2017

An analysis of long-term effects of climate change and socioeconomic activities on grassland productivity of inner Mongolia

Siyu Fan

An Analysis of Long-term Effects of Climate Change and Socioeconomic Activities on Grassland Productivity of Inner Mongolia

by

Siyu Fan

Dissertation

Submitted to the College of Technology

Eastern Michigan University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Technology

Concentration in Geographic Information System

Dissertation Committee:

Yichun Xie, Ph.D., Chair

Benedict Ilozor, Ph.D.

Xiangdong Che, Ph.D.

Xining Yang, Ph.D.

November 6, 2017

Ypsilanti, Michigan

Dedication

I dedicate my dissertation to my lovely wife, Xiaoqing Ge, for her endless support to the family and warm encouragement while I was tied to finishing up this piece. This dissertation is also dedicated to my parents to express my gratitude for their support throughout my life. Finally, my dedication extends to my dissertation committee members for their time and patience devoted to helping me improve my work. Especially, my appreciation goes to my academic advisor, Dr. Yichun Xie. I have learned so much under his supervision and would never have accomplished this much without his kind guidance.

Abstract

In recent years, researchers have recognized the complexity of the interactions between the ecological system and the economic development of human society. However, the complicated relationships overwhelm traditional statistical procedures and require an innovative approach to investigate their dynamics. We proposed this study to provide a unique perspective in analyzing the long-term causal relationships between the grassland productivity, climate change, and socioeconomic development of Inner Mongolia Autonomous Region (IMAR) of China. Our attempt began with acquiring remotely sensed satellite imagery, climatic variations, and aggregated annual reports of the socio-economy of the IMAR in vegetation growing seasons for 15 years. The spatial and temporal dissimilarities of the raw observations prevented us from exploiting the potential of this valuable dataset; thus, we interpolated and extrapolated the data to generate a panel dataset with consistent spatial and temporal resolutions. Then, we took another step to preprocess the panel data by applying a signal filter to isolate the long-term trend of change from the inter- and intra-annual cyclic patterns and used the trends as the input for a panel data model.

The results from our statistical analysis indicated that the independent variables explained the variations in the dependent variable extremely well, while the polynomial terms of climatic variables were significant with limited marginal effect and most of the climatic variables showed negative linear impact on the grassland productivity. In the meantime, we found not all socioeconomic variables we attempted to include into the model significantly affected grassland productivity, especially the variables describing the financial status of the IMAR residents.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Problem Statement

Concealed by annual cyclic patterns, the long-term coupled effects of climate changes and human activities on grassland growth in Inner Mongolia of China are not understood.

Background

As defined by NASA (2016), climate change "is a change in the typical or average weather of a region or city" (n.p.). Climate change could be a shift in a region's average annual rainfall, or it could be global, affecting the earth's average temperature. Climate change has been an ongoing process for centuries. One of the results of global climate change is variations in temperature. For example, in a simulation conducted by Shukla, Nobre, and Sellers (1990), the surface and soil temperature increased 1 °C to 3 °C in the Amazonian forest as a result of deforestation. Based on research by Lim, Cai, Kalnay, and Zhou (2005), the urban land type shows a positive impact on the surface warming process by 0.26 °C per decade. As summarized by Solomon (2007), many researchers have recorded that the trend of rising temperature has increased dramatically in recent years. The increasing trend is 0.063 °C per decade from 1850 to 2005 for the land surface temperature of the northern hemisphere (Brohan, Kennedy, Harris, Tett, & Jones, 2006). If measured from 1901, the trend becomes 0.089 °C per decade, whereas this increasing trend has exacerbated significantly since the late 20th century in which the average temperature increase is 0.328 °C per decade from 1979 to 2005. The accelerated global and regional climate changing process impacts the ecological and environmental settings of human civilization, such as the rise of sea level, variation in patterns of atmospheric circulation, change of global mean precipitation, changes of climate

zones of regions, deforestation of the Amazon rainforest, and the rise of carbon dioxide percentage in the atmosphere (Solomon, 2007).

The grasslands, usually located between the arid and humid climate zones, cover about one-quarter of the land surface on earth. Ecologically, grasslands play a critical role in carbon storage, nitrogen fixation, and water and soil conservation (Gill et al., 2002; Huss-Danell, Chaia, & Carlsson, 2007). Economically, grasslands are used to produce animal products, like milk and meat (Fu, Bo, Du, & Zheng, 2012). For the residents of the Inner Mongolia Autonomous Region (IMAR) of China, the grassland located in the IMAR has been their primary source of income and food for centuries. Traditionally, this area was used for nomadic grazing for sheep, goats, horses, and cattle. Since the late 20th century, the ecosystem of grasslands in China is facing critical challenges: degradation and desertification caused by the coupled effects from accelerated social and economic development and dramatic climate changes. However, our ability to understand the ecological consequences of global or regional climate change and socioeconomic development is limited due to the complexity of their interactions. For example, uncertainties are associated with predicting climate change (Marin, 2010) and interactions between species, communities, and humans (Hagerman, Dowlatabadi, Satterfield, & McDaniels, 2010).

China, the third largest country on earth, has a vast area of territory designated for pastoral production. The six major pastoral regions, occupying over three million square kilometers mainly of the steppe, stretch across northern China from Inner Mongolia on the east to the Tarim Basin in Xinjiang on the west. These pastoral grasslands are mainly distributed within the range of longitude 74°E to 119°E and 35°N to 50°N, affected by arid, semi-arid, and sub-humid climates from west to east (Squres, Hua, Zhang, & Li, 2010). Its

regional climate change has been noticed by different researchers. In Xie, Jia, Qin, Shen, and Chang's study (2016), they acquired both daily and monthly temperature and precipitation data between 1982 and 2011 from 73 ground stations across the Loess Plateau of China, using the China Meteorological Data Sharing Service System. Their analysis pointed out an uprising trend over the past three decades in temperature by 0.05 °C per year, while the precipitation for the same period decreased by 0.72 mm annually, which was less statistically significant than the temperature change. In a study conducted by Jiapaer, Liang, Yi, and Liu (2015), they used monthly cumulative precipitation and mean temperature data covering the Xinjiang province of China from 1982 to 2012. Their finding indicates that, though the magnitude of temperature change slightly varied around 0.1 °C/year across the studied region, the increasing trend was consistent. The variation of precipitation demonstrates significant spatial differences, ranging from 5.767 mm/year to -1.523 mm/year.

The IMAR, located in northern China, is strongly influenced by the Asian monsoon climate with increasing precipitation from the southwest to the northeast. It has one of the largest remaining natural grasslands in the world, covering an area of up to 791,000 square kilometers (S. Li et al., 2013). The majority coverage of the IMAR is grassland, which is primarily concentrated in the central part of the IMAR, while most of the forest is located in the northeastern part, dominated by broad-leaf and needle-leaf forests, and cropland in the southern and eastern regions (Li, Cui, Liu, Shi, & Qin, 2013). Traditionally, nomads owned the grasslands of the IMAR for grazing activities. From the late 20th century, the sustainability of grassland ecosystem has been challenged by severe degradation and desertification. The grassland degradation process has been emphasized and studied by many

domestic and international researchers in the past 50 years, and two leading causes of grassland degradation have been asserted: climate changes and unsustainable consumption.

Global climate change is considered one of the primary factors that has influenced the ecosystem of the grasslands (e.g., global warming, increased/decreased precipitation, and extreme weather events; Crabbe, 2008; Kyselý, Beguería, Beranová, Gaál, & López-Moreno, 2012; Piras, Mascaro, Deidda, & Vivoni, 2015; Naidu et al., 2011; Ribalaygua et al., 2013). Affected by global climate variations in the past decades, the Mongolia Plateau is getting warmer and drier. According to many researchers, the annual mean temperature of the Mongolia Plateau has been increasing (Wang, Brown, & Agrawal, 2013b; Zhang et al., 2013). Yatagai and Yasunari (1994) noticed the average temperature of Mongolia rose by 1.5°C to 2.5°C in the 1980s and 1990s, while the global average during the same time span increased only approximately one-third to a half of the level in Mongolia. The annual mean temperature and precipitation in 2009 increased about 2.1°C and decreased about 7.0%, respectively (Wang, Brown, & Agrawal, 2013a). Lu, Wilske, Ni, John, and Chen (2009) analyzed 50 years of climatic data from 51 meteorological stations in Inner Mongolia and observed an increasing temperature and a vapor pressure deficit. The vast spatial extent and biological diversity of the grassland make it delicate to global and regional climate changes and attract much attention from scholars investigating vulnerability and sensitivity of the ecosystem. S. Li et al. (2013) and Shiyomi et al. (2011) used regression analysis to attribute the grassland condition to meteorological factors, such as rainfall, temperature, and sunshine. Galvin, Thornton, Boone, and Sunderland (2004) documented that eastern African pastoralists have also been tracking climate variabilities, including the interannual and intra-annual droughts and floods. However, local pastoralists have not been able to successfully

implement their strategies partially due to the nature of the changing climate itself. This fact indicates that the relationship between climate changes and ecosystem has not been fully understood yet, and this topic still requires more attention.

Since the early 1980s, when the national rural reform started, the private property rights arrangement, also known as the Household Production Responsibility System (HPRS), was implemented in the IMAR, which improved livestock productivity significantly as the policy provided attractive incentives encouraging farmers to increase their production efficiency. As a result of rapid economic development and population growth, the demand for agricultural output increased significantly and caused higher stocking density and a shift in agricultural policy (Shiyomi et al., 2011). These changes triggered biological succession, which in turn caused partial degradation and desertification of the grassland (Zhao, 1994). According to the Xilingol Statistic Bureau (1997), the number of livestock in Xilingol (a subdivision of the IMAR) nearly doubled from 12.6 million in 1980 to 22.7 million in 1997. Reported by Qi (2001), the average grassland availability per sheep unit decreased from 1.42 ha in 1980 to 1.05 ha in 1990, and the degraded area represented 48.6% of overall Xilingol grassland. Various human-introduced causes of grassland degradation have been summarized, such as animal overgrazing, conversion to cropland, inappropriate grassland management, and collecting of wood for fuel and herbs for medicine (Akiyama & Kawamura, 2007; Squires, 2009; Squires et al., 2010; Wu, Zhang, Li, & Liang, 2015).

Besides consumption of the grasslands, another example of human intervention to the ecosystem is the construction of a transportation network. Literature reflects that road construction and transportation have had both negative and positive impact on the destruction of the IMAR grassland (Deng, Huang, Huang, Rozelle, & Gibson, 2011). The construction of

roads separates the initially integrated ecological system into smaller and more vulnerable individuals, and the road traffic becomes a significant obstacle for commuting species. In addition, major roads are likely to be attractive for economic development, such as suburban areas, which causes critical land cover type changes. On the contrary, some researchers have found that road transportation led to higher quality of grassland, as they believe the roads provide better access for implementation of grassland restoration projects.

While the biophysical and socioeconomic systems are seemingly independent of each other, evidence suggests that the interactions between these two systems are usually nonlinear across space and time, especially when observed at different scales (Liu et al., 2007). Facing the increasingly deteriorated sustainability around the world in recent decades, it becomes one of the major challenges for scientific research and for society to understand the complex relationships between human and nature. The interactions between human and nature are expressed in the form of feedbacks. The term *feedback* came from electric circuits dating back to the 1900s and was defined as "when a stimulus is fed back to its origin through one or a series of interactions" (Berryman, 1989, p. 231). Today, the stimulus may take many forms, such as biological, physical, or social interactions, and feedbacks can involve many different types of matter, energy, or information. In ecological studies, positive and negative are two major types of feedbacks. The positive and negative do not refer to "good" or "bad," but rather to the change of intensity. Positive feedbacks happen when the stimulus amplifies the effect or reinforces the changes, whereas negative feedbacks occur when the stimulus causes a dampening effect or reverses the change. While the feedbacks from nature affect socioeconomic development (Aaheim, Amundsen, Dokken & Wei, 2012), the eco-system has also been heavily influenced by human activities (Lafortezza et al., 2015).

One of the 10 big questions from "The Big Questions in Geography," which was proposed by a group of prominent geographers, was "How has the Earth been transformed by human action?" (Cutter, Golledge, & Graf, 2002). As pointed out by Moser (2010) in a summative report, a number of researchers have contributed to geographic questions surrounding climate change, which steered the direction of research to exploration of vulnerability, sustainability, and adaptation of human-nature interactions.

The coupling effects of climate change and economy are drawing researchers' attention on global and regional scales. Wang, Chen, and Dong's (2006) evaluated the key contributions to desertification and rehabilitation around the Otindag Desert in northern China based on data from the 1950s to 2000s. It was concluded that though previous studies might have emphasized the impact of human activities on grassland evolvement, attributions from environmental factors were equally important. Aaheim et al. (2012) applied a macroeconomic general equilibrium model to integrate the impacts of climate changes on the economy of different sub-regions of Europe. It was found that the economic market and business behaviors are adaptive to climate changes, and the adaptation varies depending on the magnitude of the climate change. The gross domestic product (GDP) is impacted positively when the temperature increases by 2 °C throughout Europe in general, with exceptions of a minor recession in some subregions. However, if the temperature were to increase by 4 °C, the GDP would be negatively affected for the entire region, with the most significant impacts on the southern parts where the annual decline may reach up to 0.7%. For developing countries heavily dependent on the agricultural industry, the effect imposed by climate changes tends to be more significant than on developed countries, and the impact could even be disastrous. As reported by Arndt, Asante, and Thurlow (2015), the economy of

Ghana is highly vulnerable to climate change because of its high "dependence on rain-fed agriculture, hydropower and unpaved rural roads" (p. 7214). The national welfare of Ghana is reduced as a result of climate change, and the poorer households and the northern savannah zone suffer the most. As a part of the Mongolia Plateau, Mongolia experienced the worst droughts and Dzuds (severe winter snowstorms) consecutively in 2000, 2001, and 2002. The extreme weather condition affected over 50% of the total territory and cost herders about 12 million livestock in those periods (Angerer, Han, Fujisaki, & Havstad, 2008; Wang et al., 2013b). By applying the Markov chain cellular automata model, Tong, Sun, Ranatunga, He, and Yang (2012) predicted the interactions between land use change, hydrological variation, climate change, and human activity of the Little Miami River watershed in 2050 using hypothetical scenarios. Their experiment indicated that the coupling effect from climate change and human-induced land use change would lead to an approximately 10 to 50% increase in the stream flow speed. Depending on the magnitude of variations, the land use change could either magnify the impact of climate change on the hydrological model or mitigate the increasing flow speed. The mixed effect of the coupled variables supported the claim that complexity resides among human-nature interactions.

Although many studies have attempted to examine human-nature interactions, progress in understanding the relationships has been lacking, largely due to the complexity of the coupled systems (Liu et al., 2007). In ecology studies, the complicated effects from human-nature interactions are often referred to as the concept of coupled natural and human system (CNH; Chen et al., 2015) or the framework of coupled human and natural systems (CHANS; Hull, Tuanmu, & Liu, 2015). Either conceptual framework focuses on the feedbacks from human and nature at different scales to better understand and predict the

relationships between environmental changes and socioeconomic development. Given the complexity of CNH, results from research integrating ecological and social science vary across space and time. Contributions to the current literature discussing the understandings and models are needed, especially considering the escalating rate and magnitude of environmental change challenging sustainable socioeconomic progress.

Significance

The importance of this study relies on its adaptation of newly developed research approaches from other fields of study and the integration of multi-domain data and methodologies into a spatial-temporal analysis of the coupled human-nature system. Our literature review identified gaps to be filled.

First of all, one of the shortcomings shared among many publications is the limited data quality. Here, data quality refers to the temporal and spatial resolutions of observations. As observed, the data used by Meng et al. (2011), Xie et al. (2016), Jiapaer et al. (2015), and Onema and Taigbenu (2009) shared the same problem in that the spatial and temporal resolutions were restricted. In some of the research mentioned above, vegetation indices (VIs) data with 8 km^2 spatial resolution, or even 0.5°, were assembled in a periodic manner, monthly or annually, and used to evaluate vegetation variations. While half a degree not only covers a vast area on the ground, the actual square mileage also varies among different latitudes, where the linear distance across 0.5° could range from almost 66 kilometers near the equator to zero when precisely placed at the poles. From the perspective of observation frequency, different approaches were used in previous studies to obtain a manageable number of measurements of precipitation and temperature, including the monthly average and even the annual mean. When choosing data acquisition dates, some studies (Piao, Fang,

Zhou, Ciais, & Zhu, 2006; Xie et al., 2016) preferred to use VI data of the entire year non-discriminatively. Influenced by the continental climate, the IMAR has hot summers and cold winters, and the majority of the IMAR is usually covered by a substantial amount of snow for months. The seasonal land cover variation means that the VI products collected during the winter season are merely representations of snow and ice instead of the phenomenon of vegetation. While Prentice et al. (1992) determined that the temperature threshold of plants' growing season starts from 5 °C, Piao et al. (2006) observed the growth temperature for temperate grasslands of China is 9.7 °C. Admittedly, the coarser spatial/temporal resolution successfully manages the scope of the research to a controllable scale for an inferential analysis. However, it oversimplifies the data and loses granularity because the generalized data omits the variations of sampled observations within the spatial extent and during the period between data collection. Meanwhile, the undistinguished selection of the sampling period in vegetation index analysis apparently may introduce faulty observations of the dependent variable and leads to biased conclusions.

Secondly, climate change, economic development, and grassland growth together act as a complicated and integrated system, but current publications barely investigate the relationships between the trajectory of individual components. In Figure 1, we present a partial preview of the dataset to be used in this study. In this diagram, we demonstrate how the values of our observed variables change over time. More importantly, as a critical issue caused by the collinearity between climatic variables pointed out by Shi, Tao, and Zhang (2013), we argue that the conclusions drawn from previous studies could be flawed by the concurrency between observations of variables. Given how variables vary along time in Figure 1, it is clearly visible that while three of these variables-enhanced vegetation index

(EVI), precipitation (*PRE*), and temperature (*TEM*)-share a strong periodical and repetitive pattern of change, the other variable, gross domestic production (*GDP*), has a strong and increasing trend over time. As we know, for statistical analysis, researchers draw conclusions from sampled data to estimate what the population might be, and this process makes a judgment of the probabilities heavily based on the sample variations. According to our literature review, few published studies paid attention to the tendency of variables. Most of the analysis focused on regression analysis using raw data or original observations. Undoubtedly, the statistical procedures captured the dynamics between the dependent and independent variables and successfully explained causal relationships between phenomena. However, if we conduct a regression analysis using these variables, it is arguable that the common cyclic pattern among EVI, *PRE*, and *TEM*, or even more periodic patterns buried within the data, may alternate the relationships between these variables and further significantly sway the inferential conclusions drawn from the statistical analysis.

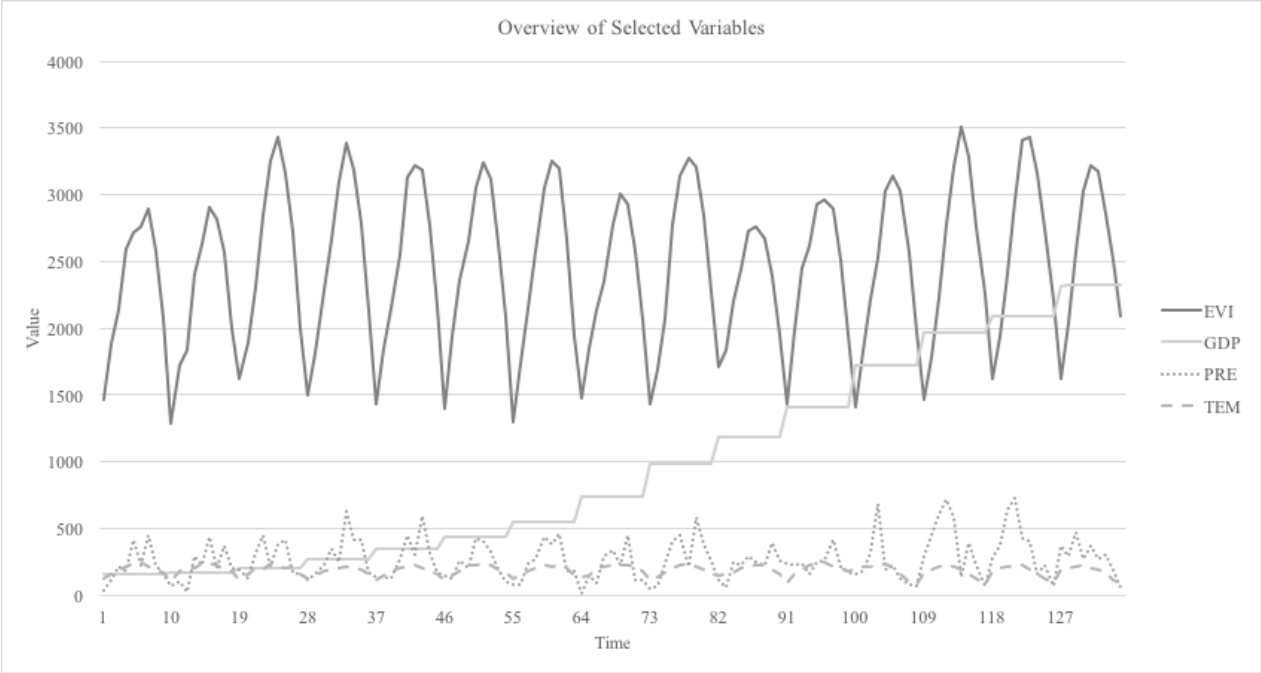

Figure 1. Time-series of selected variables.

Last but not least, we argue about the suitability of statistical models used to draw conclusions. Researchers have attempted to understand the relationships using various models, e.g., ordinary linear regression, linear least-square regression, and structural equation model. However, it remains a significant, common disadvantage of these models that they only allow either cross-sectional or time-series analysis at a time instead of simultaneously capturing the varying relationship in both cross-sectional and temporal scales. When it comes to using these models to analyze "big data" that consists of longitudinal observations of a set of individuals, time-series analysis only estimates the relationship of a single individual or treats all individuals as a whole and omits the unique properties of individuals, whereas cross-sectional analysis addresses the differences among individuals but ignores how individuals develop throughout time.

This study aims to address these issues by uncovering the long-term human-nature interactions through an advanced statistical model with additional data processing techniques on data of higher quality. The richness and quality of data were achieved by multi-dimensional and multivariate longitudinal observations. To capture the joint effect of climate change and human activities on grassland growth, this study acquired a dataset compiled from continuous observations of different instrument readings from meteorological stations, socioeconomic indicators from annual statistical yearbooks, and the remote sensed vegetation index over a 15-year period. The spatial distribution of these meteorological stations is illustrated in Figure 2. During this period, we employed the VI and climatic data with 250-meter spatial and 16-day temporal resolutions and limited our observations within April and September to avoid excessive data. Tens of thousands of observations were recorded for more than a dozen variables to subjectively describe the dynamic connections between events

of this sophisticated system. By collecting EVI and climatic data with higher spatial and temporal resolutions, we were able to use the data that depicted reality in more detail and supplemented additional within-individual and between-individual variations to support the inferential analysis.

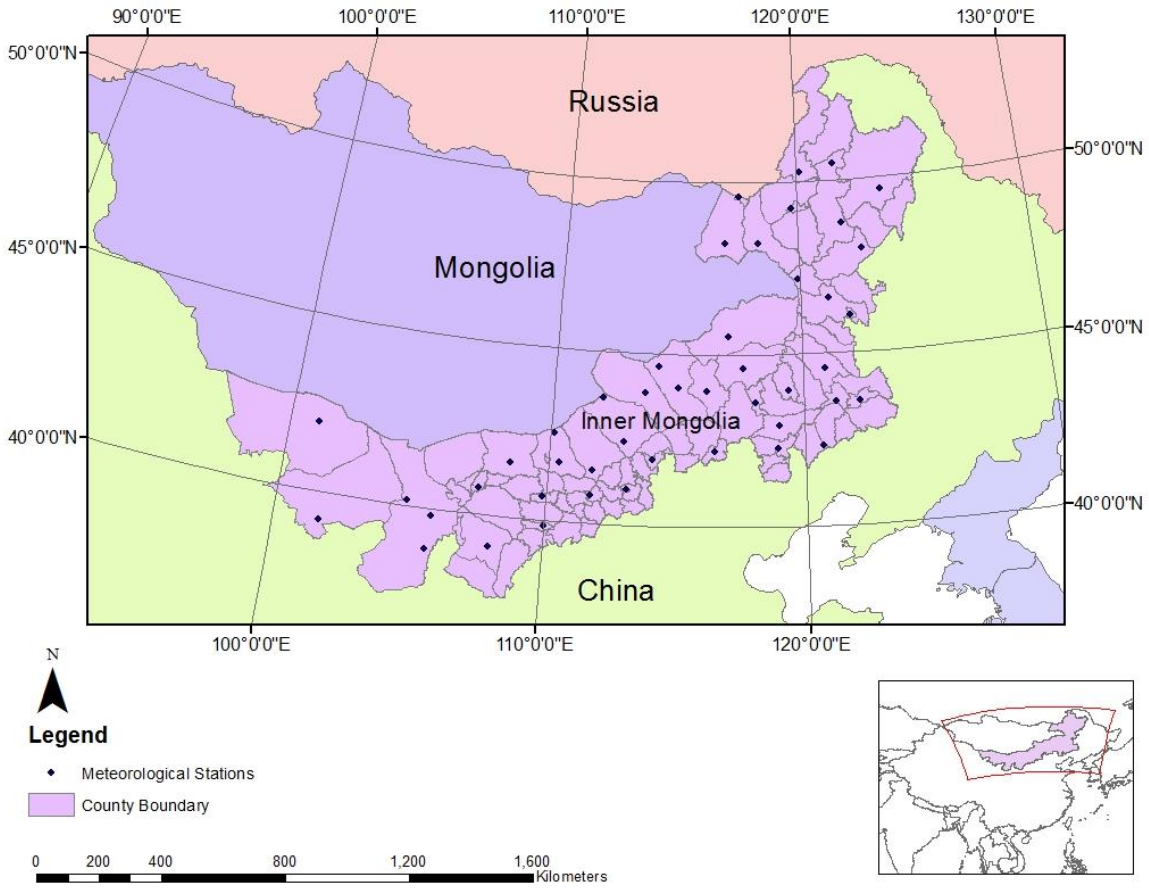

Figure 2. Mapping the study area

The uniqueness of this research also rests on the assimilation of fruitful data and advanced data analysis strategies. Based on empirical inspection of the data and the potential existence of periodical patterns, we suggest that the relationships concluded from using raw observations between factors of the coupled human-nature system should be reconsidered. By applying a filtering method, we introduce a new approach of analysis that focuses on the

tendency of variables to avoid false conclusions caused by the high collinearity among variables. Instead of selecting a traditional statistical model as discussed in previous paragraphs, we chose to advance our data analysis so that it depends on panel data modeling. Though the panel data model has been widely accepted in econometrics, its application in ecological and environmental studies has been limited. While either data filtering or panel data analysis has been widely applied in its own realm, few attempts have been made to incorporate both methodologies into a comprehensive tactic to address the undiscovered principles of the temporal-spatial phenomenon, particularly not the sophisticated coupled human-nature system. Our integration of data filtering and advanced statistical modeling provides an innovative perspective in analyzing the interactive dynamics between climate, human activities, and grassland growth.

Given the above reasoning, it is identified that there is a missing piece in the domain of geographic information system, and it is the determination of this study to contribute to the literature by investigating the long-term relationships between the trend of natural and social changes using a statistical spatial-temporal analysis.

Purpose and Objectives of the Study

This study aims to seek an explanation for the coupled effect of climate change and socioeconomic development on the IMAR grassland growth from an alternative perspective through analysis of the long-term trends.

More specifically, we itemized the objectives into the following list:

- Objective 1: Compile a comprehensive spatial-temporal dataset describing the dynamics of grassland productivity, climate change, and socioeconomic activity of the IMAR.

- Objective 2: Identify long-term trends of grassland productivity, climate change, and socioeconomic activity by eliminating cyclic patterns concealed in the time-series.
- Objective 3: Analyze the long-term causal relationships between grassland productivity, climate change, and socioeconomic activity of the IMAR using statistical modeling.

Hypotheses

- Hypothesis 1: There is no long-term trend of change concealed within the grassland growth time-series dataset.
- Hypothesis 2: There is no long-term trend of change concealed within the climate change time-series dataset.
- Hypothesis 3: There is no long-term trend of change concealed within the socioeconomic development time-series dataset.
- Hypothesis 4: The climate changes and human activities do not impose coupled effects on the IMAR grassland growth.

Limitations and Delimitations

Limitations. The time frame of this study is from 2000 to 2014 (for 15 years) and climatic data collected between April and September every year were used.

The geographic extent of this study is limited within the political boundary of the IMAR of China.

As documented by Piao et al. (2006) and recommended by subject professionals, the growing season of the IMAR grassland vegetation starts approximately from late April and ends by the end of September.

The EVI products of moderate resolution imaging spectroradiometer (MODIS), MOD13Q1, acquired between April and September were used as the dependent variable for the panel data analysis.

This study used seven climatic factors as independent variables: precipitation, barometric pressure, evaporation, temperature, sunshine duration, relative humidity, and wind speed.

In measuring the socioeconomic development of the IMAR, it is impractical to locate direct measurements quantitatively describing the phenomenon. In this study, we utilized a set of proxies from annual yearbooks as the socioeconomic variables, including the density of the arable area, the density of grain production, the density of livestock, the density of highway length, the share of farming income, the share of governmental revenue, the share of governmental investment, GDP per capita, and proportion of the rural population. These proxies were used as independent variables for the panel data analysis. The socioeconomic variables are not real measurements of the phenomenon of human activity during the period of observation, but statistically compiled proxies of annual performance. Though losing granularity, the panel data model balances missing observations and accounts for unobserved variations. Moreover, the compiled and normalized variables compensate for the disparities caused by differences in sizes of counties.

Given the variability of vegetation types, it is not feasible to create dummy variables for individual vegetation type of each county to examine the exact effect of environmental and economic impact on grassland ecosystem. Thus, we averaged the EVI of the entire county to represent grassland variation over time.

Delimitations. As the EVI data are systematically processed by well-established algorithms and then were downloaded from the U.S. Geological Survey (USGS)'s website, the quality is assured. However, due to the processing and composition procedure, there is less flexibility of available dates (every 16 days) and spatial resolution for users.

While climate is continuous around the globe, the measurements of meteorological variables are collected at discrete locations, sparsely distributed in the IMAR. It is unmanageable to collect meteorological variations in every location of the ground surface. Thus, the optimal method to represent the continuous surface for meteorological factors is to use spatial interpolation based on instrument readings from ground meteorological stations.

Assumptions

- The stationarity of normalized difference vegetation index (NDVI) concluded by Tucker et al. (2005) is correct and EVI time-series possesses the same property.
- The meteorological variables are both linear and polynomial correlated with EVI.
- The socioeconomic variables are linearly correlated with EVI.
- No human-introduced errors during the meteorological data collection process challenge the internal validity of this study.
- The EVI values at any given time point are representations of grassland growth in response to climate changes in the prior 16 days.

Definition of Terms

EVI. It is the abbreviation for enhanced vegetation index. It is an indicator derived from blue, red, and near-infrared bands and has higher sensitivity to variation in dense vegetation than NDVI (Huete et al., 2002).

IMF. This is the abbreviation for intrinsic mode function. The empirical mode decomposition (EMD) consists of a series of IMFs, and these IMFs "serve as the basis of the expansion which can be linear or nonlinear as dictated by the data" (Huang et al., 1998, p. 906). IMFs are solely based on and derived from the data, and they are "complete and almost orthogonal" (Huang et al., 1998, p. 906).

Cross-Sectional Data. Cross-sectional data are collected from a random sample of cases at the same point in time, and each observation is from a different case (Pickup, 2015).

Time-Series Data. Time-series data are observations collected from the same sample of cases at a series of time point. (Pickup, 2015).

Organization of Chapters 2-5

- Chapter 2 reviews the literature.
- Chapter 3 gives an overview of the methodology used for this study, which includes the design of the research, samples, population, data collection, and data analysis process.
- Chapter 4 presents the results from our analysis and extends to the discussion based on the results.
- Chapter 5 concludes this study by providing a summary of the findings and possible directions for the future.

Chapter 2. Review of Literature

Estimating Grassland Productivity

To investigate quantitatively the coupled effect of climate change and economy on the grassland degradation and restoration, aboveground biomass (AGB) is one of the major and widely adopted indicators of carbon sink on the ground surface (Anaya, Chuvieco, & Palacios-Orueta, 2009). In ecological studies, researchers usually use the net primary productivity (NPP) as a representation of AGB. NPP is defined as the accumulation of dry matter by green plants per unit time and space (Li, et al., 2013). NPP is a major component in the vegetation carbon cycle and a key indicator of ecosystem performance because it represents the ability of plants to fix atmospheric carbon as biomass (Li et al., 2014; Seaquist, Olsson, & Ardö, 2003). NPP data provide an approach to understanding ecosystem dynamics through factors, such as heterotrophic respiration through herbivory and decomposition, to determine the net bio-spherical exchange of carbon (Scurlock, Johnson, & Olson, 2002). As described by the Institute of Botany of Mongolia (2011), the traditional approach to measuring NPP is considered destructive because it measures the weight of dried matter, which is usually collected through in-field surveys. The Institute of Botany of Mongolia (2011) reported that average grassland biomass decreased from 804 kg/ha in 1961 to 369 kg/ha in 2010 according to a large-scale field ecological survey.

The conventional methodology for estimating AGB is applicable when the scale of study is manageable, for example, agricultural studies across several experimental farm fields. However, when being applied to research at larger spatial scales, it becomes time-consuming and uneconomical, especially for monitoring the grassland ecological systems (Ren, Zhou, & Zhang, 2011). Compared to in-situ vegetation sample collection, remote

sensing observation offers a cost-effective method and makes it possible to study spatial distribution and seasonal changes in vegetation from multi-temporal and multi-spectral perspectives (S. Li et al., 2013; Ren et al., 2011). As pointed out by Huete et al. (2002), one of the significantly distinct advantages of using satellite-based remote sensing systems for environmental research is the ability to understand the Earth as a system through large-scale observations. The Earth Observing System (EOS) program has been primarily used to study the role of terrestrial vegetation in the global process. In application, remote sensing vegetation biomass estimation relies heavily on using vegetation indices (VIs) as approximations of productivity. The VIs are one set of unit-less products that are derived from remotely sensed imagery. VIs are directly calculated using two or more bands of spectrums captured by remote sensors, without bias or assumption regarding land cover classes, soil types, or climatic conditions. The VI enhances the contribution of vegetation properties and enables consistent spatial-temporal comparison of terrestrial photosynthetic activities and canopy structural variation, thus allowing scientists to monitor seasonal, interannual, and long-term variations of plant structural, phenological, and biophysical parameters.

Between various remote sensing systems and products, VI products from the moderate resolution imaging spectroradiometer (MODIS) have been recognized as consistent measurements of global vegetation photosynthetic activities spatially and temporally. Two MODIS VIs, the normalized difference vegetation index (NDVI) and the EVI, provide global coverage up to 250-meter resolution for every 16-day period. NDVI is a normalized ratio of the near-infrared (NIR) and the red bands, calculated by using the following equation:

$$NDVI = (\rho_{NIR} - \rho_{Red})/(\rho_{NIR} + \rho_{Red}),$$

where ρ_{NIR} and ρ_{Red} are the surface bidirectional reflectance factors for MODIS NIR and red bands, respectively. The equation for NDVI makes it sensitive to the chlorophyll content of vegetation since only red and NIR bands are considered. The MODIS NDVI product is referred to as the *continuity index* because of its uninterrupted Earth observation records for over 30 years. Its broad spatial and temporal coverage makes it preferred in various applications, including land cover/land use classification; health and epidemiology; drought detection; land degradation, desertification, and deforestation; change detection; and monitoring.

A number of ecological studies using remote sensing technology have specifically focused in China. To investigate the impact of climate change on vegetation variations, Meng, Ni, and Zong (2011) used NDVI derived from the Pathfinder advanced very high-resolution radiometer (AVHRR) dataset with global coverage as a proxy for ground vegetation. In their study, they used 10-day NDVI time-series data with 8-kilometer spatial resolution as the dependent variable and monthly mean climatic data, including surface temperature and precipitation with 0.5° spatial resolution, as independent variables. They observed significant monthly variations in climatic variables sharing the same temporal structure, with abrupt changes. Meanwhile, the NDVI demonstrated an overall increasing trend over the studied 19 years from 1982 to 2000, whereas Decembers presented the largest change in variability. Based on their correlational analysis, Meng et al. (2011) concluded that the temporal pattern of climatic variables imposed more substantial impacts on vegetation variation than the magnitude of change in the same independent variables.

Using a similar dataset with a longer temporal span, from 1982 to 2011, Xie et al. (2016) examined the vegetation and climate change dynamics of Loess Plateau of China. To

investigate the dynamics of climate changes and NDVI, they used linear least square regression to understand the relationship between NDVI from AVHRR and annual mean values of temperature and precipitation. It was observed that temperature and precipitation are both positively correlated with NDVI, indicating rising temperature and precipitation are beneficial for vegetation growth. Over the studied three decades, temperature exhibited an increasing trend with an annual change of 0.05 °C, whereas the changes in precipitation showed a non-significant decreasing tendency. Comparing to climatic variables, NDVI presented a more complex pattern across the same period. NDVI significantly increased in growing season with homogeneous temporal patterns, though the annual vegetation growth trend represented by NDVI was spatially heterogeneous ranging from -0.05 to 0.05 across the plateau. They also noticed the influence of climatic variables on NDVI was both seasonal and regional. Others also ascertained this relationship between vegetation indices and temperature and precipitation variables. Instead of using NDVI, Jiapaer et al. (2015) analyzed the correlation between leaf area index (LAI), which is another type of useful VI, and climate change in Xinjiang of China using linear least square regression as well. They observed a similar positive correlation between LAI and precipitation and temperature and pointed out the season-dependent pattern of impact on LAI. Precipitation had a significant impact during the winter seasons, and temperature affected LAI strongly during spring and summer seasons. In other regions of the world, such as central Africa, the correlation between meteorological variables and VIs were also examined (Onema & Taigbenu, 2009), and a positive relationship between NDVI and precipitation was observed.

Usually used as an indicator of relative biomass and greenness, NDVI has been widely accepted as the primary source for estimating the NPP of vegetation using the remote

sensing approach. Since the calculation of NDVI is solely based on both near infrared (ρ_{NIR}) and red (ρ_{Red}) bands, the NDVI values saturate in multilayer closed canopies. The critical disadvantage of NDVI is its limited sensitivity to both atmospheric aerosols and soil background (S. Li et al., 2013). To overcome the drawback of NDVI's limited performance for canopy background, another standard MODIS product, EVI, plays a significant part in predicting the NPP (Qiu, Zeng, Tang, & Chen, 2013; Sjöström et al., 2011) by optimizing the vegetation signal with improved sensitivity in high biomass regions. EVI enhances vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmospheric influences. The equation for EVI is given as

$$EVI = G\,(\rho_{NIR} - \rho_{Red})/(\rho_{NIR} + C_1 \times \rho_{Red} - C_2 \times \rho_{Blue} + L),$$

where *G* is the gain factor, C_1 and C_2 are coefficients of the aerosol resistance terms for red and blue bands, and *L* is the canopy background adjustment. Since EVI takes the blue band into the calculation, it corrects the distortions caused by the canopy background and aerosol influences. This correction makes EVI atmospheric resistant (Kaufman & Tanre, 1992), which is why it performs well in aerosol and canopy-intensive conditions. Given the advantage of EVI, it can be used to monitor long-term variations of vegetation growth.

As both NDVI and EVI are derived from the red and near-infrared bands of the vegetation canopy and processed to reduce the adverse effects of environmental factors, such as atmospheric conditions, soil background, and a wide range of sensor views and sun angle conditions from MODIS (Shen et al., 2010; Son, Chen, Chen, Minh, & Trung, 2014), they share similar characteristics, and both present accurate and valuable snapshots of ground cover and vegetation. Huete et al. (2002) compared the MODIS VI products against in situ field biophysical data collected over four validation test sites representing a variety of land

surface biome types for the growing season in the year 2000 in an experimental research study. Among those land surface types, one of the sites, the Walnut Gulch Experimental Watershed, a United States Department of Agriculture (USDA) research station located in southeastern Arizona, was primarily used for grazing purpose. In their multi-temporal comparison, they discovered that the EVI, NDVI, and in situ collection values agreed with each other. Additionally, EVI was found to be sensitive to seasonal vegetation variations, land cover changes, and biophysical parameter changes, and the EVI annual profiles also depicted the growing season of the various biomes fairly well. These advantages of EVI indicate that EVI can be used as a valid proxy of the variations of land surface biomass and can provide better performance than NDVI.

Statistical Models

For quantitative studies, data serve as the only link we have to represent reality. Therefore, data analysis is the only way that we have to look for the underlying mechanism of any given phenomenon, and it is also the critical link in the scientific research cycle of observing, analyzing, synthesizing, and theorizing (Huang & Attoh-Okine, 2005), which unavoidably involves the selection of a statistical model. In cross-sectional studies using VIs, the most common data analysis approach is to compare the differences between descriptive statistics collected from spatial heterogeneous samples. When VIs are used to examine the transition of the phenomenon, the time-series analysis is usually favored. To describe quantitatively the coupling relationships between human and nature using cross-sectional time-series data, researchers have experimented with various advanced models in recent decades.

The structural equation model (SEM) is one of the innovative models used to explore the interactive effects of human and nature. SEM is a multivariate statistical method allowing evaluation of inter-correlated dependent and independent variables, with the ability to address latent variables, where the latent variables are usually referred to as unobserved variables. The structural equation model is specified as

$$Y = \beta_1 + \beta_2 X + u_Y.$$

At the same time, the independent variable *X* is related to *Y*

$$X = \alpha_1 + \alpha_2 Y + \alpha_3 U + u_X.$$

This model can be rewritten as

$$Y = \frac{\alpha_1 + \alpha_2\beta_1 + \alpha_3 U + u_X + \alpha_2 u_Y}{1 - \alpha_2\beta_2},$$

where u_Y and u_X denote the disturbance in each equation (Dougherty, 2011, pp. 332-333).

A significant strength of SEM is its ability to incorporate unobserved variables and measurement errors explicitly (Fornell & Larcker, 1981). Malaeb, Summers, and Pugesek (2000) reported that their structural equation modeling of 11 environmental variables revealed complex direct and indirect interactions between natural variability and growth potential. Arthonditsis et al. (2006) used SEM to explore the dynamics of ecological structures and demonstrated SEM's ability to provide a convenient means for assessing the relative roles of ecological processes. A variation of SEM was introduced by Lamb et al. (2014) to analyze the spatial information commonly found, but not addressed adequately in environmental studies. Their spatially explicit SEM of variance/covariance matrices provided interpretable plots of change in path coefficients across scale, and the application of the model effectively evaluated the broad spatial relationships in the sample data set. In a recent application, SEM was also used to analyze the effect of human intervention on wetland

integrity. Schweiger, Grace, Cooper, Bobowski, and Britten (2016) applied SEM to investigate how anthropogenic disturbance affects the wetland ecology of Rocky Mountain National Park in the long-term. Their findings indicated that, though evidence showed mixed indirect effects of human disturbance on wetland integrity, a higher level of human disturbance closely correlated with more severe biological degradation.

When correctly applied, SEM procedures produce solid preferable results over principal components analysis or factor analysis for their greater flexibility in modeling relationships between variables, controlling errors introduced by measurements and unobserved variables, and statistically testing a priori and assumptions against empirical data. The introduction of SEM allows social scientists to perform path analysis with unobserved variables, and Fornell and Larcker (1987) even described this approach as an example of the "second generation of multivariate analysis" (p. 408). However, application of SEM requires a great level of knowledge about the conditions and assumptions for appropriate usage. While mathematicians, statisticians, and other SEM experts are commonly aware of the SEM requirements, and discussions of the requirements can be found in textbooks (Hoyle, 1995), lacking due consideration can lead to flawed or invalid results and conclusions. Chin (1998) suggested that numbers of published SEM applications are suffering from serious flaws caused by insufficient understanding of the requirements. More importantly, SEM is incapable of simultaneously investigating cross-sectional and time-series effects. SEM application can either explore the cross-sectional or time-series effects, but, when a data set consists of longitudinal observations of multiple units (e.g., the cross-sectional and time-series data collection employed in this research examining the coupled effects of climate

change and socioeconomic transformation on grassland productivity), the limitation of SEM drives us to seek appropriate alternatives.

The cross-sectional and time-series observations of samples in this study create panel data. Panel data are longitudinal datasets that follow a given number of individuals from the population over time, thus providing multiple observations on the individuals in the sample (Hsiao, 2014). In other words, panel data are obtained through repetitive observations of the same set of individuals from a population. While traditional statistical data analysis procedures are used to describe the relationships between variables in quantitative and qualitative ways, their abilities are limited to strictly cross-sectional or time-series observations and reach a threshold when being applied to analyzing panel data. They cannot take both cross-sectional and longitudinal information into consideration simultaneously. For data with both spatial and temporal properties, a new method is required to understand the causal relationship between the dependent and independent variables.

In detailed examinations of panel data analysis, Hsiao (2007, 2014) pointed out that, by including the within-individual dynamics and between-individual differences, panel data have several advantages over cross-sectional or time-series data. First of all, panel data can generate a more accurate inference of model parameters by containing more degrees of freedom and more sample variabilities through a larger size of sample observations than cross-sectional or time-series data, thus improving the efficiency of statistical estimates. Secondly, panel data possess greater capabilities of capturing the complexity of behaviors than a single cross-sectional or time-series dataset. Panel data pool the data of heterogeneity to assess more precise individual outcomes and test more complicated behavioral hypotheses rather than generating predictions using data on the individual in question. Panel data

produce model parameter estimations using all sampled individuals simultaneously and provide micro foundations for aggregate data analysis by investigating homogeneity and heterogeneity issues. Thirdly, the use of panel data controls the impact of omitted or unobserved variables, which are not unusual during statistical analysis. Panel data use both the intertemporal dynamics and the individuality of the observations being investigated to control the effects of missing or unobserved variables and uncover dynamic relationships. Last but not least, panel data simplify the computation and statistical inference as they include two dimensions, cross-sectional and time-series, thus making an analysis of non-stationary time-series, measurement of errors, and dynamic Tobit models possible (Hsiao, 2007).

A generic panel data regression model is defined as follows:

$$Y_{it} = \alpha + \sum_{k=1}^{K} X_{itk}\beta_{itk} + \delta_i + \gamma_t + \varepsilon_{it},$$

where Y_{it} is the dependent variable for unit *i* at time *t*, X_{itk} is the *k*-th independent variable for unit *i* at time *t*, and *K* is the number of independent variables. Parameter-wise, α is the intercept, which is the overall constant of the model, β_{itk} is the slope coefficient for the *k*-th independent variable for unit *i* at time *t*, which can be either cross-sectional, period, or a combination of both, δ_i represents the cross-sectional effects, γ_t represents the temporal-specific effects, and ε_{it} captures the errors that vary across sections throughout the studied period and determines whether the model is a fixed-effect or a random-effect model. In a fixed-effect model specification, ε_{it} is assumed to vary non-stochastically over *i* or *t*, making the fixed-effect model analogous to a dummy variable model in one dimension. In a random-effect model specification, ε_{it} is assumed to vary stochastically over *i* or *t*, requiring special treatment of the error variance matrix. In other words, the fixed-effect model assumes that

the effect upon individuals varies over the time period but is consistent for all cross-sectional individuals at any given time, and the random-effect model assumes the effect varies both cross-sectionally and temporally for individuals.

Both fixed-effect and random-effect models have their unique advantages. The fixed-effect model specification allows the individual and/or time-specific effects to be correlated with explanatory variables X_{itk}, whereas the advantages of random-effect model specification are as follows (a) the number of parameters stays constant when sample size increases, (b) it allows the derivation of coefficient estimators that make use of both within and between group variations, and (c) it allows the estimation of the impact of time-invariant variables. On the other hand, the advantages of one model are exactly the disadvantages of the other. For the random-effect model, it requires an error variance matrix, which is unobservable; the fixed-effect model does not allow the estimation of the coefficients to be time-invariant, and the number of unknown parameters increases as the number of sample observations increases. In practice, fixed-effect model parameters are obtained by using the generalized method of moments technique, and we use the generalized least square to calculate random-effect model estimators. The selection of a fixed-effect or a random-effect model depends on the objective of an analysis and the trend of the explanatory variables, and the Hausman (1978) test is often used to test which model is a more appropriate fit for the sample observation. The Hausman test examines if the individual effects are uncorrelated with other independent variables in the model. The null hypothesis of the Hausman test assumes the individual effects are not correlated with any other independent variables. Under the null hypothesis, the coefficients from the random-effect model are no longer the best linear unbiased estimators because the individual effects are part of the error term; thus, the

fixed-effect model is preferred. If the null hypothesis is rejected, then the random-effect model is favored.

Being widely applied in econometrical studies for its capability of capturing multi-dimensional relationships for cross-sectional and time-series variables, panel data analysis has been favored in recent environmental and ecological research as an efficient regression technique that evaluates multi-scaled temporal-spatial relationships. Liu and Xie (2013) used a panel data model-derived filtering method to examine the spatial-temporal effect of land use and water resources on the economic growth of China from regional and national scales. This article addressed endogeneity and exogeneity with different specifications of the model and identified spatial and temporal autocorrelation among the explanatory variables. LeSage (2014) specified panel data models in a Bayesian approach, spatial Durbin model, and spatial Durbin error model to investigate the global and local spatial spillovers. Estimated using Markov chain Monte-Carlo simulation, its application of state-level cigarette demand from 49 US states over 16 years included several socioeconomic variables as independent variables and successfully captured the presence of a significant cross-border shopping effect from smokers.

Overall, in terms of describing the causal relationships between dependent and independent variables for a cross-sectional longitudinal dataset, the panel data model not only controls the unobserved variables but also provides superior accuracy in determining the model parameters due to its estimating mechanism.

Data Processing Techniques

In regression analysis of cross-sectional time-series data, the number of available methods has been limited, which relegates the crucial phase of data analysis to *data*

processing, where certain well-established algorithms can be applied to extract useful information to guide data analysis. Per this study, the panel data set, including the EVI from remotely sensed observations and climatic factors from ground stations, makes panel data analysis a natural and solitary fit. However, the time-series data contain noises (e.g., seasonal variations) and discontinuities resulting from disturbance events (e.g., extreme meteorological event; De Beurs & Henebry, 2005). Baho, Futter, Johnson, and Angeler (2015) used asymmetric eigenvector maps (AEM) and Moran's eigenvector maps (MEM) to reveal the significant temporal structure of variables measuring water quality, yet were not able to describe the patterns quantitatively. The complexity of time-series datasets creates a roadblock in analyzing long-term trends using unprocessed data and makes it difficult to extract tendencies using conventional data processing techniques. An advanced data processing method is required to facilitate our panel data analysis.

The Fourier transformation (FT) is the most well-known signal processing technique used to uncover the global harmonics dominated by the oscillatory behavior of data. The most critical restriction before applying FT to data analysis is that the data must be linear, and the data must be strictly periodic or stationary. Otherwise, the results will make little physical sense (Huang et al., 1998). However, as pointed out by Tucker et al. (2005), the NDVI time-series is non-stationary. Stationarity is a unique property of time-series data. The usual form of a stationary time-series data set is described as

$$X_t = \beta_1 + \beta_2 X_{t-1} + \varepsilon_t,$$

where β_1 is the intercept, β_2 is the coefficient and $|\beta_2| < 1$, and ε_t is independently and identically distributed with zero mean and finite variance. In general, a time-series data set is stationary if it satisfies three conditions:

1. The mean of the distribution is independent of time;
2. The variance of the distribution is independent of time; and
3. The covariance between its values at any two time points depends only on the distance between those points, and not on time. (Dougherty, 2011, pp. 463-469)

On the contrary, time-series data that violate any of these three conditions for stationarity is defined as being non-stationary (Dougherty, 2011). To address the drawbacks of FT, an improved technique, the short-time Fourier transformation (STFT), was introduced. The earliest application of this method was discussed in Allen and Rabiner (1977). This approach separates the temporal signal into a series of small, overlapping sub-series, and each individual sub-series uses a sliding window then further Fourier transformed, from which comes the major drawback. Considering the similarity between NDVI and EVI, as both of them are derived products from spectral transformation despite the different applications due to their specialties, it is arguable that FT is an appropriate data processing technique that can be used in decomposing the EVI time-series due to stationarity.

The singular spectrum analysis (SSA) experimented with by Ghil et al. (2002) aims to address time-series data of Southern Oscillation Index. SSA involves decomposing the time-series data into smaller segments of signals according to some choice of an embedding dimension, locating the empirical orthogonal functions (EOF) from the segments, and then projecting the time-series using the EOFs to define the principal components of the original data. Vincent, Giebel, Pinson, and Madsen (2010) evaluated this method and concluded that this approach is useful for analyzing nonlinear time-series. However, this approach requires the researchers to define an underlying periodicity in the data presumably. The relationship between the projection of the time-series and the set of globally defined EOFs is impaired

because of the significant difference in the statistical properties of the segments and the original time-series data (Vincent et al., 2010).

Another approach to understanding time-series data is to apply the wavelet transformation, which is a spectral analysis method. A wavelet function can be stretched or dilated then projected onto the original data to find the most valuable frequencies at each time step (Labat, 2005). Wavelet transformation has been considered to be able to capture the changing spectral behavior in non-stationary data and is powerful in studying geophysical time-series, such as turbulence measurements (Barthlott, Drobinski, Fesquet, Dubos, & Pietras, 2007). Martinez and Gilabert (2009) also applied wavelet transformation in their NDVI time-series analysis, and they successfully decomposed short-term, seasonal, and long-term variations from the non-stationary dataset. However, one of the critical shortcomings of wavelet transformation is that some a priori decisions must be made about the amplitudes and frequencies of the wavelet functions, which limits the adaptation of this method (Vincent et al., 2010).

A comparably new signal processing technique for analyzing nonlinear and non-stationary time-series data, empirical mode decomposition (EMD) was first introduced by Huang et al. (1998) and has rapidly gained researchers' attention. EMD is based on Hilbert-Huang transformation (HHT) and is an entirely data-adaptive technique consisting of an empirical filter that decomposes any completed dataset into a linear combination of a finite and often small number of intrinsic mode functions (IMFs). The IMFs represent the instantaneous amplitudes and frequencies of each component that are imposed by an array of seasonal, short-term, and long-term patterns and a residual, which represents the trend of the original time-series. Unlike the Fourier transformation that uses a global filtering function

and strictly requires the time-series data to be stationary, EMD can be self-adaptively applied to nonlinear and non-stationary data. Unlike the wavelet transformation, which requires a priori decision about the wavelet functions, the decomposition algorithm of EMD is based on the local characteristics of the time-series, making it empirical, nonparametric, and highly efficient. To demonstrate the potential advantages of EMD in capturing time-evolving frequencies in time-series analysis, Huang et al. (1998) systematically compared the EMD to wavelet transformation in three different applications. Compared to SSA, the advantage of EMD is that it is an entirely local method that can describe the changing statistical properties of non-stationary time-series.

Since its introduction in 1998, EMD has been successfully applied to various problems in different fields where non-stationary time-series data are involved. Veltcheva and Soares (2004) used this method to study abnormal ocean waves, and Peng, Peter, and Chu (2005) found it to be a useful strategy for analyzing vibrations generated by industrial machinery. In Duffy's (2004) application, EMD was successfully used to identify the regular diurnal cycles residing in 48,000 hours of continuous observations of sea level heights for more than 6 years. Shen et al. (2005) used HHT to analyze air temperature and sea surface temperature using a time-series dataset spanning 55 years and clearly differentiated the annual, interannual, and multi-decade cycles and a long-term trend from the original data. Rao and Hsu (2008) applied this technique to a series of hydrological and meteorological time-series and systematically demonstrated the differences between the Fourier analysis and EMD. In Vincent et al.'s (2010) time-series analysis of wind speed with 10-minute temporal resolution over a 4-year study period, cyclic patterns of 1-to-3 and 3-to-ten hours were isolated from the data. In a time-series analysis using monthly rainfall data of sub-divisional

India, Reddy and Abarsh (2016) used EMD to identify successfully multiple cyclic patterns from rainfall time-series data collected uninterruptedly over more than a century from southern Indian sub-divisions. They observed repeating patterns of 3 years, 5-to-7 years, and 11 years, which associated with the sunspot cycle, concealed in the historical rainfall dataset that had been continuously collected for over a century.

To summarize, the introduction of EMD into ecological and environmental studies enabled scientists across fields to uncover in an efficient and approachable manner more information hidden behind the cyclic patterns.

Chapter 3. Methods

This quantitative study investigated the long-term relationships of the grassland growth, climate changes, and socioeconomic development of the Inner Mongolia Autonomous Region (IMAR), China, between 2000 and 2014. Remotely sensed moderate resolution imaging spectroradiometer (MODIS) vegetation index (VI) product, enhanced vegetation index (EVI) imagery, between April and September within the study period, were downloaded from the United States Geological Survey's (USGS) website with a 16-day interval. Meteorological data, including seven factors, and were acquired from the Chinese Academy of Science, collected from 45 ground stations on a daily basis, then compiled to coincide with the temporal frequency of the EVI data. Annual socioeconomic data were retrieved from the yearbook of the IMAR, which provides an overview of the economic development in each county. To examine statistically the long-term relationships, the empirical mode decomposition (EMD) technique was applied to the EVI and meteorological data to eliminate the cyclic seasonal and interannual patterns, and then the panel data analysis was applied to the process the dataset for inferential analysis.

Research Design

This study is quantitative and, to be more specific, descriptive, aiming to understand the joint effects of climate changes and human activities on the grassland degradation process in the IMAR, using the EMD signal processing and panel data analysis.

Population, Sample, and Subjects

The population for the EVI is all the values of pixels on the collected EVI imagery from MODIS MOD13Q1 product between 2000 and 2014. As pointed out by Piao et al. (2006) and recommended by biological professionals, the growing season of the IMAR

grassland starts in late April and ends in late September. During the rest of a year, limited vegetation grows, and the majority of the IMAR grassland is covered by snow and ice, which makes the EVI data from winters and springs unusable in estimating vegetation status. Thus, instead of randomly selecting images from the entire population, this study used a purposive sampling strategy (Leedy & Ormrod, 2010) to gather the data. We carefully chose high-quality EVI imagery dated between April and September to fit the growing season window as the foundation and then aggregated the EVI values based on their geographic locations to create one EVI image sample per county per period.

The population for meteorological factors is the daily measurement instrument readings over the 15-year period. Since this study was only interested in climate changes from April to September in the IMAR, we used the purposive sampling strategy (Leedy & Ormrod, 2010) to collect data from 45 ground stations in the IMAR for 15 years, which are the most accurate available representation of climate changes within this region. Then data from specific days were chosen from the growing seasons and reconciled to coincide with the EVI 16-day periods temporally.

For the variables representing socioeconomic development, the population is all counties of the IMAR. Due to the limited number of counties, no sampling process was involved. This study used all counties for data analysis.

Together, the sampled EVI, meteorological, and socioeconomic data comprise the subjects, which include cross-sectional time-series data using seven climatic factors and EVI values to represent climate changes and fluctuations of in vegetation productivity over 15 years.

No human/animal subjects were involved in this study.

Data Collection

In this section, we discuss how data were collected and pre-processed before entering the panel data model. The list of variable names, descriptions, and units of measurement can be found in Table 1. As a remote and less developed region, there has been a sparse number of variables collected and available for the IMAR. Besides, the lengthy time frame of this study makes it so that even fewer variables can be utilized, as only few data collection efforts have been consistent. Our selection of variables, both dependent and independent, was limited by the availability of existing data collections and was also a reflection of existing literature (S. Li et al., 2013; Xie, Crary, Bai, Cui, & Zhang, 2016).

EVI. The data collection process for EVI started with capturing the Earth-oriented observations as imagery using space-borne remote sensors of MODIS. MODIS is a whiskbroom sensor, which, depending on the scan angle of the sensors, may cause the actual area caught by one pixel to increase by as much as a factor of four. Meanwhile, the quality of remotely sensed observations can also be affected by a set of uncontrollable variables (e.g., cloud-coverage, aerosol particles in the atmosphere, time of a day). To avoid geometric distortions and atmospheric disturbances, EVI is composited using a per-pixel-based algorithm that relies on multiple observations over a 16-day period. As the satellite surveys the Earth in an overlapping orbital manner on a daily basis, a maximum of 64 observations of one spot on the Earth can be recorded over a 16-day cycle. Once all 16 days of observations are collected, the algorithm applies a filter to the data based on quality, cloud, and viewing geometry of the image, and to ensure minimal residuals, the best pixel is reconciled using only the high-quality, cloud-free, and good viewing geometry observations within this 16-day period. To composite EVI, the number of acceptable pixels over a 16-day period is

usually less than 10 and sometimes even less than 5 (Huete et al., 2002). Then, the EVI data is published on the USGS website (https://earthexplorer.usgs.gov/) to allow users from around the globe to download. Within the span of the growing season of the IMAR grassland, approximately 11 EVI images are available between April and September. It was determined that nine EVI images from nine consecutive periods would be collected every year, which is consistent with the frequency from S. Li et al. (2013). Our colleague at the Institute for Geospatial Research and Education (IGRE), Lishen Mao, assisted in downloading EVI imagery from the USGS website.

After the EVI data were downloaded, they were first re-projected to the Universal Transverse Mercator (UTM) 50N projection, as the original VI products were in the Sinusoidal projection, which is a pseudo-cylindrical equal area projection. When distances along the equator and the central meridian are preserved in the Sinusoidal projection, distortions increase as the region on the map deviates away from those two lines. Given the fact that the IMAR is geographically located away from the equator and the central meridian, it could be expected that the Sinusoidal projection would introduce a significant level of distortions if used for this study. Thus, the VI products were re-projected to a cylinder-based UTM projection system to prevent spatial distortion of the study area. After re-projection, the EVI images capturing the same period were mosaicked to create one image with full coverage of the IMAR. At that time, the pixel size was transformed to approximately 230 meters, which was not an ideal resolution for calculation or measurement. The satellite imagery were then resampled to 250-meter spatial resolution. As discussed in previous chapters, the values of EVI fall between zero and one, and this value range is comparably

smaller than other independent variables. We transformed the original value range of zero to one to a new domain, zero to 10,000, by simply multiplying the original EVI by 10,000.

Given the vast spatial extent of this research, the extent of each county on the EVI imagery consisted of hundreds of pixels carrying different EVI values. Though EVI values are considered valid proxies of vegetation productivity, we were not able to consume all pixels from satellite imagery as the input of our panel data analysis. We needed to find an alternative representation of a county's EVI. Thus, a single value, the averaged pixel values within the boundary of a county, was assigned to every county at any given time point. Using this methodology, a total of 12,015 observations of EVI (89 counties × 15 years × nine periods per year × one image per period) was collected from the complete EVI data covering the entire IMAR. EVI data processing and sampling were performed using ArcGIS Desktop 10.

Climatic variables. The Chinese Academy of Science collected the meteorological data from 45 ground stations across the IMAR, and the owner of this dataset kindly agreed to share the data with us for this study. Then, we used purposive sampling strategy (Leedy & Ormrod, 2010) to select the exact dates when the EVI data were collected by the remote sensing system, as it was necessary to keep the temporal frequency of the meteorological data consistent with the EVI data. For long-term climate change studies of the IMAR, seven standard meteorological variables, including barometric pressure (*PRS*), relative humidity (*RHU*), precipitation (*PRE*), sunshine duration (*SSD*), temperature (*TEM*), vapor pressure (*EVP*), and wind speed (*WIN*), are measured at each ground station using consistent instruments on a daily basis (Bai et al., 2008). This set of variables was previously used by S. Li et al. (2013) and Xie et al. (2016), though in different temporal settings. For each variable,

basic descriptive statistics (daily minimal, maximal, and mean) are included in the data collection. The average value of the daily measurement was chosen for its representativeness in describing the variation of the factor. To be temporally consistent with the EVI data collection, the daily measurements of meteorological variables were averaged for every 16-day period, except for precipitation, which was summarized so that the cumulative effects of precipitation could be assessed. Considering the lagged response of vegetation to the changes in meteorological events (Richard & Poccard, 1998; Wang, Price, & Rich, 2001), there was a 16-day offset in selecting samples. For example, if the EVI imagery were collected on May 15, then the corresponding period for climatic variables would be between April 30 and May 15. This offset ensures that vegetation responses to climatic changes are correctly captured. A total of 42,525 (45 stations × 15 years × nine periods per year × seven meteorological variables) observations of the meteorological sample was selected.

Although climate variables were collected from discrete geographic locations in the IMAR, climate is spatially continuous across the world without breaks. Also, the limited number of meteorological ground stations, which is fewer than the number of counties, indicates that some counties have no direct climate measurements. A spatial interpolation method, the inverse distance weighted (IDW) interpolation, was used to generate a continuous climate grid surface for each climatic variable per period. Logically, the meteorological phenomenon at a given point on the surface is more likely affected by its surroundings, which makes the distance between the prediction and samples a critical factor in interpolation. The IDW assumes that the influence of events decreases with the distance from the sampled location (Philip & Watson, 1982; Watson & Philip, 1985), which makes it a more preferred interpolation method than Kriging or Spline, as it creates a smoother surface

without abrupt changes. Philip and Watson (1982) also pointed out that influence of an input point on an interpolated value is distance related and isotropic, which made us believe honoring the distance and obtaining consistent predictions was more important than introducing more uncontrollable variations by using other interpolation methods. Another reason for us to choose IDW interpolation is that Kriging or Spline may create negative values during interpolation, which seems impractical for meteorological observations. Our colleagues, Lishen Mao and Yuchen Li, contributed greatly in calculating the optimal parameters for our IDW interpolation. As a result, 945 (seven variables × 15 years × nine periods per year) climate grid maps were created using IDW. Then, similar to the calculation of EVI values for an individual county, an averaged value of each county was computed using the cells within the county's boundary for every climatic variable during each period. Hence, a total of 84,105 (89 counties × nine periods per year × 15 years × seven variables) climatic samples were collected for the balanced panel data model.

To evaluate the accuracy of our spatial interpolation results, we used the Cross Validation tool from the Geostatistical Analyst extension of ArcGIS Desktop. Instead of selecting a subset of observations as the training samples and another subset for validation, the cross-validation method "remove one or more data locations and predict their associated data using the data at the rest of the locations" (Environmental Systems Research Institute, Inc. [Esri], 2017) and repeats this remove-and-predict process for all data locations. To help users understand the effectiveness of the given interpolation algorithm, the returned result contains a table of the measured values, predicted values, and error. For our study, we calculated the average of each climatic variable at each individual station as the representation of our observations and used this data set for the cross-validation test to obtain

an overview understanding of the test performance. As shown in Table 1, the mean errors and root mean square errors from the validation were no more than 1% of the mean values.

Table 1

Cross-Validation Result for Averaged Climatic Variables

Variable	Mean	Mean Error	Root Mean Square Error
EVP	91.1748	0.2874	6.0962
PRE	254.8028	0.1247	44.2221
PRS	9068.1536	3.7806	238.0732
RHU	50.2446	0.0187	4.3286
SSD	90.5595	0.0717	6.2650
TEM	188.4995	0.1667	17.1104
WIN	27.7921	0.1611	5.5204

Besides examining the size of errors, we also plotted the distribution of the errors for each climatic variable so that we could better perceive the characteristics of the errors. Shown in Figure 3 through Figure 9, each chart demonstrates a histogram of the predicted errors for climatic variables at all ground stations. Clearly, all of these histograms follow a bell-shaped normal distribution curve. The normally distributed errors with limited magnitude indicated that the spatial interpolation we chose was appropriate for this study.

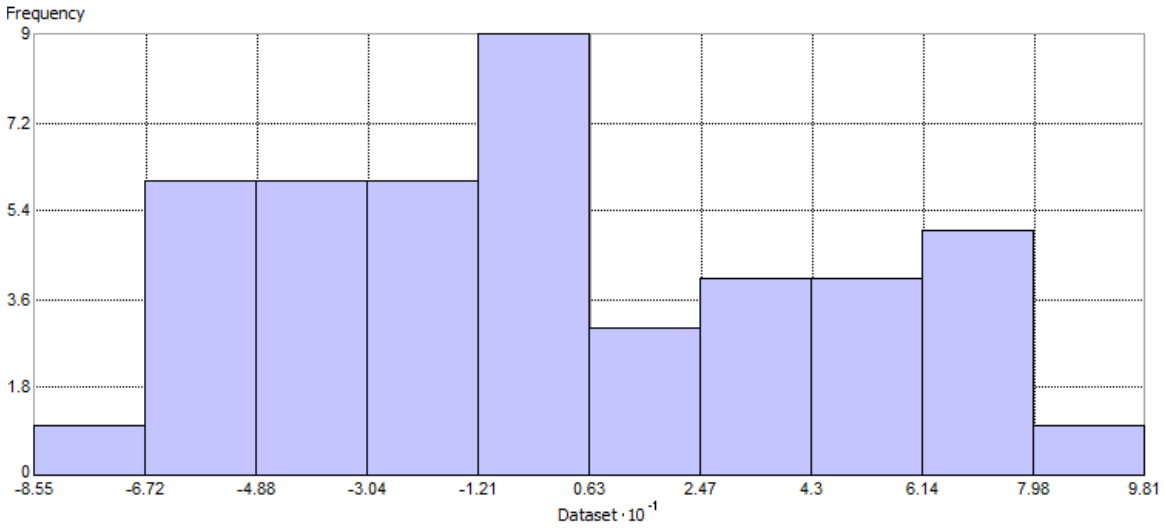

Figure 3. Histogram of error for station averaged *EVP*

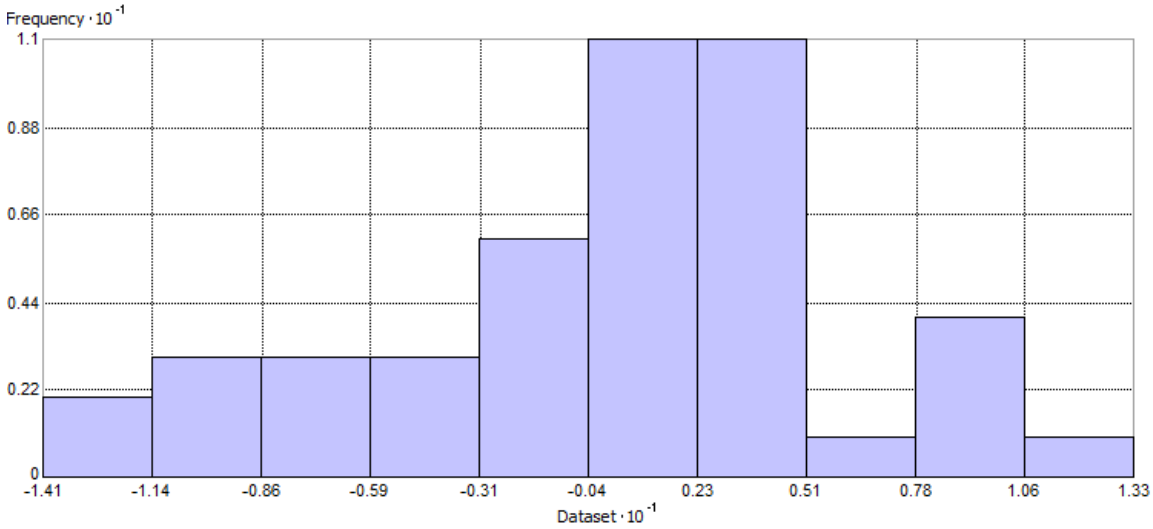

Figure 4. Histogram of error for station averaged *PRE*

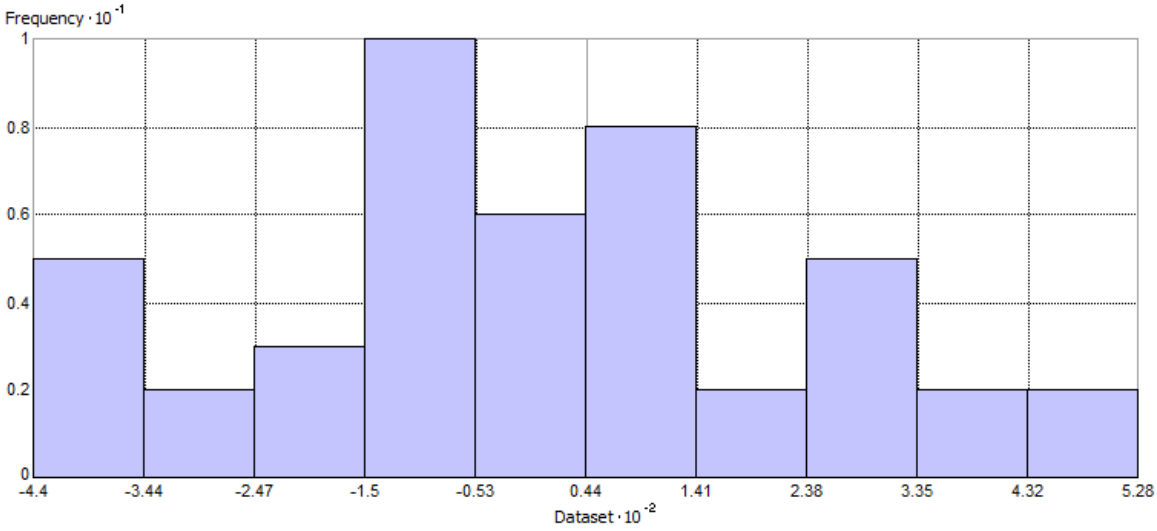

Figure 5. Histogram of error for station averaged *PRS*

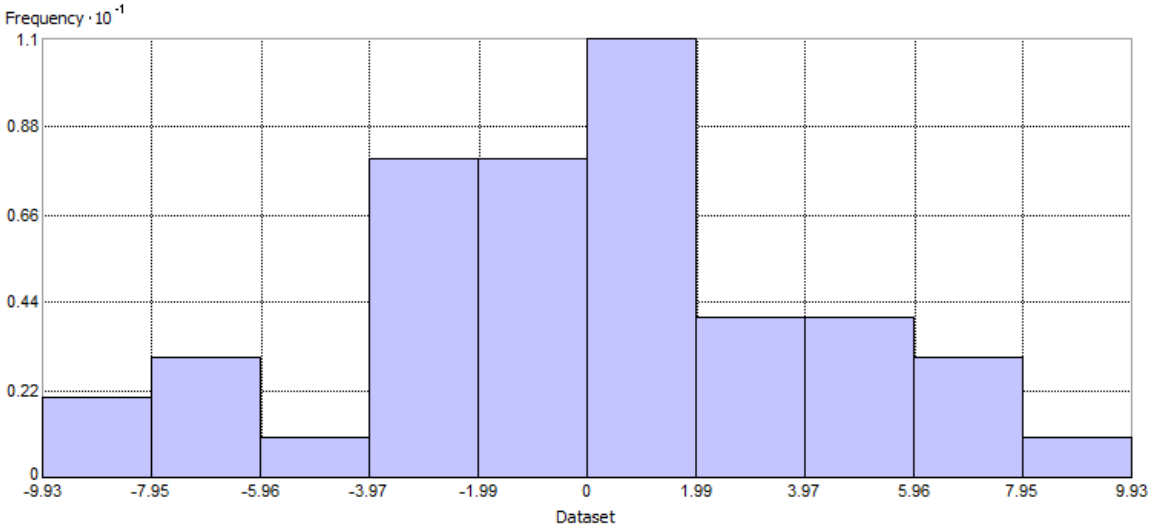

Figure 6. Histogram of error for station averaged *RHU*

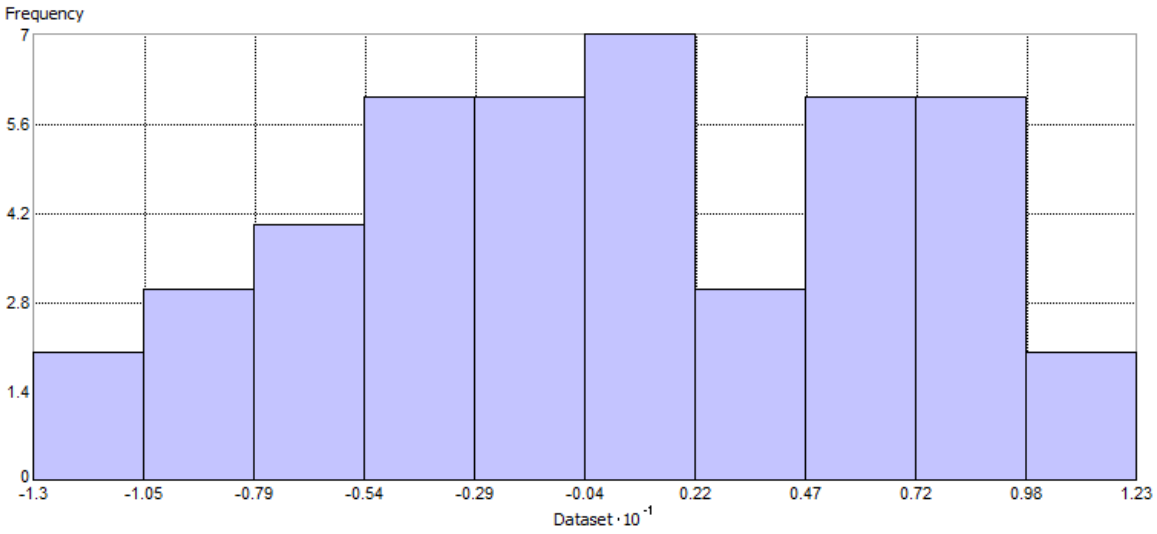

Figure 7. Histogram of error for station averaged *SSD*

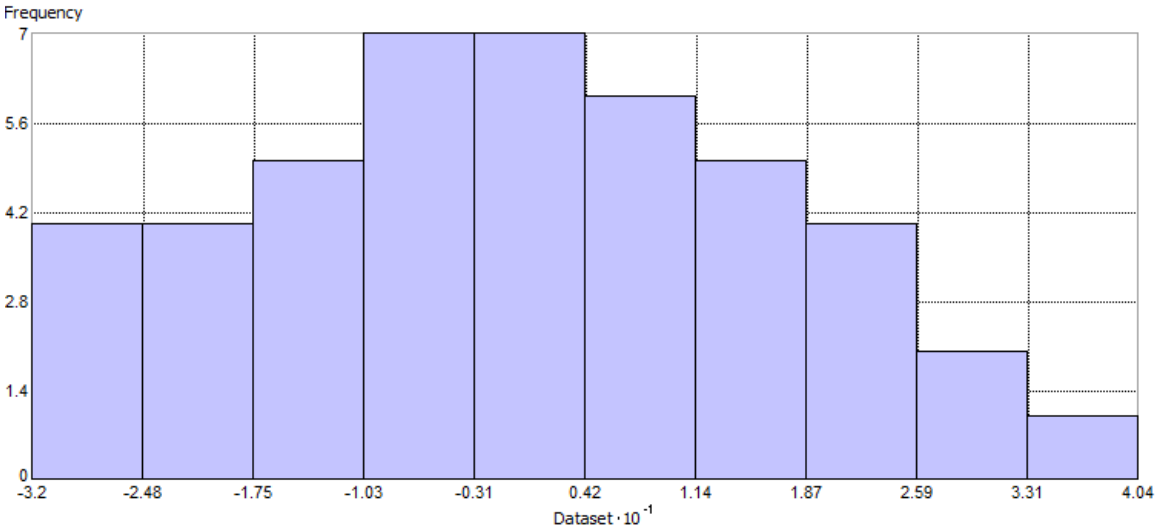

Figure 8. Histogram of error for station averaged *TEM*

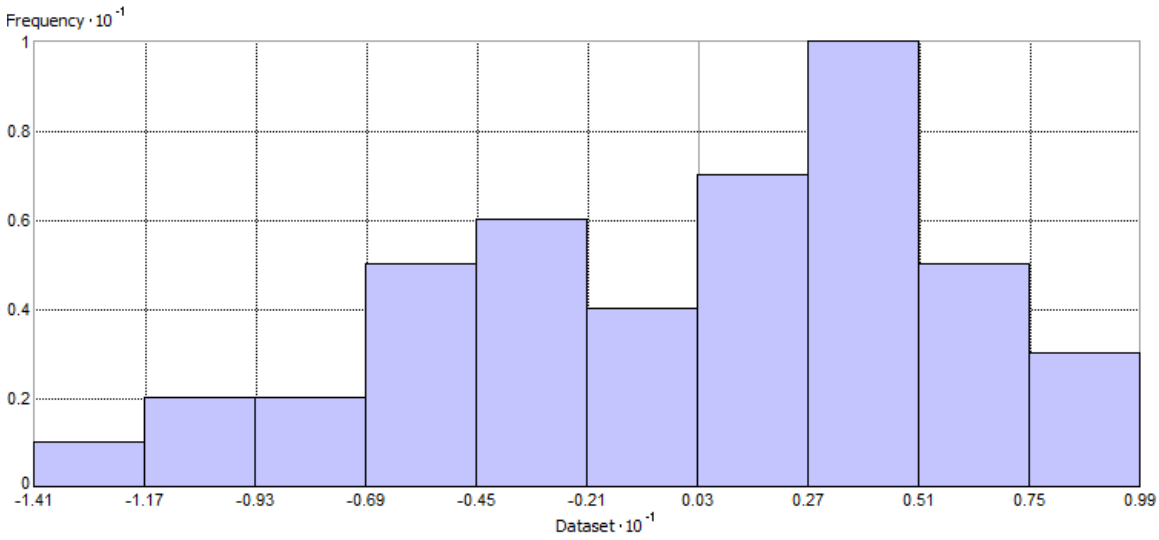

Figure 9. Histogram of error for station averaged *WIN*

Socioeconomic variables. Our collaborators at the Chinese Academy of Science gathered the social and economic variables from the statistic yearbooks of the IMAR, from 2000 to 2014. The annual journal contains brief descriptive statistics of the socioeconomic status of each county, including total population, area, arable area, gross domestic product (GDP), grain production, number of livestock, farming income, length of highway, rural population, local government revenue, and governmental investment. Bearing in mind that EVI and climatic data are collected multiple times per year, and socioeconomic data are only collected once annually, it is necessary to extrapolate the socioeconomic data so that they can be fitted into the panel data model with the same temporal frequency. Considering the inconsistent characteristics of counties (e.g., area, population), these socioeconomic data may not depict the county from an objective perspective, and direct use of these variables for inferential analysis may lead to biased conclusions. To normalize the socioeconomic data to obtain unified measurements, standardized variables, instead of the original variables, were used in the panel data analysis. A similar normalization method was also described in Xie et al. (2016). The following table describes how the normalized variables were generated using

the original data set. Based on the normalized data, we picked nine variables from those shown in Table 2 to represent the characteristics of the IMAR's economy from three perspectives. We used *gdppc* as a snapshot of a county's economic status at the individual level; *daa, dgr, srural,* and *sfarm* to capture the ratio of farming industry to a county's economic performance; and *dhw, slgov,* and *sinv* to indicate how policy affects the county. A similar set of derived socioeconomic variables was piloted in Xie et al.'s (2016) study.

The dependent variable name, description, number of observations, sample size, and units of all dependent variables can be found in the following Table 2.

Table 2

Dependent Variables and Measurement Units

Variable Name	Description	Number of Observations	Sample Size	Unit
EVP	Evaporation	12,015	12,015	0.1 mm
PRE	Precipitation	12,015	12,015	0.1 mm
PRS	Barometric Pressure	12,015	12,015	0.1 hPa
RHU	Relative Humidity	12,015	12,015	1 %
SSD	Sunshine Duration	12,015	12,015	0.1 Hour
TEM	Temperature	12,015	12,015	0.1 °C
WIN	Wind Speed	11,214	12,015	0.1 m/s
gdppc	GPD per capita	1,246	12,015	1,000 Yuan
daa	Density of arable area	1,246	12,015	Percentage (%)
dgr	Density of grain production	1,246	12,015	Ton/KM^2
dls	Density of livestock	1,246	12,015	100 Head/KM^2
dhw	Density of highways	1,246	12,015	KM/KM^2
srural	Share of rural population	1,246	12,015	Percentage (%)
slgov	Local government revenue as share of GDP	1,246	12,015	Percentage (%)
sinv	Governmental investment as share of GDP	1,246	12,015	Percentage (%)
sfarm	All agricultural income as share of GDP	1,246	12,015	Percentage (%)

To capture the polynomial relationships between climate change and grassland productivity, we derived the polynomial terms of climatic variables based on their linear terms. Thus, eventually, our dataset consisted of 24 variables covering 89 counties for 135 periods in 15 years, which formed an outstanding dataset of 288,360 observations. Among these observations, missing values were nearly inevitable. A critical drawback of missing values in a balanced panel model is that the model will omit the subject with a missing value in estimating the model parameters. However, as discussed by Dougherty (2011), "if a balanced panel has been created artificially by eliminating all units of observation with missing observations, the resulting data set may not be representative of its population" (p. 515). Thus, instead of ignoring the subjects with missing values, we used a formula to interpolate the missing observations based on the observed values with similar chronological and spatial characteristics. We assumed a value of climatic variables is linearly correlated with the previous and next observations of the same variable at the same station. Thus, given a missing observation of variable v_c at period t of station i, the unobserved value v_{cit} was calculated by

$$v_{cit} = \frac{v_{ci(t-1)} + v_{ci(t+1)}}{2}.$$

In the strictest examination of our dataset by identifying all missing values and zeros, we located 17,781 observations, which was 6.2% of our 288,360 observations. Considering the minimal percentage of missing data and how we handled them using the best available linear interpolation, we were confident that the missing values would not impose a detrimental impact on our analysis.

Data Analysis

For quantitative studies, data serve as the only link we have to represent reality. Therefore, data analysis is the only way that we have to look for the underlying mechanism of any given phenomenon and is also the critical link in the scientific research cycle of observing, analyzing, synthesizing, and theorizing (Huang & Wu, 2008).

Descriptive data analysis. Data analysis starts with basic descriptive statistics to obtain an overview for a general understanding of the data. The descriptive statistics include the covariance matrix; number of observations; mean, minimal, maximal, and standard deviation of EVI; seven climatic variables; and the socioeconomic variables over the 15-year period.

Stationarity testing. As discussed in the literature review, because of one of the drawbacks of Fourier transformation in time-series data processing, it should only be applied when data are stationary. Unit root test has been used in econometrics for formal stationarity test with no exception. A generalized *p*-th-order difference specification uses the following formula:

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \cdots + \beta_{p+1} Y_{t-p} + \varepsilon_t \ .$$

When a condition for stationarity is $\left|\beta_2 + \cdots + \beta_{p+1}\right| < 1$, then the equation can be easily transformed into the following for convenience:

$$\Delta Y_t = \beta_1 + (\beta_2^* - 1) Y_{t-1} + \beta_3^* \Delta Y_{t-1} + \cdots + \beta_{p+1}^* \Delta Y_{t-p} + \varepsilon_t \ ,$$

where $\beta_2^* = \beta_2 + \cdots + \beta_{p+1}$, and $\beta_3^*, \cdots, \beta_{p+1}^*$ are appropriate linear combinations of $\beta_2, \cdots, \beta_{p+1}$. Under the null hypothesis of non-stationarity, the t-test, also well-known as augmented Dickey-Fuller (ADF) test, has been widely adapted in application (Dougherty, 2011). Li. S. et al. (2013) pointed out that the orders of difference should be kept relatively

low to prevent losing information from the original data, and only first- and second-order differences were performed in their study. In this present study, the first- and second-order differences unit root tests were performed on the EVI data to assess if Tucker et al.'s (2005) claim on the stationarity of normalized difference vegetation index (NDVI) can also be applied to EVI, and these two tests were also applied to the climatic variables to examine their stationarity. These tests were used to examine Hypotheses 1 and 2.

Trend extraction. After descriptive statistics, the EMD technique was applied to the EVI and climatic variables to isolate the long-term trend. This study used an EMD package for MatLab that was developed by Rilling (2017) and was shared publicly for download. Using EMD, the original dataset was decomposed into a set of IMFs representing the periodic patterns and a residual representing the overall trend, which helped us identify the long-term trend of changes visually to examine Hypotheses 1, 2, and 3. This study only focused on the trend of EVI and climate changes, and the trends were carried over to the following data analysis steps.

Panel data analysis. The panel data analysis was conducted using a well-known statistical analysis software, Stata. There were 89 panels, and each panel contained 135 time-periods. In the panel data model, only post-EMD time-series were used. EVI was used as the dependent variable, whereas the climatic variables, which capture the linear correlation, the squared climatic variables, which explain the polynomial effects, and the normalized socioeconomic variables were used as independent variables. The model is described as

$$EVI_{ti} = f(EVP_{ti}, HUM_{ti}, PRE_{ti}, PRS_{ti}, SSD_{ti}, TEM_{ti}, WIN_{ti},$$

$$EVP_{ti}^2, HUM_{ti}^2, PRE_{ti}^2, PRS_{ti}^2, SSD_{ti}^2, TEM_{ti}^2, WIN_{ti}^2,$$

$$daa_{ti}, dgr_{ti}, dls_{ti}, d\not{h}w_{ti}, gdppc_{ti}, sfarm_{ti}, slgov_{ti}, sinv_{ti}, srural_{ti}) + \varepsilon_{ti},$$

where ε_{ti} is the error term of the model, the subscript *t* indicates the *t*-th period of the overall time-series, and *i* identifies the *i*-th county of the IMAR. By examining the coefficients of the regression model, Hypothesis 4 was tested. The panel data model was also applied to the pre-EMD time-series for references.

In ecological analysis, one of the critical considerations researchers need to bear in mind is that climate change is systematic, and the interactions between variables undoubtedly exist. For example, when precipitation increases, it not only stimulates the growth of vegetation, but it also typically leads to a higher level of humidity and a shorter duration of sunshine. Because of less solar radiation, temperature is expected to be lowered, and this further affects evaporation, barometric pressure, and wind formation. Thus, variables other than precipitation either directly or indirectly affect the vegetation growth while the level of precipitation varies, and this kind of direct/indirect effect of applies not only to precipitation. One of the improvements we introduced in this study is the inclusion of a set of additional climatic factors as control variables (*EVP*, *PRS*, *RHU*, *SSD*, and *WIN*) for their causal effect instead of only focusing on the impact of variations in precipitation and temperature, which had been applied in multiple publications (Meng et al., 2011; Shi et al., 2013; Jiapaer et al., 2015; Xie, Jia, et al., 2016). Without control variables, the effect of these variables will be transformed into a part of the effect of precipitation and/or temperature and a portion of the error term, either enhancing or weakening, which in turn generates biased and inaccurate estimation of the model. By explicitly including the control variables in the equation, we systematically improved the efficiency and accuracy of the estimation because the coefficients of these control variables honestly capture their specific direct and indirect effects on the vegetation growth and reserve the error term particularly for the omitted

variables. The existence of additional variables in the equation helps us obtain not only more precise estimators for precipitation and temperature but also a more subjective understanding of the system.

When analyzing the relationships between climate changes and vegetation coverage variation in Guangdong Province of China, Li, Kuang, Huang, and Zhang (2013) found that the existence of an inverted N-shape correlation between population and vegetation. Shi et al. (2013) also noticed that the relationship between temperature and crop yield is more complicated than a basic linear correlation. To examine the possibility of a polynomial relationship between the dependent and independent variables, the t-test was used to evaluate the significance of the coefficients of squared explanatory variables during panel data analysis. Hausman test was practiced to examine whether the fixed-effect or the random-effect specification was an appropriate fit of the data. Based on the selected model, the analysis of coefficient significance and discussions were carried out to determine the dynamics of grassland growth, climate changes, and socioeconomic development. There have been controversial relationships between vegetation responses to different climatic/socioeconomic drivers. For example, it seems logical that vegetation growth performs better with higher temperature. However, in arid regions, like the IMAR, Chuai, Huang, Wang, and Bao (2013) found NDVI was negatively correlated with temperature, both within seasons and between seasons. Deng et al. (2011) identified the positive effects of transportation on grassland ecosystem, yet considered negative by Akiyama and Kawamura (2007). This panel data analysis using long-term trends from 15-year observations is beneficial in justifying the debates.

Chapter 4. Results and Discussion

Understanding the Original Data

Descriptive data analysis. In the pre-empirical mode decomposition (EMD) data analysis, we obtained a general understanding of the nature of the data by calculating covariance matrix, mean, standard deviation, minimum, and maximum of the samples.

The basic descriptive statistics, including a number of observations, mean, standard deviation, minimum, and maximum, of the enhanced vegetation index (EVI), climatic variables, and socioeconomic variables are reported in Table 3. As described in the table, the dependent (EVI) and independent variables (barometric pressure [*PRS*], relative humidity [*RHU*], precipitation [*PRE*], sunshine duration [*SSD*], temperature [*TEM*], evaporation [*EVP*], wind speed [*WIN*], squared terms of *EVP*, *PRE*, *PRS*, *RHU*, *SSD*, *TEM*, *WIN* [*EVP2*, *PRE2*, *PRS2*, *RHU2*, *SSD2*, *TEM2*, *WIN2*], density of arable area [*daa*], density of grain production [*dgp*], density of livestock [*dls*], density of highways [*dhw*], share of rural population [*srural*], local government revenue as share of gross domestic product [GDP] [*slgov*], governmental investment as share of GDP [*sinv*], and all agricultural income as share of GDP [*sfarm*]) created a balanced panel data set, with identical spatial and temporal dimensions.

Table 3

Descriptive Statistics of Variables

Variable	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
EVI	12,015	2,462.57	1,133.10	365.70	6,283.18
EVP	12,015	91.71	47.54	0.00	239.33
PRE	12,015	268.46	236.94	0.00	2,096.57
PRS	12,015	9,039.57	396.60	7,844.14	9,913.63
RHU	12,015	51.79	13.33	15.30	86.74
SSD	12,015	88.89	12.91	26.86	125.68
TEM	12,015	187.08	43.93	10.41	304.45
WIN	12,015	25.71	6.65	10.79	57.21
EVP2	12,015	10,669.71	9,806.97	3.78E-06	57,276.52
PRE2	12,015	128,211.00	251,668.00	0.00	4395,593.00
PRS2	12,015	8.19E+07	7240,163.00	6.15E+07	9.83E+07
RHU2	12,015	2,859.49	1,358.15	233.98	7,523.14
SSD2	12,015	8,068.32	2,238.96	721.68	15,795.83
TEM2	12,015	36,927.97	15,249.51	108.35	92,686.75
WIN2	12,015	705.10	382.48	116.41	3,272.54
gdppc	12,015	48.94	104.68	0.59	1,113.69
daa	12,015	15.84	14.90	0.01	81.08
dgr	12,015	45.29	64.33	0.00	389.00
dls	12,015	1.34	4.81	0.01	170.28
srural	12,015	55.14	26.98	0.29	97.00
slgov	12,015	8.28	31.24	0.04	900.97
sinv	12,015	89.56	247.85	0.01	8,086.83
sfarm	12,015	40.95	38.83	0.04	486.73

While the descriptive statistics provided an overview of the panel data, they failed to portray how the sample observations vary over time. However, given the number of cross-sectional units and the high frequency of observation, it is not feasible to examine the distribution of any variable in a visual manner. Thus, we only picked the dependent variable, EVI, as an example and averaged its cross-sectional value, then plotted the time-series in a

box-plot diagram to demonstrate the variation of observations and obtain a visual understanding of the generic temporal structure of the data, as shown in Figure 10. In this box-plot, the x-axis represents the sequence of EVI observation, and the y-axis is the magnitude of EVI values.

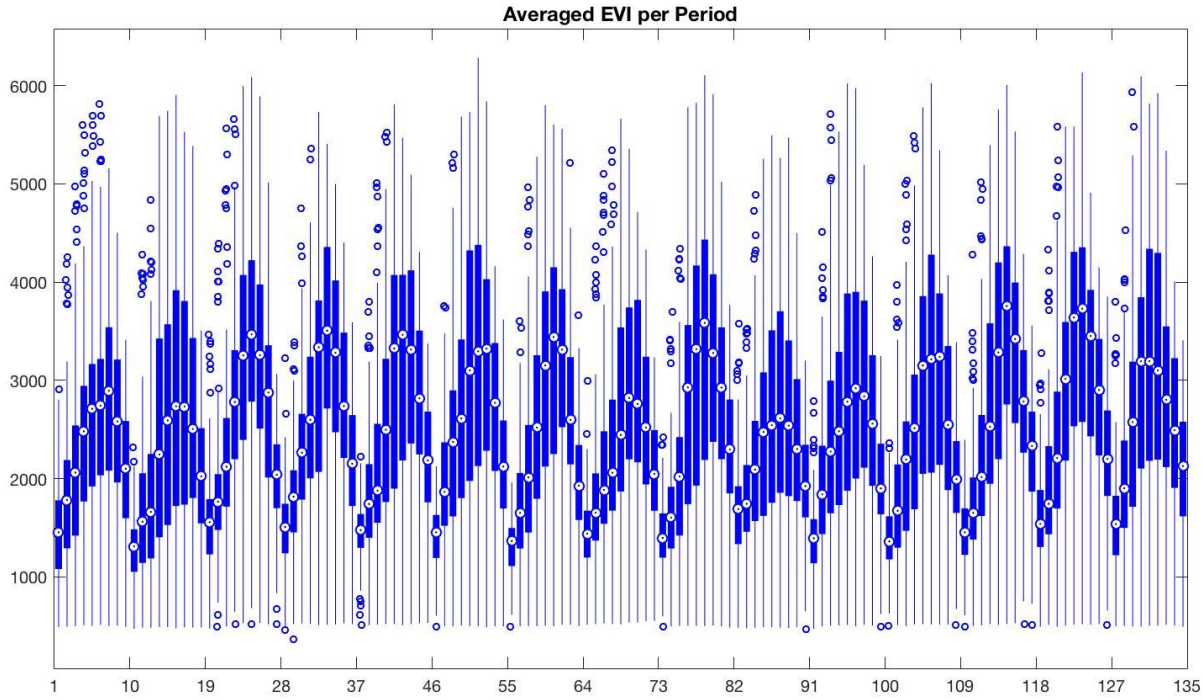

Figure 10. Box-plot of summarized EVI

Based on the descriptive statistics and the box-plot chart, some noticeable properties were easily identified. First, as revealed by computing the square value of the climatic variables, the value ranges and standard deviations of these squared climatic variables were magnified as compared to the original variables. Secondly, the value range of the observations for each variable covered a large domain, and it was arguable whether these extrema were outliners of the population. However, due to the considerable size of the dataset and the limited methods available to verify and exclude the outliners, we chose to preserve the data and use the EMD filter to smooth out the dramatic variations.

Extract the Trend

After obtaining an overview of the characteristics of the data from the pre-EMD data analysis, the EMD filtering technique was applied to the time-series of every variable of each county, which means EMD was applied 2,047 times. For each application of EMD, the original time-series was adaptively decomposed into a finite but uncertain number of IMFs. While the application of EMD seems straightforward, the analysis of EMD results can be challenging due to their complexity. The complexity comes in trifold. First, the adaptiveness, considered as one of the most significant advantages of EMD, also introduces ambiguity because of the uncertainty of how many IMFs will be generated from the application. The differences between the number of IMFs from EMD results makes researchers lack a systematic solution to address heterogeneity. On the other hand, though each intrinsic mode function (IMF) demonstrates a strong temporal signature, the continuous signal is simply a set of values instead of mathematically described, which leaves the understanding of IMFs to researchers' interpretation, mostly through visual observations. Last but not least, the scale of the dataset exceeds the capacity of visual examination. As a result of EMD application on our dataset, tens of thousands of IMFs were created, and the unsystematic signals required researchers to interpret the characteristics on an individual basis then draw conclusions from the unorganized data pool. In summary, the approach of examination, interpretation, and correlation of all IMFs from EMD is not feasible in all perspectives, unless the IMFs are generated in a structural manner or analytical technique significantly improves handling data with less integrity.

Given the incapability of analyzing all IMFs as a whole and the necessity of acquiring an understanding of EMD results, we summarized each variable over space and applied EMD

to derive a global assessment of the dataset. The following diagrams, Figure 11 to Figure 27 showing the EVI and all non-polynomial independent variables respectively, help us to depict the properties of variable time-series. While the y-axis indicates the magnitude of the observation, the x-axis marks the timestamp of the observation, with values from timestamp 1 through 9 representing the observations from the first year and values from timestamp 10 through 18 representing the observations from the second year, etc. For each diagram, the time-series shown in the top row symbolizes the original data. Starting from the second row until the second to the last row at the bottom, each IMF represents one temporal component decomposed from the original data, with various temporal structures. The last IMF positioned as the last time-series from the bottom with a comparably smoother curve is the trend of changes isolated from the 15-year observations. As this research only focused on analyzing the time-series trends and also due to the challenge in interpreting all IMFs, our examination and discussion of the decomposition are limited to the summarized variables in the following paragraphs.

To begin with, the shape of the original EVI time-series seems appealing, as well as the decomposed IMFs. It is easily observable that the unprocessed data distribution is affected by a well-paced energy with a recurring cycle approximately every nine periods, which is exactly the number of observations per year. Shown in Figure 11, the pulse-like temporal structure was successfully identified as IMF 1, and each nine-period segment shares similar magnitude of variation. The periodical pattern continues in IMF 2, 3, and 4, though the intervals of cycling become longer and less definitive. IMF 2 indicates a repeating pattern that occurs approximately every 36 periods, which is 4 years, whereas IMF 3 and IMF 4

show a 5-year and an 8-year pattern, respectively. After all, the last IMF, which is labeled as Residual in the diagram, demonstrates an increasing trend of EVI over the observed 15 years.

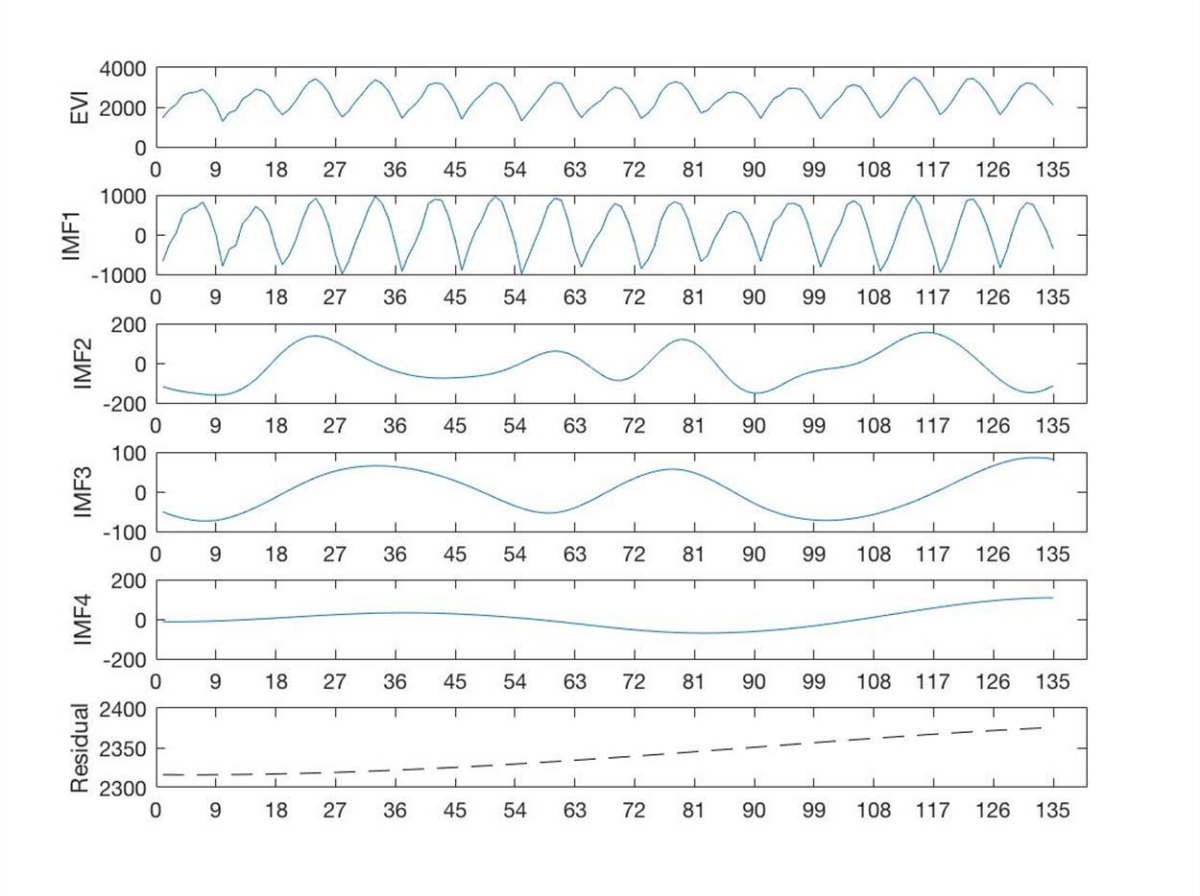

Figure 11. EMD result for EVI

The EMD result of each climatic variable, including *EVP*, *PRE*, *PRS*, *RHU*, *SSD*, *TEM*, and *WIN*, presents the factors composing the original time-series uniquely and are shown in Figure 12 to Figure 18. Though the adaptation of EMD application varies among variables, the results share some universal properties. Overall, the temporal structures they revealed are less distinctive than the ones from EVI. As observable from the diagrams, the fluctuations of the original time-series within a year are much more frequent than the EVI. Unlike the EVI decomposition where the yearly cyclic pattern is filtered out as the first IMF, the first IMF from EMD application on climatic variables represents a subtle periodic pattern

attributed to the fluctuation within a year. The annual pattern is usually listed as the second IMF, like *PRE*, *PRS*, *TEM*, and *WIN*. Meanwhile, in some instances, a semi-annual or bi-annual pattern is attributed to the variation at the same time, like IMF 2 and 4 of *EVP*, IMF 3 of *PRE*, IMF 3 of *PRS*, IMF 2 and 3 of *RHU*, IMF 2 and 3 of *SSD*, IMF 3 of *TEM*, and IMF 3 of *WIN*. The combination of semi-annual, annual, and bi-annual cyclic patterns strongly affects the distribution of our observations from a temporal perspective. Starting from IMF 4 to IMF 6, if present, these IMFs denote patterns that last for a longer period of time, varying from 5 to 12 years. The last component from the EMD application, Residual, is the trend of change during the observation period. Based on the shapes of the residual, we categorized their trends into four groups in Table 4: increase (I), decrease (D), increase-decrease (I-D), and decrease-increase (D-I).

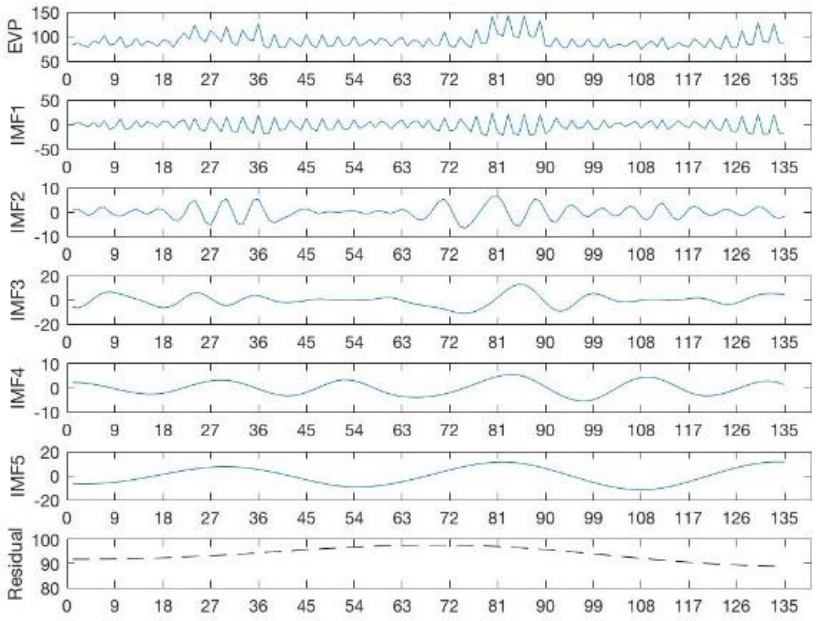

Figure 12. EMD result for *EVP*

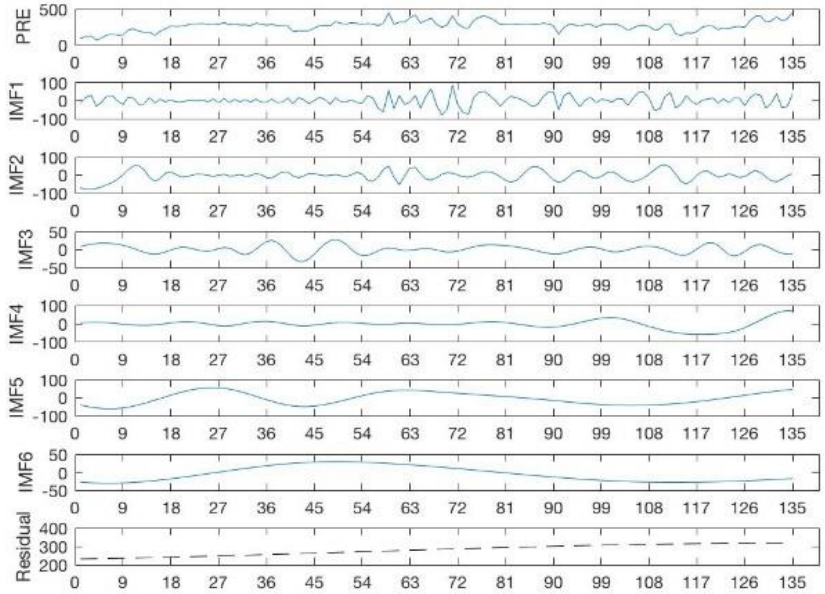

Figure 13. EMD result for *PRE*

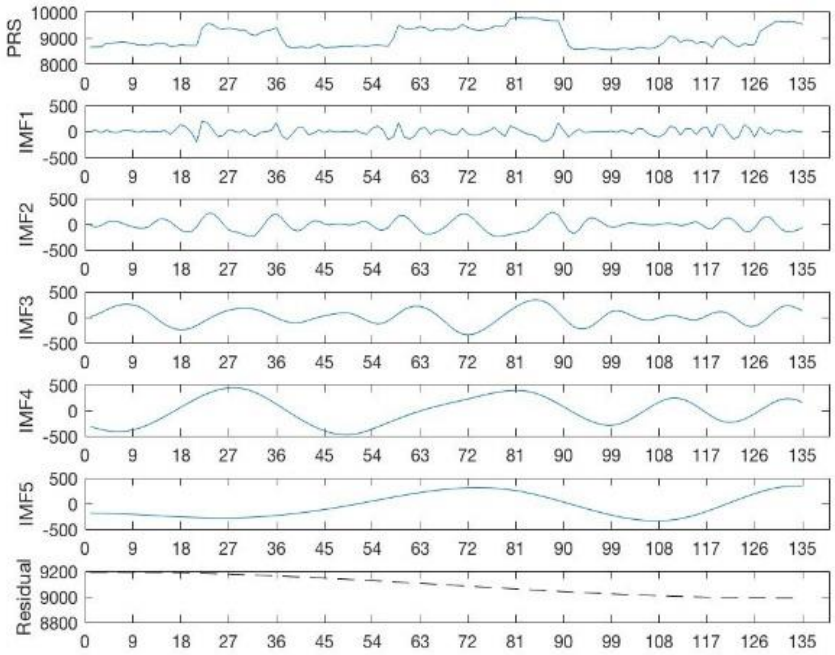

Figure 14. EMD result for *PRS*

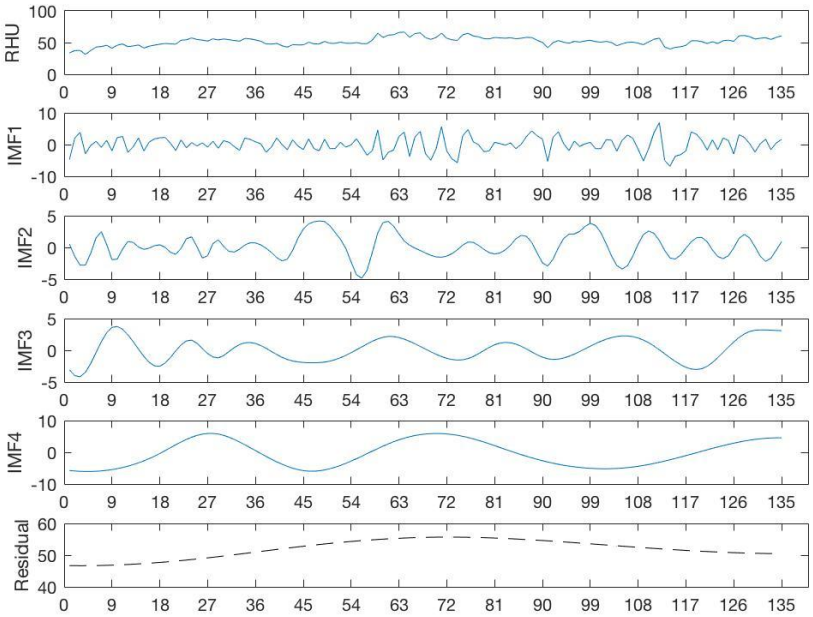

Figure 15. EMD result for *RHU*

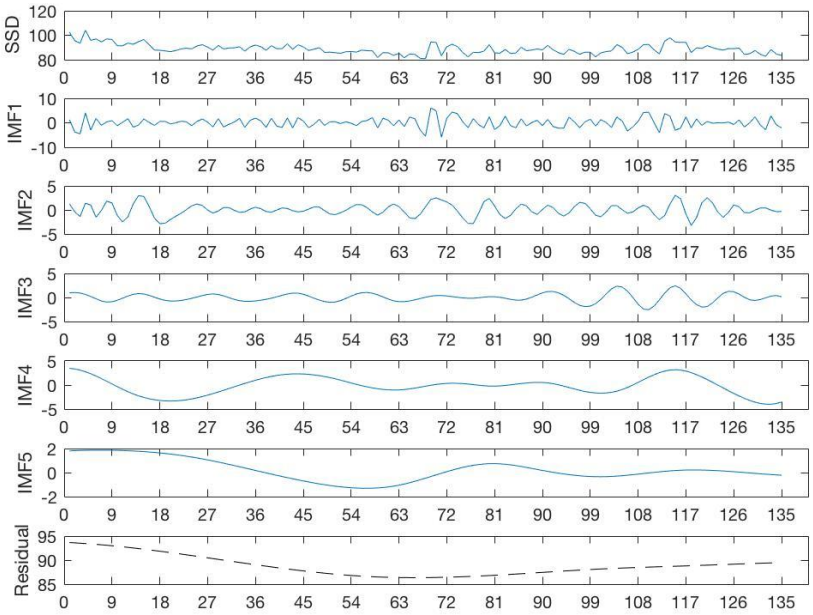

Figure 16. EMD result for *SSD*

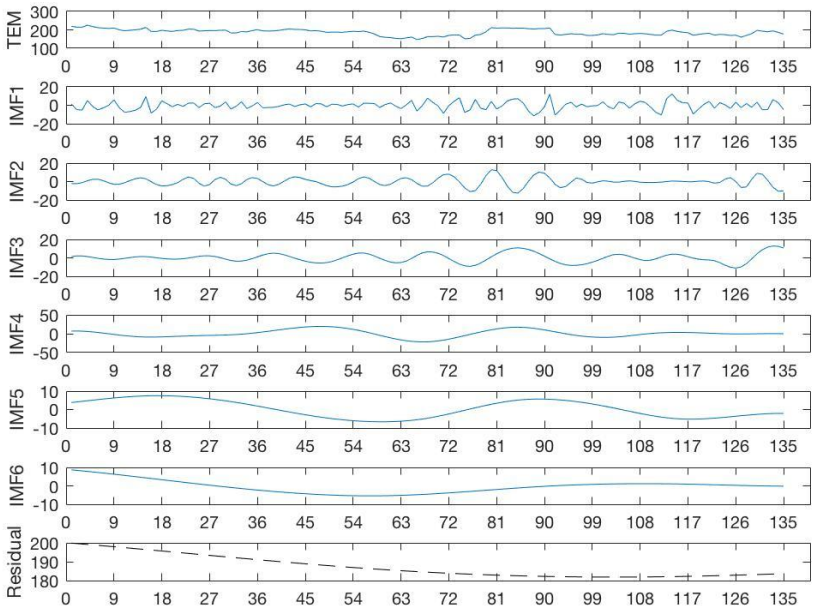

Figure 17. EMD result for *TEM*

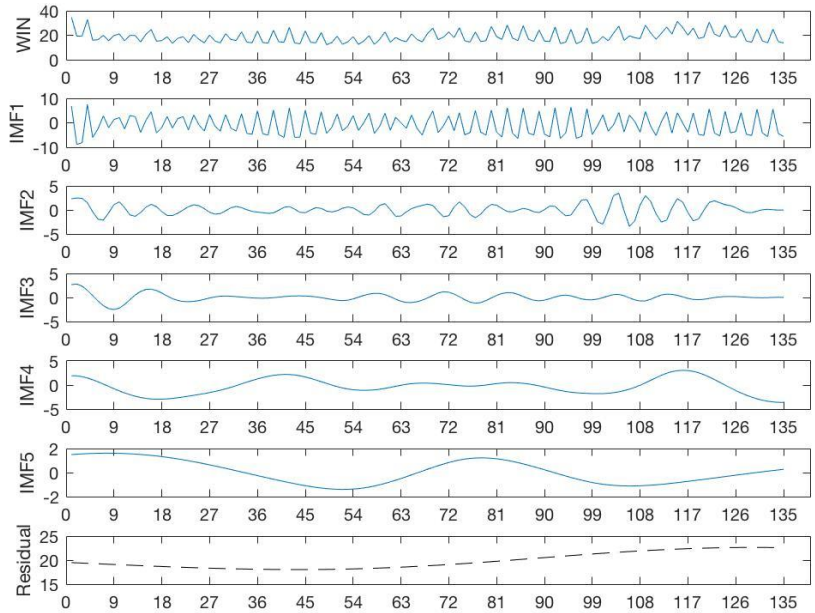

Figure 18. EMD result for *WIN*

Table 4

Visual Interpretation of Climatic Variable Trend

Trends	**I**	**D**	**I-D**	**D-I**
Variable Name	*PRE*	*PRS*	*EVP, RHU*	*SSD, TEM, WIN*

Given the shape of the trends, it is not difficult to comprehend that evaporation is positively correlated with humidity, and the temperature is positively correlated with sunshine duration; meanwhile, higher temperature will possibly lead to a higher level of evaporation and lower humidity. However, a lack of ecological and meteorological knowledge prevents us from further interpreting the cause of changes in the trends' directions.

For the socioeconomic variables, the time-series demonstrates fewer within-year variations due to the extrapolation technique we used in creating the balanced panel, and the between-year differences seem abrupt in some cases. As shown in Figure 19 to Figure 27, the application of EMD not only successfully smoothed out the sudden change in the original time-series but also identified the hidden temporal structure of periodical changes and the overall trend, except for variable *gdppc*. The *gdppc* was a unique exception because the value of observations continued to increase and did not meet the criteria for EMD, so the original time-series was considered the data trend. For other variables, the early IMFs filtered out the abrupt fluctuations, and the IMFs of lower order captured the temporal cyclic properties, which varied from 2-year to 10-year. Though the socioeconomic variables did not share a cyclic pattern, we still determined that the Residuals depicted the trends of change in a cleaner and more elegant manner.

To summarize, although the EMD application did not uncover a consistently universal fitted temporal pattern across heterogeneous variables, the last component of the decomposition, the trend of time-series over the 15-year observation period, which is exactly the focus of this study, was successfully identified for further statistical analysis.

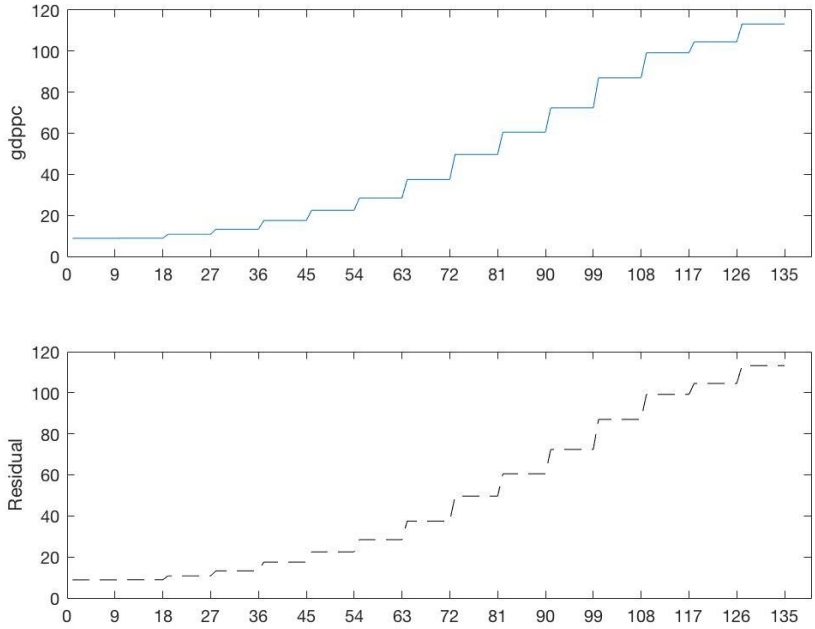

Figure 19. EMD result for *gdppc*

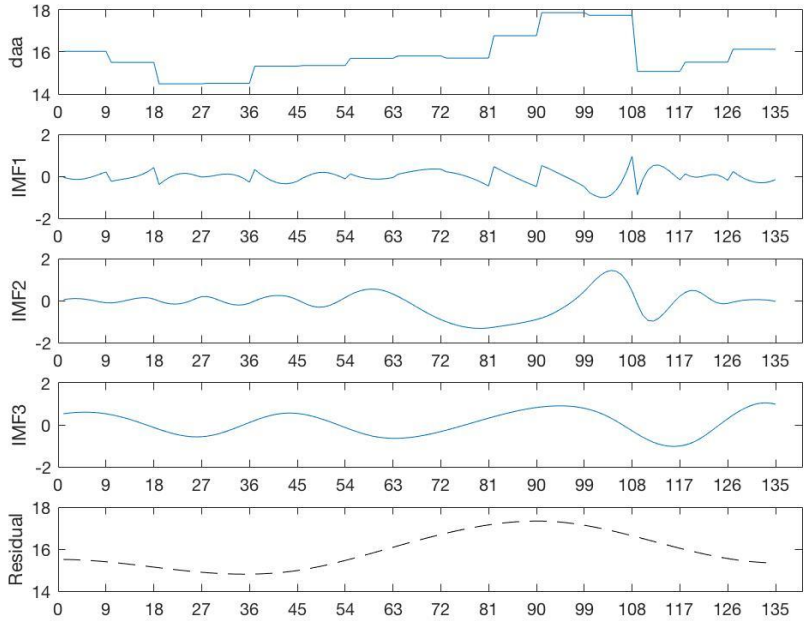

Figure 20. EMD result for *daa*

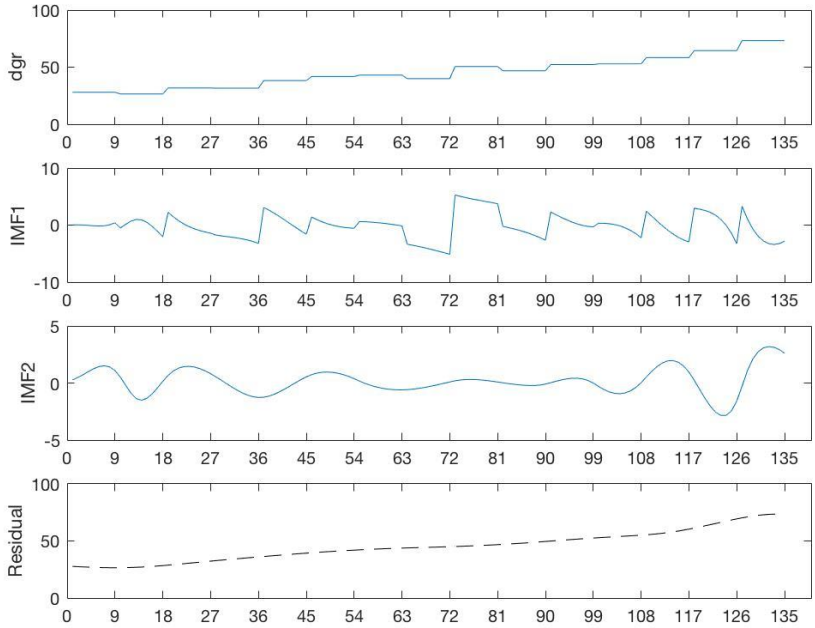

Figure 21. EMD result for *dgr*

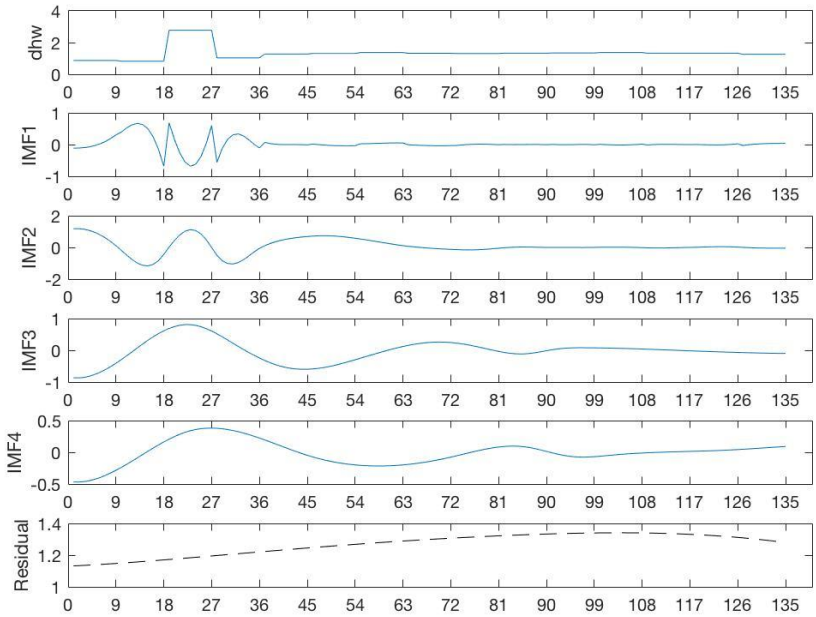

Figure 22. EMD result for *dhw*

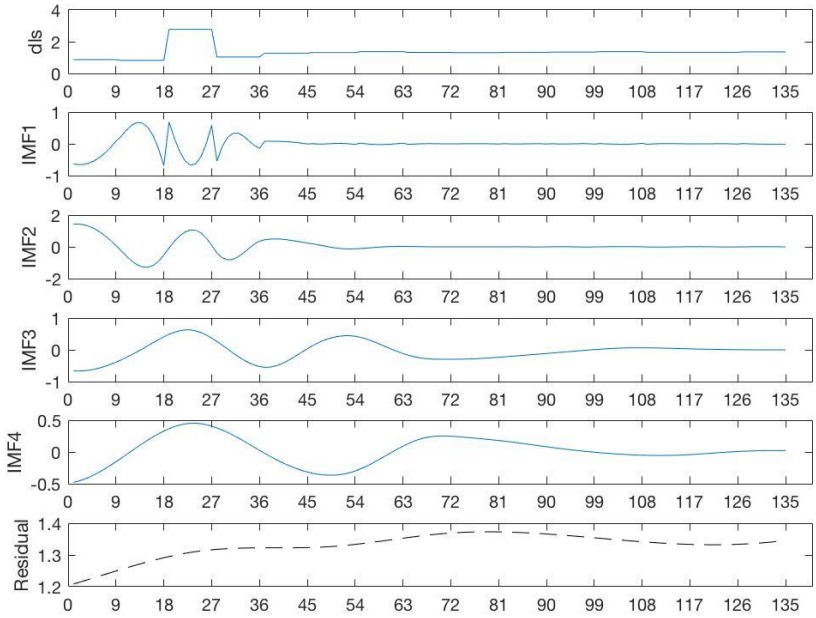

Figure 23. EMD result for *dls*

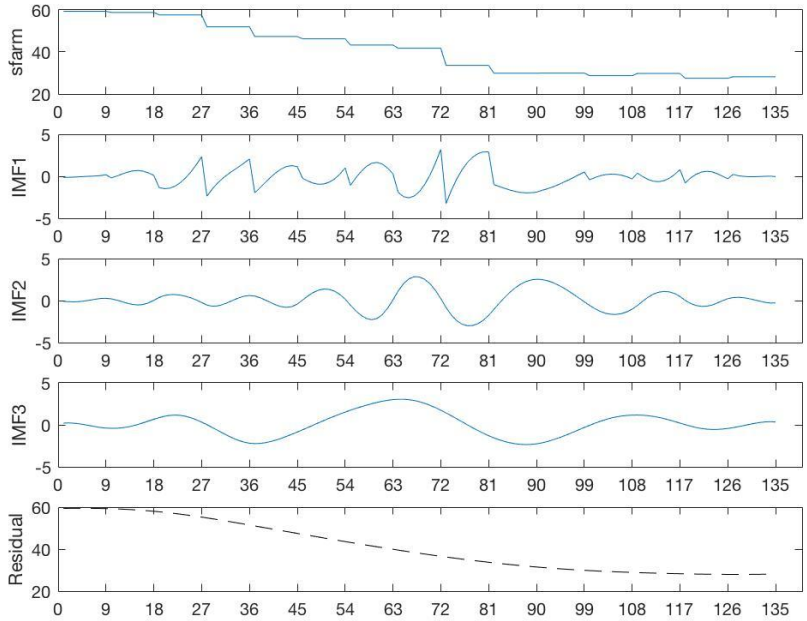

Figure 24. EMD result for *sfarm*

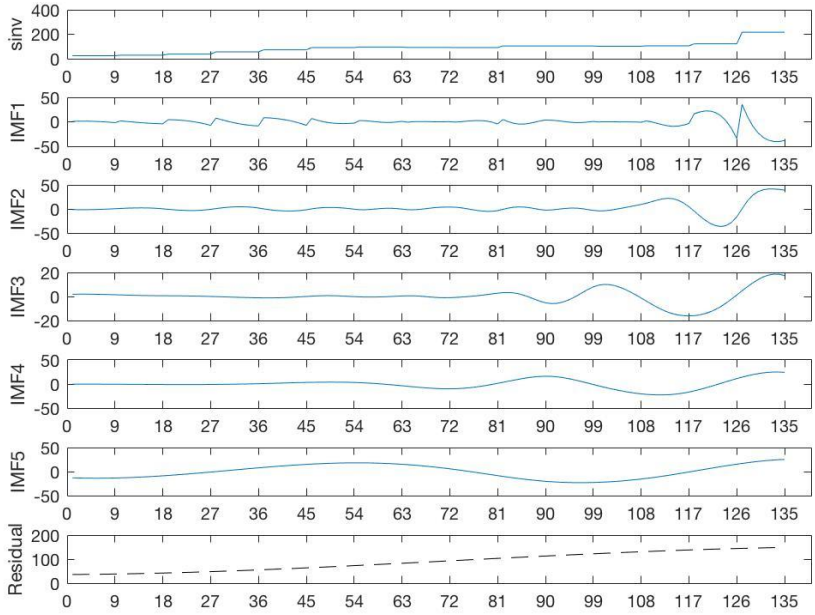

Figure 25. EMD result for *sinv*

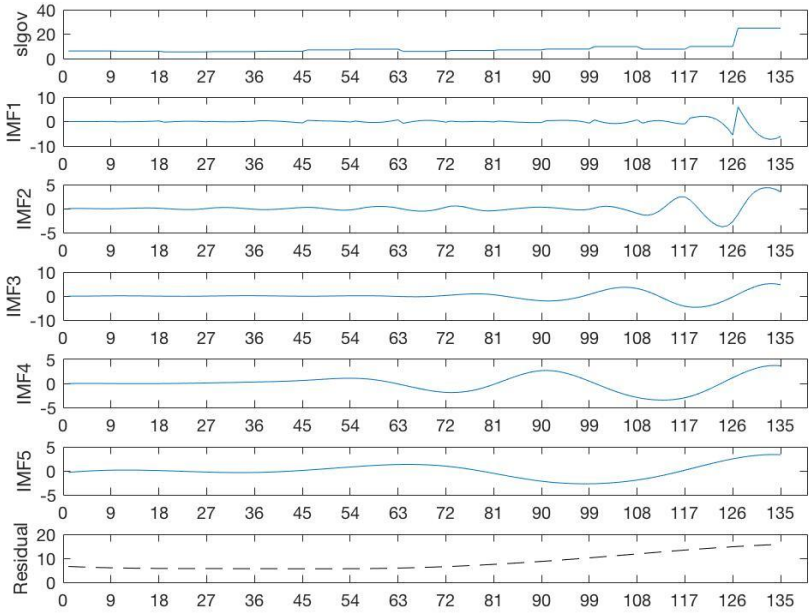

Figure 26. EMD result for *slgov*

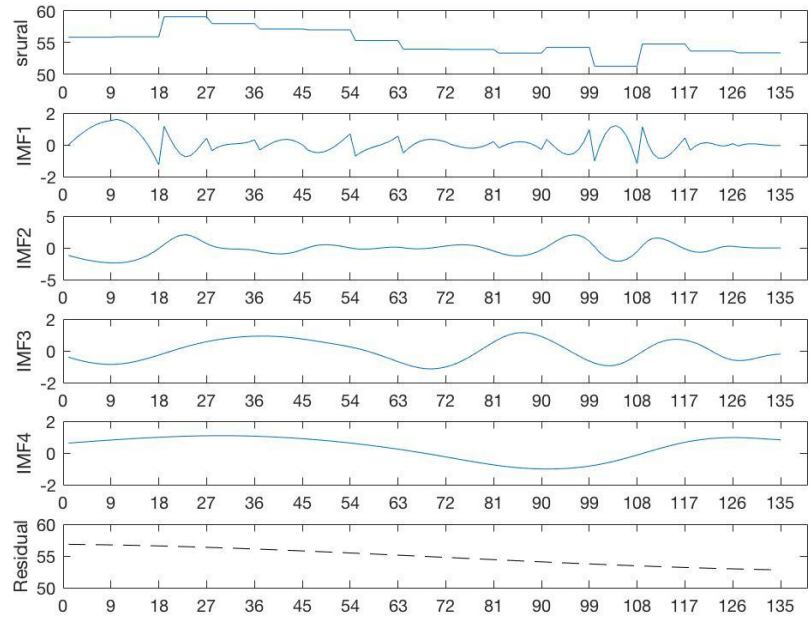

Figure 27. EMD result for *sural*

The Panel Data Analysis

The analysis based on the post-EMD variables is reported in this section. We continue to use the same abbreviations of pre-EMD variables to represent the post-EMD ones, though these two groups are dissimilar in nature. If not explicitly specified, the abbreviations mentioned in this and next sections refer to the post-EMD variables.

Re-visit the descriptive statistics. Our understanding of the post-EMD data begins with its descriptive statistics shown in Table 5. If we compare the results with the pre-EMD descriptive statistics, we observe that the means of the variables were well preserved by the EMD filtering technique, while the standard deviation of the data was largely reduced, which means the distribution of the post-EMD data is more centralized than the distribution of the original data. The filtering effect of EMD was also reflected in the minimal and maximal values as the filter smoothed out the impact of outliers. As we expected, the minimal of post-EMD variables were typically larger and the maximum was smaller than the pre-EMD dataset. Unanticipated exceptions created negative values for certain variables, e.g., precipitation, squared precipitation, squared humidity, livestock density, share of governmental revenue, and share of governmental investment. Apparently, the negative values of these variables are not realistic, but we preserved these values for the panel data analysis because they represent of the longitudinal variation of the variables.

Table 5

Descriptive Statistics for Post-EMD Variables

Variable	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
EVI	12,015	2,471.12	600.23	960.41	4,039.35
EVP	12,015	91.64	44.02	11.85	180.20
PRE	12,015	269.86	155.30	-19.01	789.60
PRS	12,015	9,033.28	69.41	8,789.52	9,408.68
RHU	12,015	51.73	9.66	20.62	73.61
SSD	12,015	88.78	7.59	60.44	108.98
TEM	12,015	187.88	38.52	70.96	257.45
WIN	12,015	25.83	4.56	15.20	40.45
EVP2	12,015	10,710.95	8,675.57	105.24	33,540.51
PRE2	12,015	135,354.40	155,798.20	-60,089.00	1542,157.00
PRS2	12,015	8.18E+07	1341,864.00	7.77E+07	9.02E+07
RHU2	12,015	2,865.32	980.77	-235.83	5,452.63
SSD2	12,015	8,035.64	1,305.34	3,718.16	12,390.08
TEM2	12,015	37,090.21	13,257.57	5,320.06	66,239.29
WIN2	12,015	714.67	262.11	151.60	1,672.37
gdppc	12,015	46.51	35.19	4.29	152.53
daa	12,015	17.16	2.18	10.31	27.73
dgr	12,015	45.06	13.47	14.80	86.79
dls	12,015	1.29	0.37	-1.28	4.89
srural	12,015	54.68	3.49	38.00	72.88
slgov	12,015	8.51	3.85	-2.06	42.02
sinv	12,015	85.62	44.65	-131.83	459.06
sfarm	12,015	43.78	12.90	22.00	100.91

Stationarity testing. As discussed by S. Li et al. (2013), variables of interest for a panel model should be stationary, especially when the dataset contains a large number of periods. Under the definition of stationarity, it is required that the values are time independent. However, it is highly arguable whether our pre-EMD dataset satisfies this requirement. Observed from Figure 11 to Figure 27, most of our time-series were clearly

affected by the period of observation. If we summarize the EVI according to the period of observation within a year, as shown in Figure 28, it is easily discovered that the EVI value follows a reversed U-shape with lower values at the beginning and end of a year and peak values around the sixth period. Based on the temporal signature of the time-series, we conclude that the original time-series were non-stationary and conducted the stationary test only on the post-EMD trends of variables.

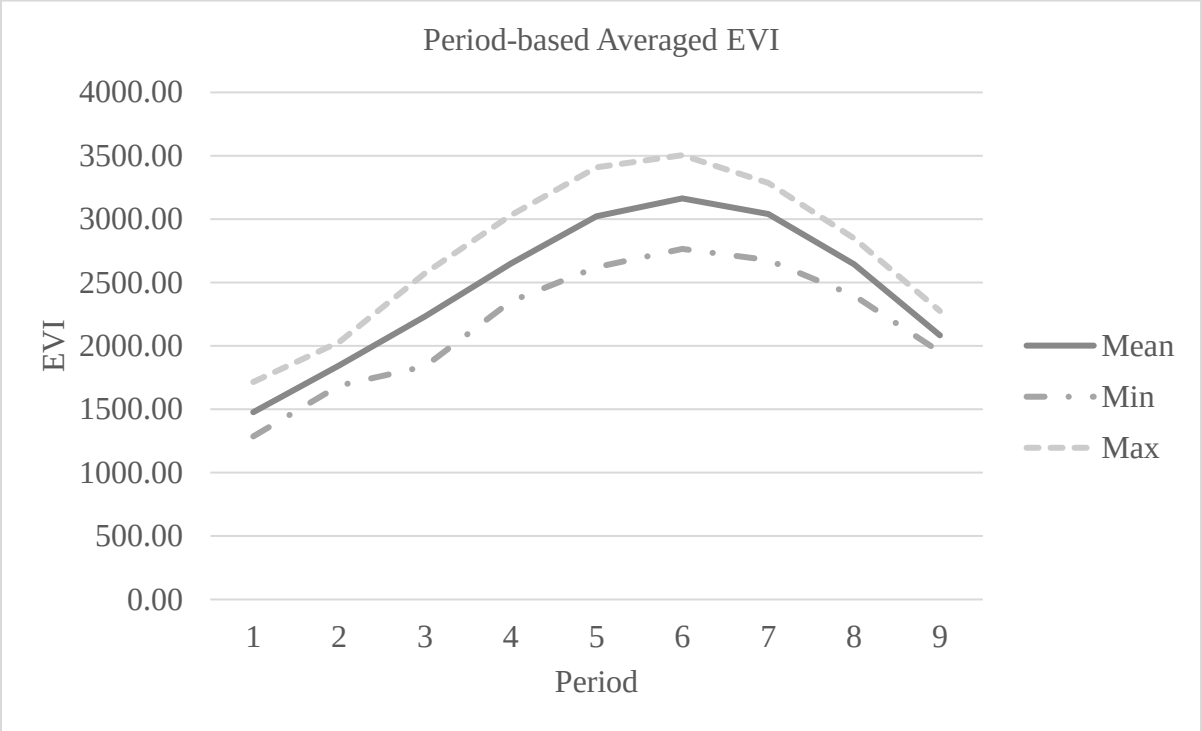

Figure 28. Annual variation of EVI

The unit root test with a low order of difference provides an approach to examine the stationarity of panel data. The concept of unit root test covers a wide range of theories and techniques in detecting the stationarity of time-series. Among various unit test methods, the Levin-Lin-Chu (LLC, 2002) approach has been widely adopted by researchers and their studies. In application, the LLC approach uses t-statistics to test the null hypothesis that the panels contain unit roots against the alternative that the panels are stationary. It is worthwhile

to note that the LLC unit root test is a large sample or asymptotic test that requires the ratio of a number of panels to time periods asymptotically tending to zero, which means the model should have a relatively smaller number of panels and a larger number of time periods. Because our panel data model contains 135 periods and 89 panels, it bears many more periods than panels. Therefore, we argue that the LLC is an appropriate test for this model.

Under the null hypothesis, some panels contain unit roots, and a stochastic process causes the variations, which makes the panels non-stationary, whereas the alternative hypothesis assumes that the panels are stationary. When the degree of freedom exceeds 30, if the t-statistic from LLC test is less than -1.65, it is safe to reject the null hypothesis and favor the alternative one that the variable is stationary at the significance level of 0.05. Given the significance level, the unit test results are compared against the cut-off t-statistic value. In fact, our tests performed extremely well. The results of LLC tests showed the p-value equals zero for all post-EMD variables; thus, we rejected the null hypothesis that all panels contain unit roots and favor the alternative and concluded that our panels are stationary.

Model selection. The stationary test encourages us to include all variables in our panel data model to investigate the causal relationship between the EVI, climatic, and socioeconomic variables using the long-term trends extracted from EMD filtering. For the panel data model, we assumed that, at any given period, the effect of unobservable and omitted variables (e.g., the influence of grassland administrative policies from higher levels of governance) was consistent across all units. Under such assumption, the fixed-effect model should be a more appropriate fit for the data. Before we determined the better model selection, both a fixed-effect and a random-effect specification were employed to the data with all variables and the estimated model parameters are reported in the following Table 6.

In this table, where the coefficient of a variable is reported, the value in parentheses indicates the standard error of the estimation.

Table 6

Comparison of Results from Fixed-Effect and Random-Effect Panel Data Model Specifications

Dependent Variable	Fixed-Effect	Random-Effect
EVP	-0.1989*** (0.0758)	-0.1634** (0.0764)
PRE	0.3044*** (0.0311)	0.3589*** (0.0311)
PRS	-1.6653*** (0.1519)	-1.0319*** (0.1446)
RHU	-1.1663* (0.6017)	-0.2470 (0.6042)
SSD	-1.2411** (0.5171)	-1.2404** (0.5221)
TEM	-1.4895*** (0.1891)	-1.3139*** (0.1897)
WIN	-8.5146*** (0.8115)	-8.4638*** (0.8195)
EVP2	-0.0003 (0.0004)	-0.0003 (0.0004)
PRE2	0.0001** (0.0000)	0.0001** (0.0000)
PRS2	0.0001*** (0.0000)	0.0001*** (0.0000)
RHU2	0.1187*** (0.0066)	0.1120*** (0.0067)
SSD2	-0.0073** (0.0034)	-0.0059* (0.0034)
TEM2	-0.0108*** (0.0007)	-0.0107*** (0.0007)
WIN2	-0.1004*** (0.0150)	-0.1009*** (0.0152)
gdppc	0.0078 (0.0083)	0.0026 (0.0084)
daa	-0.0552*** (0.0037)	-0.0592*** (0.0037)

Table 6 *continued*,

Dependent Variable	Fixed-Effect	Random-Effect
dgr	0.0023*** (0.0002)	0.0023*** (0.0002)
dls	0.0632** (0.0291)	0.0679** (0.0294)
dhw	-0.0394*** (0.0025)	-0.0390*** (0.0025)
srural	-0.2455*** (0.0143)	-0.2424*** (0.0144)
slgov	0.2643*** (0.0262)	0.2557*** (0.0265)
sinv	-0.0723*** (0.0149)	-0.0634*** (0.0150)
sfarm	0.0021 (0.0018)	0.0008 (0.0018)
Intercept	13,380.0700*** (1,004.7540)	5,959.8940*** (822.0747)
R-squared	0.9926	Not Available

(*** Significant at the significance level of 0.01; ** Significant at the significance level of 0.05; * Significant at the significance level of 0.1)

The results are very encouraging. Under the fixed-effect specification, the overall performance of the model was exceptionally good fit for the data. The R-squared value of the fixed-effect model reached 0.9926, which indicates that 99.26% of variations in the EVI are explained by the variations in 23 independent variables across 89 counties and 135 periods. For the random-effect model, due to the limitation of the model itself, a pseudo-R^2 value was calculated during the estimation process but should not be considered an indicator of model fitness. Therefore, it was not reported as a part of the results. If we examine the effects of individual variable, it is clearly indicated that most independent variables impose statistically significant effects of the same direction on the dependent variable EVI despite the minor differences in magnitude between model specifications, except for *RHU*, *EVP2*, *gdppc*, and

sfarm, which are insignificant in both specifications, and *SSD2*, which is insignificant in the random-effect specification.

To systematically examine whether our assumption that the appropriateness of the fixed-effect model specification for this specific data set is valid, we needed a statistical tool to verify our notion. The Hausman test, using chi-square statistics, is a widely accepted approach to help researchers determine the panel data model specification. For a Hausman test, the null hypothesis is stated as the difference in the coefficients is not systematic, which means the difference is caused by the unobserved variables randomly and indicates the random-effect specification is a better choice. The alternative hypothesis considers the difference in coefficients is systematic, which favors the fixed-effect model specification. Using our fixed- and random-effect specifications, the Hausman test returned a chi-square value of 251.01, and this evidence strongly supported the rejection of the null hypothesis and favored the alternative that the fixed-effect model was a more appropriate specification. The selection of model specification supported our assumption that effect varies across time but consistently across units. We used the result from the fixed-effect model specification as the basis to continue our discussion.

First, we focused on the linear effect of climatic variables. At the significance level of 0.05, all climatic variables were significant except for *RHU*, which was significant at the significance level of 0.1. Among these variables, only *PRE* is positively correlated with EVI, and 0.1 mm of increase in accumulated precipitation in a 16-day period will, ceteris paribus, increase the EVI by 0.30. Under the same ceteris paribus assumption, one unit of increase in observations of *EVP*, *PRS*, *RHU*, *SSD*, *TEM*, and *WIN* will cause a decrease in EVI by 0.20,

1.67, 1.17, 1.24, 1.49, and 8.51, respectively. Notably, considering we amplified the value domain of EVI by 10,000, climate change introduces a small impact on the EVI variation.

As for the squared climatic variables, similar to their linear terms, almost all variables significantly affect EVI except *EVP2*. Thus, we excluded *EVP2* from the model. For other polynomial terms, the positive direction of effect indicates that the correlation between the dependent and independent variables follows a U-shape, whereas the negative direction of effect specifies a reversed U-shape. Thus, the EVI value will decline first and then ascend in response to increases in *PRE2*, *PRS2*, or *RHU2*, or increase first and then decrease along with the growth in *SSD2*, *TEM2*, or *WIN2*. However, we argue that these regressors are theoretically valid but could be empirically misleading. Our argument is rooted in the value ranges of variables. Given the actual value range of these post-EMD independent variables, only *PRE2* and *RHU2* followed their reversed and regular U-shapes because of the existence of negative values. Other variables, *PRS2*, *SSD2*, *TEM2*, and *WIN2*, strictly followed a solely descending or ascending curve due to their value range. In reality, the observation of temperature is the only possible scenario allowing for negative values. The selection of an observation window of a year ruled out the only possibility and maintained all observations of temperature above 0 °C. Meanwhile, if we analyze the effect combining the linear and polynomial factors as an integral part, it is noticeable that the sizes of effect are extremely small, especially for *PRE2* and *PRS2*. Arguably, the size of effect may be too small to be included in a statistical analysis. However, due to their excessive domain, the effect still could be substantial. In the end, we kept these polynomial terms, except *EVP2*, in the final model.

The third group of independent variables contained nine socioeconomic factors, and most of these variables were significant with a small size of effect. With one unit of increase in *daa*, *dgr*, *dls*, *dhw*, *sinv*, *slgov*, or *srural*, ceteris paribus, the post-EMD EVI was expected to change by approximately -0.055, 0.002, 0.063, -0.039, -0.072, 0.264, and -0.246, respectively. Unexpectedly, one of the key proxies of economic development, the GDP per capita, was insignificant to the variation of EVI. Though a strong increasing trend, it imposed little statistically significant impact on grassland productivity. A possible explanation is that recent adjustments in policy making and grassland administration have improved the regional economy continuously while simultaneously successfully maintaining grassland sustainability. In the long-term, the grassland ecosystem may be stronger and more adaptive to changes than we thought if humans stop destructive exploitation, especially compared to the short-term observations. The other statistically insignificant socioeconomic variable was *sfarm*. Similar to *gdppc*, but in the opposite direction, the rate of change for *sfarm* was more dramatic than other socioeconomic variables such that the model was unable to correlate the consecutive drop of *sfarm* to the fluctuation of EVI.

Given the significance analysis of variables, we proposed an improved panel data model using only the tested statistically significant variables. By excluding the insignificant variables from the fixed-effect panel data model, our revised regression model would be denoted as

$$EVI_{ti} = f(EVP_{ti}, PRE_{ti}, PRS_{ti}, HUM_{ti}, SSD_{ti}, TEM_{ti}, WIN_{ti},$$
$$PRE_{ti}^2, PRS_{ti}^2, HUM_{ti}^2, SSD_{ti}^2, TEM_{ti}^2, WIN_{ti}^2,$$
$$daa_{ti}, dgr_{ti}, dhw_{ti}, dls_{ti}, sinv_{ti}, slgov_{ti}, srural_{ti}) + \varepsilon_{ti},$$

Analysis of trends. After excluding the insignificant variables from the model, we reduced the number of our model independent variables from 23 to 20, and we used the shortened list of variables as the input to our fixed-effect panel data model. We continued our analysis based on the results from the model reported in Table 7. While the left most column of Table *7* lists the variable names, the second column in this table presents the statistical output from the updated fixed-effect panel data model using post-EMD data. As shown in the last row of the second column in Table *7*, our post-EMD fixed-effect panel data model reached an exceptionally high R-squared value of 0.9926, which means more than 99% of variation in the dependent variable, EVI, can be explained using the variations of the independent variables.

Along with a near-perfect model fitness, our result from the post-EMD panel data model aligns with findings from previous studies (S. Li et al., 2013; Meng et al., 2011; Xie et al., 2016) that climate change influenced the grassland productivity significantly. Almost all climatic dependent variables were statistically significant at the 95% confidence level except *RHU*, and several of those were even significant at the 99% confidence level. Among these linear relationships between climate change and EVI, precipitation is the sole factor that affects EVI positively. If we hold all other variables constant, increasing 0.1 mm of precipitation will likely to cause the EVI rise by 0.2981. It is understandable that in arid and semi-arid regions such as the Inner Mongolia Autonomous Region (IMAR), the amount of precipitation is critical to the prosperity of vegetation. More precipitation usually leads to a higher grassland productivity. Another widely recognized variable that impacts vegetation growth is the temperature, which is negatively correlated with EVI, causing a decrease in vegetation productivity when the temperature rises. Per unit of change in temperature, 0.1 °C,

will likely to cause a change of 1.547 on EVI in the opposite direction, ceteris paribus. Considering the temperature in growing season we selected for this study is above 0 °C, besides the impact from the desert-covered areas on the western side of the IMAR, this coefficient demonstrated how increasing temperature could be detrimental for the grassland productivity. The significant directional impact from precipitation and temperature is one of the similarities that our results share with many previous studies. Besides confirming the causal relationships with precipitation and temperature, our statistical test also presents the significant impact of our control variables. Because of the complexity of the ecological system, evaporation, barometric pressure, relative humidity, sunshine duration, and wind speed are either directly or indirectly interacting with precipitation, temperature, and the growth of grassland. For example, the longer sunshine duration will increase the temperature, intensify the evaporation, reduce the relative humidity, and most likely related to less precipitation. However, there has been a limited number of publication successfully and quantitatively identified the causal relationships between these variables and grassland productivity. If we omit these variables, their effects on EVI will only be partially captured and transferred to alter the coefficients of precipitation, temperature, and the error term, which generates biased estimators for precipitation and temperature. By including the linear terms of *EVP*, *PRS*, *RHU*, *SSD*, and *WIN* as controlling variables, we were able to capture their direct effects on EVI, which reduced the transitional effects and produced more accurate model parameters. Based on our results, the linear term of *EVP*, *PRS*, *RHU*, *SSD*, and *WIN* strongly affect the vegetation growth in Inner Mongolia in a negative direction.

Table 7

Comparison of Statistical Analysis Results from Three Models

Variable	Post-EMD Fixed-Effect Panel	Pre-EMD Fixed-Effect Panel	Post-EMD OLS	Pre-EMD OLS
EVP	-0.2392*** (0.0503)	0.5061*** (0.1475)	2.3206*** (0.1563)	0.5642*** (0.1861)
PRE	0.2981*** (0.0307)	0.0960*** (0.0314)	-0.0526 (0.1184)	0. 3821*** (0.0813)
PRS	-1.6663*** (0.1519)	-0.2112 (0.1362)	7.5989*** (0.4592)	24.6657*** (1.0757)
RHU	-1.0376* (0.5956)	-22.6243*** (2.5430)	4.5785** (2.1923)	-18.1280*** (3.3126)
SSD	-1.2603** (0.5166)	28.6335*** (3.4358)	2.9161 (2.0951)	29.3569*** (4.6736)
TEM	-1.5470*** (0.1853)	3.9065*** (0.6577)	-2.5075*** (0.6969)	0.9686 (0.8734)
WIN	-8.3353*** (0.7732)	-73.2550*** (5.0298)	-12.5174*** (3.3265)	-44.9123*** (6.2773)
PRE2	0.0001** (0.0000)	*Omitted*	0.0005*** (0.0001)	-0.0001* (0.0001)
PRS2	0.0001*** (0.0000)	*Omitted*	-0.0004*** (0.0000)	-0.0013*** (0.0001)
RHU2	0.1183*** (0.0066)	0.4828*** (0.0258)	0.4456*** (0.0232)	0.6160*** (0.0333)
SSD2	-0.0076** (0.0034)	-0.1257*** (0.0201)	-0.0904*** (0.0129)	-0.0832*** (0.0273)
TEM2	-0.0108*** (0.0007)	0.0011 (0.0019)	-0.0133*** (0.0022)	-0.0035 (0.0025)
WIN2	-0.1035*** (0.0145)	0.6859*** (0.0832)	-0.4740*** (0.0581)	0.2383** (0.1059)
daa	-0.0553*** (0.0037)	-7.5449*** (1.0897)	-0.0663*** (0.0155)	10.1204*** (0.6839)
dgr	0.0023*** (0.0002)	2.2224*** (0.1998)	0.0048*** (0.0007)	1.8289*** (0.1522)
dls	0.0935*** (0.0099)	2.6142** (1.0278)	0.2616*** (0.0438)	1.0470 (1.3066)
dhw	-0.0393*** (0.0025)	0.0421* (0.0252)	0.0591*** (0.0078)	-0.1852*** (0.0199)
srural	-0.2467*** (0.0142)	-0.7552 (0.5602)	0.3155*** (0.0585)	0.1519 (0.2688)
slgov	0.2628*** (0.0261)	-0.3793 (0.3602)	0.0655 (0.1202)	0.8213* (0.4334)

Table 7 *continued*,

Variable	Post-EMD Fixed-effect Panel	Pre-EMD Fixed-effect Panel	Post-EMD OLS	Pre-EMD OLS
sinv	-0.0714*** (0.0148)	0.1896*** (0.0718)	-0.1263*** (0.0419)	-0.0807 (0.0737)
Intercept	13,358.9500*** (1,003.7260)	3,222.7080*** (1,265.5750)	-34,848.6400*** (2,085.1720)	-115,449.300*** (4,909.037)
R-squared	0.9926	0.8100	0.8260	0.6567

(*** Significant at the significance level of 0.01; ** Significant at the significance level of 0.05; * Significant at the significance level of 0.1)

Discussed by other researchers (Shi et al., 2013), the correlation between vegetation growth and climate change is usually more complicated than merely linear. Our statistical analysis supported this claim by demonstrating the significant impacts from the square terms of climatic dependent variables. However, though the impacts were significant, the sizes were much smaller than their linear terms as the largest magnitude barely exceeded 0.1, and most of them were around or below 0.01, while Xie et al. (2016) found the sizes of impact from polynomial terms were similar to the linear ones. Another characteristic of our results is that the polynomial relationships are strictly linear within the domain of dependent variables, in either the positive or negative direction. Unlike previous research, the effect of threshold of the variables, or the U-shape of a relationship, is not reflected in our analysis. We argue that our selection of sampling period, which only focused on the growing seasons instead of covering the growing and non-growing seasons of a whole year, and the value domains of the dependent and independent variables together mitigated the polynomial impacts of climatic variables. Thus, similar to the linear terms of *EVP*, *PRS*, *RHU*, *SSD*, and *WIN*, the included polynomial terms are essential and they perform more as controlling variables than

explanatory in our model to help us retrieve more accurate estimators for precipitation and temperature.

From the socioeconomic perspective, all seven variables included in the estimation exerted strong impacts on the EVI. As one of the proxies for the grassland productivity, the arable area density (*daa*) was negatively correlated with EVI, which indicated that more land used for farming caused lower grassland productivity. On the contrary, another variable measuring a different aspect of crop farming, the density of grain production (*dgr*), positively affected EVI with a greater size of effect. The results seem controversial; however, they explained the relationship between farming activities and grassland productivity well. While *daa* represented the land use density of counties, it is noticeable that the farmland is almost bare at the beginning and end of a growing season, and the bare land tends to carry a lower EVI than surfaces covered by vegetation. Apparently, our sampled EVI was heavily influenced by the bare land negatively, which means counties with a higher concentration of crops farming land use will likely to have lower EVIs. The other variable, *dgr*, represented the actual grain production. When grain production is higher, the growth of grain plants is healthier and greener. Therefore, higher values in *dgr* would be likely related to higher EVI values. For this pair of variables measuring the farming activities, while *daa* captures the effects introduced by the level of concentration of farmland, *dgr* provides a more truthful observation of grassland productivity, which makes them both legit controlling variables for farming activities.

The variable *dls* was used to capture the impact from grazing activities, and our test indicated that counties with higher livestock density usually had higher EVI values. It was easily understandable that a high amount of livestock could be raised in fertile grassland, and

only richer grasslands could support a greater scale of grazing activities. Our results also showed that the share of rural population (*srural*) negatively correlated with EVI. It could be interpreted that when there is more rural population, more animals will be raised, which will then cause heavier grazing activities and a severer grassland degradation and thus lower EVI values. This finding quantitatively supplements the claim from many previous research that the intensity of grazing is one of the major causes of fluctuations in grassland productivity. Meanwhile, it also indicates that the density of livestock should not be accounted for the degradation of grassland alone. The concentration of population, the rural population supporting grazing activities specifically, is also playing a critical role in affecting grassland productivity.

From the perspective of impacts introduced by policies, we found that the density of transportation (*dhw*) was negatively correlated with EVI, which implied that the transportation network divided the grassland into smaller patches thus increased the separation and decreased the diversity of grassland species. Meanwhile, the increased ratio of local government revenue to a county's GDP (*slgov*) trended to result in higher EVIs. We assume that the counties with more revenue perform better financially than the ones that obtain less, and they should possess more resources that could be distributed in order to maintain higher grassland productivity. In comparison to the governmental revenue, the higher ratio of investment in fixtures to a county's GDP (*sinv*) caused lower EVIs, while we expected greater grassland productivities as a result of more governmental investments. Theoretically, improvements to the transportation network would contribute to the economic development of a region, and the growth of the regional economy will in turn stimulate new establishments of the infrastructure. However, we were not able to construct a relationship

matrix between *dhw*, *slgov*, and *sinv* to address the dynamics. Given the evidence presented, we think this result is a reflectance of the socialist policy of the IMAR. When the central government allocates the investment, it is distributed equally to counties. The poorer counties, usually located near the desert region of the southeastern IMAR with a low or rare vegetation coverage, have a smaller amount of GDP and relative lower EVI. This explained why the share of investment was negatively impacting the EVI.

Though not considered a part of our final model selection, we think the two variables that were excluded due to their insignificancy, *gdppc* and *sfarm*, are still worth discussion. As discussed in previous sections, *gdppc* is the sole variable used to capture the economic status of a county. Referring to the EMD filtering results of EVI and *gdppc* demonstrated in Figure 11 and Figure 19, respectively, we found that, instead of fluctuating around certain value throughout the observed period as EVI, the shape of *gdppc* maintained a steady increasing trend, which indicated that the residents of the IMAR have been obtaining more financial resources at their disposal throughout these years. This steadily improving personal financial status of the IMAR residents is largely due to the rapidly developing economy of China. The weak relationship between *gdppc* and EVI implied that the improving financial situation of the IMAR residents is not directly related to grassland growth, which led to the exclusion of *gdppc* from the final model. The other socioeconomic variable excluded from the model, *sfarm*, represents the ratio of all agricultural income to the GDP. Unlike *gdppc* revealing a steady increasing trend, the data series of *sfarm* kept a decreasing trend throughout our study, as shown in Figure 24. On average, the percentage of agricultural income to the overall GDP was almost 60% at the beginning of the series, and it had reduced to approximately 30% near the end. The lowered ratio indicated that the socioeconomic

structure of the counties in the IMAR had shifted as a result of national economic development. More sources of income other than agriculture, e.g., commerce, trading, manufacturing, and even industry processing agricultural products, are contributing to the economy of the IMAR with increasing proportions.

To demonstrate the effectiveness of EMD filtering, we compared the coefficients obtained from the fixed-effect model using both pre- and post-EMD data to the results from the most often used regression model, ordinary least square (OLS) model. Starting from column three to column five of Table *7*, we illustrated the coefficients from pre-EMD fixed-effect panel data model, post-EMD OLS model, and pre-EMD OLS model using the optimized variables, respectively. Similar to how the statistical results were reported in the previous sections, the value in the parentheses indicates the standard error of the coefficient.

Observed from Table *7*, along with differences in the significance of coefficients, the size and standard error of a coefficient varied between models. Based on the cross-sectional and longitudinal characteristics of the data, it is certain that the panel data model was a more appropriate specification than the OLS regression. Though we may occasionally find that certain independent variables in the OLS models are significant at a lower significance level than in the panel model, e.g., *RHU*, *SSD*, and *SSD2*, these unexpected occurrences simply indicate that the OLS specifications generate more wrong estimators simply because they are inaccurate selections. Thus, we favored the panel model over the OLS model. Next, in a comparison between the two panel models, the one using post-EMD data outperformed the one using pre-EMD data by obtaining coefficients with smaller magnitudes and standard errors. Meanwhile, it is easily identified that, though the sizes of effect were limited for *PRE2* and *PRS2*, the fixed-effect panel data model using the post-EMD data captured the

significant effect of these two variables, while the same data model using pre-EMD data excluded these two variables due to high collinearity. At the same time, the pre-EMD model concluded *TEM2* was insignificant whereas the post-EMD model not only included this significant variable into EVI estimation but also corrected the size and direction of the effect. Hence, we were further convinced that the EMD filtering technique helped us understand the long-term dynamics despite the repetitive patterns that reside in the daily observations, and apparently, the post-EMD fixed-effect panel model claimed more variables significantly affected our dependent variable than any other models. Additionally, the R-squared values from these regression models supported our conclusion as well, and the differences between R-squared values of models demonstrated the advantage of model specification. While the fixed-effect panel data model using post-EMD data exhibited a near perfect R-squared value, all other three models generated considerably lower R-squared values with the pre-EMD OLS model generating the lowest R-squared value of 0.6567. All of this evidence confirmed our claim that the combination of EMD and panel data model provided superior performance in estimating the model parameters.

In addition to listing the model coefficients, our fixed-effect model included an F-test to examine the existence of individual effect of our observations. Under the null hypothesis, the observed and unobserved fixed effects are equal to zero, which leads to the conclusion that the individual effect is equal across all units. Our test result indicated $F(88, 11{,}908) = 3{,}106.50$ with the probability of 0.0000, providing strong evidence to reject the null hypothesis and favor the alternative that the fixed-effects are non-zero. Meanwhile, we used a modified Wald test to examine the heteroscedasticity of our data. In this test, the null hypothesis is stated as the data is homoscedastic, which means the variance is constant,

whereas the alternative hypothesis considers the data is heteroscedastic, which in turn will lead to the conclusion that the variability of the dependent variable varies as the values of independent variables increase. The chi-square test result from 89 degrees of freedom returned the probability of zero, providing really strong evidence for us to reject the null hypothesis and favor the alternative to conclude that our data contain heteroscedasticity.

After the comparison between models, we took a step further to analyze the impact from climatic polynomial terms alone and demonstrated the comparison in Table 8. Based on the post-EMD fixed-effect model, we re-examined the model fitness and variable coefficients while excluding those polynomial terms of climatic variables. Under such scenario, the simplified model without polynomial terms was explaining 99.22% of variations in EVI using our independent variables, which is only 0.04% less than the model containing the polynomial terms. From the perspective of individual independent variables, though the coefficients for socioeconomic variables did not change much, the coefficients for linear climatic variables differed dramatically from the previous model. The direction of effect for *RHU* was reversed from negative to positive, and the size of its effect became almost five times of what it was; the magnitudes of effect for *EVP*, *PRE*, and *TEM* almost doubled; the effect for *SSD* and *WIN* increased almost 25%; and *PRS* was the only climatic variable with a smaller size of effect than its coefficient from the post-EMD fixed-effect model with polynomial terms, which decreased by more than 75%. This comparison revealed that though the magnitude of the coefficients for polynomial climatic variables were limited, they captured and explained the variations in our dependent variables; when excluded from the model, the linear terms of the corresponding variables were accounted for the nonlinear

effects on EVI. These polynomial terms acted as control variables to help us understand the causal relationships between our variables.

Table 8

Impact of Polynomial Climatic Variables in Fixed-Effect Model

Variable	Post-EMD Fixed-effect Panel with Polynomial Terms (omitted in comparison)	Post-EMD Fixed-effect Panel without Polynomial Terms
EVP	-0.2392*** (0.0503)	-0.5649*** (0.0487)
PRE	0.2981*** (0.0307)	0.5281*** (0.0220)
PRS	-1.6663*** (0.1519)	-0.3117*** (0.1014)
RHU	-1.0376* (0.5956)	4.9928*** (0.4750)
SSD	-1.2603** (0.5166)	-1.5341*** (0.3627)
TEM	-1.5470*** (0.1853)	-2.7398*** (0.1769)
WIN	-8.3353*** (0.7732)	-12.9816*** (0.3936)
daa	-0.0553*** (0.0037)	-0.0486*** (0.0037)
dgr	0.0023*** (0.0002)	0.0018*** (0.0002)
dls	0.0935*** (0.0099)	0.0655*** (0.0100)
dhw	-0.0393*** (0.0025)	-0.0423*** (0.0025)
srural	-0.2467*** (0.0142)	-0.2336*** (0.0145)
slgov	0.2628*** (0.0261)	0.2314*** (0.0265)
sinv	-0.0714*** (0.0148)	-0.0845*** (0.0149)
Intercept	13,358.9500*** (1,003.7260)	5,864.101***(929.2121)
R-squared	0.9926	0.9922

To summarize, the preprocessing of data using the EMD filtering technique broke down the original data into several exclusive temporal components, and the fixed-effect panel data model successfully revealed the relationship between the long-term variation

trends of the dependent and independent variables of the coupled human-nature interaction in Inner Mongolia.

Spatial variation. While the coefficients from the fixed-effect panel data model described how different factors impact the EVI values over time quantitatively, the model itself was not able to project the cross-sectional variations between heterogeneous counties. To quantitatively determine how counties are different from each other, we included a set of dummy variables to capture the unique intercept of each county in the panel data model and used a map to qualitatively interpret the spatial properties of our model.

The following map (Figure 29) demonstrates how the county-based EVI values were spatially distributed in the IMAR. In this map, we arbitrarily selected County 89 as the baseline, which is identified on the map with a red boundary, and computed the difference between County 89 and all other counties. Then, we colorized the differences using a red-to-green color ramp, where the darkest red indicates the lowest EVI and the darkest green indicates highest. The spatial pattern is very clear: the EVI value gradually increases from west to east and from south to north, with an exception of counties near the central IMAR. In fact, this unique spatial pattern is not hard to explain if we take into account the land cover and environment of the IMAR. While the northeastern part of the IMAR is primarily covered by forest, the majority land cover type in the western regions of the IMAR is bare land and grassland suffering desertification. Most pastoral grasslands are concentrated in the central region, where most counties are colored as yellow and light orange on our map. The counties that form a horizontal patch across the west-central IMAR seem less affected by the land cover types. However, after we overlap our map with the national hydrological system, we noticed that this patch perfectly aligned with the path of the Yellow River, which is the

second-longest river in Asia and the sixth-longest river system in the world and nourishes 140 million people. Meanwhile, the belt of major cities of the IMAR coincidentally falls within these counties. Apparently, EVI values from the counties along the path of the Yellow River are not only benefiting from the proximity to the water resource, but are also influenced by the socioeconomic development of the region, thus gaining higher values than their neighbors.

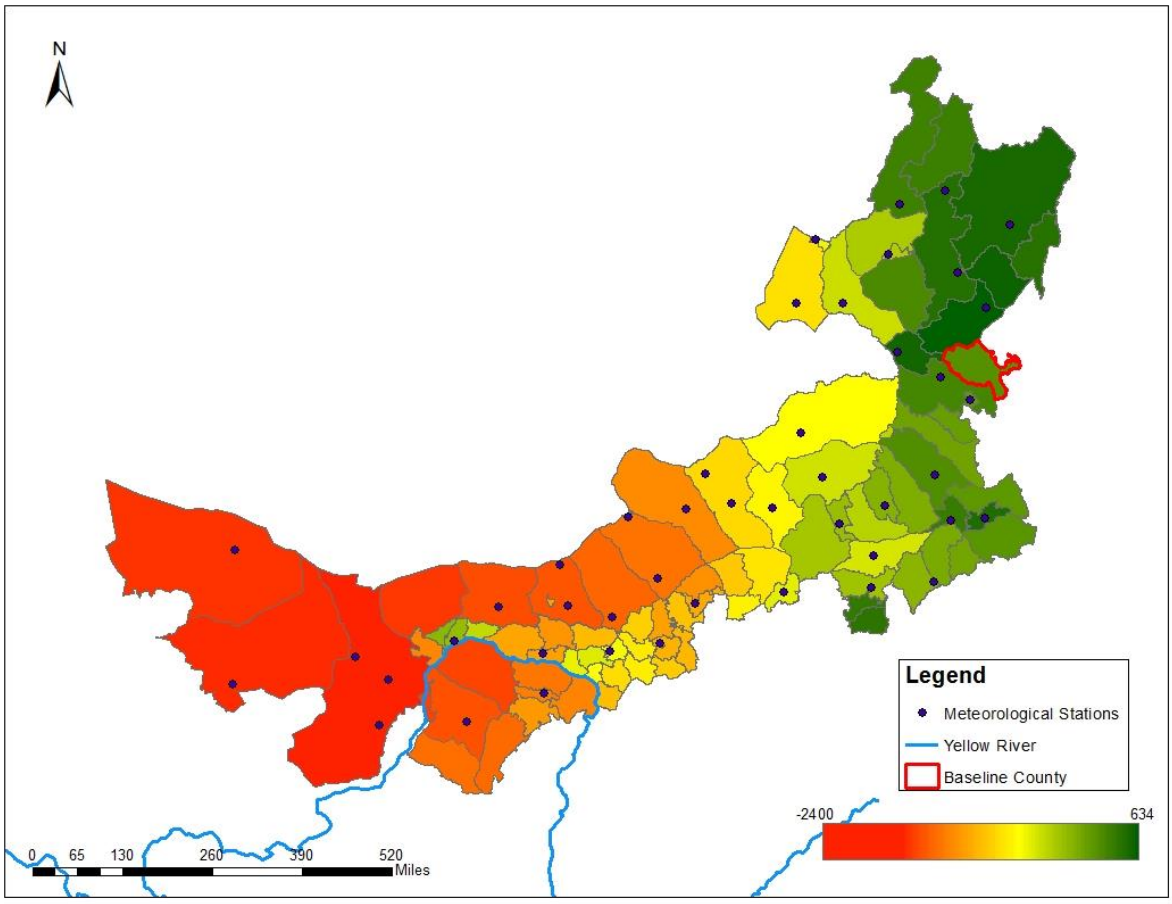

Figure 29. Spatial distribution of heterogeneity among counties

Chapter 5. Conclusion

This final chapter aims to consolidate findings presented in previous chapters to conclude our discussion, which may help identify future contributions to the body of knowledge. In the following sections, we summarize our findings, provide an overview of the methodologies and techniques, and suggest possible directions for future research.

Summary of Findings

In the past decades, research on the coupled human-nature system has been trending, and by the increasing power of analytic tools and methods, the complexity of the articulated mechanism between the participants has been gradually revealed. In this study, we successfully broke down the variations of observations into long-term trends and recursive patterns in an innovative approach by applying the empirical mode decomposition (EMD) filter and analyzed the cross-sectional and temporal impact of climate change and socioeconomic development on the Inner Mongolia Autonomous Region (IMAR) grassland productivity over 15 years through a panel data model.

To reveal how climatic and socioeconomic factors jointly affected the IMAR grassland productivity in the long-term, we collected 14 climatic variables from 45 ground stations, enhanced vegetation index (EVI) as the proxy for grassland productivity, and nine socioeconomic variables for 89 counties across the entire IMAR region for 15 years. Given the variation in collection frequency and spatial resolution, we interpolated and extrapolated the samples to create a comprehensive data set consisting of 288,360 observations over 135 periods for 89 counties. Our preliminary examination of the data identified that the time-series of variables shared a periodic pattern, which seemed both inter- and intra-annual. We argued that the temporal similarity of the time-series caused high collinearity between

variables and was likely to alter the underlying long-term trend, thus leading to a biased statistical conclusion, and our suspicion was recognized by many other researchers (Duffy, 2004; Onema & Taigbenu, 2009; Vicent et al., 2010). To isolate the long-term tendency of the time-series from the original observations, we used EMD to adaptively filter out the repetitive temporal patterns.

While the longitudinal observations of grassland productivity, climatic variables, and socioeconomic factors of the IMAR supplied a comprehensive snapshot of the dynamics between nature and human, the richness and complexity of this data set overwhelmed the capability of traditional statistical models. Ranging from the simple linear regression model to the advanced structural equation model, traditional statistical models either compress the temporal variability to focus on the differences between individuals or address the changes in time but ignore how the changes vary among samples. On the other hand, the panel data model provides an ideal solution such that the causal relationships between the dependent and independent variables of a cross-sectional and longitudinal data set can be captured simultaneously. The F-test of our panel data model strongly supported our assumption that the individuals were heteroskedastic and should not be pooled as samples from the same population and analyzed indiscriminately.

The two variants of the panel data model specification, the fixed-effect and random-effect model, are designed to accommodate whether the effects of independent variables are consistent across all units. It was one of our assumptions that all IMAR counties are equally affected by the variables, both observed and unobserved (e.g., the policies). We set up a hypothesis under such assumption that the fixed-effect model would fit our data better, and we used the Hausman test to test our hypothesis. The test result strongly supported our

hypothesis, which made the fixed-effect model specification of the panel data model a more appropriate setting for our study.

Based on our data processing results and model selection, we discovered that grassland productivity had been increasing during our study period, and the panel data model using the tendencies of the independent variables explained the trend of EVI very well, as the fitness of the model, the R-squared value, reached over 0.99. From the perspective of variables, our finding confirmed the significant long-term impact from climatic variables. We also confirmed that precipitation and temperature are the two major factors influencing the grassland productivity as discussed in many studies. Almost all 20 independent variables included in our model demonstrated significant effect on our dependent variable at the significance level of 0.05 and were included in the final model specification, and the climatic effect on the grassland was found to be two-fold, both linear and polynomial.

All linear terms significantly affected the grassland productivity at a significance level of 0.05, and most of them were even significant at the significance level of 0.01. Among these linear terms, only precipitation (*PRE*) was found to positively affect the EVI while all other variables (evaporation [*EVP*], barometric pressure [*PRS*], relative humidity [*RHU*], sunshine duration [*SSD*], temperature [*TEM*], and wind speed [*WIN*]*)* impacted the grassland productivity in the opposite direction. While *WIN* and *EVP* presented the largest and smallest sizes of effect, respectively, we argued that the magnitude of the effect could be altered by the value ranges of the variables and that comparing magnitudes of different variables hardly leads to convincing conclusions. Instead trying to explain how each climatic variable affects the grassland productivity, we only focused on the precipitation and temperature, and used *EVP*, *PRS*, *RHU*, *SSD*, and *WIN* as controlling variables to reduce the

transitional effect to *PRE* and *TEM*. Apparently, the grassland would benefit from more precipitation and lower temperature in the growing seasons.

The polynomial terms were introduced into the model to capture the U-shaped effect as described by C. Li et al. (2013). As reported in Table 8, all squared terms of the climatic variables, except squared evaporation (*EVP2*), showed significant influence on the EVI. The presence of the statistically significant polynomial coefficients qualitatively identified the nonlinear causal relationship. However, the sizes of the effects of these coefficients are limited. We also observed that the effect of a variable on EVI was either strictly negative or positive within the value range of the samples, which leaded to the conclusion that the U-shaped effect exists among the climatic variables, but the thresholds for changing direction of effects are beyond the value range of the data in our empirical study. We also found that while the linear relationships played a dominant role affecting the grassland productivity, the polynomial terms of the climatic variables also significantly contributed to the variations of grassland productivity. To qualitatively assess the impacts of polynomial terms, we revisited the post-EMD fixed-effect model without these variables. While the overall model fitness varied little, we noticed significant changes in the coefficients of the linear climatic variables and insignificant changes in the socioeconomic variables comparing to the full post-EMD fixed-effect panel data model. The difference caused by excluding the polynomial terms indicated that the U-shaped effects have significant impacts on EVI and these effects could be and should be controlled in statistical analysis. If not, the polynomial effects would still be captured by the linear variables and the panel data model would remain effective and efficient. However, the estimated coefficients of the variables could be less accurate and biased. This comparison told us that omitting the polynomial terms would lead to less

accurate estimation of the coefficients for climatic variables and convinced us that it is the best practice to keep them as controlling variables to reduce errors of our model. It indicates that it would be optimal to include the polynomial terms of climatic variables in analysis of the interactions between climate and ecology as it helps reduce the transitional impacts on the linear terms, especially when the number of independent variables is limited.

To measure the socioeconomic effect, we collected the IMAR's annual statistics from 2000 to 2014. Then we further calculated the normalized variables (gross domestic product [GPD] per capita [*gdppc*], density of arable area [*daa*], density of grain production [*dgr*], density of livestock [*dls*], share of rural population [*srural*], local government revenue as share of GDP [*slgov*], governmental investment as share of GDP [*sinv*], and all agricultural income as share of GDP [*sfarm*]) to mitigate the influence from the various sizes of counties. By using the EMD filtering and panel data model, we had interesting findings. Though we expected an individual's economic status would be an important factor in grassland productivity variations, our statistics results demonstrated otherwise. The variable we chose to represent the financial situation of the local population, *gdppc*, was recognized as insignificant in the model, which was largely due to its steady and steeply increasing tendency. While it has been recognized that the economy of the IMAR is boosted dramatically, the rapid rate of change on the financial status of the residents is not statistically significant to the variations of grassland productivity. Another variable measuring the percentage of farming industries to the GDP, *sfarm*, was not significant either and was excluded from the model as well. This means that the improved and diversified economic structure of the IMAR reduced the proportion of agricultural income in GDP, which made the share of agricultural income statistically insignificant affecting the grassland

productivity. Though our model downplayed the importance of economy to grassland productivity, we observed significant relationships between grassland and other variables. Among these variables, we found that *slgov* and *srural* imposed the largest positive and negative effect, respectively, indicating that a larger proportion of local governmental revenue to GDP will likely increase the grassland productivity while a higher percentage of rural population may lower grassland productivity. Besides these two variables, we also found *daa*, *dhw*, and *sinv* negatively affected EVI while the dependent variable was positively influenced by *dgr* and *dls*. A higher concentration of arable land led to lower grassland productivity due to the annual rotation of bare land, but a higher concentration of grain production and livestock density indicated the grassland is more productive, which means the government should encourage farms with higher production efficiency and direct research to explore methods and technology to improve unit productivity. The construction of a transportation network reduced the grassland productivity, as it not only took up the space for vegetation growth but also caused species separation.

Last but not least, we presented each county's performance according to our model through a map. In general, counties on the east side of the IMAR maintained higher EVI, and the proximity to the Yellow River seemed to increase grassland productivity.

Implications

The purpose of this study is to better understand the coupled effects of climate change and socioeconomic activities on grassland productivity. Our statistical model found significant relationships between the dependent and independent variables, and these findings discussed in the previous sections led to valuable practical implications as described below:

- All climatic variables, except *RHU*, were found significantly affecting grassland productivity in a linear fashion at the significance level of 0.05. Because of the complex interactions between the climatic variables, the climatic variables except precipitation and temperature were kept in the data model as controlling variables to reduce their transitional effects on our dependent variable, EVI.
- We confirmed that the precipitation has a positive effect on grassland productivity. It implies that increasing precipitation through irrigation for smaller scales and rainmaking for larger regions could be effective approaches to improve grassland productivity and such actions should be repeated regularly to establish a long-term effect on the ecosystem. Meanwhile, the practices of increasing precipitation should only target counties yielding high productivity instead of exhausting the precious energy and resources regardless of the land cover and vegetation types.
- We also confirmed the temperature, another commonly recognized factor that affects the grassland ecological system, has a negative effect on the grassland productivity. This indicated that the grassland is vulnerable to increasing temperature and it produces more biomass when the temperature is comparably lower. The IMAR, situated in arid and semi-arid zones with hot growing seasons, makes its grasslands sensitive to global warming as the rising temperature deteriorates the grassland productivity. This finding demonstrated how the long-term climate change is related to the economic and ecological system and provided strong support for the campaign of fighting against the global warming.

- Besides the linear impact, all included climatic variables, except *EVP*, were also influencing the grassland productivity through a polynomial relationship. The sizes of these polynomial relationships were much smaller than their linear counterparts. However, they played indispensable roles in describing the causal relationships between climate change and grassland growth.
- While the ratio of the arable area was negatively affecting grassland productivity, the higher level of grain production concentration and density of livestock helped improve grassland productivity. It suggests farming practices with higher efficiency of more grain production using less land are beneficial to grassland productivity.
- The measurement of the density of transportation networks of a county was negatively affecting the grassland productivity. Despite the potential economic incentives a transportation network could bring to a region, our observations in this study indicated that more roads led to lower grassland productivity. Our explanation for this phenomenon is that these manmade features dissect the grassland into smaller patches, which reduced the biodiversity of the grassland ecosystem and made it more vulnerable.
- Counties with more resources contributed to the grassland tend to have a higher grassland productivity in return, and evenly distributed governmental investment did not seem to be beneficial in grassland management.
- As a result of the rapid economic development of China, the income level and financial status of the IMAR residents had been increasing during our observations, and this proxy of financial status we used in this study was asserted

insignificantly causing variations in the grassland productivity. The IMAR residents are getting richer without exhausting the grassland ecology.

- The importance of agricultural industry to the economy of the IMAR had been decreasing over time, which excluded the percentage of agricultural income from the factors that significantly influenced grassland productivity in the IMAR.
- The application of EMD prior to the panel data model proved that it effectively isolated the long-term trends of variables by filtering out the periodic patterns, and this data processing technique improved the model fitness in our comparison of results from the pre- and post-EMD analysis.
- In a comparison of statistical models, our chosen panel data model outperformed the ordinary least square regression model in both the significance of coefficients and the fitness of the model, which means the panel data model is more suitable for cross-sectional and time-series dataset.

Contribution

This study quantitatively investigated the causal relationships between grassland productivity, climate change, and socioeconomic activities. After excluding some insignificant variables from the panel data model, all independent variables were significantly affecting our dependent variable, the EVI, at the significance level of 0.05. Several tests were conducted to examine the stationarity and heterogeneity of the data, the suitability of variants for a panel data model, and the hypotheses of this research. We found significant evidence to support the coupled effect of the human-nature system through an innovative approach by integrating the EMD filtering and the panel data modeling in our

statistical analysis, and we are contributing to the body of knowledge in the following perspectives:

- The polynomial terms of climatic variables were found to be significant affecting the linear causal relationships between climatic variables and grassland productivity. Though the direct effects on grassland productivity were limited, we recommend that future research should always keep the polynomial terms in the model estimation because they could reduce the bias of the estimators and improve the accuracy of their linear counterparts.
- The usage of the EMD filter helped us to identify the period patterns and isolate the long-term trends of the variables. The application of the filter in our study was highly efficient and proved its adaptiveness. When the trend extraction is necessary for a longitudinal study but the researchers only have limited background information about the disturbances buried in the time-series, the adaptive EMD could be an appropriate selection for data processing.
- The R-squared value of our panel data model reached an exceptionally high level of 99.26%, which is a great enhancement to the model's fitness, meaning that almost all variability in the EVI were explained by the variations in climatic and socioeconomic variables. We partially attribute the enhancement to the introduction of EMD in data processing. Meanwhile, the panel data model using a larger sample size with more variables together helped to explain the coupled interactions. This study fully demonstrated the potential of the EMD transformation, the panel data model, and their integration, and researchers could

use our approach of data collection, processing, and analysis as a reference for their future ecological studies.

- In terms of grassland management, the government should encourage higher efficiency farming practices by providing more incentives and promoting more investments into research that aims to improve the farming efficiency because of the important role it plays in increasing grassland productivity. Meanwhile, decision makers should consider concentrating limited resources on regions with denser farming activities and greater grassland coverages. For regions with a low grassland productivity, it might be optimal to reduce the farming and grazing intensity, or even fully suspend these activities and migrate the residents to flourish regions to improve production efficiency and restore the grassland ecological system.
- Since the construction of a transportation network reduces grassland productivity, corresponding agencies should prioritize grassland sustainability by either reconsidering the feasibility of a plan that causes patches of grassland or planning alternative migration paths for animals to maintain the biodiversity of the region at the early stage of project designs.
- Traditionally, researchers claimed that the urge to pursue more financial incentives had been the primary reason of unsustainable grassland consumption. Our results challenged the applicability of this claim in the early twenty-first century by providing the evidence that the improved financial status is independent of the fluctuation of grassland productivity.

Overall, our study provided a quantitative interpretation of the causal relationships between climate change, socioeconomic development, and grassland productivity of the IMAR counties and helped us to understand their coupled interactions. Its uniqueness not only relies on the usage of the largest dataset for the IMAR ecological analysis with the greatest spatial coverage, the most number of variables, and the longest period of study, but also because it is the first application of integrating EMD transformation and panel data models. We hope this study provided valuable pieces of evidence to assist future grassland ecological studies and support long-term grassland management strategy by helping governmental officials make informed decisions to maintain a sustainable grassland ecological system.

Assessment of Research

As mentioned in the "Significance" section of this dissertation, this study endeavors a unique approach of integrating interdisciplinary methodologies to understand how climate change and socioeconomic activities of the IMAR affected the grassland productivity in 15 years using data with finer resolutions. The assessment begins with an evaluation of the data we used in this study. This dataset consists of daily observation of climatic variations, grassland productivity for every 16-day period, and annual statistics of socioeconomic characteristics of all counties of the IMAR during the growing seasons for 15 years. While maintaining the granularity of the data, our database also provided extensive coverage in both spatial and temporal dimensions. The vast spatial extent, the substantial duration of observations, the number of variables included, and the complexity introduced by these heterogeneous variables of this study have all surpassed its predecessors (S. Li et al., 2013; Xie et al., 2016). Considering the representativeness of our database, we concluded that this

study could be viewed as the foundation, or a great source of reference, for future studies of the IMAR ecological systems. Despite the thoroughness of our database, it is worth pointing out that none of the datasets were collected from restricted data sources; they were all obtained from publicly accessible data sources or services. In another word, this database can be reproduced by experienced personnel, and databases to be built to evaluate ecological phenomenon using the similar metrics in the same area, or even in different regions around the globe, could be populated following the same data collection procedure. Raw data acquired are barely meaningful without processing and analysis, especially when the data are collected in various frequency by different agencies. We have no intention of emphasizing the data acquisition, but we want to highlight the steps we took to prepare the raw data for analysis, which includes integrating heterogeneous data acquired from different agencies and sources, normalizing them into the same spatial and temporal unit, and eliminating the disturbance caused by shared periodic patterns. From this perspective, this study demonstrates a data acquisition approach of compiling a big data using open data for long-term studies of the coupled human-nature system, and we believe that the open data access is more valuable than restricted datasets because they make the research repeatable and universal, which is welcomed and encouraged by the body of knowledge. Considering the limited value range and fluctuations due to the scope of our study, we are eager to see whether the significance level and the magnitudes of the polynomial terms increase when the sampling period expands to annual coverage instead of only the growing seasons.

From the perspective of data processing, due to the dissimilarities in both spatial and temporal scales between the raw observations of EVI, climatic, and socioeconomic variables, we employed different preprocessing techniques to consolidate these variables into the same

resolution and created a balanced panel data set. To minimize the impact of missing values and maintain the integrity of the balanced panel data, we used linear interpolation and extrapolation using the neighboring values to fill in for a particular variable, year, or county. To build the panel data set per county and period, we averaged the pixel-based EVI values of the county and used an inversed distance weighted interpolation to estimate the characteristics of the ground station-based climatic factors for counties that do not have ground stations. Though there are variants of spatial interpolation methods available, we argue that distance is the most critical factor determining the value at any unobserved location. Given the data availability, we are confident that we have applied the most available and reliable data processing techniques. On the other hand, we also believe the spatial interpolation would yield more convincing results if more geographic data could be integrated, for example, the digital elevation model.

Though both the panel data model or the EMD filter has been exploited in research individually, there has been a gap of integrating these practices into one application to reveal their capabilities. This study took advantages of the panel data model, which helped us systematically analyze the cross-sectional and time-series variables while controlling the unobserved and omitted variables. We hypothesized that the system errors would vary over time but be consistent across sections at any given time, and we used the Hausman test to test our hypothesis. The test result strongly supported our hypothesis, which made the fixed-effect model specification of the panel data model a more appropriate setting for our study. While the panel data model provides an analytic framework for data with both cross-sectional and longitudinal properties, the highly adaptive EMD filter supplements an important component to eliminate the periodic patterns. This parameter-free tendency

extraction method provides a convenient way for researchers with less engineering experience or background knowledge to decompose complex time-series. The introduction of EMD into data processing helped us to separate the repetitive, shared, annual, and interannual patterns among the observations so that our study could focus on the long-term trend of changes of our variables. In our comparison, it was clearly demonstrated that the panel data model using EMD normalized data outperformed the one using non-normalized data with a significantly higher overall model fitness. Although previous studies (C. Li et al., 2013; S. Li et al., 2013) using panel data models had revealed satisfying R-squared values, the application of EMD prior to the panel data model pushed the model's fitness to a higher level. Our results indicate that the EMD filtering is a very effective method isolating the trend of a time-series, and this filtering technique can be easily applied in application to help researchers eliminate periodical patterns as long as the prerequisites are satisfied. However, as discussed in the methodological review in previous chapters, the flexibility of EMD becomes a two-edged sword that its adaptiveness helps users to decompose the time-series into intrinsic mode functions (IMFs) while it increases the difficulties to interpret the IMFs.

The comparison of analysis results between several regression models demonstrated how the combination of EMD and panel data model enhanced our ability to create a model with more comprehensive explanations of the relationships between variables. Among the post-EMD fixed-effect panel, pre-EMD fixed-effect panel, post-EMD OLS, and pre-EMD OLS models, the post-EMD fixed-effect panel model yielded the highest R-squared value, which indicated that the independent variables in the selected model explained the variation in the dependent variable better than other candidates. Meanwhile, more independent variables in the post-EMD fixed-effect panel model were statistically significantly affecting

our dependent variable, usually at higher significance levels with smaller sizes of effect though.

Additionally, to visually understand the spatial heterogeneity between different counties of our model, we used a color ramp to symbolize the intercepts of counties. By presenting the quantitative characteristics of counties qualitatively, we were able to interpret the spatial distribution pattern of how both the ecosystem and the economy affect the grassland.

Future Directions and Lessons Learned

Admittedly, this research has its limitations and drawbacks. As mentioned in the limitations section of Chapter 1, one of the limitations is the capacity of EVI to represent grassland growth. Ideally, to capture accurately the grassland progression over years, EVI values should be recorded according to individual species of grassland vegetation within any given geographic region. However, given the vast study area and the biodiversity of the grassland ecosystem, it is not feasible to include too many dummy variables in statistical analysis. Alternatively, this study decided to focus on summarized EVI values based on the geographic location. In the future, how different species of grassland vegetation respond to the joint effect of climate change and human activities could be investigated.

Another limitation, lacking proper data validation, resides in the data preprocessing procedure. First of all, though the data interpolating and extrapolating techniques we applied to create the panel data set and handle the inconsistency of various data sources were widely recognized and easily approachable, and the cross-validation for climatic variables proved the validity of our estimation, it is noticeable that we skipped data validation for socioeconomic variables. This was largely due to the minimal percentage of missing data in

the socioeconomic data set and the fact of insufficient referencing data to compare with our estimates. Besides arguments surrounding how we handled the missing data, the ability to evaluate the accuracy of the results from EMD filters remains a challenge. The adaptiveness of EMD undoubtedly simplifies the filtering process significantly, but lack of control during the process brings uncertainties to the amplitude, frequency, and number of IMFs. Even for the isolated temporal patterns, the first IMF from the original time-series is not guaranteed to represent the annual cycle of change. As a result, the discussion about the cyclic pattern of variables was limited to visual interpretations of the generalized observations due to the absence of an applicable strategy in analyzing the IMFs. These issues may possibly be addressed in future studies.

Overall, this research staged an empirical, interdisciplinary application by applying techniques from electrical engineering and econometrics into geospatial analysis. This endeavor may help us to understand the complexity of coupled human-nature interactions and provide guidance for researchers who need to tackle heterogeneous datasets, which happens more often as analytic models evolve and the body of knowledge expands.

Reference

Aaheim, A., Amundsen, H., Dokken, T., & Wei, T. (2012). Impacts and adaptation to climate change in European economies. *Global Environmental Change*, *22*(4), 959-968.

Akiyama, T., & Kawamura, K. (2007). Grassland degradation in China: Methods of monitoring, management and restoration. *Grassland Science, 53*(1), 1-17. doi:10.1111/j.1744-697X.2007.00073.x

Allen, J. B., & Rabiner, L. R. (1977). A unified approach to short-time Fourier analysis and synthesis. *Proceedings of the IEEE*, *65*(11), 1558-1564.

Anaya, J. A., Chuvieco, E., & Palacios-Orueta, A. (2009). Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management, 257*(4), 1237-1246. doi:10.1016/j.foreco.2008.11.016

Angerer, J., Han, G., Fujisaki, I., & Havstad, K. (2008). Climate change and ecosystems of Asia with emphasis on Inner Mongolia and Mongolia. *Rangelands*, *30*(3), 46-51.

Arhonditsis, G. B., Stow, C. A., Steinberg, L. J., Kenney, M. A., Lathrop, R. C., McBride, S. J., & Reckhow, K. H. (2006). Exploring ecological patterns with structural equation modeling and Bayesian analysis. *Ecological Modelling*, *192*(3), 385-409.

Arndt, C., Asante, F., & Thurlow, J. (2015). Implications of climate change for Ghana's economy. *Sustainability*, *7*(6), 7214-7231.

Baho, D. L., Futter, M. N., Johnson, R. K., & Angeler, D. G. (2015). Assessing temporal scales and patterns in time series: Comparing methods based on redundancy analysis. *Ecological Complexity, 22*, 162-168.

Bai, Y., Wu, J., Xing, Q., Pan, Q., Huang, J., Yang, D., & Han, X. (2008). Primary production and rain use efficiency across a precipitation gradient on the Mongolia

plateau. *Ecology*, *89*(8), 2140-2153.

Barthlott, C., Drobinski, P., Fesquet, C., Dubos, T., & Pietras, C. (2007). Long-term study of coherent structures in the atmospheric surface layer. *Boundary-layer Meteorology*, *125*(1), 1-24.

Berryman, A. A. (1989). The conceptual foundations of ecological dynamics. *Bulletin of the Ecological Society of America*, *70*(4), 230-236.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F., & Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research: Atmospheres*, *111*(D12).

Chen, J., John, R., Shao, C., Fan, Y., Zhang, Y., Amarjargal, A., ... & Dong, G. (2015). Policy shifts influence the functional changes of the CNH systems on the Mongolian plateau. *Environmental Research Letters*, *10*(8), 085003.

Chin, W. W. (1998). Issues and opinion on structural equation modeling. *Management Information Systems Quarterly, 22*(1), vii-xvi.

Chuai, X. W., Huang, X. J., Wang, W. J., & Bao, G. (2013). NDVI, temperature and precipitation changes and their relationships with different vegetation types during 1998–2007 in Inner Mongolia, China. *International Journal of Climatology*, *33*(7), 1696-1706.

Crabbe, M. J. C. (2008). Climate change, global warming and coral reefs: Modelling the effects of temperature. *Computational Biology and Chemistry*, *32*(5), 311-314.

Cutter, S. L., Golledge, R., & Graf, W. L. (2002). The big questions in geography. *The Professional Geographer*, *54*(3), 305-317.

De Beurs, K. M., & Henebry, G. M. (2005). A statistical framework for the analysis of long

image time series. *International Journal of Remote Sensing*, *26*(8), 1551-1573.

Deng, X., Huang, J., Huang, Q., Rozelle, S., & Gibson, J. (2011). Do roads lead to grassland degradation or restoration? A case study in Inner Mongolia, China. *Environment and Development Economics, 16*(6), 751-773. doi:http://dx.doi.org/10.1017/S1355770X11000180

Dougherty, C. (2011). *Introduction to econometrics*. New York, NY: Oxford University Press.

Duffy, D. G. (2004). The application of Hilbert–Huang transforms to meteorological datasets. *Journal of Atmospheric and Oceanic Technology*, *21*(4), 599-611.

Environmental Systems Research Institute, Inc. (Esri). (2017, November 6). Performing cross-validation and validation. Retrieved from http://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/performing-cross-validation-and-validation.htm

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 39-50.

Fornell, C., & Larcker, D. F. (1987). A second generation of multivariate analysis: Classification of methods and implications for marketing research. *Review of Marketing*, *51*, 407-450.

Fu, L., Bo, T., Du, G., & Zheng, X. (2012). Modeling the responses of grassland vegetation coverage to grazing disturbance in an alpine meadow. *Ecological Modelling, 247*, 221-232.

Galvin, K. A., Thornton, P. K., Boone, R. B., & Sunderland, J. (2004). Climate variability

and impacts on East African livestock herders: the Maasai of Ngorongoro Conservation Area, Tanzania. *African Journal of Range and Forage Science*, *21*(3), 183-189.

Ghil, M., Allen, M.R., Dettinger, M.D., Ide, K., Kondrashov, D., Mann, M.E.,…Yiou, P. (2002). Advanced spectral methods for climatic time series. *Reviews of Geophysics*, *40*(1).

Gill, A., Polley, W., Johnson, B., Anderson, J., Maherali, H., & Jackson, B. (2002). Nonlinear grassland responses to past and future atmospheric CO_2. *Nature*, *417*(6886), 279-282.

Hagerman, S., Dowlatabadi, H., Satterfield, T., & McDaniels, T. (2010). Expert views on biodiversity conservation in an era of climate change. *Global Environmental Change*, *20*(1), 192-207.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, *46*(6), 1251-1271.

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage.

Hsiao, C. (2007). Panel data analysis—advantages and challenges. *Test*, *16*(1), 1-22.

Hsiao, C. (2014). *Analysis of panel data* (3rd ed.). New York, NY: Cambridge University Press.

Huang, N. E., & Wu, Z. (2008). A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, *46*(2).

Huang, N. E., Shen, Z., Long, R., Wu, C., Shih, H., Zheng, Q.,…Liu, H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-

stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *454*(1971), 903-995.

Huang, N. E., & Attoh-Okine, N. O. (Eds.). (2005). *The Hilbert-Huang transform in engineering*. Boca Raton, FL: CRC Press.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1), 195-213.

Hull, V., Tuanmu, M. N., & Liu, J. (2015). Synthesis of human-nature feedbacks. *Ecology and Society*, *20*(3), 17.

Huss-Danell, K., Chaia, E., & Carlsson, G. (2007). N2 fixation and nitrogen allocation to above and below ground plant parts in red clover-grasslands. *Plant and Soil*, *299*(1-2), 215-226.

Institute of Botany, Mongolia. (2011). Statistics of grassland quality in Mongolia. Institute of Botany (IOB), Ulaanbaatar, Mongolia.

Jiapaer, G., Liang, S., Yi, Q., & Liu, J. (2015). Vegetation dynamics and responses to recent climate change in Xinjiang using leaf area index as an indicator. *Ecological Indicators, 58*, 64-76. doi:10.1016/j.ecolind.2015.05.036

Kyselý, J., Beguería, S., Beranová, R., Gaál, L., & López-Moreno, J. (2012). Different patterns of climate change scenarios for short-term and multi-day precipitation extremes in the Mediterranean. *Global and Planetary Change, 98*, 63-72.

Labat, D. (2005). Recent advances in wavelet analyses: Part 1. A review of concepts. *Journal of Hydrology*, *314*(1), 275-288.

Lafortezza, R., Tanentzap, A. J., Elia, M., John, R., Sanesi, G., & Chen, J. (2015).

Prioritizing fuel management in urban interfaces threatened by wildfires. *Ecological Indicators, 48*, 342-347.

Lamb, E. G., Mengersen, K. L., Stewart, K. J., Attanayake, U., & Siciliano, S. D. (2014). Spatially explicit structural equation modeling. *Ecology*, *95*(9), 2434-2442.

Kaufman, Y. J., & Tanre, D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE transactions on Geoscience and Remote Sensing*, *30*(2), 261-270.

Leedy, P., & Ormrod, J. (2010). *Practical research: Planning and design* (9th ed.). Upper Saddle River, NJ: Pearson.

LeSage, J. P. (2014). Spatial econometric panel data model specification: A Bayesian approach. *Spatial Statistics*, *9*, 122-145.

Levin, A., Lin, C. F., & Chu, C. S. J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics*, *108*(1), 1-24.

Li, C., Kuang, Y., Huang, N., & Zhang, C. (2013). The long-term relationship between population growth and vegetation cover: An empirical analysis based on the panel data of 21 cities in Guangdong Province, China. *International Journal of Environmental Research and Public Health*, *10*(2), 660-677.

Li, J., Cui, Y., Liu, J., Shi, W., & Qin, Y. (2013). Estimation and analysis of net primary productivity by integrating MODIS remote sensing data with a light use efficiency model. *Ecological Modelling, 252*, 3-10. doi:10.1016/j.ecolmodel.2012.11.026

Li, S., Xie, Y., Brown, D., Bai, Y., Hua, J., & Judd, K. (2013). Spatial variability of the adaptation of grassland vegetation to climatic change in Inner Mongolia of China. *Applied Geography, 43*, 1-12. doi:10.1016/j.apgeog.2013.05.008

Li, Z., Liu, S., Tan, Z., Bliss, N., Young, C., West, T., & Ogle, S. (2014). Comparing cropland net primary production estimates from inventory, a satellite-based model, and a process-based model in the Midwest of the United States. *Ecological Modelling, 277*, 1-12. doi:10.1016/j.ecolmodel.2014.01.012

Lim, Y., Cai, M., Kalnay, E., & Zhou, L. (2005). Observational evidence of sensitivity of surface climate changes to land types and urbanization. *Geophysical Research Letters, 32*(22).

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E.,…Tayler, W. (2007). Complexity of coupled human and natural systems. *Science, 317*(5844), 1513-1516.

Liu, Y., & Xie, Y. (2013). Measuring the dragging effect of natural resources on economic growth: Evidence from a space–time panel filter modeling in China. *Annals of the Association of American Geographers*, *103*(6), 1539-1551.

Lu, N., Wilske, B., Ni, J., John, R., & Chen, J. (2009). Climate change in Inner Mongolia from 1955 to 2005—Trends at regional, biome and local scales. *Environmental Research Letters,4*, 045006. doi:10.1088/1748-9326/4/4/045006

Malaeb, Z. A., Summers, J. K., & Pugesek, B. H. (2000). Using structural equation modeling to investigate relationships among ecological variables. *Environmental and Ecological Statistics*, *7*(1), 93-111.

Marin, A. (2010). Riders under storms: Contributions of nomadic herders' observations to analysing climate change in Mongolia. *Global Environmental Change*, *20*(1), 162-176.

Martínez, B., & Gilabert, M. A. (2009). Vegetation dynamics from NDVI time series analysis using the wavelet transform. *Remote Sensing of Environment*, *113*(9), 1823-

1842.

Meng, M., Ni, J., & Zong, M. (2011). Impacts of changes in climate variability on regional vegetation in china: NDVI-based analysis from 1982 to 2000. *Ecological Research, 26*(2), 421-428. doi:10.1007/s11284-011-0801-z

Mertes, C. M., Schneider, A., Sulla-Menashe, D., Tatem, A. J., & Tan, B. (2015). Detecting change in urban areas at continental scales with MODIS data. *Remote Sensing of Environment, 158,* 331-347.

Moser, S. C. (2010). Now more than ever: The need for more societally relevant research on vulnerability and adaptation to climate change. *Applied Geography, 30*(4), 464-474.

Naidu, C., Durgalakshmi, K., Satyanarayana, G., Rao, L., Ramakrishna, S., Mohan, J., & Ratna, K. (2011). An observational evidence of climate change during global warming era. *Global and Planetary Change, 79*(1), 11-19.

NASA. (2016, December 17). *What Are Climate and Climate Change?* Retrieved from https://www.nasa.gov/audience/forstudents/5-8/features/nasa-knows/what-is-climate-change-58.html

Onema, J. M. K., & Taigbenu, A. (2009). NDVI–rainfall relationship in the Semliki watershed of the equatorial Nile. *Physics and Chemistry of the Earth, Parts A/B/C, 34*(13), 711-721.

Peng, Z. K., Peter, W. T., & Chu, F. L. (2005). An improved Hilbert–Huang transform and its application in vibration signal analysis. *Journal of Sound and Vibration, 286*(1), 187-205.

Philip, G. M., & Watson, D. F. (1982). A precise method for determining contoured surfaces. *Australian Petroleum Exploration Association Journal 22,* 205–212.

Piao, S., Fang, J., Zhou, L., Ciais, P., & Zhu, B. (2006). Variations in satellite-derived phenology in China's temperate vegetation. *Global Change Biology*, *12*(4), 672-685.

Pickup, M. (2015). *Introduction to time series analysis*. Los Angeles, CA: SAGE.

Piras, M., Mascaro, G., Deidda, R., & Vivoni, E. R. (2016). Impacts of climate change on precipitation and discharge extremes through the use of statistical downscaling approaches in a Mediterranean basin. *Science of the Total Environment*, *543*, 952-964.

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., & Solomon, A. M. (1992). Special paper: a global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, 117-134.

Qi, B. (2001). Xilingol Stockbreeding Record. *Inner Mongolian Renmin Press*, Huhhot (in Chinese).

Qiu, B., Zeng, C., Tang, Z., & Chen, C. (2013). Characterizing spatiotemporal non-stationarity in vegetation dynamics in China using MODIS EVI dataset. *Environmental Monitoring and Assessment*, *185*(11), 9019-9035.

Rao, A. R., & Hsu, E. C. (2008). *Hilbert-Huang transform analysis of hydrological and environmental time series* (Vol. 60). Dordrecht, Netherlands: Springer Science & Business Media.

Reddy, M. J., & Adarsh, S. (2016). Time–frequency characterization of sub-divisional scale seasonal rainfall in India using the Hilbert–Huang transform. *Stochastic Environmental Research and Risk Assessment*, *30*(4), 1063-1085.

Ren, H., Zhou, G., & Zhang, X. (2011). Estimation of green aboveground biomass of desert steppe in Inner Mongolia based on red-edge reflectance curve area

method. *Biosystems Engineering, 109*(4), 385-395. doi:10.1016/j.biosystemseng.2011.05.004

Ribalaygua, J., Pino, M., Pórtoles, J., Roldán, E., Gaitán, E., Chinarro, D., & Torres, L. (2013). Climate change scenarios for temperature and precipitation in Aragón (Spain). *Science of the Total Environment, 463*, 1015-1030.

Richard, Y., & Poccard, I. (1998). A statistical study of NDVI sensitivity to seasonal and interannual rainfall variations in Southern Africa. *International Journal of Remote Sensing*, *19*(15), 2907-2920.

Rilling, G. (2017, January 1). *Empirical mode decomposition*. Retrieved from http://perso.ens-lyon.fr/patrick.flandrin/emd.html

Schweiger, E. W., Grace, J. B., Cooper, D., Bobowski, B., & Britten, M. (2016). Using structural equation modeling to link human activities to wetland ecological integrity. *Ecosphere*, *7*(11).

Scurlock, M., Johnson, K., & Olson, J. (2002). Estimating net primary productivity from grassland biomass dynamics measurements. *Global Change Biology, 8*(8), 736-753.

Seaquist, J. W., Olsson, L., & Ardö, J. (2003). A remote sensing-based primary production model for grassland biomes. *Ecological Modelling, 169*(1), 131-155. doi:10.1016/S0304-3800(03)00267-9

Shen, M., Chen, J., Zhu, X., Tang, Y., & Chen, X. (2010). Do flowers affect biomass estimate accuracy from NDVI and EVI? *International Journal of Remote Sensing, 31*(8), 2139-2149. doi:10.1080/01431160903578812

Shen, S. S., Shu, T., Huang, N. E., Wu, Z., North, G. R., Karl, T. R., & Easterling, D. R. (2005). HHT analysis of the nonlinear and non-stationary annual cycle of daily

surface air temperature data. *Hilbert-Huang Transform and its Applications, 5,* 187-209.

Shi, W., Tao, F., & Zhang, Z. (2013). A review on statistical models for identifying climate contributions to crop yields. *Journal of Geographical Sciences, 23*(3), 567-576.

Shiyomi, M., Akiyama, T., Wang, S., Hori, Y., Chen, Z., Yasuda,…Yamamura, Y. (2011). A grassland ecosystem model of the Xilingol steppe, Inner Mongolia, China. *Ecological Modelling, 222*(13), 2073-2083.

Shukla, J., Nobre, C., & Sellers, P. (1990). Amazon deforestation and climate change. *Science (Washington), 247*(4948), 1322-1325.

Sjöström, M., Ardö, J., Arneth, A., Boulain, N., Cappelaere, B., Eklundh, L.,…Scholes, R.J. (2011). Exploring the potential of MODIS EVI for modeling gross primary production across African ecosystems. *Remote Sensing of Environment, 115*(4), 1081-1089. doi:http://dx.doi.org/10.1016/j.rse.2010.12.013

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., & Miller, H. L. (2007). *Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change, 2007*. Retrieved from https://www.ipcc.ch/publications_and_data/ar4/wg1/en/contents.html

Son, T., Chen, F., Chen, R., Minh, Q., & Trung, H. (2014). A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation. *Agricultural and Forest Meteorology, 197,* 52-64. doi:10.1016/j.agrformet.2014.06.007

Squires, V., Hua, L., Zhang, D., & Li, G. (Eds.). (2010). *Towards sustainable use of rangelands in North-West China*. Dordrecht, Netherlands: Springer Science &

Business Media.

Squires, V. R. (2009). *Rangeland degradation and recovery in China's pastoral lands*. Cabi.

Tong, S. T., Sun, Y., Ranatunga, T., He, J., & Yang, Y. J. (2012). Predicting plausible impacts of sets of climate and land use change scenarios on water resources. *Applied Geography*, *32*(2), 477-489.

Tucker, C., Pinzon, J., Brown, M., Slayback, D., Pak, E., Mahoney, R.,…El Saleous, N. (2005). An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, *26*(20), 4485-4498. doi:10.1080/01431160500168686

Veltcheva, A. D., & Soares, C. G. (2004). Identification of the components of wave spectra by the Hilbert Huang transform method. *Applied Ocean Research*, *26*(1), 1-12.

Vincent, C. L., Giebel, G., Pinson, P., & Madsen, H. (2010). Resolving nonstationary spectral information in wind speed time series using the Hilbert-Huang transform. *Journal of Applied Meteorology and Climatology*, *49*(2), 253-267.

Wang, J., Brown, D., & Agrawal, A. (2013a). Climate adaptation, local institutions, and rural livelihoods: A comparative study of herder communities in Mongolia and Inner Mongolia, China. *Global Environmental Change*, *23*(6), 1673-1683.

Wang, J., Brown, D., & Agrawal, A. (2013b). Sustainable governance of the Mongolian grasslands: comparing ecological and social-institutional changes in the context of climate change in Mongolia and Inner Mongolia, China. *Dryland Ecosystems in East Asia: State, Changes, and Future, HEP-De Gruyter*, 423–444.

Wang, J., Price, K. P., & Rich, P. M. (2001). Spatial patterns of NDVI in response to precipitation and temperature in the central Great Plains. *International Journal of*

Remote Sensing, *22*(18), 3827-3844.

Wang, X., Chen, F., & Dong, Z. (2006). The relative role of climatic and human factors in desertification in semiarid China. *Global Environmental Change*, *16*(1), 48-57.

Watson, D. F., and Philip, G. M. (1985). A refinement of inverse distance weighted interpolation. *Geoprocessing 2*, 315–327.

Wu, J., Zhang, Q., Li, A., & Liang, C. (2015). Historical landscape dynamics of Inner Mongolia: Patterns, drivers, and impacts. *Landscape Ecology*, *30*(9), 1579-1598. doi:10.1007/s10980-015-0209-1

Xie, B., Jia, X., Qin, Z., Shen, J., & Chang, Q. (2016). Vegetation dynamics and climate change on the loess plateau, china: 1982–2011. *Regional Environmental Change*, *16*(6), 1583-1594. doi:10.1007/s10113-015-0881-3

Xie, Y., Crary, D., Bai, Y., & Zhang, A. (2016). Modeling grassland ecosystem responses to coupled climate and socioeconomics influences in multi-spatial-and-temporal scales. *Journal of Environmental Informatics*. doi:10.3808/jei.201600337

Xilingol Statistic Bureau. (1997). *Xilingol Resplendence 50 Years*. Hohhot: Inner Mongolian Renmin Press.

Yatagai, A., & Yasunari, T. (1994). Trends and decadal-scale fluctuations of surface air temperature and precipitation over China and Mongolia during the recent 40 year period (1951-1990). *Journal of the Meteorological Society of Japan. ser. ii*, *72*(6), 937-957.

Zhang, F., Li, X., Wang, W., Ke, X., & Shi, Q. (2013). Impacts of future grassland changes on surface climate in Mongolia. *Advances in Meteorology*, *2013*, 1-9. doi:10.1155/2013/263746

Zhao, H. (1994). Land desertification by human activities in the steppe zone of Inner Mongolia Autonomous Region. In *Proceedings of International Symposium for Grassland Resources. Inner Mongolia University Press, Huhhot* (pp. 1333-1337).

APPENDICES

Appendix A: Fixed-Effect Model of All Post-EMD Observations

Fixed-effects (within) regression — Number of obs = 12015
Group variable: cid_order — Number of groups = 89

R-sq: within = 0.3846 — Obs per group: min = 135
between = 0.0014 — avg = 135.0
overall = 0.0027 — max = 135

F(23,11903) = 323.43
corr(u_i, Xb) = -0.2415 — Prob > F = 0.0000

p_evi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p_evp	-.1989381	.0757864	-2.62	0.009	-.3474917	-.0503844
p_pre	.3043889	.0310898	9.79	0.000	.2434478	.3653299
p_prs	-1.665338	.1518736	-10.97	0.000	-1.963035	-1.367641
p_rhu	-1.166347	.6016692	-1.94	0.053	-2.345717	.013023
p_ssd	-1.241126	.5170821	-2.40	0.016	-2.254691	-.2275606
p_tem	-1.489483	.1891225	-7.88	0.000	-1.860194	-1.118772
p_win	-8.514591	.8115484	-10.49	0.000	-10.10536	-6.923823
p_evp2	-.0002624	.0003533	-0.74	0.458	-.000955	.0004301
p_pre2	.0000589	.0000255	2.31	0.021	8.96e-06	.0001088
p_prs2	.0000596	7.72e-06	7.72	0.000	.0000444	.0000747
p_rhu2	.1187354	.0066443	17.87	0.000	.1057115	.1317592
p_ssd2	-.0072933	.0034025	-2.14	0.032	-.0139628	-.0006238
p_tem2	-.0107942	.0006599	-16.36	0.000	-.0120877	-.0095007
p_win2	-.1004166	.015043	-6.68	0.000	-.1299033	-.07093
p_gdppc	.0078041	.0082867	0.94	0.346	-.0084391	.0240473
p_daa	-.0552045	.0037204	-14.84	0.000	-.062497	-.0479119
p_dgr	.002289	.0001605	14.26	0.000	.0019743	.0026036
p_dls	.0632331	.0290869	2.17	0.030	.0062181	.1202481
p_dhw	-.0393745	.0024547	-16.04	0.000	-.0441861	-.034563
p_srural	-.2455003	.0142687	-17.21	0.000	-.2734692	-.2175313
p_slgov	.2643003	.0262091	10.08	0.000	.2129261	.3156745
p_sinv	-.07231	.0148768	-4.86	0.000	-.101471	-.0431491
p_sfarm	.0020777	.0018263	1.14	0.255	-.0015022	.0056576
_cons	13380.07	1004.754	13.32	0.000	11410.59	15349.55
sigma_u	845.30515					
sigma_e	70.791417					
rho	.99303536	(fraction of variance due to u_i)				

F test that all u_i=0: F(88, 11903) = 2996.55 — Prob > F = 0.0000

Appendix B: Random-Effect Model of All Post-EMD Observations

Random-effects GLS regression — Number of obs = **12015**
Group variable: **cid_order** — Number of groups = **89**

R-sq: within = **0.3799** — Obs per group: min = **135**
between = **0.8063** — avg = **135.0**
overall = **0.7925** — max = **135**

Wald chi2(**23**) = **7468.63**
corr(u_i, X) = **0** (assumed) — Prob > chi2 = **0.0000**

p_evi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
p_evp	-.1633995	.0764254	-2.14	0.033	-.3131905	-.0136085
p_pre	.3588743	.0311019	11.54	0.000	.2979157	.4198329
p_prs	-1.031915	.1446132	-7.14	0.000	-1.315352	-.7484783
p_rhu	-.24703	.604239	-0.41	0.683	-1.431317	.9372567
p_ssd	-1.240377	.5220931	-2.38	0.018	-2.263661	-.2170934
p_tem	-1.313925	.1897006	-6.93	0.000	-1.685731	-.9421183
p_win	-8.46375	.8194761	-10.33	0.000	-10.06989	-6.857606
p_evp2	-.0003086	.0003567	-0.87	0.387	-.0010076	.0003904
p_pre2	.0000584	.0000257	2.27	0.023	7.99e-06	.0001088
p_prs2	.0000791	7.64e-06	10.36	0.000	.0000641	.0000941
p_rhu2	.1119612	.0066711	16.78	0.000	.0988861	.1250363
p_ssd2	-.0058916	.0034303	-1.72	0.086	-.0126148	.0008316
p_tem2	-.0106897	.0006631	-16.12	0.000	-.0119895	-.00939
p_win2	-.1009483	.0151871	-6.65	0.000	-.1307145	-.0711821
p_gdppc	.0025766	.0083603	0.31	0.758	-.0138093	.0189625
p_daa	-.0591854	.0037467	-15.80	0.000	-.0665288	-.0518421
p_dgr	.0022563	.000162	13.93	0.000	.0019389	.0025738
p_dls	.0678553	.0293708	2.31	0.021	.0102896	.125421
p_dhw	-.0390176	.0024778	-15.75	0.000	-.043874	-.0341612
p_srural	-.2423971	.0144051	-16.83	0.000	-.2706306	-.2141636
p_slgov	.2556958	.0264567	9.66	0.000	.2038415	.30755
p_sinv	-.0634111	.0149915	-4.23	0.000	-.0927939	-.0340283
p_sfarm	.0008312	.0018421	0.45	0.652	-.0027791	.0044416
_cons	5959.894	822.0747	7.25	0.000	4348.657	7571.13
sigma_u	342.69372					
sigma_e	70.791417					
rho	.95907384	(fraction of variance due to u_i)				

Appendix C: Fixed-Effect Model of Reduced Post-EMD Observations

Fixed-effects (within) regression — Number of obs = 12015
Group variable: cid_order — Number of groups = 89

R-sq: within = 0.3845 — Obs per group: min = 135
between = 0.0027 — avg = 135.0
overall = 0.0044 — max = 135

F(20,11906) = 371.82
corr(u_i, Xb) = -0.2271 — Prob > F = 0.0000

p_evi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p_evp	-.2391713	.0503046	-4.75	0.000	-.3377765	-.1405661
p_pre	.2980649	.0307025	9.71	0.000	.2378831	.3582467
p_prs	-1.666286	.1518614	-10.97	0.000	-1.963959	-1.368613
p_rhu	-1.037564	.5956298	-1.74	0.082	-2.205096	.1299673
p_ssd	-1.260261	.5165513	-2.44	0.015	-2.272786	-.2477358
p_tem	-1.546985	.1852785	-8.35	0.000	-1.910161	-1.183809
p_win	-8.335261	.7731852	-10.78	0.000	-9.85083	-6.819692
p_pre2	.0000635	.0000253	2.51	0.012	.0000139	.0001131
p_prs2	.00006	7.70e-06	7.79	0.000	.0000449	.0000751
p_rhu2	.1183333	.0065952	17.94	0.000	.1054056	.1312611
p_ssd2	-.0075692	.003391	-2.23	0.026	-.014216	-.0009223
p_tem2	-.0107789	.0006583	-16.37	0.000	-.0120693	-.0094885
p_win2	-.1035486	.0144745	-7.15	0.000	-.131921	-.0751763
p_daa	-.0553004	.0037121	-14.90	0.000	-.0625768	-.048024
p_dgr	.0022898	.0001604	14.27	0.000	.0019753	.0026043
p_dls	.0934693	.009923	9.42	0.000	.0740186	.1129199
p_dhw	-.039276	.0024525	-16.01	0.000	-.0440834	-.0344687
p_srural	-.2466992	.0142347	-17.33	0.000	-.2746015	-.2187969
p_slgov	.2628346	.0260745	10.08	0.000	.2117243	.3139448
p_sinv	-.0713888	.0148463	-4.81	0.000	-.10049	-.0422877
_cons	13358.95	1003.726	13.31	0.000	11391.48	15326.41
sigma_u	841.57502					
sigma_e	70.79025					
rho	.99297415	(fraction of variance due to u_i)				

F test that all u_i=0: F(88, 11906) = 3051.24 — Prob > F = 0.0000

Appendix D: Fixed-Effect Model of Reduced Pre-EMD Observations

Fixed-effects (within) regression — Number of obs = **12015**
Group variable: **cid_order** — Number of groups = **89**

R-sq: within = **0.5331** — Obs per group: min = **135**
between = **0.4839** — avg = **135.0**
overall = **0.4022** — max = **135**

F(**20,11906**) = **679.75**
corr(u_i, Xb) = **0.2008** — Prob > F = **0.0000**

evi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
evp	.4588102	.1498658	3.06	0.002	.1650487	.7525717
pre	.1959378	.0631989	3.10	0.002	.0720576	.319818
prs	-.6394369	2.212133	-0.29	0.773	-4.975579	3.696705
rhu	-23.55495	2.593841	-9.08	0.000	-28.63931	-18.4706
ssd	27.36019	3.505754	7.80	0.000	20.48834	34.23204
tem	3.792934	.6607224	5.74	0.000	2.49781	5.088058
win	-73.93013	5.053052	-14.63	0.000	-83.83494	-64.02533
pre2	-.0000886	.0000488	-1.81	0.070	-.0001844	7.10e-06
prs2	.0000252	.0001238	0.20	0.838	-.0002173	.0002678
rhu2	.4886269	.0260409	18.76	0.000	.4375826	.5396712
ssd2	-.1180661	.0205316	-5.75	0.000	-.1583115	-.0778208
tem2	.0014685	.0019002	0.77	0.440	-.0022561	.0051931
win2	.6931208	.0833076	8.32	0.000	.5298242	.8564174
daa	-7.503046	1.089866	-6.88	0.000	-9.639362	-5.36673
dgr	2.197459	.200313	10.97	0.000	1.804812	2.590105
dls	2.580147	1.027933	2.51	0.012	.5652305	4.595063
dhw	.0415386	.0251971	1.65	0.099	-.0078519	.0909291
srural	-.7518932	.5601818	-1.34	0.180	-1.849941	.3461545
slgov	-.3781173	.3601658	-1.05	0.294	-1.084101	.3278664
sinv	.1859136	.0717821	2.59	0.010	.045209	.3266182
_cons	5121.713	9904.64	0.52	0.605	-14293	24536.43
sigma_u	742.82976					
sigma_e	496.01193					
rho	.69162639	(fraction of variance due to u_i)				

F test that all u_i=0: F(**88, 11906**) = **109.32** — Prob > F = **0.0000**

Appendix E: Ordinary Least-Square Model of Reduced Post-EMD Observations

Source	SS	df	MS		
Model	6.6700e+09	20	333498789	Number of obs	12015
Residual	1.4052e+09	11994	117161.317	F(20, 11994)	2846.49
				Prob > F	0.0000
Total	8.0752e+09	12014	672149.877	R-squared	0.8260
				Adj R-squared	0.8257
				Root MSE	342.29

p_evi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p_evp	2.320585	.1562774	14.85	0.000	2.014255	2.626914
p_pre	-.0525621	.1183715	-0.44	0.657	-.2845895	.1794652
p_prs	7.598879	.4591761	16.55	0.000	6.698819	8.498938
p_rhu	4.578538	2.192267	2.09	0.037	.2813394	8.875737
p_ssd	2.916135	2.095122	1.39	0.164	-1.190643	7.022914
p_tem	-2.507477	.6969408	-3.60	0.000	-3.873593	-1.14136
p_win	-12.51744	3.326468	-3.76	0.000	-19.03785	-5.997022
p_pre2	.0005374	.0000943	5.70	0.000	.0003525	.0007223
p_prs2	-.0003804	.0000253	-15.04	0.000	-.00043	-.0003308
p_rhu2	.4455552	.0232095	19.20	0.000	.4000608	.4910496
p_ssd2	-.0904013	.0129321	-6.99	0.000	-.1157504	-.0650522
p_tem2	-.013317	.0022252	-5.98	0.000	-.0176788	-.0089552
p_win2	-.4739636	.0580504	-8.16	0.000	-.5877517	-.3601755
p_daa	-.0663397	.015508	-4.28	0.000	-.096738	-.0359415
p_dgr	.004796	.0006943	6.91	0.000	.0034351	.0061569
p_dls	.2616048	.0438036	5.97	0.000	.1757427	.347467
p_dhw	.0590501	.0078123	7.56	0.000	.0437367	.0743636
p_srural	.3155412	.0584938	5.39	0.000	.200884	.4301984
p_slgov	.0654627	.1201551	0.54	0.586	-.1700607	.3009861
p_sinv	-.1263114	.0418607	-3.02	0.003	-.2083651	-.0442577
_cons	-34848.64	2085.172	-16.71	0.000	-38935.92	-30761.37

Appendix F: Ordinary Least-Square Model of Reduced Pre-EMD Observations

Source	SS	df	MS		
Model	1.0129e+10	20	506453992	Number of obs =	12015
Residual	5.2960e+09	11994	441550.356	F(20, 11994) =	1146.99
				Prob > F =	0.0000
Total	1.5425e+10	12014	1283921.66	R-squared =	0.6567
				Adj R-squared =	0.6561
				Root MSE =	664.49

evi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
evp	.5642446	.1861414	3.03	0.002	.1993775	.9291118
pre	.3821341	.0812901	4.70	0.000	.2227924	.5414758
prs	24.66567	1.075706	22.93	0.000	22.55711	26.77423
rhu	-18.12799	3.312575	-5.47	0.000	-24.62117	-11.6348
ssd	29.3569	4.673574	6.28	0.000	20.19594	38.51786
tem	.9686064	.8733663	1.11	0.267	-.7433328	2.680546
win	-44.91227	6.27725	-7.15	0.000	-57.2167	-32.60785
pre2	-.0001254	.0000644	-1.95	0.052	-.0002517	8.63e-07
prs2	-.0013096	.0000589	-22.23	0.000	-.0014251	-.0011941
rhu2	.6160154	.0332785	18.51	0.000	.5507842	.6812467
ssd2	-.0831743	.0273158	-3.04	0.002	-.1367178	-.0296309
tem2	-.0035158	.0024832	-1.42	0.157	-.0083832	.0013516
win2	.2382912	.1059206	2.25	0.024	.0306697	.4459128
daa	10.12036	.6839378	14.80	0.000	8.779733	11.46099
dgr	1.82891	.1522061	12.02	0.000	1.530561	2.127258
dls	1.046966	1.306638	0.80	0.423	-1.514256	3.608187
dhw	-.1851605	.0198809	-9.31	0.000	-.2241303	-.1461906
srural	.1518535	.268825	0.56	0.572	-.375087	.6787939
slgov	.8213151	.4333632	1.90	0.058	-.028147	1.670777
sinv	-.08068	.0737313	-1.09	0.274	-.2252051	.0638452
_cons	-115449.3	4909.037	-23.52	0.000	-125071.8	-105826.8

Appendix G: Linear Regression of Reduced Post-EMD Observations to Obtain R-Squared Value for The Fixed-Effect Model

Linear regression, absorbing indicators

Number of obs	12015
F(20, 11906)	371.82
Prob > F	0.0000
R-squared	0.9926
Adj R-squared	0.9925
Root MSE	70.7902

p_evi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p_evp	-.2391713	.0503046	-4.75	0.000	-.3377765	-.1405661
p_pre	.2980649	.0307025	9.71	0.000	.2378831	.3582467
p_prs	-1.666286	.1518614	-10.97	0.000	-1.963959	-1.368613
p_rhu	-1.037564	.5956298	-1.74	0.082	-2.205096	.1299673
p_ssd	-1.260261	.5165513	-2.44	0.015	-2.272786	-.2477358
p_tem	-1.546985	.1852785	-8.35	0.000	-1.910161	-1.183809
p_win	-8.335261	.7731852	-10.78	0.000	-9.85083	-6.819692
p_pre2	.0000635	.0000253	2.51	0.012	.0000139	.0001131
p_prs2	.00006	7.70e-06	7.79	0.000	.0000449	.0000751
p_rhu2	.1183333	.0065952	17.94	0.000	.1054056	.1312611
p_ssd2	-.0075692	.003391	-2.23	0.026	-.014216	-.0009223
p_tem2	-.0107789	.0006583	-16.37	0.000	-.0120693	-.0094885
p_win2	-.1035486	.0144745	-7.15	0.000	-.131921	-.0751763
p_daa	-.0553004	.0037121	-14.90	0.000	-.0625768	-.048024
p_dgr	.0022898	.0001604	14.27	0.000	.0019753	.0026043
p_dls	.0934693	.009923	9.42	0.000	.0740186	.1129199
p_dhw	-.039276	.0024525	-16.01	0.000	-.0440834	-.0344687
p_srural	-.2466992	.0142347	-17.33	0.000	-.2746015	-.2187969
p_slgov	.2628346	.0260745	10.08	0.000	.2117243	.3139448
p_sinv	-.0713888	.0148463	-4.81	0.000	-.10049	-.0422877
_cons	13358.95	1003.726	13.31	0.000	11391.48	15326.41
cid_order	F(88, 11906) =	3051.240	0.000	(89 categories)		

Appendix H: Hausman Test

	Coefficients (b) fixed	(B) random	(b−B) Difference	sqrt(diag(V_b−V_B)) S.E.
p_evp	−.2391713	−.2110934	−.0280779	.0049772
p_pre	.2980649	.3573811	−.0593162	.0046191
p_prs	−1.666286	−1.028743	−.6375426	.0513475
p_rhu	−1.037564	−.1702375	−.8673268	.0604748
p_ssd	−1.260261	−1.239997	−.0202639	.0099227
p_tem	−1.546985	−1.336807	−.210178	.0237584
p_win	−8.335261	−8.418075	.082814	.0114046
p_pre2	.0000635	.0000607	2.83e−06	6.97e−07
p_prs2	.00006	.0000793	−.0000193	1.53e−06
p_rhu2	.1183333	.1118172	.0065161	.0007135
p_ssd2	−.0075692	−.0061235	−.0014457	.0001963
p_tem2	−.0107789	−.0106914	−.0000876	.0000655
p_win2	−.1035486	−.101662	−.0018866	.0004541
p_daa	−.0553004	−.0592167	.0039164	.0002746
p_dgr	.0022898	.0022595	.0000303	6.28e−06
p_dls	.0934693	.0795275	.0139417	.0011488
p_dhw	−.039276	−.0390045	−.0002716	.0000704
p_srural	−.2466992	−.2427897	−.0039095	.0003809
p_slgov	.2628346	.2542464	.0085882	.0007867
p_sinv	−.0713888	−.062769	−.0086198	.0009792

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(16) = (b−B)'[(V_b−V_B)^(−1)](b−B)
= 250.25
Prob>chi2 = 0.0000

Appendix I: Fixed-Effect Model of Reduced Post-EMD Observations Without Polynomial Terms

Fixed-effects (within) regression — Number of obs = **12015**
Group variable: **cid_order** — Number of groups = **89**

R-sq: within = **0.3492** — Obs per group: min = **135**
between = **0.0414** — avg = **135.0**
overall = **0.0450** — max = **135**

F(**14,11912**) = **456.55**
corr(u_i, Xb) = **0.0242** — Prob > F = **0.0000**

p_evi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p_evp	-.5649382	.0486714	-11.61	0.000	-.6603421	-.4695343
p_pre	.5280664	.021976	24.03	0.000	.4849899	.5711428
p_prs	-.3117065	.1013767	-3.07	0.002	-.5104213	-.1129916
p_rhu	4.992768	.4750227	10.51	0.000	4.061646	5.92389
p_ssd	-1.534114	.3627234	-4.23	0.000	-2.245111	-.8231173
p_tem	-2.739774	.1769133	-15.49	0.000	-3.086553	-2.392995
p_win	-12.98164	.3935627	-32.98	0.000	-13.75309	-12.21019
p_daa	-.0485775	.0037181	-13.07	0.000	-.0558655	-.0412895
p_dgr	.0017517	.0001621	10.81	0.000	.0014339	.0020694
p_dls	.0655088	.0100236	6.54	0.000	.0458609	.0851567
p_dhw	-.0423388	.0024972	-16.95	0.000	-.0472337	-.0374439
p_srural	-.2335514	.0144838	-16.12	0.000	-.2619421	-.2051607
p_slgov	.2313684	.0264579	8.74	0.000	.1795065	.2832302
p_sinv	-.0845211	.0148623	-5.69	0.000	-.1136537	-.0553886
_cons	5864.101	929.2121	6.31	0.000	4042.694	7685.508
sigma_u	802.63278					
sigma_e	72.771358					
rho	.99184673	(fraction of variance due to u_i)				

F test that all u_i=0: F(**88, 11912**) = **3093.82** Prob > F = **0.0000**

Appendix J: Ordinary Least-Square Model to Esitmate Cross-Sectional Coefficients of Reduced Post-EMD Observations

Source	SS	df	MS		
Model	8.0155e+09	108	74218005.2	Number of obs	12015
Residual	59664055.2	11906	5011.25946	F(108, 11906)	14810.25
				Prob > F	0.0000
				R-squared	0.9926
				Adj R-squared	0.9925
Total	8.0752e+09	12014	672149.877	Root MSE	70.79

p_evi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p_evp	-.2391713	.0503046	-4.75	0.000	-.3377765	-.1405661
p_pre	.2980649	.0307025	9.71	0.000	.2378831	.3582467
p_prs	-1.666286	.1518614	-10.97	0.000	-1.963959	-1.368613
p_rhu	-1.037564	.5956298	-1.74	0.082	-2.205096	.1299673
p_ssd	-1.260261	.5165513	-2.44	0.015	-2.272786	-.2477358
p_tem	-1.546985	.1852785	-8.35	0.000	-1.910161	-1.183809
p_win	-8.335261	.7731852	-10.78	0.000	-9.85083	-6.819692
p_pre2	.0000635	.0000253	2.51	0.012	.0000139	.0001131
p_prs2	.00006	7.70e-06	7.79	0.000	.0000449	.0000751
p_rhu2	.1183333	.0065952	17.94	0.000	.1054056	.1312611
p_ssd2	-.0075692	.003391	-2.23	0.026	-.014216	-.0009223
p_tem2	-.0107789	.0006583	-16.37	0.000	-.0120693	-.0094885
p_win2	-.1035486	.0144745	-7.15	0.000	-.131921	-.0751763
p_daa	-.0553004	.0037121	-14.90	0.000	-.0625768	-.048024
p_dgr	.0022898	.0001604	14.27	0.000	.0019753	.0026043
p_dls	.0934693	.009923	9.42	0.000	.0740186	.1129199
p_dhw	-.039276	.0024525	-16.01	0.000	-.0440834	-.0344687
p_srural	-.2466992	.0142347	-17.33	0.000	-.2746015	-.2187969
p_slgov	.2628346	.0260745	10.08	0.000	.2117243	.3139448
p_sinv	-.0713888	.0148463	-4.81	0.000	-.10049	-.0422877
cid_order						
2	-18.8561	9.624794	-1.96	0.050	-37.72226	.0100713
3	39.90842	17.97035	2.22	0.026	4.683606	75.13324
4	2088.596	20.35925	102.59	0.000	2048.689	2128.504
5	1991.994	23.03343	86.48	0.000	1946.845	2037.144
6	44.63172	16.49122	2.71	0.007	12.30624	76.95719
7	822.3528	15.14295	54.31	0.000	792.6702	852.0355
8	290.6094	13.54743	21.45	0.000	264.0542	317.1645
9	1815.996	16.95787	107.09	0.000	1782.756	1849.236
10	601.6004	17.94556	33.52	0.000	566.4242	636.7766

11	369.6572	14.37144	25.72	0.000	341.4868	397.8275
12	779.5216	17.93444	43.47	0.000	744.3671	814.676
13	694.7361	15.52012	44.76	0.000	664.3141	725.158
14	1429.065	17.14582	83.35	0.000	1395.456	1462.674
15	2207.727	95.23767	23.18	0.000	2021.046	2394.409
16	2122.447	90.45743	23.46	0.000	1945.135	2299.758
17	1866.128	69.2203	26.96	0.000	1730.445	2001.811
18	2198.374	76.12893	28.88	0.000	2049.148	2347.599
19	1953.36	69.4749	28.12	0.000	1817.178	2089.542
20	2604.519	68.20817	38.18	0.000	2470.82	2738.218
21	1916.98	45.60175	42.04	0.000	1827.593	2006.366
22	1953.986	57.83271	33.79	0.000	1840.624	2067.347
23	2647.3	68.77756	38.49	0.000	2512.485	2782.115
24	1617.409	78.17165	20.69	0.000	1464.18	1770.638
25	564.575	15.20743	37.12	0.000	534.766	594.3841
26	589.4029	16.45006	35.83	0.000	557.158	621.6477
27	310.7563	13.35591	23.27	0.000	284.5766	336.9361
28	482.496	13.7433	35.11	0.000	455.5569	509.4351
29	161.1719	15.57837	10.35	0.000	130.6357	191.708
30	471.5563	15.14459	31.14	0.000	441.8704	501.2422
31	700.3749	15.32089	45.71	0.000	670.3435	730.4064
32	589.9733	15.61249	37.79	0.000	559.3703	620.5763
33	1130.132	16.80429	67.25	0.000	1097.193	1163.071
34	1367.972	17.41933	78.53	0.000	1333.827	1402.117
35	908.5755	16.81361	54.04	0.000	875.6181	941.5329
36	1504.17	17.64956	85.22	0.000	1469.574	1538.766
37	1287.956	16.72949	76.99	0.000	1255.164	1320.749
38	891.6696	16.37844	54.44	0.000	859.5651	923.774
39	2873.656	88.43705	32.49	0.000	2700.305	3047.007
40	1887.268	70.67786	26.70	0.000	1748.728	2025.808
41	2431.574	74.25379	32.75	0.000	2286.024	2577.123
42	2818.437	83.74813	33.65	0.000	2654.277	2982.597
43	2325.702	64.61508	35.99	0.000	2199.046	2452.358
44	2490.674	75.11464	33.16	0.000	2343.437	2637.911
45	1868.296	70.2857	26.58	0.000	1730.524	2006.067
46	1359.009	68.76184	19.76	0.000	1224.224	1493.793
47	2638.179	90.67969	29.09	0.000	2460.432	2815.926
48	1126.71	71.5582	15.75	0.000	986.4442	1266.976
49	1688.655	66.55222	25.37	0.000	1558.202	1819.108
50	2660.764	71.41643	37.26	0.000	2520.776	2800.752

51	2964.034	83.5022	35.50	0.000	2800.356	3127.712
52	1832.104	79.56622	23.03	0.000	1676.142	1988.067
53	2579.803	114.8486	22.46	0.000	2354.681	2804.925
54	2734.144	118.0203	23.17	0.000	2502.805	2965.483
55	2356.685	115.7767	20.36	0.000	2129.744	2583.626
56	2344.432	117.3004	19.99	0.000	2114.505	2574.36
57	2254.237	110.3178	20.43	0.000	2037.996	2470.478
58	2225.668	104.9973	21.20	0.000	2019.857	2431.48
59	2359.149	106.9472	22.06	0.000	2149.515	2568.783
60	151.072	11.96903	12.62	0.000	127.6107	174.5333
61	707.0319	21.55331	32.80	0.000	664.7839	749.2799
62	843.5416	22.24412	37.92	0.000	799.9394	887.1437
63	966.4809	17.68625	54.65	0.000	931.813	1001.149
64	926.16	21.49025	43.10	0.000	884.0355	968.2844
65	817.3613	24.35249	33.56	0.000	769.6265	865.0962
66	518.7876	24.56652	21.12	0.000	470.6333	566.942
67	1206.84	19.18045	62.92	0.000	1169.243	1244.437
68	895.5908	22.5485	39.72	0.000	851.392	939.7895
69	445.2288	14.14043	31.49	0.000	417.5112	472.9463
70	887.4504	20.82246	42.62	0.000	846.635	928.2658
71	1191.885	18.75247	63.56	0.000	1155.127	1228.643
72	1065.999	23.83114	44.73	0.000	1019.286	1112.712
73	1320.32	48.22213	27.38	0.000	1225.797	1414.843
74	1415.503	26.33117	53.76	0.000	1363.89	1467.116
75	508.3347	32.20592	15.78	0.000	445.2059	571.4636
76	493.5808	18.40067	26.82	0.000	457.5125	529.6491
77	600.5909	23.85615	25.18	0.000	553.8289	647.3528
78	1208.995	21.19027	57.05	0.000	1167.459	1250.532
79	1668.273	47.27233	35.29	0.000	1575.612	1760.935
80	1238.977	30.3787	40.78	0.000	1179.43	1298.524
81	629.036	17.8901	35.16	0.000	593.9684	664.1035
82	1168.756	23.00018	50.82	0.000	1123.672	1213.84
83	897.6567	19.45712	46.14	0.000	859.5175	935.7958
84	2821.053	63.09715	44.71	0.000	2697.373	2944.734
85	2433.11	87.63876	27.76	0.000	2261.324	2604.897
86	2254.009	104.0507	21.66	0.000	2050.053	2457.966
87	2293.668	99.41853	23.07	0.000	2098.792	2488.545
88	2311.904	103.1653	22.41	0.000	2109.684	2514.125
89	2340.06	91.07778	25.69	0.000	2161.533	2518.588
_cons	11931.47	976.1425	12.22	0.000	10018.07	13844.87