from CMAQ4.7.1 and incorporated the physical processes of advection, diffusion, cloud process and dry deposition.

Governing equations

Appendix D.1 lists the governing equations for fully compressible atmosphere in a generalized meteorological curvilinear coordinates $(\hat{x}^1, \hat{x}^2, \hat{x}^3, \hat{t})$ in a conformal map projection. Neglecting the molecular diffusion, conservation equation of Reynolds averaged concentration of pollen grains was formulated using the decomposed velocity (D.7d) and concentration components (Equation D.7b) through equation 4.1,

$$\begin{aligned}
&\underbrace{\frac{\partial \varphi_p^*}{\partial t}}_{(a)} + \overbrace{\boldsymbol{\nabla}_s \cdot \left[\varphi_p^* \overline{\hat{\boldsymbol{V}}}_s\right]}^{(b)} + \underbrace{\frac{\partial(\varphi_p^* \overline{\hat{v}^3})}{\partial \hat{x}^3}}_{(c)} + \overbrace{\boldsymbol{\nabla}_s \cdot \left[\overline{\rho}\sqrt{\hat{\gamma}}\hat{\boldsymbol{F}}_{q_p}\right]}^{(d)} + \underbrace{\frac{\partial(\overline{\rho}\sqrt{\hat{\gamma}}\hat{\boldsymbol{F}}^3_{q_p})}{\partial \hat{x}^3}}_{(e)} \\
&\quad = \overbrace{\sqrt{\hat{\gamma}} S_{\varphi_p}}^{(f)} + \underbrace{\left.\frac{\partial \varphi_p^*}{\partial t}\right|_{cld}}_{(g)}
\end{aligned} \tag{4.1}$$

where $\varphi_p^* = \sqrt{\hat{\gamma}}\varphi_p = (J_s/m^2)\varphi_p$ is the scaled averaged pollen concentration, and $\hat{\boldsymbol{F}}_{q_p} = \hat{\boldsymbol{i}}\hat{F}^1_{q_p} + \hat{\boldsymbol{j}}\hat{F}^2_{q_p} = \hat{\boldsymbol{i}}\overline{q_p'\hat{v}^{1\prime}} + \hat{\boldsymbol{j}}\overline{q_p'\hat{v}^{2\prime}}$, $\hat{F}^3_{q_p} = \overline{q_p'\hat{v}^{3\prime}}$ are the corresponding Reynolds turbulent flux terms. Terms in equation 4.1 are explicitly related to the modules corresponding to the science processes in CMAQ[131]: (a) time rate of change of pollen concentration, (b) horizontal advection, (c) vertical advection, (d) horizontal eddy diffusion, (e) vertical eddy diffusion, (f) emission, and (g) cloud process. Formulations of each of the above physical process are presented in Appendix D.1.

For regional meteorology and air quality model, modeling errors are often minimized by adding one artificial term in deterministic equations and nudging the modeling system toward observed data[132]. In terms of CMAQ algorithms, error growth in each process and from one process to another were tracked and controlled by adding an artificial errors term in conservative equations, such as Q_ρ in equations D.2 and D.8. At each step of modeling time, governing equations with additional diagnostic equations were solved together to guarantee mass consistence in the modeling system[131].

4.2.3 Initial and boundary conditions

For birch, oak and grass, the simulation for the CONUS domain was run from 00:00 of March 1st through 23:00 of April 30th. For ragweed and mugwort, the simulation was run from 00:00 of August 1st through to 23:00 of September 30th. March 1st generally precedes the earliest flowering day of birch, oak and grass in the CONUS; and August 1st generally precedes the earliest flowering day of ragweed and mugwort in the CONUS. Therefore, the Initial Conditions (IC) for first simulation hour can be set to zero. ICs of subsequent hours were obtained from the simulation results of the corresponding previous hours.

For simulation for the CONUS domain, Boundary Conditions (BC) were also set to zero with eastern and western boundaries of simulated domain bordering the Atlantic and Pacific oceans, and northern and southern boundaries adjoining Canada and Mexico. To investigate the influence of BC on airborne pollen concentrations, one additional simulation for the CONUS domain was run for oak pollen between March 1st and April 30th, 2004 by fixing four BC values as 10 pollen grains/m^3 at each time step of the CMAQ-Pollen model. The difference of simulated airborne pollen concentrations in the first layer between these two BCs was calculated to examine the BCs' impact on airborne pollen concentrations.

For simulation conducted on the domain of northeastern US, ICs were also set as zero at the first simulation hour; and ICs of subsequent hours were obtained from the simulation results of corresponding previous hours. The BCs for this domain were also set as zero at each model step.

4.2.4 Evaluation of model performance

Figure 4.2 presents the process for evaluating the performance of the WRF-SMOKE-CMAQ-Pollen modeling system. The simulated start date, season length, daily pollen concentrations, and sum of daily pollen concentrations during the pollen season (hereafter referred as Seasonal Count) in 2004 were firstly paired with the corresponding observations. For example, the observed start date at a monitor station was paired with the simulated start date in a grid that contains the corresponding pollen monitor station; likewise for pairing the season length, daily pollen concentrations and seasonal counts. As shown in Equation 4.2,

$$C(Day, i, j) = \frac{\sum_{hr \in Day} C(hr, 1, i, j)}{24} \tag{4.2}$$

the simulated daily pollen concentration at a given day in a grid (i,j) $C(Day, i, j)$ is defined as the average of the simulated hourly concentration of the first layer over 24 hours in that day. The choice of the first layer concentration is based on the fact that observation of pollen counts generally occurs in this layer. Layers are defined based on a pressure coordinate system; the first layer is between 0.993 and 1 atmosphere, which is about 0-60 m above the ground.

Hit and false rates, fractional bias and mean fractional bias were calculated for evaluation of the simulated daily pollen concentration. Procedures in the literature were followed to calculate the hit and false rates at three different concentrations[52,53], which are 5, 20 and 100 pollen grains/m^3, respectively. On the basis of the confusion table for classification problem of the machine learning model (Table 2.3), the hit rate (H_i) and false rate (F_i) for a given pollen concentration C_i are defined using Equation 4.3.

$$\begin{cases} H_i = \frac{TP_i}{TP_i+FN_i} \\ F_i = \frac{FP_i}{TP_i+FP_i} \end{cases} \tag{4.3}$$

The hit rate H_i is essentially the definition of Recall for the classification problem of the machine learning model. It indicates among observed airborne concentrations greater than or equal to C_i, how many are correctly predicted. The false rate F_i indicates among predicted airborne concentrations greater than or equal to C_i, how many are falsely predicted.

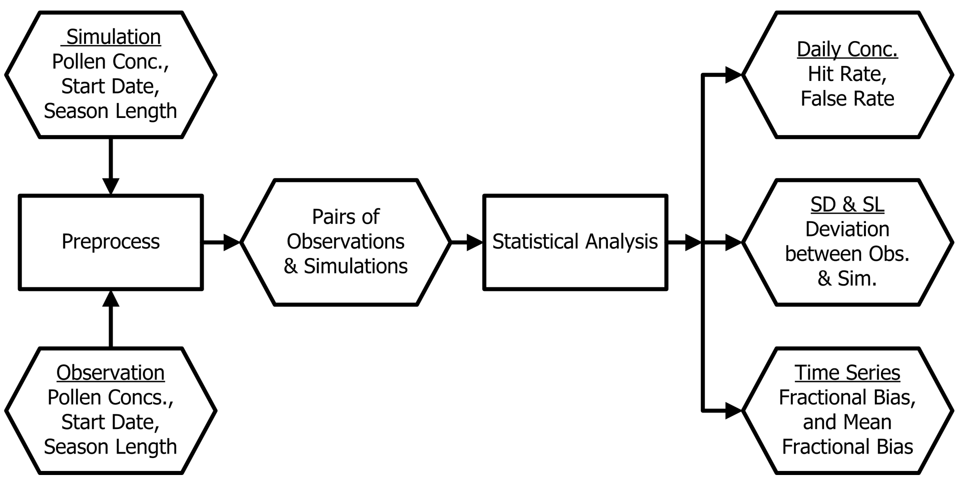

Figure 4.2: Procedures used to evaluate the performance of the WRF-SMOKE-CMAQ-Pollen modeling system based on observed pollen count.

The factional bias FB is defined in Equation 4.4,

$$FB = 2\frac{C_{Sim} - C_{Obs}}{C_{Sim} + C_{Obs}} \tag{4.4}$$

where C_{Sim} and C_{Obs} are the simulated and observed concentrations, respectively.

The difference between simulated and observed start dates was calculated to evaluate the model performance on predicting start date; likewise for evaluation of season length.

4.2.5 Process analysis of pollen transport model

The Process Analysis Preprocessor (PROCAN) was compiled together with an adapted CMAQ model to activate the process analysis function in the CMAQ-Pollen modeling system[131]. The process analysis was conducted to identify the contributions of each physical process on airborne pollen concentrations. The physical processes incorporated into the process analysis included cloud process, dry deposition, emission, horizontal and vertical advection, and horizontal and vertical diffusion. The process analysis was carried out using the time series of simulated hourly concentrations of allergenic oak pollen during the pollen season in 2004 in the grid that contains the pollen monitoring station at Springfield, New Jersey.

4.2.6 Simultaneous simulation of anthropogenic and biogenic allergens

The WRF-SMOKE-CMAQ-Pollen modeling system was applied to the Ozone Transport Commission domain in the Northeastern US with spatial resolution of 12x12 km and temporal resolution of one hour. The modeling system was run for the periods of March 1st-May 31st and July 1st-September 30th, 2007. The first period was used to simultaneously simulate the concentrations of oak pollen, ozone and particulate matter during the spring of 2007; and the second was used to simultaneously simulate the concentrations of ragweed pollen, ozone and particulate matter during the late summer and early fall.

As mentioned in section 1.6.2, the effects of vertical advection on pollen transport is potentially weak due to the large horizontal resolution (50x50 km) of meteorology data on the domain covering the CONUS. The high resolution simulations for the OTC domain also served as an example to study the impact of vertical advection and diffusion on vertical distributions of airborne pollen grains. For oak, the simulated pollen concentrations at local time of 10:00 am and 2:00 pm during April in Springfield, New Jersey were extracted and averaged at each vertical level, respectively; likewise for vertical profiles of oak pollen concentration at 10:00 am and 2:00 pm during May. For ragweed, the simulated pollen concentrations at local time of 10:00 am and 3:00 pm during August in Springfield, New Jersey were extracted and averaged at each vertical level, respectively; likewise for vertical profiles of ragweed pollen concentrations at 10:00 am and 3:00 pm during September.

4.2.7 Climate change impact on spatiotemporal distribution of allergenic pollen

Figure 4.3 presents the process for assessing climate change impact on allergenic pollen season timing and airborne levels in the CONUS. Mean hourly concentration, maximum hourly concentration, Start Date, Season Length, and number of hours exceeding threshold pollen concentrations during each of the periods of 2001-2004 and 2047-2050 were selected as five metrics to assess the impact of climate change on allergenic pollen season. The threshold concentrations for calculating the number of exceedance hours are 20, 13, 30, 30 and 20 pollen grains/m^3 for birch, oak, ragweed, mugwort and grass, respectively[133,134,15]. These threshold values were generally established based on the occurrence of first symptoms of allergy in sensitive patients[133,134,15].

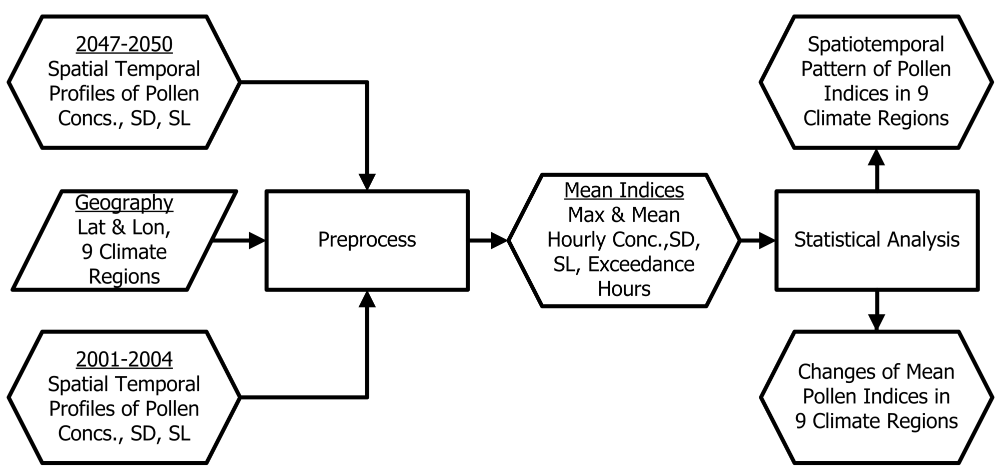

Figure 4.3: Procedures used to evaluate climate change impact on allergenic pollen season timing and airborne levels based on simulations during periods of 2001-2004 and 2047-2050.

The hourly pollen concentration during the first period *P1* (i.e., 2001-2004) for each

50x50 km grid ($C^{P1}(hr,i,j)$) was calculated through Equation 4.5,

$$C^{P1}(hr,i,j) = \frac{\sum_{yr} C^{P1}(yr,hr,1,i,j)}{N_{yr}} \tag{4.5}$$

where $C^{P1}(yr,hr,1,i,j)$ is the simulated pollen concentration of the first layer at hour (hr) in year (yr) during period *P1* in grid (*i,j*); N_{yr} is the number years in period *P1*. The reason for choosing the first layer concentration is because allergic airway disease is mainly related to the concentrations of aeroallergens in this layer.

The mean hourly concentration ($C^{P1}_{hr,Mn}(i,j)$), maximum hourly concentration ($C^{P1}_{hr,Mx}(i,j)$), start date ($SD^{P1}(i,j)$), season length ($SL^{P1}(i,j)$), and number of exceedance hours ($N^{P1}_{Exd}(i,j)$) for each grid during the first period were then calculated using the hourly pollen concentrations at each grid $C^{P1}(hr,i,j)$ through Equation 4.6,

$$\begin{cases} C^{P1}_{hr,Mn}(i,j) = \frac{\sum_{hr} C^{P1}(hr,i,j)}{N_{hr}} \\ C^{P1}_{hr,Mx}(i,j) = \max_{hr} C^{P1}(hr,i,j) \\ SD^{P1}(i,j) = \frac{\sum_{yr} SD^{P1}(yr,i,j)}{N_{yr}} \\ SL^{P1}(i,j) = \frac{\sum_{yr} SL^{P1}(yr,i,j)}{N_{yr}} \\ N^{P1}_{Exd}(i,j) = \sum_{hr} \mathbb{1}_{C^{P1}(hr,i,j) \geq C_{Thr}} \\ C^{P1}_{hr,Std}(i,j) = \frac{\sum_{hr}(C^{P1}(hr,i,j) - C^{P1}_{hr,Mn}(i,j))^2}{N_{hr}} \end{cases} \tag{4.6}$$

where N_{hr} is number of simulation hours in each year during period *P1*; $C^{P1}_{hr,Std}(i,j)$ is the standard deviation of hourly pollen concentration at grid (*i,j*) during period *P1*. $\mathbb{1}$ is the indicator function; it takes 1 as its value when the hourly concentration $C^{P1}(hr,i,j)$ is greater or equal to the threshold concentration C_{Thr}, otherwise takes 0 as its value.

Similarly, the mean hourly concentration ($C^{P2}_{hr,Mn}(i,j)$), maximum hourly concentration ($C^{P2}_{hr,Mx}(i,j)$), start date ($SD^{P2}(i,j)$), season length ($SL^{P2}(i,j)$), and number of exceedance hours ($N^{P2}_{Exd}(i,j)$) for each grid during the second period (i.e., 2047-2050) can also be calculated using Equation 4.6. The change of each of the above five metrics between periods of 2001-2004 and 2047-2050 was calculated using Equation 4.7.

$$\begin{cases} \Delta C_{hr,Mn}(i,j)/C_{hr,Mn}^{P1}(i,j) = \frac{C_{hr,Mn}^{P2}(i,j)-C_{hr,Mn}^{P1}(i,j)}{C_{hr,Mn}^{P1}(i,j)} \\ \Delta C_{hr,Mx}(i,j)/C_{hr,Mx}^{P1}(i,j) = \frac{C_{hr,Mx}^{P2}(i,j)-C_{hr,Mx}^{P1}(i,j)}{C_{hr,Mx}^{P1}(i,j)} \\ \Delta SD(i,j) = SD^{P2}(i,j) - SD^{P1}(i,j) \\ \Delta SL(i,j) = SL^{P2}(i,j) - SL^{P1}(i,j) \\ \Delta N_{Exd}(i,j)/N_{Exd}^{P1}(i,j) = \frac{N_{Exd}^{P2}(i,j)-N_{Exd}^{P1}(i,j)}{N_{Exd}^{P1}(i,j)} \end{cases} \tag{4.7}$$

The mean and standard deviation of the change of a given metric (e.g, SD) in a climate region could be therefore derived using the changes in the grids (e.g., $\Delta SD(i,j)$) in that region. For example, Equation 4.8 was used to calculate the mean and standard deviations of changes in start dates in climate region k,

$$\begin{cases} \overline{\Delta SD}_k = \frac{\sum_{(i,j)\in \text{Region}k} \Delta SD(i,j)}{N_k} \\ \Delta SD_{k,Std} = \frac{\sum_{(i,j)\in \text{Region}k} (\Delta SD(i,j) - \overline{\Delta SD}_k)^2}{N_k} \end{cases} \tag{4.8}$$

where N_k is the number of grids in climate region k. The mean and standard deviation of other metrics in other climate regions were calculated similarly following the method in Equation 4.8.

4.3 Results and Discussion

In the following discussion, oak and ragweed pollen were served as examples to demonstrate the processes and steps for evaluating model performance, and for assessing climate change impact on allergenic pollen season.

4.3.1 Ambient concentration profile of allergenic pollen

Figure 4.4 presents representative time slices of ambient concentrations of oak pollen in the first layer (i.e., 0-60 m above the ground) for 2004. Figures 4.4A and B display the spatial distribution of ambient concentrations at morning and afternoon time in an early period (March) of oak pollen season in the CONUS; and Figures 4.4C and D display concentration distributions at morning and afternoon time in a late period (April) of pollen season.

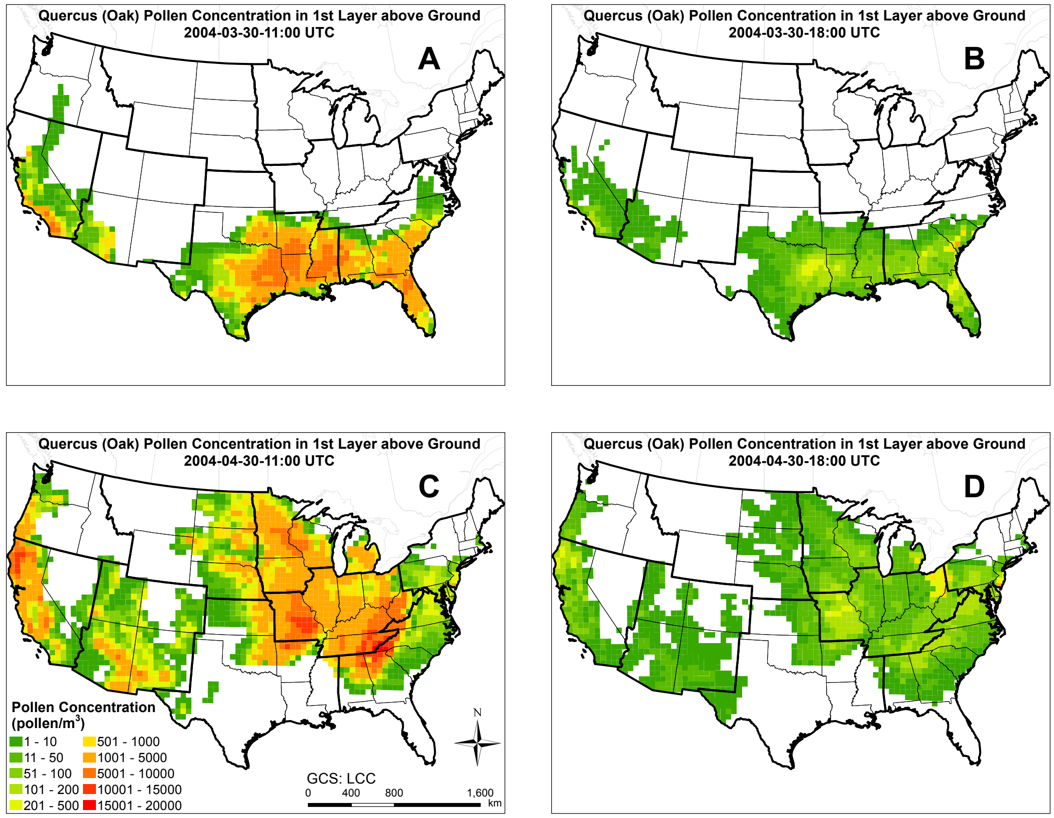

Figure 4.4: Time slices of spatiotemporal concentration profiles of oak pollen for 2004 in the first layer with spatial resolution of 50x50km. The first layer is about 0-60 m above the ground. (A) Mar. 30th 11:00 UTC, (B) Mar. 30th 18:00 UTC, (C) Apr. 30th 11:00 UTC, and (D) Apr. 30th 18:00 UTC.

As shown in Figure 4.4, ambient oak pollen grains in ground surface layer were mainly distributed in the southern and western US in March, and in the northern US in April. Higher ambient concentrations could be found in the Southeast and South climate regions in March, and Central and Eastern North Central climate regions in April. These higher ambient concentrations are consistent with area coverage of oak trees in Figure 3.5B and emission profiles in Figure 3.13. The lower ambient oak pollen concentrations at afternoon time were mainly caused by lower oak pollen emission in afternoon.

Figure 4.5 depicts the representative time slices of ambient concentrations of oak pollen in the tenth layer (i.e., 1.2-1.5 km above the ground) for 2004. As shown in Figure 4.5, oak pollen concentrations in higher layers (e.g., layer 10) were far lower than those in the first layer. However they could distribute across the entire CONUS due to long range transport of pollen grains, because of higher wind speed at higher altitudes.

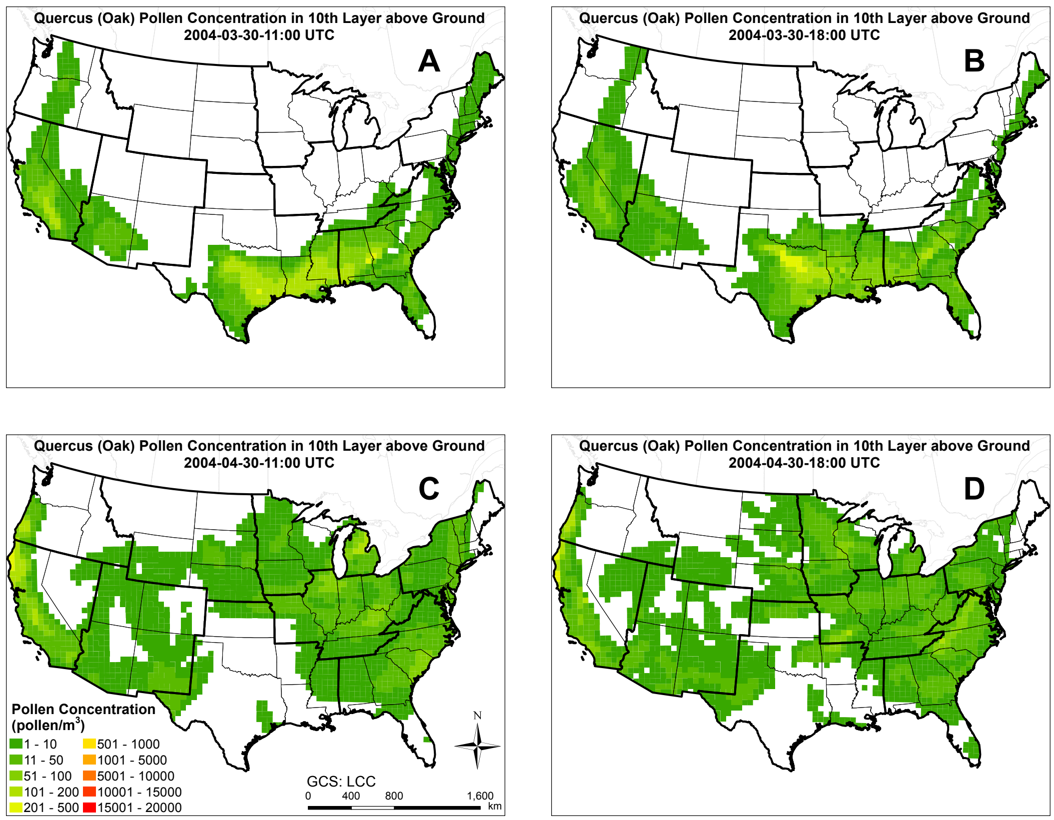

Figure 4.5: Time slices of spatiotemporal concentration profiles of oak pollen for 2004 in the tenth layer with spatial resolution of 50x50km. The tenth layer is about 1.2-1.5 km above the ground. (A) Mar. 30th 11:00 UTC, (B) Mar. 30th 18:00 UTC, (C) Apr. 30th 11:00 UTC, and (D) Apr. 30th 18:00 UTC.

Figure 4.6 presents representative time slices of ambient concentrations of ragweed pollen in the first layer for 2004. Figures 4.6A and B display the spatial distribution of ambient concentrations at morning and afternoon time in an early period (August) of ragweed pollen season in the CONUS; and Figures 4.6C and D display concentration distributions at morning and afternoon time in a late period (September) of pollen season.

As shown in Figure 4.6, ambient ragweed pollen grains in the ground surface layer were mainly distributed in the northern US in August, and in the southern and western

US in September. Higher ambient concentrations could be found in the climate regions of West North Central in August, and South, Southwest and West in September. These higher ambient concentrations are consistent with area coverage of ragweed plants in Figure 3.5C and emission profiles in Figure 3.14.

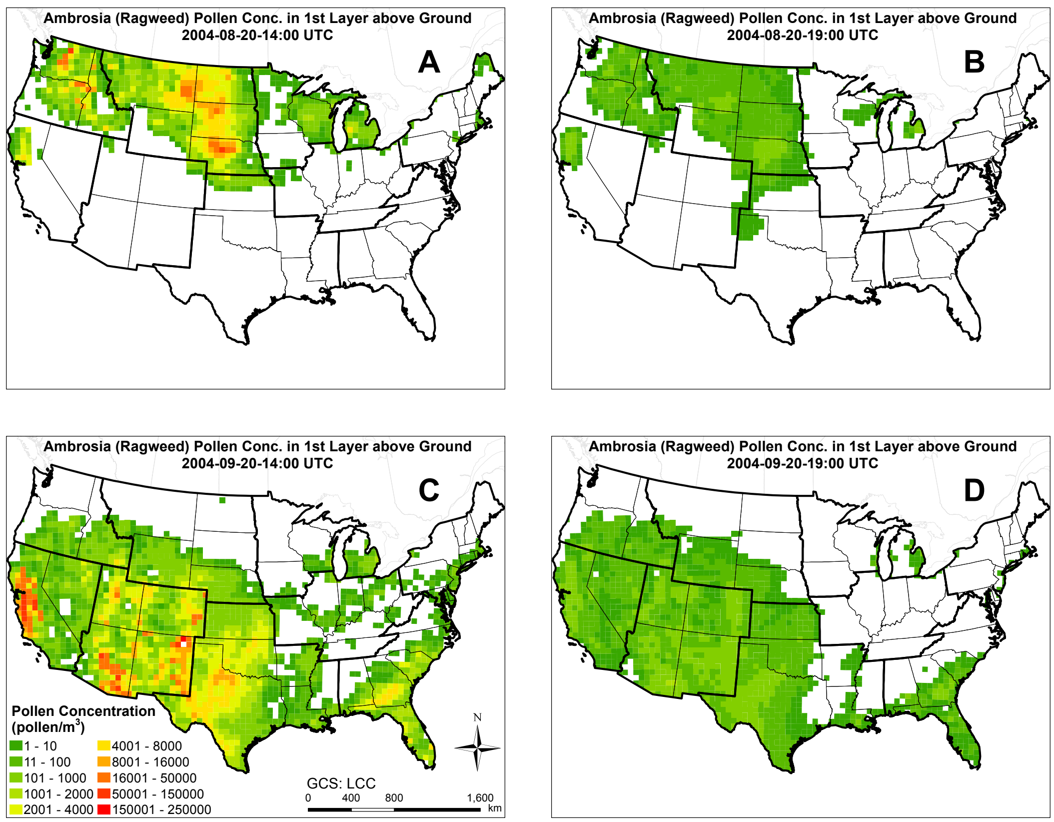

Figure 4.6: Time slices of spatiotemporal concentration profiles of ragweed pollen for 2004 in the first layer with spatial resolution of 50x50km. The first layer is about 0-60 m above the ground. (A) Aug. 20th 14:00 UTC, (B) Aug. 20th 18:00 UTC, (C) Sep. 20th 14:00 UTC, and (D) Sep. 20th 19:00 UTC.

4.3.2 Performance evaluation of pollen transport model

Evaluation of simulated pollen season timing and ambient levels

Figure 4.7 presents the hit rates and false rates for predicted and observed daily oak pollen concentrations for three pollen levels at the studied stations during 2004 across the CONUS. The hit rates for airborne oak pollen levels of 5, 20 and 100 pollen grains/m^3 were all between 0.7 and 1 for most of the studied stations, and close to

1 for the majority of stations. This indicates that the observed daily oak pollen levels at most the studied stations were correctly predicted by the WRF-SMOKE-CMAQ-Pollen modeling system. The false rates for airborne pollen levels of 5, 20 and 100

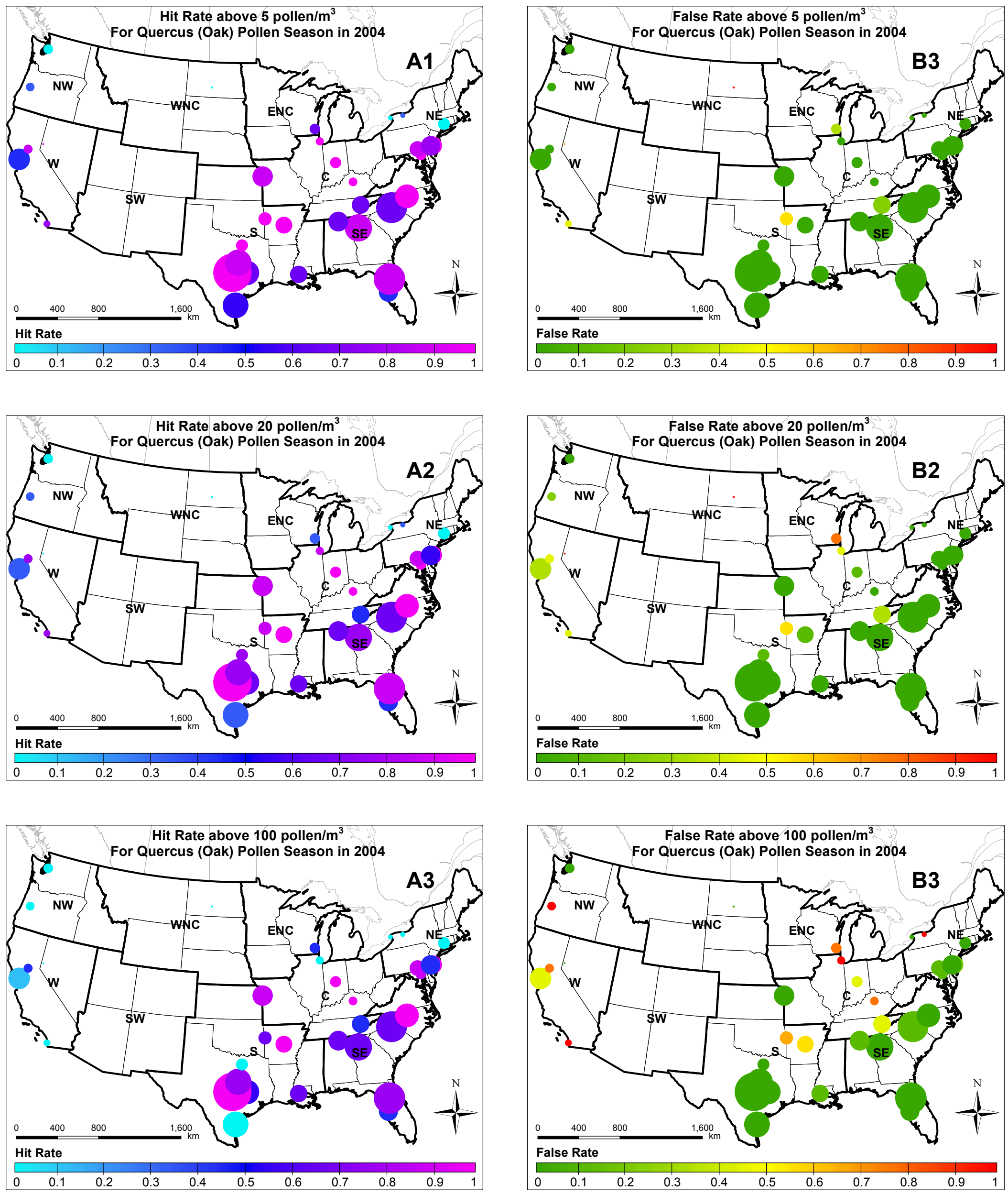

Figure 4.7: Hit and false rates for predicted and observed daily oak pollen concentration during 2004 in the CONUS. The size of the circle indicates the oak pollen abundance at that station. (A1-A2) Hit rates for 5, 20 and 100 pollen grains/m^3, respectively; (B1-B3) False rates for 5, 20 and 100 pollen grains/m^3, respectively.

pollen grains/m^3 were all between 0 and 0.1 for most of the studied stations, and close to 0 for the majority of stations. This indicates that the predicted daily oak pollen levels at most of the studied stations are not falsely estimated.

Figure 4.8A shows the mean fractional bias between the observed and simulated daily oak pollen concentrations at the studied NAB stations during 2004 across the CONUS; Figure 4.8B shows the fractional bias between simulated and observed seasonal pollen count. The mean fractional bias for daily oak pollen concentration were between

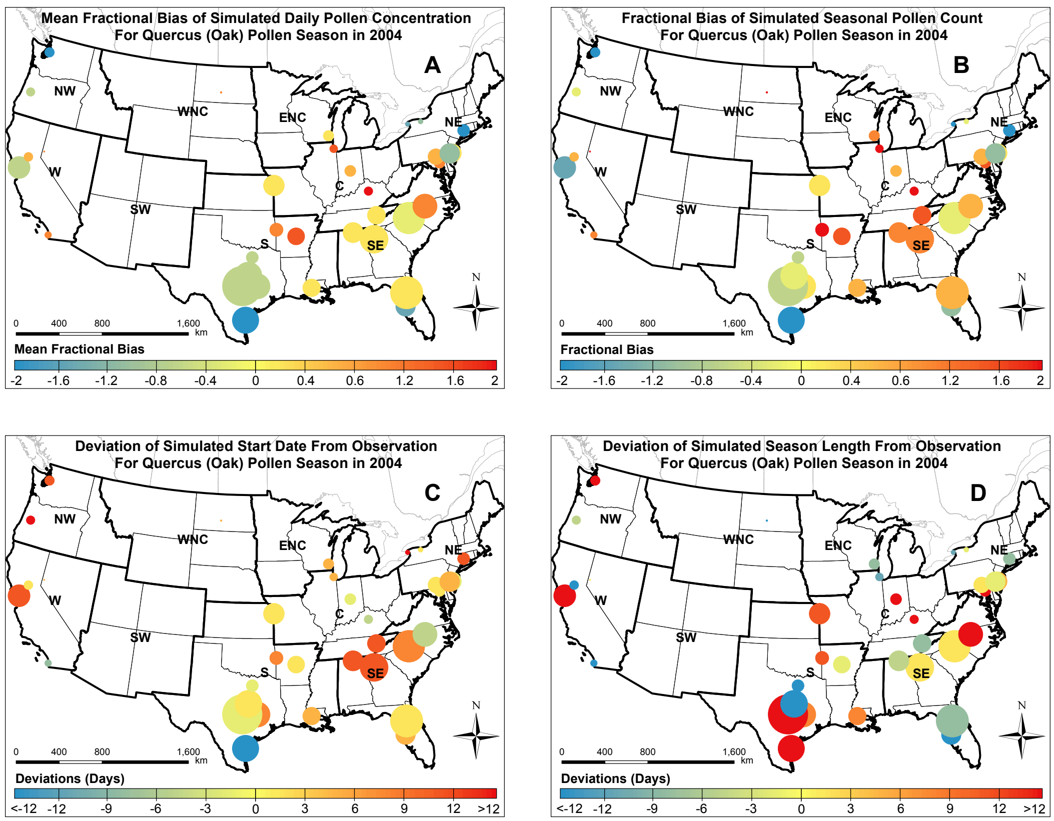

Figure 4.8: Evaluation of predicted oak pollen season during 2004 in the CONUS. The size of the circle indicates the oak pollen abundance at that station. (A) Mean Fractional Bias of daily pollen concentration, (B) Fractional Bias of seasonal pollen counts, (C) deviation between observed and simulated Start Dates, and (D) deviation between observed and simulated Season Length.

-0.4 to 0.4 for the majority of studied stations. This indicates the predicted daily oak pollen concentrations were in good agreement with the observations. The fractional bias for seasonal pollen count was not as good as that for daily pollen concentration, but still reasonable for many studied stations. The potential reason for the weak performance on seasonal pollen count may be due to the misalignment of predicted and observed oak pollen season duration (Figure 4.8D).

Figure 4.8C displays the deviation between simulated and observed start dates of oak pollen seasons at the studied pollen stations during 2004 across the CONUS. Figure 4.8D displays the deviation between simulated and observed season length at the studied pollen stations across the CONUS. The deviations of start date were generally between 0 and 6 days for the majority of studied stations. The deviations of season length were relatively large, and generally between 0 and 9 days. These are consistent with the results in Figure 3.9B and Figure 3.10B, where the SD and SL during 1994-2010 were calculated based on observed meteorology data.

Figure 4.9 presents the hit rates and false rates for predicted and observed daily ragweed pollen concentrations for three pollen levels at the studied stations across the CONUS. The hit rates for airborne ragweed pollen levels of 5, 20 and 100 pollen grains/m^3 at the stations in the central CONUS were generally between 0.7 and 1. The hit rates at stations along the east coast of the CONUS were generally low, particularly for pollen level of 100 grains/m^3. The low hit rates at the stations in eastern CONUS are mainly caused by the low area coverage of ragweed plant in these areas. As shown in Figure 3.5C, area coverage of ragweed plants along the east coast of the CONUS is lower, leading to low emission fluxes and airborne concentrations in these areas.

The false rates for airborne ragweed pollen levels of 5, 20 and 100 pollen grains/m^3 were all generally between 0 and 0.1 for most of the studied stations. The predicted ragweed pollen concentration in the eastern US were generally low due to the low area coverage, and therefore had no chance to falsely overpredict ragweed pollen concentrations in the eastern CONUS.

Figure 4.10A shows the mean fractional bias between the observed and simulated daily ragweed pollen concentrations at the studied NAB-AAAAI stations during 2004 across the CONUS; Figure 4.10B shows the fractional bias between simulated and observed seasonal pollen count. The mean fractional bias for daily ragweed pollen concentrations was between -0.4 and 0.4 for the majority of studied stations in the

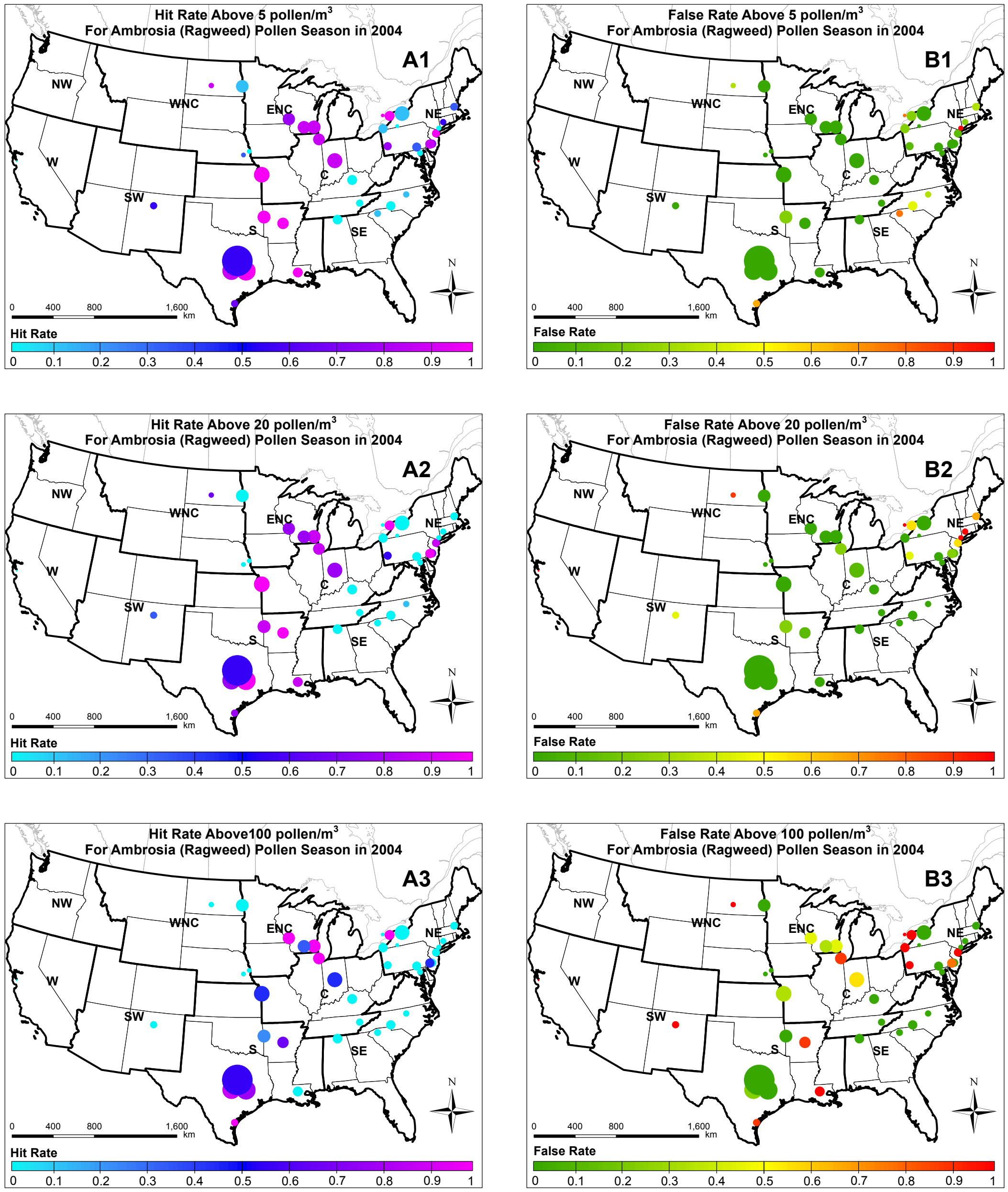

Figure 4.9: Hit and false rates for predicted and observed daily rageed pollen concentration during 2004 in the CONUS. The size of the circle indicates the ragweed pollen abundance at that station. (A1-A2) Hit rates for 5, 20 and 100 pollen grains/m^3, respectively; (B1-B3) False rates for 5, 20 and 100 pollen grains/m^3, respectively.

South, Central and East North Central climate regions. The mean fractional bias was generally low at the stations in the south east climate region. This indicates the ragweed daily concentrations in the Southeast were underestimated. It is potentially due to the low area coverage in this region. The fractional bias of ragweed seasonal count has a pattern similar to that of the mean fractional bias of daily concentrations.

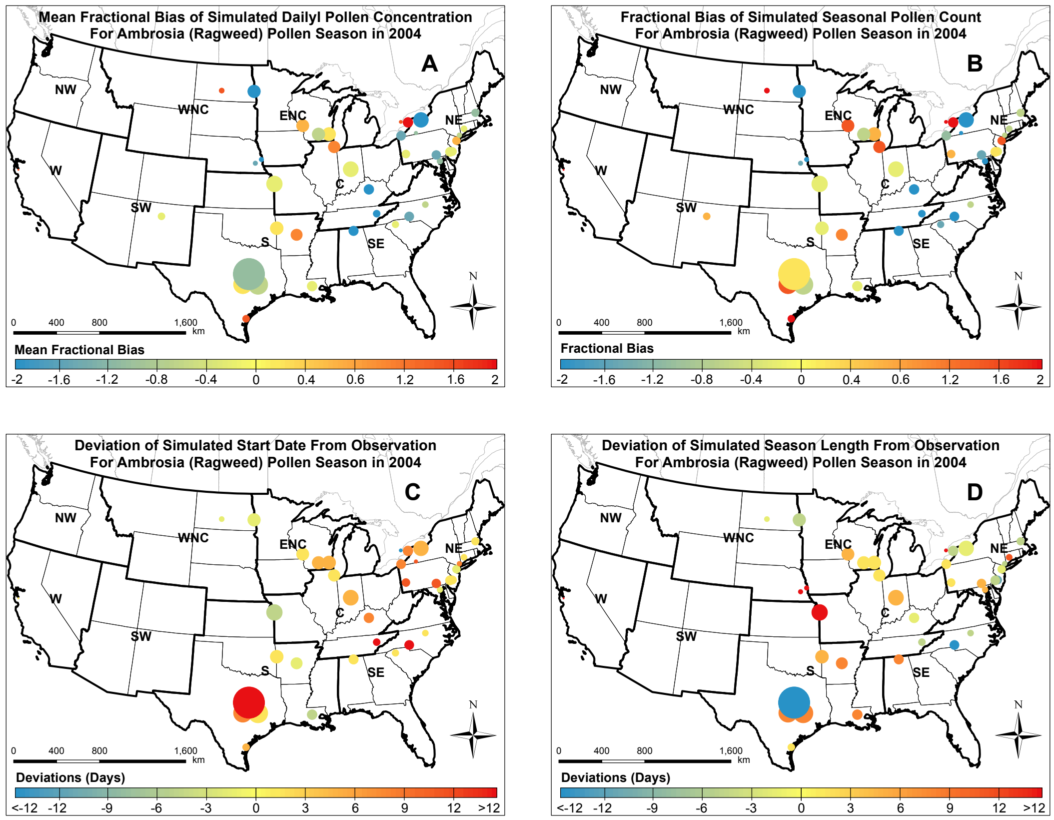

Figure 4.10: Evaluation of predicted ragweed pollen season during 2004 in the CONUS. The size of the circle indicates the ragweed pollen abundance at that station. (A) Mean Fractional Bias of daily pollen concentration, (B) Fractional Bias of seasonal pollen counts, (C) deviation between observed and simulated Start Dates, and (D) deviation between observed and simulated Season Length.

Figure 4.10C displays the deviation between simulated and observed start dates of ragweed pollen season at the studied pollen stations during 2004 across the CONUS. Figure 4.10D displays the deviation between simulated and observed season length at the studied pollen stations across the CONUS. The deviations of start date were generally between 0 and 6 days for the majority of studied stations. The deviations of season length were also generally between 0 and 6 days. These are consistent with the results in Figure 3.9C and Figure 3.10C.

The evaluation metrics for birch, mugwort and grass are presented in Figures D.1, D.2, and D.3 in the Appendix D, section D.2.

Process analysis

Figure 4.11 presents the contributions of advection, diffusion, dry deposition and cloud process on the hourly oak pollen concentrations between local time EST 5:00 pm April 11th to EST 5:00 pm April 13th, 2004 in Springfield, New Jersey. The dry deposition, emission and vertical eddy diffusion were the dominant processes determining ambient concentrations of oak pollen. The emission process continuously released pollen grains

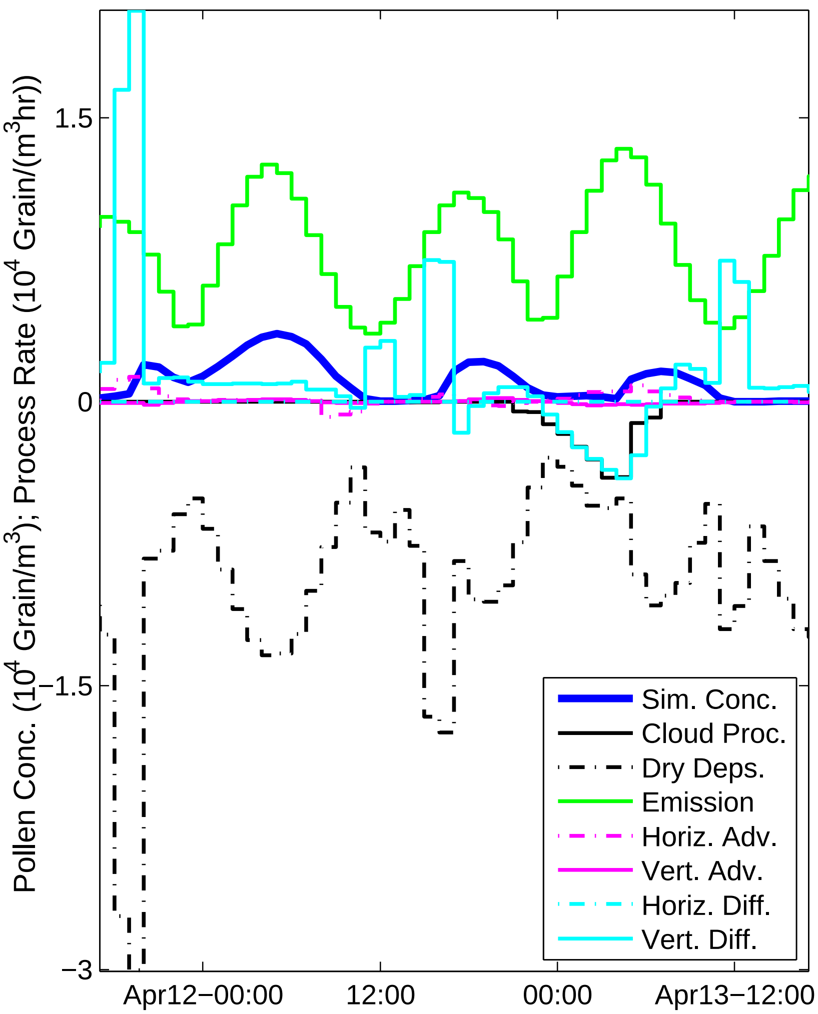

Figure 4.11: Contributions of advection, diffusion, dry deposition and cloud process on the hourly oak pollen concentrations between local time EST 17:00 April 11th to EST 17:00 April 13th, 2004 in Springfield, New Jersey.

into the air following a regular diurnal pattern. The majority of ambient pollen grains were removed from the air through dry deposition. Vertical diffusion may dominate the transport of ambient pollen grains when there is strong turbulent atmospheric movement. The vertical diffusion could bring the pollen grains down to the first layer from above layers under special weather conditions, and therefore increase the pollen concentrations in the first layer. The cloud process also played an important role through in-cloud and below-cloud scavenging during rainy time (black line between April 12th and 13th in Figure 4.11).

Influence of boundary conditions

Figure 4.12 displays the difference in ambient oak pollen concentration due to different boundary conditions. In most of the grids, the mean hourly pollen concentrations were not influenced by the boundary conditions. In some grids, the mean hourly concentrations increased around 30% because of the changes in boundary conditions. The low

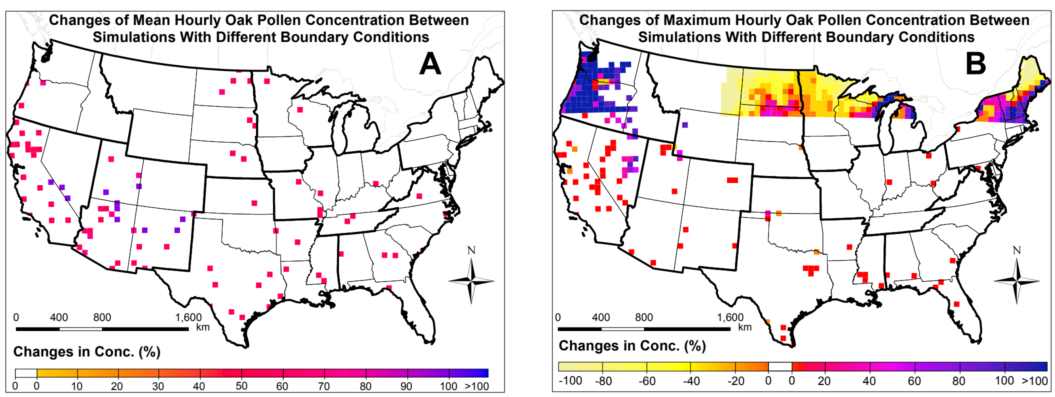

Figure 4.12: The difference in hourly concentrations of oak pollen between two different Boundary Conditions. The default BC was set as 0 pollen grains/m^3, and the other BC was set as 10 pollen grains/m^3. (A) Difference in mean hourly concentration from March 1st to April 30th, 2004; (B) Difference in maximum hourly concentration.

influence of boundary conditions on ambient pollen concentrations is consistent with the findings from the literature [48]. The maximum hourly concentrations in the states bordering Canada tend to change significantly due to the change in boundary conditions. The maximum hourly concentrations increased around 100% in Washington, Oregon and New Hampshire; and decreased around 80% in Maine, North Dakota and

Minnesota. The large variations in maximum hourly concentrations may be related to the distribution of oak trees in the grids of neighboring Canadian regions.

4.3.3 Simultaneous simulation of anthropogenic and biogenic allergen

Ambient concentrations of allergenic pollen and air pollutants

For demonstration of the WRF-SMOKE-CMAQ-Pollen modeling system's capability to simultaneously simulate distribution of anthropogenic and biogenic allergens, Figure 4.13 depicts the simulated mean hourly concentrations of oak and ragweed pollen, ozone and $PM_{2.5}$ (i.e., particulate matter with diameter less than 2.5 μm) covering the OTC domain with spatial resolution of 12x12 km. The simulations were conducted for the periods of March 1st-April 30th and July 1st-September 30th, 2007. These two periods correspond to allergenic pollen season in spring and fall, respectively.

The simultaneous simulation of biogenic and anthropogenic allergens can help to identify "hot spots" with higher concentrations of both allergenic pollen and air pollutants. For example, as shown in Figures 4.13A2-C2, New Jersey had the relatively high ambient concentrations of ragweed pollen, ozone and $PM_{2.5}$ from July 1st to September 30th, 2007.

Vertical profile of pollen concentration

The simulation results on the OTC domain with spatial resolution of 12x12 km were also used to examine the vertical profiles of oak and ragweed pollen concentrations. Figure 4.14 presents the vertical profiles of mean hourly concentrations for oak and ragweed pollen during 2007 in Springfield, New Jersey.

As shown in Figure 4.14, the ambient concentration of both oak and ragweed pollen at afternoon time (EST 2 or 3 PM) was lower in the first layer, and roughly the same in other layers. The ambient concentration of both oak and ragweed pollen at morning time (EST 10 PM) tended to increase with the increase of elevation from the first layer to a threshold height; and then tended to decrease with increase of elevation from the threshold height. Theses profiles are mainly due to temperature inversion at morning time and strong mixing of atmosphere in the vertical direction at afternoon time.

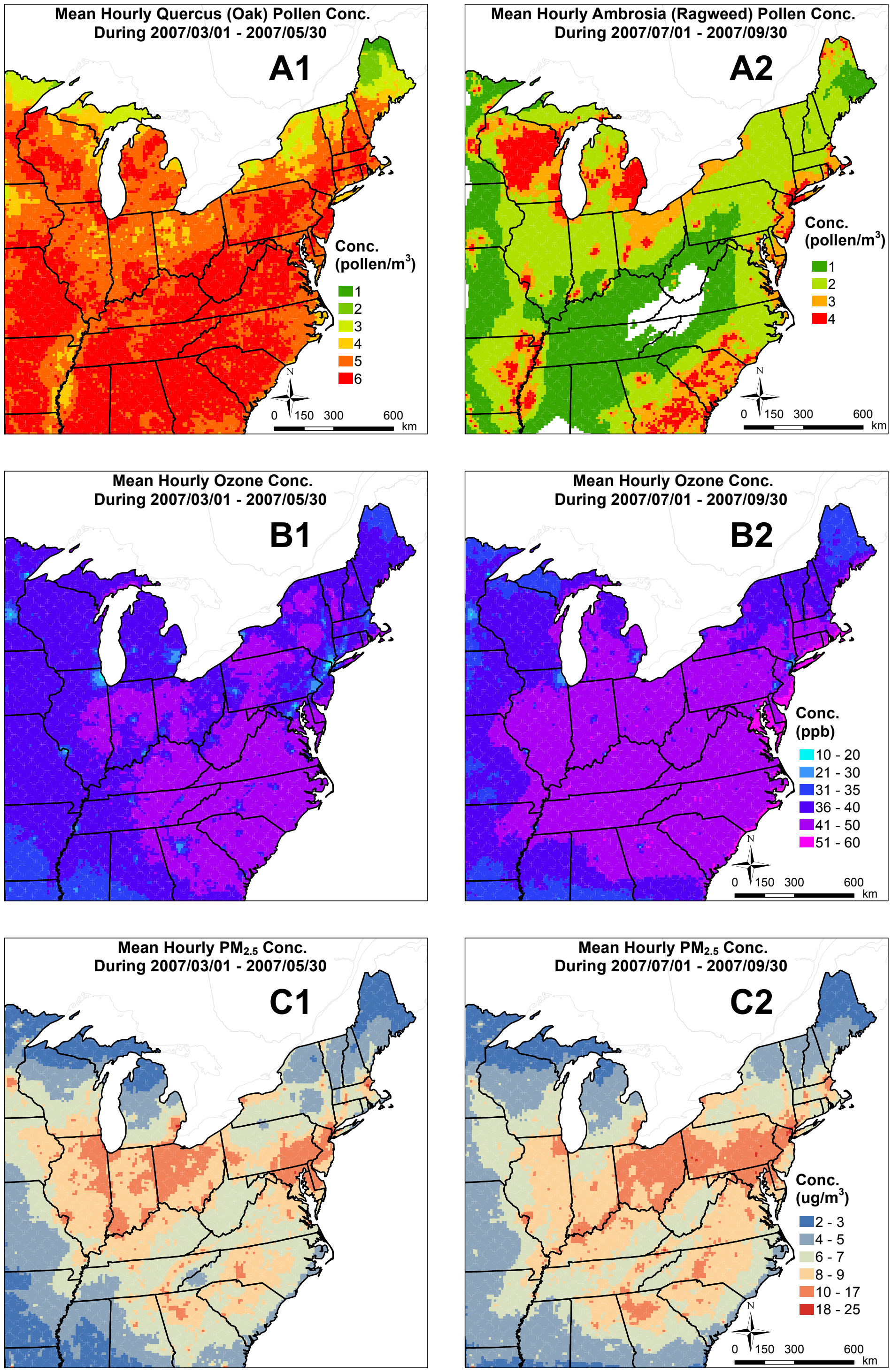

Figure 4.13: Simulated mean hourly concentrations of oak and ragweed pollen, ozone and $PM_{2.5}$ during 2007 on the Ozone Transport Commission domain. (A1-C1) Mean hourly concentrations for oak pollen, ozone and $PM_{2.5}$ from March 1st to May 31st, respectively; and (A2-C2) Mean hourly concentrations for ragweed pollen, ozone and $PM_{2.5}$ from July 1st to September 30th, respectively.

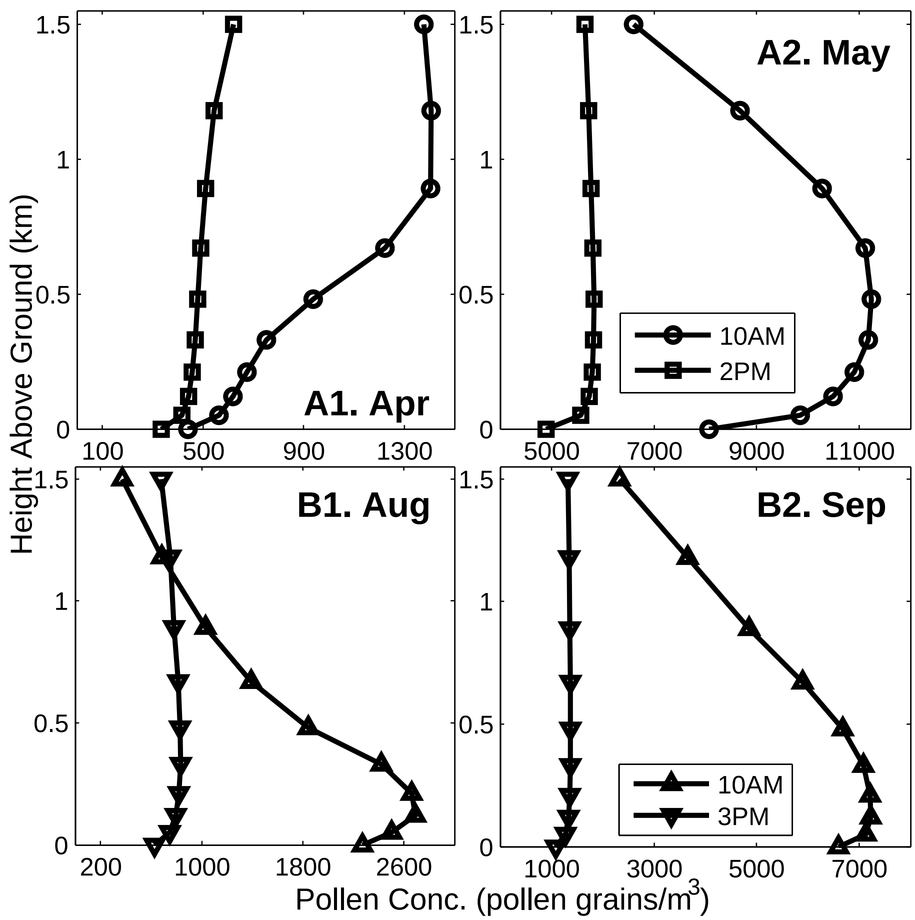

Figure 4.14: Vertical profiles of oak and ragweed pollen concentrations during 2007 in Springfield, New Jersey. (A1-A2) Mean hourly oak pollen concentrations at EST 10 AM and 2 PM in April and May, respectively; (B1-B2) Mean hourly ragweed pollen concentrations at EST 10 AM and 3 PM in August and September, respectively.

The simulated vertical profiles of pollen concentrations could not be evaluated using actual pollen counts because of the unavailability of pollen counts at multiple heights at a given monitoring station. However, these simulated vertical profiles of pollen concentrations are consistent with vertical distributions of aerosol backscatter coefficients measured using the method of Light Detection and Ranging (LIDAR) during the pollen season [135]. The vertical distributions of aerosol backscatter coefficients were reported to be associated with large amounts of pollen grains in the air during the pollen season [135].

4.3.4 Climate change impact on allergenic pollen

Changes of temperature and precipitation during periods of 2001-2004 and 2047-2050

Figure 4.15 presents the changes of average Surface Air Temperature (SAT) and cumulative precipitations during the periods of 2001-2004 and 2047-2050. Average SAT

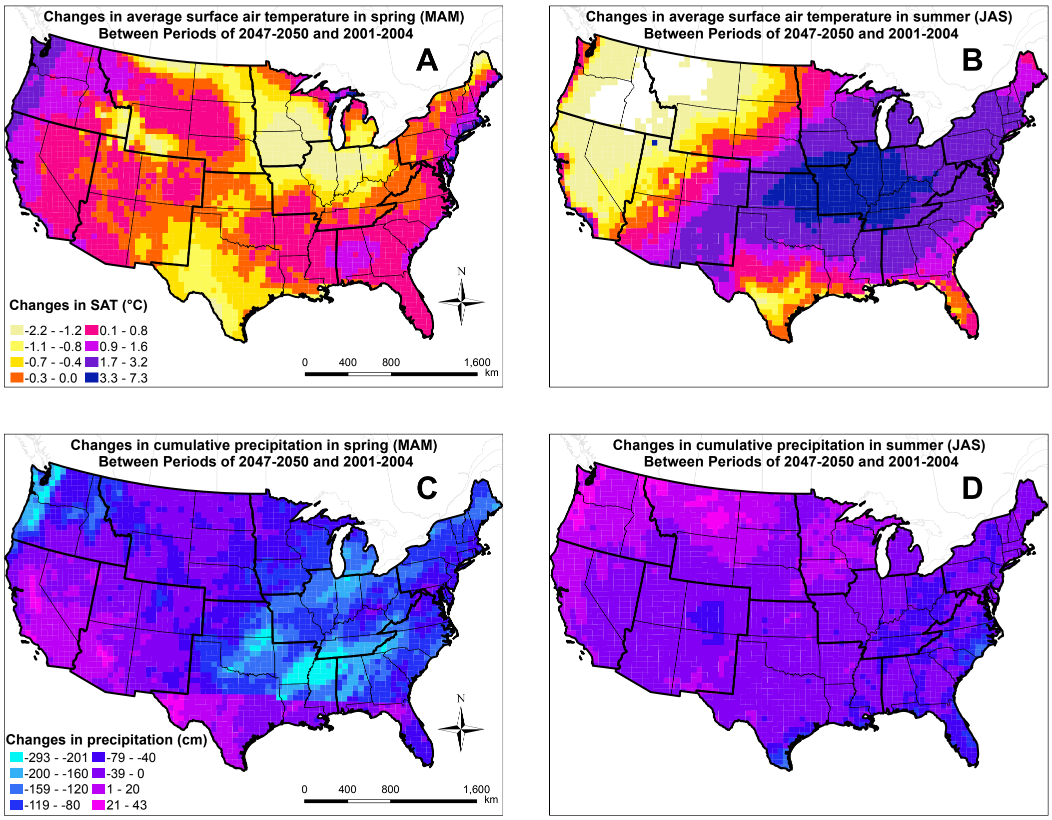

Figure 4.15: Changes of mean surface air temperature and accumulative precipitation between periods of 2001-2004 and 2047-2050. (A) Average SAT in spring (MAM: March, April and May), (B) Average SAT in summer (JAS: July, August and September), (C) Accumulative precipitation in spring, and (D) Accumulative precipitation in summer. Average SAT and cumulative precipitation were derived using the hourly simulations in each of the periods.

and cumulative precipitation were derived for spring (MAM: March, April and May) and summer (July, August and September) using the hourly simulations in each of the periods of 2001-2004 and 2047-2050. The spring temperature and precipitation are relevant to allergenic pollen season onset and airborne pollen levels of trees and grass.

The summer temprature and precipitation are relevant to the pollen season onset and airborne levels of weeds.

As shown in Figure 4.15A, the average surface air temperature in spring was predicted to increase from period of 2001-2004 to 2047-2050 along the west and east coasts of the CONUS, but to decrease in the central and the southern CONUS. The average summer SAT was predicted to increase from the period of 2001-2004 to 2047-2050 in the eastern CONUS, but to decrease in the western CONUS. The precipitation were predicted to decrease in the majority of regions during spring, but to slightly increase during the summer. The changes in these meteorology factors would driven relative changes in simulated allergenic pollen season.

The changes of simulated meteorology are expected to have substantial uncertainties. First, internal climate variability could not be captured by four years' simulations. Second, the meteorology output used in the current study were from only one regional meteorology model and one global climate model. The one ensemble member model may not be able to represent the full atmospheric physics and ocean dynamics.

Distributions of allergenic pollen during periods of 2001-2004 and 2047-2050

Figure 4.16 presents the spatial distributions of the mean and maximum hourly concentrations of oak pollen during the periods of 2001-2004 and 2047-2050 in the CONUS. The standard deviation is also presented to characterize variation of hourly concentrations of oak pollen in each grid during these two periods. These three metrics in each grid during each period were calculated using Equation 4.6 based on the simulated hourly concentrations of the WRF-SMOKE-CMAQ-Pollen modeling system. The spatial distributions of oak pollen concentrations in both periods followed the pattern of area coverage of oak trees in Figure 3.5B.

The mean hourly concentration of oak pollen varied among different regions. It ranged from 0-10 pollen grains/m^3 in the of Northwest and West North Central climate regions, to 1201-1800 pollen grains/m^3 in the Southeast and Central climate regions. The maximum hourly concentrations range from 0-500 pollen grains/m^3 in the Northwest and West North Central climate regions, to 2x10^4-3x10^4 pollen grains/m^3 in the

Southeast and Central climate regions. The hourly pollen concentrations in the Southeast climate region had the highest variation during pollen season, ranging from 3001 to 4000 pollen grains/m^3.

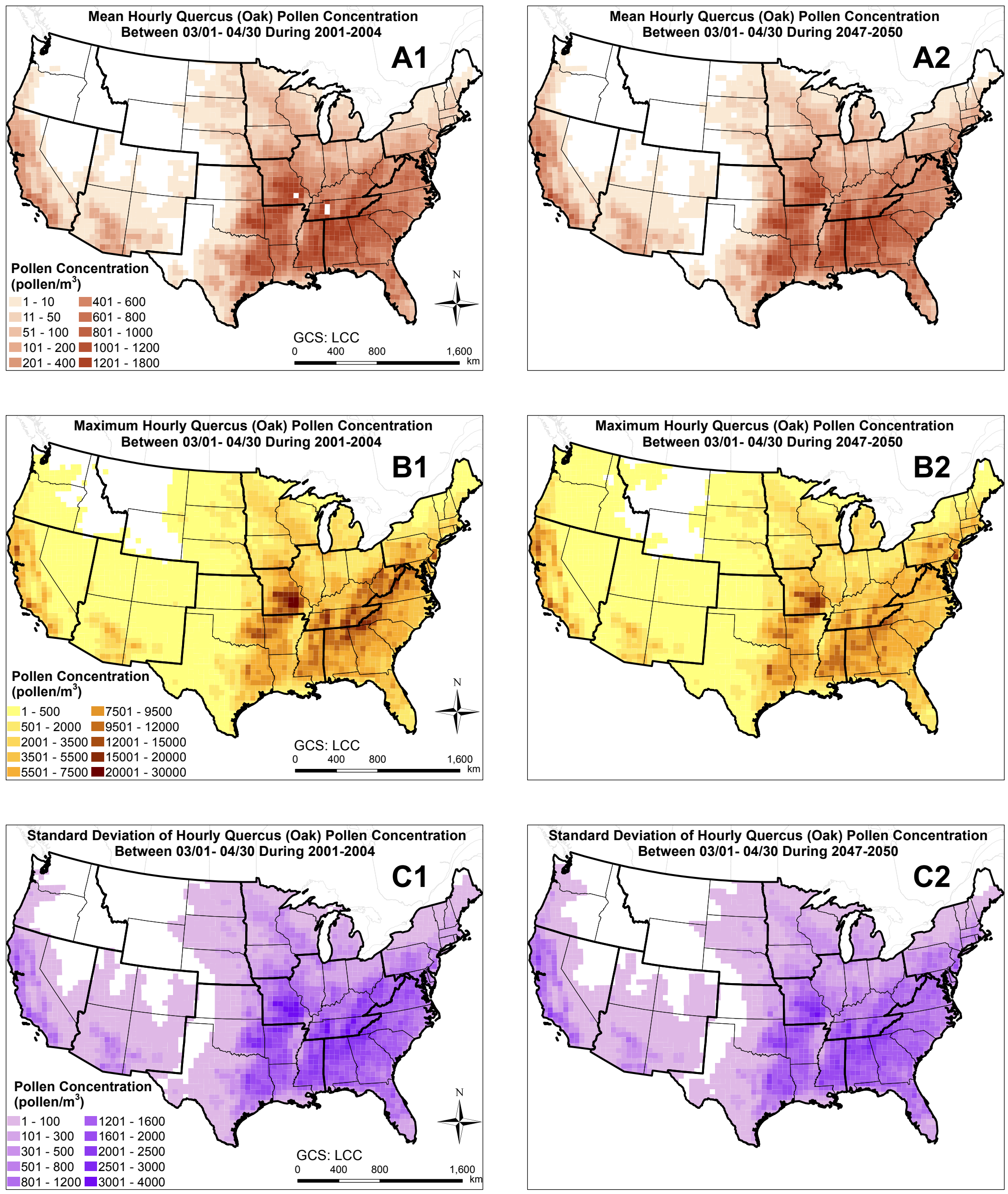

Figure 4.16: Mean, maximum and standard deviation of the simulated hourly concentrations of oak pollen during periods of 2001-2004 and 2047-2050. (A1-C1) Mean, maximum and standard deviation during 2001-2004, and (A2-C2) Mean, maximum and standard deviation during 2047-2050.

Figure 4.17 displays the average start date and season length of oak pollen season during the periods of 2001-2004 and 2047-2050 in the CONUS. These two metrics in each grid during each period were calculated using Equation 4.6 based on the simulated start date and season length. The oak pollen season in both periods started around March in the southern US, and around April in the northern US. The season length ranged between 14-22 days in the northern US, and 39-46 days in the southern US.

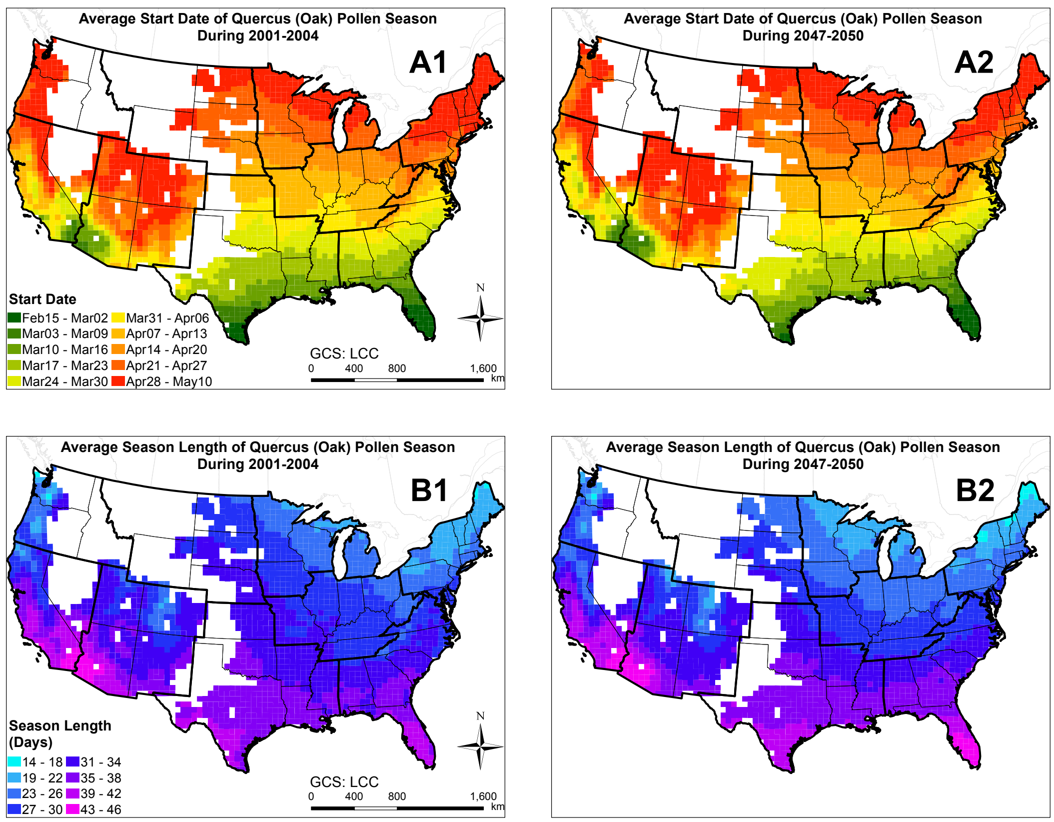

Figure 4.17: Average Start Date and Season Length of oak pollen season during periods of 2001-2004 and 2047-2050. Data were mapped only on cells in which the area coverage of oak trees is greater than zero. (A1-B1) Average SD and SL during 2001-2004, and (A2-B2) Average SD and SL during 2047-2050.

Figure 4.18 displays the number of hours in which oak pollen concentration exceeds 13 pollen grains/m^3 during the periods of 2001-2004 and 2047-2050 in the CONUS. The exceedance hour in each grid during each period was calculated using Equation 4.6 based on the simulated hourly concentrations of the WRF-SMOKE-CMAQ-Pollen modeling system. The exceedance hours in the Southeast climate region ranged from 900 to 1200 hours per year. This was around one hundred times higher than those in the Northwest and North West Central climate regions.

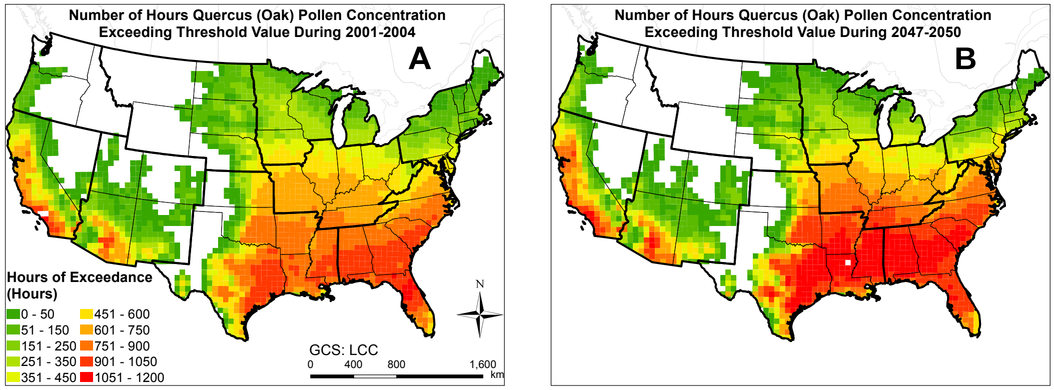

Figure 4.18: Number of hours in which oak pollen concentrations exceed 13 pollen grains/m^3 during periods of 2001-2004 and 2047-2050. (A) Average exceedance hours during 2001-2004 and (B) Average exceedance hours during 2047-2050.

The spatial profiles of ragweed pollen concentrations, start date, season length and exceedance hours are displayed in Figures D.4, D.5 and D.6, respectively, in Appendix D.

Changes of allergenic pollen season between periods of 2001-2004 and 2047-2050

Figure 4.19 presents changes of mean and maximum hourly concentrations, start date, season length and exceedance hours for oak pollen between periods of 2001-2004 and 2047-2050 in the CONUS. The change of each metric in each grid between the two periods was calculated using Equation 4.7 based on the simulation results from the WRF-SMOKE-CMAQ-Pollen modeling system. Table 4.2 summarizes the regional average and standard deviation of the changes in each of the five metrics above for oak pollen. These regional averages and standard deviations were calculated using Equation 4.8.

As shown in both Figure 4.19 and Table 4.2, the response of oak pollen season to climate change varies in different regions. The mean and maximum hourly concentrations of oak pollen were estimated to increase in the Northwest, North East, West and Southwest climate regions during 2047-2050, but to decrease in other climate regions. In particular, the hourly mean concentration in the Northeast climate region was estimated to increase 57.3% on average. The oak pollen season during 2047-2050 was predicted to start early in the Northwest, Northeast, West and Southeast climate regions, but to start late in other climate regions. The oak pollen season length was estimated to be short in six of nine climate regions. The number of hours in which oak pollen concentrations exceed the allergy-threshold value was estimated to increase in seven out of nine climate regions. The increase of exceedance hours ranges from 2.3% in the south climate region, to 31.7% in Northwest climate region.

The late start date and shorter duration of oak pollen seasons in the central and southern CONUS are possibly caused by the predicted lower surface air temperatures in these areas (Figures 4.19 C and D). These results are consistent with the changes of average spring SAT as shown in Figure 4.15 A.

Table 4.2: Regional average and standard deviation of the changes in mean and maximum hourly concentrations, start date, season length and exceedance hours for oak (*Quercus*) pollen. These changes were calculated using the simulation data between periods of 2001-2004 and 2047-2050 in nine climate regions across the CONUS (mean ± standard deviation).

Region[a]	Mean Hourly (%)	Max Hourly (%)	Start Date (Days)	Season Length (Days)	Exceedance Hours[b] (%)
NW	28.2 ± 92.6	113.4 ± 194.2	-3.3 ± 2.6	0.3 ± 1.4	31.7 ± 122.9
WNC	-10.7 ± 27	-20.8 ± 34.4	0.3 ± 1.1	-2.4 ± 1	-11.4 ± 30.3
ENC	-19.5 ± 60.7	-11.9 ± 66.8	1.4 ± 2.1	-2.6 ± 1.2	-7 ± 66.6
NE	57.3 ± 169.4	23.3 ± 84.9	-1.8 ± 3.2	-0.1 ± 1.1	25.9 ± 66.7
W	20.9 ± 33.8	33.5 ± 85.2	-2.3 ± 3.4	-0.4 ± 1.3	15.4 ± 95.6
C	-20.8 ± 8.6	-30.8 ± 10.6	3 ± 1.4	-1.5 ± 1.4	14.7 ± 10.6
SW	1.6 ± 37.8	56.6 ± 404.6	0.7 ± 1.3	-1 ± 1.1	8.6 ± 47.1
SE	-2.2 ± 10.2	-9.4 ± 17.3	-0.4 ± 1.4	1 ± 1.4	15.4 ± 10.2
S	-8.8 ± 18.8	-16.5 ± 30.1	3 ± 1.6	0.1 ± 1	2.3 ± 31.5

[a] The nine climate regions: South (S), Southeast (SE), Southwest (SW), Central (C), West (W), Northeast (NE), East North Central (ENC), West North Central (WNC), and Northwest (NW).
[b] Threshold pollen concentration: 13 pollen grains/m^3.

Figure 4.20 presents the changes of the mean and maximum hourly concentrations, start date, season length and exceedance hours for ragweed pollen between the periods of 2001-2004 and 2047-2050 in the CONUS. Table 4.3 summarizes the regional average and standard deviation of the changes in each of the five metrics above for ragweed pollen. As shown in both Figure 4.20 and Table 4.3, the response of ragweed pollen

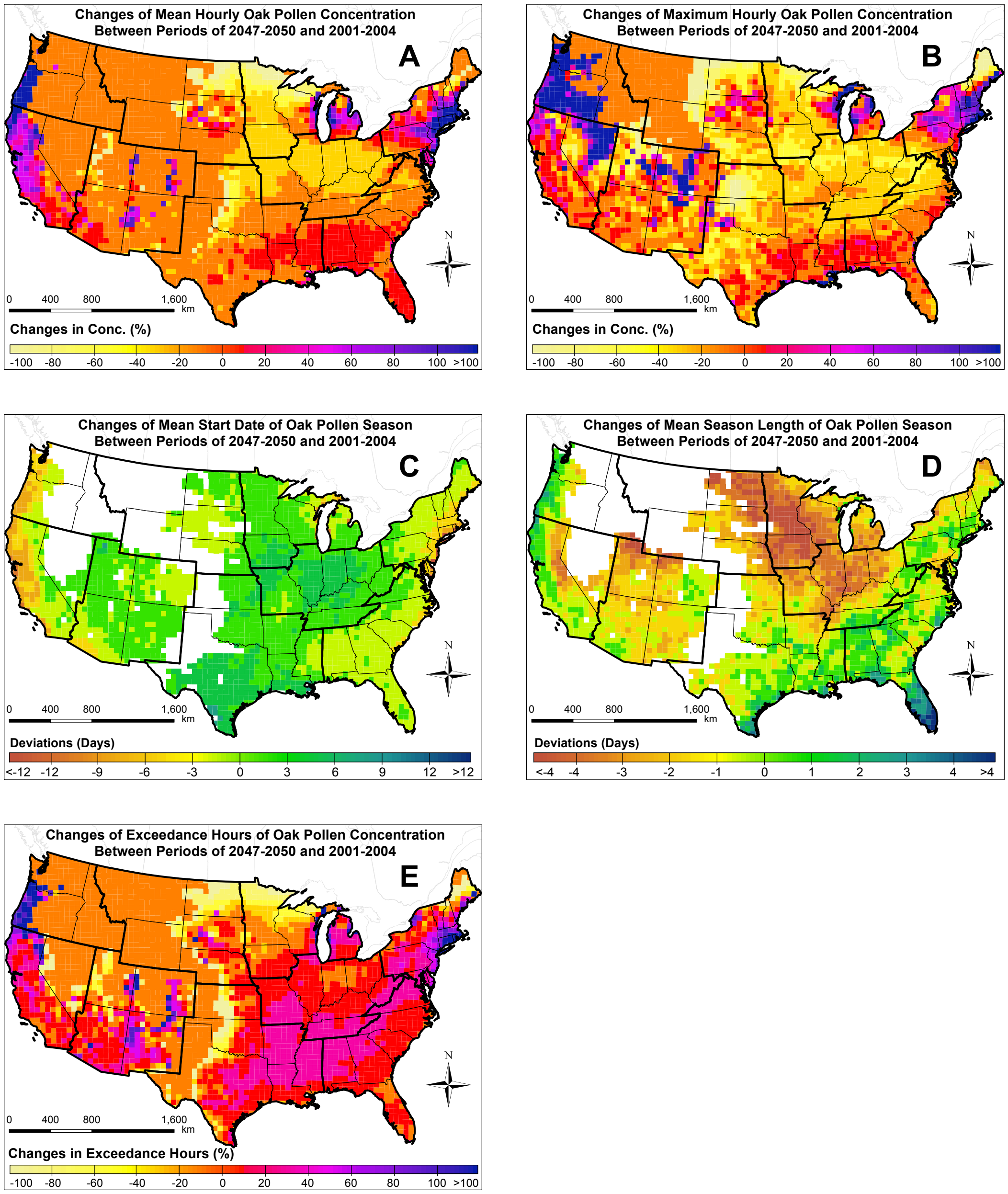

Figure 4.19: Changes of oak pollen season between periods of 2001-2004 and 2047-2050. (A) Mean hourly concentrations, (B) Maximum hourly concentrations, (C) Start Date, (D) Season Length, and (E) Exceedance hours

season to climate change also varies in different regions.

The mean hourly concentration of ragweed pollen was estimated to decrease in six out of nine climate regions during 2047-2050. However, the maximum hourly concentration was predicted to increase in five out of nine regions. In particular, the maximum hourly concentration was estimated to increase 34.7% on average in the East North Central climate region. The ragweed pollen season during 2047-2050 was predicted to start late in seven out of nine climate regions. The ragweed pollen season length was estimated to be short in all nine climate regions. The number of hours in which ragweed pollen concentrations exceed the allergy-threshold value was estimated to decrease in six out of nine climate regions.

Table 4.3: Regional average and standard deviation of the changes in mean and maximum hourly concentrations, start date, season length and exceedance hours for ragweed (*Ambrosia*) pollen. These changes were calculated using the simulation data between periods of 2001-2004 and 2047-2050 in nine climate regions across the CONUS (mean ± standard deviation).

Region[a]	Mean Hourly (%)	Max Hourly (%)	Start Date (Days)	Season Length (Days)	Exceedance Hours[b] (%)
NW	-8.7 ± 26.2	-10.6 ± 41	1.2 ± 1.3	-1.1 ± 1	-5.5 ± 25.7
WNC	0.8 ± 22.2	0.9 ± 45.3	1.7 ± 1	-2 ± 0.9	3.5 ± 47.1
ENC	-3.4 ± 16.5	34.7 ± 98.3	1.8 ± 1.2	-3.7 ± 0.9	-3 ± 17.3
NE	-9.3 ± 43.8	-19.3 ± 74.7	-0.9 ± 1.5	-0.9 ± 0.8	-6.2 ± 111.8
W	-36.4 ± 19.1	-30.5 ± 23	0.5 ± 1.2	-0.3 ± 1	-11 ± 25.7
C	2 ± 30.7	13.8 ± 91	1 ± 1.1	-3.8 ± 1	4.1 ± 81.1
SW	15.2 ± 39.7	13.1 ± 78	0.1 ± 1.1	-0.7 ± 0.8	6.2 ± 42.7
SE	-9.6 ± 34.3	-23 ± 59.6	-1.1 ± 0.7	-0.2 ± 0.5	-8.6 ± 46.2
S	-1.8 ± 30.3	17.7 ± 128.7	0.3 ± 1	-1.4 ± 0.7	-2.8 ± 28.9

[a] The nine climate regions: South (S), Southeast (SE), Southwest (SW), Central (C), West (W), Northeast (NE), East North Central (ENC), West North Central (WNC), and Northwest (NW).
[b] Threshold pollen concentration: 30 pollen grains/m^3.

Figure 4.21 presents the changes of the mean and maximum hourly concentrations, start date, season length and exceedance hours for birch pollen between periods of 2001-2004 and 2047-2050 in the CONUS. Table 4.4 summarizes the regional average and standard deviation of the changes in each of the five metrics above for birch pollen. As shown in both Figure 4.21 and Table 4.4, the response of birch pollen season to climate change also varies in different regions.

The mean and maximum hourly concentrations of birch pollen were estimated to decrease in the majority of climate regions during 2047-2050. The birch pollen season during 2047-2050 was predicted to start early in the Northwest, Northeast and West climate regions, but to start late in other climate regions. The season length was

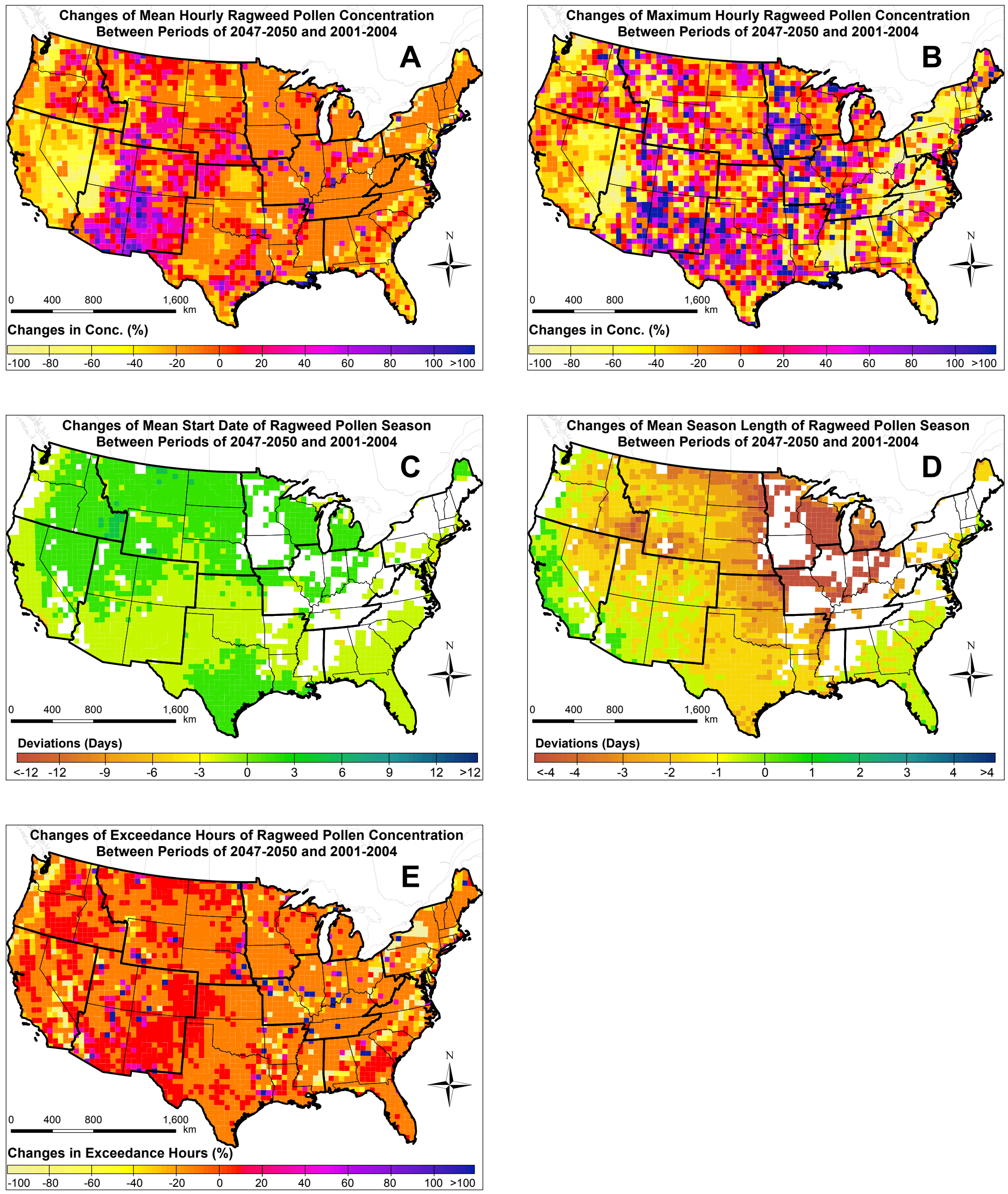

Figure 4.20: Changes of ragweed pollen season between periods of 2001-2004 and 2047-2050. (A) Mean hourly concentrations, (B) Maximum hourly concentrations, (C) Start Date, (D) Season Length, and (E) Exceedance hours

estimated to be short in eight out of nine climate regions. The exceedance hours were predicted to increase in the Northwest, West North Central, and Northeast climate regions; but to decrease in the East North Central and Central climate regions.

Figure 4.22 presents the changes of the mean and maximum hourly concentrations,

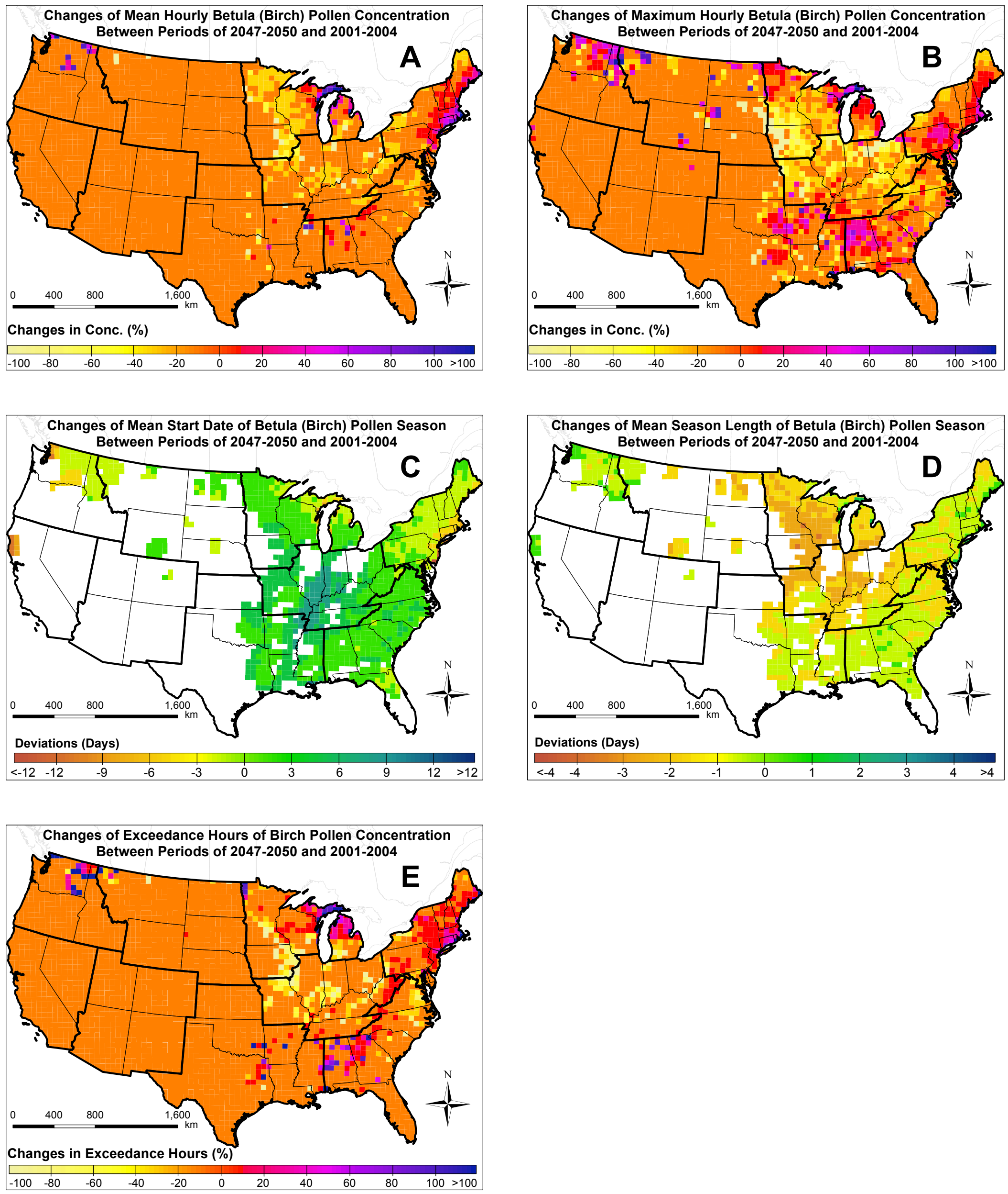

Figure 4.21: Changes of birch pollen season between periods of 2001-2004 and 2047-2050. (A) Mean hourly concentrations, (B) Maximum hourly concentrations, (C) Start Date, (D) Season Length, and (E) Exceedance hours.

start date, season length and exceedance hours for mugwort pollen between periods of 2001-2004 and 2047-2050 in the CONUS. Table 4.5 summarizes the regional average and standard deviation of the changes in each of the five metrics above for mugwort pollen. As shown in both Figure 4.22 and Table 4.5, the response of mugwort pollen

Table 4.4: Regional average and standard deviation of the changes in mean and maximum hourly concentrations, start date, season length and exceedance hours for birch (*Betula*) pollen. These changes were calculated using the simulation data between periods of 2001-2004 and 2047-2050 in nine climate regions across the CONUS (mean ± standard deviation).

Region[a]	Mean Hourly (%)	Max Hourly (%)	Start Date (Days)	Season Length (Days)	Exceedance Hours[b] (%)
NW	3.5 ± 18.1	4.4 ± 23.4	-2.1 ± 2	0.2 ± 0.5	8.3 ± 37.8
WNC	-0.7 ± 8.6	-0.9 ± 26.9	0.3 ± 1.2	-0.7 ± 0.7	0.6 ± 22.7
ENC	-7.5 ± 37.7	-15.7 ± 34.2	1.5 ± 2.1	-1.4 ± 0.8	-5 ± 42.2
NE	1.6 ± 25.9	-1.6 ± 19.1	-0.9 ± 2.3	-0.3 ± 0.7	3 ± 31.8
W	0 ± 0	0.2 ± 2.4	-8.7 ± 1.4	0.9 ± 0.3	0 ± 0
C	-11 ± 21.2	-21.6 ± 24.1	4.6 ± 2	-1.1 ± 0.7	-14.1 ± 27.6
SW	0 ± 0	0.1 ± 1.8	0.2 ± 0.8	-0.2 ± 0.4	0 ± 0
SE	-2 ± 12.1	-1.7 ± 24.6	2.2 ± 1.2	-0.2 ± 0.5	0.7 ± 29.4
S	-0.7 ± 11.1	-1.1 ± 19.7	4 ± 1.4	-0.3 ± 0.5	2.6 ± 31.5

[a] The nine climate regions: South (S), Southeast (SE), Southwest (SW), Central (C), West (W), Northeast (NE), East North Central (ENC), West North Central (WNC), and Northwest (NW).
[b] Threshold pollen concentration: 20 pollen grains/m^3.

season to climate change also varies in different regions.

The mean and maximum hourly concentrations of mugwort pollen were estimated to decrease in the majority of climate regions during 2047-2050. The mugwort pollen season during 2047-2050 was predicted to start early in eight out the nine climate regions. The season length was estimated to be short in six out of nine climate regions. The exceedance hours were predicted to increase in the West, Southwest, Southeast and South climate regions; but to decrease in other climate regions.

Figure 4.23 presents the changes of the mean and maximum hourly concentrations, start date, season length and exceedance hours for grass pollen between periods of 2001-2004 and 2047-2050 in the CONUS. Table 4.6 summarizes the regional average and standard deviation of the changes in each of the five metrics above for grass pollen. As shown in both Figure 4.23 and Table 4.6, the response of grass pollen season to climate change also varies in different regions.

The mean and maximum hourly concentrations of grass pollen were estimated to decrease in most of the climate regions during 2047-2050. The grass pollen season during 2047-2050 was predicted to start late in seven out nine climate regions. The

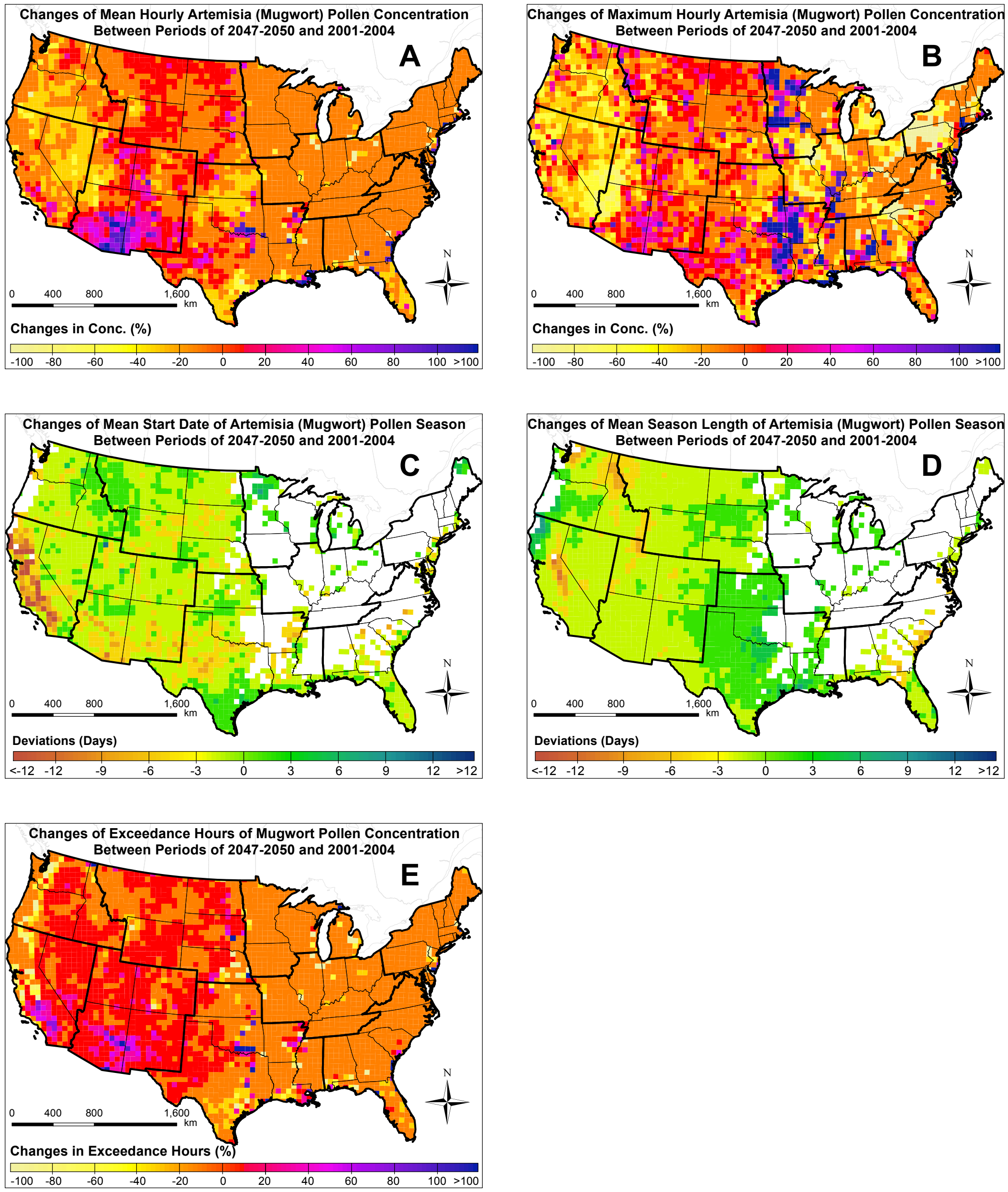

Figure 4.22: Changes of mugwort pollen season between periods of 2001-2004 and 2047-2050. (A) Mean hourly concentrations, (B) Maximum hourly concentrations, (C) Start Date, (D) Season Length, and (E) Exceedance hours.

season length was estimated to be long in all of the nine climate regions. The exceedance hours were predicted to decrease in eight out of nine climate regions.

It should be noted that the regional average of each of the five metrics used above is generally noticeably lower than its corresponding standard deviations in Tables 4.2-4.6.

Table 4.5: Regional average and standard deviation of the changes in mean and maximum hourly concentrations, start date, season length and exceedance hours for mugwort (*Artemisia*) pollen. These changes were calculated using the simulation data between periods of 2001-2004 and 2047-2050 in nine climate regions across the CONUS (mean ± standard deviation).

Region[a]	Mean Hourly (%)	Max Hourly (%)	Start Date (Days)	Season Length (Days)	Exceedance Hours[b] (%)
NW	-12 ± 15.2	-17.6 ± 24.9	-0.5 ± 1.6	-1.1 ± 2.5	-4.5 ± 26.6
WNC	0.3 ± 10.9	-1 ± 20.4	-0.6 ± 1.5	-0.3 ± 1.3	0 ± 20.8
ENC	-1 ± 9.4	15.7 ± 69.7	2 ± 2.1	1.1 ± 0.9	-2.4 ± 15.3
NE	0.9 ± 24.3	-27.3 ± 56.1	-0.1 ± 3.4	-0.5 ± 1.3	-1.6 ± 23.9
W	-15.5 ± 17	-25.2 ± 24.5	-3.4 ± 4.1	-0.9 ± 2.5	1 ± 26.4
C	-3.6 ± 15.7	-13.9 ± 43.5	-0.3 ± 1.7	1.8 ± 1.5	-3.3 ± 15.5
SW	14 ± 33.9	-3.7 ± 28.2	-1.5 ± 2	-0.6 ± 1.2	11 ± 16.9
SE	-0.8 ± 26.9	0.5 ± 68.9	-0.8 ± 1.9	-1.7 ± 2.3	8.1 ± 173.3
S	-2.4 ± 29.7	16.7 ± 63	-1 ± 2.2	1.8 ± 1.4	5.3 ± 104.6

[a] The nine climate regions: South (S), Southeast (SE), Southwest (SW), Central (C), West (W), Northeast (NE), East North Central (ENC), West North Central (WNC), and Northwest (NW).
[b] Threshold pollen concentration: 30 pollen grains/m^3.

Table 4.6: Regional average and standard deviation of the changes in mean and maximum hourly concentrations, start date, season length and exceedance hours for grass (*Poaceae*) pollen. These changes were calculated using the simulation data between periods of 2001-2004 and 2047-2050 in nine climate regions across the CONUS (mean ± standard deviation).

Region[a]	Mean Hourly (%)	Max Hourly (%)	Start Date (Days)	Season Length (Days)	Exceedance Hours[b] (%)
NW	-23.9 ± 17	-16.9 ± 17.5	3.3 ± 3	4.6 ± 3.4	-19 ± 20.5
WNC	-26.2 ± 16.8	-12.8 ± 16	4.3 ± 2.5	3.7 ± 2.4	-18.2 ± 24.6
ENC	-7.9 ± 24.7	0.3 ± 49.9	5 ± 3.2	3.8 ± 5.8	-3 ± 16.7
NE	-3.2 ± 15.2	-30.6 ± 42.6	-4.3 ± 4.2	12.2 ± 12.6	-0.9 ± 9.6
W	-10.1 ± 23	-11.6 ± 22	2.8 ± 3	0.2 ± 4.4	-6.1 ± 32.2
C	-4.1 ± 16.2	-3.9 ± 49.1	3 ± 2.1	5.8 ± 3.6	-3.2 ± 14
SW	-4.4 ± 15.4	-6.3 ± 15.6	0.1 ± 2.7	3.7 ± 2.6	-0.4 ± 20
SE	-4 ± 15.5	0.7 ± 36.9	-2.5 ± 1.7	6 ± 4.8	-1 ± 19.6
S	-8.4 ± 18.3	-5.5 ± 22.3	1 ± 2.4	3.8 ± 7.2	0.5 ± 53.2

[a] The nine climate regions: South (S), Southeast (SE), Southwest (SW), Central (C), West (W), Northeast (NE), East North Central (ENC), West North Central (WNC), and Northwest (NW).
[b] Threshold pollen concentration: 20 pollen grains/m^3.

This indicates high uncertainties in the predicted regional averages of the changes of allergenic pollen season between periods of 2001-2004 and 2047-2050. The uncertainties may be from the high heterogeneity of pollen distribution within a climate region, or from the modeling errors.

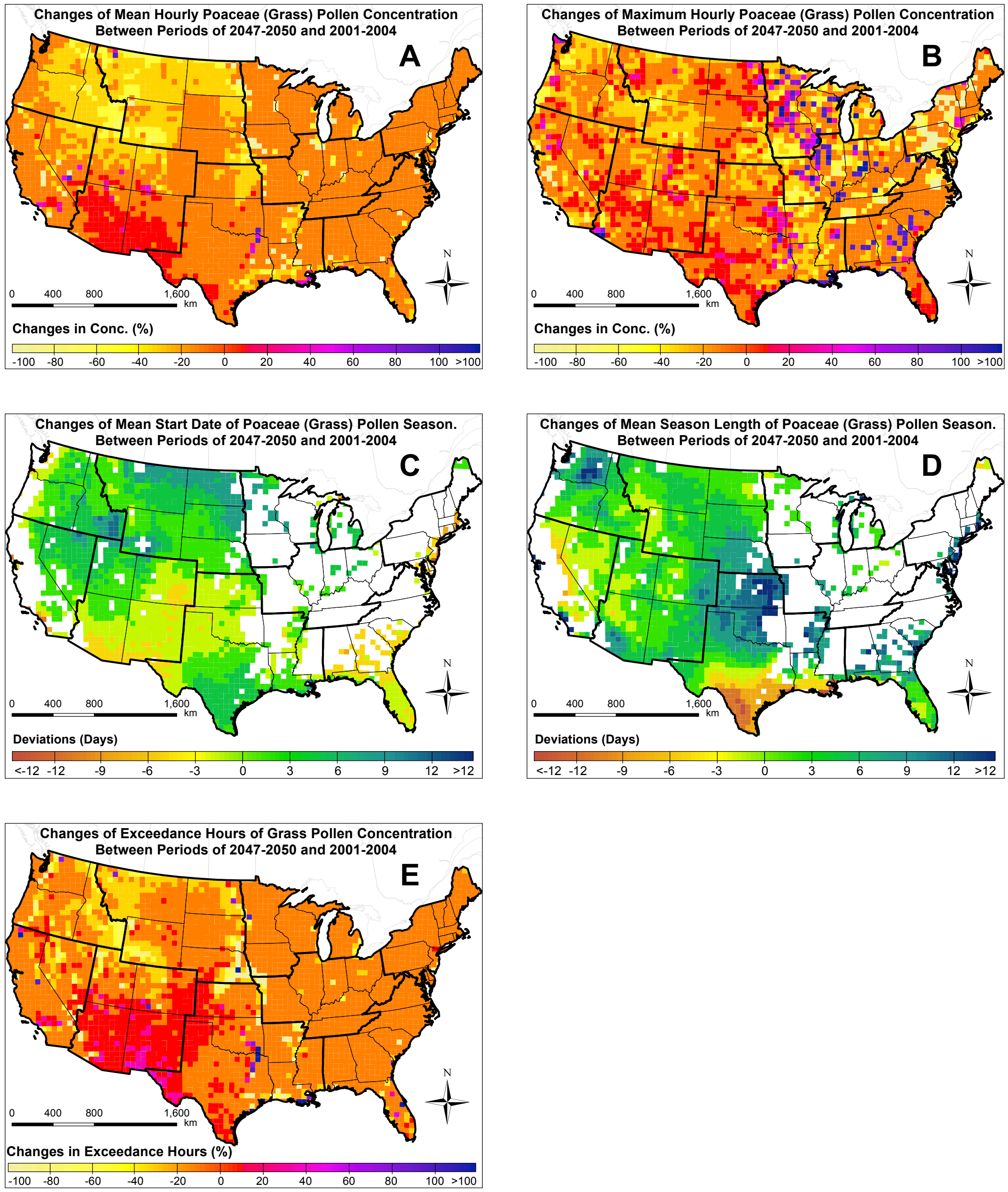

Figure 4.23: Changes of grass pollen season between periods of 2001-2004 and 2047-2050. (A) Mean hourly concentrations, (B) Maximum hourly concentrations, (C) Start Date, (D) Season Length, and (E) Exceedance hours.

It should be emphasized that the allergenic pollen season variation simulated in the current study is just one of many possible realizations of future scenarios. There are substantial uncertainties existing in the simulations results. First, simulation of only four years in each period cannot capture the internal variation of climate and meteorology[136]. Second, the input meteorology in the current study was from the output of only one regional meteorology model and one global climate model. This single RCM or GCM cannot capture the full atmospheric physics and external driving forces.

Furthermore, in the current study, the vegetation coverage and the intrinsic annual emission flux (q_e) were assumed unchanged between periods of 2001-2004 and 2047-2050. The WRF-CMAQ-SMOKE-Pollen modeling system only accounted for meteorology influence on the spatiotemporal distributions of allergenic pollen. These factors should be taken into consideration when interpreting and using the predictions from the WRF-CMAQ-SMOKE-Pollen modeling system. The uncertainties in the deterministic modeling system are discussed in section 4.3.5.

4.3.5 Uncertainty in the WRF-SMOKE-CMAQ-Pollen modeling system

Uncertainties generally pervade in the entire modeling process[137]. They may result from different components and modules of the modeling framework[138]. There are substantial uncertainties in each of the components and modules of the WRF-SMOKE-CMAQ-Pollen modeling system. For each of the model components and its modules, the uncertainty has mainly resulted from the model formulations, parameters and the input data. Table 4.7 presents the sources of uncertainties for each component and module in the WRF-SMOKE-CMAQ-Pollen modeling system, and the corresponding general treatments.

In the current study, great effort has been made to identify and reduce the uncertainties in each of the model components and modules based on different methods. For the climate model CCSM3 in NARCCAP, there are one ensemble member for historical control years and one member for future years[55,56]. The quality of the CCSM3 data have been evaluated by NARCCAP to ensure the consistency between simulations and the observed long term climate trends.

For the meteorology simulations from the WRF model, quality control measures have been applied by NARCCAP to guarantee the quality of the archived meteorology data[55,56]. Similarly, the observed meteorology factors from the NOAA stations have

Table 4.7: Sources of uncertainties for each component and module in the the deterministic modeling system, and the corresponding general treatments.

Uncertainty Source		Model Formulations	Model Parameters	Input Data	Notes
Emission Model	Flowering likelihood	Relevant	Relevant	Relevant	Developed in the current study
	Vegetation coverage	Relevant	Relevant	Relevant	
	Start date and season length	Relevant	Relevant	Relevant	
	Meteorology adjustment factors	Relevant	Relevant	Not relevant	
Transport Model	Advection	Relevant	Relevant	Not relevant	Adapted from CMAQ4.7.1
	Diffusion	Relevant	Relevant	Not relevant	
	Cloud process	Relevant	Relevant	Not relevant	
	Dry deposition	Relevant	Relevant	Not relevant	
	Initial, Boundary conditions	Not relevant	Not relevant	Relevant	
Meteorology Model	Microphysics	Relevant	Relevant	Relevant	NARCCAP archived WRF2.1.1 output; NJDEP WRF3.1.1 output
	Long and short wave radiation	Relevant	Relevant	Relevant	
	PBL scheme, land scheme, etc.	Relevant	Relevant	Relevant	
	Initial, Boundary conditions	Not relevant	Not relevant	Relevant	
	Ensemble members	Relevant	Relevant	Relevant	
Climate Model	Atmosphere physics	Relevant	Relevant	Relevant	NARCCAP archived CCSM3 output
	Ocean circulation	Relevant	Relevant	Relevant	
	Phases of ENSO and AMO	Relevant	Relevant	Relevant	
	Ensemble members	Relevant	Relevant	Relevant	
General Treatment		New model methods or modified formulations	Global uncertainty, sensitivity analyses	Statistical techniques to clean data	

also been checked rigorously to maintain the high quality. The observed pollen counts from NAB-AAAAI stations were examined carefully according to rules set in section 2.2.2 to ensure data quality [45].

For the developed pollen emission model, global sensitivity analysis was conducted to identify sensitive and interactive input parameters based on Morris' design [114]. The values of highly sensitive and interactive parameters were generally carefully picked from the literature, or parameterized using literature data. Many iterations of the emission model have been tried to ensure the consistency and quality of the simulated pollen emission data.

For the pollen transport model, process analysis has been conducted to identify the contributions of each physical process on the airborne pollen concentrations (section 4.2.5). Different boundary conditions were applied to investigate BC's influence on airborne pollen concentrations.

4.4 Summary

A modeling system incorporating a meteorology model (WRF), a pollutant emission model (SMOKE), a pollen emission model and an air quality model (CMAQ) have been developed to simulate the spatiotemporal distributions of allergenic pollen and anthropogenic air pollutants. This WRF-SMOKE-CMAQ-Pollen modeling system has been applied to a domain covering the contiguous United States during the periods of 2001-2004 and 2047-2050 to investigate the climate change impact on allergenic pollen from representative trees, weeds and grass.

The performance of the WRF-SMOKE-CMAQ-Pollen modeling system was evaluated operationally through calculation of hit and false rates, fractional bias, and deviations between observed and simulated start dates and season length based on the observed pollen count at monitor stations across the CONUS in 2004. For oak and ragweed pollen, the hit rates for three ambient pollen levels were generally between 70% and 100% at the majority of studied stations; the false rate were generally between 0% and 10%; the fractional bias were between -0.4 and 0.4; the deviation of observed and simulated start dates and season length were generally between 0 to 6 days at the majority of studied stations. This indicated that the WRF-SMOKE-CMAQ-Pollen modeling system correctly predicted the observed pollen season start date and duration, and airborne pollen levels at the majority of monitor stations. The WRF-SMOKE-CMAQ-Pollen modeling system could capture the variations in start date, season length and airborne levels of birch, mugwort and grass pollen. However it did not perform as well as for oak and ragweed pollen.

The modeling system was also evaluated diagnostically through process analysis and boundary condition analysis. The dry deposition, emission and vertical eddy diffusion were the dominant processes determining ambient pollen concentrations. The boundary condition exerted less influence on mean pollen concentrations, but significant influence on maximum pollen concentrations in some northern states bordering Canada.

The response of allergenic pollen season to climate change varies in different climate region for different taxa. For ragweed, mugwort and grass, the regional average pollen concentrations were predicted to decrease in the majority of climate regions during the period of 2047-2050. For oak and birch, although there were not remarkable increases of airborne pollen concentrations during the period of 2047-2050; the number of hours in which pollen concentrations exceed the threshold values for triggering allergy, has been predicted to increase in the majority of climate regions.

Chapter 5
EXPOSURES TO AIRBORNE ALLERGENIC POLLEN

5.1 Introduction

This chapter examines population exposures to allergenic pollen and their spatiotemporal patterns in nine climate regions of the CONUS. These nine climate regions are defined by NOAA based on long term observations of precipitation and temperature, and consist of South (S), Southeast (SE), Southwest (SW), Central (C), West (W), Northeast (NE), East North Central (ENC), West North Central (WNC), and Northwest (NW) (Figure 2.2). Exposures to allergenic pollen in the nine climate regions during the 1990s (1994-2000) and the past decade (2001-2010) were calculated based on probabilistic methods by sampling from observed pollen concentrations, demographics, time spent indoors and outdoors, and inhalation rates for different activity levels in indoor and outdoor environments (Figure 5.1).

5.2 Methods

Figure 5.1 presents a diagram of the pollen exposure modeling system. A "virtual subject" is randomly assigned an age and gender according to the demographic data relevant to a given period and climate region. The pollen concentrations to which the individual is exposed depend on the day of year, period and climate region. His/her time spent in indoor and outdoor environments on a given day in a given climate region were simulated according to the Consolidated Human Activity Database (CHAD)[3]. Exposure factors, such as inhalation rates and exposed human skin surface area, were derived from the EPA Exposure Factor Handbook[4]. The methods applied for each component are described in the following subsections.

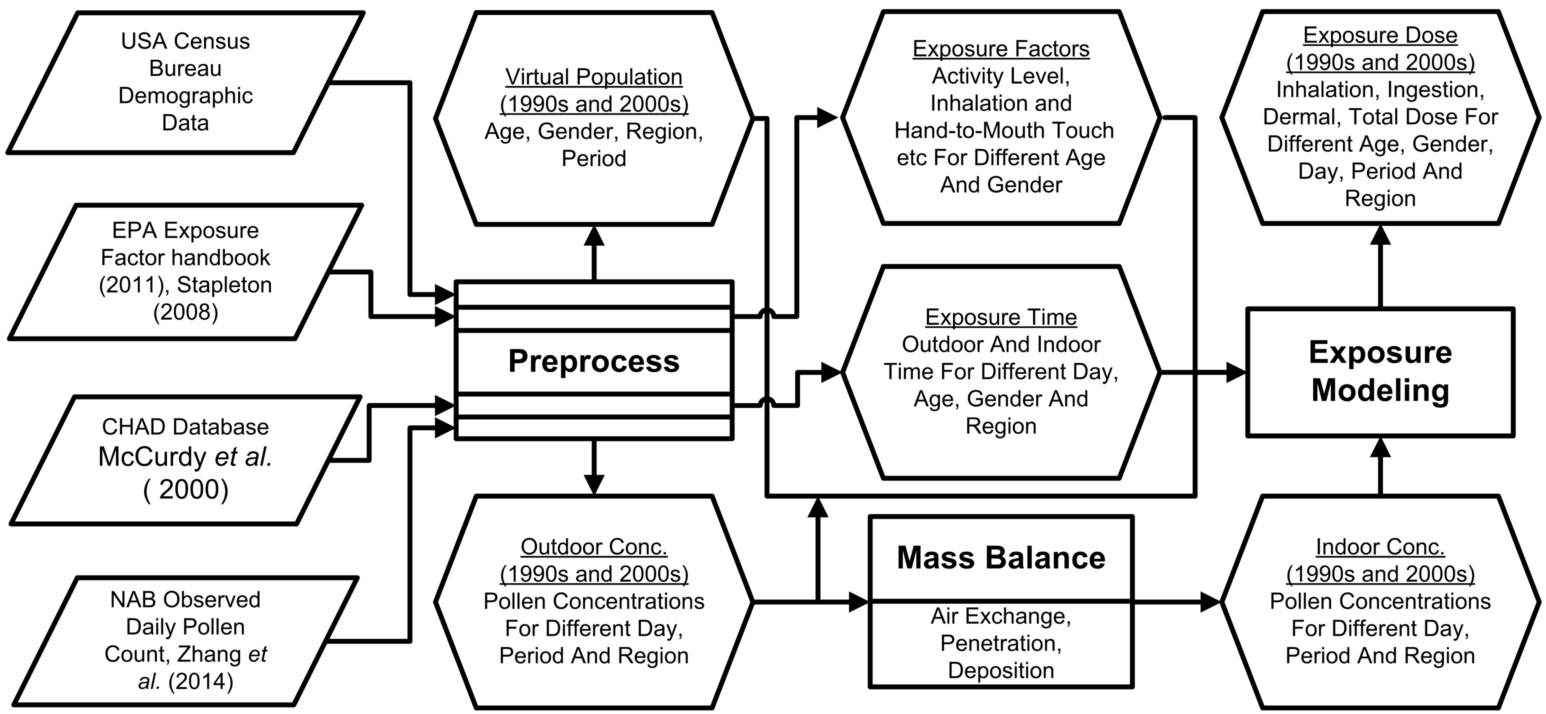

Figure 5.1: Schematic representation of exposure modeling system. Exposures to allergenic pollen in nine climate regions during periods of 1994-2000 and 2001-2010, were calculated based on probabilistic methods by sampling from observed pollen concentrations, demographics, exposure time and inhalation rate for different activity levels in indoor and outdoor environments.

5.2.1 Virtual population

For simulation of population exposure to allergenic pollen, 3,000 "virtual subjects" were generated for each of the nine CONUS climate regions and for each period considered, based on demographic data from the US Census Bureau[139]. The size of the "virtual population" was selected on the basis of preliminary studies, which indicated that 3,000 "virtual subjects" were sufficient to generate the statistics of population exposure to allergenic pollen (Figure E.1 in the Appendix E). The "virtual population" for the periods of 1994-2000 and 2001-2010 were sampled from demographic data for 2000 and 2010, respectively. A "virtual subject" during period j in climate region k $s(a, g; j, k)$ was assigned an age a and gender g by randomly sampling from demographic data $S(j, k)$ in the corresponding climate region and period (equation 5.1).

$$s(a, g; j, k) \sim S(j, k) \tag{5.1}$$

5.2.2 Exposure concentrations of airborne pollen

Outdoor daily pollen concentrations $c_{out}(i, j, k)$ (pollen grain/m^3) for a "virtual subject" on day i during period j in climate region k were sampled from the collection of observed daily pollen counts $C_{out}(i, j, k)$ on day i during period j from monitoring stations in climate region k (equation 5.2).

$$c_{out}(i, j, k) \sim C_{out}(i, j, k) \tag{5.2}$$

Indoor daily pollen concentrations $c_{in}(i, j, k)$ (pollen grain/m^3) for the same person were derived from outdoor daily pollen concentration $c_{out}(i, j, k)$ based on a mass balance[140] equation 5.3,

$$c_{in}(i, j, k) = \frac{P_f(i, k) \times \mathrm{ACH(i, k)}}{\mathrm{ACH(i, k)} + v_{dr}(i, k)} c_{out}(i, j, k) \tag{5.3}$$

where P_f (dimensionless) is the penetration factor, ACH (h^{-1}) the air exchange rate, v_{dr} (h^{-1}) the deposition rate. The distributions of P_f, ACH and v_{dr} vary in different climate regions and seasons for pollen grains of different taxa. In the current study, these three parameters were lumped into one parameter as a ratio of indoor to outdoor pollen concentrations (Equation 5.4).

$$r_{IO}(i, k) = \frac{P_f(i, k) \times \mathrm{ACH(i, k)}}{\mathrm{ACH(i, k)} + v_{dr}(i, k)} \tag{5.4}$$

The ratio of indoor to outdoor pollen concentrations for species Sp on day i in climate region j ($r_{IO}(i,k;Sp)$) was uniformly sampled from a range ($R_{IO}(k;Sp)$) as presented in Equation 5.5,

$$\begin{cases} r_{IO}(i,k;Sp) \sim R_{IO}(k;Sp) \\ R_{IO}(k;Sp) = [R_{IO}^{L}(k;Sp), \quad R_{IO}^{U}(k;Sp)] \end{cases} \tag{5.5}$$

where $R_{IO}^{L}(k;Sp)$ and $R_{IO}^{U}(k;Sp)$ are lower and upper boundaries of the uniform distribution of ratio, respectively.

The upper and lower boundaries of the ratio of indoor to outdoor pollen concentrations were further parameterized using Equation 5.6,

$$\begin{cases} R_{IO}^{U}(k;Sp) = C_{Sp} R_{IO}(k;PM_{10}) \\ R_{IO}^{L}(k;Sp) = \frac{C_{Sp} R_{IO}^{Min}}{C_{Oak}} \end{cases} \tag{5.6}$$

where $R_{IO}(k;PM_{10})$ is the ratio of indoor to outdoor PM_{10} concentrations at climate region k[141]. R_{IO}^{Min} is the observed minimum ratio of indoor to outdoor pollen concentrations[142]. C_{Sp} is the species specific constant to adjust the $R_{IO}(k;PM_{10})$ for pollen grains from different species. In general, higher density and larger diameter of pollen grains lead to a relatively lower C_{Sp}. In the current study, C_{Sp} was assumed to be 0.65, 0.60, 0.75, 0.70 and 0.50 for birch, oak, ragweed, mugwort and grass pollen, respectively.

The upper boundary of the ratio of indoor to outdoor pollen concentration ($R_{IO}^{U}(k;Sp)$) was derived using the data on PM_{10} reported in the literature. Chen *et al.* investigated indoor exposures to PM_{10} (particulate matter with diameter smaller than 10 μm) in seven different regions in the CONUS[141]. They reported region specific ratios for calculating indoor PM_{10} concentrations from the corresponding outdoor PM_{10} concentrations. For a given region, the ratio of indoor PM_{10} to outdoor PM_{10} took a fixed value by incorporation of information of residence characteristics. This information includes: (1) whether the house has central air conditioning (2) how long the central AC is operated, and (3) how long the windows are open or closed. These region specific PM_{10} ratios were used to derive the upper boundaries of region specific ratios of indoor to outdoor pollen concentrations (Equation 5.6).

The lower boundary of the ratio of indoor to outdoor pollen concentrations ($R^L_{IO}(k;Sp)$) was derived based on the minimum ratio of observed indoor to outdoor pollen concentrations. Spengler *et al.* reviewed studies regarding the ratio of indoor to outdoor pollen concentrations[142]. They reported that under some conditions (e.g., window and doors closed through a day), the indoor pollen concentrations could be few hundreds times lower than the corresponding outdoor concentrations. In the current study, the minimum ratio R^{Min}_{IO} was assumed to be 0.005 for oak pollen based on the data and descriptions from the literature[142,143]. The derived ranges of $R_{IO}(k;Sp)$ are listed in Table 5.1 for nine climate regions and five allergenic taxa. These ranges are consistent with the observations in Sofia, Bulgaria[143].

Table 5.1: Ratios of indoor to outdoor pollen concentrations ($R_{IO}(k;Sp)$) for nine climate regions and five allergenic taxa in the CONUS.

Region	Birch (%)	Oak (%)	Ragweed (%)	Mugwort (%)	Grass (%)
NW	0.54-9.10	0.50-8.40	0.63-11.25	0.58-10.50	0.42-8.50
WNC	0.54-12.35	0.50-11.40	0.63-14.25	0.58-13.30	0.42-10.50
ENC	0.54-16.25	0.50-15.00	0.63-19.5	0.58-18.20	0.42-13.50
NE	0.54-15.60	0.50-14.40	0.63-18.75	0.58-17.50	0.42-13.50
W	0.54-14.30	0.50-13.20	0.63-18.00	0.58-16.80	0.42-11.50
C	0.54-15.60	0.50-14.40	0.63-18.00	0.58-16.80	0.42-12.50
SW	0.54-10.40	0.50-9.60	0.63-12.00	0.58-11.20	0.42-7.50
SE	0.54-14.30	0.50-13.20	0.63-16.50	0.58-15.40	0.42-11.00
S	0.54-12.35	0.50-11.40	0.63-14.25	0.58-13.30	0.42-9.50

The observed pollen counts at each station are resolved daily, representing the average airborne pollen concentrations within 24 hours at the corresponding monitoring station. Hence, the simulated daily concentrations c_{out} and c_{in} represent the average airborne pollen concentrations within 24 hours of a calendar day in outdoor and indoor environments, respectively.

5.2.3 Exposure time

Outdoor and indoor exposure times per day for a "virtual subject" depend on the gender, age, climate region and day of the year. The outdoor exposure time $t_{out}(i,k)$ (hours) for a "virtual subject" on day i in climate regions k was sampled from the collection of corresponding observed outdoor exposure times $T_{out}(i,k;a,g)$ as shown in equation 5.7.

$$t_{out}(i,k) \sim T_{out}(i,k;a,g) \tag{5.7}$$

The indoor exposure time $t_{in}(i,k)$ (hours) was then calculated using equation 5.8.

$$t_{in}(i,k) = 24 - t_{out}(i,k) \tag{5.8}$$

5.2.4 Exposure factors

Exposure factors include inhalation rate, body weight, exposure body surface area and hand-to-mouth touch frequency, etc. Inhalation rate is described here while other exposure factors are described in section 5.2.5 and listed in table 5.2. Indoor inhalation rate $IH_{in}(a,g)$ (m^3/(hr kg BW)) for a "virtual subject" depends on age a, gender g and activity level as presented in equation 5.9,

$$\begin{aligned} IH_{in}(a,g) = IH(a,g,L_P)f_{Pin} + IH(a,g,L_L)f_{Lin} + \\ IH(a,g,L_M)f_{Min} + IH(a,g,L_H)f_{Hin} \end{aligned} \tag{5.9}$$

where L_P, L_L, L_M and L_H indicate passive, low, moderate and high activity levels, respectively; and f_{Pin}, f_{Lin}, f_{Min} and f_{Hin} are fractions of time spent in indoor environments at passive, low, moderate and high activity levels, respectively. The inhalation rate for any given age, gender and activity level (e.g., $IH(a,g,L_P)$), and the fraction of time spent at the corresponding activity level were derived based on data from the Exposure Factors Handbook [4].

Similarly, the outdoor inhalation rate can be calculated using equation 5.10,

$$\begin{aligned} IH_{out}(a,g) = IH(a,g,L_P)f_{Pout} + IH(a,g,L_L)f_{Lout} + \\ IH(a,g,L_M)f_{Mout} + IH(a,g,L_H)f_{Hout} \end{aligned} \tag{5.10}$$

where f_{Pout}, f_{Lout}, f_{Mout} and f_{Hout} are fractions of time spent in outdoor environments at passive, low, moderate and high activity levels, respectively.

5.2.5 Exposures to allergenic pollen

Equation 5.11 calculates the aggregated exposure to allergenic pollen for a given "virtual subject" with age a and gender g. The aggregated exposure $E_A(\Delta t_E; j,k,a,g)$ (pollen grains/(day kg BW)) during time interval Δt_E in climate region k and period j is the sum of inhalation $E_{inha}(\Delta t_E; j, k, a, g)$, deposition on exposed skin $E_{derm}(\Delta t_E; j, k, a, g)$, and unintentional ingestion $E_{inge}(\Delta t_E; j, k, a, g)$ through hand-to-mouth transfer.

$$E_A(\Delta t_E; j, k, a, g) = E_{inha}(\Delta t_E; j, k, a, g) + E_{derm}(\Delta t_E; j, k, a, g) + E_{inge}(\Delta t_E; j, k, a, g) \tag{5.11}$$

The inhalation exposures to allergenic pollen were obtained through equation 5.12:

$$E_{inha}(\Delta t_E; j, k, a, g) = \int_t^{t+\Delta t_E} [\underbrace{IH_{in}(t) c_{in}(t) I_{in}(t)}_{\text{Indoor}} + \underbrace{IH_{out}(t) c_{out}(t) I_{out}(t)}_{\text{Outdoor}}] dt \tag{5.12}$$

The first and second term on the right hand side of equation 5.12 represents indoor and outdoor exposures, respectively. The inhalation rates in equations 5.9 and 5.10 have been normalized by body weight BW, and so have inhalation exposures in equation 5.12. The $I_{in}(t)$ and $I_{out}(t)$ are indicator functions, which take either 1 or 0 as value as shown in equation 5.13,

$$I_{in}(t) = \begin{cases} 1, & t \in t_{in} \\ 0, & t \notin t_{in} \end{cases} \qquad I_{out}(t) = \begin{cases} 1, & t \in t_{out} \\ 0, & t \notin t_{out} \end{cases} \tag{5.13}$$

where $I_{in}(t)$ takes 1 as its value when the time is spent in an indoor environment, and 0 as its value when in an outdoor environment; likewise for $I_{out}(t)$.

The dermal exposure to allergenic pollen was obtained through equation 5.14,

$$E_{derm}(\Delta t_E; j, k, a, g) = \frac{1}{BW} \int_t^{t+\Delta t_E} TE_{DS} S_{hum} [\underbrace{v_d c_{in}(t) I_{in(t)}}_{\text{Indoor}} + \underbrace{v_d c_{out}(t) I_{out}(t)}_{\text{Outdoor}}] dt \tag{5.14}$$

where TE_{DS} is the adhesion efficiency on skin (%), v_d is deposition velocity (m/h), and BW is body weight (kg). BW of a "virtual subject" with age a and gender g ($BW(a,g)$) was uniformly sampled from a collection of body weights for the corresponding gender and age groups ($BW_C(a, g)$) based on data from the Exposure Factors Handbook[4] (Equation 5.15).

$$BW(a, g) \sim BW_C(a, g) \tag{5.15}$$

S_{hum} is the exposed human skin surface area, which can be derived from equation 5.16,

$$S_{hum} = BW \times F_{BS} P_{EX} \tag{5.16}$$

where F_{BS} is the ratio of the skin surface area to weight (m^2/kg) of the human body, and P_{EX} is the percentage of exposed skin surface are.

The unintentional ingestion exposure to allergenic pollen was obtained through equation 5.17,

$$E_{inge}(\Delta t_E; j,k,a,g) = \frac{1}{BW}\int_t^{t+\Delta t_E} P_{OAR}P_H S_H(1-(1-TE_{HM})^{TN_{HM}}) \\ [\underbrace{v_d c_{in}(t) I_{in}(t)}_{\text{Indoor}} + \underbrace{v_d c_{out}(t) I_{out}(t)}_{\text{Outdoor}}]dt \tag{5.17}$$

where P_{OAR} is the oral adsorption rate (%), P_H is the proportion of the hand area contacting the mouth for each touch event (%), TE_{HM} is the transfer efficiency from hand to mouth for a single touch event (%), TN_{HM} is the number of hand-to-mouth touch events per hour (1/h). S_H is surface area of hand (m^3), which can be calculated based on equation 5.18,

$$S_H = BW \times F_{BS}F_{HB} \tag{5.18}$$

where F_{HB} (%) is the fraction of hand surface area with respect to body surface area.

The exposure duration Δt_E can be set to different values for assessing exposures associated with different time durations. In the current study, Δt_E was set as 24 hours to calculate daily exposures to allergenic pollen. Daily exposures to allergenic pollen through inhalation, unintentional ingestion and deposition on skin surface were calculated for each day during the pollen season. The duration of pollen season is determined using the observed pollen counts in each of the climate regions; roughly from March 1st to June 30th for birch, oak and grass and July 1st to October 31st for ragweed and mugwort. The simulated daily exposure represents the intake of allergenic pollen during 24 hours for a calendar day for a "virtual subject".

Maximum daily exposures for each "virtual subject" during pollen season were selected as a metric for further comparison of exposures among different regions, periods, exposure routes and sources, ages and genders. For example, the maximum daily exposure through inhalation for a "virtual subject" with age a and gender g during period of j in climate region k (E^M_{inha}(j,k,a,g)) was calculated using equation 5.19.

$$E^M_{inha}(j,k,a,g) = \max_i E_{inha}(i,j,k,a,g) \tag{5.19}$$

Similarly, the maximum daily exposure through dermal deposition and unintentional ingestion can also be calculated using the same method as shown in equation 5.19. The choice of this metric is explained in section 5.3.3, and referred to as "peak exposure" in later discussion.

The difference in peak exposure between periods of 1994-2000 and 2001-2010 was calculated to investigate the climate change impact on exposure to allergenic pollen during these two periods. Since "virtual population" during periods of 1994-2000 and 2001-2010 were sampled from US demographic data in 2000 and 2010, respectively; there are not strict corresponding relationships between "virtual populations" during these two periods. For calculation of relative difference of exposure between two periods, percentiles (1^{st} to 100^{th}) of maximum daily exposures were calculated from exposures of 3,000 "virtual subject" in each climate region. The relative difference of exposure during two periods in a climate region, for example the relative difference of inhalation exposure in climate region k (i.e.,$\mathrm{E}^{D}_{inha}(\mathrm{k})$), was then calculated based on the corresponding percentiles as presented in equation 5.20,

$$E^{D}_{inha}(k) = \frac{\mathrm{Pct}(E^{M}_{inha}(P2,k)) - \mathrm{Pct}(E^{M}_{inha}(P1,k))}{\mathrm{Pct}(E^{M}_{inha}(P1,k))} \tag{5.20}$$

where $\mathrm{Pct}(E^{M}_{inha}(P2,k))$ and $\mathrm{Pct}(E^{M}_{inha}(P1,k))$ are percentiles of maximum daily inhalation exposures in climate region k during periods of 2001-2010 and 1994-2000, respectively. The relative difference of dermal and ingestion exposures during two periods can be calculated using the same method in equation 5.20.

5.2.6 Sensitivity analysis

Global sensitivity analyses were performed to test the sensitivity of the exposure model to multiple inputs and parameters based on Morris' design [114]. This design estimates the main effect of a parameter by computing a number of local sensitivities at random points of the parameter space. The mean of these randomized local sensitivities indicates the overall influence of a given parameter on the output metric, while the corresponding standard deviation indicates the effects of interaction and nonlinearity [115].

Three thousand "virtual subjects" were generated for the Northeast climate region.

Table 5.2: Parameters for calculating the exposures to allergenic pollen. These parameters were listed either as fixed values, known distributions or unknown empirical distributions derived from the literature.

Parameter	**ID**	**Distribution**	**Mean (STD)**	**Range**	**Reference**
r_{IO}, ratio of indoor to outdoor conc.(%)	1	Equation 5.5		region and species specific	[142,143]
t_{out}, outdoor exposure time (hrs)	2	Equation 5.7		age, gender and region specific	[3]
IN_{in}, indoor inhalation rate ($10^{-5}m^3$/(hr kg BW))	3	Equation 5.9		age and gender specific	[4]
IN_{out}, outdoor inhalation rate ($10^{-5}m^3$/(hr kg BW))	4	Equation 5.10		age and gender specific	[4]
TE_{DS}, adhesion efficiency on skin (%)	5	Uniform	10	5-15	[144]
S_{hum}, exposed human skin surface area (m^2)	6	Equation 5.16		age and gender specific	[144]
v_d, deposition velocity (m/h)	7	Equation 3.23		species specific	[113]
BW, body weight (kg)	8	Equation 5.15		age and gender specific	[4]
F_{BS}, ratio of body surface to weight (m^2/kg)	9	Norm	0.05(0.02)	0.03-0.08	[4,145]
P_{EX}, percent of exposed body surface area(%)	10	Uniform	29.0	18.5-39.5	[4]
P_{OAR}, oral adsorption rate (%)	11	Uniform	62.5	40-85	[146]
P_H, portion of hand surface touching mouth (%)	12	Fixed	10		[147]
S_H, hand surface area (cm^2)	13	Equation 5.18		age and gender specific	[4]
TE_{HM}, hand-mouth touch transfer efficiency (%)	14	Uniform	50	40-60	[4]
TN_{HM}, hand-mouth touch number (1/h)	15	Uniform	2	1-3	[4]
F_{HB}, fraction of hand surface to body surface area, female (%)	16	Empirical		4.8-5.1	[4]
F_{HB}, fraction of hand surface to body surface area, male (%)	16	Empirical		4.5-5.2	[4]

The daily exposure to allergenic oak pollen for each of the 3000 "virtual subjects" in the Northeast region were simulated for each day during a pollen season in the period of 2001-2010. The maximum daily exposure (peak exposure) was then calculated for each "virtual subject" during the pollen season. Mean peak exposures of these 3000 "virtual subjects" was selected as a metric for testing the exposure model's sensitivity to multiple inputs and parameters.

In the current study, each of the 16 parameters (Table 5.2) was sampled 8,500 times according to Morris' method from 500 random trajectories (each has 17 steps) in the parameter space[114,115]. Each of the parameters was perturbed between 50% and 150% of its base value or distribution while keeping other parameters unchanged. Equation 5.21 was used to calculate the Normalized Sensitivity Coefficient (NSC) for peak exposure through inhalation NSC_{inha} at a local point:

$$NSC_{inha} = \frac{\Delta \overline{E^M_{inha}} / \overline{E^M_{inha}}}{\Delta P / P} \tag{5.21}$$

where $\overline{E^M_{inha}}$ and P are the mean peak exposure through inhalation and the input parameters, respectively; and $\Delta \overline{E^M_{inha}}$ and ΔP are the perturbations in the exposure and input parameters, respectively. The local NSCs for exposure through unintentional ingestion and dermal deposition were calculated in the same way as in equation 5.21.

The global NSC of a parameter, NSCg, is defined as the mean of the corresponding local sensitivities. The average absolute global NSC, $\overline{|NSCg|}$, for each parameter and exposure route can be derived based on means of the absolute NSCg. Similarly, the standard deviations averaged over each parameter and exposure route ($\overline{STD}$) can be obtained to evaluate the interaction and nonlinearity effect of input parameters on modeling output.

5.3 Results and Discussion

5.3.1 Indoor and outdoor exposure time

Figure 5.2 depicts the distribution of indoor and outdoor exposure time of different age groups in spring, summer, fall and winter based on an analysis of the information contained in CHAD[3]. The population is divided into five age groups according to the

Exposure Factors Handbook[4], which are 1 to 4 years old, 5 to 11 years old, 12 to 17 years old, 18 to 64 years old, and older than 64 years.

As shown in figure 5.2, people tend to spend more time outdoors during summer and fall, and less time outdoors during winter. In particular, the third age group (12-17 years old) spends on average around 9.4 hours per day in outdoor environments in summer; and the fourth age group (18-64 years old) spends on average around 4.9 hours per day in outdoor environments in fall. Furthermore, the third age group spends on average 4.3 hours per day in outdoor environments, which is the longest outdoor exposure time among the five age groups during spring. Since the outdoor and indoor time add up to 24 hours, Figure 5.2 shows a high degree of symmetry between indoor and outdoor exposure time. As a result of this symmetry, further discussion only focuses on distribution of outdoor exposure time.

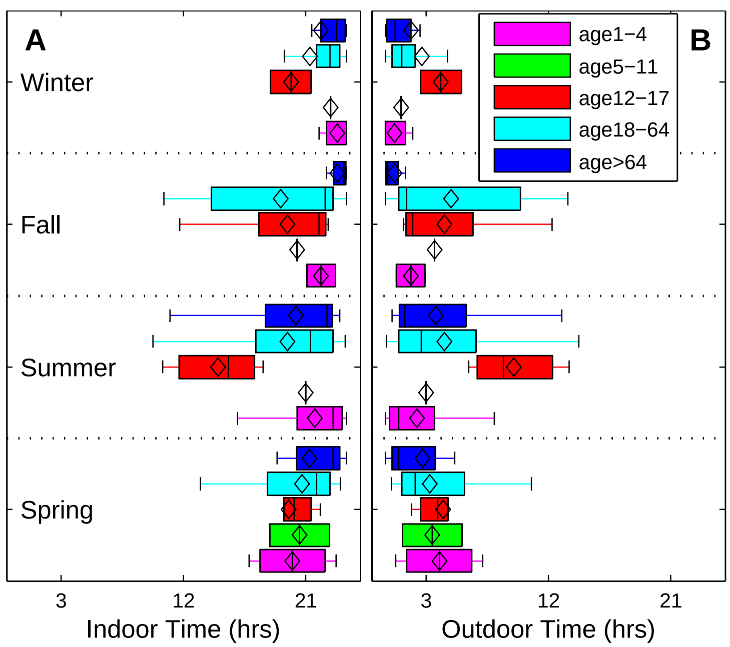

Figure 5.2: Indoor and outdoor exposure times for five age groups during spring, summer, fall and winter based on the Consolidated Human Activity Database[3]. In each box plot the central black line is the median; the black diamond is the mean; two sides are the 25^{th} (q_1) and 75^{th} (q_3) percentiles; the whiskers represent $q_3+1.5(q_3-q_1)$ and $q_1-1.5(q_3-q_1)$, respectively.(A) Indoor and (B) Outdoor.

Figure 5.3 presents the distribution of outdoor exposure time for five age groups in nine climate regions during spring, summer, fall and winter across the CONUS. Similar to Figure 5.2, the outdoor exposure times during summer and fall for people for all age groups are greater than those during winter; and the third age group (12-17 years) spends the longest time in outdoor environments among five age groups during spring across the majority of climate regions in the CONUS.

Outdoor exposure time appears similar among different climate regions in winter and spring, but appears distinct among regions in summer and fall. Specifically, the outdoor exposure time in northern regions (e.g., North West and West North Central) are longer than those in southern regions (e.g., South and South East). In particular, the second and third age groups (5-17 years old) spend longer times in outdoor environments than other age groups in most of the climate regions during summer.

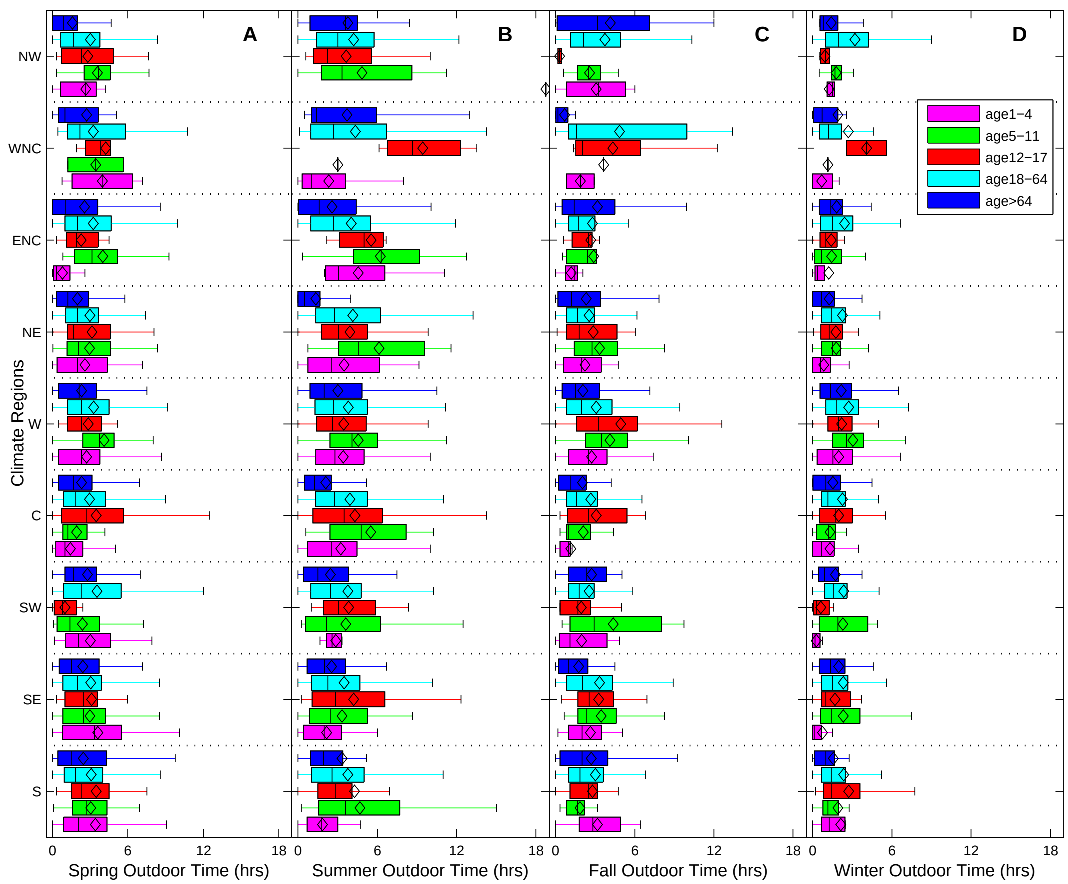

Figure 5.3: Outdoor exposure times for five age groups in nine climate regions during spring, summer, fall and winter based on the Consolidated Human Activity Database [3]. The symbols are same as defined in Figure 5.2. (A) Spring, (B) Summer, (C) Fall, and (D) Winter.

5.3.2 Distribution of inhalation rates

Figure 5.4 shows the distribution of inhalation rates according to age groups, genders and activity levels. The distribution was derived using measurement data from the Exposure Factors Handbook [4]. The inhalation rate has been normalized by body weight. The indoor and outdoor inhalation rates for a given "virtual subject" were simulated through equations 5.9 and 5.10, respectively.

Figure 5.4 shows that inhalation rate increases from lower (passive) to higher (high) activity level. On average, it ranges from 3.4×10^{-3}(m^3/(hr kg BW)) for male from the first age group (1-4 years old) at high activity level to 6.6×10^{-5}(m^3/(hr kg BW)) for female from the fourth age group (18-64 years old) at passive activity level. The normalized inhalation rate seems similar between female and male populations. At all activity levels, the first age group has the highest normalized inhalation rate among the five age groups; their inhalation rates on average are around two to five times higher than those of other age groups at the same activity level.

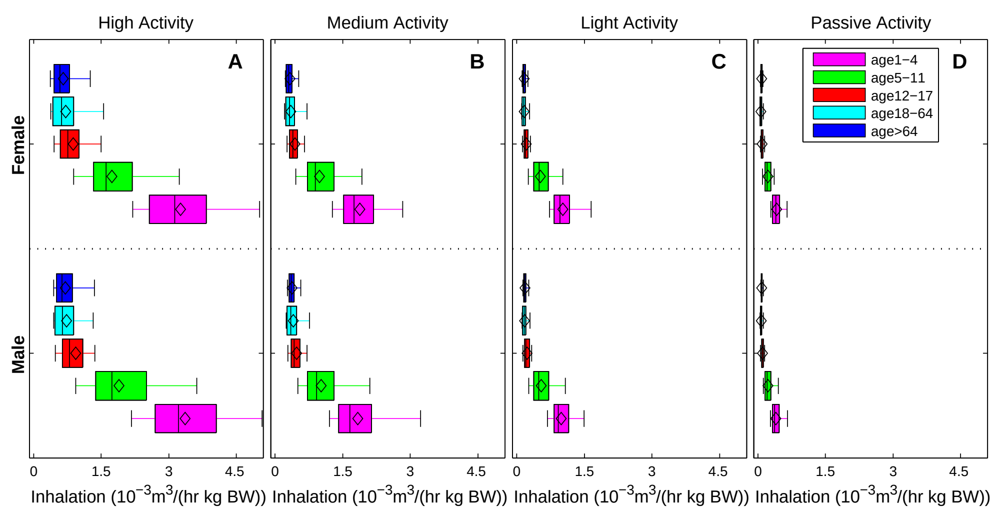

Figure 5.4: Distribution of inhalation rates according to age groups, genders, and activity levels based on the Exposure Factors Handbook [4]. The symbols are the same as defined in Figure 5.2. (A) High Activity, (B) Medium Activity, (C) Low Activity, and (D) Passive Activity.

5.3.3 Exposures to allergenic pollen

As an example, figure 5.5 illustrates the representative profiles of the simulated daily oak pollen concentrations, exposure times, inhalation rates, and inhalation exposures. The simulation assumed a scenario of a "virtual subject" in late spring and early summer during period of 2001-2010 in the Northeastern CONUS. The simulated "virtual subject" is a 67 years old male, whose indoor and outdoor activity was tracked from April 1st to June 15th. Since the simulations were essentially random samples from the observed daily pollen counts, exposure times and inhalation rates, the simulated time series can capture the variations of daily inhalation exposure to pollen.

Figure 5.6 shows the time series of inhalation exposures to oak pollen in nine climate regions during periods of 1994-2000 and 2001-2010. For each climate region, a unique "virtual subject" was tracked to calculate his/her indoor and outdoor inhalation exposures to oak pollen between March 1^{st} and June 30^{th}. Figure 5.6 indicates that the inhalation exposure to oak pollen in different climate regions occurred at different times and lasted for different durations. In general, the exposures to oak pollen in southern and southeastern CONUS were higher and lasted for longer time.

Maximum daily exposure (Peak Exposure) during pollen season was selected as a metric for further comparison of exposure to pollen among different regions, periods, exposure routes and sources, ages and genders. First, as shown in figure 5.6, the duration of pollen season are different in different regions, and even not aligned for the same region in different periods (e.g., East North Central and West North Central regions). This makes other metrics, such as seasonal mean and total, biased to compare pollen exposures among different regions and periods. Furthermore, maximum daily exposure is a good candidate to indicate the worst scenario for population exposures to allergic pollen during pollen season in spring and fall.

Similar to Figures 5.5 and 5.6, 3,000 "virtual subjects" and their exposures to pollen were simulated for each of the nine climate regions during each of the periods of 1994-2000 and 2001-2010. These simulation results were then used to generate the statistics (percentiles, maximum, mean, median etc.) of exposures to pollen. Oak pollen exposure during 2001-2010 was used as an example to compare the exposures among different

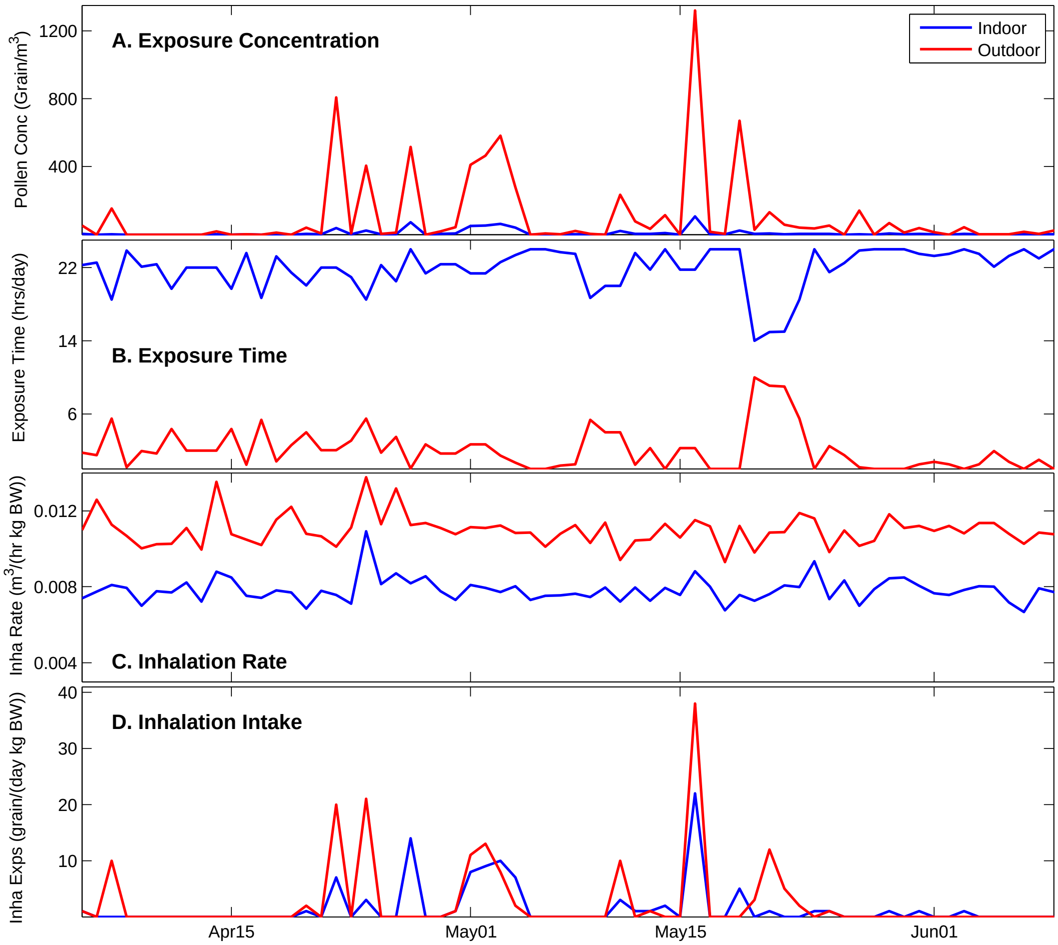

Figure 5.5: Representative simulated time series of daily oak pollen concentration, exposure time, inhalation rate, and inhalation exposure. The simulated "virtual subject" is a 67 years old male, whose indoor and outdoor activity was simulated for April 1st to June 15th in a given year during 2001-2010 in the Northeastern climate region. (A) Pollen Concentration, (B) Exposure Time, (C) Inhalation Rate, and (D) Inhalation Intake.

ages, genders and exposure sources in nine climate regions.

Figure 5.7 compares the exposure to oak pollen between female and male populations in different climate regions. The exposure to oak pollen for males is slightly higher than for females. The differences are prominent in the Northeast, South and Southeast climate regions. The aggregated exposure for males on average is between 69 pollen grains/(day kg BW) in the Northwest region and 1400 pollen grains/(day kg BW) in the South region; while the aggregated exposure for females is between 69 and 1330 pollen grains/(day kg BW).

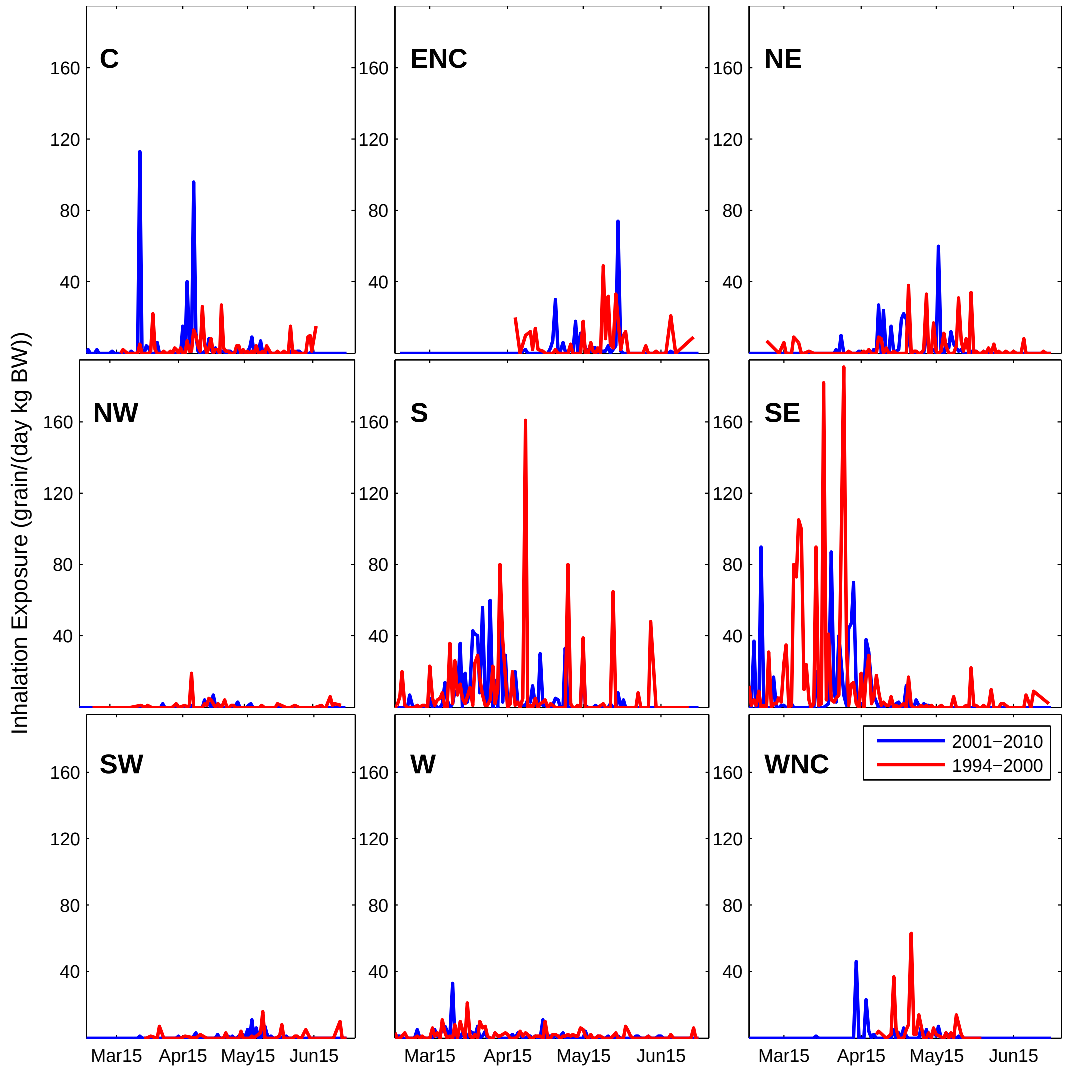

Figure 5.6: Time series of inhalation exposure to oak pollen in nine climate regions during periods of 1994-2000 and 2001-2010. The exposure includes both outdoor and indoor source.

The exposure to oak pollen in the southern CONUS (e.g., Southeast and South regions) is much higher than that in the northern CONUS (e.g., Northwest and East North Central regions). This is consistent with area coverage of oak trees and the distribution of oak pollen concentrations in different regions. Surprisingly, exposure through oak pollen deposition on exposed skin surface is more than twice as high as through inhalation and unintentional ingestion. However, the effect of dermal exposure on allergy is much lower than exposure through inhalation and unintentional ingestion.

Figure 5.8 compares the exposure to oak pollen among five age groups in nine

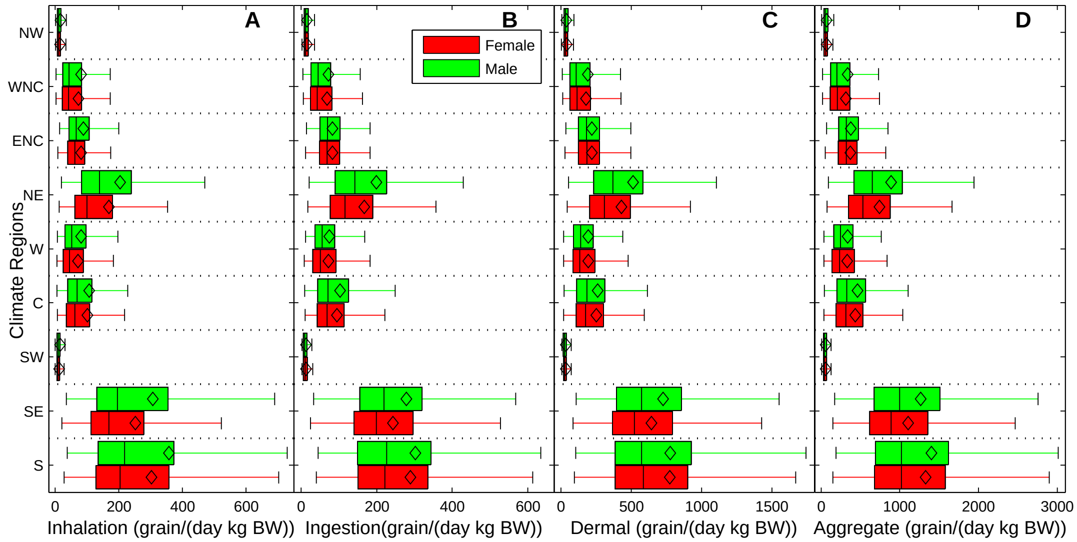

Figure 5.7: Comparison of exposure to oak pollen between female and male populations in nine climate regions. In each of the nine climate regions, 3,000 "virtual subjects" were divided into female and male groups to generate the box plot. In each box plot the central black line is the median; the black diamond is the mean; two sides are the 25^{th} (q_1) and 75^{th} (q_3) percentiles; the whiskers represent $q_3+1.5(q_3\text{-}q_1)$ and $q_1\text{-}1.5(q_3\text{-}q_1)$, respectively. (A) Inhalation, (B) Ingestion, (C) Dermal, and (D) Total.

climate regions. The difference is not distinct among age groups for exposures to pollen through unintentional ingestion and deposition on exposed skin surface. However, the inhalation and aggregated exposures to oak pollen for the first age group (1-4 years old) are higher than those for other age groups. In particular, inhalation exposure for the first age group is on average from 42 pollen grains/(day kg BW) in the Southwest region to 1073 pollen grains/(day kg BW) in the South region. These inhalation exposures for the first age group is two to five times higher than for other age groups in each of the nine climate regions in the CONUS.

The exposure to oak pollen in figure 5.8 also shows similar spatial pattern as shown in figure 5.7, which is that exposure in southern regions of US is generally larger than those in northern regions. It also indicates that dermal and inhalation are the dominant exposure routes, which contribute the majority of aggregated exposure to allergenic pollen.

Figure 5.9 compares the exposure to oak pollen in indoor and outdoor environments in nine climate regions. Exposure to oak pollen in outdoor environments is much

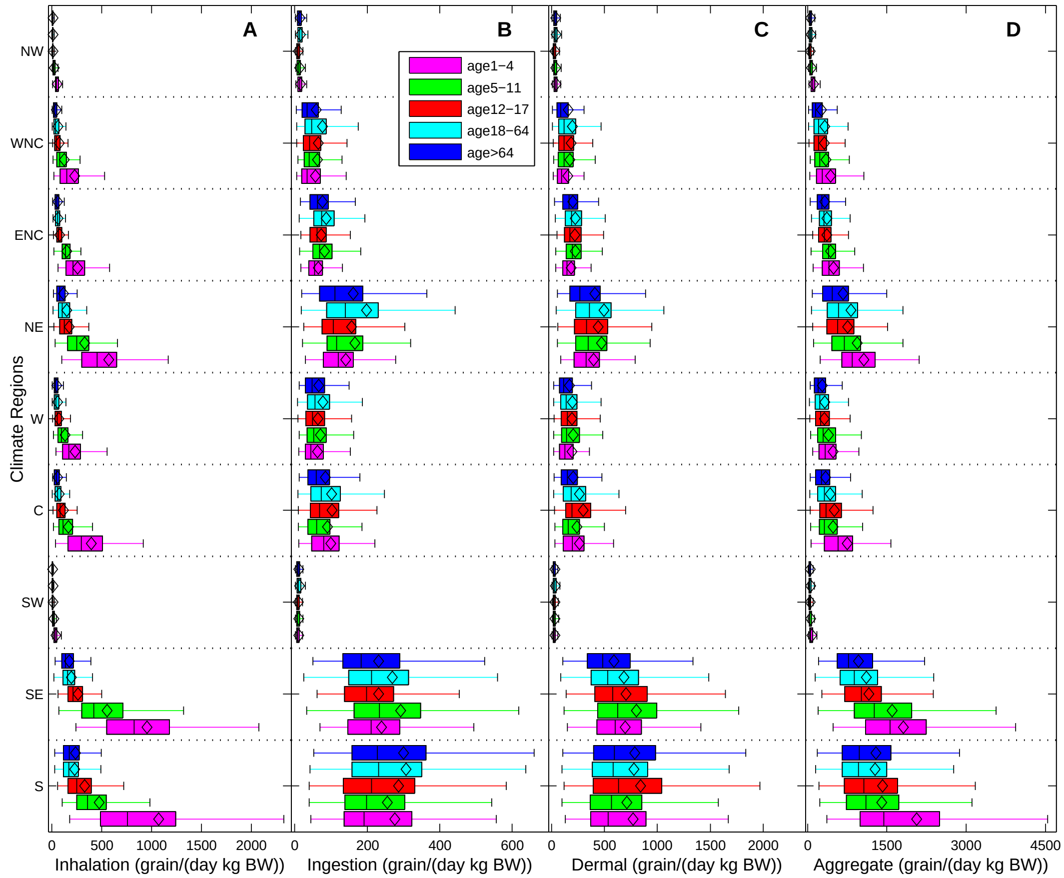

Figure 5.8: Comparison of exposure to oak pollen among five age groups in nine climate regions. In each of the nine climate regions, 3,000 "virtual subjects" were divided into three age groups to generate the box plot. The symbols are the same as defined in Figure 5.7. (A) Inhalation, (B) Ingestion, (C) Dermal, and (D) Total.

higher than in indoor environments for all exposure routes. The aggregated exposure in outdoor environments ranges from 51 pollen grains/(day kg BW) in southwest region to 1127 pollen grains/(day kg BW) in the south region. The aggregated exposure in outdoor environments is on average two to three times more than in indoor environments in each of the nine climate regions in the CONUS. Similar patterns in space and exposure routes, as shown in figure 5.7 and 5.8, can also be identified from figure 5.9.

Figures 5.10 to 5.14 present the comparison of exposures to allergenic pollen between periods of 1994-2000 (1990s) and 2001-2010 (2000s) for oak, birch, ragweed, mugwort and grass, respectively. The presented exposures are aggregations of exposures from different ages, genders, and indoor and outdoor environments. The box plots in panel

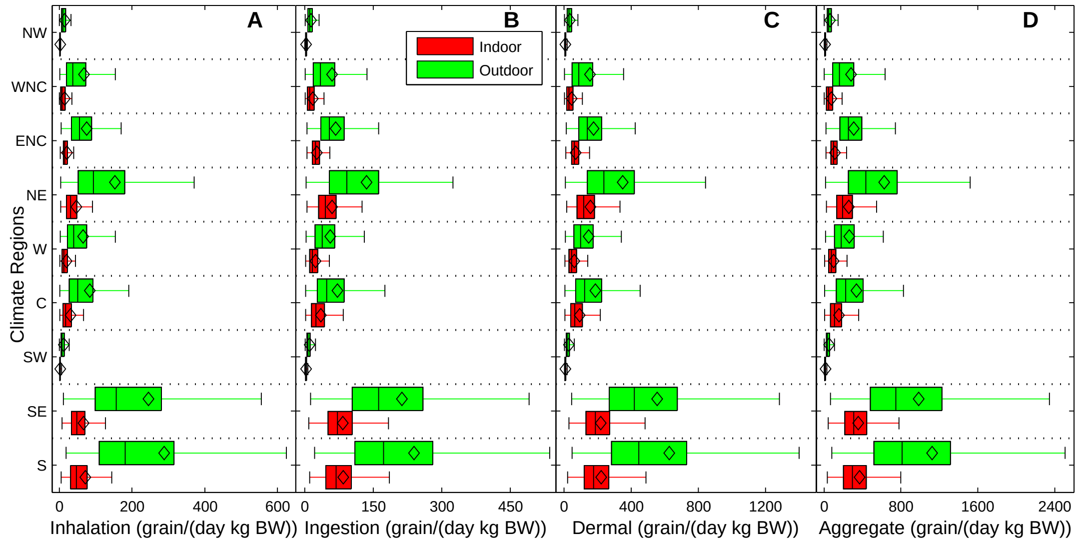

Figure 5.9: Comparison of exposure to oak pollen in indoor and outdoor environments in nine climate regions. In each of the nine climate regions, exposures of each of the 3,000 "virtual subjects" were divided into indoor and outdoor sources to generate the box plot. The symbols are the same as defined in Figure 5.7. (A) Inhalation, (B) Ingestion, (C) Dermal, and (D) Total.

(A) were generated based on maximum daily exposures of each "virtual subject" from the 3,000 "virtual subjects" during period of 1994-2000 in each climate region; and box plots in panel (B) show simulation results during period of 2001-2010. The box plots in panel (C) were generated according to equation 5.20 based on corresponding percentiles of maximum daily exposures during periods of 1994-2000 and 2001-2010.

As shown in figures 5.10 to 5.14, dermal deposition and inhalation were the dominant exposure routes. They contributed to the majority of exposures to allergenic pollen for birch, oak, ragweed and grass during two periods in nine climate regions. Ingestion and inhalation exposures to grass pollen are comparable with each other (Figure 5.14). This may be due to the large diameter of grass pollen grains (35 μm). The large diameter leads to less penetration of grass pollen into indoor environments, and quick deposition on exposed skin surface. Exposures and their difference between periods of 1994-2000 and 2001-2010 varied in different climate regions. The relative difference of exposures are summarized in figure 5.15.

For exposure to oak pollen (Figure 5.10), the exposures were higher in southern

CONUS (e.g., South and Southeast regions) and lower in northern CONUS (e.g., Northwest and East North Central regions). The exposures during the period of 2001-2010 increased in seven out of nine climate regions. Compared with the 1990s, the aggregated exposures in the 2000s increased, on average, between 1% in the West region and 85% in the South region; while the exposures in the Southeast and West North Central regions decreased by 17% and 4%, respectively.

For exposure to birch pollen (Figure 5.11), the exposures during the past decade decreased compared with the 1990s in the North West, West North Central, West, Central, and South climate regions; but increased in the East North Central, North East, and South East climate regions. In particular, the inhalation and aggregated exposures in the North East region during the 2000s increased by 124% and 130%, respectively, compared with those during the 1990s.

For exposure to ragweed pollen (Figure 5.12), the exposures during the past decade decreased compared with the 1990s in the West North Central, East North Central, North East, Central, and South East climate regions; but increased in the West, South West, and South climate regions. In particular, the inhalation and aggregated exposures in the West region during the 2000s increased by 540% and 538%, respectively, compared with those during the 1990s.

For exposure to mugwort pollen (Figure 5.13), data were missing in six out of nine climate regions. Among three regions with available data, the exposures during the 2000s increased compared with the 1990s in the West, and South West climate regions. In particular, the inhalation and aggregated exposures in the South West region during the past decade increased by 328% and 335%, respectively, compared with those during the 1990s.

For exposure to grass pollen (Figure 5.14), the exposures during the 2000s decreased compared with the 1990 in the West North Central, East North Central, West, and South East climate regions; but increased in the North West, North East, Central, South West, and South climate regions. In particular, the inhalation and aggregated exposures in the South region during the past decade increased by 448% and 445%, respectively, compared with those during the 1990s.

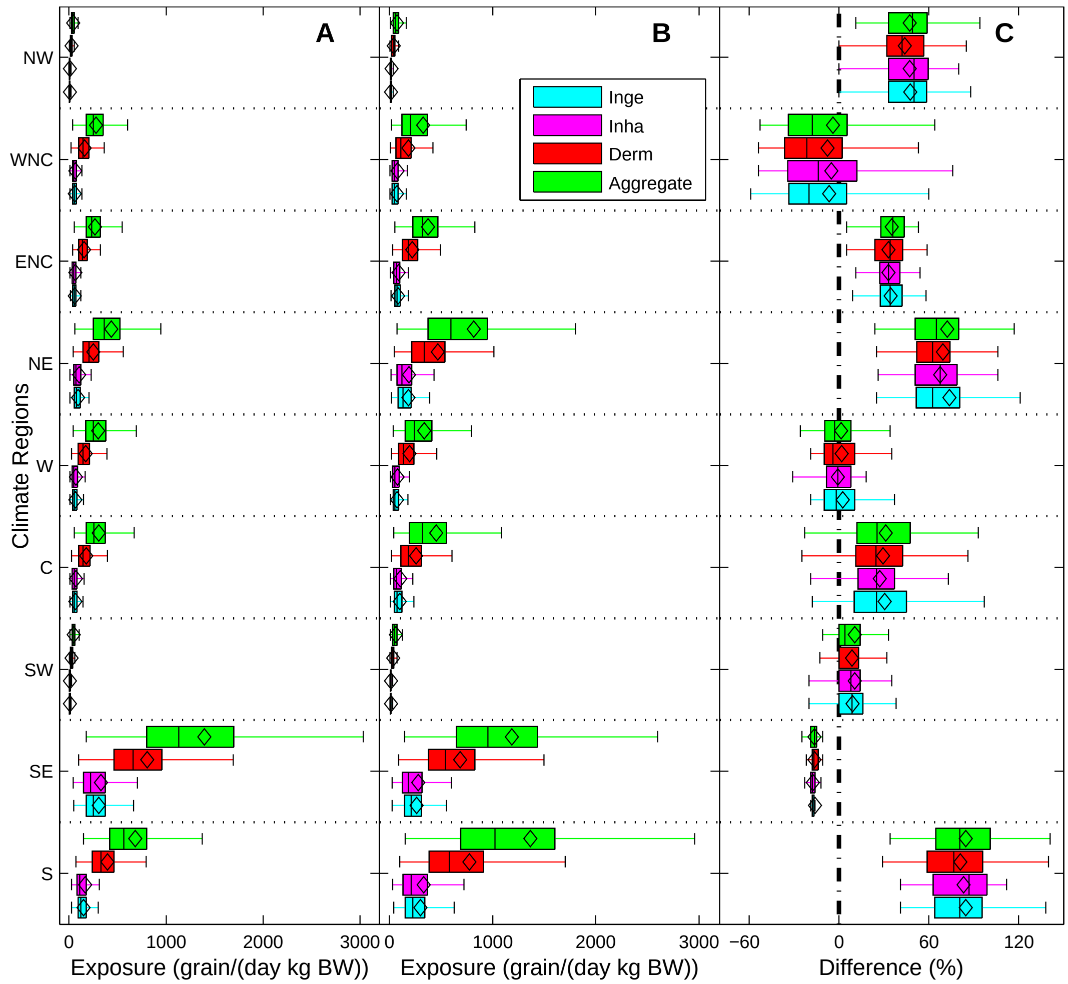

Figure 5.10: Comparison of exposure to oak pollen between periods of 1994-2000 and 2001-2010 in nine climate regions. In each of the nine climate regions, 3,000 "virtual subjects" and their exposures were simulated during periods of 1994-2000 and 2001-2010 to generate the box plot. The symbols are the same as defined in Figure 5.7. (A) Exposure during 1994-2000, (B) Exposure during 2001-2010, (C) Difference between two periods

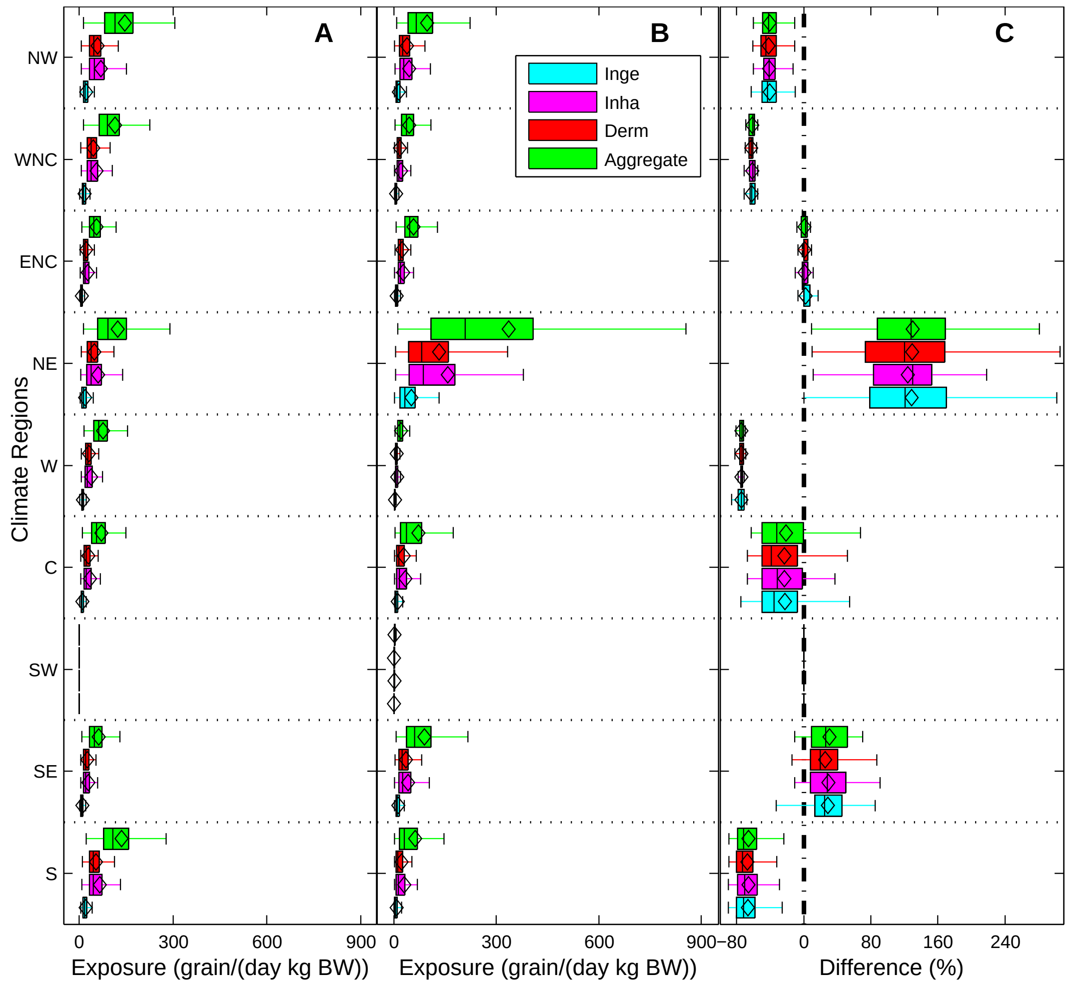

Figure 5.11: Comparison of exposure to birch pollen between periods of 1994-2000 and 2001-2010 in nine climate regions. In each of the nine climate regions, 3,000 "virtual subjects" and their exposures were simulated during periods of 1994-2000 and 2001-2010 to generate the box plot. The symbols are the same as defined in Figure 5.7. (A) Exposure during 1994-2000, (B) Exposure during 2001-2010, (C) Difference between two periods

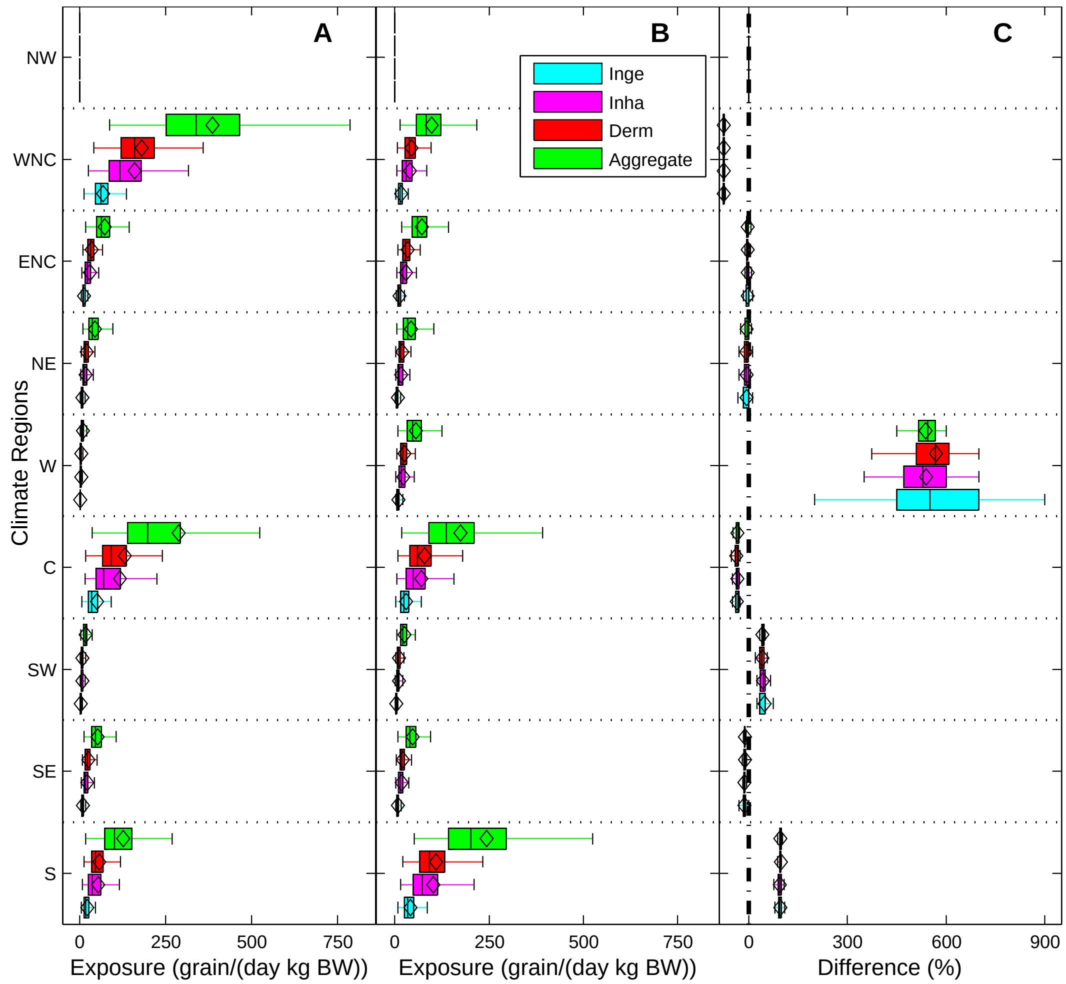

Figure 5.12: Comparison of exposure to ragweed pollen between periods of 1994-2000 and 2001-2010 in nine climate regions. In each of the nine climate regions, 3,000 "virtual subjects" and their exposures were simulated during periods of 1994-2000 and 2001-2010 to generate the box plot. The symbols are the same as defined in Figure 5.7. (A) Exposure during 1994-2000, (B) Exposure during 2001-2010, (C) Difference between two periods

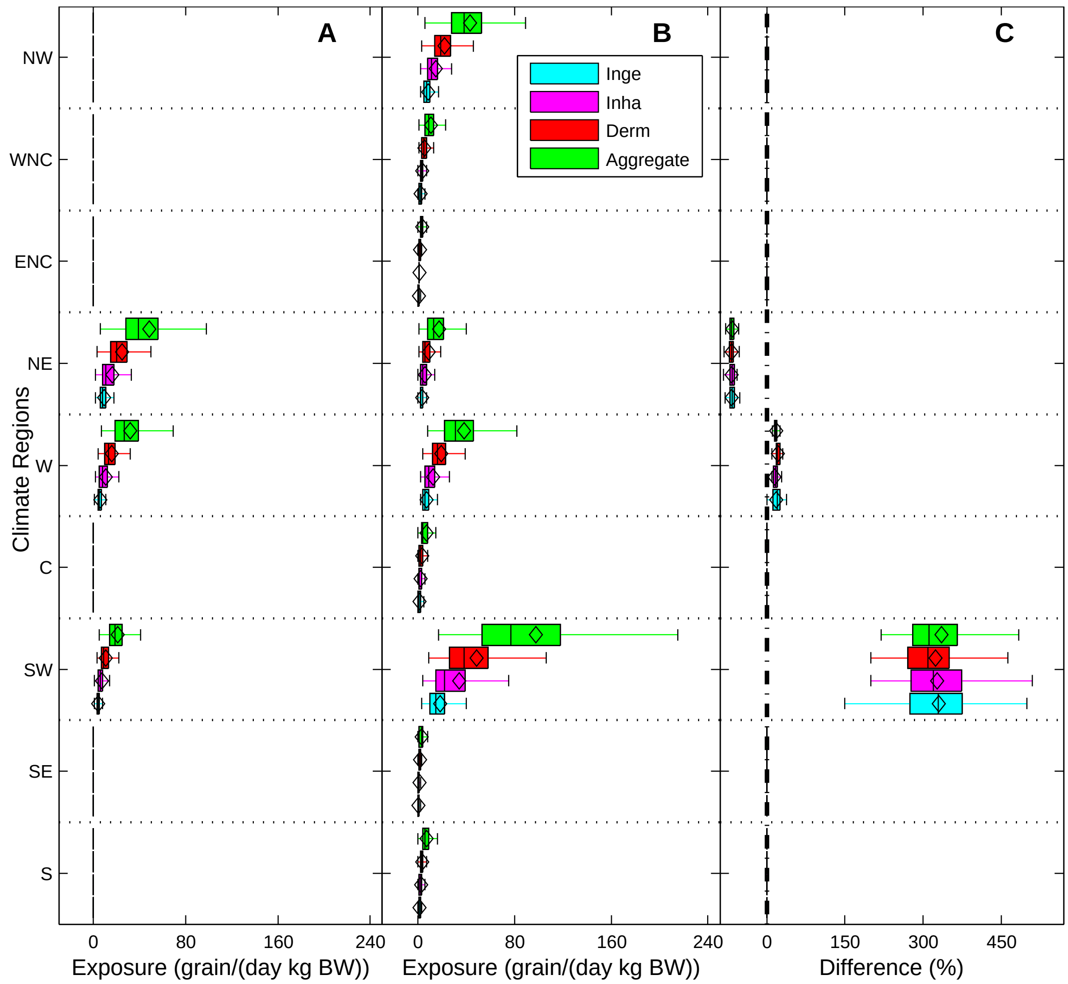

Figure 5.13: Comparison of exposure to mugwort pollen between periods of 1994-2000 and 2001-2010 in nine climate regions. In each of the nine climate regions, 3,000 "virtual subjects" and their exposures were simulated during periods of 1994-2000 and 2001-2010 to generate the box plot. The symbols are the same as defined in Figure 5.7. (A) Exposure during 1994-2000, (B) Exposure during 2001-2010, (C) Difference between two periods

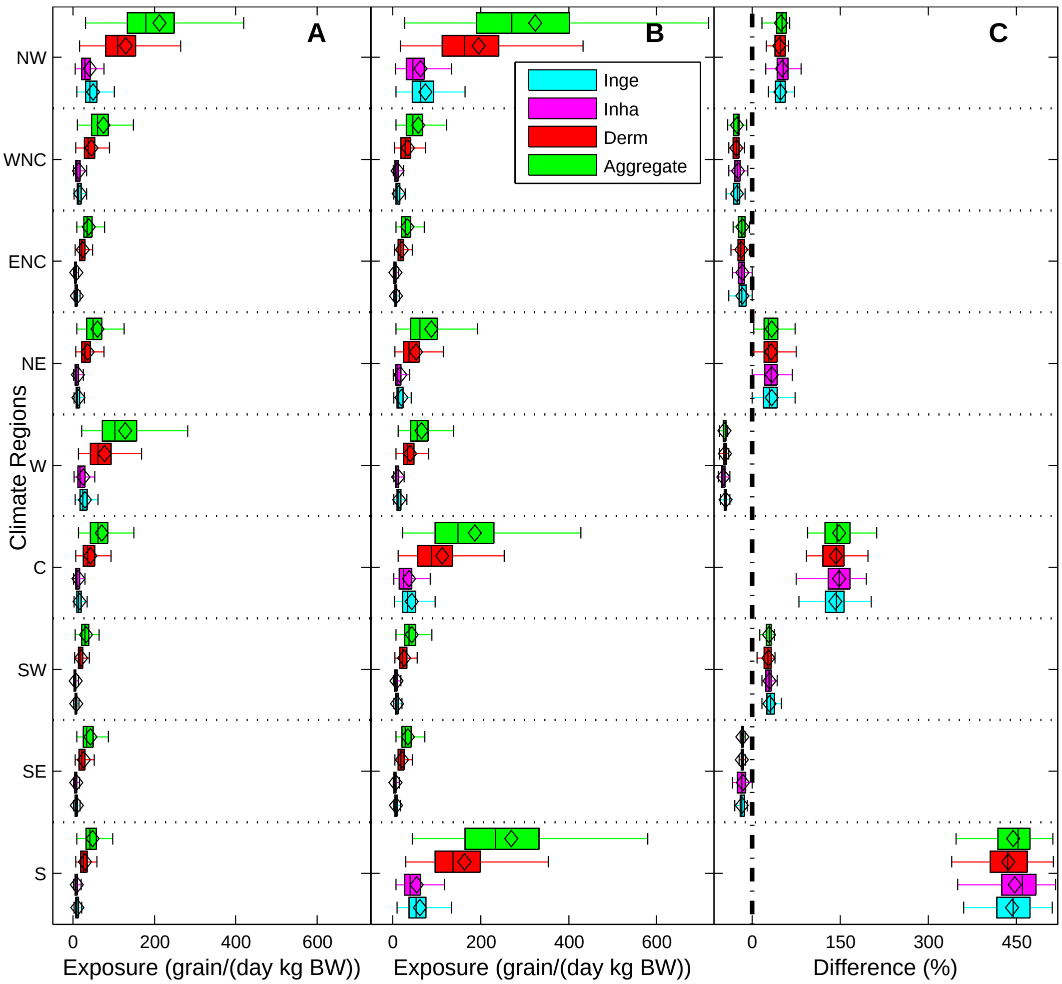

Figure 5.14: Comparison of exposure to grass pollen between periods of 1994-2000 and 2001-2010 in nine climate regions. In each of the nine climate regions, 3,000 "virtual subjects" and their exposures were simulated during periods of 1994-2000 and 2001-2010 to generate the box plot. The symbols are the same as defined in Figure 5.7. (A) Exposure during 1994-2000, (B) Exposure during 2001-2010, (C) Difference between two periods

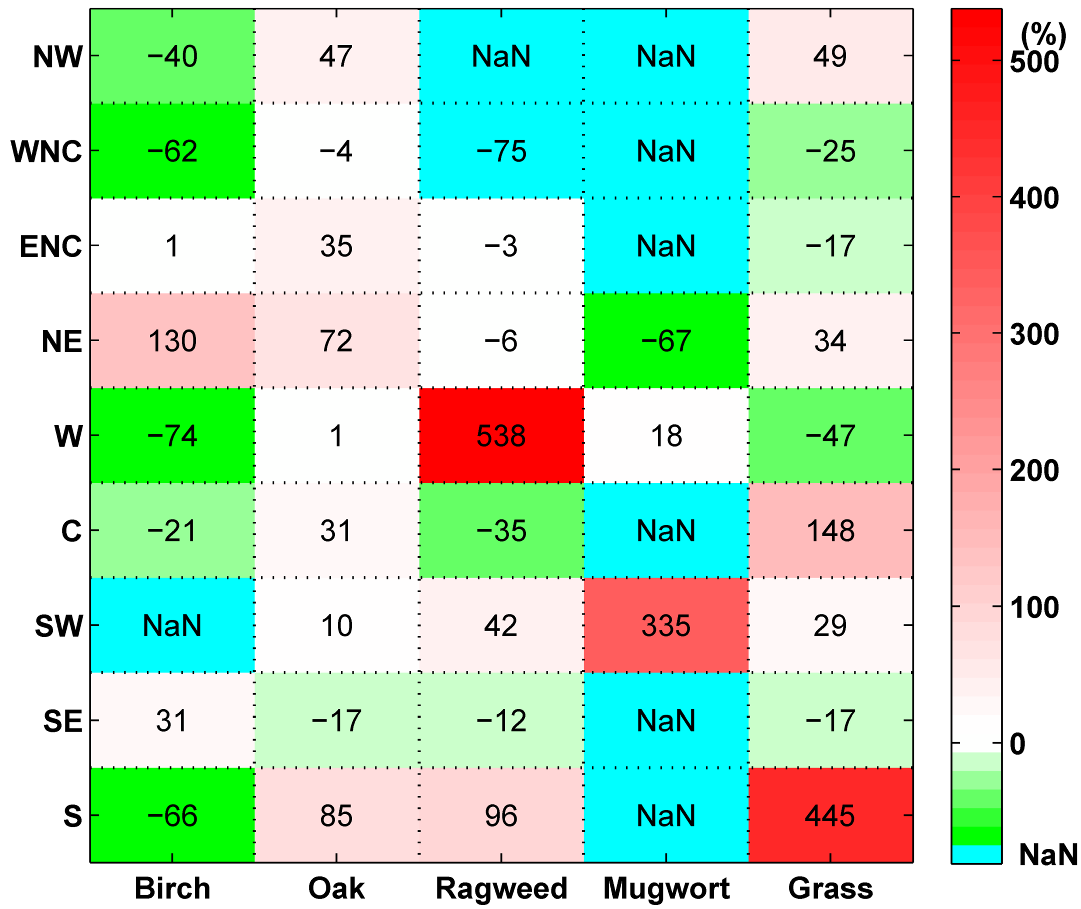

Figure 5.15: Changes in exposures to allergenic pollen during period of 2001-2010 from those during period of 1994-2000 across the CONUS. The metric of exposure used here is the average peak value of daily exposure of 3,000 "virtual subjects" in each climate region. The number in each cell indicates the percentage of increase (positive) or decrease (negative).

5.3.4 Sensitivity analysis

The global sensitivity of the simulated exposures to different parameters is presented in Figure 5.16. Overall, the global NSC of all parameters varied between -0.12 and 0.08, indicating the robustness of the modeling approach. Ingestion and dermal exposures were more sensitive to parameter perturbations, with average absolute global NSC, $\overline{|NSCg|}$, being 0.05 and 0.03, respectively. Sensitive parameters included: exposed hu-

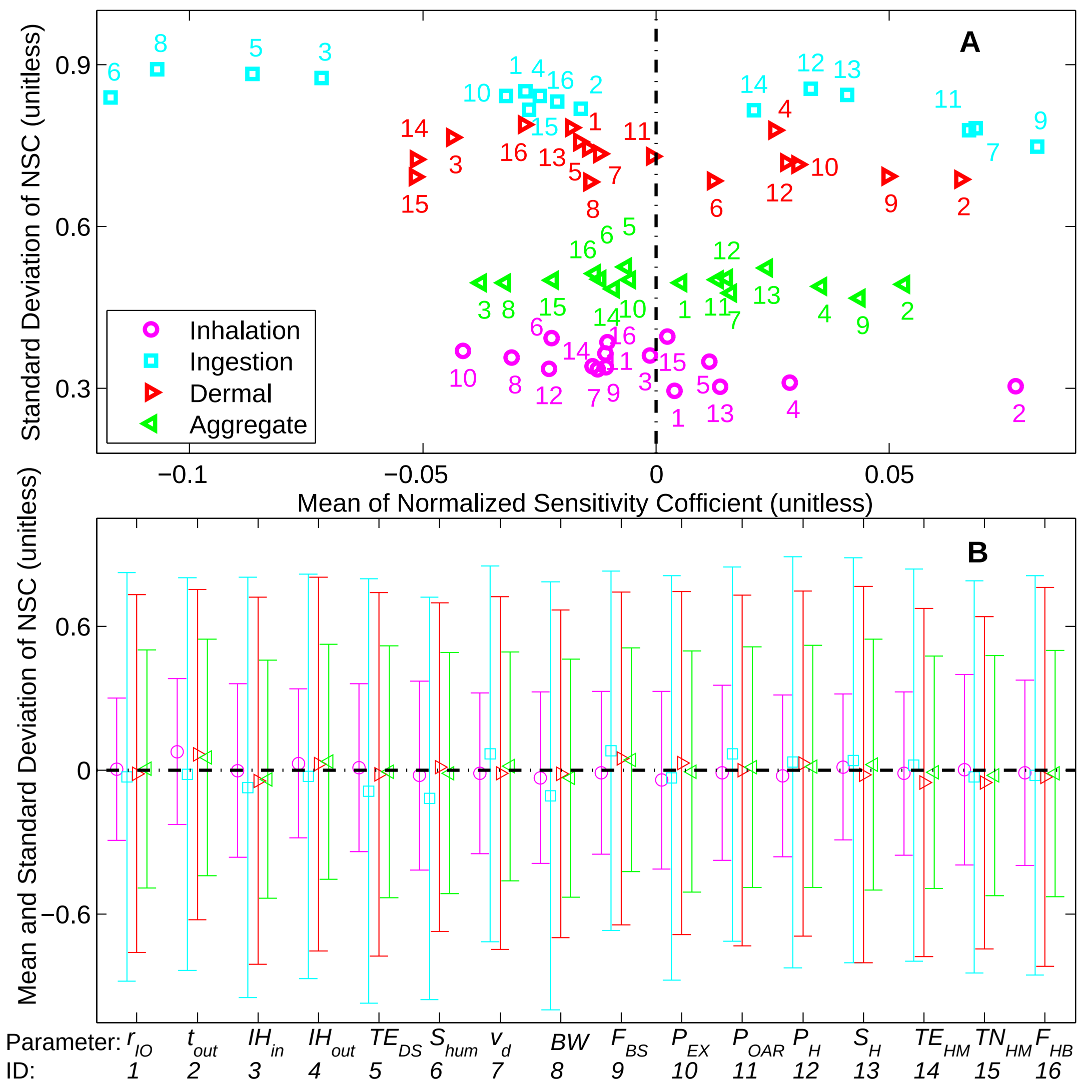

Figure 5.16: Mean and standard deviation of Normalized Sensitivity Coefficient (NSC) for each parameter for exposures to oak pollen through unintentional ingestion, inhalation and dermal deposition. The vertical dashed lines represent the NSC values of 0. All parameters are described in table 5.2.

man skin surface area (S_{hum}), body weight (BW), adhesion efficiency on skin (TE_{DS}), the ratio of the skin surface area to body weight (F_{BS}), and outdoor exposure time (t_{out}). The inlation and aggregated exposures are sensitive to outdoor exposure time and inhalation rate.

High interaction and nonlinearity effects among parameters were found for exposures through unintentional ingestion and dermal deposition, with average interaction effects $\overline{STD}$ being 0.82 and 0.73, respectively. Parameters with high interaction and nonlinearity effects included: body weight (BW), adhesion efficiency on skin (TE_{DS}), portion of hand surface touch mouth (P_H), and indoor inhalation rate (IH_{in}).

Uncertainties in sensitive and interactive input parameters result in large deviations of model predictions. Parameters derived from large population studies, such as distribution of body weight and inhalation rates are believed to bear low uncertainties. High uncertainties are expected for sensitive parameters: exposed human skin surface area (S_{hum}), the ratio of the skin surface area to body weight (F_{BS}), and outdoor exposure time (t_{out}); and interactive parameters: adhesion efficiency on skin (TE_{DS}), and portion of hand surface touch mouth (P_H).

In the current study, probabilistic distributions were used to capture the variability of the ratio of indoor to outdoor pollen concentrations in different climate regions for different taxa (Tables 5.1 and 5.2). The global sensitivity analysis indicates that the developed exposure model is generally not sensitive and interactive to the perturbations in the ratio of indoor to outdoor pollen concentrations.

5.4 Summary

A probabilistic model has been developed to simulate exposure to allergenic pollen based on the framework of the Modeling Environment for Total Risk studies (MENTOR)[60]. This exposure model was mainly driven by observation data of airborne pollen counts, demographics, and time spent indoors and outdoors in nine climate regions in the CONUS. It also incorporates information of air exchange rates, penetration factors, and inhalation rates at different activity levels for different genders and age groups. The exposure model was applied to study the changes and the spatiotemporal pattern of exposures to allergenic birch, oak, ragweed, mugwort and grass pollen during the

1990s (1994-2000) and the 2000s (2001-2010) in nine climate regions in the CONUS.

Inhalation and dermal deposition were the dominant exposure routes for allergenic pollen. The aggregated exposure to allergenic pollen in outdoor environments was around two to three times more than that in indoor environments. Exposures to allergenic pollen were not distinct between females and males. The inhalation exposures for children of 1-4 years old was two to five times higher than those for other age groups in each of the nine climate regions in the CONUS. Changes in exposures to allergenic pollen between periods of 1994-2000 and 2000-2010 varied for different climate regions and allergenic taxa. The aggregated exposure to oak pollen during the 2000s was 1%-85% higher than those during the 1990s in seven climate regions; and 4% and 12% lower in the West North Central and Southeastern regions, respectively. In particular, the aggregated exposure to birch, oak, ragweed, mugwort and grass pollen during the 2000s was 130%, 85%, 538%, 335% and 445% higher than those during the 1990s in North East, South, West, South West, and South climate regions, respectively.

Ingestion and dermal exposures were more sensitive and interactive to parameter perturbations. Sensitive parameters included exposed human skin surface area, body weight, adhesion efficiency on skin, the ratio of the skin surface area to body weight, and outdoor exposure time. Parameters with high interaction and nonlinearity effects included body weight, adhesion efficiency on skin, portion of hand surface touch mouth, and indoor inhalation rate.

Chapter 6
CONCLUSIONS AND RECOMMENDATIONS

This chapter presents the main findings and recommendations for future research. In this dissertation, climate change impact on allergenic pollen was investigated through statistical analysis and modeling of observed airborne pollen counts and climatic factors, and through simulation using a deterministic modeling system.

6.1 Main Findings

6.1.1 Model development

First, statistical analysis was carried out to examine the trends and changes in the observed pollen season start date, duration and airborne pollen level based on the airborne pollen counts during the periods of 1994-2010 from the National Allergy Bureau (NAB) of the American Academy of Allergy (AAAAI), Asthma & Immunology monitor stations across the contiguous United States. The Bayesian statistics and machine learning models were used to identify the relationships among the observed pollen season start date, duration and airborne pollen levels, and the meteorological, phenological and geographical factors.

Second, a comprehensive deterministic modeling system was developed to support integrated studies of climate change effect on airborne allergens, which include the biogenic allergenic pollen from trees, weeds and grasses, and air pollutants ozone and particulate matter. The modeling system consists of the Weather Research and Forecast (WRF) model, the Sparse Matrix Operator Kernel Emissions (SMOKE) model, the pollen emission model developed in the current study, and an expanded version of the Community Multiscale Air Quality (CMAQ) model. A mechanistic pollen emission model was developed for this deterministic WRF-SMOKE-CMAQ-Pollen modeling system based on the vegetation area coverage, pollen season onset and duration, and daily and hourly flowering likelihood.

Third, a probabilistic exposure model was developed to study the changes and the spatiotemporal patterns of exposures to allergenic birch, oak, ragweed, mugwort and grass pollen during the 1990s (1994-2000) and the 2000s (2001-2010) in nine climate regions in the CONUS.

6.1.2 Observed allergenic pollen season variations under changing climate

The allergenic pollen seasons of representative trees, weeds and grasses during the 2000s across the CONUS have been observed to start 3.0 (95% Confidence Interval [CI], 1.1-4.9) days earlier on average than in the 1990s. The average peak value and annual total of daily counted airborne pollen have increased by 42.4% (95% CI, 21.9%-62.9%) and 46.0% (95% CI, 21.5%-70.5%), respectively. Changes of the observed pollen season timing and airborne level depend on latitude, and are associated with changes of growing degree days, frost free days, and precipitation. These changes are likely due to recent climate change and particularly the enhanced warming and precipitation at higher latitudes in the CONUS. The observed pollen season start date, season length and airborne pollen level could be correctly predicted using Bayesian and machine learning models based on the locally observed meteorological factors.

6.1.3 Pollen emission model

A mechanistic pollen emission model has been developed based on mass balance of pollen grain fluxes in the near surrounding of allergenic plants. The emission model consists of direct emission and resuspension, and accounts for influences of temperature, wind velocity and relative humidity. Modules of this emission model have been developed and parameterized to provide pollen season onset and duration, daily and hourly flowering likelihood, and vegetation coverage for birch, oak, ragweed, mugwort and grass.

The emission model is robust with respect to the pollen emissions of oak, ragweed, mugwort and grass, but highly sensitive and interactive to perturbations in input parameters for birch pollen emission. The sensitive and interactive parameters included the threshold temperature and wind speed, the density of pollen grain, the aerodynamic resistance and quasi-laminar resistance, and the flowering likelihood.

6.1.4 Performance of WRF-SMOKE-CMAQ-Pollen modeling system

The performance of the WRF-SMOKE-CMAQ-Pollen modeling system was evaluated operationally based on the observed pollen counts at monitor stations across the CONUS in 2004. For oak and ragweed pollen, this modeling system correctly predicted the observed pollen season start date and duration, and airborne pollen level at the majority of monitor stations. The WRF-SMOKE-CMAQ-Pollen modeling system could capture the variations in start date, season length and airborne level of birch, mugwort and grass pollen. However it did not perform as well as for oak and ragweed pollen. The dry deposition, emission and vertical eddy diffusion were the dominant processes determining the ambient pollen concentrations. The boundary condition exerted less influence on mean pollen concentrations, but remarkable influence on maximum pollen concentrations in some northern states bordering Canada.

6.1.5 Climate change impact on allergenic pollen

The response of allergenic pollen season to climate change varies in different climate regions for different taxa. For ragweed, mugwort and grass, the regional average of pollen concentrations was predicted to decrease in the majority of climate regions during the period of 2047-2050. For oak and birch, although there were not remarkable increases of airborne pollen concentrations during the period of 2047-2050, the number of hours in which pollen concentrations exceed the threshold values for triggering allergy was predicted to increase in the majority of climate regions.

6.1.6 Exposure to airborne allergenic pollen

Inhalation and dermal deposition were the dominant exposure routes for allergenic pollen. The aggregated exposure to allergenic pollen in outdoor environments was around two to three times more than that that in indoor environments. The inhalation exposures for children of 1-4 years old was two to five times higher than for other age groups. Changes in exposures to allergenic pollen between the periods of 1994-2000 and 2000-2010 varied in different climate regions for different taxa. In particular, the aggregated exposure to birch, oak, ragweed, mugwort and grass pollen during the 2000s was 130%, 85%, 538%, 335% and 445% higher than those during the 1990s in North East, South, West, South West, and South climate regions, respectively.

6.2 Future Research Recommendation

6.2.1 Improve the performance of the WRF-SMOKE-CMAQ modeling system

(1) Develop mechanistic daily and hourly flowering likelihood functions for birch, mugwort, grass and oak, so that the flowering likelihood functions can incorporate the influence of meteorology factors and geographical factors.

(2) Develop dynamic vegetation coverage of allergenic plants.

(3) Run the latest version of WRF to generate high resolution meteorology simulations.

6.2.2 Expand the model capability

(1) Adjust and parameterize the developed emission model, and apply it to simulate emission and transport of mold and spores.

(2) Develop modules to simulate start date, duration, and daily and hourly emission patterns for mold and spores.

(3) Observations of mold and spore counts could be obtained from the NAB-AAAAI monitor stations.

6.2.3 Utilize the satellite data

(1) Find sensitive characteristic spectral (reflection or absorption) of airborne allergenic pollen, mold and spores.

(2) Find suitable satellite products (e.g., fluorecence plant map from GOSAT) which have wave bands containing the characteristic spectra of airborne allergenic pollen, mold and spores.

(3) Establish an empirical relationship between satellite data and observed concentrations of airborne allergenic pollen, mold and spores using machine learning models.

(4) Derive sptiotemporal distributions of allergenic pollen, mold and spores from satellite images or products.

6.2.4 Relate the allergen distribution to allergic airway disease

(1) Observed O_3 and $PM_{2.5}$ concentrations could be obtained from AQS Data Mart.

(2) Hospitalization data of allergy patients could be obtained from NHANS dataset.

(3) Establish a statistical or empirical relationship between Allergic Airway Disease (e.g., Asthma) and airborne anthropogenic allergens ($PM_{2.5}$ and O_3), and biogenic allergens (pollen from birch, oak, ragweed, mugwort, grass, mold and spores).

(4) Assess the health effects of anthropogenic and biogenic allergens using climate models, regional meteorology and air quality models based on the above derived relationships.

6.2.5 Expand the exposure model to study indoor air quality under changing climate

(1) Study changes in background concentrations of anthropogenic and biogenic air stressors due to climate change.

(2) Study changes in building design and structure due to climate change.

(3) Study changes in humidity, temperature and ventilation rate for indoor environments due to climate change.

(4) Study changes in emission pattern of air stressors in indoor environments due to climate change.

(5) Study changes in human activity patterns in indoor environments due to climate change.

(6) Study empirical relationships between broad indoor air stressors (ventilation rate, humidity, temperature, air pollutants and allergens, *etc.*) and broad health effects under climate change scenarios.

REFERENCES

[1] Thomas Karl and Walter James Koss. *Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983*. Historical climatology series 4-3. National Climatic Data Center, Asheville, NC, 1984.

[2] Alison M Eyth and K Habisak. The mims spatial allocator: a tool for generating emission surrogates without a geographic information system. In *Proceedings, 12th International Emission Inventory Conference, San Diego*, 2003.

[3] T. McCurdy, G. Glen, L. Smith, and Y. Lakkadi. The National Exposure Research Laboratory's consolidated human activity database. *Journal of Exposure Analysis and Environmental Epidemiology*, 10(6):566–578, Nov-Dec 2000.

[4] U.S. EPA. *Exposure Factors Handbook 2011 Edition.* U.S. Environmental Protection Agency, Washington, DC,, 2011 edition, 2011.

[5] Sonia Altizer, Richard S Ostfeld, Pieter TJ Johnson, Susan Kutz, and C Drew Harvell. Climate change and infectious diseases: from evidence to a predictive framework. *science*, 341(6145):514–519, 2013.

[6] Alistair Boxall, Anthony Hardy, Sabine Beulke, Tatiana Boucard, Laura Burgin, Peter D Falloon, Philip M Haygarth, Thomas Hutchinson, R Sari Kovats, Giovanni Leonardi, et al. Impacts of climate change on indirect human exposure to pathogens and chemicals from agriculture. *Environmental Health Perspectives*, 117(4):508–514, 2008.

[7] Paul R Epstein. Climate change and human health. *New England Journal of Medicine*, 353(14):1433–1436, 2005.

[8] Anthony J McMichael, Rosalie E Woodruff, and Simon Hales. Climate change

and human health: present and future risks. *The Lancet*, 367(9513):859–869, 2006.

[9] Jonathan A Patz, Diarmid Campbell-Lendrum, Tracey Holloway, and Jonathan A Foley. Impact of regional climate change on human health. *Nature*, 438(7066):310–317, 2005.

[10] Thomas Frei and Ewald Gassner. Climate change and its impact on birch pollen quantities and the start of the pollen season an example from Switzerland for the period 1969?2006. *International Journal of Biometeorology*, 52(7):667–674, 2008.

[11] H. Garcia-Mozo, C. Galan, P. Alcazar, C. de la Guardia, D. Nieto-Lugilde, M. Recio, P. Hidalgo, F. Gonzalez-Minero, L. Ruiz, and E. Dominguez-Vilches. Trends in grass pollen season in southern Spain. *Aerobiologia*, 26(2):157–169, Jun 2010.

[12] Patrick L Kinney. Climate change, air quality, and human health. *American journal of preventive medicine*, 35(5):459–467, 2008.

[13] Shiqiang Wan, Tong Yuan, Sarah Bowdish, Linda Wallace, Scott D. Russell, and Yiqi Luo. Response of an allergenic species, Ambrosia psilostachya (Asteraceae), to experimental warming and clipping: implications for public health. *Am. J. Bot.*, 89(11):1843–1846, November 1, 2002 2002.

[14] Colleen Reid and Janet Gamble. Aeroallergens, Allergic Disease, and Climate Change: Impacts and Adaptation. *EcoHealth*, 6(3):458–470, 2009.

[15] Mikhail Sofiev and Karl-Christian Bergmann. *Allergenic Pollen: A Review of the Production, Release, Distribution and Health Impacts.* Springer, ISBN: 9400748817, 2013.

[16] Sabit Cakmak, Robert E Dales, and Frances Coates. Does air pollution increase the effect of aeroallergens on hospitalization for asthma? *Journal of Allergy and Clinical Immunology*, 129(1):228–231, 2012.

[17] Charles E. Lamb, Paul H. Ratner, Clarion E. Johnson, Ambarish J. Ambegaonkar, Ashish V. Joshi, David Day, Najah Sampson, and Benjamin Eng. Economic

impact of workplace productivity losses due to allergic rhinitis compared with select medical conditions in the United States from an employer perspective. *Current Medical Research and Opinion*, 22(6):1203–1210, 2006.

[18] G. T. Slavov, S. P. DiFazio, and S. H. Strauss. Gene flow in forest trees: gene migration patterns and landscape modelling of transgene dispersal in hybrid poplar. pages 89–106. CABI Publishing, Wallingford, 2004. Author Affiliation: Oregon State University, Corvallis, OR 97331, USA. Author Email: Steve.Strauss@oregonstate.edu.

[19] P. J. Beggs. Impacts of climate change on aeroallergens: past and future. *Clinical & Experimental Allergy*, 34(10):1507–1513, 2004.

[20] Leonard Bielory, Kevin Lyons, and Robert Goldberg. Climate change and allergic disease. *Current allergy and asthma reports*, 12(6):485–494, 2012.

[21] Gina Dapul-Hidalgo and Leonard Bielory. Climate change and allergic diseases. *Annals of Allergy, Asthma & Immunology*, 109(3):166–172, 2012.

[22] IPCC. Climate change 2013: the physical science basis. Technical report, Cambridge University Press, doi:10.7930/J0Z31WJ2., 2013.

[23] A. H. Fitter and R. S. R. Fitter. Rapid Changes in Flowering Time in British Plants. *Science*, 296(5573):1689–1691, 2002.

[24] Jerry M. Melillo, Richmond Terese, and Gary W. Yohe. Climate change impacts in the united states: the third national climate assessment. Technical report, U.S. Global Change Research Program, doi:10.7930/J0Z31WJ2., 2014.

[25] Thomas Frei and Ewald Gassner. Trends in prevalence of allergic rhinitis and correlation with pollen counts in Switzerland. *International Journal of Biometeorology*, 52(8):841–847, 2008.

[26] J. F.; Aira J. M. Jato, V.; Rodr?gues-Rajo. Use of phenological and pollen-production data for interpreting atmospheric birch pollen curves. *Annals of Agricultural and Environmental Medicine*, 14:271–280, 2007.

[27] Pilvi Siljamo, Mikhail Sofiev, Elena Severova, Hanna Ranta, Jaakko Kukkonen, Svetlana Polevova, Eero Kubin, and Alexander Minin. Sources, impact and exchange of early-spring birch pollen in the Moscow region and Finland. *Aerobiologia*, 24(4):211–230, 2008.

[28] Michael D. Martin, Marcelo Chamecki, Grace S. Brush, Charles Meneveau, and Marc B. Parlange. Pollen clumping and wind dispersal in an invasive angiosperm. *Am. J. Bot.*, 96(9):1703–1711, September 1, 2009 2009.

[29] Aslog Dahl and Sven-Olov Strandhede. Predicting the intensity of the birch pollen season. *Aerobiologia*, 12(1):97–106, 1996.

[30] J. Emberlin, M. Detandt, R. Gehrig, S. Jaeger, N. Nolard, and L. Rantio. Responses in the start of Betula (birch) pollen seasons to recent changes in spring temperatures across Europe. *International Journal of Biometeorology*, 46(4):159–170, 2002.

[31] Alix Rasmussen. The effects of climate change on the birch pollen season in Denmark. *Aerobiologia*, 18(3):253–265, 2002.

[32] Eija Yli-Panula, Desta Bey Fekedulegn, Brett James Green, and Hanna Ranta. Analysis of Airborne Betula Pollen in Finland; a 31-Year Perspective. *International Journal of Environmental Research and Public Health*, 6(6):1706–1723, 2009.

[33] Hanna Ranta, Tatu Hokkanen, Tapio Linkosalo, Liisa Laukkanen, Kristoffer Bondestam, and Annukka Oksanen. Male flowering of birch: Spatial synchronization, year-to-year variation and relation of catkin numbers and airborne pollen counts. *Forest Ecology and Management*, 255(3-4):643–650, 2008.

[34] E. Tedeschini, F. J. Rodriguez-Rajo, R. Caramiello, V. Jato, and G. Frenguelli. The influence of climate changes in Platanus spp. pollination in Spain and Italy. *Grana*, 45(3):222–229, Sep 2006.

[35] Lewis Ziska, Kim Knowlton, Christine Rogers, Dan Dalan, Nicole Tierney, Mary Ann Elder, Warren Filley, Jeanne Shropshire, Linda B. Ford, Curtis Hedberg, Pamela Fleetwood, Kim T. Hovanky, Tony Kavanaugh, George Fulford, Rose F. Vrtis, Jonathan A. Patz, Jay Portnoy, Frances Coates, Leonard Bielory, and David Frenz. Recent warming by latitude associated with increased length of ragweed pollen season in central North America. *Proceedings of the National Academy of Sciences*, 108(10):4248–4251, March 8, 2011 2011.

[36] T. Frei. The effects of climate change in Switzerland 1969-1996 on airborne pollen quantities from hazel, birch and grass. *Grana*, 37(3):172–179, 1998.

[37] H. Garcia-Mozo, C. Galan, V. Jato, J. Belmonte, C. D. de la Guardia, D. Fernandez, M. Gutierrez, M. J. Aira, J. M. Roure, L. Ruiz, M. M. Trigo, and E. Dominguez-Vilches. Quercus pollen season dynamics in the Iberian Peninsula: Response to meteorological parameters and possible consequences of climate change. *Annals of Agricultural and Environmental Medicine*, 13(2):209–224, 2006.

[38] C. Efstathiou, S. Isukapalli, and P. Georgopoulos. A mechanistic modeling system for estimating large-scale emissions and transport of pollen and co-allergens. *Atmospheric Environment*, 45(13):2260–2276, Apr 2011.

[39] S. Kawashima and Y. Takahashi. An improved simulation of mesoscale dispersion of airborne cedar pollen using a flowering-time map. *Grana*, 38(5):316–324, 1999.

[40] Silvio Schueler and Katharina Schl?nzen. Modeling of oak pollen dispersal on the landscape level with a mesoscale atmospheric model. *Environmental Modeling and Assessment*, 11(3):179–194, 2006.

[41] Donald E. Aylor. Quantifying maize pollen movement in a maize canopy. *Agricultural and Forest Meteorology*, 131(3-4):247–256, 2005.

[42] Robert Pasken and Joseph A. Pietrowicz. Using dispersion and mesoscale meteorological models to forecast pollen concentrations. *Atmospheric Environment*, 39(40):7689–7701, 2005.

[43] Chiara Ziello, Tim H Sparks, Nicole Estrella, Jordina Belmonte, Karl C Bergmann, Edith Bucher, Maria Antonia Brighetti, Athanasios Damialis, Monique Detandt, Carmen Galán, et al. Changes to airborne pollen counts across europe. *PLoS One*, 7(4):e34076, 2012.

[44] Simon G Haberle, David MJS Bowman, Rewi M Newnham, Fay H Johnston, Paul J Beggs, Jeroen Buters, Bradley Campbell, Bircan Erbas, Ian Godwin, Brett J Green, et al. The macroecology of airborne pollen in australian and new zealand urban areas. *PloS one*, 9(5):e97925, 2014.

[45] Yong Zhang, Leonard Bielory, Zhongyuan Mi, Ting Cai, Alan Robock, and Panos Georgopoulos. Allergenic pollen season variations in the past two decades under changing climate in the united states. *Global change biology*, DOI:10.1111/gcb.12755., 2014.

[46] Nora Helbig, Bernhard Vogel, Heike Vogel, and Franz Fiedler. Numerical modelling of pollen dispersion on the regional scale. *Aerobiologia*, 20(1):3–19, 2004.

[47] Heike Vogel, Andreas Pauling, and Bernhard Vogel. Numerical simulation of birch pollen dispersion with an operational weather forecast system. *International Journal of Biometeorology*, 52(8):805–814, 2008.

[48] Rui Zhang, Tiffany Duhl, Muhammad T Salam, James M House, Richard C Flagan, Edward L Avol, Frank D Gilliland, Alex Guenther, Serena H Chung, Brian K Lamb, et al. Development of a regional-scale pollen emission and transport modeling framework for investigating the impact of climate change on allergic airway disease. *Biogeosciences (Online)*, 10(3):3977, 2013.

[49] TR Duhl, R Zhang, A Guenther, SH Chung, MT Salam, JM House, RC Flagan, EL Avol, FD Gilliland, BK Lamb, et al. The simulator of the timing and magnitude of pollen season (stamps) model: a pollen production model for regional emission and transport modeling. *Geoscientific Model Development Discussions*, 6(2):2325–2368, 2013.

[50] M Sofiev, P Siljamo, H Ranta, T Linkosalo, S Jaeger, A Rasmussen, A Rantio-Lehtimaki, E Severova, and J Kukkonen. A numerical model of birch pollen emission and dispersion in the atmosphere. description of the emission module. *International journal of biometeorology*, 57(1):45–58, 2013.

[51] Pilvi Siljamo, Mikhail Sofiev, Elena Filatova, Łukasz Grewling, Siegfried Jäger, Ekaterina Khoreva, Tapio Linkosalo, Sara Ortega Jimenez, Hanna Ranta, Auli Rantio-Lehtimäki, et al. A numerical model of birch pollen emission and dispersion in the atmosphere. model evaluation and sensitivity analysis. *International journal of biometeorology*, 57(1):125–136, 2013.

[52] Katrin Zink, Heike Vogel, Bernhard Vogel, Donát Magyar, and Christoph Kottmeier. Modeling the dispersion of ambrosia artemisiifolia l. pollen with the model system cosmo-art. *International journal of biometeorology*, 56(4):669–680, 2012.

[53] Marje Prank, Daniel S Chapman, James M Bullock, Jordina Belmonte, Uwe Berger, Aslog Dahl, Siegfried Jäger, Irina Kovtunenko, Donát Magyar, Sami Niemelä, et al. An operational model for forecasting ragweed pollen release and dispersion in europe. *Agricultural and Forest Meteorology*, 182:43–53, 2013.

[54] M. Sofiev, P. Siljamo, H. Ranta, and A. Rantio-Lehtimki. Towards numerical forecasting of long-range air transport of birch pollen: theoretical considerations and a feasibility study. *International Journal of Biometeorology*, 50(6):392–402, 2006.

[55] Linda O. Mearns, William Gutowski, Richard Jones, Ruby Leung, Seth McGinnis, Ana Nunes, and Yun Qian. A Regional Climate Change Assessment Program for North America. *Eos Trans. AGU*, 90(36), 2009.

[56] Linda Mearns, Ruby Leung, James Correia, and Yun Qian. The north american regional climate change assessment program dataset. National Center for Atmospheric Research Earth System Grid data portal, 2011/05/01 2011.

[57] Christian Hogrefe, Prakash Doraiswamy, Brian Colle, KennethL. Demerjian, Winston Hao, Mark Beauharnois, Michael Erickson, Matthew Souders, Jia-Yeong Ku, and Gopal Sistla. Ensemble air quality forecasting: Effects of perturbations in meteorology and emissions. In Douw G. Steyn and Silvia Trini Castelli, editors, *Air Pollution Modeling and its Application XXI*, NATO Science for Peace and Security Series C: Environmental Security, pages 411–415. Springer Netherlands, 2012.

[58] Jung-Hun Woo, Shan He, Praveen Amar, Efthimios Tagaris, Kasemsan Manomaiphiboon, Kao-Jen Liao, and Armistead G Russell. Development of mid-century anthropogenic emissions inventory in support of regional air quality modeling under influence of climate change. In *15th Annual Emission Inventory Conference Reinventing Inventories New Ideas in New Orleans*, pages 16–18, 2008.

[59] Kinnee E.J. Geron C.D. Pierce, T.E. Development of a 1-km resolvedvegetatio n cover data base for regional air quality modeling. In *23rd Conference on Agricultural and Forest Meteorology, American Meteorological Society*, pages 329–332, Boston, MA,, 1998.

[60] Panos G. Georgopoulos and Paul J. Lioy. From a Theoretical Framework of Human Exposure and Dose Assessment to Computational System Implementation: The Modeling ENvironment for TOtal Risk Studies (MENTOR). *Journal of Toxicology and Environmental Health, Part B*, 9(6):457–483, 2006/01/01 2006.

[61] N. Nakicenvoic, O. Davidson, G. Davis, A. Grbler, T. Kram, E. Rovere, M. Metz, T. Morita, W. Pepper, H. Pitcher, A. Sankovski, P. Shukla, R. Swart, R. Watson, and Z. Dadi. Special report on emissions scenarios: A special report of working group iii of the intergovernmental panel on climate change. Technical report, IPCC, Cambridge, UK, Cambridge University Press, 2000.

[62] T. L. Otte and J. E. Pleim. The Meteorology-Chemistry Interface Processor

(MCIP) for the CMAQ modeling system: updates through MCIPv3.4.1. *Geosci. Model Dev.*, 3(1):243–256, 2010.

[63] G. D'Amato, L. Cecchi, S. Bonini, C. Nunes, I. Annesi-Maesano, H. Behrendt, G. Liccardi, T. Popov, and P. Van Cauwenberge. Allergenic pollen and pollen allergy in Europe. *Allergy*, 62(9):976–990, 2007.

[64] M. Smith, C. A. Skjoth, D. Myszkowska, A. Uruska, M. Puc, A. Stach, Z. Balwierz, K. Chlopek, K. Piotrowska, I. Kasprzyk, and J. Brandt. Long-range transport of Ambrosia pollen to Poland. *Agricultural and Forest Meteorology*, 148(10):1402–1411, Sep 2008.

[65] William Deen, Clarence J. Swanton, and L. Anthony Hunt. A mechanistic growth and development model of common ragweed. *Weed Science*, 49(6):723–731, 2001/11/01 2001.

[66] M Black and PF Wareing. Photoperiodic control of germination in seed of birch (betula pubescens ehrh.). 1954.

[67] Chunyang Li, Olavi Junttila, Arild Ernstsen, Pekka Heino, and E Tapio Palva. Photoperiodic control of growth, cold acclimation and dormancy development in silver birch (betula pendula) ecotypes. *Physiologia Plantarum*, 117(2):206–212, 2003.

[68] Yong Zhang, Leonard Bielory, and Panos G Georgopoulos. Climate change effect on betula (birch) and quercus (oak) pollen seasons in the united states. *International journal of biometeorology*, pages 1–11, 2013.

[69] H Garcia-Mozo, I Chuine, MJ Aira, J Belmonte, D Bermejo, C Díaz de la Guardia, B Elvira, M Gutiérrez, J Rodríguez-Rajo, L Ruiz, et al. Regional phenological models for forecasting the start and peak of the¡ i¿ quercus¡/i¿ pollen season in spain. *agricultural and forest meteorology*, 148(3):372–380, 2008.

[70] Yong Zhang, Sastry S Isukapalli, Leonard Bielory, and Panos G Georgopoulos.

Bayesian analysis of climate change effects on observed and projected airborne levels of birch pollen. *Atmospheric Environment*, 68:64–73, 2013.

[71] László Makra, István Matyasovszky, Anna Páldy, and Áron József Deák. The influence of extreme high and low temperatures and precipitation totals on pollen seasons of ambrosia, poaceae and populus in szeged, southern hungary. *Grana*, 51(3):215–227, 2012.

[72] Clara Deser, Reto Knutti, Susan Solomon, and Adam S. Phillips. Communication of the role of natural variability in future North American climate. *Nature Clim. Change*, 2(11):775–779, 2012.

[73] Kazuhiko Masaka and Shirou Maguchi. Modelling the Masting Behaviour of Betula platyphylla var. japonica using the Resource Budget Model. *Annals of Botany*, 88(6):1049–1055, December 1, 2001 2001.

[74] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.

[75] Emmanuel Gringarten and Clayton V Deutsch. Teacher's aide variogram interpretation and modeling. *Mathematical Geology*, 33(4):507–534, 2001.

[76] Jean-Michel Marin and Christian P. Robert. Regression and variable selection. In *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*, Springer Texts in Statistics, pages 47–84. Springer New York, 2007a.

[77] Jean-Michel Marin and Christian P. Robert. Normal models. In *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*, Springer Texts in Statistics, pages 15–46. Springer New York, 2007b.

[78] V Vapnik and AJ Lerner. Generalized portrait method for pattern recognition. *Automation and Remote Control*, 24(6):774–780, 1963.

[79] Vladimir Vapnik. *The nature of statistical learning theory*. springer, New York, 2000.

[80] Dimitris Voukantsis, Harri Niska, Kostas Karatzas, Marina Riga, Athanasios Damialis, and Despoina Vokou. Forecasting daily pollen concentrations using data-driven modeling methods in thessaloniki, greece. *Atmospheric Environment*, 44(39):5101–5111, 2010.

[81] Simon Haykin. *Neural networks: a Comprehensive Foundation, Second ed.* Prentice Hall, 1999.

[82] L Breiman, JH Friedman, RA Olshen, and C.J. Stone. *Classification and Regression Trees*. Wadsworth, Belmont, California, 1984.

[83] John R Quinlan et al. Learning with continuous classes. In *Proceedings of the 5th Australian joint Conference on Artificial Intelligence*, volume 92, pages 343–348. Singapore, 1992.

[84] Y. Zhang, P.G. Georgopoulos, and L. Bielory. Climate change effects on ragweed (ambrosia) and mugwort (artemisia) pollen seasons in usa. In *Proceedings of the Air and Waste Management Association's Annual Conference and Exhibition, AWMA*, volume 4, pages 2983–2993, 2013.

[85] Ellen Kinnee, Chris Geron, and Thomas Pierce. UNITED STATES LAND USE INVENTORY FOR ESTIMATING BIOGENIC OZONE PRECURSOR EMISSIONS. *Ecological Applications*, 7(1):46–58, 1997.

[86] Anna Bock, Tim H. Sparks, Nicole Estrella, Nigel Jee, Andrew Casebow, Christian Schunk, Michael Leuchner, and Annette Menzel. Changes in first flowering dates and flowering duration of 232 plant species on the island of guernsey. *Global Change Biology*, DOI: 10.1111/gcb.12579, 2014.

[87] Camille Parmesan and Gary Yohe. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421(6918):37–42, 2003.

[88] Terry L Root, Jeff T Price, Kimberly R Hall, Stephen H Schneider, Cynthia Rosenzweig, and J Alan Pounds. Fingerprints of global warming on wild animals and plants. *Nature*, 421(6918):57–60, 2003.

[89] Chiara Ziello, Andreas Böck, Nicole Estrella, Donna Ankerst, and Annette Menzel. First flowering of wind-pollinated species with the greatest phenological advances in europe. *Ecography*, 35(11):1017–1023, 2012.

[90] David W Inouye, Billy Barr, Kenneth B Armitage, and Brian D Inouye. Climate change is affecting altitudinal migrants and hibernating species. *Proceedings of the National Academy of Sciences*, 97(4):1630–1633, 2000.

[91] Gian-Reto Walther, Eric Post, Peter Convey, Annette Menzel, Camille Parmesan, Trevor JC Beebee, Jean-Marc Fromentin, Ove Hoegh-Guldberg, and Franz Bairlein. Ecological responses to recent climate change. *Nature*, 416(6879):389–395, 2002.

[92] Jessica L Blois, Phoebe L Zarnetske, Matthew C Fitzpatrick, and Seth Finnegan. Climate change and the past, present, and future of biotic interactions. *Science*, 341(6145):499–504, 2013.

[93] Jonathan T Overpeck, David Rind, and Richard Goldberg. Climate-induced changes in forest disturbance and vegetation. 1990.

[94] John Pastor and WM Post. Response of northern forests to co2-induced climate change. 1988.

[95] IPCC. Chapter 10 Global Climate Projections. Technical report, Cambridge, United Kingdom and New York, NY, USA, 2007 2007c.

[96] Eric Post, Mads C. Forchhammer, M. Syndonia Bret-Harte, Terry V. Callaghan, Torben R. Christensen, Bo Elberling, Anthony D. Fox, Olivier Gilg, David S. Hik, Toke T. Hye, Rolf A. Ims, Erik Jeppesen, David R. Klein, Jesper Madsen, A. David McGuire, Sren Rysgaard, Daniel E. Schindler, Ian Stirling, Mikkel P. Tamstorf, Nicholas J.C. Tyler, Rene van der Wal, Jeffrey Welker, Philip A. Wookey, Niels Martin Schmidt, and Peter Aastrup. Ecological dynamics across the arctic associated with recent climate change. *Science*, 325(5946):1355–1358, 2009.

[97] Lewis H Ziska and Paul J Beggs. Anthropogenic climate change and allergen exposure: the role of plant biology. *Journal of Allergy and Clinical Immunology*, 129(1):27–32, 2012.

[98] Paul B English, Amber H Sinclair, Zev Ross, Henry Anderson, Vicki Boothe, Christine Davis, Kristie Ebi, Betsy Kagey, Kristen Malecki, Rebecca Shultz, et al. Environmental health indicators of climate change for the united states: findings from the state environmental health indicator collaborative. *Environ Health Perspect*, 117(11):1673–81, 2009.

[99] Christine A Rogers, Julie Wolf, Nichole R O'Neill, Christine A Rogers, Michael L Muilenberg, and Lewis H Ziska. Elevated atmospheric carbon dioxide concentrations amplify alternaria alternata sporulation and total antigen production. *Environmental health perspectives*, 118(9):1223–1228, 2010.

[100] L. H. Ziska, P. R. Epstein, and W. H. Schlesinger. Rising CO2, Climate Change, and Public Health: Exploring the Links to Plant Biology. *Environmental Health Perspectives*, 117(2):155–158, Feb 2009.

[101] L. H. Ziska, D. E. Gebhard, D. A. Frenz, S. Faulkner, B. D. Singer, and J. G. Straka. Cities as harbingers of climate change: Common ragweed, urbanization, and public health. *Journal of Allergy and Clinical Immunology*, 111(2):290–295, Feb 2003.

[102] Carsten A. Skjoth, Matt Smith, Branko ?Sikoparija, Alicja Stach, Dorota Myszkowska, Idalia Kasprzyk, Predrag Radisic, Barbara Stjepanovic, Ivana Hrga, Dora Apatini, Donat Magyar, Anna Paldy, and Nicoleta Ianovici. A method for producing airborne pollen source inventories: An example of Ambrosia (ragweed) on the Pannonian Plain. *Agricultural and Forest Meteorology*, 150(9):1203–1210, 2010.

[103] Gbor Csornai, Gbor Mikus, Gizella Ndor, Irn Hubik, Istvn Lszl?, and Zsuzsanna Suba. Integration of hightech components for operating ragweed mapping and control system in hungary using remote sensing and gis. In Georg Gartner and

Felix Ortag, editors, *Cartography in Central and Eastern Europe*, Lecture Notes in Geoinformation and Cartography, pages 405–415. Springer Berlin Heidelberg, 2010.

[104] A. Pauling, M. Rotach, R. Gehrig, B. Clot, and Contributors to the European Aeroallergen Network. A method to derive vegetation distribution maps for pollen dispersion models using birch as an example. *International Journal of Biometeorology*, pages 1–10.

[105] Michael D. Martin, Marcelo Chamecki, and Grace S. Brush. Anthesis synchronization and floral morphology determine diurnal patterns of ragweed pollen dispersal. *Agricultural and Forest Meteorology*, 150(9):1307 – 1317, 2010.

[106] Malgorzata Latalowa, Agnieszka Uruska, Anna Pedziszewska, Malgorzata Gora, and Anna Dawidowska. Diurnal patterns of airborne pollen concentration of the selected tree and herb taxa in gdańsk (northern poland). *Grana*, 44(3):192–201, 2005.

[107] Peer-Gunther von Wahl and Klaus Puls. The emission of mugwort pollen (<i>Artemisia vulgaris</i> L.) and its flight in the air. *Aerobiologia*, 5(1):55–63, 1989.

[108] Isabelle Chuine. A unified model for budburst of trees. *Journal of theoretical biology*, 207(3):337–347, 2000.

[109] Peter JM Van Laarhoven and Emile HL Aarts. *Simulated annealing*. Springer, 1987.

[110] MA Brighetti, C Costa, P Menesatti, F Antonucci, S Tripodi, and A Travaglini. Multivariate statistical forecasting modeling to predict poaceae pollen critical concentrations by meteoclimatic data. *Aerobiologia*, 30(1):25–33, 2014.

[111] Letty A de Weger, Thijs Beerthuizen, Pieter S Hiemstra, and Jacob K Sont. Development and validation of a 5-day-ahead hay fever forecast for patients with

grass-pollen-induced allergic rhinitis. *International journal of biometeorology*, pages 1–9, 2013.

[112] Yaping Shao and Hua Lu. A simple expression for wind erosion threshold friction velocity. *J. Geophys. Res.*, 105(D17):22437–22443, 2000.

[113] John H Seinfeld and Spyros N Pandis. *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons, 2012.

[114] Max D Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[115] Andrea Saltelli, Karen Chan, E Marian Scott, et al. *Sensitivity analysis*, volume 134. Wiley New York, 2000.

[116] Merit Kund, Aivo Vares, Allan Sims, Hardi Tullus, and Veiko Uri. Early growth and development of silver birch (Betula pendula Roth.) plantations on abandoned agricultural land. *European Journal of Forest Research*, 129(4):679–688, 2010.

[117] Patrick Brose, David Van Lear, and Roderick Cooper. Using shelterwood harvests and prescribed fire to regenerate oak stands on productive upland sites. *Forest Ecology and Management*, 113(2-3):125–141, 1999.

[118] M. Chamecki, C. Meneveau, and M. B. Parlange. Large eddy simulation of pollen transport in the atmospheric boundary layer. *Journal of Aerosol Science*, 40(3):241–255, Mar 2009.

[119] Jacob N Barney, Antonio Di Tommaso, and Leslie A Weston. Differences in invasibility of two contrasting habitats and invasiveness of two mugwort artemisia vulgaris populations. *Journal of applied ecology*, 42(3):567–576, 2005.

[120] Boris Fumanal, Bruno Chauvel, and François Bretagnolle. Estimation of pollen and seed production of common ragweed in france. *Annals of Agricultural and Environmental Medicine*, 14(2):233, 2007.

[121] Thomas W Jurik. Population distributions of plant size and light environment of giant ragweed (ambrosia trifida l.) at three densities. *Oecologia*, 87(4):539–550, 1991.

[122] Hong-Jun Li and Zhi-Bin Zhang. Effect of rodents on acorn dispersal and survival of the liaodong oak (¡ i¿ quercus liaotungensis¡/i¿ koidz.). *Forest Ecology and Management*, 176(1):387–396, 2003.

[123] Margaret Mead Foster, Peter M Vitousek, and Polley Ann Randolph. The effects of ragweed (ambrosia artemisiifolia l.) on nutrient cycling in a 1st-year old-field. *American Midland Naturalist*, pages 106–113, 1980.

[124] N. Aboulaich, H. Bouziane, M. Kadiri, M. D. Trigo, H. Riadi, M. Kazzaz, and A. Merzouki. Pollen production in anemophilous species of the Poaceae family in Tetouan (NW Morocco). *Aerobiologia*, 25(1):27–38, Mar 2009.

[125] J. Heiskanen. Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data. *International Journal of Remote Sensing*, 27(6):1135–1158, 2006/03/20 2006.

[126] Y. S. Wang, D. R. Miller, J. M. Welles, and G. M. Heisler. Spatial Variability of Canopy Foliage in an Oak Forest Estimated with Fisheye Sensors. *Forest Science*, 38(4):854–865, Nov 1992.

[127] Stith T. Gower, Chris J. Kucharik, and John M. Norman. Direct and Indirect Estimation of Leaf Area Index, fAPAR, and Net Primary Production of Terrestrial Ecosystems. *Remote Sensing of Environment*, 70(1):29–51, 1999.

[128] W. B. Cohen, T. K. Maiersperger, D. P. Turner, W. D. Ritts, D. Pflugmacher, R. E. Kennedy, A. Kirschbaum, S. W. Running, M. Costa, and S. T. Gower. MODIS land cover and LAI collection 4 product quality across nine sites in the western hemisphere. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(7):1843–1857, 2006.

[129] Margaret Bryan Davis and Linda B. Brubaker. Differential Sedimentation of Pollen Grains in Lakes. *Limnology and Oceanography*, 18(4):635–646, 1973.

[130] M.L. Laucks, G. Roll, G. Schweiger, and E.J. Davis. Physical and chemical (raman) characterization of bioaerosols?pollen. *Journal of Aerosol Science*, 31(3):307 – 319, 2000.

[131] D. Byun and K. L. Schere. Review of the governing equations, computational algorithms, and other components of the models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Reviews*, 59(1-6):51–77, 2006.

[132] Thomas Tomkins Warner. *Numerical Weather And Climate Prediction*, pages 298–299. Cambridge University Press, New York, 1st edition, 2011.

[133] P Rapiejko, W Stanlaewicz, K Szczygielski, and D Jurkiewicz. [threshold pollen count necessary to evoke allergic symptoms]. *Otolaryngologia polska. The Polish otolaryngology*, 61(4):591–594, 2006.

[134] M Thibaudon. [the pollen-associated allergic risk in france]. *European annals of allergy and clinical immunology*, 35(5):170–172, 2003.

[135] YM Noh, H Lee, D Mueller, K Lee, D Shin, S Shin, TJ Choi, YJ Choi, and KR Kim. Investigation of the diurnal pattern of the vertical distribution of pollen in the lower troposphere using lidar. *Atmospheric Chemistry and Physics*, 13(15):7619–7629, 2013.

[136] Clara Deser, Adam S Phillips, Michael A Alexander, and Brian V Smoliak. Projecting north american climate over the next 50 years: Uncertainty due to internal variability*. *Journal of Climate*, 27(6):2271–2296, 2014.

[137] Colin M Beale and Jack J Lennon. Incorporating uncertainty in predictive species distribution modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1586):247–258, 2012.

[138] Helen M Regan, Mark Colyvan, and Mark A Burgman. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological applications*, 12(2):618–628, 2002.

[139] USCB. United States Census 2010. http://www.census.gov/, 2012. Accessed March 28, 2014.

[140] P. G. Georgopoulos, S. W. Wang, V. M. Vyas, Q. Sun, J. Burke, R. Vedantham, T. McCurdy, and H. Ozkaynak. A source-to-dose assessment of population exposures to fine PM and ozone in Philadelphia, PA, during a summer 1999 episode. *Journal of Exposure Analysis and Environmental Epidemiology*, 15(5):439–457, Sep 2005.

[141] Chun Chen, Bin Zhao, and Charles J Weschler. Indoor exposure to "outdoor pm10": Assessing its influence on the relationship between pm10 and short-term mortality in us cities. *Epidemiology*, 23(6):870–878, 2012.

[142] J.D. Spengler, J.F. McCarthy, and J.M. Samet. *Indoor Air Quality Handbook*. Engineering Handbks. Mcgraw-hill, 2000.

[143] R Yankova. Outdoor and indoor pollen grains in sofia. *Grana*, 30(1):171–176, 1991.

[144] Yong Zhang, Sastry Isukapalli, Panos Georgopoulos, and Clifford Weisel. Modeling flight attendants' exposures to pesticide in disinsected aircraft cabins. *Environmental science & technology*, 47(24):14275–14281, 2013.

[145] Linda J Phillips, Robert J Fares, and LG Schweer. Distributions of total skin surface area to body weight ratios for use in dermal exposure assessments. *Journal of exposure analysis and environmental epidemiology*, 3(3):331–338, 1992.

[146] Stephen J Godin, Michael J DeVito, Michael F Hughes, David G Ross, Edward J Scollon, James M Starr, R Woodrow Setzer, Rory B Conolly, and Rogelio Tornero-Velez. Physiologically based pharmacokinetic modeling of deltamethrin:

development of a rat and human diffusion-limited model. *Toxicological Sciences*, page kfq051, 2010.

[147] Heather M Stapleton, Shannon M Kelly, Joseph G Allen, Michael D McClean, and Thomas F Webster. Measurement of polybrominated diphenyl ethers on hand wipes: estimating exposure from hand-to-mouth contact. *Environmental science & technology*, 42(9):3329–3334, 2008.

[148] A. Zellner. *An introduction to Bayesian inference in econometrics*, pages 18–20. Wiley-Interscience, New York, 1st edition, 1971.

[149] Jonathan W. Silvertown. The evolutionary ecology of mast seeding in trees. *Biological Journal of the Linnean Society*, 14(2):235–250, 1980.

[150] Daewon W. Byun. Dynamically Consistent Formulations in Meteorological and Air Quality Models for Multiscale Atmospheric Studies. Part I: Governing Equations in a Generalized Coordinate System. *Journal of the Atmospheric Sciences*, 56(21):3789–3807, 1999/11/01 1999.

[151] D. W. Byun, J. Young, J. Pleim, M. T. Odman, and K. Alapaty. *Numerical Transport Algorithms for the Community Multiscale Air Quality (CMAQ) Chemical Transport Model in Generalized Coordinates*, chapter 7, pages 1–55. National Exposure Research Laboratory, U.S. EPA, 1999.

[152] J. Smagorinsky. General circulation experiments with the primitive equations. *Monthly Weather Review*, 91(3):99–164, 1963/03/01 1963.

[153] J. E. Pleim. A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model. *Journal of Applied Meteorology and Climatology*, 46(9):1396–1409, Sep 2007.

[154] M.L. Wesely. Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmospheric Environment (1967)*, 23(6):1293 – 1304, 1989.

Appendix A

LIST OF ACRONYMS

AAAAI	American Academy of Allergy, Asthma & Immunology
AAD	Allergic Airway Diseases
AM	Annual Mean
AMO	Atlantic Multidecadal Oscillation
AOGCM	Atmosphere-Ocean General Circulation Model
AP	Annual Production
BC	Boundary Condition
BELD3.1	Biogenic Emissions Landuse Database, version 3.1
BEIS	Biogenic Emission Inventory System
CCSM	Community Climate System Model
CCTM	CMAQ Chemical Transport Model
CHAD	Consolidated Human Activity Database
CMAQ	Community Multiscale Air Quality
DOE	Department of Energy
ENSO	El Nino Southern Oscillation
GCM	General Circulation Model
GDD	Growing Degree Days
GDH	Growing Degree Hours
IC	Initial Condition
ID	Initial Date
IPCC	Intergovernmental Panel on Climate Change
JPROC	Photolysis rate preprocessor
LAI	Leaf Area Index
LULC	Land Use and Land Coverage
M3TOOL	Models-3 Tools
MCIP3.6	Meteorology-Chemistry Interface Processor
MENTOR	Modeling Environment for Total Risk studies
MSE	Mean Square Error
NAB	National Allergy Bureau
NARCCAP	North American Regional Climate Change Assessment Program
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NJDEP	New Jersey Department of Environmental Protection
NOAA	National Oceanic and Atmospheric Administration
OTC	Ozone Transport Commission
PM	Particulate Matter

PD	Peak Date
PDF	Probability Distribution Function
PV	Peak Value
RCM	Regional Climate Model
SAT	Surface Air Temperature
SD	Start Date
SL	Season Length
SMOKE	Sparse Matrix Operator Kernel Emissions
SRES	Special Report on Emissions Scenarios
WRF	Weather Research and Forecasting model
NCAR	National Center for Atmospheric Research
USGS	U.S. Geological Survey
USEPA	US Environmental Protection Agency

Appendix B

SUPPLEMENTARY DATA AND FORMULAS FOR ANALYSIS OF HISTORICAL OBSERVATIONS

B.1 Pollen Monitoring Stations

Table B.1 lists the Coordinates, elevations, main climate characteristics and years of data for the studied stations.

B.2 Definition of start and end dates

The definitions of the start and end dates were demonstrated using the observed daily pollen count during the allergenic pollen season in 2010 at the monitor station in Springfield, New Jersey (B.1).

B.3 Bayesian Analysis

Equation B.1 is the likelihood function of the Bayesian model.

$$f(Y, X|\beta, \sigma^2) = (2\Pi\sigma^2)^{-\frac{n}{2}} exp[-\frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)] \tag{B.1}$$

Zellner's informative G-priors[148] are assumed for β and σ^2 as shown in equation B.2,

$$\begin{cases} (\beta|\sigma^2, X) \sim N_{k+1}(\widetilde{\beta}, c\sigma^2(X^TX)^{-1}) \\ \Pi(\sigma^2|X) \propto \sigma^{-2} \end{cases} \tag{B.2}$$

where $\widetilde{\beta}$ and c are further assumed to be $\mathbf{0}_{k+1}$ and 100 respectively so that parameterizations are mainly dependent on the explanatory matrix X. In this study $c = 100$, the prior gets a weight corresponding to 1% of the sample.

Variable Selection:

Multiple climatic factors were first prescreened by regressing each individual pollen index against each individual climatic factor of a given month for historical data of twenty

Table B.1: Coordinates, elevations, main climate characteristics and years of data for the studied stations.

ID	Station Name	Latitude (°N)	Longitude (°W)	Elevation (m)	Mean Temp. (°C)	Annual Precip. (mm)	Years of Data (Yrs.)				
							B	O	R	M	G
1	Seattle, WA	47.66	122.29	20	11.9	603	13	13	-	4	14
2	Fargo, ND	46.84	96.87	277	5.9	569	11	11	12	7	12
3	Vancouver, WA	45.62	122.50	89	12.3	960	7	4	-	-	6
4	Eugene, OR	44.04	123.09	129	11.3	1065	8	13	-	-	12
5	LaCrosse, WI	43.88	91.19	216	9.0	905	10	10	9	-	10
6	Rochester, NY	43.10	77.58	148	9.3	878	14	14	14	7	13
7	**Niagara Falls, ON , CA**	43.09	79.09	188	9.3	893	-	-	4	-	-
8	Madison, WI	43.08	89.43	263	8.7	909	7	7	7	7	7
9	Waukesha, WI	43.02	88.24	270	9.6	557	4	6	6	4	6
10	**London, ON, CA**	42.99	81.25	250	8.3	476	-	4	4	-	4
11	Albany, NY	42.68	73.77	72	9.4	992	4	5	-	-	-
12	Chelmsford, MA	42.60	71.35	37	10.0	814	9	9	8	5	9
13	St. Clair Shores, MI	42.51	82.9	180	9.8	863	7	6	7	-	7
14	Salem, MA	42.50	70.92	42	10.9	1082	9	10	10	10	10
15	Erie, PA	42.10	80.13	215	10.1	1002	12	9	13	6	12
16	Olean, NY	42.09	78.43	433	7.3	974	8	8	14	-	7
17	Chicago, IL	41.91	87.77	189	11.0	617	7	7	7	6	7
18	Waterbury, CT	41.55	73.07	140	11.8	665	10	10	10	8	8
19	Omaha, NE	41.14	95.97	305	10.9	854	7	12	13	4	12
20	Armonk, NY	41.13	73.73	187	11.1	865	6	7	7	6	7
21	Lincoln, NE	40.82	96.64	371	11.0	699	4	4	5	4	4
22	Springfield, NJ	40.74	74.19	43	13.0	1213	10	10	13	7	9
23	Pittsburgh, PA	40.47	79.95	287	11.2	858	5	5	7	-	5
24	Philadelphia, PA	39.96	75.16	12	13.5	1106	11	11	10	6	11
25	York, PA	39.94	76.71	195	13.0	948	6	6	7	4	-
26	Cherry Hill, NJ	39.94	74.91	13	12.7	550	13	13	14	7	12
27	Indianapolis, IN	39.91	86.2	254	12.0	1095	7	11	11	4	11
28	New Castle, DE	39.66	75.57	3	13.5	1106	4	4	5	5	4
29	Reno, NV	39.56	119.77	1382	12.1	195	8	6	-	4	4
30	Baltimore, MD	39.37	76.47	36	13.3	1117	8	6	10	-	-
31	Kansas City, MO	39.08	94.58	288	13.9	750	7	8	8	-	8
32	Colorado Springs 2, CO	38.87	104.83	1868	9.6	372	-	-	-	5	4
33	Colorado Springs 1, CO	38.87	104.82	1867	9.8	346	-	5	4	4	6
34	Roseville, CA	38.76	121.27	57	17.0	637	7	10	-	-	10
35	Lexington, KY	38.04	84.5	299	13.1	1225	5	8	9	4	8
36	Pleasanton, CA	37.69	121.91	100	14.2	256	10	10	-	-	13
37	San Jose 1, CA	37.33	121.94	35	15.7	234	7	10	-	-	10
38	San Jose 2, CA	37.31	121.97	47	15.7	234	4	6	-	-	6
39	Las Vegas, NV	36.17	115.15	620	20.9	105	-	-	-	-	5
40	Durham, NC	36.05	78.9	110	15.7	1160	9	9	8	5	9
41	Tulsa 1, OK	36.03	95.87	207	16.2	1072	9	4	5	-	-
42	Knoxville, TN	35.95	84.01	305	15.0	1285	9	13	12	-	13
43	Los Alamos, NM	35.88	106.32	2227	11.8	323	4	6	-	6	5
44	Oklahoma City, OK	35.61	97.6	340	15.9	886	-	7	6	-	7
45	Fort Smith, AR	35.35	94.39	186	16.5	1149	6	4	-	-	8
46	Charlotte, NC	35.3	80.75	229	16.0	1097	8	8	7	-	13
47	Little Rock, AR	34.75	92.39	115	17.3	1198	6	8	8	-	9
48	Huntsville, AL	34.73	86.59	191	16.3	1325	12	12	13	4	14
49	Santa Barbara, CA	34.44	119.76	57	14.9	354	-	7	-	-	8
50	Atlanta, GA	33.97	84.55	366	16.8	1286	14	14	-	-	-
51	Orange, CA	33.78	117.86	53	17.9	170	-	4	-	-	-
52	Dallas, TX	33.04	96.83	207	19.3	912	-	7	7	-	-
53	Waco, TX	31.51	97.2	185	19.4	945	-	4	-	-	5
54	Georgetown, TX	30.64	97.76	269	20.3	1009	6	7	7	-	7
55	College Station, TX	30.64	96.31	91	19.5	509	6	10	10	4	9
56	Tallahassee, FL	30.44	84.28	62	19.7	1478	4	6	6	-	6
57	Tampa, FL	28.06	82.43	12	22.7	1101	-	7	-	-	8
58	Corpus Christi, TX	27.80	97.4	2	22.2	794	-	7	6	-	7

B, Birch; O, Oak; R, Ragweed; M, Mugwort; G: Grass

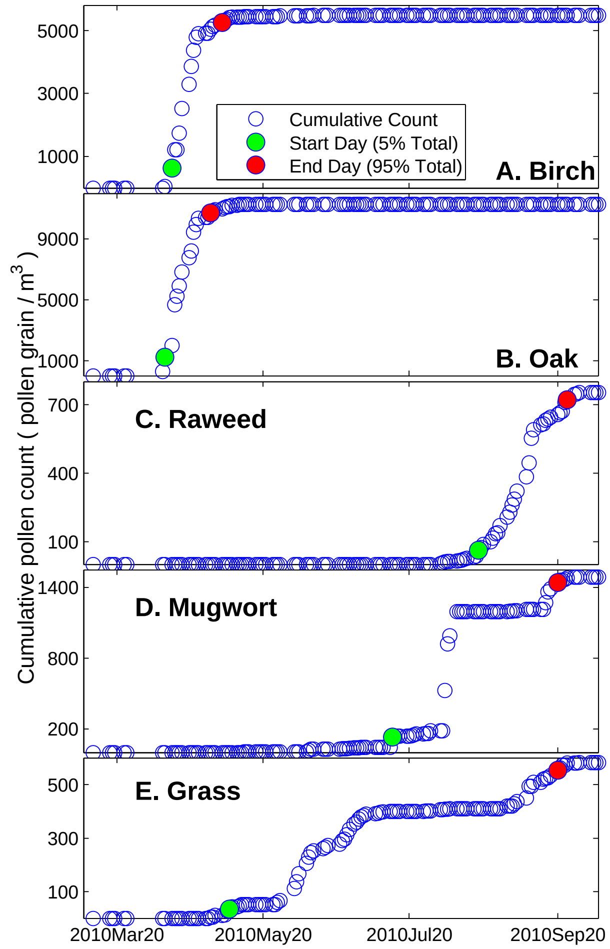

Figure B.1: Definition of start and end dates. With day 1 being January 1st, the SD (days from January 1st) of a pollen season is defined as the day when the cumulative pollen count reaches 5% and the end date is the day when it reaches 95% of annual total count. Pollen data are from the monitoring station at Springfield, NJ.

years. Climatic factors in two periods influence the pollen indices[73]: (1) initiation of flower primordial during the burst period in spring and early summer of the current year; and (2) development of flower inflorescences in autumn and winter of the previous year. In this study, monthly climatic factors for CO_2, temperature, precipitation, cloud coverage, and sunshine hours in June to December of previous year and January to May of current year were taken into account in the prescreening stage. First, multiple monthly climatic factors were consecutively screened starting from the smallest P value and the largest R^2; then monthly climatic factors in consecutive months were lumped together to form nine preselected variables for each pollen index.

The preselected climatic variables were further selected and assessed by calculating the probability of each sub-model and the probability of inclusion of each variable in the full model.

Calculation of sub-model probability is obtained through equation B.3,

$$\begin{aligned}\Pi(\gamma|Y,X) \propto &(c+1)^{-\frac{(q_\gamma+1)}{2}}[Y^TY - \frac{c}{c+1}Y^TX_\gamma(X_\gamma^TX_\gamma)^{-1}X_\gamma^TY \\ &- \frac{1}{c+1}\widetilde{\beta}_\gamma^TX_\gamma^TX_\gamma\widetilde{\beta}_\gamma]^{-\frac{n}{2}}\end{aligned} \tag{B.3}$$

where binary indicator vector $\gamma \in \Gamma = \{0,1\}^k$, γ_i=1 means variable x_i is included in the model while γ_i=0 means x_i not included in the model; β_γ, X_γ, q_γ are sub-vectors, sub-matrix and number of variables in the sub-model, respectively.

A Gibbs sampling algorithm, as shown in the following, was used to calculate inclusion probabilities. It is a Markov chain, and after a large number of iterations, its output can be used to approximate the posterior probabilities $P(\gamma_i = 1|Y,X)$ based on the Monte Carlo method in the form of equation B.4,

$$\widehat{P}(\gamma_i = 1|Y,X) = \frac{1}{T - T_0 + 1}\sum_{t=T_0}^{T} I_{\gamma_i^{(t)}=1} \tag{B.4}$$

where T is the number of total iterations, and T_0 is the "burn-in" period, such that the first T_0 values are eliminated to guarantee convergence. In this work, T was set to be 20,000 and T_0 to be 10,000.

Initialization: draw γ^0 from the uniform distribution on Γ.
(1). draw $\gamma_1^{(t)}$ according to $\Pi(\gamma_1|Y,\gamma_2^{(t-1)},\cdots,\gamma_k^{(t-1)},X)$,
(2). draw $\gamma_2^{(t)}$ according to $\Pi(\gamma_1|Y,\gamma_1^{(t)},\gamma_3^{(t-1)},\cdots,\gamma_k^{(t-1)},X)$,
$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$
(k). draw $\gamma_k^{(t)}$ according to $\Pi(\gamma_1|Y,\gamma_1^{(t)},\cdots,\gamma_{k-1}^{(t)},\gamma_k^{(t-1)},X)$,

Parameterization:

The corresponding Bayesian estimator of expectations of β and σ^2 are presented in equations B.5 and B.6:

$$E(\beta|Y,X) = \frac{\widetilde{\beta} + c\widehat{\beta}}{c+1} \tag{B.5}$$

$$E(\sigma^2|Y,X) = \frac{s^2 + (\widetilde{\beta} - \widehat{\beta})^T X^T X(\widetilde{\beta} - \widehat{\beta})/(c+1)}{n-2} \tag{B.6}$$

P values were calculated based on F-statistics. The highest posterior density regions (HPD) in Bayesian statistics are the sections of the parameter space where the parameters most likely take values. HPD of β were calculated to characterize the regions of most probable variations of predicted pollen indices. A Bayes factor[77] B_{10}^{Π} was constructed through null hypothesis $H_0 : \beta_i = 0$.

Prediction:

The future vector $\widetilde{Y}$ based on the posterior and future explanatory matrix $\widetilde{X}$ has a Gaussian distribution[76] and its expectation can be predicted by equation B.7:

$$E(\widetilde{Y}|\sigma^2, Y, X, \widetilde{X}) = \widetilde{X}\frac{\widetilde{\beta} + c\widehat{\beta}}{c+1} \tag{B.7}$$

The pollen indices for base year (2000) were obtained by averaging over the corresponding five year overlapping means of pollen indices from Basel (Switzerland), Turku (Finland), and New Jersey and North Dakota (US). These locations span different climate zones, geographical regions and forest vegetations in the northern hemisphere. Birch pollen levels from these locations are sufficiently representative to generate and analyze future plausible pollen indices and their mean trends.

The intrinsic inter-annual variation of the pollen index has been observed by many researchers[149]. According to[73], the mast and sparse years occur alternately due to evolution stress. As for the intrinsic inter-annual variation of pollen indices in the current study, the data from[32] were first normalized using their mean values, and then fit using equation B.8,

$$Y_{NP}(i) = P_1 i + P_2 \sin(P_3 i) + P_4 \tag{B.8}$$

where $Y_{NP}(i)$ is the normalized pollen index in year i, and parameters P_1, P_2, P_3 and P_4 are to be determined. The first and fourth terms describe the mean trends of the pollen index. The second and third terms in equation B.8 characterize the intrinsic inter-annual variation of pollen indices and are used to simulate the fluctuations around the mean trends obtained through equation B.7.

B.4 Mean and Standard Deviations of Pollen Indices

Tables B.2 and B.3 list the mean pollen indices (average over 1994-2010) and their standard deviations at the studied stations.

B.5 Normalized Semi-variogram

Figure B.2 displays the Normalized semi-variogram at different spatial lags for mean pollen indices during the periods of 1994-2000 and 2001-2010. Mean pollen indices during periods of 1994-2000 and 2001-2010 were first normalized using overall mean pollen indices over the entire observation time of 1994-2010; the variogram calculated using a normalized pollen index was further normalized using the largest variogram of that pollen index.

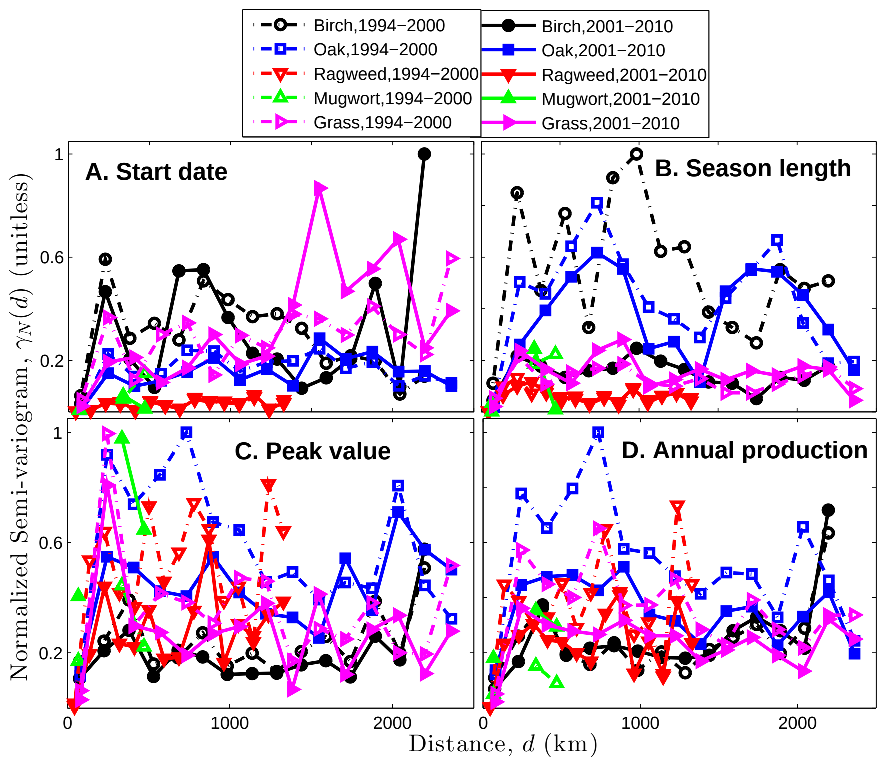

Figure B.2: Normalized semi-variogram for mean pollen indices during the periods of 1994-2000 and 2001-2010. Smaller variogram indicates higher similarity or synchronization among mean pollen indices from different regions; suggesting pollen seasons in different regions appear to synchronize their start times, duration and pollen production. (A) Start Date; (B) Season Length; (C) Peak Value; and (D) Annual Production.

Table B.2: Mean and standard deviation of observed start date and season length of allergenic pollen season at the studied stations during 1994-2010.

ID	Start Date (mean ± std), Days from Jan 1st					Season Length (mean ± std), Days				
	Birch	Oak	Ragweed	Mugwort	Grass	Birch	Oak	Ragweed	Mugwort	Grass
1	85 ± 8	113 ± 7		209 ± 10	135 ± 10	25 ± 9	16 ± 4		20 ± 15	71 ± 9
2	118 ± 8	123 ± 7	225 ± 6	213 ± 4	155 ± 6	21 ± 9	21 ± 7	27 ± 7	45 ± 8	60 ± 19
3	53 ± 23	105 ± 6			136 ± 8	45 ± 19	54 ± 12			69 ± 8
4	74 ± 17	112 ± 9			150 ± 6	35 ± 14	28 ± 12			39 ± 7
5	113 ± 8	120 ± 10	226 ± 3	238 ± 5	146 ± 8	24 ± 8	25 ± 10	34 ± 4	26 ± 1	69 ± 26
6	116 ± 7	124 ± 11	229 ± 3	238 ± 8	148 ± 7	31 ± 10	24 ± 10	36 ± 7	32 ± 7	56 ± 12
7	117 ± 4	119 ± 11	228 ± 3		147 ± 8	32 ± 9	33 ± 8	38 ± 3		101 ± 8
8	103 ± 12	121 ± 8	223 ± 3	222 ± 18	149 ± 5	36 ± 10	19 ± 8	38 ± 5	52 ± 22	72 ± 14
9	116 ± 8	122 ± 10	224 ± 4	229 ± 18	146 ± 4	27 ± 4	21 ± 9	45 ± 9	40 ± 19	68 ± 15
10	123 ± 12	113 ± 2	232 ± 3		154 ± 6	29 ± 10	40 ± 7	42 ± 9		70 ± 17
11	116 ± 3	128 ± 4	228 ± 3			35 ± 8	20 ± 3	29 ± 3		
12	120 ± 6	131 ± 6	231 ± 10	235 ± 9	130 ± 35	27 ± 6	20 ± 4	50 ± 11	43 ± 7	98 ± 26
13	111 ± 12	120 ± 16	231 ± 3	239 ± 2	146 ± 7	32 ± 10	26 ± 13	33 ± 2	25 ± 1	67 ± 29
14	117 ± 7	132 ± 9	233 ± 6	240 ± 5	150 ± 9	28 ± 6	17 ± 6	37 ± 5	27 ± 6	116 ± 10
15	113 ± 8	126 ± 13	232 ± 5	233 ± 9	137 ± 7	32 ± 8	19 ± 9	37 ± 7	43 ± 21	95 ± 23
16	114 ± 18	121 ± 12	230 ± 5		144 ± 9	29 ± 19	34 ± 12	41 ± 14		57 ± 7
17	99 ± 5	102 ± 6	227 ± 4	232 ± 6	132 ± 9	48 ± 11	42 ± 10	38 ± 7	35 ± 10	97 ± 22
18	115 ± 7	119 ± 9	230 ± 3	234 ± 17	145 ± 21	24 ± 8	30 ± 16	38 ± 6	29 ± 12	96 ± 22
19	102 ± 7	113 ± 9	235 ± 8	236 ± 16	142 ± 7	58 ± 9	25 ± 9	41 ± 8	54 ± 23	119 ± 8
20	114 ± 3	119 ± 9	227 ± 2	244 ± 9	138 ± 8	20 ± 7	19 ± 5	41 ± 8	37 ± 14	120 ± 12
21	95 ± 15	101 ± 15	239 ± 11	239 ± 13	136 ± 5	28 ± 14	26 ± 16	43 ± 14	41 ± 16	133 ± 9
22	116 ± 7	116 ± 8	232 ± 3	244 ± 22	136 ± 9	24 ± 7	25 ± 9	42 ± 6	34 ± 16	126 ± 12
23	108 ± 8	127 ± 14	231 ± 5		143 ± 9	34 ± 11	21 ± 10	37 ± 8		65 ± 10
24	103 ± 9	111 ± 8	231 ± 5	230 ± 18	111 ± 6	38 ± 11	35 ± 8	46 ± 10	54 ± 17	108 ± 40
25	93 ± 16	108 ± 6	229 ± 4	217 ± 14		46 ± 13	39 ± 10	37 ± 8	54 ± 14	
26	106 ± 8	111 ± 6	230 ± 5	236 ± 8	112 ± 6	30 ± 7	29 ± 9	51 ± 9	56 ± 11	134 ± 23
27	94 ± 8	116 ± 7	231 ± 3	242 ± 14	137 ± 7	36 ± 15	19 ± 5	34 ± 4	26 ± 14	89 ± 25
28	102 ± 11	103 ± 9	230 ± 3	248 ± 15	134 ± 6	41 ± 14	35 ± 9	47 ± 7	26 ± 15	123 ± 8
29	93 ± 18	98 ± 30	270 ± 0	262 ± 20	131 ± 7	33 ± 21	39 ± 17	47 ± 0	31 ± 15	96 ± 46
30	95 ± 15	109 ± 3	237 ± 3		135 ± 6	29 ± 11	25 ± 9	37 ± 5		120 ± 5
31	91 ± 4	99 ± 7	238 ± 3	216 ± 0	120 ± 12	22 ± 13	23 ± 6	34 ± 6	59 ± 0	128 ± 19
32	111 ± 5	131 ± 26	216 ± 7	221 ± 8	144 ± 19	26 ± 0	39 ± 21	58 ± 12	64 ± 7	111 ± 16
33	101 ± 0	109 ± 20	214 ± 9	214 ± 8	150 ± 8	27 ± 0	52 ± 13	54 ± 4	64 ± 9	108 ± 11
34	63 ± 21	81 ± 7	263 ± 59		93 ± 15	43 ± 25	36 ± 12	33 ± 25		130 ± 41
35	91 ± 7	104 ± 12	233 ± 7	234 ± 20	131 ± 8	40 ± 17	26 ± 16	38 ± 8	44 ± 21	101 ± 29
36	76 ± 7	79 ± 7	251 ± 0		103 ± 17	34 ± 6	43 ± 18	76 ± 0		82 ± 20
37	81 ± 12	82 ± 9	249 ± 14	266 ± 20	107 ± 10	29 ± 9	50 ± 12	39 ± 11	58 ± 8	83 ± 23
38	79 ± 8	85 ± 11			114 ± 24	36 ± 15	47 ± 18			112 ± 38
39		91 ± 9			97 ± 14		52 ± 6			169 ± 52
40	90 ± 6	92 ± 8	234 ± 4	246 ± 11	108 ± 9	32 ± 13	23 ± 9	61 ± 8	33 ± 16	142 ± 15
41	84 ± 15	84 ± 6	242 ± 6	247 ± 9	98 ± 26	37 ± 14	24 ± 10	39 ± 7	37 ± 10	156 ± 10
42	86 ± 11	92 ± 13	224 ± 21		122 ± 14	41 ± 20	34 ± 13	48 ± 14		113 ± 35
43	89 ± 18	111 ± 13	232 ± 35	241 ± 17	130 ± 20	49 ± 21	32 ± 14	53 ± 18	35 ± 12	119 ± 38
44	90 ± 3	87 ± 11	244 ± 6	245 ± 4	91 ± 11	44 ± 10	30 ± 13	47 ± 4	61 ± 12	180 ± 18
45	67 ± 21	88 ± 3	255 ± 20			58 ± 12	23 ± 7	39 ± 7		
46	83 ± 13	89 ± 11	231 ± 5		117 ± 9	47 ± 13	28 ± 10	59 ± 6		124 ± 18
47	78 ± 6	86 ± 8	242 ± 6	245 ± 20	101 ± 13	22 ± 9	24 ± 9	45 ± 12	28 ± 14	145 ± 22
48	81 ± 10	91 ± 10	244 ± 4	252 ± 8	104 ± 11	33 ± 18	23 ± 8	42 ± 7	24 ± 10	155 ± 24
49	60 ± 0	67 ± 10		236 ± 3	100 ± 25	32 ± 0	63 ± 13		67 ± 7	95 ± 50
50	75 ± 11	88 ± 9		286 ± 0	95 ± 8	47 ± 17	25 ± 9		19 ± 0	169 ± 16
51		89 ± 25	264 ± 0		73 ± 33		50 ± 25	64 ± 0		231 ± 53
52	94 ± 9	79 ± 6	255 ± 9			9 ± 1	42 ± 9	45 ± 9		
53		69 ± 8	245 ± 10		110 ± 16		69 ± 4	65 ± 7		89 ± 11
54	71 ± 2	83 ± 5	254 ± 6		84 ± 30	23 ± 15	25 ± 3	39 ± 6		215 ± 25
55	68 ± 11	69 ± 7	260 ± 5	252 ± 7	83 ± 13	31 ± 11	26 ± 8	35 ± 6	56 ± 7	188 ± 21
56	59 ± 5	71 ± 7	251 ± 24		100 ± 10	44 ± 8	35 ± 10	56 ± 17		193 ± 10
57	45 ± 4	56 ± 7		260 ± 0	101 ± 34	55 ± 1	49 ± 15		47 ± 0	203 ± 43
58	55 ± 9	75 ± 4	266 ± 8		88 ± 17	50 ± 8	30 ± 7	37 ± 7		202 ± 57

Table B.3: Mean and standard deviation of observed seasonal total count and maximum daily count at the studied stations during 1994-2010.

ID	Annual Production (mean ± std), pollen/m^3					Peak Value (mean ± std), pollen/m^3				
	Birch	Oak	Ragweed	Mugwort	Grass	Birch	Oak	Ragweed	Mugwort	Grass
1	2834 ± 1175	602 ± 261		5 ± 3	668 ± 146	807 ± 448	135 ± 35		2 ± 1	54 ± 13
2	1055 ± 565	1993 ± 1130	3135 ± 2153	323 ± 173	723 ± 269	341 ± 253	651 ± 459	567 ± 410	37 ± 15	106 ± 73
3	1732 ± 1198	168 ± 30			2750 ± 780	502 ± 374	44 ± 6			289 ± 114
4	716 ± 633	539 ± 501			6268 ± 4347	190 ± 138	133 ± 120			658 ± 225
5	1016 ± 584	2531 ± 760	1742 ± 595	57 ± 10	321 ± 172	258 ± 130	727 ± 249	225 ± 61	14 ± 1	48 ± 24
6	2018 ± 2128	2411 ± 2259	2264 ± 1178	424 ± 142	1269 ± 723	429 ± 349	550 ± 418	275 ± 93	94 ± 37	180 ± 95
7	542 ± 236	1076 ± 169	1012 ± 174		1272 ± 332	101 ± 59	253 ± 91	158 ± 120		106 ± 38
8	1075 ± 482	3278 ± 1326	1686 ± 267	45 ± 28	332 ± 85	261 ± 107	843 ± 399	211 ± 64	6 ± 3	37 ± 14
9	614 ± 434	2329 ± 1186	968 ± 761	34 ± 14	340 ± 197	164 ± 84	739 ± 281	180 ± 97	10 ± 2	72 ± 33
10	655 ± 403	1027 ± 609	2077 ± 801		754 ± 272	92 ± 47	147 ± 75	201 ± 66		97 ± 45
11	1260 ± 576	7139 ± 4299	890 ± 70			284 ± 47	1278 ± 508	99 ± 6		
12	2842 ± 1747	5310 ± 1744	546 ± 173	533 ± 349	513 ± 182	665 ± 387	1130 ± 380	56 ± 19	97 ± 113	57 ± 67
13	749 ± 234	1442 ± 965	1574 ± 394	54 ± 31	384 ± 136	165 ± 60	349 ± 161	236 ± 70	10 ± 5	58 ± 28
14	1660 ± 1120	3058 ± 1229	314 ± 112	493 ± 232	441 ± 116	369 ± 235	762 ± 325	45 ± 15	88 ± 34	69 ± 44
15	1287 ± 1223	2953 ± 1484	1036 ± 291	76 ± 27	777 ± 362	296 ± 248	1099 ± 689	136 ± 44	20 ± 12	100 ± 63
16	1996 ± 1169	1618 ± 676	278 ± 187		490 ± 147	620 ± 369	411 ± 263	54 ± 34		71 ± 17
17	761 ± 674	995 ± 720	896 ± 604	72 ± 18	693 ± 311	126 ± 100	155 ± 113	101 ± 71	11 ± 4	42 ± 10
18	8180 ± 7731	5491 ± 5087	478 ± 191	79 ± 41	136 ± 52	1947 ± 1727	1503 ± 1462	60 ± 20	23 ± 9	18 ± 11
19	1096 ± 624	1298 ± 642	4181 ± 1512	133 ± 84	1012 ± 355	157 ± 41	268 ± 105	369 ± 135	12 ± 5	88 ± 35
20	13501 ± 9087	13695 ± 5761	319 ± 127	128 ± 56	417 ± 235	4296 ± 3658	4834 ± 4649	53 ± 25	42 ± 21	47 ± 21
21	102 ± 16	4047 ± 3516	2159 ± 1610	76 ± 37	590 ± 197	33 ± 16	1507 ± 1252	250 ± 177	12 ± 6	52 ± 24
22	5520 ± 3253	9271 ± 4074	1108 ± 585	1021 ± 449	999 ± 359	1162 ± 852	1471 ± 583	141 ± 91	273 ± 145	124 ± 67
23	660 ± 464	3258 ± 1645	841 ± 420		513 ± 151	132 ± 86	1227 ± 798	162 ± 67		69 ± 22
24	1152 ± 1079	4327 ± 2913	736 ± 389	193 ± 110	699 ± 397	226 ± 236	816 ± 650	88 ± 34	70 ± 52	63 ± 22
25	1097 ± 842	4301 ± 1553	1232 ± 758	104 ± 60		276 ± 297	540 ± 180	137 ± 66	12 ± 6	
26	1298 ± 1147	7133 ± 4859	769 ± 409	80 ± 17	1030 ± 435	272 ± 193	1409 ± 1085	77 ± 30	23 ± 14	73 ± 36
27	285 ± 167	1860 ± 1476	4650 ± 2252	83 ± 46	591 ± 262	97 ± 48	457 ± 307	576 ± 165	59 ± 44	78 ± 25
28	1324 ± 1060	3794 ± 455	618 ± 192	164 ± 56	1905 ± 388	461 ± 509	1117 ± 281	138 ± 83	53 ± 29	661 ± 199
29	144 ± 113	847 ± 1530	2253 ± 0	331 ± 283	246 ± 163	77 ± 60	297 ± 538	365 ± 0	98 ± 84	31 ± 12
30	561 ± 481	2714 ± 1654	566 ± 192		734 ± 24	96 ± 74	464 ± 250	71 ± 23		67 ± 36
31	887 ± 703	7058 ± 5936	9863 ± 9359	231 ± 0	3635 ± 2996	366 ± 302	1600 ± 1533	2046 ± 2683	32 ± 0	416 ± 261
32	72 ± 29	572 ± 498	1002 ± 217	2087 ± 1012	2037 ± 1413	14 ± 2	61 ± 38	123 ± 40	148 ± 47	81 ± 42
33	139 ± 0	617 ± 508	488 ± 109	705 ± 312	706 ± 394	18 ± 0	102 ± 45	54 ± 18	57 ± 22	47 ± 25
34	57 ± 21	642 ± 508	25 ± 13		464 ± 361	27 ± 18	229 ± 216	7 ± 3		80 ± 52
35	188 ± 269	1893 ± 2481	2002 ± 1873	6 ± 1	917 ± 641	70 ± 109	370 ± 478	282 ± 230	2 ± 1	133 ± 93
36	294 ± 214	2829 ± 2095	3 ± 0		813 ± 354	76 ± 52	862 ± 803	1 ± 0		128 ± 75
37	156 ± 97	1937 ± 1419	26 ± 30	10 ± 3	757 ± 1070	50 ± 36	483 ± 320	15 ± 18	4 ± 2	117 ± 148
38	979 ± 1162	2689 ± 1996			2144 ± 1017	259 ± 259	499 ± 385			265 ± 173
39		263 ± 127			307 ± 230		37 ± 15			20 ± 14
40	839 ± 393	7504 ± 4964	366 ± 272	7 ± 5	439 ± 250	170 ± 81	1835 ± 1141	43 ± 36	2 ± 1	50 ± 29
41	3653 ± 2549	15626 ± 8670	8832 ± 3743	106 ± 48	1837 ± 863	800 ± 473	5998 ± 5434	1184 ± 819	24 ± 4	142 ± 58
42	997 ± 1101	1635 ± 1309	547 ± 379		1610 ± 1911	335 ± 460	427 ± 391	127 ± 117		166 ± 200
43	6 ± 5	454 ± 162	243 ± 175	1558 ± 1701	270 ± 108	4 ± 5	187 ± 128	45 ± 37	236 ± 244	42 ± 10
44	980 ± 120	2868 ± 1612	3570 ± 2822	196 ± 54	1973 ± 675	251 ± 35	825 ± 675	444 ± 300	39 ± 22	170 ± 51
45	836 ± 575	4345 ± 2921	1834 ± 1205			140 ± 78	1107 ± 163	269 ± 192		
46	1437 ± 849	12233 ± 7994	1886 ± 620		1207 ± 305	216 ± 121	2678 ± 1927	182 ± 80		128 ± 55
47	375 ± 253	6456 ± 4239	871 ± 428	29 ± 23	446 ± 237	159 ± 106	1386 ± 938	139 ± 80	14 ± 13	53 ± 26
48	1358 ± 1576	4172 ± 2397	945 ± 266	39 ± 16	783 ± 268	442 ± 495	1066 ± 602	141 ± 40	14 ± 7	116 ± 101
49	7 ± 0	1056 ± 552		449 ± 156	616 ± 634	5 ± 0	213 ± 122		41 ± 14	70 ± 38
50	928 ± 549	11663 ± 4778		3 ± 0	624 ± 222	165 ± 125	2355 ± 1323		1 ± 0	56 ± 29
51		951 ± 583	24 ± 0		530 ± 258		120 ± 93	11 ± 0		30 ± 17
52	2793 ± 1048	2344 ± 502	7851 ± 2735			785 ± 23	481 ± 544	676 ± 153		
53		26968 ± 12960	19110 ± 3986		16170 ± 4313		974 ± 311	856 ± 207		509 ± 189
54	921 ± 870	36066 ± 6305	6255 ± 2172		2079 ± 1123	470 ± 544	5440 ± 2044	655 ± 218		121 ± 31
55	182 ± 52	13474 ± 6355	5738 ± 1630	33 ± 11	1301 ± 465	55 ± 26	2479 ± 1110	669 ± 285	13 ± 7	109 ± 43
56	1389 ± 1567	25830 ± 25080	1832 ± 1389		712 ± 453	162 ± 149	2713 ± 2544	139 ± 112		37 ± 26
57	814 ± 547	21989 ± 18927		4 ± 0	460 ± 224	85 ± 56	2468 ± 1772		2 ± 0	24 ± 12
58	556 ± 381	11743 ± 3964	1958 ± 1205		590 ± 269	88 ± 71	1579 ± 865	272 ± 170		22 ± 6

B.6 Structures of Machine Learning Models

Figures B.3-B.6 present the structures of machine learning models.

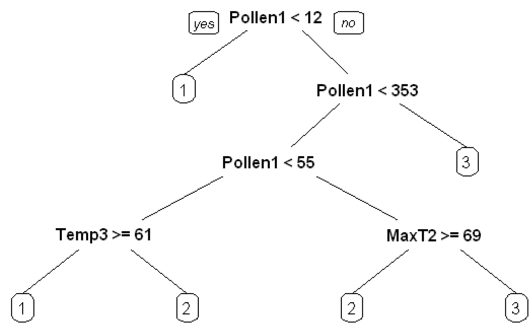

Figure B.3: Model structure of decision tree for prediction of daily airborne pollen levels. The nodes represent the conditions and input variables in Table 2.2; the leaves represent airborne pollen levels.

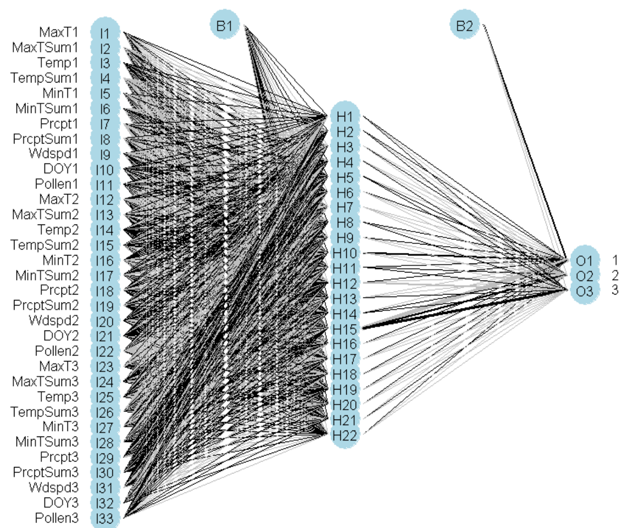

Figure B.4: Model structure of neural network for prediction of daily airborne pollen levels. The first layer is the input layer consisting of input variables in Table 2.2. The second layer is the hidden layer. The third layer is the output layer containing three airborne pollen levels. The B1 and B2 are auxiliary units.

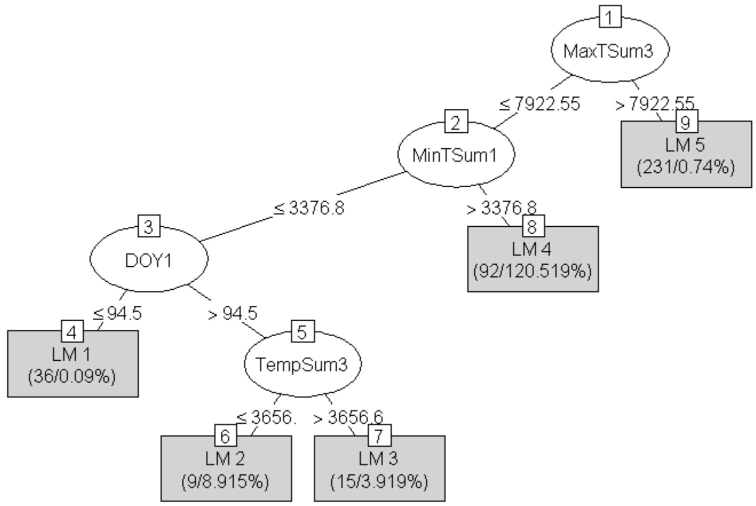

Figure B.5: Model structure of regression tree for prediction of daily airborne pollen concentrations. The nodes represent the conditions and input variables in Table 2.2. The leaves represent stagewise linear models. Under each leaf, the first number is the number of training instances falling into this leaf and the second number is the root mean squared error of the linear model on these training examples divided by the global absolute deviation.

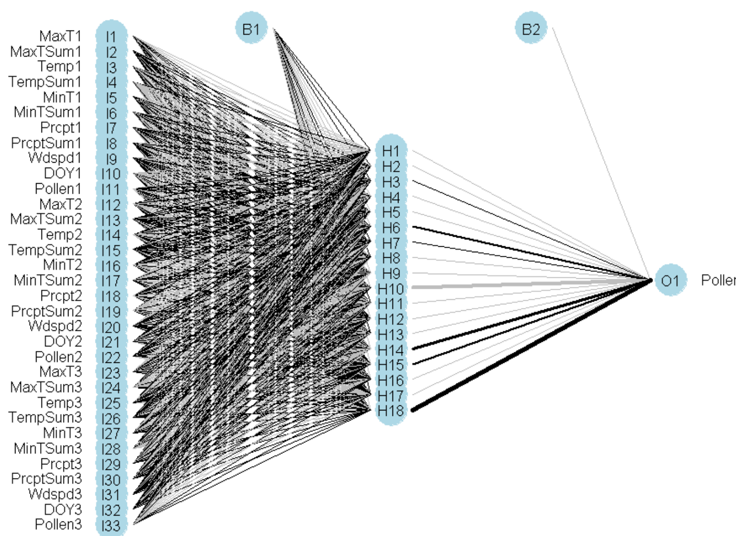

Figure B.6: Model structure of neural network for prediction of daily airborne pollen concentrations. The first layer is the input layer consisting of input variables in Table 2.2. The second layer is the hidden layer. The third layer outputs the airborne pollen concentration. The B1 and B2 are auxiliary units.

Appendix C
SUPPLEMENTARY DATA FOR EMISSION MODEL

C.1 Supplementary Data for Prediction of Start Date and Season Length

Figure C.1 illustrates the relationship between observed start dates and annual mean temperature. The data shown are from selected stations at which there are at least ten years' records for ragweed pollen. It shows that at a given location, higher annual temperature leads to an earlier start dates of allergenic pollen season.

Figure C.2 presents the comparison between observed and simulated SD and SL at each station for each individual year during 1994-2010. For cross validation, data in one year (e.g. 1994) at all stations were held out as a validation set, and the data in other years (e.g. 1995-2010) were used as a training set. M1 model (Table 3.4) was first trained using the training set, and then used to predict the SD and SL for the validation year. The process was repeated for all years.

C.2 Supplementary Data for Pollen Emission Pattern

The Figures C.5, C.6 and C.7 display the mean, maximum, seasonal total and standard deviation of simulated hourly pollen emission for birch, mugwort and grass, respectively.

C.3 Supplementary Data for Sensitivity Analysis of Emission Model

Figure C.8 presents the global sensitivity analysis results for oak pollen emission using four different regional emission metrics (Equation 3.27).

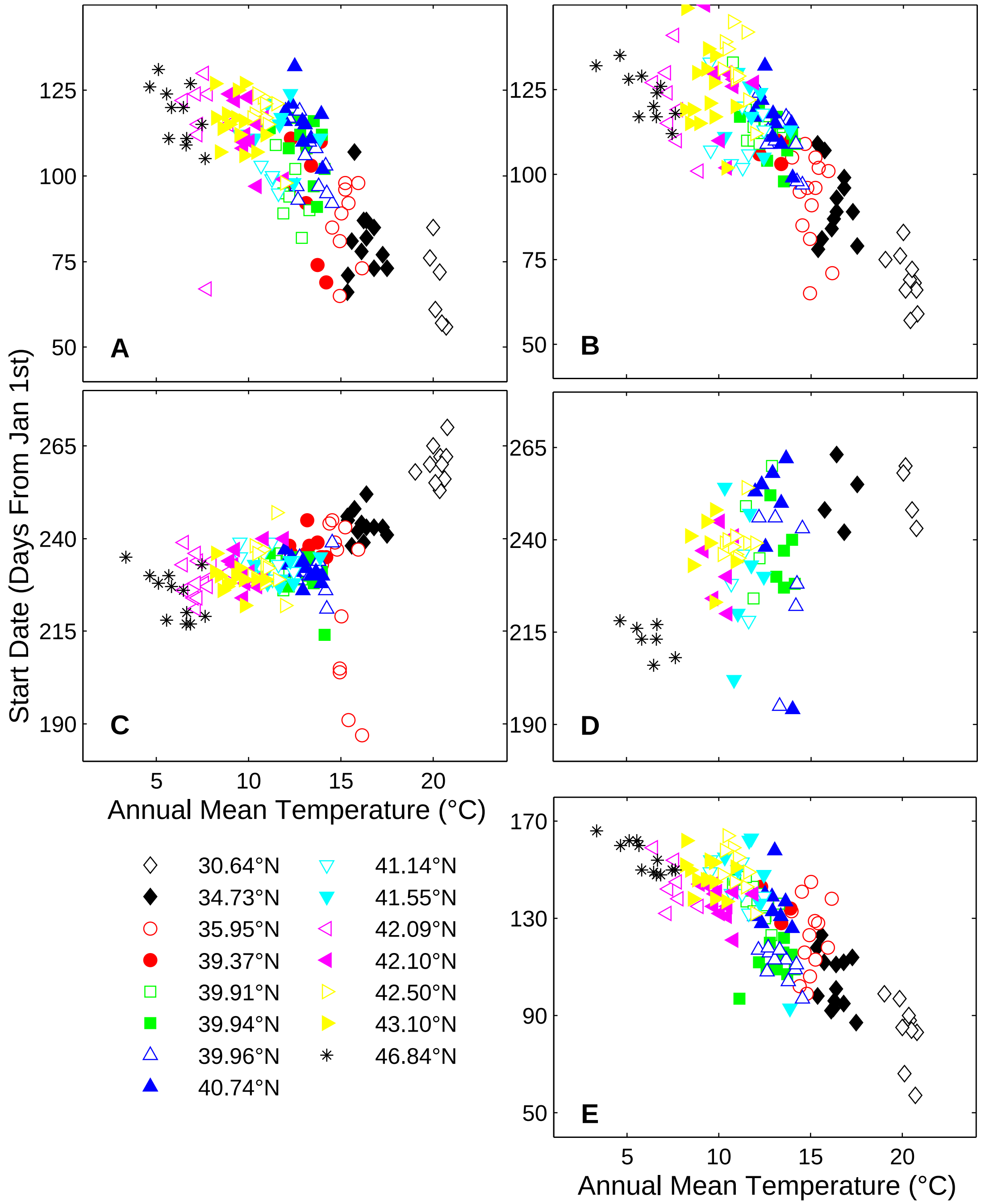

Figure C.1: Relationship between Start Date (SD) and annual mean temperature (T_C) at different latitudes. Each data point in each plot corresponds to an observed SD and a T_C in one year at a pollen monitoring station, which has a unique latitude. (A) birch, (B) oak, (C) ragweed, (D) mugwort, and (E) grass. The data shown are from selected stations at which there are at least ten years' records for ragweed pollen.

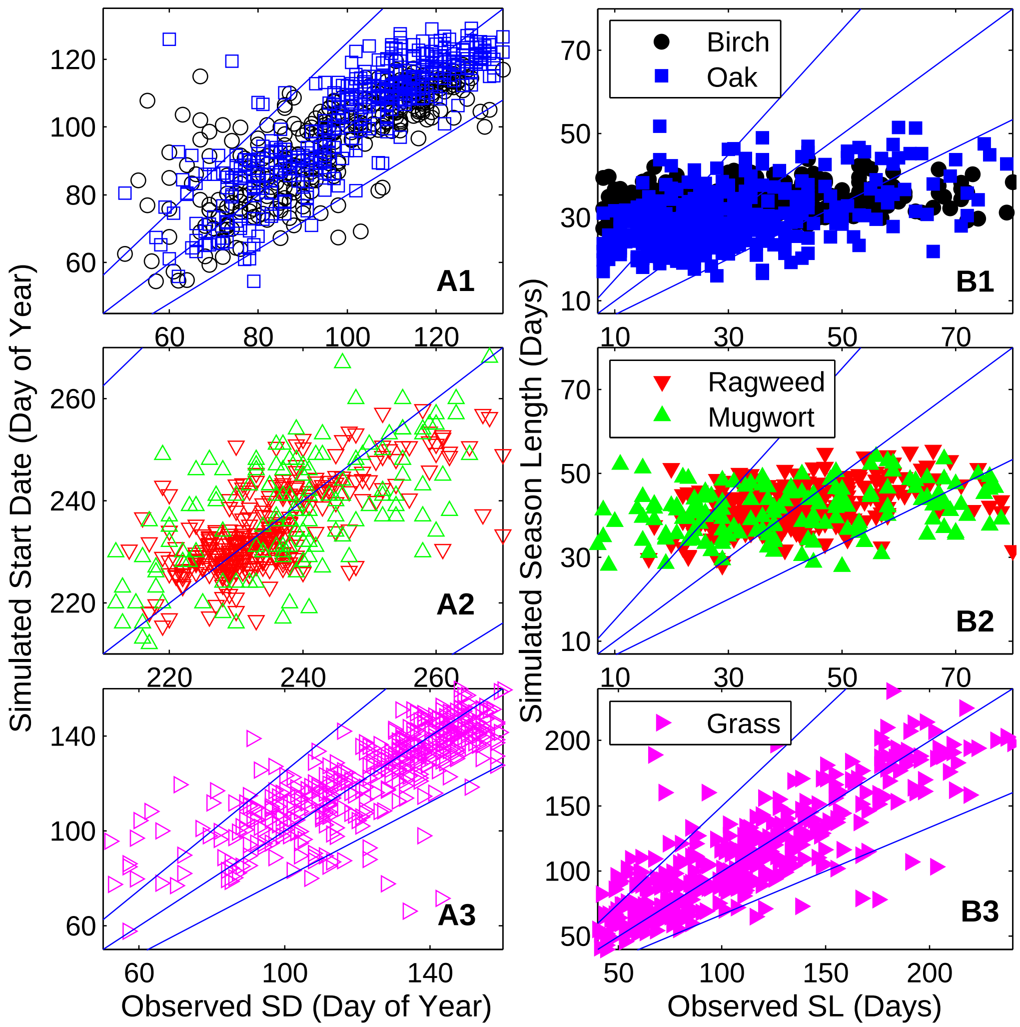

Figure C.2: Cross validation of observed and simulated SD and SL at each station for each individual year during 1994-2010. For cross validation, data in one year (e.g. 1994) at all stations were held out as a validation set, and the data in other years (e.g. 1995-2010) were used as a training set. M1 model (Table 3.4) was first trained using the training set, and then used to predict the SD and SL for the validation year. The process was repeated for all years. Three diagonal lines have been plotted in each panel: the middle line has a slope of unity; the upper line has a slope of 1.20 or 1.50 for SD or SL, respectively; the lower line has a slope of 0.80 or 0.67 for SD or SL, respectively. (A1) SD for birch and oak, (A2) SD for ragweed and mugwort, (A3) SD for grass, (B1) SL for birch and oak, (B2) SL for ragweed and mugwort, and (B3) SL for grass.

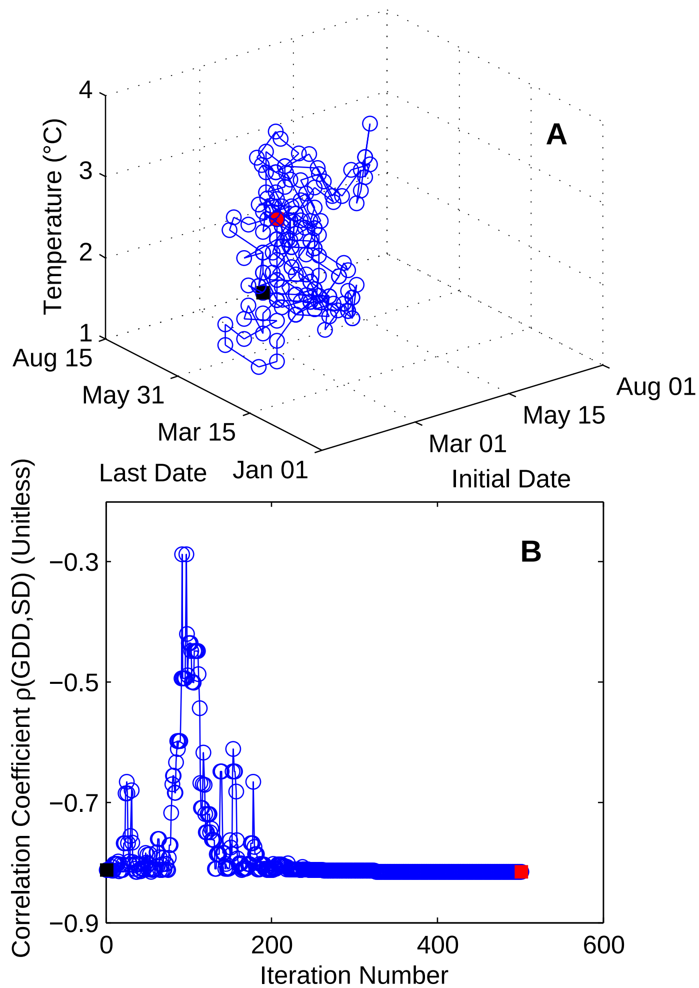

Figure C.3: Searching for optimum Initial Date (ID), Last Date (LD) and base temperature (Tb) for Start Date (SD) of birch pollen season based on Simulated Annealing method. These parameters are used to calculate the fixed-period Growing Degree Day (GDD), which has the highest correlation with SD. The black and red solid square in the plots represent the start and end points, respectively. The start point was determined on the basis of optimum parameters from "coarse grid" search. (A) trajectory in the parameter space, and (B) convergence of the correlation coefficient.

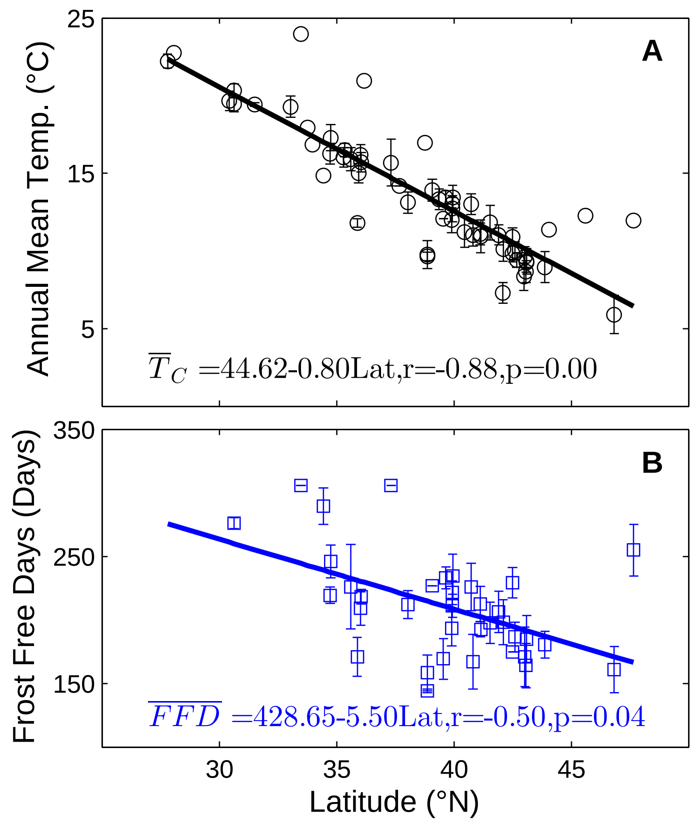

Figure C.4: Averaged annual mean temperature (TC) and Frost Free Days (FFD) across latitudes. The error bar represents the corresponding standard deviation. The average is defined as the mean over the observation period of 1994-2010 at a pollen monitoring station, which has a unique latitude. (A) averaged annual mean temperature, and (B) averaged FFD.

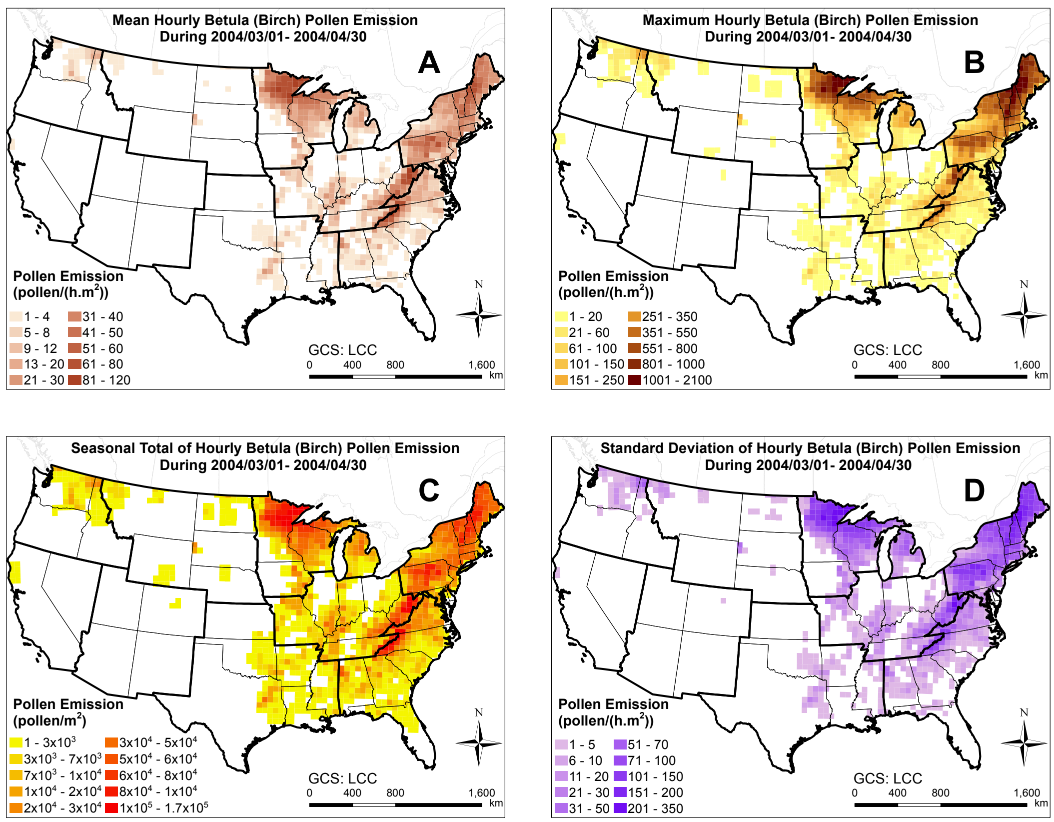

Figure C.5: Spatial pattern of mean, maximum, seasonal total and standard deviation of hourly emission of birch pollen. (A) Hourly mean, (B) Hourly maximum, (C) Seasonal total, and (D) Standard deviation.

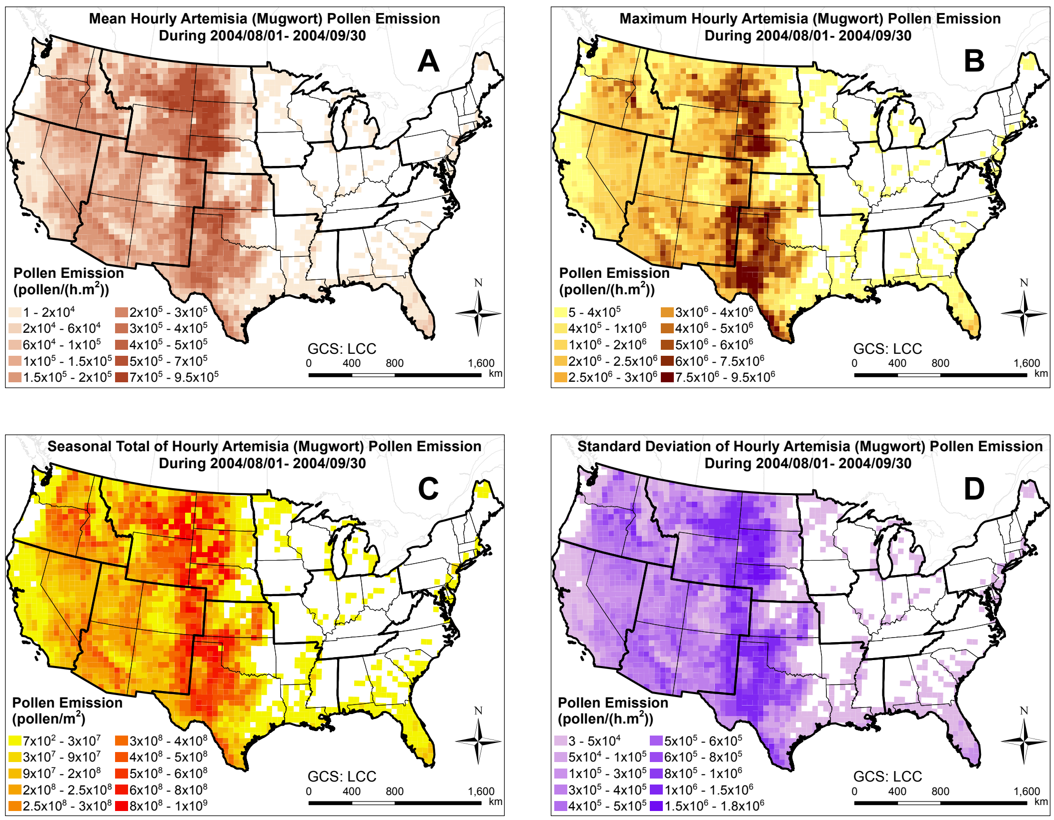

Figure C.6: Spatial pattern of mean, maximum, seasonal total and standard deviation of hourly emission of mugwort pollen. (A) Hourly mean, (B) Hourly maximum, (C) Seasonal total, and (D) Standard deviation.

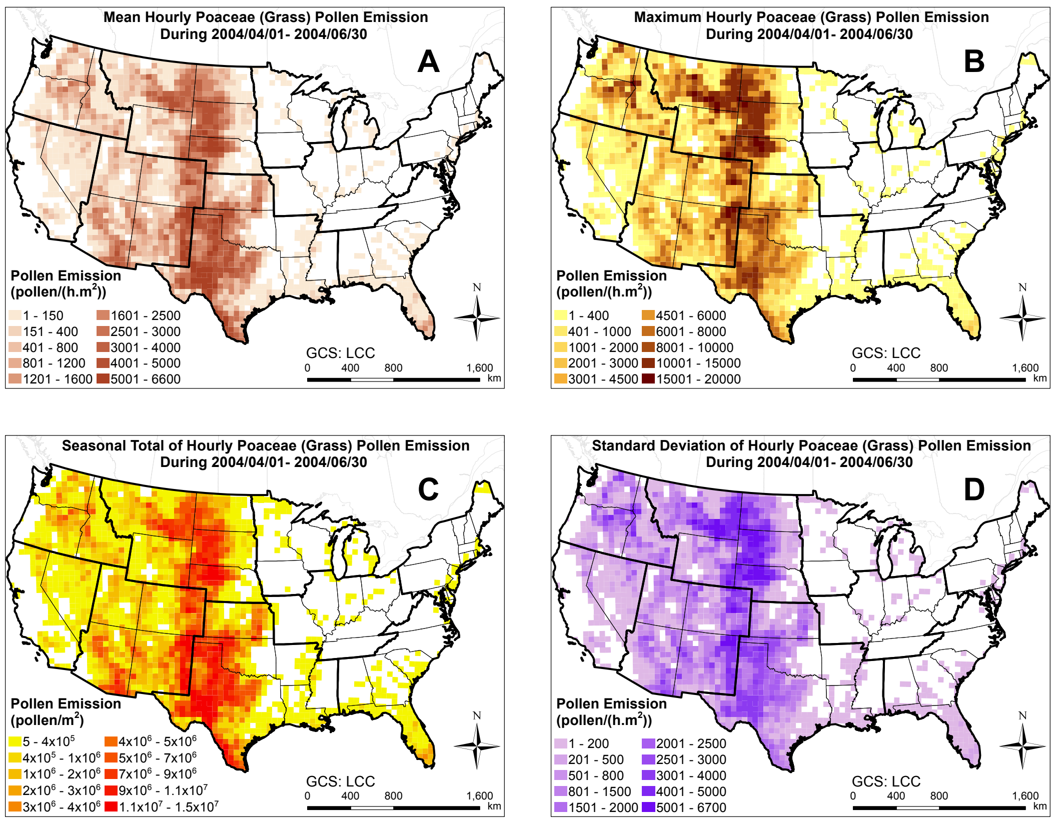

Figure C.7: Spatial pattern of mean, maximum, seasonal total and standard deviation of hourly emission of grass pollen. (A) Hourly mean, (B) Hourly maximum, (C) Seasonal total, and (D) Standard deviation.

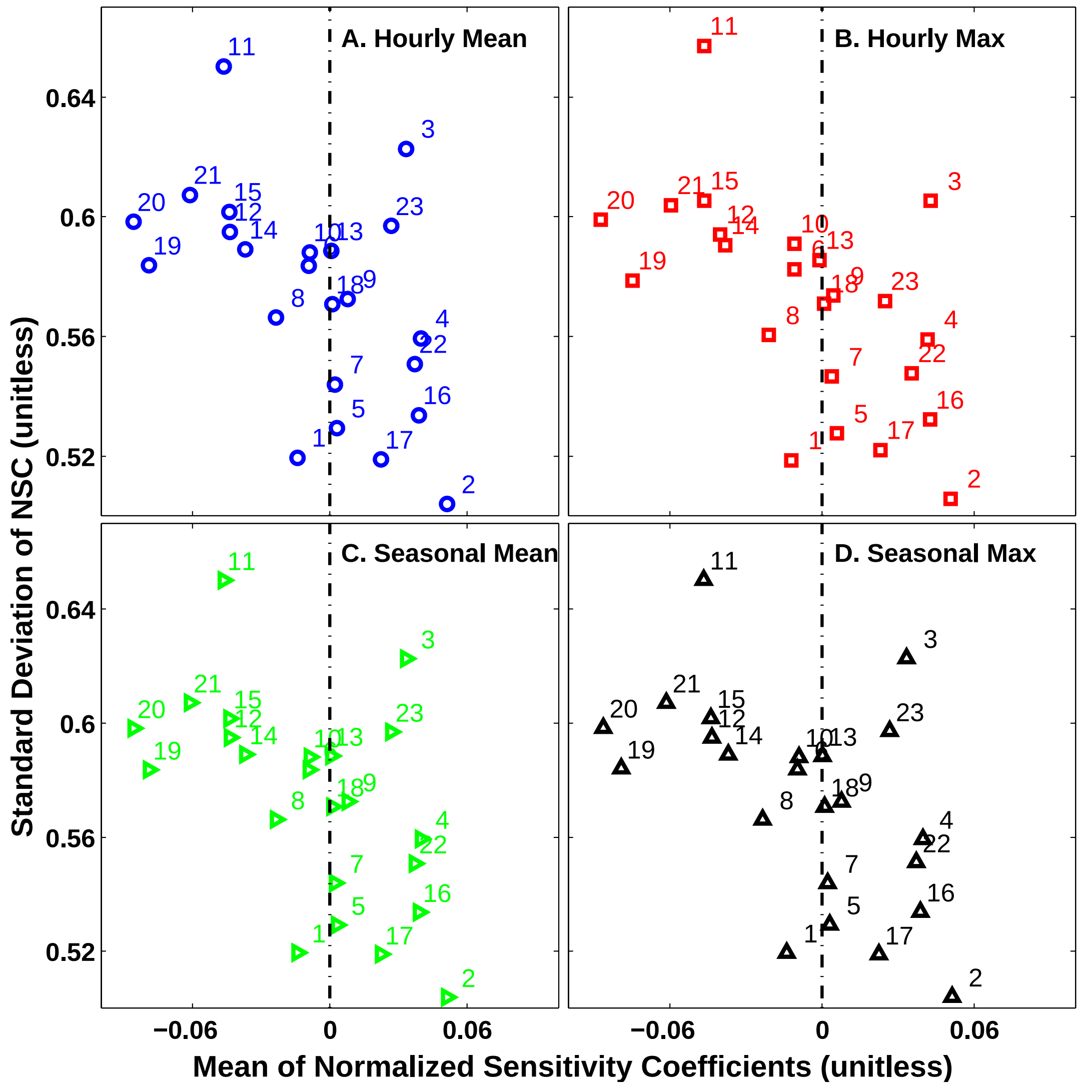

Figure C.8: Mean and standard deviation of Normalized Sensitivity Coefficient (NSC) for each input parameter of oak pollen emission model. (A) Regional hourly mean emission, (B) Regional hourly maximum emission, (C) Regional seasonal mean emission, and (D) Regional seasonal maximum emission. All parameters are described in Table 3.1.

Appendix D
SUPPLEMENTARY DATA FOR TRANSPORT MODEL

D.1 Transport Model Equations

Governing Equations

Byun *et al.* presented the governing equations for fully compressible atmosphere in a generalized meteorological curvilinear coordinates $(\hat{x}^1, \hat{x}^2, \hat{x}^3, \hat{t})$ in a conformal map projection [131]. These general coordinates can be related with the rotated earth-tangential coordinates (x, y, z, t) through equations D.1a and D.1b,

$$\begin{cases} \hat{x}^1 = mx \\ \hat{x}^2 = my \\ \hat{x}^3 = s \\ \hat{t} = t \end{cases} \tag{D.1a}$$

$$\begin{cases} x = m^{-1}\hat{x}^1 \\ y = m^{-1}\hat{x}^2 \\ z = h(\hat{x}^1, \hat{x}^2, \hat{x}^3, \hat{t}) = h_{AGL} + Z_{sfc}(\hat{x}^1, \hat{x}^2) \\ t = \hat{t} \end{cases} \tag{D.1b}$$

where m is the map scale factor, s is the generalized meteorological vertical coordinate, Z_{sfc} is the topographic height, h is the geometric height, and h_{AGL} is the height above the ground. The generalized coordinates facilitate transformations among various horizontal map projections and vertical coordinates adopted by different meteorology and climate modeling system. Since some vertical coordinates (e.g. nondimensional hydrostatic pressure and step-mountain ETA) in meteorological modeling system depends on the atmospheric pressure, the coordinate values decreases with height. In CMAQ, the vertical coordinate is redifined to have a positive definite coordinate [131].

The conservation equations for air density ρ, entropy density ζ (entropy per unit volume) and species concentrations φ_i (mass per unit volume) are shown in equations D.2, D.3 and D.4,

$$\frac{\partial(\rho J_s)}{\partial t} + m^2 \boldsymbol{\nabla}_s \cdot \left(\frac{\rho J_s \hat{\boldsymbol{V}}_s}{m^2} \right) + \frac{\partial(\rho J_s \hat{v}^3)}{\partial s} = J_s Q_\rho \tag{D.2}$$

$$\frac{\partial(\zeta J_s)}{\partial t} + m^2 \boldsymbol{\nabla}_s \cdot \left(\frac{\zeta J_s \hat{\boldsymbol{V}}_s}{m^2} \right) + \frac{\partial(\zeta J_s \hat{v}^3)}{\partial s} = J_s Q_\zeta \tag{D.3}$$

$$\frac{\partial(\varphi_i J_s)}{\partial t} + m^2 \boldsymbol{\nabla}_s \cdot \left(\frac{\varphi_i J_s \hat{\boldsymbol{V}}_s}{m^2} \right) + \frac{\partial(\varphi_i J_s \hat{v}^3)}{\partial s} = J_s Q_{\varphi_i} \tag{D.4}$$

where $J_s = |\partial h / \partial s|$ is the vertical derivative, $\hat{\boldsymbol{V}}_s = \hat{v}^1 \boldsymbol{i} + \hat{v}^2 \boldsymbol{j}$ is the horizontal contravariant wind vector on conformal map coordinates, $\boldsymbol{\nabla}_s = \hat{\boldsymbol{i}} \partial / \partial \hat{x}^1|_s + \hat{\boldsymbol{j}} \partial / \partial \hat{x}^2|_s$, and $\hat{v}^3$ is the contravariant vertical velocity component, Q terms represent sources and sinks of each conservative property.

Entropy density ζ can be diagnostically related with the thermodynamic variables such as temperature and pressure through ideal gas law [150]. It is formulated as,

$$\zeta = \rho C_{vd} \ln(T/T_{oo}) - \rho R_d \ln(\rho/\rho_{oo}) \tag{D.5}$$

where C_{vd} is the specific heat capacity for dry air at constant volume, R_d is the dry air gas constant, T_{oo} is the reference temperature at reference pressure $P_{oo} = 10^5$Pa.

The conservation equation D.4 for pollen grains is detailed in the following paragraphs. Since the stochastic nature of atmospheric motion, the conservation equation is averaged to form a deterministic form through modified Reynolds decomposition.

Air density and pollen concentration φ_p (μg/m^3) are decomposed into mean and turbulent terms as,

$$\rho = \overline{\rho} + \rho' \tag{D.6a}$$

$$\varphi_p = \overline{\varphi}_p + \varphi'_p \tag{D.6b}$$

where $\overline{\rho}$ and ρ' are the mean and turbulent components of air density and $\overline{\varphi}_p$ and φ'_p the same for pollen concentration, respectively. Since some parameters are nolinearly related in the conservation equation, direct Reynolds decomposition to these parameters will introduce covariance terms that complicate the turbulence equations. Instead, averaged mixing ratio $\overline{q}_p$, averaged contravariant wind components $\overline{\hat{v}^k}$ and their fluctuation terms q'_p and $(\hat{v}^k)'$ are defined through equations D.7a to D.7d based on Reynolds decomposition in equations D.6a and D.6b.

$$\overline{q}_p \equiv \overline{\varphi}_p/\overline{\rho} \tag{D.7a}$$

$$q'_p \equiv \varphi'_p/\overline{\rho} \tag{D.7b}$$

$$\overline{\hat{v}^k} \equiv \overline{\rho\hat{v}^k}/\overline{\rho} \tag{D.7c}$$

$$(\hat{v}^k)' \equiv \hat{v}^k - \overline{\hat{v}^k} \tag{D.7d}$$

Advection Process

The advection process depends on the mass conservation characteristics of the continuity equation. Mass consistence error may be arisen from the different sources such as physics, dynamics, numerical process adopted by the meteorology modeling system. A term $\varphi_p^*\frac{Q_\rho}{\rho}$ is added into advection equation D.8 to account for the mass consistency issues.

$$\frac{\partial\varphi_p^*}{\partial t} = -\boldsymbol{\nabla}_s \cdot \left(\varphi_p^*\overline{\hat{\boldsymbol{V}}}_s\right) - \frac{\partial(\varphi_p^*\overline{\hat{v}^3})}{\partial\hat{x}^3} + \varphi_p^*\frac{Q_\rho}{\rho} \tag{D.8}$$

Horizontal Diffusion

Horizontal diffusion equation is presented in equation D.9 by parameterizing the turbulent flux terms via gradient transport theory (K theory)[151],

$$\left.\frac{\partial\varphi_p^*}{\partial t}\right|_{hdiff} = \frac{\partial}{\partial\hat{x}^1}\left[\hat{K}_*^{11}\frac{\partial\overline{q}_p}{\partial\hat{x}^1}\right] + \frac{\partial}{\partial\hat{x}^2}\left[\hat{K}_*^{22}\frac{\partial\overline{q}_p}{\partial\hat{x}^2}\right] \tag{D.9}$$

where $\hat{K}_*^{11}$ and $\hat{K}_*^{22}$ are the horizontal components of the contravariant eddy diffusivity multiplied with factor $\sqrt{\hat{\gamma}}\overline{\rho}$. They are related to the Cartesian counterpart as $\hat{K}_*^{11} = \hat{K}_*^{22} = mk_H$. The K_H is formulated in detail in the following paragraphs.

A heurostic method by[131] is used to parameterize the horizontal diffucivity as shown in D.10. This method can capture the diffusion effects (K_{HT}) on grid-scale process, and reduce the effect of numerical diffusion (K_{HN}) on sub-grid-scale process.

For larger grid size, it minimized the effects of transportive disspersion, whereas for small grid size it reduces the impact of the numerical diffusion.

$$1/K_H = 1/K_{HT} + 1/K_{HN} \tag{D.10}$$

K_{HT} is based on[152] horizontal diffusivity algorithm which accounts for the stretching and shearing deformation characteristics of wind flows

$$K_{HT} = 2\alpha_o^2 (S_\Gamma^2 + S_\Lambda^2)^{1/2} (\triangle x)^2 \tag{D.11}$$

where $\alpha_o^2 \cong 0.28$, stretching strength (S_Γ) and shearing strength (S_Λ) are formulated by

$$S_\Gamma = \frac{1}{2}\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) \tag{D.12a}$$

$$S_\Lambda = \frac{1}{2}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \tag{D.12b}$$

where Cartesian velocity component (u, v) and coordinates (x, y) can be transformed to/from the generalized counterparts using the rules defined by equations D.1a and D.1b.

K_{HN} is parameterized in equation D.13 to counteract the numerical diffusion for simulations with larger grid size[151].

$$K_{HN}(\triangle x) = K_{Hf}(\triangle x_f)\left(\frac{\triangle x_f}{\triangle x}\right)^2 \tag{D.13}$$

where $K_{Hf}(\triangle x_f)$ represents a uniform diffusivity at a fixed resolution $\triangle x_f$. In CMAQ, $K_{Hf}|_{\triangle x_f = 4km} = 2000 m^2 s^{-1}$ is used.

Vertical diffussion

Shown in equation D.14 is the vertical diffusion formulated in terms of mixing ratio,

$$\left.\frac{\partial \overline{q}_p}{\partial t}\right|_{vdiff} = -\frac{\partial \hat{F}^3_{q_p}}{\partial \hat{x}^3} + \frac{Q_{\varphi_p}}{\overline{\rho}} - \hat{F}^3_{q_p}\frac{\partial[\ln(\sqrt{\hat{\gamma}}\overline{\rho})]}{\partial \hat{x}^3} \tag{D.14}$$

where $\frac{Q_{\varphi_p}}{\overline{\rho}}$ is mass consistency error term, last term represents the coordinate divergence term. Effects of the vertical gradient of $\sqrt{\hat{\gamma}}\overline{\rho}$ on the turbulence flux is related with the type of vertical coordinates[131]. Similar as in equation D.9, the flux term $\hat{F}^3_{q_p}$ can be parameterized based K theory. In the situation where eddies are larger than the grid size, K theory fails to adequately represent vertical mixing, CMAQ also provides a nonlocal mixing parameterization shceme based on[153].

Dry deposition

Neglecting the sidewall deposition flux, the effect of dry deposition on pollen concentration is accounted as follows,

$$\left.\frac{\partial \varphi_p^*}{\partial t}\right|_{dep} \approx -\left.\frac{v_d}{h_{dep}}\varphi_p^*\right|_{layer1} \tag{D.15}$$

where h_{dep} is the depth of the lowest model layer in the geometric height coordinate. Dry deposition velocity v_d from[154] is used as default option for CMAQ.

Cloud process

CMAQ cloud module includes parameterizations for subgrid convective precipitation and non-precipitation clouds, and grid-scale resolved clouds. Effects of cloud process on change of pollen concentrations is given by equation D.16,

$$\left.\frac{\partial \overline{\varphi}_p}{\partial t}\right|_{cld} = \left.\frac{\partial \overline{\varphi}_p}{\partial t}\right|_{subcld} + \left.\frac{\partial \overline{\varphi}_p}{\partial t}\right|_{rescld} \tag{D.16}$$

where subscript cld, subcld and rescld represent cloud, sub-grid-scale cloud and resolved cloud, respectively. The effects of subgrid clouds and resolved clouds on average concentrations of pollen are modeled by processes of mixing, scavenging and wet deposition.

Equation D.17 is used to account for the scavenging effects of cloud on rate of change of in-cloud pollen concentrations ($\overline{\varphi_p^{cld}}$) following the cloud time scale (τ_{cld}),

$$\left.\frac{\partial \overline{\varphi}_p^{cld}}{\partial t}\right|_{scav} = \overline{\varphi}_p^{cld}\left(\frac{e^{-\alpha_p \tau_{cld}} - 1}{\tau_{cld}}\right) \tag{D.17}$$

where α_p is the scavenging coefficient for pollen. τ_{cld} is 1 hr for sub-grid convective clouds, and it is equal to CMAQ's synchronization time step for grid resolved clouds. Pollen grains are treated as coarse mode aerosols and assumed to be completely absorbed by the cloud and rain water. Therefore, scavenging coefficients for pollen grains is simply a function of the washout time $\tau_{washout}$ as shown in equation D.18.

$$\alpha_p = 1/\tau_{washout} \tag{D.18}$$

Washout time indicates the amount of time needed to remove all the water from the could volume at the specified precipitation rate (P_r) given the cloud thickness Δz_{cld}.

It is parameterized as follows,

$$\tau_{washout} = \frac{\overline{W}_T \Delta z_{cld}}{\rho_{H_2O} P_r} \tag{D.19}$$

where $\overline{W}_T$ is the mean total water content, and ρ_{H_2O} is the density of water.

Wet deposition:

Wet deposition of pollen grains ($wdep_p$) is related with the precipitation rate (P_r) and the cloud water concentration $\overline{\varphi}_p^{cld}$ as formulated in equation D.20.

$$wdep_p = \int_0^{\tau_{cld}} \overline{\varphi}_p^{cld} P_r dt \tag{D.20}$$

D.2 Performance Evaluation of Transport Model

Figures D.1, D.2, and D.3 present the results for evaluating the simulated allergenic pollen season timing and ambient levels for birch, mugwort and grass, respectively. The WRF-SMOKE-CMAQ-Pollen modeling system could capture the variations in start date, season length and airborne levels of birch, mugwort and grass pollen. However it did not perform as well as for oak and ragweed pollen. The potential reasons was discussed in the section 4.3.5.

D.3 Climate Change Impact on Allergenic Pollen

Figure D.4 presents the mean, maximum and standard deviation of hourly concentration of ragweed pollen during periods of 2001-2004 and 2047-2050 in the CONUS. Figure D.5 displays the average start date and season length of ragweed pollen season. Figure D.6 shows the number of hours in which ragweed pollen concentration exceed 30 pollen grain/m^3.

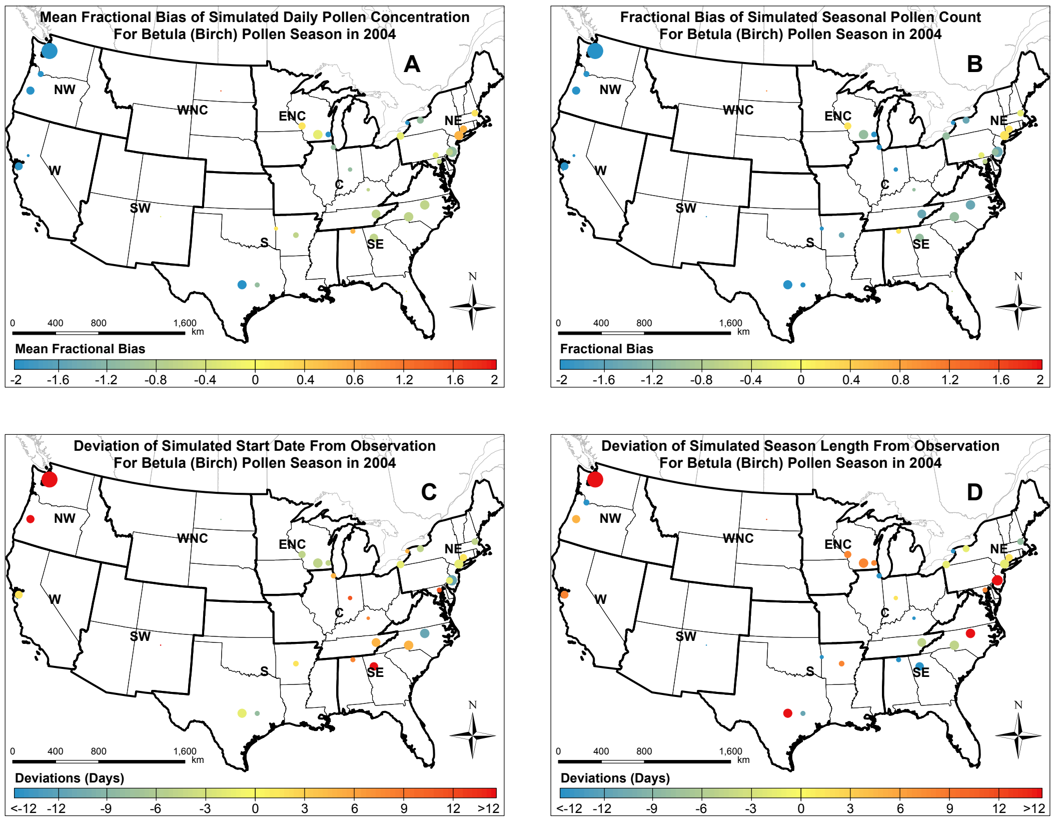

Figure D.1: Evaluation of predicted birch pollen season during 2004 in the CONUS. The size of the circle indicates the birch pollen abundance at that station. (A) Mean Fractional Bias of daily pollen concentration, (B) Fractional Bias of seasonal pollen counts, (C) deviation between observed and simulated Start Dates, and (D) deviation between observed and simulated Season Length.

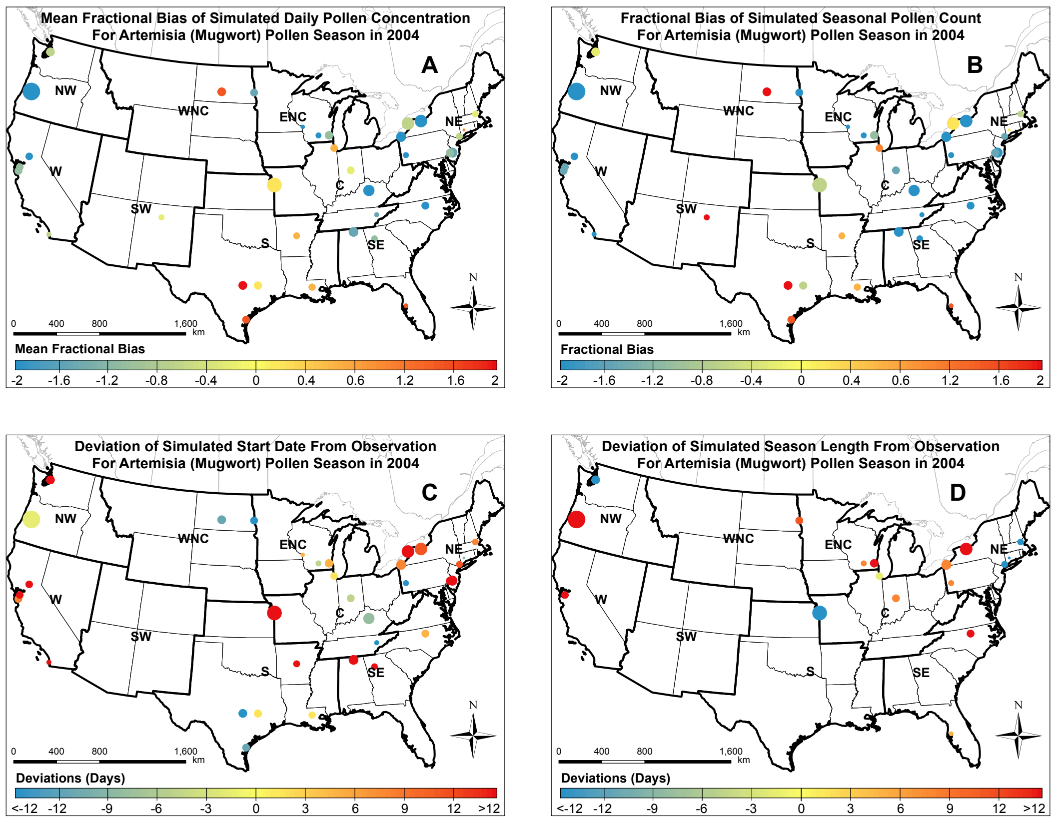

Figure D.2: Evaluation of predicted mugwort pollen season during 2004 in the CONUS. The size of the circle indicates the mugwort pollen abundance at that station. (A) Mean Fractional Bias of daily pollen concentration, (B) Fractional Bias of seasonal pollen counts, (C) deviation between observed and simulated Start Dates, and (D) deviation between observed and simulated Season Length.

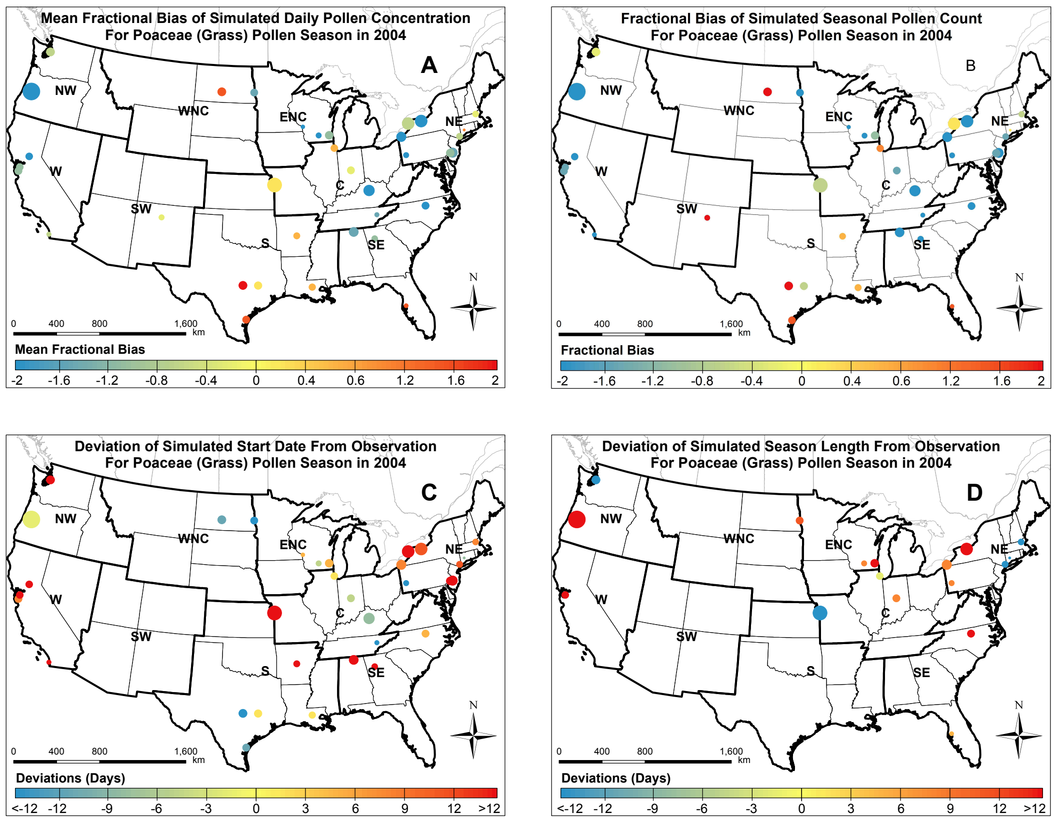

Figure D.3: Evaluation of predicted grass pollen season during 2004 in the CONUS. The size of the circle indicates the grass pollen abundance at that station. (A) Mean Fractional Bias of daily pollen concentration, (B) Fractional Bias of seasonal pollen counts, (C) deviation between observed and simulated Start Dates, and (D) deviation between observed and simulated Season Length.

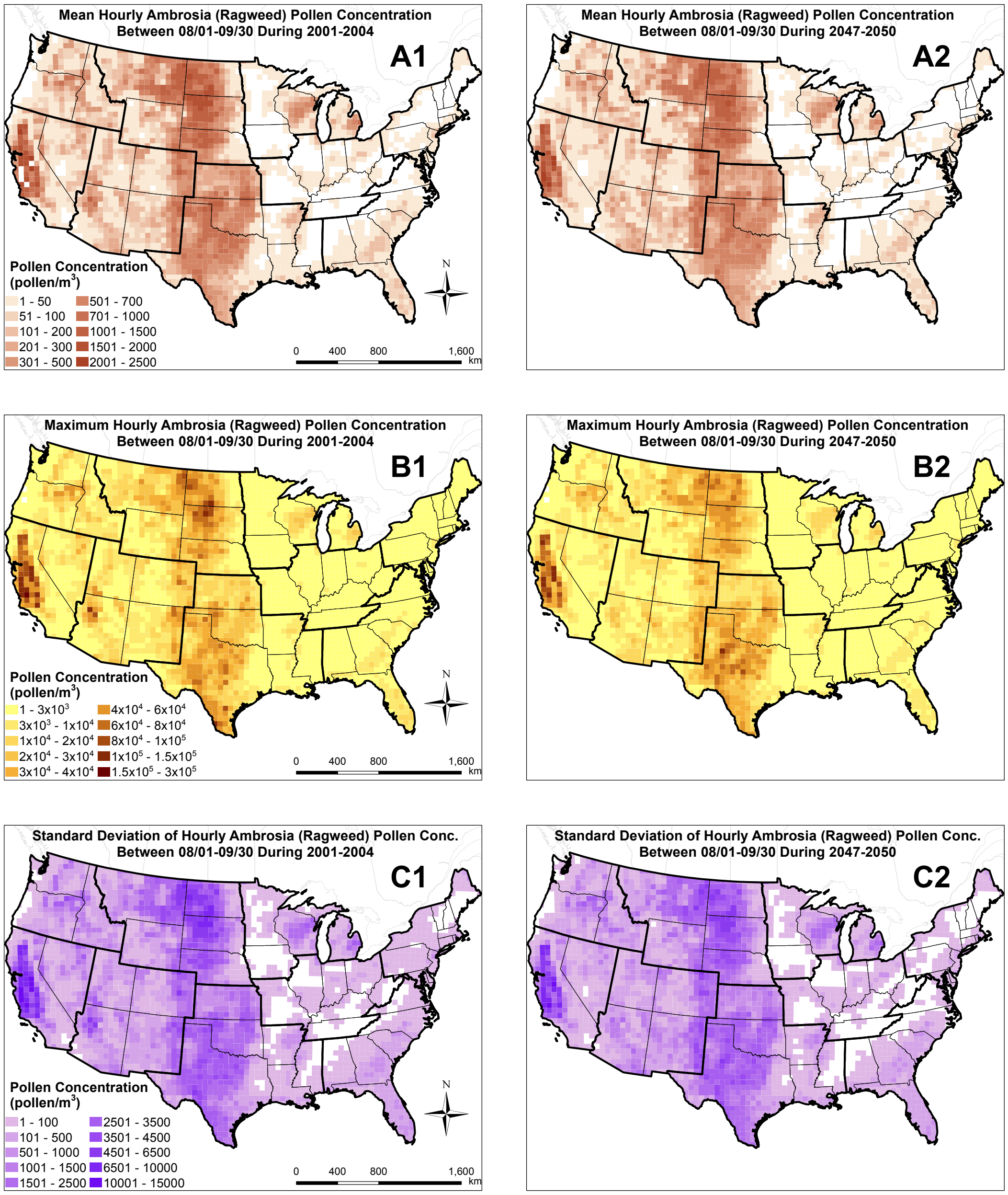

Figure D.4: Mean, maximum and the standard deviation of the simulated hourly concentrations of ragweed pollen during periods of 2001-2004 and 2047-2050. (A1-C1) Mean, maximum and standard deviation during 2001-2004, and (A2-C2) Mean, maximum and standard deviation during 2047-2050.

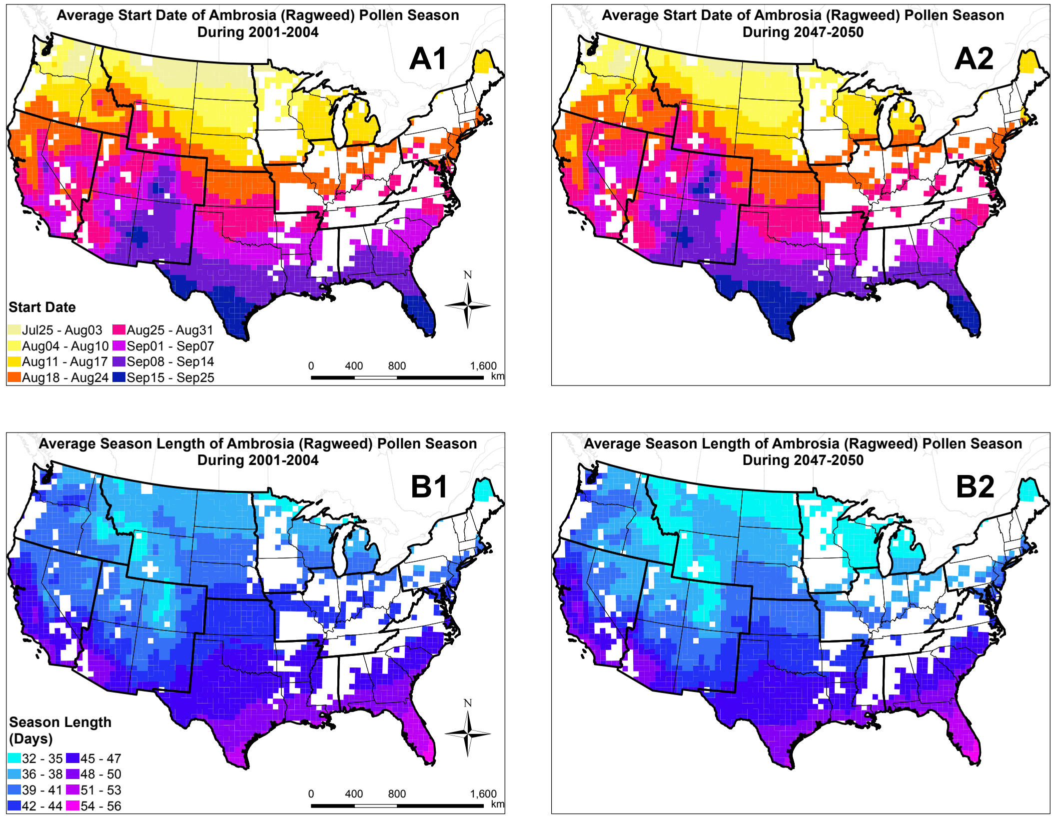

Figure D.5: Average Start Date and Season Length of ragweed pollen season during periods of 2001-2004 and 2047-2050. Data were mapped only on cells in which the area coverage of ragweed plant is greater than zero. (A1-B1) Average SD and SL during 2001-2004, and (A2-B2) Average SD and SL during 2047-2050.

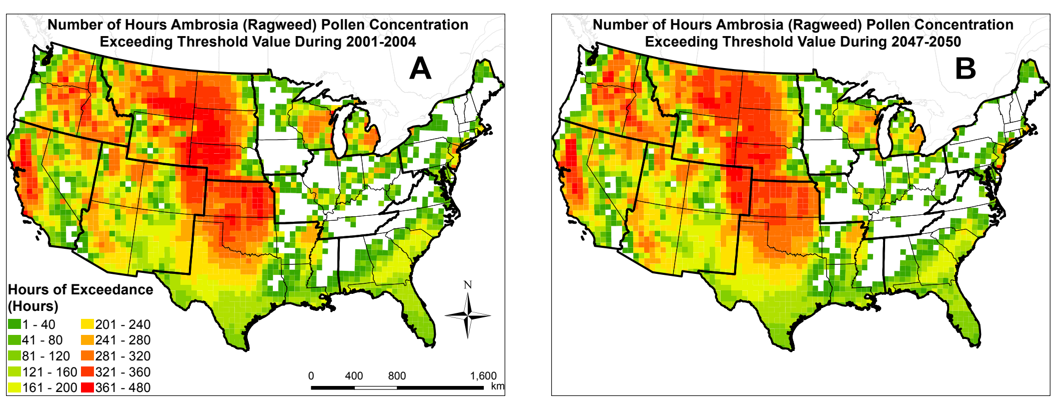

Figure D.6: Number of hours in which ragweed pollen concentration exceeds 30 pollen grains/m^3 during periods of 2001-2004 and 2047-2050. (A) Average exceedance hours during 2001-2004, and (B) Average exceedance hours during 2047-2050.

Appendix E
SUPPLEMENTARY DATA FOR EXPOSURE MODEL

E.1 Number of Virtual Subjects

As an example, Figure E.1 presents the relationship between the simulated mean exposure and the size of simulated population sample. In Figure E.1, mean inhalation exposure to oak pollen grains during 2001-2010 in south region converges around the population size of 2500. Similar convergences around population size of 2500-3000 were found for exposures to allergenic pollen of other taxa, through other routes, in other regions, and during period of 1994-2000. This indicated that population size of 3000 in each of the nine climate regions is sufficient to generate the representative statics of exposures to allergenic pollen.

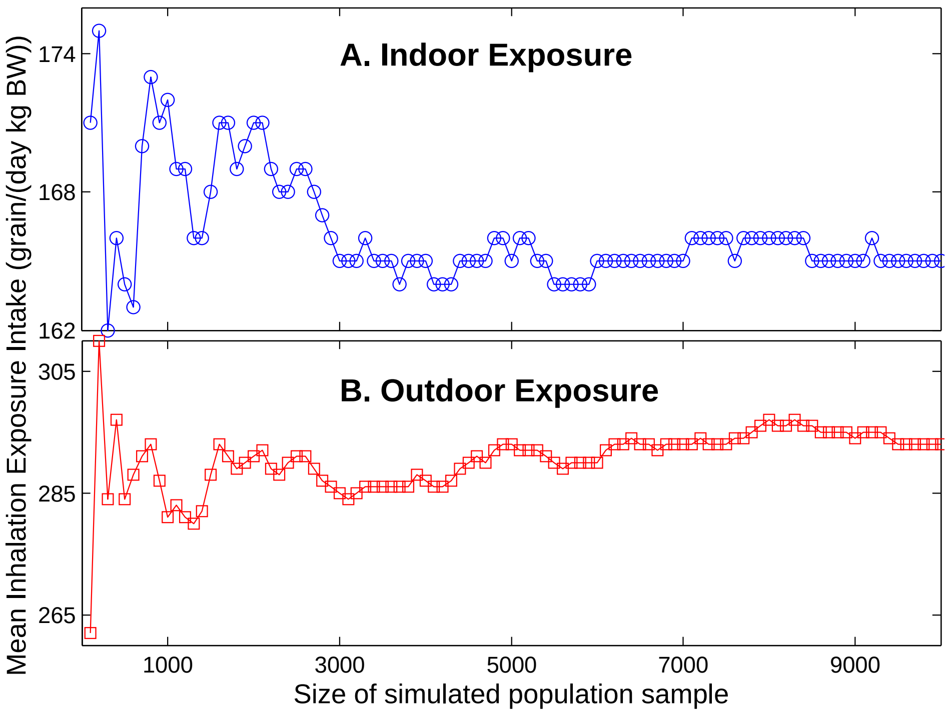

Figure E.1: Mean inhalation exposure to allergenic pollen and the size of simulated population sample. Mean inhalation exposure to oak pollen grains during 2001-2010 in south region converges around the population size of 2500.