2017

Monitoring Pollen Counts and Pollen Allergy Index Using Satellite Observations in East Coast of the United States

Murat Cagatay Kececi
South Dakota State University

MONITORING POLLEN COUNTS AND POLLEN ALLERGY INDEX USING SATELLITE OBSERVATIONS IN EAST COAST OF THE UNITED STATES

BY

MURAT CAGATAY KECECI

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Geography

South Dakota State University

2017

MONITORING POLLEN COUNTS AND POLLEN ALLERGY INDEX USING SATELLITE OBSERVATIONS IN EAST COAST OF THE UNITED STATES

This thesis is approved as a creditable and independent investigation by a candidate for the Master of Science in Geography degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this thesis does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Dr. Xiaoayng Zhang Date

Thesis Advisor

Dr. Bob Watrel Date

Head, Geography

Dean, Graduate School Date

ACKNOWLEDGEMENTS

First, I would like to express sincere gratitude to my thesis advisor, Dr. Xiaoayng Zhang, for his constant guidance and encouragement in the way of achieving my goals and elaboration of my thesis research. Also, I would like to thank my committee members, Dr. Evren Celik Wiltse, Dr. Jamie Spinney, and Dr. Bob Watrel for the support in progressing and evaluating of my thesis. In addition to these, I would like to thank to Dr. Lingling Liu for helping on EVI2 data and Dr. Dong Yan for helping on precipitation data from TRMM.

Without the financial support of Ministry of Food, Agriculture and Livestock and Ministry of National Education of Republic of Turkey which offered me a great scholarship to do graduate studies, this work and my dreams would not have been possibly persuaded. I would like to express my heartfelt gratitude to the crew of these ministries and Office of Education Attaché to The Turkish Consulate General in New York for interest and help before/during my graduate studies.

I would also like to a lot of thanks to one of the my best friends, Mehmet Ozturk, for his encouragement and moral support which made me stronger and my studies very enjoyable, and thanks to Dinesh Shrestha and Atilla Polat for their continuous support.

In addition, I would like to thanks to my family including my father, mother, sister, and mother's brother for all their supports and sacrifices. Moreover, I would like to big thanks to one of the my oldest and best friends, Burak Karatas, for his moral support in every level of my life.

TABLE OF CONTENTS

LIST OF ABBREVATIONS

AAAAI: American Academy of Allergy Asthma and Immunology

ANN: Artificial Neural Network

ARIMA: Autoregressive Integrated Moving Average

AVHRR: Advance Very High-Resolution Radiometer

CDC: The Center of Disease Control

CT: Connecticut State

DC: District of Columbia State

DE: Delaware State

EVI: Enhance Vegetation Index

EVI2: Two-band Enhance Vegetation Index

FL: Florida State

GA: Georgia State

HDM: House Dust Mites

HPLM: Hybrid Piecewise Logistic Model

ºK: Kelvin Degree

LSP: Land Surface Phenology

LST: Land Surface Temperature

MA: Massachusetts State

ME: Maine State

MD: Maryland State

MODIS: Moderate-Resolution Imaging Radiometer

MSE: Mean Square Error

NDVI: Normalized Difference Vegetation Index

NC: North Carolina State

NCSU: North Carolina State University

NH: New Hampshire State

NIR: Near Infrared

NJ: New Jersey State

NY: New York State

NYC: New York City

OK: Oklahoma State

RI: Rhode Island State

RMSE: Root Mean Square Error

SC: South Carolina State

TRMM: Tropical Rainfall Measurement Mission

US: United States

USD: United States Dollar

VA: Virginia State

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

MONITORING POLLEN COUNTS AND POLLEN ALLERGY INDEX USING SATELLITE OBSERVATIONS IN EAST COAST OF THE UNITED STATES

MURAT CAGATAY KECECI

2017

Allergic diseases have become increasingly common over the world during the last four decades, and they are affecting millions of people. Pollination is an important process in the life cycle of plants. However, pollen exposure is associated with allergic diseases such as asthma and seasonal allergic rhinitis (hay fever). As a result, the total annual expenditure for asthma-associated morbidity is about $56 billion in the United States, and the overall cost of allergic diseases is over $18 billion annually. For allergic rhinitis, the annual medical cost is approximately $3.4 billion. The intensity and frequency of the pollen exposures can be easily affected by many factors such as climate, vegetation, and topography, which are difficult to predict in large scales. Vegetation is very important as a pollen source, and the amount and time of pollinations depend on the flowering and growth of plants. With optimal water and temperature, vegetation can reach a maximum growth and flowering during a growing season, which means that maximum amount of pollen can be released from the plants. However, if the requirements of water and temperature cannot be met in the specific times within the growing season, pollen dispersal will be affected negatively. It is an urgent need to develop models or systems for predicting pollen events at large scales and providing early warning to prevent pollen effects on people. Unlike manual pollen counting at local sites, remote sensing facilitates the pollen estimates at large scales with temporally and spatially distributed observations, which significantly reduces the time and labor costs. With remotely sensed observations, Artificial Neural Network (ANN) helps us fill the gaps in understanding of the relationships between environmental variables and pollen concentration. At this point, I investigated pollen estimates from satellite observations in the states of East Coast United States with short and long-term data. This region is highly populated with a population of 104 million. In addition, this region has a great variety of temperature, precipitation, and vegetation. The final goal of this project is to investigate the relationships between satellite-derived variables (precipitation, land surface temperature (LST), and enhance vegetation index (EVI2)) and pollen count and further to generate a model for the prediction of pollen counts at high temporal and spatial resolutions. For this purpose, to predict pollen concentration using environmental variables, a Neural Network Analysis was performed. The results showed that strong correlations existed between pollen counts and environmental variables, except for precipitation in most locations. The validation analysis using regression models revealed strongly significant relationships between the observed and predicted pollen concentrations obtained for short and long-term data. The R squares (R^2) for long term pollen counts were mostly higher than 0.5, ranging from 0.5542 for Olean, NY to 0.8589 for Savannah, GA. For short term predictions of pollen allergy index, R^2 ranged from 0.53 to 0.966 except for a few sites, especially in southern Florida. The pollen distribution was mostly affected by precipitation in the southern part, whereas it was influenced by temperature in the northern part. Moreover, results demonstrated that ANN is a suitable tool for complicated statistical analysis and EVI2 combining with LST and precipitation is a reliable predictor of pollen variation. Overall the results provide a better understanding of pollen variation with vegetation seasonality and climate variables, which could assist an approach towards the establishment of an early warning system for allergy patients.

Key Words: Pollen Count, Pollen Allergy, Artificial Neural Network, Remote Sensing Data

CHAPTER 1: INTRODUCTION

1.1. Problem Statement and Description

Pollen is a very fine powder produced by flowers on trees, grasses, and weeds, which carried by insects or wind to fertilize other plants of the same species (Kerr 2016). The pollen grain is a part of the flowering plant life cycle, and is a specialized structure that harbors the flowering plant male gametes. Its biological function is to fertilize the female gametophyte (Taketomi et al. 2006). Despite of the key role of pollen for plants, pollen is a serious problem in our daily life with causing different allergic diseases leading to sneezing, wheezing, coughing, snorting, and itching (UCB Institute of Allergy 1999). While approximately 31 million Americans each year suffer from different allergic diseases in the past (Davis 1972), 60 million Americans suffer from allergies every day in the present (Meltzer et al. 2009). It is also reported that while allergies were affecting 1% of the population in Europe at the beginning of the century, they are now affecting about 20% of people, and it is predicted that 35% of the population will be affected within next twenty years with an estimated cost of around 30 billion Euro (UCB Institute of Allergy 1999; Ranzi et al. 2003). It is substantial to be aware that allergic diseases and their medical costs can be a heavy burden for allergic sufferers. Therefore, pollen forecasts must be improved for preventing allergic episodes and reducing economic costs owing to allergic diseases (Ranzi et al. 2003). In the science world, the awareness of pollen diseases among the scientists are increasing and more studies are conducted on pollen diseases. However, current studies are mostly focused on local scales due to the difficulties of pollen counting, so it is needed to improve models to predict pollen intensity in large scales.

Pollen counts can be affected directly or indirectly by environmental factors (relative humidity, temperature, and precipitation) (Taketomi et al. 2006; Valecia et al. 2001), human-induced factors (industrialization and urbanization) (Takai et al. 2015; Jalbert et al. 2015), vegetation types (grass, trees etc.) and topography (slope, elevation) (Jin et al. 2008). Some factors are very changeable in short times, which makes counting pollen and predicting intensity of pollen distribution difficult, so it is important to develop innovative approaches to predict pollen effectively.

Pollen counting is most commonly performed manually by counting pollens one by one using a haemocytometer. This method could provide accurate results, but it takes too much time and needs tremendous labors (Mudd and Arathi 2012). Moreover, pollens collected for manual counting through pollen traps can represent the area where they are placed (Skjøth et al. 2012), which means this method can only be effective for small-scaled forecast of pollen. However, new monitoring devices such as remote sensing systems are generating vast amounts of spatio-temporal data with the wider accessibility (Turner et al. 2006) thus providing opportunities for large-scaled forecasts as well as reducing labor. For example, Normalized Difference Vegetation Index (NDVI) and a two band EVI (EVI2) derived from Advanced Very High-Resolution Radiometer (AVHRR) and Moderate-Resolution Imaging Spectroradiometer (MODIS) might be used for observing long-term dynamics of the vegetated land surface and climate impacts since it enables the longest time series of global satellite measurements (Huete et al. 2006 and Jiang et al. 2008).

Models to estimate pollen intensity in the air are based on the interactions between atmospheric weather and pollen (Arizmendi et al. 1993). Several statistical approaches have been developed to investigate pollen concentration from environmental factors but showed limited successes (Goldberg et al. 1988; Arizmendi et al. 1993). Artificial Neural Network (ANN) is a universal statistical tool for the problems related to complex or poorly understood phenomena (Castellano-Méndez et al. 2005), which is able to model the dynamic characteristics of time series of atmospheric pollen (Bianchi, Arizmendi, and Sanchez 1992) and to predict pollen concentration (Arizmendi et al. 1993) in anarchic time series (Lapedes and Farber 1987).

The difference of ANNs from the other statistical programs for algorithmic processing of information is the ability to generalize knowledge over new, previously unknown data not presented in the process of learning (Puc 2012). This characteristic of ANNs made it popular in the science world and it has been used in various pollen studies such as a grass pollen in the southern part of the Iberian Peninsula (Sanchez-Mesa et al. 2002), betula pollen in different parts of Europe (Castellano-Méndez et al. 2005), determining the relationship between betula pollen and meteorological factors in Szczecin (Poland) (Puc 2012), airborne castanea pollen forecasting model for ecological and allergological implementation (Astray et al. 2016), and using machine learning to estimate atmospheric Ambrosia pollen concentrations in Tulsa, OK (Liu et al 2017).

1.2. Thesis Statement and Research Objectives

In this study, I researched the linkage of climatic factors (such as temperature and precipitation) and vegetation growth seasonality to the variation of seasonal pollen count in the East Coast of the United States. In particular, I focused on the prediction of short and long-term pollen concentration using satellite observations and Neural Network Analysis towards an early warning system for allergy sufferers.

In this case, to better understand the vegetation seasonality and climate impacts on pollen releases, I followed the research questions below:

- What is the association of pollen allergy index or pollen count with climate variables and vegetation growth condition?
- How to predict seasonal variation in pollen counts and pollen allergy index based on vegetation phenology, temperature, and precipitation?
- How to integrate vegetation phenology data and climate parameters to predict pollen count?
- How to validate the pollen prediction of model?

I collected two different data sets in this research project. The first is long term pollen count data observed in field sites from 2002 to 2015, and second is short term data of pollen allergy index from May to December 2016 across the East Coast of the US. The aim of this study is to reveal the strength of relationship between pollen and other variables (land surface temperature, precipitation and EVI2) and then create the best model to predict pollen count or pollen allergy using these variables in ANN. This research is expected to show that the pollen counts and pollen allergy index could be predicted spatially and temporally from satellite observations.

CHAPTER 2: LITERATURE REVIEW

2.1. Pollen

The dictionary definition of pollen is a fine powdery substance, typically yellow, consisting of microscopic grains discharged from the male part of a flower or from a male cocoon. The pollen grain is a part of the flowering plant life cycle, and is a specialized structure that harbors the flowering plant male gametes. Its biological function is to fertilize the female gametophyte (Taketomi et al. 2006). Pollination is a substantial process because pollen, looking like insignificant yellow dust, bears a plant's male sex cells and is a vital link in the reproductive cycle. Almost all of the seed plants in the world need to be pollinated. With adequate pollination, wildflowers (Huang and Giray 2012):

- Reproduce and produce enough seeds for dispersal and propagation.
- Maintain genetic diversity within a population.
- Develop adequate fruits to entice seed dispersers.

Pollination provides sustainability for plants and thus contribute to carbon cycling/sequestration, clean air, and purification of water as well as prevent soil erosion (Knapp et al. 2002).

Although it is an important process for nature, pollen is one of the allergen sources such as house dust mites (HDM) and fungi, HDM fecal pellets, food, and cat dander, all of which produce or include allergen components (Takai and Izuhara 2015). It also plays a vital role in improvement and escalation of allergic diseases (Kizilpinar et al. 2010). Airborne pollen is a significant cause of asthma and rhinitis (Rojo et al. 2016).

While about 300 million people worldwide suffer from asthma (UCB institute of allergy 1999), 31 million Americans each year suffer from some form of allergic disease, including 8.6 million from asthma (Davis 1972). And then, the number of sufferers increased to 60 million based on the Meltzer's research in 2009 (Meltzer et al. 2009). These diseases are also common in Europe (Rojo et al. 2016; Emberlin et al. 1999; Emberlin et al. 2000; Ranzi et al. 2003) as results of urbanization, industrialization and climate change affecting the prevalence and management of allergic diseases (Jalbert and Golebiowski 2015). 5-30 % of the population of industrialized countries (European countries) is influenced by pollen-related allergic diseases (Cecchi 2012). Allergy owing to pollen might be ranked as one of the most common diseases during the next century among European population (Ranzi et al. 2003), and the awareness on this global issue is increasing among scientists from various countries such as Spain, France, Turkey, Argentina, United States and Australia (D'Amato et al. 2007; Beggs et al. 2015; Valencia-Barrera et al. 2001; Emberlin et al. 2000; Kizilpinar et al. 2010).

2.2. Factors Affecting Pollen Dispersal

Pollen dispersal is a natural event and can be affected by various factors. These factors can be grouped into environment, human-induced factors, vegetation types, and topography.

2.2.1. Environment

Environmental factors that affect pollen allergen release in the air are relative humidity, precipitation, temperature and wind speed. In high humidity, allergens are released from the pollen grain in a process similar to that occurs in physiological pollinating conditions (Taketomi et al. 2006). Rarely, such as in thunderstorms, pollen grains might rupture as a result of osmotic shock, releasing allergen-containing particles (Taketomi et al. 2006). In Leon (Spain), the results of a correlation analysis between the meteorological parameters and the daily pollen concentrations reveal a positive and significant correlation of pollen with maximum and minimum temperatures and sunshine hours, but a negative (but still significant) correlation with relative humidity and wind speed (Valencia et al. 2001). One study claim that allergy sufferers were at stake from February through October due to the high airborne pollen concentrations, which only showed a temporary decline when the temperature decreased or there was precipitation (Peternel 2004). Also in the previous studies, heavy and long-term rainfalls were generally informed to keenly decrease airborne pollen concentration but correlations were poor (Perez et al. 2009; Green et al. 2004; Gottardini and Fabiana 1997; Barnes et al. 2001; Bartková-Ščevková, 2003) because a sustained rain or a short but heavy rain washes the air of pollen away. Raindrops falling to the ground with the force of gravity take pollen with them and therefore plants are more prolific at releasing pollen during warm, dry weather (Korpella 2017).

2.2.2. Human-induced factors

Human-induced changes in environments resulting from urbanization and industrialization have vital implications for the prevalence and management of allergic diseases (Takai and Izuhara 2015; Jalbert and Golebiowski 2015). For instance, 5-30% of the population in industrialized countries are affected by pollen-related allergic diseases (Cecchi 2012). Furthermore, urbanization has an indirect impact on pollen concentration effecting some factors such as temperature and soil moisture. Many studies attributed advancement in flowering in the cities to the urban heat island effect (Neil and Jianguo

2006), also different vegetation types showed different reactions to this change in temperature (Zhang et al. 2004). Urbanization can alter humidity either by escalating or reducing surface water or count of plant (Adebayo 1991; Lipfert et al. 1991; Jonsson 2004). Experimental research conducted in the Mediterranean stated a strong correlation between decreased moisture availability and delayed flowering (Penuelas et al. 2004).

2.2.3. Vegetation

Pollen could be released from trees, grasses, and weeds (Table 1). Weeds are the most prolific producers of allergenic pollen among North American plants. Ragweed, sagebrush, redroot pigweed, lamb's quarters, Russian thistle (tumbleweed), and English plantain are the major sources for pollen (Kosisky, Marks, and Nelson 2010). Grasses and trees are also important sources of allergenic pollens. Although more than 1,000 species of grass grow in North America, the highly allergenic pollen is only produced from a few species such as timothy grass, Kentucky bluegrass, Johnson grass, Bermuda grass, redtop grass, orchard grass, and sweet vernal grass (Kosisky, Marks, and Nelson 2010). This shows that different vegetation types release different amount pollen causing different allergic diseases. Therefore, there are studies trying to build models to estimate the composition of vegetation from pollen data (Kujawa et al. 2016).

Table 1: Type of Allergens by Species Source: Right Diagnosis from Healthgrades

Tree Pollen Allergies	**Weed Pollen Allergies**	**Grass Pollen Allergies**	**Other Pollen Allergies**
Birch	Ragweed	Timothy Grass	Walnut Tree
Oak	Sagebrush	Kentucky Bluegrass	Maple Tree
Ash	Redroot Pigweed	Bermuda Grass	Sorrel
Pecan	Lamb's Quarters	Johnson Grass	Rapeseed Oil
Hickory	Russian Thistle	Orchard Grass	Beet
Mountain Cedar	English Plantain	Sweet Vernal Grass	Sunflower

2.2.4. Topography

The impacts of climate and topology on pollen dispersal have already been revealed in different species with both direct techniques and indirect techniques of pollen observation (Streiff et al. 1999). Spatial distribution of vegetation cover is generally affected by elevation, aspect and slope (Jin et al. 2008). Also, temperature and wind change depending on elevation, aspect and slope.

2.3. Pollen Allergy and Measurements

While grass pollen is one of the substantial allergens related with atopic disease all over the World, other pollens are especially significant in some areas, e.g. birch pollen in Scandinavia, ragweed in the United States and some parts of Europe, and olive in Mediterranean countries (Burr et al. 2003). The symptoms of hay fever become more frequent and more severe when the pollen count increases beyond a certain threshold (Burr et al. 2003). However, the geographical relationship is unclear between pollen exposure and the underlying prevalence of rhinitis (Burr et al. 2003) because weather conditions may rapidly change and different vegetation species can release a different amount of pollen. Surveys in several countries illustrate that the extensity of rhinitis and atopy is apt to be lower in rural areas (Gergen et al. 1987; Åberg 1989; Bråbäck et al. 1994) specially on farms (Åberg 1989; Braun-Fahrländer 1999; Von Ehrenstein et al. 2000; Riedler et al. 2000; Kilpelainen et al. 2000). On the other hand, in some surveys the occurrence of rhinitis has been similar in urban and rural areas, or higher in an area with high pollen exposure (Burr et al. 2003). Briefly, high pollen exposure means high occurrence of allergic diseases.

Despite its significance, only a few studies have examined statewide variation and determinants of allergic diseases as well as the role of climate on pollen dispersal (Silverberg, Braunstein, and Wong 2014). Knowing the amount of the pollen in a specific time is significant to determine the effect of allergic diseases and establish an early warning system for people who are vulnerable to allergic diseases. Early forecasting of pollen concentration in the atmosphere is very important for medical applications due to the increasing occurrence of allergic diseases induced by allergenic pollen.

During the last two decades, the allergic diseases inducted by allergenic pollen have dramatically increased, as well as the severity of allergic symptoms. Consequently, the social cost of pollen related to diseases has increased greatly. One of the main characteristics of the pollen allergies is the seasonal nature, due to the pollination period that characterizes each plant. The information of seasonal variation may be used to support preventive allergic therapy. In particular, early forecasting of daily airborne pollen concentration could greatly support a better application of preventive allergic therapy (Arca et al. 2002).

However, variations in the composition of outdoor aeroallergens change geographically and seasonally with different vegetation types (Jalbert and Golebiowski 2015; Beggs et al. 2015; Haberle et al. 2014). For example, vegetation has a particular importance on allergic diseases caused by pollen. Diverse types of plants can cause different allergic diseases in different areas. While grass pollens usually cause hay fever in Europe (van Vliet et al. 2002), indigenous vegetation such as Eucalyptus, Acacia and Sorthum grass can contribute to allergic diseases in Australia (Jalbert and Golebiowski 2015).

Briefly, allergic diseases are widespread all over the World, so it is an urgent necessity to create early warning systems for allergy sufferers. For the allergy diseases, pollen data was used in the previous studies (Darrow et al. 2012; Ito et al. 2015; Jariwala et al. 2014). However, large-scaled measurement of pollen has always become one of the challenges because of unpredictable weather conditions and large distribution of vegetation types. In addition, start and peak days of flowering control the time of pollen season and the amount of pollen counts (Bogawski et al. 2015; Garcia-Mozo et al. 2008). Manual pollen count is the most common method used among researchers, but it needs lots of labor and time. Also, the results of manual pollen count are applicable to the place where traps for pollen are installed, so it is difficult to use this data for a large-scale study. In addition, weather data are sometimes collected from weather stations away from pollen traps, so they cannot represent the area well due to changing weather conditions in short distances. In contrast, remote sensing systems are generating large amounts of spatio-temporal data with better accessibility (Turner et al. 2006), which facilitates large-scaled forecasts of precipitation, temperature and vegetation distribution. Lately, there have been some studies which focus on observing pollen releases from specific plants in local areas using satellite data (Karlsen et al. 2009; Luvall et al. 2012; Peng et al. 2013). It was claimed that land surface phenology (LSP) is an important indicator of flowering progress and flowering duration is correlated with temporal vegetation index variation in a specific plant (Karlsen et al., 2009) Also, statistical models such as Neural network facilitate the monitoring of the interactions between environmental variables and pollen concentration or vegetation and pollen concentration.

2.3.1. Remote Sensing and Neural Network

2.3.1.1. Remote Sensing

The dynamics of terrestrial ecosystem can be monitored from satellite observations, and vegetation index from these observed data is primarily used and a powerful parameter for analyzing the features of vegetated land surface (Zhang 2015). The Advanced Very High-Resolution Radiometer (AVHRR) series is the first satellite sensors that has been used and is suitable to monitor land surface vegetation phenology across large areas (Tan et al. 2011). The Moderate Resolution Imaging Spectroradiometer (MODIS) enables a more extensive data source to examine land surface phenology with high quality observations at continental and global scales (Tan et al. 2011, Zhang 2015). The time series satellite observations are capable of monitoring the seasonal development of vegetation growth and pollen seasonality.

Vegetation growth and seasonality can be characterized using several vegetation indices. Normalized difference vegetation index (NDVI) is widely used in remote sensing studies such as examining long-term dynamics of the vegetated land surface and climate impacts. NDVI is calculated with formula below (NIR: Near-infrared band; Red: Red Band):

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

NDVI is sensitive to brightness of soil background and scattering by the atmosphere from highly variable aerosols. The noises in the long-term NDVI time series are essentially resulted from clouds and optically thick aerosols, directional illumination and viewing variations, shadows, topographic variation, geometric errors, and ground snow conditions (Goward et al. 1991; Pinty and Verstraete 1992; Roujean et al. 1992).

The Enhance Vegetation Index (EVI) coordinates with several elements of ecosystem dynamics such as leaf area index, biomass, canopy cover, and the fraction of absorbed photosynthetically active radiation (Boegh et al. 2002; Huete et al. 2006). EVI has been developed to reduce the NDVI limitations. Using EVI, vegetation signal was optimized with improved sensitivity in high biomass regions and improved vegetation monitoring as well as a reduction in atmospheric influences (Houborg, Soegaard, & Boegh, 2007; Zhang 2014). EVI is also less affected by saturation compared to the other vegetation indexes such as NDVI (Huete et al. 2006), so it is more effective for monitoring seasonal, inter-annual, and long-term variation in vegetation structure. EVI is calculated using the following formula (ρNIR: reflectance in the near infrared waveband; ρRED: reflectance in the red waveband):

$$\mathrm{EVI} = 2.5\frac{\rho_{\mathrm{nir}} - \rho_{\mathrm{red}}}{\rho_{\mathrm{nir}} + 6 \times \rho_{\mathrm{red}} - 7.5 \times \rho_{\mathrm{blue}} + 1}$$

More recently, a two band enhance vegetation index (EVI2) has also developed. It is calculated by taking advantage of the autocorrelative properties of surface reflectance spectra between the red and blue wavelengths in the EVI calculation equation:

$$\mathrm{EVI2} = 2.5\frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{RED}}}{\rho_{\mathrm{NIR}} + 2.4\rho_{\mathrm{RED}} + 1}$$

where ρNIR is reflectance in the near infrared waveband and ρRED is reflectance in the red waveband. EVI2 is equivalent to EVI and has advantages over NDVI. EVI2 produces values within an error of ±0.02 of EVI over most land cover types and during different seasons at local and global scales (Jiang et al. 2008), which is fully compatible with EVI (Huete et al. 2006; Jiang et al. 2008; Zhang 2015)

Zhang (2015) renovated temporal EVI2 trajectory using the Hybrid Piecewise Logistic Model (HPLM). The HPLM algorithm is described as:

$$EVI2(t) = \begin{cases} \dfrac{c_1}{1 + e^{a_1 + b_1 t}} + EVI2_b \\ \dfrac{\mathrm{c}_2 + dt}{1 + \mathrm{e}^{\mathrm{a}_2 + b_2 t}} + EVI2_b \end{cases}$$

where t is time in the day of year (DOY), a is related to the vegetation growth time, b is associated with the rate of plant leaf development, c is the amplitude of EVI2 variation, d is a vegetation stress factor, and $EVI2_b$ is the background EVI2 value.

Similar to EVI, two band (red and near-infrared) enhance vegetation index (EVI2) is less sensitive to noises unlike NDVI (Huete et al. 2006; Rocha and Shaver 2009). Moreover, EVI2 is better than NDVI in terms of classifying land cover variety (Mondal 2011), predicting crop yield (Bolton and Friedl 2013), monitoring gross primary productivity (Xiao et al. 2005), capturing subtle changes in vegetation condition and structure (Rocha and Shaver 2009), and detecting land surface phenology (Zhang et al. 2014).

2.3.1.2. Neural Network

Researchers have been trying to understand the correction between pollen and climate, yet their relationship is still poorly understood. There are various methods in analyzing pollen data. The Box-Jenkins methodology (Box and Jenkins, 1976) is one of them, in which the autoregressive integrated moving average (ARIMA) models forecast a variable by a linear combination of the previous state of the variable and the previous forecast errors (Arca et al. 2004). The essential restriction in ARIMA methodologies is the linear structure of the model (Zhang, 2003), which are not capable of capturing nonlinear patterns that affect many environmental phenomena (Arca et al. 2004).

In general, all ANN models present better predictions than other kinds of models like linear regressions (Astray et al. 2016). Unlike the previous statistical models, artificial neural network (ANN) is a model-based nonlinear method (Zhang 1998) with an ability to learn (Puc 2012; Mesa, Carmen, and Cesar 2005), which is capable of performing nonlinear modelling without a priori knowledge about the relationships between input and output variables (Zhang 1998). The ability of learning algorithm helps ANNs learn the existing relationship between input and output (Mesa, Carmen, and Cesar 2005). When the ANNs has “learnt” to carry out the desired function, input values for which the output is unknown can be entered, and the neural network will execute the output (Mesa, Carmen, and Cesar 2005). ANNs has a wide range of use such as solar energy potential (Sozen et al. 2005) and prediction of the fracture parameters of concrete (Ince 2004) in Engineering, developing information system in Computer Science (Heiat 2002), prediction of the river flow forecast in reservoir management in Hydrology (Baratti et al. 2003), gene expression data analysis in Medicine (Tan and Pan 2005).

Recent researches have demonstrated that ANNs is a powerful tool (Zhang et al. 1998; Castellano-Méndez et al. 2005) and forecasting of pollen is one of the major application areas (Sharda, 1994; Sánchez-Mesa et al. 2002; Castellano-Méndez et al. 2005; Puc 2012; Tomassetti et al. 2013; Astray et al. 2016; Liu et al 2017). For instance, it was used to estimate atmospheric ambrosia pollen concentration in Tulsa, OK (Liu et al. 2017) and forecast airborne castanea pollen for ecological and allergological implementation (Astray et al. 2016). The former studies demonstrated relationships between climatic factors and the dispersal of pollen in the air were successfully modelled and analyzed by ANNs (Puc 2012; Liu et al. 2017). In addition, a previous study stated that pollen can be predicted using machine learning and a suite of environmental data from meteorological stations and remote sensing (Liu, 2017).

ANNs is an attractive tool because of the following reasons. Contrary to the traditional model-based methods, ANNs are data-driven self-adaptive methods in that there are few a priori assumptions about the models for problems under study (Zhang et al. 1998). It can be utilized as an effective and an alternative method for the experimental studies whose the mathematical model cannot be formed (Tosun and Ozler 2002; Zain et al. 2012). It has the ability to model more complex nonlinearities and interactions than linear and exponential regression models can offer (Zain et al. 2012).

CHAPTER 3: FRAMING THE PROBLEM AND THEORY

3.1. Framing the Problem

Asthma and eczema are the gravest of the allergic diseases. They are affecting increasing numbers of people every day along with increasing costs. Allergic diseases cost approximately 6 billion USD annually in USA, 3 billion USD in Germany, and 1.6 billion USD in Britain, whereas more than half of this expenses is spent on hospital care and 80% of the entire bill is attributable to the 20% of patients who require the most treatment (Cookson 1999). Therefore, allergic diseases are destroying the social life of people due to fear of death from an asthma attack or anaphylactic shock (Kizilpinar et al. 2010, 623).

Despite its significance, studies on the role of climate on pollen dispersal are limited in large scales (Silverberg, Braunstein, and Wong 2014). Previous studies are mostly in small-scales since pollens manually collected by pollen traps which represent the area where the traps are placed (Skjøth et al. 2012; Goldberg et al. 1988; Charaborty et al. 1992; Arca et al. 2004; Taylor et al. 2014; Kosisky 2010). Even though pollen counting performed manually with using a haemocytometer gives more accurate results, it takes too much time and needs more labor (Mudd and Arathi 2012). There are only 48 stations (National Allergy Bureau) that count the pollen around the United States. As mentioned above, the amount of pollen can be affected by climate, geographical location and time, so the pollen counts from these stations are not able to represent the entire United States. Therefore, it is needed to develop innovative approaches for large-scaled measurements of pollen variation.

Remote sensing technology is getting popular in environmental sciences due to the accurate results and wider accessibility. Our method based on satellite observations will provide satisfactory results because remote sensing data are updated timely and distributed spatially in a high spatial resolution. Remote sensing data provide not only data in vegetation growth, but also climate data (precipitation and temperature). This provides us a good opportunity in understanding of pollen distributions. Furthermore, remote sensing data could help establishing an early warning system for allergic diseases. Thus, people will be aware of what to do or how to protect themselves from these diseases.

The interactions between pollen and environmental factors are not clearly understood. Although different methodologies are used to predict pollen amounts, some of them failed (Goldberg et al. 1988; Arizmendi et al. 1993) due to linear approaches (Zhang, 2003). ANN is a nonlinear statistical analysis, which has been successfully used in the previous studies.

3.2. Theory

To estimate the amount of pollen in a large scale has always become an issue as the intensity of pollen is severely affected by environmental factors because these factors are highly changeable in temporal and spatial. In the former studies, it has been reported that airborne pollen concentration is related to climatic factors. Temperature plays a key role on the characteristics of the vegetative and reproductive growth of plants, and determines the time of flowering. Frenguelli et al. (1991) stated that higher temperature accelerates the process of ripening of flowers, thus the outset of pollen season will be earlier. They also point out two critical issues;

1. Trees need chilling before termination of dormancy permitting bud break.
2. Trees also require a heat before actual growth is resumed and pollination occurs.

Another research indicates that the high airborne pollen concentrations showed a temporary decline when the temperature went down (Peternel 2004). This explains that increasing temperature causes an escalation in pollen amount in late spring. In addition, in the other studies, heavy and long-term rainfalls were reported to sharply decrease airborne pollen concentration but correlations were not strong (Perez et al. 2009; Green et al. 2004; Gottardini and Fabiana 1997; Barnes et al. 2001; Bartková - Ščevková, 2003). Especially a sustained rain or a short but heavy rain washes the air of pollen away. Raindrops falling to the ground with the force of gravity take pollen with them and therefore plants are more prolific at releasing pollen during warm, dry weather (Korpella 2017). Briefly, while there was a strong positive correlation between temperature and pollen, the correlation was weak and negative between precipitation and pollen concentration. It is obvious to say based on the information above, climatic factors (temperature and precipitation) are related with pollen concentration and have a significant role to predict pollen count.

Temporal vegetation index reflects the vegetation growing cycles. Vegetation is very important as a pollen sources, and the amount and time of pollinations depend on the flowering and growth of plants. Flowing/pollen could appear in at the beginning of a growing season for some trees or grasses while in late spring or autumn for others. With optimal water and temperature, vegetation can reach a maximum growth and flowering during a growth season, which means that maximum amount of pollen can be released from the plants. However, if the requirements of water and temperature cannot be provided in the specific times within the growth season, pollen dispersal will be affected negatively or positively depending on increases and decreases in precipitation and temperature. Moreover, vegetation type (or species) has a key role in the intensity of pollen concentration and the amount of pollen released from plants because various vegetation types may release different amounts of pollen in different time of the year. For example, pollen could appear in at the beginning of a growing season for some trees or grasses while in late spring or autumn for others. Type of vegetation can be affected by environmental factors easily (topography, wind, temperature, precipitation etc.), so it is a big challenging to separate various vegetation species at large scales in a year.

New monitoring devices such as remote sensing systems are generating vast amounts of spatio-temporal data with the better accessibility (Turner et al. 2006) thus providing opportunities to create large-scaled models for pollen forecasts. NDVI and EVI are vegetation indexes that facilitate to the measurement of plant covers. Although both indexes are capable of measuring vegetation cover, EVI has shown a better performance in the previous studies (Mondal 2011; Friedl 2013; Xiao et al. 2005; Rocha and Shaver 2009; Yan et al. 2016; Zhang et al. 2014). Also, the 5 km EVI could contain various types of vegetation.

Based on the knowledge mentioned above, the theories used in this project will focus on how pollen dispersal is affected by different environmental factors such as temperature, precipitation and vegetation. It is possible to predict pollen dispersal using the environmental factors in statistical models (generalized regression, neural network analysis etc.). Unlike the other statistical models, artificial neural network (ANN) is a nonlinear statistical method (Zhang 1998) which has the ability of learning (Puc 2012; Mesa, Carmen, and Cesar 2005). This ability helps ANNs learn the existing relationship between input and output (Mesa, Carmen, and Cesar 2005). Some of the statistical models using to predict pollen concentration have been resulted in failures (Goldberg et al. 1988; Arizmendi 1993), so it is important to choose suitable model. The former studies indicate that relationships between climatic factors and the dispersal of pollen in the air were successfully analyzed by ANNs (Puc 2012; Liu et al. 2017).

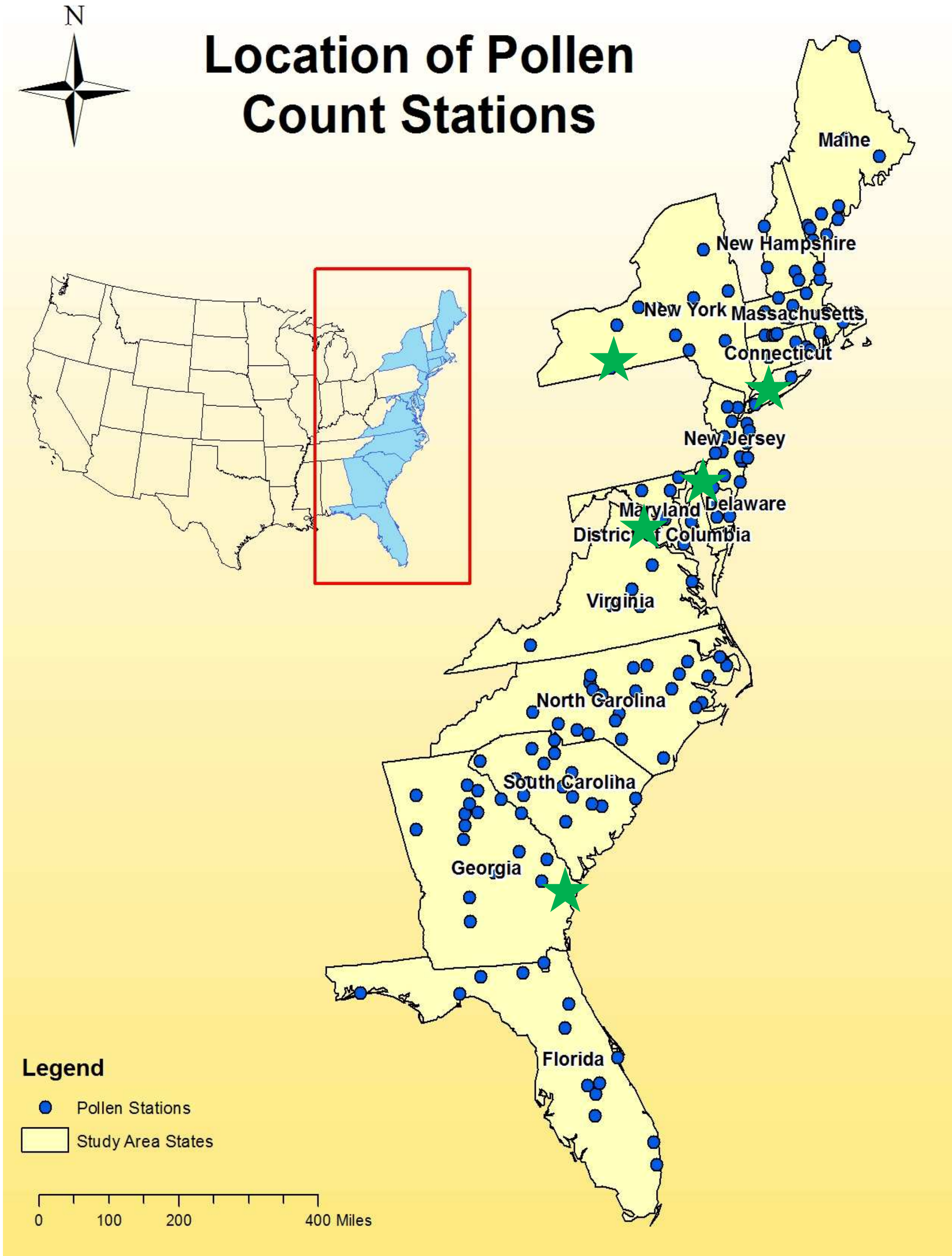
CHAPTER 4: METHOD

4.1. Study Area

The study area covers all East Coast of the United States (15 states and 224 cities, see Figure 1 for the states and Appendix 1 for the all cities). The Eastern Region (especially East Coast of the US) is the most geographically, ecologically, and socially different territory in the United States (USDA 2017). Approximately 105 million people live in the East Coast of the United States, which is recognized as one of the most urbanized and populated area in the United States because this region contains approximately 35% of total population of the country (Table 2).

There are fifteen states which have coastal access to the Atlantic Ocean in this region. The states of East Coast region are: Maine, New Hampshire, Massachusetts, Rhode Island, Connecticut, New York, District of Columbia, New Jersey, Delaware, Maryland, Virginia, North Carolina, South Carolina, Georgia, and Florida (Figure 1). The region from Maine to almost central Connecticut has a continental climate, with warm summers and long, cold and snowy winters. The region from southern Connecticut to almost the Virginia Eastern Shore has a temperate climate with hot summers and cool winters with a mix of rain and snow. The region from southeastern Virginia to central Florida has a humid subtropical climate, with long hot summers and mild winters. The far southern portion of the East Coast from south-central Florida southward has a tropical climate, which is frost free and is warm to hot all year.

Figure 1: Map of Study Area for East Coast of the United States



In northeast of the study area, mean annual precipitation changes up to around 20 inches throughout the Northeast with the highest amounts observed in coastal and some mountainous regions, and also, winters have bitter cold and frozen precipitation, especially in the north with repeated storms (Horton et al. 2014). Summers are warm and humid, especially to the south. The Northeast is periodically exposed to extreme events such as ice storms, floods, droughts, heat waves, hurricanes, and major storms in the Atlantic Ocean (Horton et al. 2014). In the Southeast, there is more precipitation in March and less in October and November, and June through August is a wet period for all states in this area (NCSU 2012). This is due in part to heavy rainfall produced by summer thunderstorms (NCSU 2012).

Moreover, there are roughly 10,000 lakes, over 10,000 miles of streams, and approximately 2 million acres of wetlands in this region. The 10 million acres of national forest system lands are among the largest contiguous blocks of public lands (USDA 2017).

Table 2: Some Information of Study Area by States (Source: US Census of 2010 and 2016)

STATE NAME	POPULATION *(2016)*	TOTAL AREA *(sq. mi) (2010)*	DENSITY *(per sq. mi)*
Maine	1,331,479	35,380	37.63
New Hampshire	1,334,795	9,349	142.77
Massachusetts	6,811,779	10,554	645.42
Rhode Island	1,056,426	1,545	683.77
Connecticut	3,576,452	5,543	645.22
New York	19,745,289	54,555	361.93
New Jersey	8,944,469	8,723	1,025.39
District of Columbia	681,170	68	10,017.21
Delaware	952,065	2,489	382.51
Maryland	6,016,447	12,406	484.96
Virginia	8,411,808	42,775	196.65
North Carolina	10,146,788	53,819	188.54
South Carolina	4,961,119	32,020	154.95
Georgia	10,310,371	59,425	173.5
Florida	20,612,439	65,758	313.46
TOTAL	104,892,896	394,409	265.95

4.2. Data Collection and Their Descriptions

I collected several types of datasets such as precipitation, temperature, pollen count, pollen allergy index and EVI2 to achieve my purposes. The first data I collected were daily real pollen count data from 2002 to 2015 from 5 different stations (National Allergy Bureau) of AAAAI (American Academy of Allergy Asthma and Immunology). Second, I collected pollen allergy index data from pollen.com from April to December of 2016. Pollen.com is a pollen forecasting website for entire United States, and provides us to get daily (present) and historical pollen allergy index values for the periods of 1, 3, and 6 months. I extracted a short-term dataset of pollen allergy index from 225 different locations.

I acquired the daily rainfall time series from the Tropical Rainfall Measurement Mission (TRMM) product 3B42 (post-real-time, Version 7). The TRMM product 3B42 is available as 3-hourly rainfall rate (mm/h) with a spatial resolution of 0.25°. I also obtained land surface temperature (K) (LST) in my research. This data is a 3-day LST data which is extracted from 5 km MODIS LST product.

Vegetation properties can be characterized in several ways such as Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) deriving from MODIS data. In this research, we used EVI2 that has better capability in quantifying seasonal vegetation variation than NDVI (Zhang 2015).

4.3. Neural Network Analyses

After the data collection, my main purpose is to determine the relationship between variables (temperature, precipitation, and EVI2) and pollen. There are different statistical models for this purpose, but neural network analysis is most common and suitable for this task. First of all, correlation analysis between variables was done using R language to see the interactions.

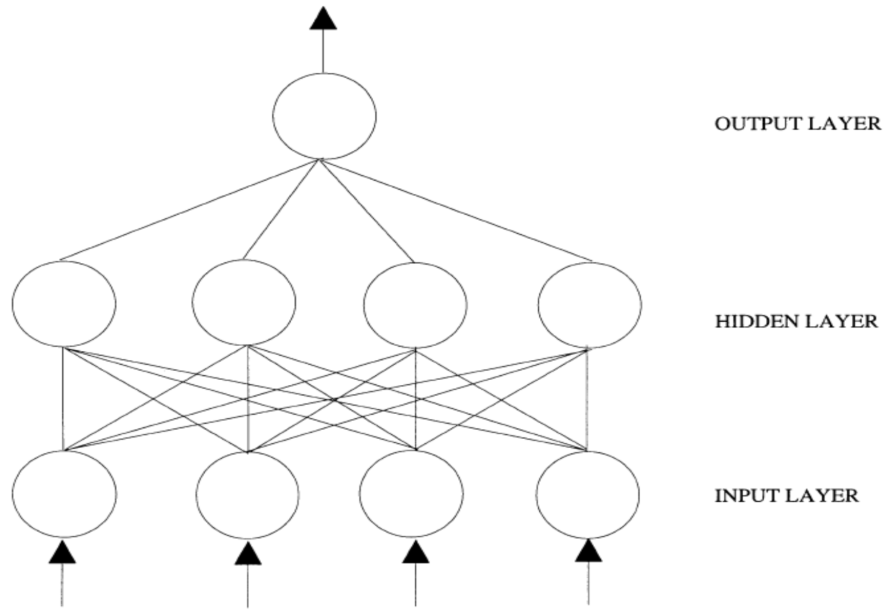
A typical ANNs consists of (Michael 2017) (Figure 2):

- Input layers: Layers that take inputs based on existing data.
- Hidden layers: Layers that use backpropagation to optimize the weights of the input variables in order to improve the predictive power of the model.
- Output layers: Output of predictions based on the data from the input and hidden

layers.

The basic schema of ANNs is shown in Figure 2. Our input variables are land surface temperature, precipitation and EVI2, while output data is pollen count.

Figure 2: A typical feed forward neural network (Zhang 1997)



Package “neuralnet” was used in R environment to execute the model. Data normalization was performed before the training process begins to avoid computational problems (Lapedes and Farber 1988). In the data normalization, basic idea is to squash the variables to typically (0,1) or (-1,1). The approaches of normalization are (Zhang et al. 1998);

- Linear Transformation to [0, 1]: $X_n = (X_0 – X_{min}) / (X_{max} – X_{min})$
- Linear Transformation to [a, b]: $Xn = (b-a) (X_0 – X_{min}) (X_{max} – X_{min}) + a$
- Statistical Normalization: $X_n = (X_0 - \bar{X}) / S$
- Simple Normalization: $X_n = X_0 / X_{max}$

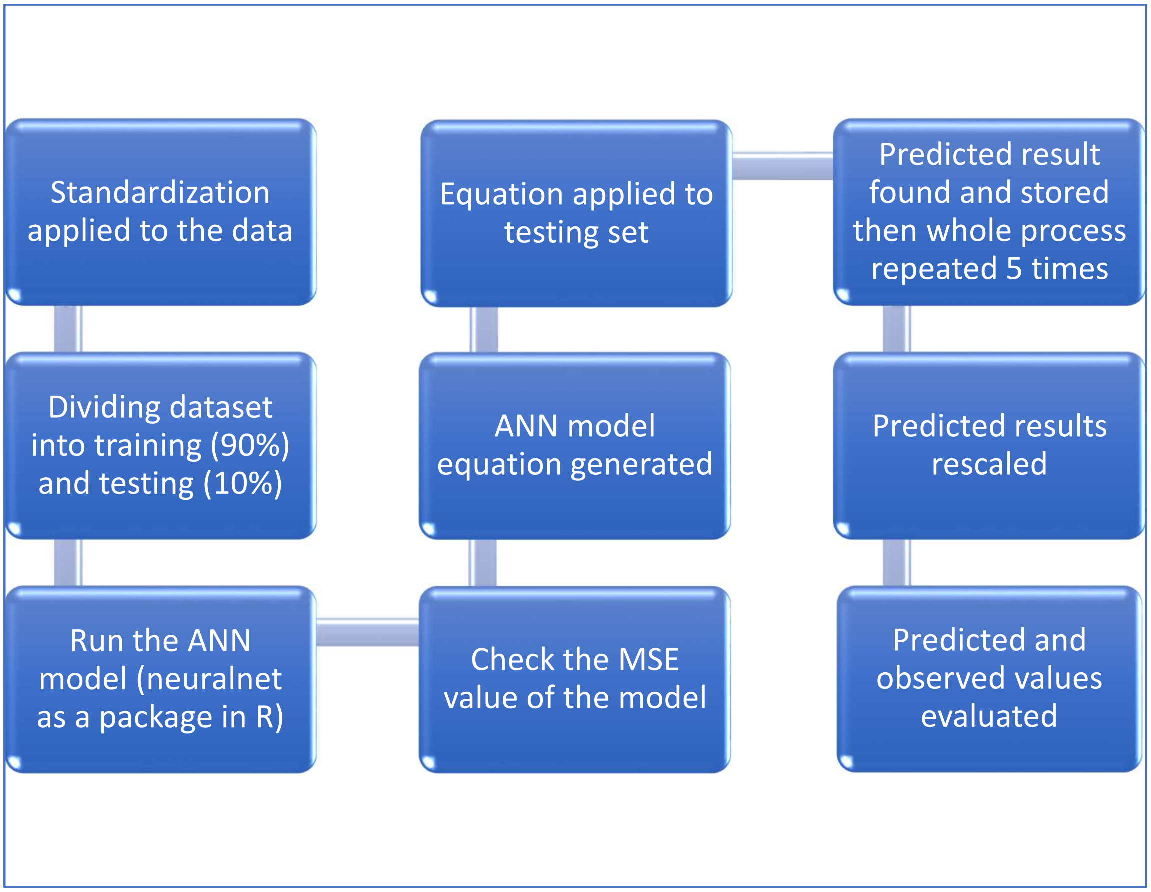
In this study, we used the statistical normalization as it provided better results compared to the other methods. In the previous studies, dataset was classified as training

(70%), validation (10%) and test (20%) (Smith 2005) or 19 years of dataset divided into validation (2011-2012) and training (1993-2010) was used to create ANNs (Astray et al. 2016). It was also reported that most authors use the rule of 90% vs. 10%, 80% vs. 20% or 70% vs. 30%, etc. for training and testing of ANN models (Zhang et al.1998).

4.3.1. Processing Short Term Pollen Allergy Index

A limited term dataset of all cities was collected to run ANN. There are about 60 measurements (every measurement represents 3 days) of daily allergy index for each city, which is later divided into 90% as training and 10% as testing to run ANNs. Standardization is important for ANNs, and the described method in the methodology from R Package was applied to the whole dataset before running ANNs (Figure 3). Package "neuralnet" in R was used to run ANNs after setting suitable parameters of neural net function. Until I acquired a small value of MSE, I changed the parameters (number of hidden layers, stepmax, threshold, etc) in the function and reran ANNs. After obtaining a low MSE, I recorded the value. Then the equation which was acquired from the ANNs was applied to the test data, and the result was recorded. I repeated this entire process 5 times and stored 5 best results (6 measurements, a total of 30 measurements) with the lowest MSE. Finally, the predicted results recorded from ANNs and the observed results were compared by using regression analysis to evaluate how well our model works.

Figure 3: Process of the Short Term Neural Network Analysis



4.3.2. Processing Long Term Pollen Count Data

We have long-term dataset that include precipitation, land surface temperature (LST), EVI2 and daily pollen count for 5 different locations (Table 3).

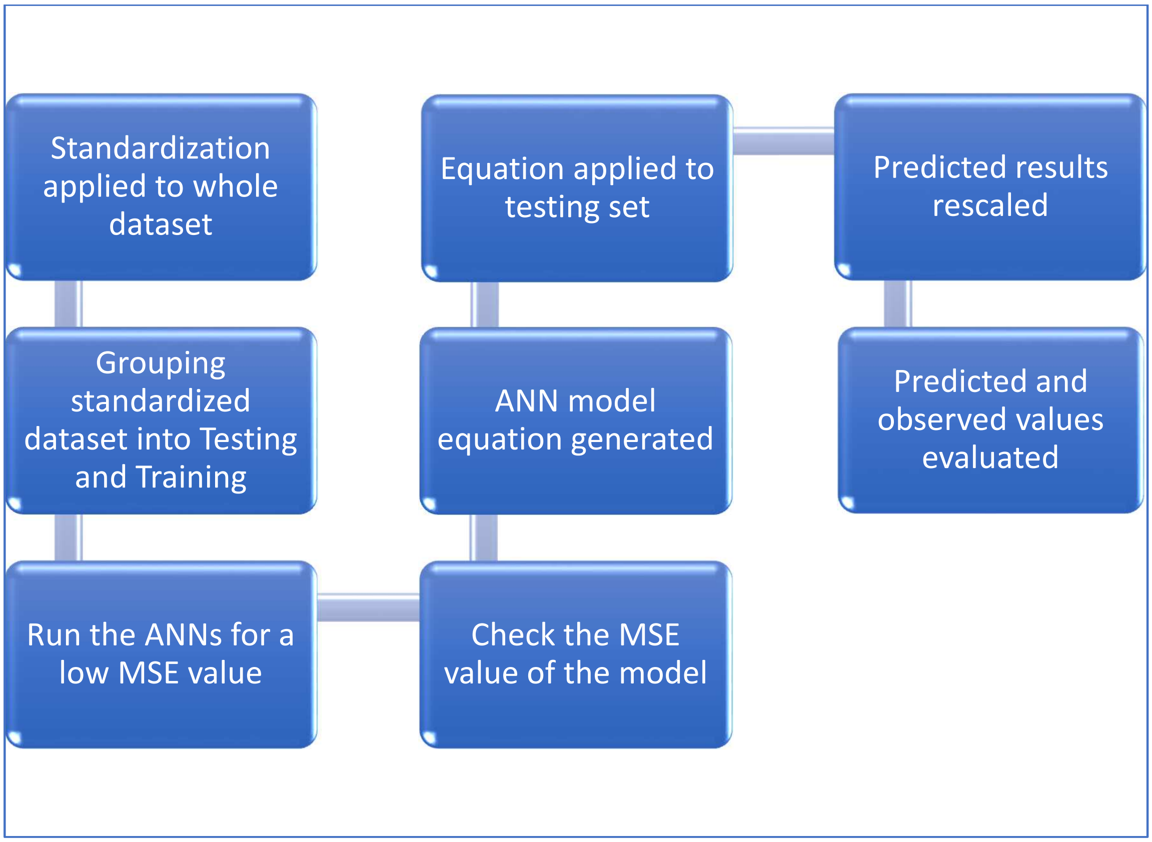
Table 3: Information of Data Range, Testing, and Training for Long Term Locations which using in ANN processes

Long Term Location	**Data Range**	**Testing**	**Training**
New York, NY	11 years	1 years	10 years
Olean, NY	8 years	1 years	7 years
New Castle, DE	12 years	1 years	11 years
Washington, DC	8 years	1 years	7 years
Savannah, GA	14 years	1 years	13 years

Long term datasets are mostly from 2002 to 2015. After collecting precipitation, LST, EVI2 and daily pollen count, every three-day value grouped into one value by taking average of 3 days because every EVI2 and LST data represent every three days. ANNs processing for long-term prediction was similar to ANN analysis to short term data, except for repeating process 5 times because there was enough dataset to run ANNs successfully. There were some missing values in the long-term data, which caused poor result in implementing ANNs. Therefore, average of three days at before or after the missing observations was used to fill the missing value. In this way, the missing values that was replaced with the average values provided better results. After data preparation, I ran the ANN model in R language. The flow chart illustrates the process of ANN for long term data below (Figure 4). First, standardization was applied to whole dataset to see them in a same range. Then, I divided the dataset into testing and training depending on which year gives better result. To do this, different combinations were tried to get better results. For example, I chose one year of data for testing and the rest was chosen for training. I reran the ANN with the different sets of testing and training choosing different years because previous studies related with pollen prediction using ANN suggest 10%, 20%, and 30% or different percentages as a testing. Different combinations were tested to get better results, then the best result stored as a final result.

As with the short-term analysis, mean square error (MSE) value was determinative for the long-term analysis. When the MSE value was high after running ANN with training data, ANN was repeated until to get a lower MSE value. The result with the lowest MSE was recorded and the equation was applied to the testing data. The extracted result from the testing process was compared to the observed pollen concentration by using regression analysis to evaluate how well our model predicted.

Figure 4: The Process of Neural Network Analysis for long term data



CHAPTER 5: RESULTS AND DISCUSSION

The results of our research will help understand interactions between climatic variables and prediction of pollen count so that we can help improving the estimation of pollen distribution during a year and relieving allergy sufferers' problems. For a better understanding of this study, we categorized the results based on short and long-term data analysis.

5.1. Short Term Analysis

In this section, Neural Network Analysis is applied to 225 cities, and the results of 2 cities are unveiled (for the results of the rest of the cities, see appendix 1 in APPENDICES). As it was mentioned in the Methodology, short term data includes rainfall (mm), land surface temperature (LST) (K), EVI2 and pollen allergy index. Annual distribution and comparison line charts of these variables for Adirondack, NY and Gainesville, FL are demonstrated in Figure 5 with the other variables. First graph (Figure 5A) describes the pollen allergy index which was higher from late June up to late September for Adirondack, NY and Gainesville, FL. However, pollen allergy index was in a similar level except some days of the data for Gainesville, FL.

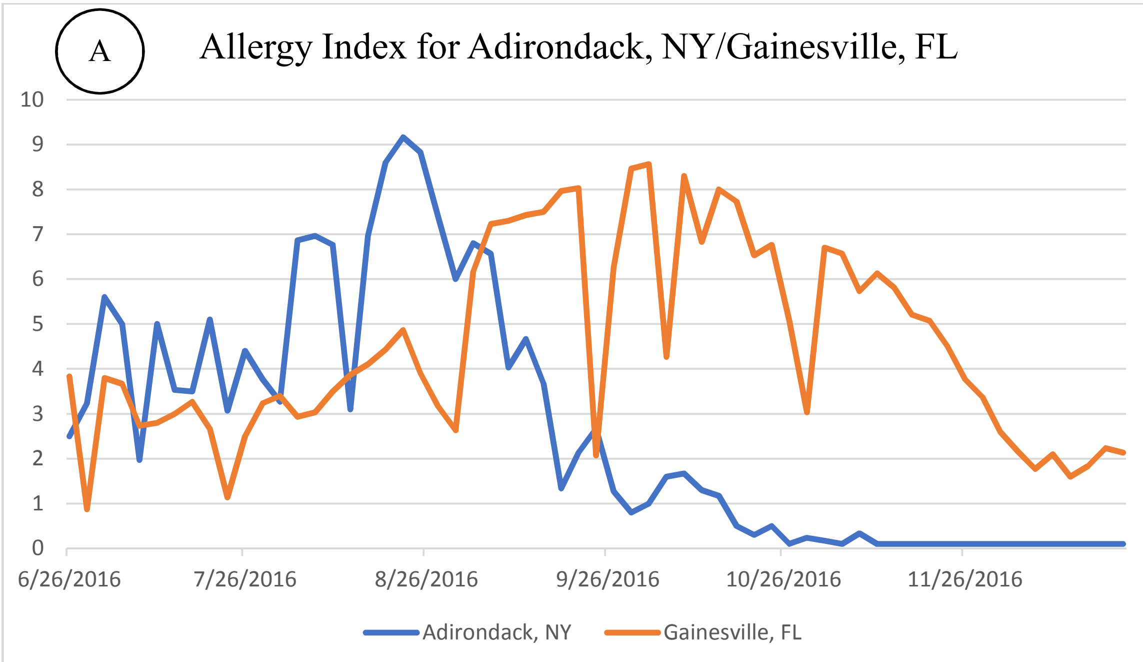
Second graph (Figure 5B) describes about LST. The distribution of the LST was higher in summer months and going down through fall and winter months in Adirondack, NY. The LST pattern for Gainesville, FL shows similarities with the Adirondack, NY. The only difference between them is level of the LST. For example, the difference between summer and winter is about 40 K in Adirondack, NY and 15 K in Gainesville, FL because of their geological locations.

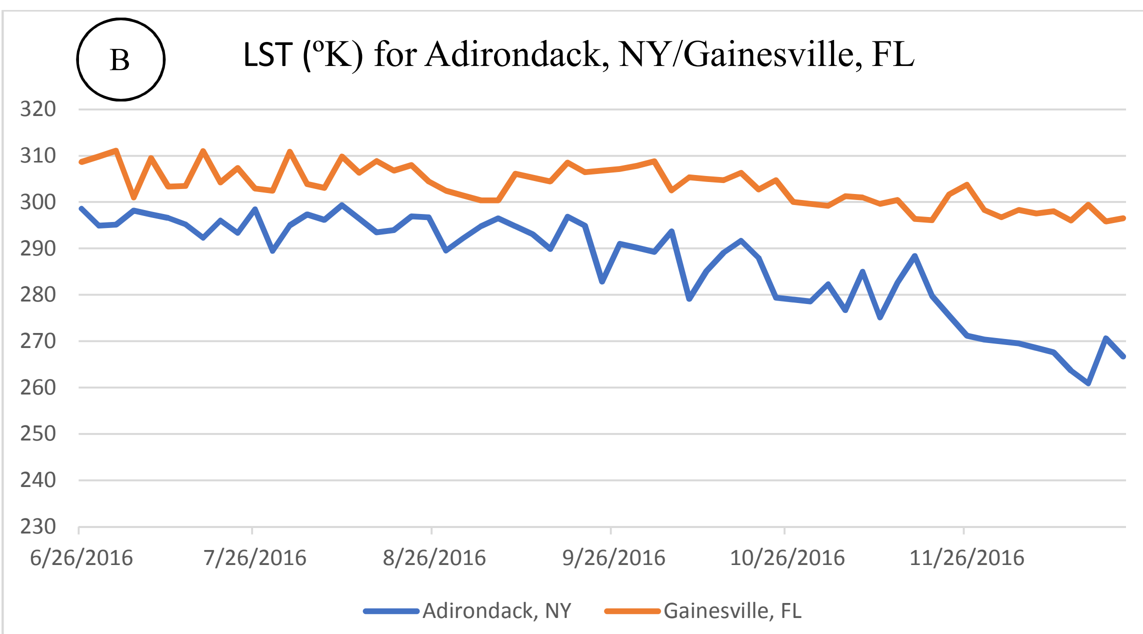
Third graph (Figure 5C) explains for EVI2 which is an indicator of vegetation cover. EVI2 was at higher level and going down through the September in Adirondack, NY. However, EVI2 is not changing much in a year for Gainesville, FL. The most important factor in this difference might be temperature because pollination rate increases with the temperature rises (the late spring and summer times) for cities in the north of the study area such as Adirondack, NY.

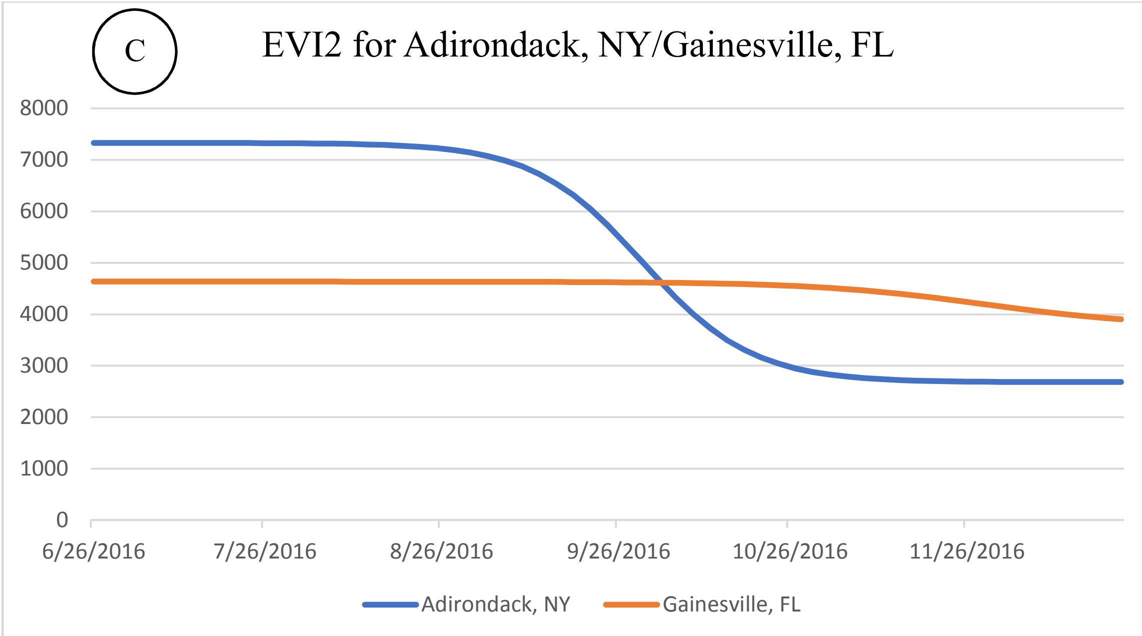
Last variable (Figure 5D) is a rainfall in millimeter (mm). The distribution of the rainfall level is generally low, but it has some peak points (in June and beginning of the July) for Adirondack, NY. Unlike Adirondack, Gainesville has higher level pattern and peaks (in June, late August, and late September are the highest peak times) for rainfall. The reason of the difference between rainfall levels might be ocean effect to Gainesville, FL.

When a correlation analysis was performed between these variables (Figure 6), we found that there was a strong positive correlation between EVI2 and temperature (89%), EVI2 and pollen allergy index (83%), temperature and pollen allergy index (82%). However, rainfall has a negative, weak relationship with the other variables as it was mentioned in some previous studies. Correlations for the other cities in this research show similar results, which are positive, strong relationships between temperature-EVI2, pollen allergy index-EVI2, temperature-pollen allergy index and a negative, weak relationship between rainfall and the others.

Figure 5: Temporal variation of variables. (A) Pollen Allergy Index, (B) land surface temperature, (C) EVI2, and (D) rainfall in time for Adirondack, NY and Gainesville, FL in Adirondack, NY and Gainesville, FL






C
EVI2 for Adirondack, NY/Gainesville, FL
8000
7000
6000
5000
4000
3000
2000
1000
0
6/26/2016
7/26/2016
8/26/2016
9/26/2016
10/26/2016
11/26/2016
Adirondack, NY
Gainesville, FL

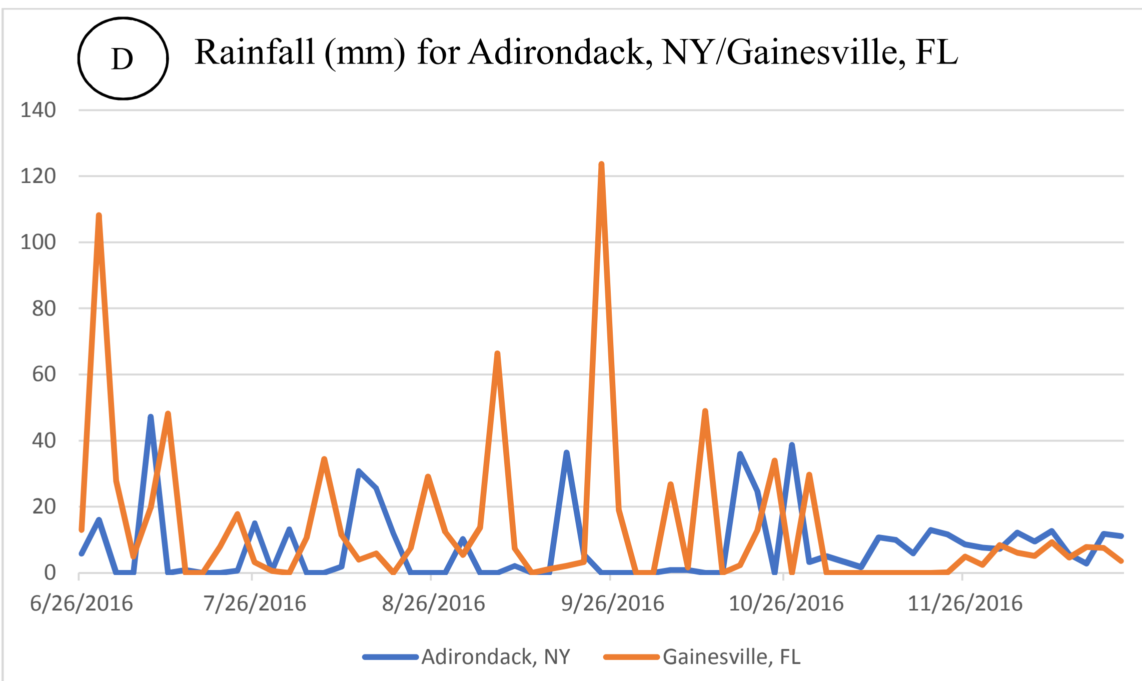
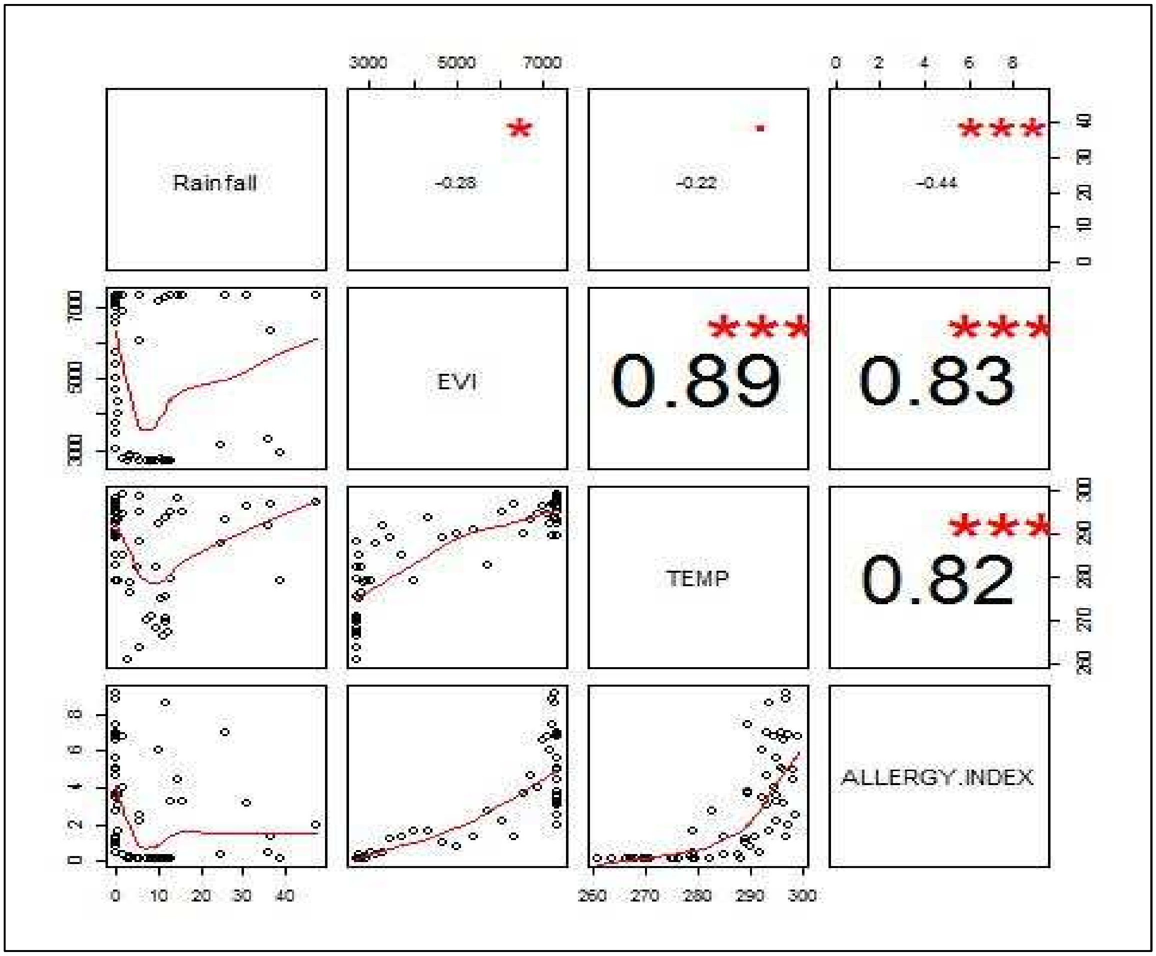
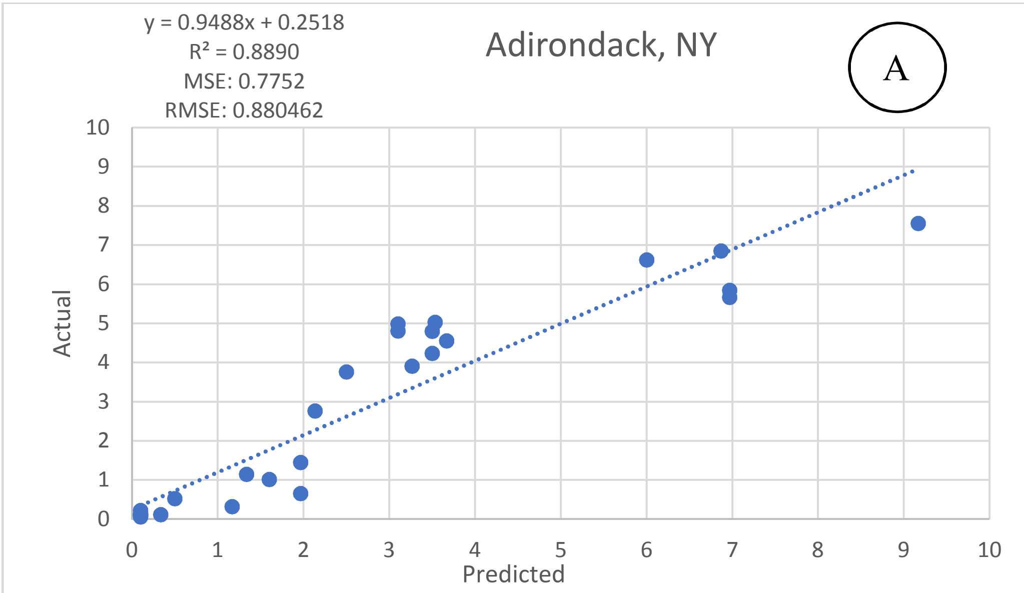

D
Rainfall (mm) for Adirondack, NY/Gainesville, FL
140
120
100
80
60
40
20
0
6/26/2016
7/26/2016
8/26/2016
9/26/2016
10/26/2016
11/26/2016
Adirondack, NY
Gainesville, FL

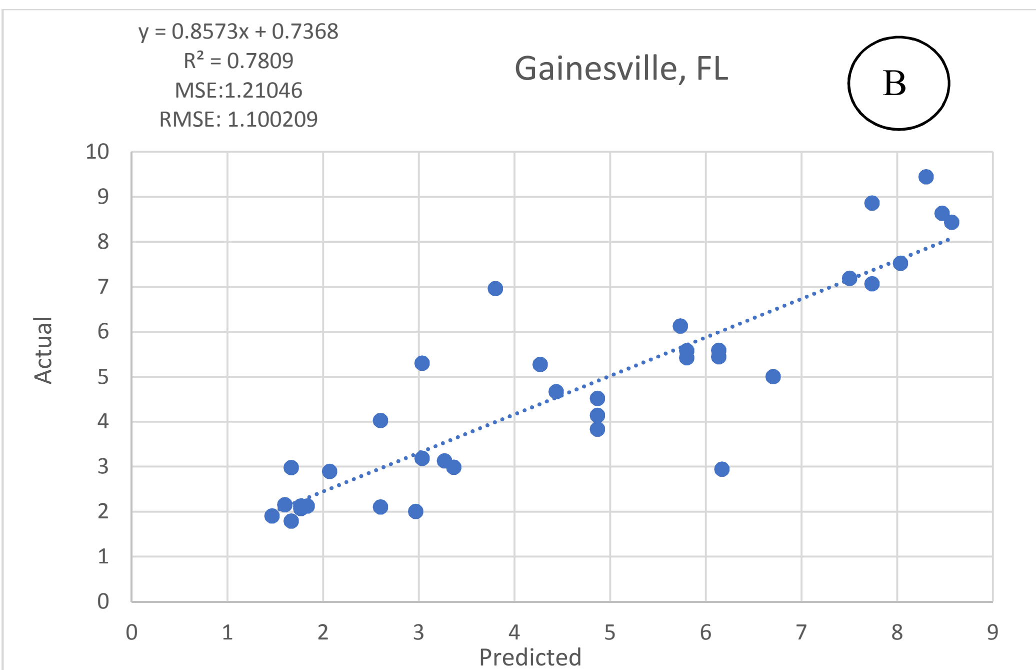
Figure 6: Interactions of variables for Adirondack, NY (rainfall, EVI2, land surface temperature (LST), and pollen allergy index).



Correlations between the variables and pollen allergy index prove that pollen distribution can be estimated from these variables. To prove this idea, the variables were used as inputs in ANN model. The observed and ANN predicted pollen allergy index from two cities (Adirondack, NY and Gainesville, FL) were plotted to manifest the capability of this research (Figure 7). Validation of prediction against the observations was revealed in the regression analysis. The regression plot provided detailed information of slope equations, R squares (R^2), mean square errors (MSE) and root mean square errors (RMSE). The results show low MSEs (0.7752 for Adirondack, NY and 1.21046 for Gainesville, FL) and high R square values (0.8890 for Adirondack, NY and 0.7809 for Gainesville, FL). ANN results for the rest of the cities show similar correlations between observations and predictions although R square varies slightly (see appendix 1). Overall, the results demonstrate that the pollen allergy index could be accurately predicted.

Figure 7: Linear regression between actual and predicted pollen allergy index for Adirondack, NY (A) and Gainesville, FL (B)





Nevertheless, some cities had statistically meaningful results with lower R^2 compared to the other cities. This difference might be due to the fluctuations or the stability of environmental variables within a year. For example, the cities of Florida mostly presented a relatively higher MSE and lower R^2, where the distribution of EVI2 values were pretty stable during the year (see Gainesville). In contrast, EVI2 shows high temporal variation in Adirondack (Figure 6), which associates with high prediction quality. Here, it may be expected that temporal variation of pollen concentration should be stable depending on stable variation of EVI2 and partly changing temperature in Gainesville, but pollen concentration changes over time. The reason behind the unstable pollen concentration may be correlated with slightly changing temperature and extreme variation of rainfall as well as the number of thunderstorms occurring during a year in Florida.

Figure 8: Spatial variation of correlation (R^2) between actual and predicted pollen allergy index in East Coast of the United States

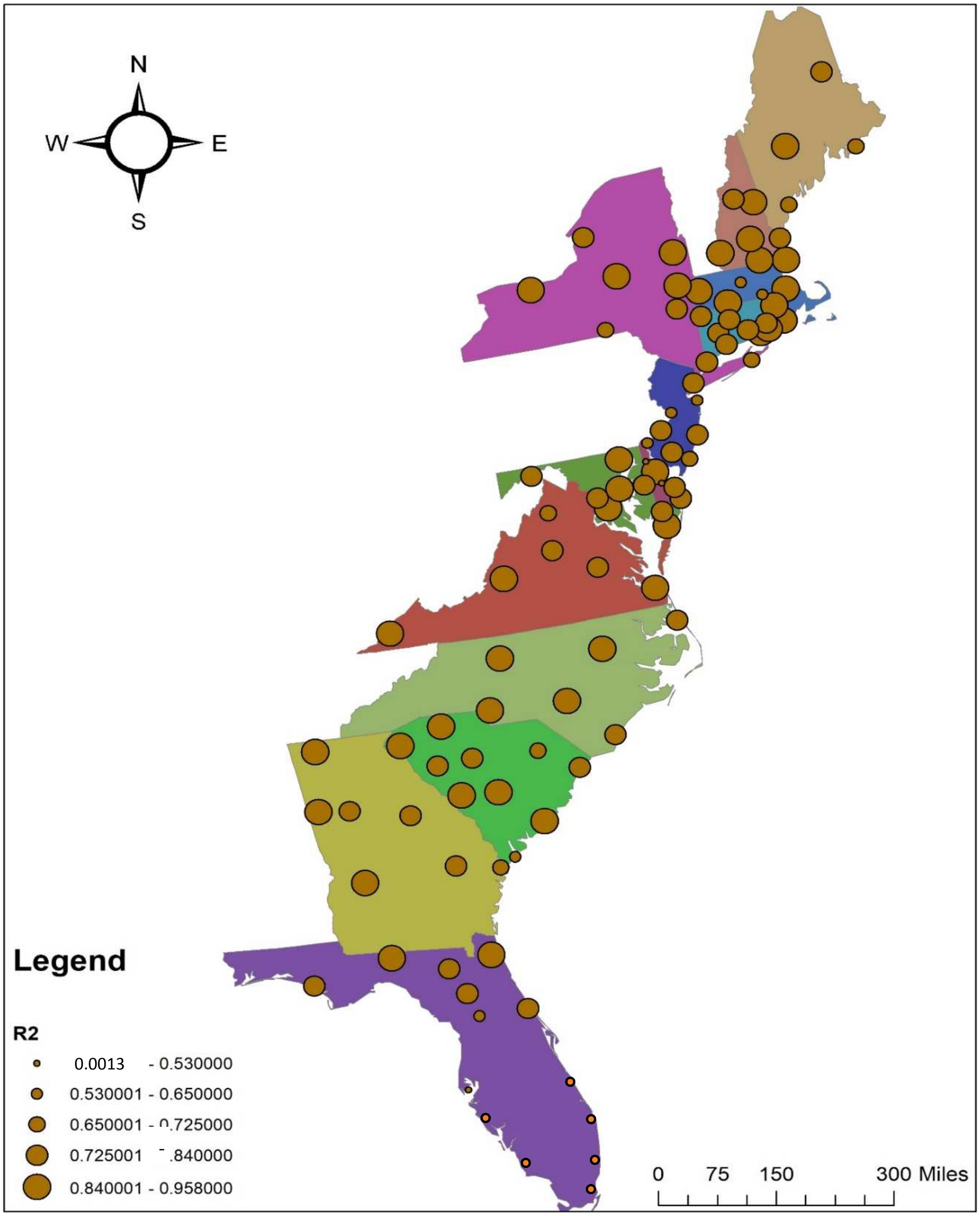
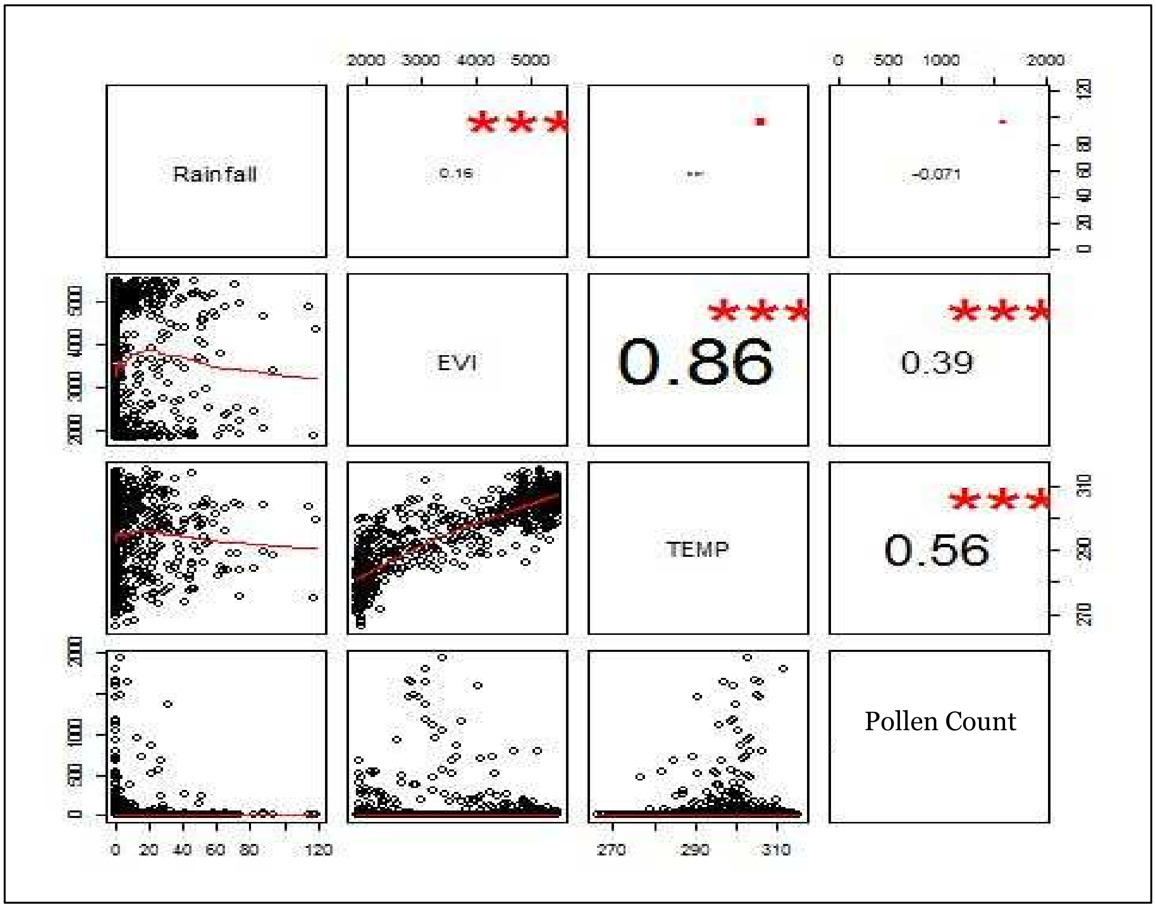


Figure 8 presents the spatial variation of the correlation between actual pollen allergy index and ANN predictions. Based on the map, our results are between 0.53 and 0.95 except for some cities which are mostly in southeast and southwest coast of the Florida. We found very good predictions for many locations, which proves that our model and satellite data (precipitation, land surface temperature, and EVI2) are effective in predicting pollen allergy index. The reason of the low R^2 values might be attributed to heavy and high-level rain and other climatic events. In addition to this, total amount of precipitation used in the analyses might not be good index. Future work is needed to investigate this problem.

5.2. Long term Analysis

Two long-term datasets out of five were explained in this section (the rest of the result, see appendix 2 and 3), which were from for Washington, DC and New York, NY. The same variables as in short term analysis were used to execute ANN. But the time series of pollen count data from stations instead of pollen allergy index were used for long term. Relationships between these variables are shown in Figure 9. The results of correlation analyses are similar to those in short terms, which are positive correlations between pollen count and EVI2 (0.39), temperature (LST) (0.56), except for rainfall (-0.071). Further, positive and strong correlation (0.86) was found between EVI2 and LST. Similar relationship was also found for many locations in rest of the analysis.

Figure 9: Correlation between long term variables of Washington, DC



One-year pollen distribution was chosen to simplify the graph representation from New York, NY (2003) and Washington, DC (2002). Figures 10 and 11 illustrate the seasonal pollen distribution from actual observations and ANN predictions. Pollen concentration was higher from the spring to early summer for New York City, NY on actual values, and our predicted results show mostly similar pattern. On the other hand, pollen concentration was higher from late spring to early summer in Washington, DC on both actual and predicted results, which was similar to New York City, NY.

R square value between actual and predicted pollen counts is 0.8319 for New York and 0.8093 for Washington, which reveals the reliability of ANN model prediction.

The high R square indicates that the LST, EVI2 and precipitation can explain very well the temporal variation in pollen counts although the pollen releases could also be impacted by other environmental variables including high winds, storms, tornadoes, and heavy rain and their irregular behaviors in different years.

Figure 10. Temporal variation in actual and predicted pollen counts (2003) for NYC, NY

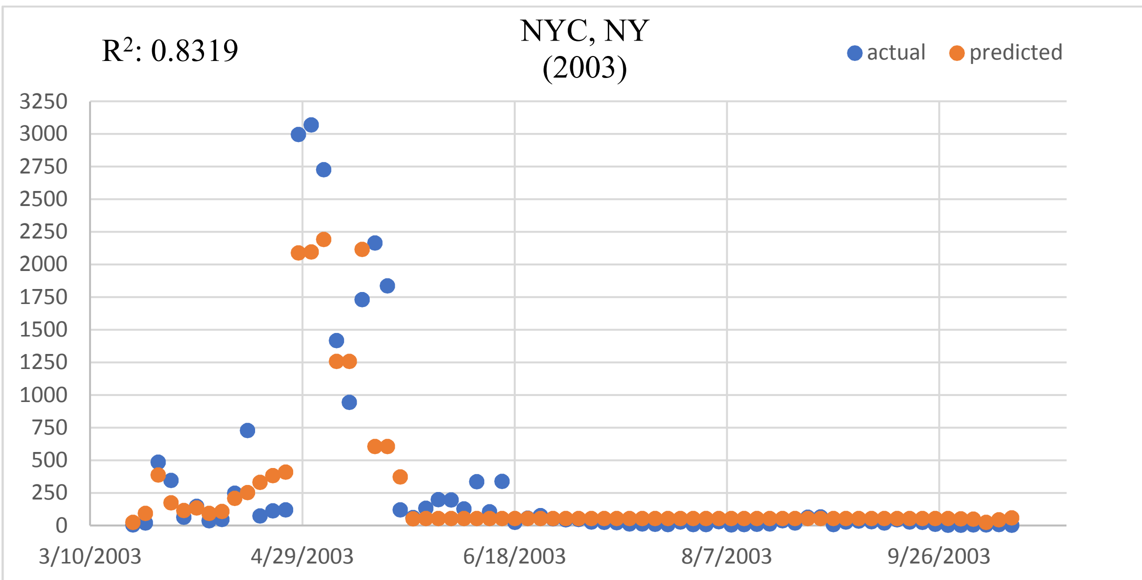
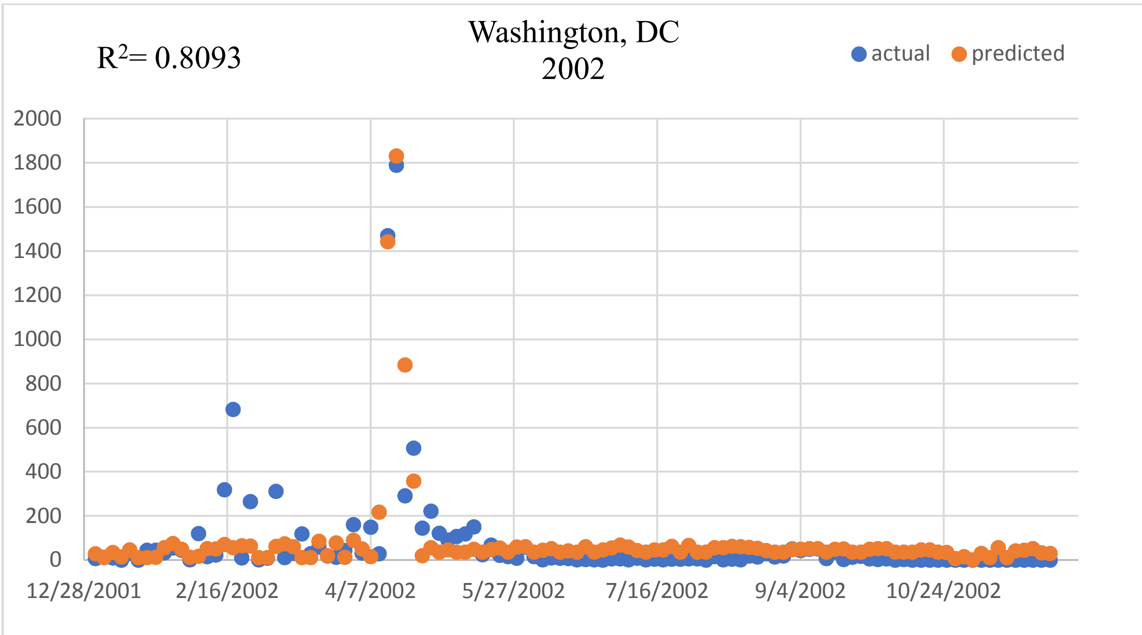


Figure 11. Temporal variation in actual and predicted pollen counts (2002) for Washington, DC



Overall, our ANN models enabled useful estimations of pollen concentrations for long term data. The intensity of predicted and observed pollen for both cities was highest though the spring months, it was pretty stable in the rest of the time, the value close to zero (Figures 10 and 11). As it was found in the short-term analysis, LST, precipitation and EVI2 have significant impacts on pollen concentration in long terms too.

CHAPTER 6: CONCLUSION

Many people are suffering from allergic diseases. This puts people in unsocial lives, which causes stress. Also, allergic sufferers pay lots of money for allergy relief in prescription medications that are costly. Despite its importance, studies are limited because the amount of pollen can easily change depending on environmental factors (temperature, precipitation and vegetation), human-induced factors (industrialization-urbanization), and collecting and counting pollen grains take a long time as well as need more labor.

Remote sensing system is a popular tool due to its multifunctional features and accurate results. It also facilitates researchers to deal with large datasets like environmental and vegetation data as well as provides high accuracy of the results in large scales. The consequences obtained will hopefully provide the information needed for establishing and improving better early warning systems that show intensive pollen releases. These systems may help people be aware of the days or specific times when the intensity of pollen is high.

In general, R^2 values were pretty high (over 0.5) for both short and long-term analysis, but there were remarkable differences between higher values of northeast cities and lower values of southeast cities in the study area. These differences can be explained by correlation analysis between the variables because it was a strong hint of how pollen concentration interacts with environmental variables and vegetation. Although the correlation between rainfall and pollen was low in most locations, it didn't mean there was no relationship between them. While pollen concentration changed depending mostly on precipitation, temperature, and vegetation in the northeast, it changed depending mostly on precipitation in the Southeast. This proves that pollen concentration is closely related to precipitation, temperature and vegetation seasonality, and shows EVI2 data derived from MODIS is reliable for modelling pollen counts. Moreover, the results show that ANN is a suitable statistical technique for this study.

APPENDICES

Appendix 1: Table of the All Short-Term Results

CITY	**STATE**	R^2	MSE	RMSE
Adirondack	NY	0.889	0.775213	0.880461811
Aiken	SC	0.9057	0.627645	0.792240494
Albany	GA	0.8473	0.77122	0.878191323
Albany	NY	0.8903	1.111584	1.05431684
Americus	GA	0.9155	0.688272	0.8296216
Anderson	SC	0.8751	0.792722	0.89034937
Annapolis	MD	0.8494	1.380698	1.175031063
Appalachia	VA	0.9449	0.379342	0.615907461
Arlington	VA	0.7953	1.580713	1.257264093
Armonk	NY	0.2069	8.819109	2.96969847
Asbury Park	NJ	0.5657	4.690998	2.165871187
Asheville	NC	0.7524	1.264272	1.124398506
Athens	GA	0.9536	0.404657	0.63612656
Atlanta	GA	0.8974	0.711065	0.843246702
Auburn	ME	0.668	2.744066	1.656522261
Auburn	NY	0.936	0.678256	0.823562991
Augusta	GA	0.6567	1.941652	1.393431735
Augusta	ME	0.6367	2.037555	1.427429508
Baltimore	MD	0.6531	3.261483	1.805957641
Barre	MA	0.6375	2.181897	1.477124572
Bear Creek	NC	0.9598	0.380396	0.616762515
Beaufort	SC	0.6238	1.768403	1.329813145
Bel Air	MD	0.8811	1.328857	1.1527606

Bethany Beach	DE	0.7937	1.864611	1.365507598
Binghamton	NY	0.7214	3.232053	1.797791145
Blue Hill	ME	0.7164	1.684499	1.297882506
Boston	MA	0.7304	1.738843	1.318651963
Bozrah	CT	0.6742	4.002811	2.000702627
Bradford	RI	0.8802	0.904114	0.950849094
Bridgeport	CT	0.6798	2.87499	1.695579547
Bridgewater	ME	0.7381	1.57317	1.254260738
Brockton	MA	0.864	1.128103	1.062121933
Brooklin	ME	0.7513	1.653191	1.285764753
Brownville Junction	ME	0.739	1.44674	1.202805055
Brunswick	GA	0.3701	4.01896	2.004734396
Buffalo	NY	0.7812	2.544715	1.595216286
Burlington	NC	0.9045	0.850648	0.922305806
Camden	NJ	0.8084	2.338867	1.529335477
Canaan	CT	0.8987	1.149104	1.071962686
Carrollton	GA	0.8796	0.806182	0.897876383
Cartersville	GA	0.9102	0.703274	0.838614333
Catskill	NY	0.8351	2.204745	1.484838375
Charleston	SC	0.8526	0.827145	0.909475123
Charlotte	NC	0.8588	0.696138	0.834348848
Charlottesville	VA	0.7513	1.833537	1.354081608
Chesapeake Beach	MD	0.8392	1.97329	1.40473841
Chesterfield	NH	0.5963	3.314934	1.820696021
Clearwater	FL	0.4156	1.712019	1.308441439
Cocoa	FL	0.4936	1.822692	1.350071109
Colebrook	CT	0.8174	2.337165	1.528778925
Collins	GA	0.8318	0.638885	0.799302821

Columbia	NC	0.4624	3.792525	1.947440628
Columbia	SC	0.8941	0.591842	0.76931268
Columbus	GA	0.9338	0.678398	0.823649197
Compton	MD	0.8735	1.59861	1.24436084
Concord	NH	0.8573	1.130795	1.063388452
Cornish	ME	0.8479	0.999008	0.999503877
Covington	GA	0.9053	0.658555	0.811514017
Dalton	GA	0.8582	1.031506	1.015630838
Daytona Beach	FL	0.7625	0.917328	0.957772416
Deland	FL	0.6347	0.875057	0.935444814
Douglasville	GA	0.7669	1.055038	1.027150427
Dover	DE	0.9139	1.231441	1.109703113
Durham	NC	0.736	2.150664	1.466514235
Eagle Rock	VA	0.9196	0.500005	0.707110317
Elizabeth City	NC	0.7945	2.563687	1.601151773
Elmira	NY	0.9474	0.363315	0.60275617
Essex	MA	0.9034	1.113506	1.055227937
Fair Haven	NJ	0.6398	3.684501	1.919505405
Falls Village	CT	0.7808	2.63487	1.623228265
Fayetteville	NC	0.8543	0.726979	0.852630635
Fernandina Beach	FL	0.0597	5.434548	2.331211702
Florence	SC	0.7233	1.588046	1.260176972
Fort Lauderdale	FL	0.0661	1.076802	1.037690705
Fort McCoy	FL	0.6995	0.93977	0.969417351
Fort Myers	FL	0.2941	1.972829	1.404574313
Fort Pierce	FL	0.0013	1.732174	1.316120815
Frederick	MD	0.9043	1.066881	1.032899317
Fredericksburg	VA	0.7987	1.589667	1.260819971

Freeport	ME	0.7746	1.683809	1.297616661
Fremont	NH	0.821	1.570683	1.253268926
Gaffney	SC	0.8336	0.870688	0.933106639
Gainesville	FL	0.7809	1.210461	1.100209525
Gainesville	GA	0.9595	0.389272	0.623916661
Gastonia	NC	0.8851	0.829029	0.910510296
Georgetown	SC	0.7325	1.490246	1.220756323
Glens Falls	NY	0.8839	1.210684	1.100310865
Gloversville	NY	0.8676	1.54063	1.241221173
Goldsboro	NC	0.9158	0.770813	0.877959566
Greenfield	MA	0.8644	1.32946	1.153022116
Greensboro	NC	0.9006	0.845342	0.91942482
Greenville	NC	0.9098	0.76191	0.872874561
Greenville	SC	0.8448	1.100978	1.049274988
Greenwood	SC	0.834	0.571518	0.755988095
Griffin	GA	0.9106	0.642087	0.801303313
Hagerstown	MD	0.6109	4.501673	2.121714637
Hartford	CT	0.7694	2.3798	1.54266004
Haverhill	MA	0.7	2.255915	1.501970373
Hilton Head Island	SC	0.6416	1.327881	1.15233719
Homestead	FL	0.2294	1.096261	1.047024833
Hudson	NY	0.8347	1.95246	1.397304548
Island Falls	ME	0.7857	0.930441	0.964593697
Ithaca	NY	0.6946	2.002182	1.414984806
Jacksonville	FL	0.9034	0.598043	0.773332399
Jarvisburg	NC	0.8092	1.947814	1.395641071
Jonesboro	GA	0.797	1.523365	1.234246734
Kennebunk	ME	0.651	1.746087	1.321395853

Lake City	FL	0.8288	1.043848	1.021688798
Lakeland	FL	0.4632	2.362954	1.537190294
Lakeville	CT	0.9063	1.038142	1.018892536
Lancaster	SC	0.9227	0.623823	0.789824664
Lawrenceville	GA	0.7514	1.414839	1.18947005
Lee	FL	0.8207	1.068701	1.033779957
Lewes	DE	0.7749	2.42043	1.55577312
Lexington	NC	0.9184	0.653355	0.80830378
Lincoln	NH	0.8367	1.196895	1.094026965
Littleton	MA	0.867	1.341217	1.158109235
Lynchburg	VA	0.831	1.302381	1.141219085
Macon	GA	0.9272	0.595953	0.771979922
Madison	ME	0.8697	0.653873	0.80862414
Manahawkin	NJ	0.7408	2.63629	1.623665606
Manchester	NH	0.529	2.747193	1.657465837
Marietta	GA	0.8872	0.674329	0.821175377
Matthews	NC	0.893	0.695472	0.833949639
Melbourne	FL	0.2375	2.334319	1.527847833
Miami	FL	0.0981	1.325042	1.151104687
Midway Park	NC	0.6756	1.083909	1.041109504
Milford	DE	0.528	4.642564	2.154660994
Monticello	FL	0.9124	0.473432	0.688063951
Morristown	NJ	0.5858	4.070183	2.017469455
MT. Laurel	NJ	0.9661	0.327478	0.572256935
Myrtle Beach	SC	0.8334	1.233075	1.110439102
Mystic	CT	0.8993	1.083832	1.041072524
Naples	FL	0.1136	1.644999	1.282575144
Nashua	NH	0.8622	1.107296	1.052281331

New Bern	NC	0.6861	1.942256	1.393648449
New Castle	DE	0.8979	1.244131	1.115406204
New Hampton	NH	0.7055	2.874723	1.695500811
New Haven	CT	0.7685	1.732633	1.31629518
Newark	NJ	0.6263	4.001109	2.000277231
Newport	RI	0.9025	1.424412	1.193487327
Norfolk	VA	0.8653	1.421544	1.192285201
North Uxbridge	MA	0.6041	3.987268	1.996814463
NYC	NY	0.7731	2.165166	1.471450305
Ocala	FL	0.5561	1.501329	1.225287313
Ocean City	NJ	0.7109	2.545223	1.595375504
Oldtown	MD	0.8029	2.156333	1.468445777
Olean	NY	0.2488	5.454993	2.335592644
Orangeburg	SC	0.8969	0.560121	0.74841232
Orlando	FL	0.4259	1.594995	1.262931115
Panama City	FL	0.8287	0.811351	0.900750243
Peak	SC	0.828	1.136145	1.065901027
Pelion	SC	0.4033	3.693523	1.921854053
Pensacola	FL	0.4474	4.082976	2.020637523
Peterborough	NH	0.8451	1.163276	1.078552734
Pittsfield	MA	0.905	0.849744	0.9218156
Plymouth	MA	0.6255	3.438414	1.854296093
Pocomoke City	MD	0.9565	0.597608	0.773051098
Portland	ME	0.6774	2.351497	1.533459162
Portsmouth	NH	0.8071	1.576757	1.25568985
Poughkeepsie	NY	0.877	1.58207	1.257803641
Prospect Harbor	ME	0.7032	1.695983	1.302299121
Providence	RI	0.8526	1.624144	1.274419083

Raleigh	NC	0.9231	0.727575	0.85298007
Rehoboth Beach	DE	0.6686	2.803271	1.674297166
Richmond	VA	0.8145	2.002357	1.415046642
Ridgely	MD	0.8347	2.241742	1.497244803
Rileyville	VA	0.6553	4.21415	2.052839497
Roanoke	VA	0.7509	2.127448	1.458577389
Rochester	NH	0.8015	1.542539	1.241989936
Rochester	NY	0.9004	1.203113	1.096865078
Rock Hill	SC	0.918	0.671339	0.819352793
Rockport	ME	0.9182	0.618409	0.786389852
Rockville	RI	0.8057	1.39828	1.182488901
Rocky Mount	NC	0.881	0.9356	0.967264183
Rome	GA	0.9139	0.667231	0.81684209
Saint Augustine	FL	0.185	3.398241	1.843431854
Saint George	ME	0.6164	2.260272	1.503420101
Saint Petersburg	FL	0.4146	1.93454	1.390877421
Salem	CT	0.8039	2.091745	1.446286624
Salisbury	MD	0.8133	1.822864	1.350134808
Salisbury	NC	0.9258	0.819697	0.905371195
Sarasota	FL	0.3008	1.727416	1.314311987
Savannah	GA	0.6895	1.179488	1.086042356
Seneca	SC	0.8559	0.9909766	0.995478076
Silver Lake	NH	0.9313	0.662978	0.81423461
Silver Spring	MD	0.8099	1.574555	1.254812735
Smithfield	NC	0.8749	0.819306	0.905155235
Southampton	NY	0.6993	2.828814	1.681907845
Sparta	GA	0.8117	1.388243	1.178237243
Spartanburg	SC	0.8544	1.000256	1.000127992

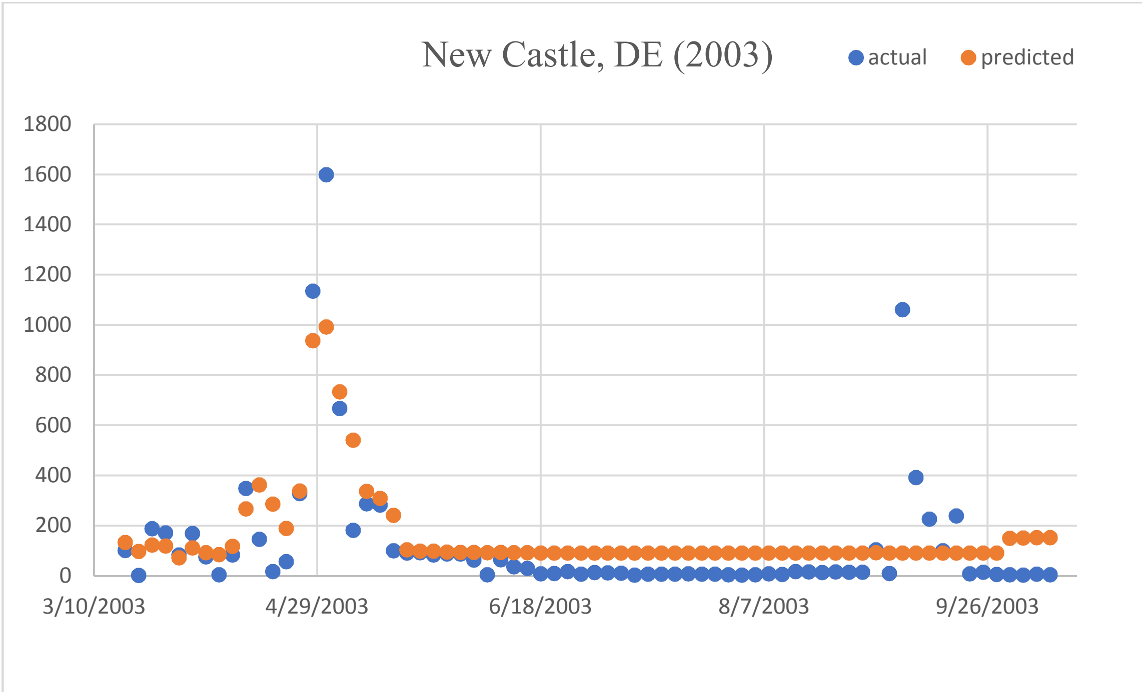
Springfield	MA	0.8756	1.142344	1.06880494
Springfield	NJ	0.8796	1.479976	1.216542642
Stamford	CT	0.8219	2.769432	1.66416105
Statesboro	GA	0.7166	1.012334	1.0061481
Stockholm	ME	0.5426	3.742638	1.934589879
Sumter	SC	0.6542	1.5216	1.233531516
Sunbury	NC	0.8578	1.058965	1.029060251
Swainsboro	GA	0.76	1.646097	1.283003118
Syracuse	NY	0.8245	1.212214	1.101005904
Tallahassee	FL	0.7283	1.668293	1.291624171
Tampa	FL	0.1752	2.58116	1.606598892
Toms River	NJ	0.7285	2.687894	1.639479796
Townsend	DE	0.5174	5.105356	2.259503485
Trenton	NJ	0.6419	3.536916	1.88066903
Utica	NY	0.8886	0.937756	0.968378025
Ventnor City	NJ	0.9627	0.25641	0.50636943
Vineland	NJ	0.7663	2.055529	1.433711617
Waldorf	MD	0.8887	0.938095	0.968553044
Waleska	GA	0.8568	0.776716	0.881314927
Walpole	NH	0.8998	0.754686	0.868726654
Warrenton	VA	0.9292	1.305093	1.14240667
Waterbury	CT	0.796	1.856321	1.362468715
Watertown	NY	0.7469	1.501987	1.225555792
West Ossipee	NH	0.6479	1.84098	1.356827181
West Palm Beach	FL	0.0518	1.240575	1.113811025
West Point	NY	0.8197	1.767346	1.329415661
West Townsend	MA	0.8905	1.708086	1.306937642
Wilmington	DE	0.5909	2.931524	1.712169384

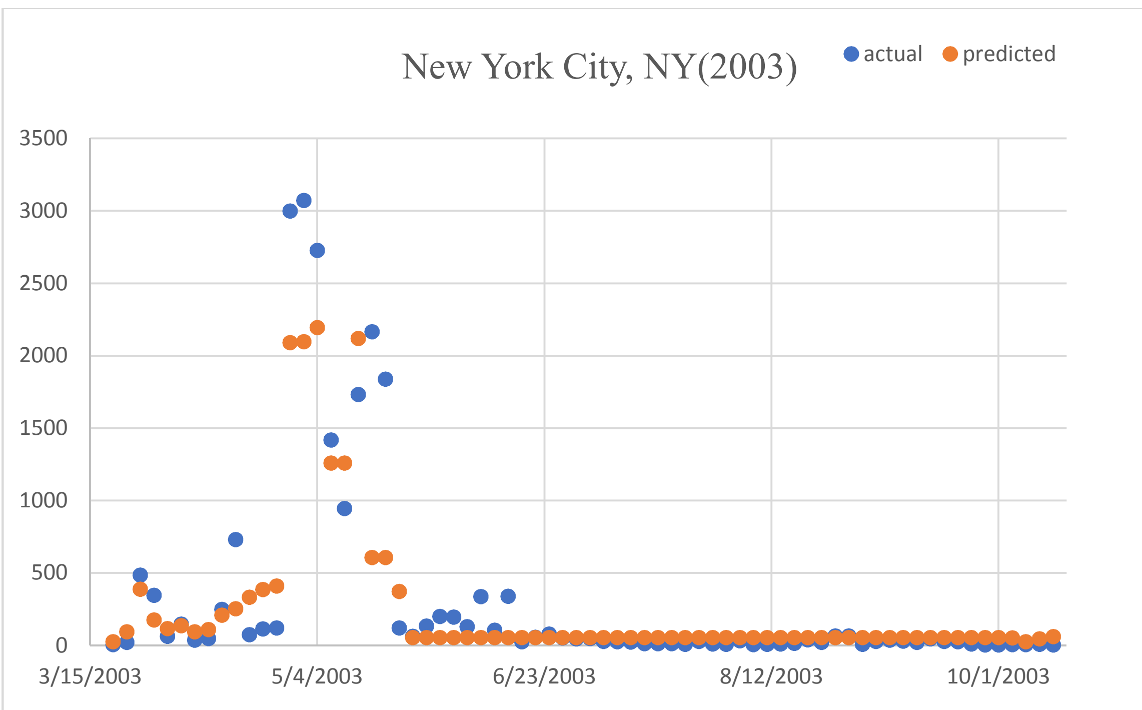
Wilmington	NC	0.7771	1.0934	1.045657688
Wilson	NC	0.827	1.237107	1.112253119
Winston Salem	NC	0.8764	0.861612	0.928230575
Wolfeboro	NH	0.7068	2.107241	1.451633907
Worcester	MA	0.5496	3.817855	1.953933213

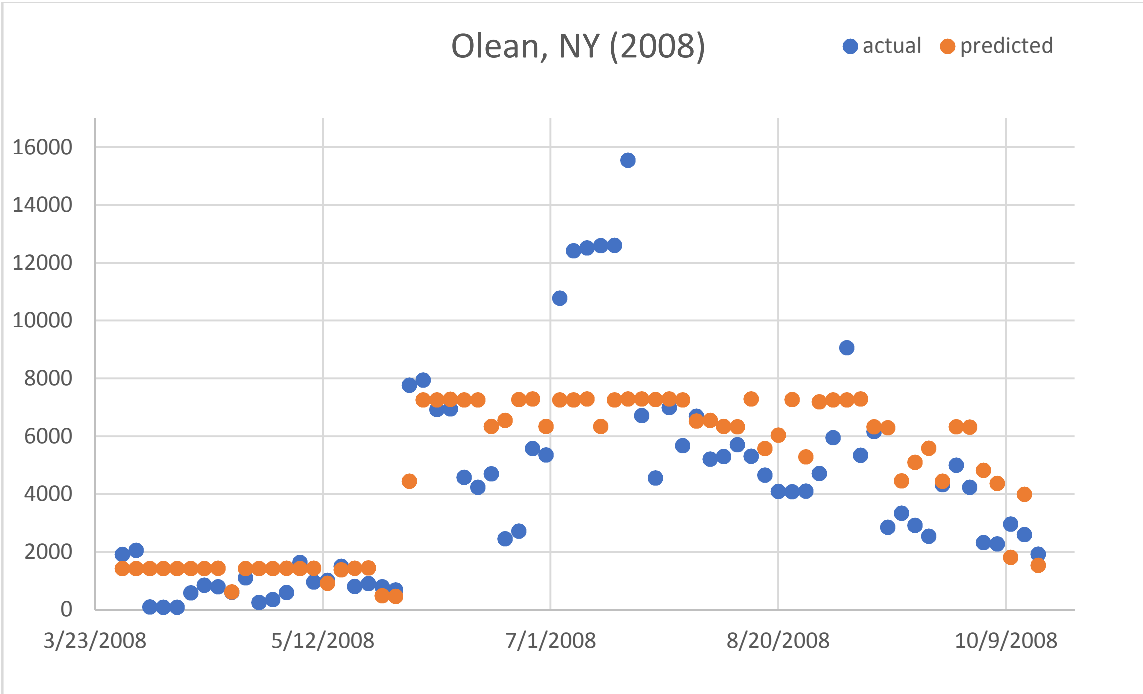
Appendix 2: Table of the All Long-Term Results

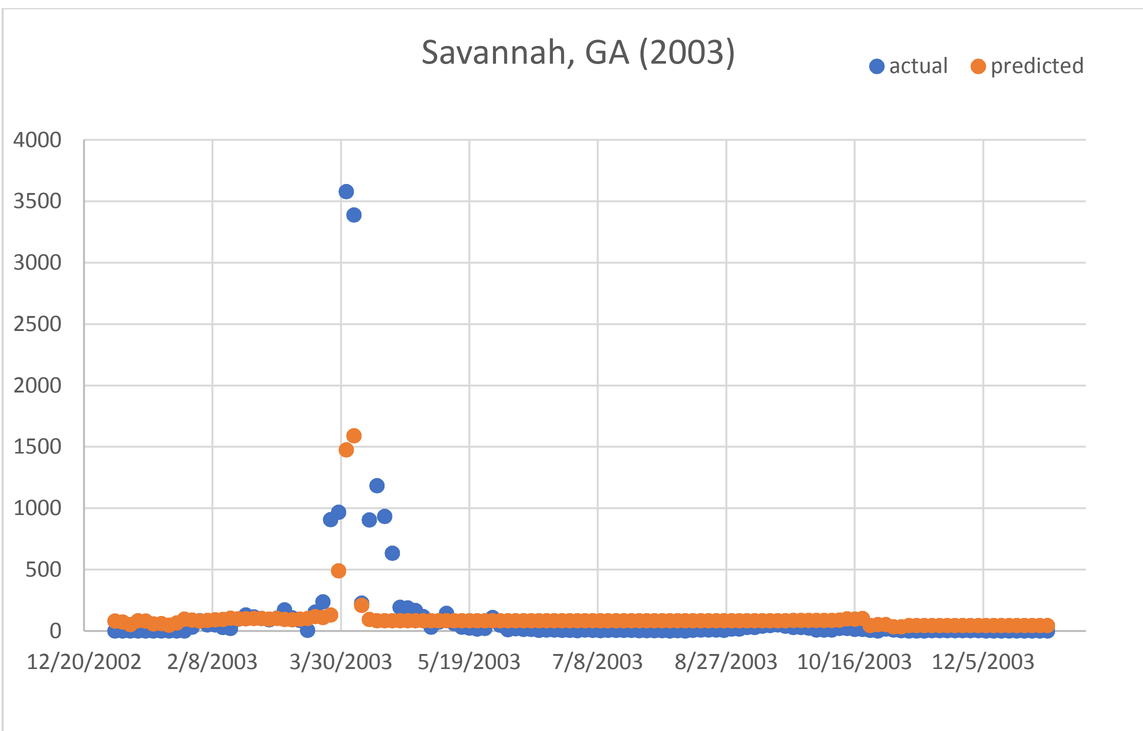
City	**State**	**R^2**	**MSE**	**RMSE**
New Castle	Delaware	0.6242	30116.36	173.54
New York City	New York	0.8319	109865.5	331.46
Olean	New York	0.4845	6898466	2626.49
Savannah	Georgia	0.8589	104001.2	322.49
Washington	District of Columbia	0.8093	13744.74	117.24

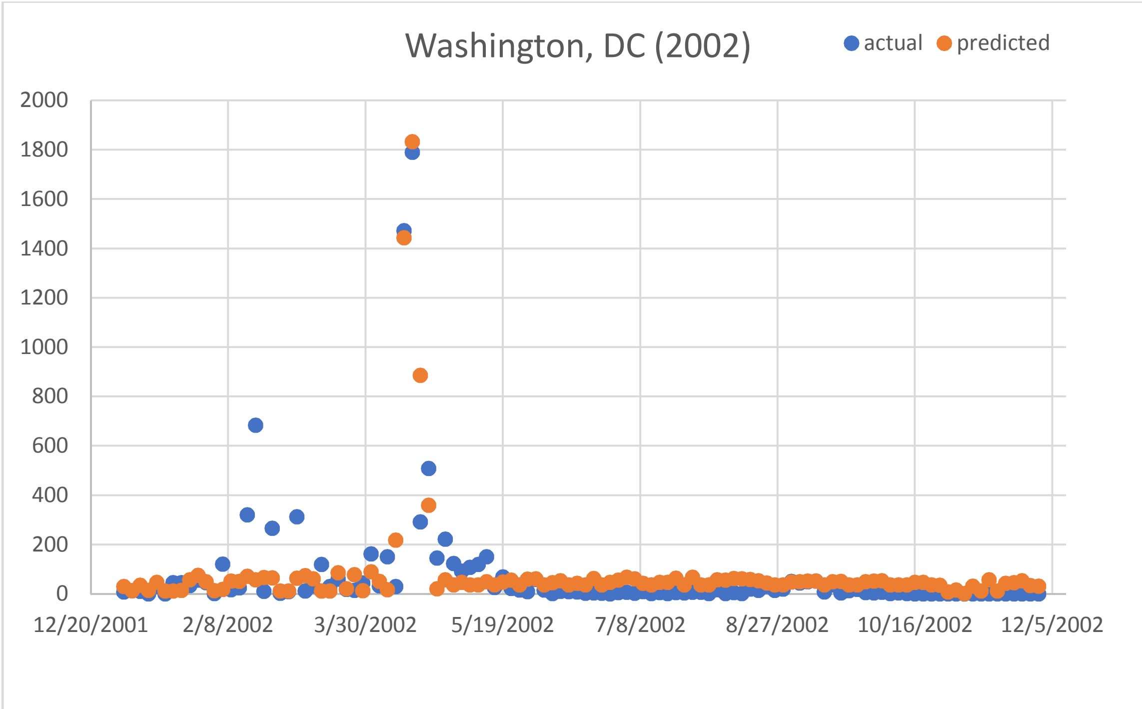
Appendix 3: Distributions of Long-Term Results in time






Olean, NY (2008)
actual
predicted
16000
14000
12000
10000
8000
6000
4000
2000
0
3/23/2008
5/12/2008
7/1/2008
8/20/2008
10/9/2008


Savannah, GA (2003)
actual
predicted
4000
3500
3000
2500
2000
1500
1000
500
0
12/20/2002
2/8/2003
3/30/2003
5/19/2003
7/8/2003
8/27/2003
10/16/2003
12/5/2003


Washington, DC (2002)
actual
predicted
2000
1800
1600
1400
1200
1000
800
600
400
200
0
12/20/2001
2/8/2002
3/30/2002
5/19/2002
7/8/2002
8/27/2002
10/16/2002
12/5/2002

REFERENCES

Åberg, N. 1989. Asthma and allergic rhinitis in Swedish conscripts. *Clinical and Experimental Allergy* 19(1):59-63.

Adebayo, Y. R. 1991. Day-time effects of urbanization on relative humidity and vapour pressure in a tropical city. *Theoretical and applied climatology* 43(1): 17-30.

Arca, B., G. Pellizzaro, A. Canu, and G. Vargius. 2002. Airborne pollen forecasting: Evaluation of ARIMA and neural network models. *In 15th conference on biometeorology and aerobiology joint with the 16th International Congress on Biometeorology* 194-196.

Arca, B, Grazia Pellizzaro, Annalisa Canu, and Arnoldo Vargiu. 2004. Use of neural networks to short-term forecast of airborne pollen data. *Institute of Biometeorology* 4 pages

Astray, G., M. Fernández-González, F. J. Rodríguez-Rajo, D. López, and J. C. Mejuto. 2016. Airborne castanea pollen forecasting model for ecological and allergological implementation. *Science of the Total Environment* 548:110-121.

Arizmendi, C. M., Sanchez J. R., Ramos N. E., and Ramos G. I. 1993. Time series predictions with neural nets: application to airborne pollen forecasting. *International journal of biometeorology* 37(3):139-144.

Baratti, R., Cannas B., Fanni A., Pintus M., Sechi G.M., and Toreno N. 2003. River flow forecast for reservoir management through neural networks. *Neurocomputing* 55:421–437.

Barnes, Charles, Freddy Pacheco, Julie Landuyt, Frank Hu, and Jay Portnoy. 2001. The effect of temperature, relative humidity and rainfall on airborne ragweed pollen concentrations. *Aerobiologia* 17(1):61-68.

Bartková - Ščevková, J. 2003. The influence of temperature, relative humidity and rainfall on the occurrence of pollen allergens (Betula, Poaceae, Ambrosia artemisiifolia) in the atmosphere of Bratislava (Slovakia). *International Journal of Biometeorology* 48(1): 1-5.

Beggs, Paul J., Constance H. Katelaris, Danielle Medek, Fay H. Johnston, Pamela K. Burton, Bradley Campbell, Alison K. Jaggard, Don Vicendese, David M. J. S. Bowman, Ian Godwin, Alfredo R. Huete, Bircan Erbas, Brett J. Green, Rewi M. Newnham, Ed Newbigin, Simon G. Haberle, and Janet M. Davies. 2015. Differences in grass pollen allergen exposure across Australia. *Australian and New Zealand Journal of Public Health* 39(1): 51-55.

Bianchi, M.M., Arizmendi C.M., and Sanchez J.R. 1992. Detection of chaos: New approach to atmospheric pollen time-series analysis. *International journal of biometeorology* 36(3):172-175.

Bolton, Douglas K. and Mark A. Friedl. 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology* 173:74-84

Bogawski, P., Grewling, Ł. and Frątczak, A. 2016. Flowering phenology and potential pollen emission of three Artemisia species in relation to airborne pollen data in Poznań (western Poland). *Aerobiologia 32*(2):265-276.

Box, G. E. P. and Jenkins G.M., 1976. Time series analysis. Forecasting and control. *In Holden-Day Series in Time Series Analysis, Revised ed., San Francisco: Holden-Day* 1-729.

Bråbäck, L., Breborowicz A., Dreborg S., Knutsson A., Pieklik H., and Björksten B. 1994. Atopic sensitization and respiratory symptoms among Polish and Swedish school children. *Clinical and Experimental Allergy* 24(9):826-835.

Braun-Fahrländer, C.H., Gassner M., Grize L., Neu U., Sennhauser F.H., Varonier H.S., Vuille J.C. and Wüthrich B. 1999. Prevalence of hay fever and allergic sensitization in farmer's children and their peers living in the same rural community. *Clinical and Experimental allergy* 29(1):28-34.

Burr, M. L., J. C. Emberlin, R. Treu, S. Cheng, N. E. Pearce and the ISAAC Phase One Study Group. 2003. Pollen counts in relation to the prevalence of allergic rhinoconjunctivitis, asthma, and atopic eczema in the International Study of Asthma and Allergies in Childhood (ISAAC). Clinical and Experimental Allergy. 33:1675–1680.

Castellano-Méndez, M., M. J. Aira, I. Iglesias, V. Jato, and W. González-Manteiga. 2005. Artificial neural networks as a useful tool to predict the risk level of Betula pollen in the air. *International Journal of Biometeorology* 49(5):310-316

Chakraborty, Kanad, Kishan Mehrotra, Chilukuri K. Mohan, and Sanjay Ranka. 1992. Forecasting the behavior of multivariate time series using neural networks. *Neural networks* 5(6):961-970.

Cecchi L. 2012. Allergenic pollen: a review of the production, release, distribution and health impacts. Introduction of Book of Allergenic Pollen. *Springer Science & Business Media* 1–8.

Chen, Jin, Per Jönsson, Masayuki Tamura, Zhihui Gu, Bunkei Matsushita, and Lars Eklundh. 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote sensing of Environment* 91(3): 332-344.

Cookson, W. 1999. The alliance of genes and environment in asthma and allergy. *Nature* 402:5-11.

D'Amato J., L. Cecchi, S. Bonini, C. Nunes, I. Annesi-Maesano, H. Behrendt, G. Liccardi, T. Popov, and P. van Cauwenberge. 2007. Allergenic pollen and pollen allergy in Europe. *Allergy* 1-15.

Darrow, L.A., Hess, J., Rogers, C.A., Tolbert, P.E., Klein, M. and Sarnat, S.E. 2012. Ambient pollen concentrations and emergency department visits for asthma and wheeze. *Journal of Allergy and Clinical Immunology* 130(3):630-638.

Davis, D. J. 1972. NIAID initiatives in allergy research. *Journal of Allergy and Clinical Immunology* 49(6):323-328.

Emberlin, J., J. Mullins, J. Corden, S. Jones, W. Millington, M. Brooke, and M. Savage. 1999. Regional variations in grass pollen seasons in the UK, long-term trends and forecast models. *Clinical and Experimental Allergy* 29:347-356.

Emberlin, J., Siegfried Jaeger, Eugenio Dominguez-Vilches, Carmen Galan Soldevilla, Lene Hodal. Paolo Mandrioli, Auli Rantio Lehtimäki, Mike Savage, Frits Th. Spieksma, and Catherine Bartlett. 2000. Temporal and geographical variations in grass pollen seasons in areas of Western Europe: an analysis of season dates at sites of the European pollen information system. *Aerobiologia* 16:373-379.

García-Mozo, H., Chuine, I., Aira, M.J., Belmonte, J., Bermejo, D., de la Guardia, C.D., Elvira, B., Gutiérrez, M., Rodríguez-Rajo, J., Ruiz, L. and Trigo, M.M. 2008. Regional phenological models for forecasting the start and peak of the Quercus pollen season in Spain. *agricultural and forest meteorology*, *148*(3):372-380.

Gergen, P.J., Turkeltaub P.C., and Kovar M.G. 1987. The prevalence of allergic skin test reactivity to eight common aeroallergens in the US population: results from the second National Health and Nutrition Examination Survey. *Journal of Allergy and Clinical Immunology* 80(5):669-679.

Green, Brett James, Mary Dettmann, Eija Yli-Panula, Shannon Rutherford, and Rod Simpson. 2004. Atmospheric Poaceae pollen frequencies and associations with meteorological parameters in Brisbane, Australia: a 5-year record, 1994–1999. *International Journal of Biometeorology* 48(4):172-178.

Goldberg, C., Buch H., Moseholm L., and Weeke E. R. 1988. Airborne pollen records in Denmark, 1977–1986. *Grana* 27(3):209-217

Gottardini, Elena, and Fabiana Cristofolini. 1997. Spring airborne pollen data in two sites in Trentino (Northern Italy): A comparison with meteorological data. *Aerobiologia* 13(3): 199-204.

Goward, S. N., Markham B., Dye D. G., Dulaney W., and Yang, A. J. 1991. Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer. *Remote Sensing of Environment* 35:257–277.

Haberle, Simon G., David M. J. S. Bowman, Rewi M. Newnham, Fay H. Johnston, Paul J. Beggs, Jeroen Buters, Bradley Campbell, Bircan Erbas, Ian Godwin, Brett J. Green, Alfredo Huete, Alison K. Jaggard, Danielle Medek, Frank Murray, Ed Newbigin, Michel Thibaudon, Don Vicendese, Grant J. Williamson, and Janet M. Davies. 2014. The macroecology of airborne pollen in Australian and New Zealand urban areas. *Plos One* 9(5):1-11

Heiat, A. 2002. Comparison of artificial neural network and regression models for estimating software development effort. *Information and Software Technology* 44:911–922.

Horton, R., Gary Y., William E., Robert K., Matthias R., Edna S., Adam W., and David W. 2014. Climate Change Impacts in the United States: Northeast. U.S. Global Change Research Program. 371-385.

Houborg, R., Soegaard, H. and Boegh, E., 2007. Combining vegetation index and model inversion methods for the extraction of key vegetation biophysical parameters using Terra and Aqua MODIS reflectance data. *Remote Sensing of Environment* 106(1):39-58.

Huang, Zachary Y. and Tugrul Giray. 2012. Factors Affecting Pollinators and Pollination. *Psyche* 2012 Article ID 302409:3 pages.

Huete, Alfredo R., Kamel Didan, Yosio E. Shimabukuro, Piyachat Ratana, Scott R. Saleska, Lucy R. Hutyra, Wenze Yang, Ramakrishna R. Nemani, and Ranga Myneni. 2006. Amazon rainforests green□up with sunlight in dry season. *Geophysical research letters* 33(6):1-4.

Ince, Ragib. 2004. Prediction of fracture parameters of concrete by Artificial Neural Networks. *Engineering Fracture Mechanics* 71(15):2143–2159.

Ito, K., Weinberger, K.R., Robinson, G.S., Sheffield, P.E., Lall, R., Mathes, R., Ross, Z., Kinney, P.L. and Matte, T.D. 2015. The associations between daily spring pollen counts, over-the-counter allergy medication sales, and asthma syndrome emergency department visits in New York City, 2002-2012. *Environmental Health* 14(1):71.

Jalbert, I. and Golebiowski B. 2015. Environmental aeroallergens and allergic rhino-conjunctivitis. *Current Opinion in Allergy and Clinical Immunology* 15:476-481.

Jariwala, S., Toh, J., Shum, M., de Vos, G., Zou, K., Sindher, S., Patel, P., Geevarghese, A., Tavdy, A., & Rosenstreich, D. 2014. The association between asthma-related emergency department visits and pollen and mold spore concentrations in the Bronx, 2001-2008. *Journal of Asthma* 51:79-83.

Jiang, Zhangyan, Alfredo R. Huete, Kamel Didan, and Tomoaki Miura. 2008. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment* 112(10):3833-3845.

Jin, X. M., Y. –K. Zhang, M. E. Schaepman, J. G. P. W. Clevers, and Z. Su. 2008. Impact of elevation and aspect on the spatial distribution of vegetation in the Qilian mountain area with remote sensing data. The International Archives of Photogrammetry. *Remote Sensing and Spatial Information Sciences* 37(B7):1385-1390

Jonsson P. 2004. Vegetation as an urban climate control in the subtropical city of Gaborone, Botswana. *Intern J Climat* 24:1307–1322

Karlsen, S.R., Ramfjord, H., Høgda, K.A., Johansen, B., Danks, F.S. and Brobakk, T.E. 2009. A satellite-based map of onset of birch (Betula) flowering in Norway. *Aerobiologia* 25(1):15.

Kerr, Michael. 2016. Pollen Allergies. *Health Line*. Retrieved from http://www.healthline.com/health/allergies/pollen#Overview1.

Kilpelainen, M., Terho E.O., Helenius H., and Koskenvuo M. 2000. Farm environment in childhood prevents the development of allergies. *Clinical and experimental allergy* 30(2):201-208.

Kizilpinar, I., Civelek E., Tuncer A., Dogan C., Karabulut E., Sahiner M. U., Yavuz S. T., and Sackesen C. 2010. Pollen counts and their relationship to meteorological factors in Ankara, Turkey during 2005–2008. *International Journal of Biometeorology* 55:623-631.

Knapp, Alan K., Philip A. Fay, John M. Blair, Scott L. Collins, Melinda D. Smith, Jonathan D. Carlisle, Christopher W. Harper, Brett T. Danner, Michelle S. Lett, and James K. Mccarron. 2002. Rainfall Variability, Carbon Cycling, and Plant Species Diversity in a Mesic Grassland. *Science* 2202-2205

Korpella, Robert. 2017. Does Rain Raise or Lower the Pollen Count. *Sciencing*. Retrieved from http://sciencing.com/rain-raise-lower-pollen-count-23009.html.

Kosisky, S. E., Marks, M. S., & Nelson, M. R. 2010. Pollen aeroallergens in the Washington, DC, metropolitan area: a 10-year volumetric survey (1998-2007). Annals of Allergy, Asthma & Immunology, 104(3): 223-235.

Kujawa, Ellen Ruth, Simon Goring, Andria Dawson, Randy Calcote, Eric C. Grimm, Sara C. Hotchkiss, Stephen T. Jackson et al. 2016. The effects of anthropogenic land cover change on pollen-vegetation relationships in the American Midwest. *Anthropocene* 15:60-71.

Lapedes. A. and Farber. R. 1987. Nonlinear signal processing using neural networks: Prediction and system modeling. *Technical Report LA-UR-87-2662. Los Alamos National Laboratory, Los Alamos, NM* 1-52.

Lipfert, F. W., S. Cohen, L. R. Dupuis, and J. Peters. 1991. Relative humidity predictor equations based on environmental factors. Atmospheric Environment. Part B. *Urban Atmosphere* 25(3): 435-441.

Liu, Xun, Daji Wu, Gebreab K. Zewdie, Lakitha Wijerante, Christopher I. Timms, Alexander Riley, Estelle Levetin, and David J. Lary. 2017.Using machine learning to estimate atmospheric Ambrosia pollen concentrations in Tulsa, OK. *Environmental Health Insights* 11:1178630217699399.

Luvall, J., Levetin, E., Bunderson, L., VanDeWater, P., Sprigg, B., Prasad, A., Nickovic, S., Vukovic, A., Pejanovic, G., Crimminis, T. and Myers, O., 2013. Integration of Airborne Dust Prediction Systems and Vegetation Phenology to Track Pollen for Asthma Alerts in Public Health Decision Support Systems. Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20130013979.pdf.

Meltzer, E.O., Blaiss M.S., Derebery M.J., Mahr T.A., Gordon B.R., Sheth K.K., Simmons A.L., Wingertzahn M.A., and Boyle J.M. 2009. Burden of allergic rhinitis: results from the Pediatric Allergies in America survey. *Journal of Allergy and Clinical Immunology* 124(3):43-70.

Mesa, Juan Antonio Sánchez, Carmen Galán, and César Hervás. 2005. The use of discriminant analysis and neural networks to forecast the severity of the Poaceae pollen season in a region with a typical Mediterranean climate. *International Journal of Biometeorology* 49(6):355-362.

Michael Grogan. 2017. Neural Network Modelling Using R. Retrieved from http://www.michaeljgrogan.com/neural-network-modelling-using-r/.

Mondal, Pinki. 2011. Quantifying surface gradients with a 2-band Enhanced Vegetation Index (EVI2). *Ecological Indicators* 11:918-924.

Mudd, Samuel J. and H. S. Arathi. 2012. Image analysis protocol for detecting and counting viable and inviable pollen grains. *Journal of Plant Studies* 1(2):158-167.

NCSU 2012. Southeast Precipitation. Climate Education for K-12. Retreived from http://climate.ncsu.edu/edu/k12/.SEPrecip.

Neil, Kaesha, and Jianguo Wu. 2006. Effects of urbanization on plant flowering phenology: A review. *Urban Ecosystems* 9(3): 243-257.

Peñuelas, Josep, Iolanda Filella, Xiaoyang Zhang, Laura Llorens, Romà Ogaya, Francisco Lloret, Pere Comas, Marc Estiarte, and Jaume Terradas. 2004. Complex spatiotemporal phenological shifts as a response to rainfall changes. *New Phytologist* 161(3): 837-846.

Peng, D., Jiang, Z., Huete, A.R., Ponce-Campos, G.E., Nguyen, U. and Luvall, J.C. 2013. Response of spectral reflectances and vegetation indices on varying juniper cone densities. *Remote Sensing*, *5*(10):5330-5345.

Pérez, Claudio F., Maria I. Gassmann, and Mauro Covi. 2009. An evaluation of the airborne pollen–precipitation relationship with the superposed epoch method. *Aerobiologia* 25(4): 313-320.

Peternel, Renata, Lidija Srnec, Josip Čulig, Ksenija Zaninović, Božena Mitić, and Ivan Vukušić. 2004. Atmospheric pollen season in Zagreb (Croatia) and its relationship with temperature and precipitation. *International journal of biometeorology* 48(4): 186-191.

Pinty, B., and M. M. Verstraete.1992. GEMI: a non-linear index to monitor global vegetation from satellites. *Plant ecology* 101(1):15-20.

Puc, Małgorzata. 2012. Artificial neural network model of the relationship between Betula pollen and meteorological factors in Szczecin (Poland). *International journal of biometeorology* 56(2):395-401.

Ranzi, A., Paolo Lauriola, Vittorio Marletto, and Franco Zinoni. 2003. Forecasting airborne pollen concentrations: Development of local models. *Aerobiologia* 19:39-45.

Riedler, J., Eder W., Oberfeld G., and Schreuer M. 2000. Austrian children living on a farm have less hay fever, asthma and allergic sensitization. *Clinical and Experimental Allergy* 30(2):194-200.

Rocha, A. V. and Shaver G. R. 2009. Advantages of a two band EVI calculated from solar and photosynthetically active radiation fluxes. *Agricultural and Forest Meteorology* 149:1560–1563.

Rojo, J., Rapp A., Lara B., Sabariego S., González F. F., and Badia. R. P. 2016. Characterization of the airborne pollen spectrum in Guadalajara (central Spain) and estimation of the potential allergy risk. *Environment Monitoring Assessment* 188(130):1-13.

Roujean, J. L., Leroy M., Deschamps P. Y., and Podaire A. 1992. Evidence of surface reflectance bidirectional effects from a NOAA/AVHRR multitemporal data set. *International Journal of Remote Sensing* 13:685–698.

Sánchez□Mesa, J. A., C. Galán, J. A. Martínez□Heras, and C. Hervás□Martínez. 2002. The use of a neural network to forecast daily grass pollen concentration in a Mediterranean region: the southern part of the Iberian Peninsula. *Clinical & Experimental Allergy* 32(11):1606-1612.

Sharda, R., 1994. Neural networks for the MS/OR analyst: An application bibliography. *Interfaces* 24(2):116–130.

Silverberg, J.I., Braunstein, M. and Wong, M., L. 2015. Association between climate factors, pollen counts, and childhood hay fever prevalence in the United States. *Journal of Allergy and Clinical Immunology* 135:463-469.

Skjøth, C. A., P. V. Ørby, , T. Becker, C. Geels, V. Schlünssen,T. Sigsgaard, J. H. Bønløkke, J. Sommer, P. Søgaard, and O. Hertel. 2013. Identifying urban sources as cause to elevated grass pollen concentrations using GIS and remote sensing. *Biogeosciences* 10:541-554.

Smith, J.T., 2005. Neural Network Verification. In B.J. Taylor (Ed.), Methods and Procedures for the Verification and Validation of Artificial Neural Networks. Springer.

Sozen, A., Arcaklioglu E., Özalp M., and Çaglar N. 2005. Forecasting based on neural network approach of solar potential in Turkey. *Renew. Energy* 30:1075–1090.

Streiff, R., Ducousso, A., Lexer, C., Steinkellner, H., Gloessl, J. and Kremer, A., 1999. Pollen dispersal inferred from paternity analysis in a mixed oak stand of Quercus robur L. and Q. petraea(Matt.) Liebl. *Molecular Ecology* 8(5):831-841.

Takai, T., and Izuhara K. 2015. Allergens in modern society: Updated catalogs and future prospects. Allergology International 64(4):293-294.

Taketomi, E. A., M. C. Sopelete, P. F. de Souza Moreira, F. A. M. Vieira. 2006. Pollen allergic diseases: pollens and its major allergens. *Brasilian Journal of Otorrinolaringology* 72(4):562-567.

Tan, A. H. and Pan H. 2005. Predictive neural networks for gene expression data analysis. *Neural Network* 18:297–306.

Tan, Bin, Jeffrey T. Morisette, Robert E. Wolfe, Feng Gao, Gregory A. Ederer, Joanne Nightingale, and Jeffrey A. Pedelty. 2011. An Enhanced TIMESAT Algorithm for Estimating Vegetation Phenology Metrics from MODIS Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4(2):361-371.

Taylor, J. A., A. D. Mowat, A. F. Bollen, and B. M. Whelan. Early season detection and mapping of Pseudomonas syringae pv. actinidae infected kiwifruit (Actinidia sp.) orchards. 2014. *New Zealand Journal of Crop and Horticultural Science* 42(4):303-311.

Tomassetti, B., Lombardi A., Cerasani E., Di Sabatino A., Pace L., Ammazzalorso D., Verdecchia M. 2013. Mapping of alternaria and pleospora concentrations in central Italy using meteorological forecast and neural network estimator. *Aerobiologia* 29(1):55–70.

Tosun N. and Ozler L. 2002 A study of tool life in hot machining using artificial neural networks and regression analysis method. J. Mater. Process. Technol. 124:99–104.

Turner, R., Somayajulu S., Ehud R., and Ian P. D. 2006. Generating Spatio-Temporal Descriptions in Pollen Forecasts. *EACL '06 Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations* 163-166.

UCB Institute of Allergy. 1999. European Allergy White Paper Update. 1-238.

USDA 2017. Eastern Region. Forest Service. Retrieved from https://www.fs.fed.us/wildflowers/regions/eastern/.

Valencia-Barrera, R. M., P Comtois, and D. Fernandez-Gonzalez. 2001. Biogeography and bioclimatology in pollen forecasting. *Grana* 40(4-5):223-229.

van Vliet A. J. H., Aart Vliet, Rudolf S. De Groot, Adrie F. G. Jacobs, and Frits T. M. Spieksma. 2002. The influence of temperature and climate change on the timing of pollen release in the Netherlands. *International Journal of Climatolology* 22:1757–1767.

von Ehrenstein, O.S., Von Mutius E., Illi S., Baumann L., Bohm O. and Von Kries R. 2000. Reduced risk of hay fever and asthma among children of farmers. *Clinical and experimental allergy* 30(2):187-193.

Xiao, Xiangming, Qingyuan Zhang, Scott Saleska, Lucy Hutyra, Plinio De Camargo, Steven Wofsy, Stephen Frolking, Stephen Boles, Michael Keller, and Berrien Moore. 2005. Satellite-based modeling of gross primary production in a seasonally moist tropical evergreen forest. *Remote Sensing of Environment* 94(1):105-122.

Zain, Azlan Mohd, Habibollah Haron, Sultan Noman Qasem, and Safian Sharif. 2012. Regression and ANN models for estimating minimum value of machining performance. Applied Mathematical Modelling 36(4):1477-1492.

Zhang Guoqiang, B. Eddy Patuwo, and Michael Y. Hu. 1998. Forecasting with artificial neural network: The state of the art. International Journal of Forecasting 14:35-62

Zhang, G.P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50:159-175.

Zhang, Xiaoyang, Mark A. Friedl, Crystal B. Schaaf, and Alan H. Strahler. 2004. Climate controls on vegetation phenological patterns in northern mid□and high latitudes inferred from MODIS data. *Global Change Biology* 10(7):1133-1145.

Zhang, X., Tan B., and Yu Y. 2014. Interannual variations and trends in global land surface phenology derived from enhanced vegetation index during 1982–2010. *International Journal of Biometeorology* 58:547–564.

Zhang, Xiaoyang. 2015. Reconstruction of a complete global time series of daily vegetation index trajectory from long-term AVHRR data. *Remote Sensing of Environment* 156:457-472.

Ziska, Lewis H., Paul R. E., and Christine A. R. 2007. Climate change, aerobiology, and public health in the Northeast United States. *Mitig. Adapt. Strat. Glob. Change* 13:607-613.