Essays on the Environmental Economic History of China During the Ming and Qing Dynasties

By

Kai Cheng

A thesis submitted to
the University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Department of Economics
Birmingham Business School
College of Social Sciences
University of Birmingham
September 2021

ACKNOWLEDGMENTS

First and foremost, I am extremely grateful to my supervisors, Prof. Robert Elliott, Prof. Eric Strobl and Dr Sami Bensassi, for their invaluable advice and patience. Without their complete trust and sufficient help, I would never have overcome plenty of difficulties to complete the thesis. Furthermore, their immense knowledge and ample experience have encouraged me in all the time of my academic research and daily life. What is more, my supervisors have provided abundant resources and training opportunities to improve my expertise and have helped me proofreading, which I would appreciate deeply.

I would like to extend my sincere thanks to all Professors and academic researchers in workshops of the department and CAGE summer school 2021 for their insightful comments and suggestions. Furthermore, I would like to express a special thanks to the administrative staff in the Department of Economics and Birmingham Business School. Besides, my appreciation also goes to all my colleagues for their help across my PhD journey.

I am also profoundly grateful to my family and friends for their encouragement and support throughout my studies. Lastly, I would like to thank my supervisors and department for the studentship and funding from my family that allowed me to conduct this thesis.

Contents

List of Figures

List of Tables

Chapter One

Introduction

In recent years concerns over the state of the environment have received increasing attention with fears growing about the impact of climate change and the perceived increase and intensity of climate induced natural disasters. For example, at the policy level, since the signing of the Kyoto Protocol, Europe has gradually established the European Emissions Trading System (EUETS) covering all of Europe (Ellerman and Buchner, 2007), which is by far the world's largest emissions trading market. However, people have understood the damaging impact of changes in environmental conditions for hundreds if not thousands of years. Despite this depth of knowledge, environmental history was not considered an independent subject until 1972 although there were a number of pioneering studies that came before. The more commonly used definition of environmental history is "the study, through historical time, of the interface where specifically human systems meet with other natural systems" (Elvin and Liu, 1998).

Economic historians also recognised the potential importance of environmental factors. As early as the 18th century, Malthus, Winch, and James (1798) argued that environmental resources could limit population growth. Many studies have also found a strong correlation between human systems and natural systems. As a result, environmental factors have been regarded as a key variable for reinterpreting old topics.

Environmental economic history combines the research results and research methods of environmental history and economic history, focusing on the interaction between ancient environmental change and economic activities in the hope of discovering new economic mechanisms. Since historic environmental data helps us to understand past environmental conditions, collecting and cleaning historical environmental data usually forms the basis of environmental history and economic history and is dependent on the related achievements from the disciplines of palaeoclimatology and history. After establishing an accurate historical environment database, with the help of the Geographic Information System (GIS) and quantitative analysis, the connections between humans and the environment can be investigated.

In a Chinese context, research into environmental history really began in the 1980s. By 1998, the first "Chinese Ecological Environmental History Colloquium" was organised, and published ‘Sediments of Time: Environment and Society in Chinese History’ (Elvin and Liu, 1998). Since then, the study of environmental history has gradually become more widespread. The main thrust of this work has been to analyse historical records and to create a consistent historical record of environmental events.

China's environmental economic history also overlaps with economic history and environmental history. Most of the essential topics in China's economic history focus on the Great Divergence and economic changes in more recent times. Among them, research on the Great Divergence can be classified as traditionalist (Allen, 2009; Allen et al., 2011) and revisionist (Pomeranz, 2001). Discussions usually concentrate on the time points, the compatible regions and the magnitude of the Great Divergence. The scope of discussion covers most of the period after the 14th century and before the 20th century and most of China's geographical, mainly to the south of the Yangtze River. This period also covers the Little Ice Age when the environment under went significant changes, and the frequency of disasters has also increased significantly. Therefore, it is important to consider whether environmental factors can supplement our understanding from previous studies which concentrate on political and economic regimes as the reason of the economic stagnation of historical China (Huang, 1990; Ma and Rubin, 2017; Li, 2000).

In addition to the Great Divergence, there are other topics where environmental factors are likely to be important. For example, research into the relationship between the environment and warfare. Zhang, Jim, et al. (2006) tested the correlations between "average temperature anomalies and the number of wars in each decade" and found a statistically significant relationship between cooler temperature and conflict. Bai and Kung (2011) found that adverse rainfall shocks had a significant and positive relationship on nomadic incursions into China. Finally, Jia (2014) identified that the adoption of the sweet potato reduced the possibility of a revolt breaking out.

This thesis contributes and expand the literature on Chinese environmental economic history through the creation of a new and detailed environmental data set. More specifically, chapter two combines environmental records from two different sources, which are "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*" and "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*". We digitise these records using Artificial Intelligence (AI) techniques based on optical character recognition (OCR) techniques and parametrise them to generate weighted frequencies of floods, droughts, wind, and cold. At the same time, we visualise the environmental data in GIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Our data set is at the county level for the period 1368 to 1911 and has a higher level of precision compared to other data sets and covers all of China based on the boundaries in 1911. In order to verify the credibility of the data set constructed in chapter two, chapter three provides several validation methods.

Following the creation of our data set the next two chapters address two different questions related to the economic history of China. Chapter four uses the county-level environmental data and town number data for the traditional agricultural regions of China from 1820 to 1911 to estimate the relationship between environmental events and rural-to-urban migration. The Harris-Todaro model is employed to illustrate how people decide to move to the urban area when faced with weather shocks. Furthermore, empirical results using a cross-sectional model suggest that flood events encourage rural-to-urban migration. However, the estimation is different at the prefecture level and indicates that county-level analysis is important to correctly identify the rural-urban migration mechanism. This thesis also contributes to market town ("Shi Zhen") research, which seldom provides empirical evidence at the county level.

Chapter five also uses the data created in Chapter 2 and employs an annual panel of traditional agricultural regions of China from 1368 to 1911 to test the weather-conflict relationship. Although the data on conflicts in well documented, there is no county-level estimations in the current literature. This thesis extends the research to the county level and adopts the Spatial Durbin Model to investigate the role of spatial spillovers. The results re-affirm that prefecture-level estimations may be biased and suggests that the non-spatial model might be violated. Specifically, in the study of the weather-conflict connection in historical China, a non-spatial model at the prefecture level may overestimate the influence of droughts and underestimate the influence of floods. Floods turn out to generate more conflicts locally and across different areas.

Finally, chapter six concludes the thesis and discusses areas for future research.

Chapter Two

Constructing a County Level Extreme Environmental Events Dataset for China during the Ming and Qing Dynasties

2.1 Introduction

2.1.1 Environmental economic history

In recent decades, attention paid to environmental influences has increased due to the possibility of more environmental crises in the future. For example, more and more people have started to take measures to prevent global warming, while many governments have imposed relevant policies or established institutions to protect the environment (Roberts, 2004), such as the “European Emissions Trade System” launched in 2005, which is a trading scheme mainly to control the emissions of greenhouse gas (Ellerman and Buchner, 2007). However, it is not the first time for people to realise the environmental impact on human activities. Traced back to the beginning of the 19th century, the “Conservation Movement” in the United States rapidly grew to an environmental and political movement, which inspired people to “preserve and protect America’s wildlife, wetlands, and other natural resources” (Library of Congress, 2002). Simultaneously, similar concepts and activities were encouraged worldwide, and prolonged their influences on the academic field. Impressed by this wide spread movement, Nash (1972) used the term “environmental history” to establish a new course at the University of California Santa Barbara in 1972. Since then environmental history has become an individual discipline in its own right. Generally environmental history mainly focuses on “the interaction between human cultures and the environment in the past” (Worster, 2008). Several studies of the environment before 1972, such as Turner (1921) and Webb (1931), have in this regard pioneered the study of environmental history. In this regard, the historical data used in the analysis of the environment can stretch back to ancient times across the world (Zhang, 2004b; Elvin and Liu, 1998).

As the studies on environmental history grew, a deeper understanding of environmental history evolved, taking a more global outlook (James and Kala, 2017). Elvin and Liu (1998) have argued that “environmental history is more precisely defined as the study, through historical time, of the interface where specifically human systems meet with other natural systems”. In this regard, nature’s systems encompass a number of different aspects, including “climates, topographies, rocks and soils, water, vegetation, animals and microorganisms, or, to put it another way, the biogeochemical systems on and near the Earth's surface that produce and process energy and accessible resources and recycle waste products". Studies that have concentrated on Africa (Beinart, 1984; Fairhead and Leach, 1996; Alsan, 2015), Asia(McNeill, 1998; Elvin, 2006) or taken a global perspective (Crosby, 1972) have illustrated the dynamic and static interaction between different environmental systems and human systems.

Economists have identified the relationship between environmental conditions and the economy even earlier. Around the 1800s, Malthus, Winch, and James (1798) noted that natural resources can set an upper bound on the total population within a given region. Unsurprisingly then, many topics in economic history have involved environmental factors (Sandmo, 2015). Filled with literature from environmental history, environmental economic history has been expanded. The study of environmental economic history usually focuses on the causal effects of each environmental factor on economic output (James and Kala, 2017; Hornbeck, 2012), or uses the environmental impact as an instrument to identify the causal effect of an economic factor (Williams, 1994; Hornbeck and Naidu, 2014). In this regard, appropriate environmental data is the essential requirement for economic analysis, and the research of environmental economic history often contributes "the generation and assembly of new Geographic Information Systems (GIS) databases" (Alsan, 2015; Ashraf and Michalopoulos, 2015; Fiszbein, 2017) to the literature.

2.1.2 Situation of China

Triggered by the research of other countries, the environmental history of China has evolved considerably since the 1980s. In 1993, the first international conference that focused on the topics from the environmental history of China was held in Hongkong by the Academia Sinica of Taiwan and Australian National University under the name of “Chinese Ecological Environment History Colloquium” (Han, 2016). Papers presented at the conference were published as a book called ‘Sediments of Time: Environment and Society in Chinese History’ (Elvin and Liu, 1998), consisting of a collection of 24 papers from 8 relevant sub-topics. As a landmark publisher for environmental history in China, this conference not only summarised the previous exploration, but also provided a guideline for future topics. McNeill (1998) argued that due to the integration of intensive inland waterways across a large range of latitude, China was unique even from the worldwide perspectives. What is more, he suggested a list of future research venues, including works on marine ecosystems, biological invasion, data reconstruction, war and political violence, export trade, and air pollution. However, the methodology of environmental history in China is characterized by different features compared with research elsewhere (Xu, 2014). One of the main aspects of studies for China is that the research relies substantially on records in historical documents since there are many historical records available over a long time period for historical China.

Similar to environmental history, research of economic history in China also requires the proficiency in investigating historical documents. Activated by the introspection of the economic stagnation of modern China, the economic history of China has developed a vast literature on the topic of the divergence between west and east. In this regard, there has been much debate as to when and how did the stagnation in China occur. In a recent paper, Ma and Rubin (2017) established a perfect refined sub-game equilibrium to prove that unlimited monarchical power in the Qing dynasty would reduce the taxation ability of the central government and thus encumber the economic development. Many academics also agree that the Yangtze Delta was economically comparable to England before the 18th century (Bernhofen et al., 2017; Shiue and Keller, 2007; Edwards, 2013), but England, as well as the rest of Europe, has advanced relatively more since the 19th century. Moreover, some academics instead argued that the earlier divergence might have emerged as early as the 17th century (Li and Zanden, 2012). Meanwhile, research on China's environmental history has denoted a continuous cooling period called the "Little Ice Age" across the 15th to 19th centuries (Chu, 1973; Marks, 1998). During this period, the number of disasters rose dramatically according to the historical record (Marks, 2011; McNeill, 1998). Thus there was considerable overlap between the time of the Great Divergence and this worsening of environmental conditions. Therefore, it seems of interest to incorporate such environmental factors into an economic analysis of the reasons for the Great Divergence.

Economists have also investigated topics other than the Great divergence in terms of the environmental economic history of China as the field developed. For example, several papers have discussed the Great Famine of the 1960s in order to obtain insight from the deep interaction between crop outputs and institutions (Gooch, 2017; Kasahara and Li, 2017; Meng, Qian, and Yared, 2015), while some studies have examined the connection between weather and the Great Famine (Bai and Kung, 2014). These papers have employed some theoretic models or have adopted a proxy index to estimate the output of agriculture, since historical data on agriculture production are sparse. While, compared with agriculture output, population data is available for more extensive regions and time periods, it is still difficult to capture the temporal variation caused by climate shocks (Li and Lin, 2015; Li, An, et al., 2015) due to missing data over the very long term. In contrast, connections between climate and conflict are more popular topics in environmental history in China, since records on conflict are much more extensive than economic data. Although the channel through which environmental variation has an impact on conflict is usually suspected to be agriculture, environmental factors might also impose direct damage on the economy, such as the destruction of physical and human resources, which in turn may also trigger conflicts. For instance, Bai and Kung (2011) assumed that rainfall would have facilitated food gathering for nomadic herds and motivated the trade within a nomadic economy, but would also have encouraged sedentary attacks. To test this potential relationship between Sino-nomadic conflicts and rainfall the authors used station level grades of wetness and dryness as indicators of high and low precipitation, respectively. As far as we know, their work is the first empirical estimation of the reason for conflicts between two different civilizations across more than two 2000 years. Relatedly, (Ma, 2011) also discovered the possible influences of environmental impact on the nomadic economy but lacked an empirical approach to examine the connection.

Since readily available data has been a limiting factor of research on historical China, academics often have constructed new environmental data sets for their research purposes, as has been done elsewhere. More precisely, Chen (2015) constructed region level flood and drought indicators (region of Yellow River) to verify the causal links between domestic conflicts and adverse climate shocks. Since the peasant uprisings are commonly considered as determinants of the dynasty cycle, the effects of extreme environmental events were hypothesized to accelerate the process of dynasty transitions "primarily through the channel of severe famine". Additionally, Jia (2014) employed the timing of sweet potato cultivation as an instrument to denote the food supplements at the prefecture level to test the relationship between climate shocks and peasant revolts. Her research provided convincing evidence of the role of agriculture production in the interaction between the nature and the human systems. In another study, migrants data was used to "examine economic effects on migrants of the Manchuria Plague of 1910–11", where it was found that "migrant households that moved to plague-hit villages soon after the plague ended prospered the most" (Li and Li, 2017). Importantly, much of the data used in these just cited studies above were generated from historical documents.

2.1.3 Works and issues on data

As environmental risks increase, it is essential for academics and policymakers to expand the investigation to the impact of environmental events on a more broader set of human activities. To this end obtaining accurate historical data on environmental events is often the first step. As mentioned earlier, data collection with regard to environmental history and environmental economic history in China tends to rely heavily on historical documents. Thus, further developing research tools to extract such information from historical documents is arguably crucial for the field to advance.

China has an extensive recorded history, where relevant environmental records are widely available in many of the historical documents. For example, there were generally reports from local governments to the central government when severe disasters occurred. What is more, famous local people often recorded abnormal environmental variation in their dairies (Xia, 2015). The majority of the meteorological phenomena were also recorded in plenty of local annals, poems, personal or official letters, novels, official documents, lyrics and personal diaries. Since such historical records are extensive and continuous, historians can compare different narratives of the same event, and then identify the occurrence and estimate the size of a given environmental event. In this regard, the Chinese historical records can be considered one of the best available data sets of the last few centuries, especially in the environmental field. Compared with other environmental proxies, such as the tree-ring and ice-core (Yang, Braeuning, et al., 2002), historical records contain specific time and location, which is an advantage for the reconstruction of a historical climate database (Zhang, 1998). However, relevant materials from historical China are impressively large making it a near impossible task for individual researcher to sort through all of the available data.

Due to the complexity of sources of historical data, recognising and collecting climate data from the historical information is important, so that academics can access these easily. Many original works have verified the validation of different sources and provided validating methods to undertake this (Zhang, Ge, and Zheng, 2002; Zhang, 2005). Research institutions and academics have extracted historical climate records from validated sources and have published their collections (Zhang, 1984; Zhang, 2004b; Wen, 2006; Wang, Xie, and Wen, 1983). These works have supplemented the historical climate database to cover the whole country across a long historical period.

As historical climate records are usually some qualitative descriptions, an important task is to parametrise organised records into acceptable indices that can then be employed in quantitative analysis. There are several methods to parametrise the environmental data (Zhang, 1984; Brooks, 2007; Yang, Man, and Zheng, 2006; Zheng, Zhang, and Zhou, 1993). Initially researchers calculated the frequency of each environmental disaster from different sources. However, Chu (1973) argued that stylistic rules and layouts of different sources varied, which could lead to non-uniformity and violate the frequency method. Accordingly, parametrisation is normally employed on similar source groups. Many works then focus on reclassifying the description of environmental events (Yang, Man, and Zheng, 2006; China Meteorological Administration Institute of Meteorology, 1981; Chen, 1987). However, these works require extensive experience in historical research and ignore the quantitative information that can be gathered from the number of records. Other works have investigated the frequency approach to calibrate environmental indices, such as the number of counties where the disaster occurred (Zheng, Zhang, and Zhou, 1993). However, all of these methods suffer from important shortcomings. For example, while the number of waterlogged counties prevents perception error from ranking historical descriptions of precipitation, it is not able to provide county-level data, since the number of waterlogged counties can only be calculated within regions which contain multiple counties. What is more, the number of cold years is an efficient index which could be applied at the county level, but the annual variation is not possible to be captured. As a result, current methods indeed guarantee uniformity to an acceptable extent but only represent a few parts of existing sources and involve a sacrifice in accuracy on spatial or time variation.

This study adopts two new data sources that cover the whole history of the entire Chine before 1911. We combine these two sources and apply a new approach to count each county's frequency of environmental events. Compared with previous frequency approaches, this study assigns various weights on records from different sources to avoid the non-uniformity and generate county-level indices for each category of extreme environmental events. Arguably this combination method can serve as a template for future work to parametrise environmental information from various sources. The period under scrutiny mainly covers the whole of the Ming and Qing dynasties from 1368 to 1911, as these two dynasties the most recent two dynasties to modern China when the local gazetteer systems of each county were universally developed. The research region includes the boundaries of the whole country based on the administrative boundaries in 1911 (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016), and the data itself is constructed at the county level. Moreover, two new categories of environmental disasters are included. The approach in compiling the data involves the use of newly developed AI techniques, including new optical character recognition and lexical analysis based on machine learning, word frequency analysis, and other text related operations by coding with Python. Application of these techniques facilitates the digitalisation work when converting paper form sources into digital data. This approach increases the efficiency of future research, especially when the digital source is not shared. Finally, this study also takes care in presenting the data generated more clearly. More precisely, in order to implement better visualisation, ArcGIS is adopted to map the spatial and temporal traits of the climatic events.

The structure of this paper is organised as follow: Section I introduces the primary background of the data collection in environmental economic history. Section II reviews previous works of the establishment of historical environmental databases in China. Section III introduces the data sources employed in this research and the complete method for data processing. Section IV presents the features of the new data set. The last section contains the conclusion and a discussion of the limitations of this research.

2.2 Literature Review

2.2.1 Topics and Data Collection in the Global Literature

Generally, the environmental aspects covered in environmental economic history can be organised into three groups: natural endowments, disasters, and climate change. As a result, fundamental research such as data reconstruction and collection of historical documents can also generally be classified according.

Natural endowments (which can also be termed environmental endowments) include both geographical endowments and biological endowments. Literature in this direction focuses on the different consequences of endowment distribution. Alsan (2015) obtained data on the historical TseTse fly distribution in Africa, to investigate "why was land in historical Africa relatively abundant", where abundant land in Africa was assumed to be one of the determinants of why Africa is undeveloped. The TseTse fly distribution was generated from the TseTse suitability index, which assumes the possible population of TseTse fly at a steady-state level from 1870. Alsan (2015) proposed a hypothesis that pre-colonial agriculture in Africa was affected by the TseTse fly and that this worsened economic output through channels that link agriculture with other sections of the economy. Alsan's study showed supportive evidence that the TseTse fly "has a negative correlation with current economic outcomes as measured by satellite light density or the observed cattle distribution in Africa" (Alsan, 2015). One should note that typically cross-sectional data and models are employed as empirical approaches when environmental endowments are taken into consideration. Meanwhile, Alesina, Giuliano, and Nunn (2013) also applied ethnographic data for modern time and land suitability for different crops to investigate the origins of the differences in gender roles. They claimed that cultivation work based on different tools would implement divergent incentives on women's willingness to work outside. Plough cultivation requires more "upper body strength, grip strength, and bursts of power," which is appropriate for men to do farming and assigns more power to men. However, shifting cultivation does not have such requirements so that. women and men are approximately identically productive for farming work in such a context. Accordingly, agricultural practices have formed the shape of different attitudes to gender. A cross-sectional approach was adopted also in Alesina's study to reveal the prolonged influences of environmental endowments on human cultural formation.

Apart from the research above, literature for relevant topics usually focuses on environmental endowments that have time-invariant characteristics during a given period, where the duration of the examined period is typically very long. For example, some studies have investigated ultraviolet radiation (Andersen, Dalgaard, and Selaya, 2016) and spatial differentiation of ancient temperature (Ashraf and Michalopoulos, 2015). Meanwhile, Geographic Information System (GIS) databases for environmental data have been developed to support further investigation of geographical endowments (Bailey, 2011). For example, it has been argued that geographic features played a role in the distribution of slavery, the shape of population density (Fenske, 2013), and trade (Fenske, 2014). For all of these topics relevant information has been organised in GIS databases. Meanwhile, it is also possible to adopt GIS databases to help illustrate the formation of culture, where studies have generated such databases to investigate the attitudes of females to work outside (Alesina, Giuliano, and Nunn, 2013), ethnolinguistic diversity (Michalopoulos, 2012) and religion plausibility (Michalopoulos, Naghavi, and Prarolo, 2016). Additionally, geographic information facilitates research on long term influences of spatial differences of any type of endowment on relevant economic output. Besides, Bleakley (2007) applied hookworm infection rates to instrument the status of personal health and compared the spatial variation of the infection rates before and after an eradication campaign. The variation of infection rates across different areas was found to respond to spatial deviations of education investment. Compared with other studies looking at the time-invariant environmental endowments, Bleakly's work focused on spatial traits and employed a panel data model to simulate the dynamic process.

According to previous research, the key point to collect data of environmental endowments, no matter whether they are time-invariant or time-varying endowments, is to identify appropriate geographical units. For endowments that are invariant across time to some extent, the historical distribution can be generated by using current data. In contrast, for endowments that cannot be estimated from current distributions, historical records are collected to construct environmental indices, which are usually dummies(Ashraf and Michalopoulos, 2015; Alesina, Giuliano, and Nunn, 2013). It is possible to use available data directly from historical records, in some special cases for limited time spans, but records before the last century do generally not sufficiently exist. This makes the parametrisation of historical records for environmental endowments an attractive alternative.

Data collection for disasters reveals different characteristics. Disasters are extreme events that usually occur suddenly in a short period, or in the case of droughts, sometimes over long periods. Disasters can leave direct and indirect damage to the economy by destroying physical and human capital. Since disasters are time varying by nature and tend not to be location specific, spatially and time varying data is generally required for research purposes. The economics of natural disasters has developed considerably over the last several decades. For example, Mohan and Strobl (2013) compiled historical sugar exports and records of hurricane frequencies in the Caribbean to explore the connection between hurricanes and the sugar industry from 1700 to 1960. Their analysis confirmed "historical documents that argued that hurricanes had a major impact on the sugar industry in the Caribbean" (Mohan and Strobl, 2013). However, investigation of the economic impact of natural disasters in a historical context is more complicated due to the limitation of data availability.

Economists have considered several different ways to deal with data generation and then to estimate their hypotheses. In this regard, using indicator variables for environmental events has played an important role in capturing the historical features, since it is easier to convert historical information into quantitative records in this way. Arthi (2014) made the occurrence of American Dust Bowl into dummy variable as a treatment to measure the influences of dust bowl on the accumulation of human capital in the long term. Hansen and Libecap (2004) showed that the dust bowl would impose negative impacts on the outcome of firms and then would lead to a significant reduction in the economy, where regions had been assigned a dummy to represent whether a given location was affected by the dust bowl. Hornbeck and Naidu (2014) investigated the Great Mississippi Flood in 1927 and noticed that the collapse of the levee encouraged black out-migration. The authors argued that there should be a strong connection between disasters and agricultural modernisation. To investigate this they also used a dummy to denote whether a region experienced a broken levee. One should notice is zthat the cited studies focus on individual cases and often contain few periods (typically two periods). Accordingly, it is impossible to illustrate dynamic features across the long term for environmental events. Simultaneously, using a dummy is not always ideal to capture spatial and temporal variation. Moreover, since previous studies have usually focused on the partial effects of a single type of disaster event, Caruso (2017) state that it would be insightful to also estimate the comprehensive consequence of different types of disasters. He generated the index of all kinds of disasters to test the integrated effects and confirmed that disasters in Latin America had affected the aggregation of human capital significantly over the last 100 years. Caruso's work provided an approach to estimate the all-inclusive influences of natural disasters. However, the establishment of the index was highly constrained by the availability of environmental records, and was not able to identify the actual districts affected by the disasters.

Commonly, extreme events such as floods and droughts are investigated the same way as other severe disasters, since both are likely to have significant impacts within a short time period. Such events are usually recognised as environmental shocks. There is influential research on the "judge-coordinated revolt" against the sovereign rule on the Nile, indicating that environmental shocks weakened the power of the secular monarchy and reduced the social stability in Egypt (Chaney, 2013). The strength of his work was that the author could employ records of the height of the Nile directly from historical documents. According to his results, he found that in the periods of flood, religious structures were incited to expand the scale of construction compared to secular structures, so that religious power was strengthened when destruction by nature overwhelmed human ability. Another study supported Chaney's conclusion by inspecting the relationship between precipitation and Roman emperor assassinations (Christian and Elbourne, 2017). More precisely, precipitation data was generated from tree-ring series constructed by palaeontologist, while the assassination records of ancient Rome were collected from historical documents. The result was that "lower rainfall in northern provinces, in Gaul and Germania, increases the likelihood of the Roman emperor assassination" (Christian and Elbourne, 2017). These studies have implied that the past stability of governance was determined by environmental conditions. More generally it becomes clear that research on ancient periods in these contexts crucially relies on not only the existence of historical records but also the progress of data reconstruction from other fields. Notably, the majority of studies have only been able to identify environmental shocks or climate changes by using temperature or precipitation across long time periods previously because decadal temperature series and precipitation series have not been generated gradually since the last decade. Economists have started to update the data in history based on the reconstruction work done by paleoclimate scientists and historians to estimate the long-run impact of environmental shocks on agriculture (Fiszbein, 2017; Christian and Fenske, 2015), demography (Fishback et al., 2011) and conflicts (Hsiang, Burke, and Miguel, 2013; Iyigun, Nunn, and Qian, 2017). However, the existence of historical records and reconstruction work from other scientists have not covered sufficient regions and period and thus far limited the historical contexts that could be explored. Moreover, available historical data usually constitute time series without spatial variance, and the time intervals between periods tends to be relatively long.

2.2.2 Data Collection in China

Following in the steps of foreign colleagues, many Chinese environmental historians have embarked on the reconstruction of paleoclimate. The temperature was the first factor being considered, since the methodology to reconstruct temperature has been well established and employed all over the world, and the historical information of temperature in China is straightforward to access. Combined with other climate shock indicators, such as droughts and floods, temperature has been employed to estimate the dynamic impact of dynasty cycles and civil conflicts in China. In order to generate ancient temperature series, tree-ring and ice-core measurements have been employed to reconstruct temperature series at the regional level (Yang, Braeuning, et al., 2002; Ge, Zheng, Hao, Shao, et al., 2010; Wang, Liu, and Wang, 2015). Later research (Zhang, Zhang, et al., 2007) that argued that geographic conditions would lead to heterogeneous spatial consequence isolated the temperature series for northern China and southern China. The temperature data has been widely employed in other research (Zhang, Zhang, et al., 2007; Chen, 2015; Zhang, Tian, et al., 2010). However, due to the advantage of historical records in China, the central concentration has always been on the reconstruction of environmental records from historical documents.

Studying historical documents is crucial in terms of reconstructing the environmental history of China. As a matter of fact, Zhang (1998) claimed that historical documents are the primary source of paleoclimate reconstruction. Therefore, it is necessary to collect and process textual information contained in these documents to enhance the credibility of the information source. Zhang, Ge, and Zheng (2002) analysed the environmental information from the content of different sources, including the local gazetteers and government archives of the Qing dynasty and Modern China, and proposed a criteria to validate the historical sources. Other studies have verified the validation of informal historical documents and non-literacy materials (Qian, 2014; Wang, 2013), and recommended that employing such sources requires a high level of historical knowledge as well as expertise across different fields. Based on the collection and reconstruction from validated sources, several achievements boosted the incentive further development in this regard (Li, 2017; Xia, 2015).

After the founding of the People's Republic of China, for the sake of national economic development and social security, governments at all levels and relevant scientific research institutions have highly valued the importance of the history of disasters in China. They have mobilised a large number of scientific research forces across the country. Information on waterlogs, droughts, earthquakes, storm surge and other disasters were collected, sorted and compiled at an unprecedented scale. There are now several works combining information from the official history and local topography. One of these is the compilation of Chinese earthquake history data. In 1954, the famous geologist and Vice President of the Chinese Academy of Sciences Li Siguang proposed to collect historical data on earthquakes in China. With the support of the Institute of Geophysics of the Chinese Academy of Sciences and relevant organisations, the staff of the third institute of the Institute of History of the Chinese Academy of Sciences (now the Institute of Modern History of the Chinese Academy of Social Sciences) has investigated from more than 8,000 official histories, alternative histories, notes, miscellaneous records, poetry collections, local chronicles, archives, and other historical documents. Nearly 10,000 earthquakes from 1177 BC to 1955 AD were recognised and compiled into "Chronological Table of China Earthquake Data" (History Group, Earthquake Working Committee, Chinese Academy of Sciences, 1956). From 1977, sponsored by the Chinese Academy of Sciences, the Chinese Academy of Social Sciences, and the National Seismological Bureau, government organised historians and seismologists to conduct a more extensive collection and excavation of historical earthquake materials, revised and expanded the existing chronology. It took 5 years to complete a five-volume compilation of “China Earthquake Historical Data” (Wang, Xie, and Wen, 1983) covering the period from ancient times to 1980.

Another spectrum is historical climate data. Due to the requirement of water conservancy construction, academies, Meteorological Bureau, Cultural History Museum, Water Conservancy Bureau and Academy of Agricultural Sciences of different regions started to engage with the compilation of wetness and dryness data from the 1950s. The data is collected from historical documents including local chronicles, “Mingshilu”, “Qingshilu”, “Qing history draft (in Chinese)”, as well as civil affairs data of each province, surveys of drought and flood disasters of each provincial meteorological bureau, and precipitation data from modern observations. In 1975, “Historical data on drought and flood in North of China and Northeast of China for nearly 500 years” was published, and then in 1978, “Historical climate data in East of China for nearly 500 years” (Xia, 2015).

Moreover, Zhang (2004b) organised many resources to reorganise the climate information in various local chronicles while supplementing nearly 1,0000 local chronicles, including those scattered in Taiwan and the Library of Congress. The result is “*Chinese Three Thousand years Meteorological Record Collection*” which is based on 8228 official histories and gazetteers, and contains over 220 thousand records from up to three thousand years ago (Zhang, 2005). This source is compiled in the same style and is widely accepted as the most comprehensive collection of official history. All the records collected in this book are well verified with detailed references.

The most recent data set is the collection that involves rainstorms, droughts, cold damage, and unbroken spells of wet weather, frosts, gale and hail, lightning strikes, dense fog, rainfall-triggered geologic hazard, and forest fire (Wen, 2006). The book used current administrative divisions as geographic units and collected historical data on different meteorological disasters from the pre-Qin period to 2000. The government of each province held the collective work, and individual volumes were published separately, such as "*China Meteorological Disaster Dictionary: Volume of Hunan*" (Zeng, 2006) which is the volume for Hunan province. It is considered as a practical reference book.

All these works based on official historical documents are not constructed using the entire set of available records. For example, although the compilation of wetness and dryness data used around 2100 local chronicles, the number of employed documents is still less than one-fifth of the total number of existing chronicles. "Chinese Three Thousand years Meteorological Record Collection" adopted 7855 kinds of literature. However, the private diaries and archives for climate data in the Qing dynasty were not involved due to different styles of records. "China Earthquake Historical Data" contained the largest number of categories for sources, but still could be further supplemented (Xia, 2015). Moreover, many works have collected data directly from historical documents but have failed to validate the different sources. For example, the criteria of the Qing dynasty to report a disaster differs from other periods. Fake reports of disasters due to corruption would affect the accuracy of environmental records, especially in Gansu province where severe corruption occurred around the 1780s (Ge, 2012). "*China Meteorological Disaster Dictionary (in Chinese)*" is representative of this issue. More precisely, the book combined the data from different sources but did not provide references in most volumes.

Apart from the research on historical documents, it should be noted that field research also occasionally plays a role. For example, "*The Great Flood of Chinese History (in Chinese)*" (Luo, 2006) summarised precipitation records, including floods of different channel segments from 1700s-1980 according to field surveys and field research. In this paper, field research will, however, not be considered as due to limitations of time.

Although there is large space for improvement in the data collection from historical evidence, previous efforts have provided adequate information for many topics. Data processing from historical records is important to be able to explore the role of the historical environment quantitatively (Xu, 2014). Early works have built several analytic tables for disaster frequency or spatial distribution statistics. The most outstanding collection in this regard is the "The Chronology of Natural Disasters, and Human Calamities in Chinese History (in Chinese)" (Chen, 1986), which constructed the disaster records from two aspects, namely natural disasters and human calamities. The type of disaster, location (mainly at the province level), and the time are included in the published tables. This kind of processing reclassifies original records and extracts useful information from the content to present the historical disasters more intuitively. However, the table does not provide enough quantitative indices to be used in quantitative analysis (Zhang and Crowley, 1989). In addition, it is not straightforward to identify the spatial distribution of disasters.

Recent works have delved into data processing that could be employed to present environmental information visually and digitally. One of the most iconic works is the "*Historical Seismological Atlas of China (in Chines)*" (State Seismological Bureau Institute of Geophysics and Fudan University Institute of Chinese History and Geography, 1990) . The atlas parameterised historical earthquake records from 2300 B.C. to 1911 A.D. and thus constitutes the most complete earthquake data compilation. If the magnitude of earthquake was greater than 4.75, data of the longitudes and latitudes of epicentre, earthquake magnitude, epicentre strength and the earthquake severity are available and presented in maps. It is hard to know details regarding the methodology of parameterisation, the achievement is nevertheless considered credible since the work was conducted by authorised organisations in China.

Another prominent data set is "*The Atlas of Wetness and Dryness Distribution in China in the Past Five Hundred Years (in Chinese)*"(China Meteorological Administration Institute of Meteorology, 1981) published in the 1970s. This atlas contains 120 climate series of the grade of wetness and dryness from 1470 to 1979 converted from the compilation of wetness and dryness data. The index of wetness and dryness grade is parameterised by the description of wetness and dryness in historical records when there is no record for accurate precipitation, since most of the actual records of precipitation exist after 1911. Additionally, before the Republic of China, this atlas parameterised records from some 2100 local gazetteers from the over 10 thousand existing local gazetteers. Since the historical grade of precipitation is based on analyses of the description for environmental events in the records, it is possible to contain perspective errors of the severity of the included environmental events.

There are other methods to parameterise historical climate records apart from the description reclassification. One approach is to count the frequency of environmental events. Zhang (1984) defined a winter-temperature index:

$$T = \{n_1 + 0.3[10 - (n_1 + n_2)]\}$$

where n_1 and n_2 are years of cold and mild winter for one decade, respectively. This method can only be used for the long term, ignoring any annual variation. An example is the study by Zhang (1984) which derived a frequency series of dust falls from historical documents. The earliest dust fall was recorded in 1150 B.C., but the records have become more frequent and viable for processing only for the last 1000 years. The year in which a dust rain is recorded is called a “dust rain year”, although several dust rains could occur within one year. The number of dust rain years in different decades has been calculated, but the frequencies form a series at most for 500 years. This method actually sacrifices some information on time variation in order to imply probability of the occurrence of environmental events for a given region, and is only available at the decade level.

Zheng, Zhang, and Zhou (1993) adopted a different method with the number of flood counties minus those affected by drought. As the number of counties can change considerably, the differences may not represent the magnitude of the flood or drought. However, Zheng, Zhang, and Zhou (1993) confirmed the reliability of the number of flood and drought counties in the Ming and Qing dynasties, since the number of counties across these two dynasties does not vary significantly. Besides, due to the constant compilation of local chronicles, variation in the frequency of environmental events could exclude the influences from increasing historical documents. In order to validate the reliability of his new approach for dryness/wetness index, Zheng parameterised the actual records of precipitation into the grade of precipitation in the Beijing region from 1724 to 1950 and compared the grade of precipitation with the grade of dryness and wetness of Beijing. The comparison showed that the grade from the number of drought/flood counties does not differ significantly from the grade generated from actual precipitation records. Therefore, the approach of the number of counties that were affected by environmental events is reliable as a parameterisation of the historical records.

To summarise, using both the number of years affected by environmental events and the number of counties affected by environmental events are eligible methods to process data during the Ming and Qing dynasty. "*The Atlas of Wetness and Dryness Distribution in China in the Past Five Hundred Years (in Chinese)*"(China Meteorological Administration Institute of Meteorology, 1981) provides a prominent application in an influential study of the periods after 1470 (Bai and Kung, 2011). In this respect, the historical climate records of the Ming and Qing dynasty indicate a high potential for broader quantitative analysis. Frequency indices for this period can be adopted to build a grade index which could be validated by the existing grade index. However, current methods to process the data do not capture the complete information from historical records. The grade index from the atlas ranks the level of severity description to determine the severity level of environmental events, but ignores severity information suggested by the number of records (Zheng, Zhang, and Zhou, 1993) especially when the historical records are considered sufficient. Besides, the grade index combines records of floods and droughts to indicate the wetness or dryness of a given region. As a result, the impact of floods or droughts alone cannot be estimated separately by a quantitative approach. Moreover, the number of years affected by environmental events and the number of counties affected by environmental events can denote the severity quantitatively to some extent. However, the former can only be employed at decadal frequency and the latter cannot be adopted at the county level. Consequently, annual county-level environmental data cannot be generated by the previous approach from current sources, unless there are town level collections of records. Based on existing sources, it is necessary to develop a new approach to integrate different sources to construct annual county-level data set.

2.2.3 Existing Database

The development of the digitalisation of historical documents has experienced a great revolution in China from the 1980s. Following the concept of Chu (1926), the digitalisation process in the 1980s has focused on the parameterisation of disaster information to construct climate indices and eliminate any non-uniformity. Chen (1987) argue that the purpose of digitalisation for historical climate records is to convert qualitative information into quantitative time series. Since the stylistic rules and layouts of sources are different, it is necessary to ensure that all the records are adopted under the same criteria to keep uniformity (Zheng, Ge, Fang, et al., 2007). After the 1990s, the process of digitalisation and the establishment of databases has involved more information from the historical documents. Zheng, Zhang, and Jian (1992) proposed a digitalisation approach based on the historical archive of the Republic of China. They extracted necessary fields from every single record in the historical documents and coded these fields. The main fields contain the location of the disaster, its range, time, weather condition, severity, influence, as well as the impact on the output of agriculture, the social response, etc.

Later in the 21st century, specialised research on the historical disaster database has developed. The database of attributes and database of spatial characteristics were linked as a part of disaster information system. Furthermore, Zheng, Hao, and Di (2002) designed a database for the climate variation over the past 2000 years by combining different environmental indices and historical documents into a single system. According to their work, data and documents were managed together and could be used to implement spatial analysis. What is more, China Meteorological Administration developed the fundamental data system of historical climate based on the work of "*Chinese Three Thousand years Meteorological Records Collection*" (Zhang, 2005). It should also be noted that recent studies of the environmental database have focused on the structure of data rather than the parameterisation of historical information. Additionally, these databases required a considerable investment of time and manpower since the digitalisation process was conducted manually, which would be impossible for an individual researcher to undertake. However, only some results based on these databases are shared with the public (Zhang, 2005; Xia, 2015), which severely limits its academic use.

There are number of other databases being constructed at present. One of the most anticipated of these is the integrated information database of disasters and famines in the Qing dynasty proposed by the Qing History Institute of Renmin University in 2014 (Xia, 2015). The goal of this venture is to construct a dynamic information system that can be shared and updated when new findings are discovered and can cover as many as possible records for the Qing dynasty. It is expected that this database will connect the historical records across different sources and to inspire further investigations on the Qing dynasty. While the database is not completed yet, it will likely be a better source of historical records according to current progress (National Office for Philosophy and Social Science, 2018). Thus, the parametrisation approach and the digitalisation process to input data are still attractive tools to be able to facilitate better data construction in the environmental historical field of China.

2.3 Data Sources and Methodology

2.3.1 Data Sources

The main source of environmental records in this study is "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*". This source systematically collected the records of weather conditions, climate status, atmospheric physical phenomenon across entire China from 1300s B.C. to 1911 A.D. The book was published in analytic style and divided into four volumes. Records for the Ming and Qing dynasties are organised in Volume 2, Volume 3 and Volume 4, which are the ones mainly employed in this research. The original source contains a literal narrative of environmental factors, including flood, drought, rainfall, snow, cold, warm, ice, freeze, frost, hail, wind, haze, storm surge, lightning, time and location for the occurrence. Besides, the damage of natural disaster, severity, relevant relief, famine and disease which are related to the environmental shocks are also carefully collected (Zhang, 2005). What is more, records from the Ming dynasty (1368-1911) are separated by each province and are linked with modern county names in order to facilitate the referencing. This work examined 8228 types of Chinese official history, and 7835 types of these were adopted to construct the environmental chronicle, among which 7713 are local gazetteers and 28 are local biographies. Except for the official records, there are some records stored in the personal diaries and palace archives of the Qing dynasty, however their styles differ significantly with official history. These data are not included in this source. Thus one can argue that records in this source are all in the same style with the same criteria and thus the uniformity of this source should not be a significant issue.

The reliability of records in "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*" can be confirmed from several directions. First of all, original sources of this work covered most of official history, which represents one of the best collections of local gazetteers among the rest of current existing data sets. Secondly, this book extracted over 200 thousand pieces of original records, and each of these records was precisely referenced. Records were compared among multiple sources to find the correct version. If the same fact was recorded in different documents, this book would cite the earliest origin. What is more, errors in historical records were corrected as well. Finally, this was a project supported and accepted by the Ministry of Science and Technology Department of China (Zhang, 2005) and has been appreciated by many influential academics (Xia, 2015; Li, 2017). As a result, records in this book will be employed directly without any modification.

Although "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*" is one of the best collections of official history, records in this collection need to be improved to reveal more specific features of environmental conditions in the historical context. In this regard, it is known that records in personal diaries and palace archives were not included in this book. That is, this book is more conservative in this vein by including only original sources and ignoring the record if it raised issues of non-uniformity. Therefore, another source, "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*" is here adopted to supplement the records for this research.

The supplementary sources are from "*China Meteorological Disaster Dictionary (in Chinese)*" (Wen, 2006). This project was the most recent national scale compilation of historical data. Since each province hosted the its own collection, the quality and criteria used for the identification of environmental information differed a lot across different volumes. Compared with "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*" which can be considered as a conservative source, "*China Meteorological Disaster Dictionary (in Chinese)*" does not provide references for every single record except for the volume of the Shaanxi province. Nevertheless, the supplementary sources do not consider the uniformity but pools different types of records together into an annual style. Thus, it is necessary to estimate the reliability of each volume to modify the records and eliminate the non-uniformity to an acceptable extent.
The situation of each volume is shown in Table 2.1.

Table 2.1: Information of each volume in *China Meteorological Disaster Dictionary (in Chinese)*

Province	Compilation time	Reference information	Volume	Other information
Anhui	2001-2006	10 secondary sources	382 pages, 600 thousand words	Records are from Ming history, Qing history, local history, old scriptures and meteorological research, while the detailed list of historical documents is not given. Data without confirmation are cited.
Beijing	2001-2004	13 secondary sources	285 pages	There is no detail about the types of historical documents employed.
Chongqing	2001-2007	27 secondary sources	382 pages	
Fujian	2001-2007	15 secondary sources	392 pages	Local gazetteers are original sources.
Gansu	2001-2003	39 secondary sources	450 pages	
Guangdong	2001-2006	7 secondary sources	351 pages	
Guangxi	2001-2007	27 secondary sources	413 pages	Original sources contain literature, local gazetteers, journals, yearbooks and books
Guizhou	2001-2004		297 pages	
Hainan		13 secondary sources	284 pages	Local histories are original sources
Hebei	2001-2005		372 pages	Over 100 documents were investigated but the references of all these materials are not provided
He'nan	2001-2003		398 pages	Local gazetteers and observations from meteorology equipment are claimed to be the original sources.
Heilongjiang		38 secondary sources	329 pages	Original sources are local gazetteers and historical archives
Hubei	2001-2007	89 gazetteers and 26 secondary sources	426 pages	
Hu'nan	2001-2005	65 gazetteers and 70 secondary sources	509 pages	
Jilin	2001-2008	4 local gazetteers and 11 secondary sources	394 pages	
Jiansu		69 local gazetteers	249 pages	
Jiangxi	2001-2002	96 local gazetteers and 27 other sources	470 pages	
Liaoning	2001-2004		322 pages	
Neimenggu			384 pages	
Ningxia		16 sources	273 pages	
Qinghai	2001-2005	17 sources	256 pages	
Shandong	2001-2005	29 sources	648 pages	Uncertain records are cited.

Shanxi	2001-2005		919 pages	Every record is identified by the original source.
Shaanxi	2000-2005		211 pages	
Sichuan	2001-2005	27 sources	600 pages	
Tianjin	2003-2008	24 sources	262 pages	
Xizang	2001-2006	15 sources	217 pages	
Xinjiang	2001-2006	8 sources	340 pages	
Yunnan	2002-2005	22 sources	540 pages	
Zhejiang	2001-2003		287 pages	

One should note that historical records in all of these volumes were collected from the historical documents by default, but the majority of these works do not provide detailed information of their historical sources. References in the bibliography are previous research or some internal materials which are not published. Based on the situation of each volume, it is possible to evaluate the effort of each volume and then to assess the quality of each work.

For the environmental data, records from the Ming and Qing dynasty (1368-1911) were employed and were categorised into drought, flood, cold, and wind event. Based on the quality assessment of supplementary sources, records were parameterised and combined to build new indices of environmental events via the frequency approach.

To implement visualisation of historical environment, relevant research needs to linked the geographical attributes to the environmental data. Therefore, a simple GIS database can be created to manage the geographic characteristics of the historical environment and to make maps of the historical environmental events distribution. The historical geographical attributes come from a public GIS database called "*The China Historical Geographic Information System*" (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). This is an open source project launched in 2001 and held by Fudan University and Harvard University. Data can be obtained from the main project website or the DataVerse Repository at Harvard University. This project provides a common framework for georeferencing historical materials and is the most frequently used GIS platform for spatio-temporal analysis of historical China (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).Since Version 6 of this GIS system is the newest, the administrative boundaries of 1911 in version 6 are employed here. The reason to adopt boundaries of 1911 is that differences between the 1911 boundaries of counties and modern boundaries of counties are not significant, and the coverage of 1911 territories is the largest during the Ming and Qing dynasty, which facilitates the work to identify locations in different sources to the same associated geographic unit.

2.3.2 Digitalisation Process

Our main sources of historical environmental records are available only in paper form, and there is no access to editable digital archives. Hence, we needed to find another way to transfer paper books into digital files. As manual typing would have taken too much time and is beyond the time limits of this research, we use Artificial Intelligence (AI) techniques for optical character recognition (OCR) to capture the characters from the books. After that, it is possible to edit the text into the form that is easier to code.

To extract the environmental information from historical texts, the time, location, and the type of relevant environmental events were marked to be recognised by the programme. The challenge is that the names of the location and the type of environmental event are not standardised. What is more, the structure of the "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*" and the "*China Meteorological Disaster Dictionary (in Chinese)*" are different. Therefore, methods to treat these two sources are kept separately. Developing editable digital text from paper books can also mean some loss of accuracy, which may lead to misclassification of the environmental information. For example, the word "夏旱" which means drought in summer, has been recognised as "夏早" which means summer comes earlier. In this regard, we double-checked with original sources, a time and patience consuming task. Fortunately, AI OCR provides a high accuracy, which can be verified using spot tests. In the spot test, 100 pages were randomly selected from the whole book to compare with the corresponding digital versions and the accuracy was around 99% regardless of some missing punctuation.

Process for "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*"

The two primary sources of environmental records are organised in different ways. In "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*", every single record is identified with modern county names. What is more, since duplicate records which refer to the same event have already been eliminated, it is rational to ignore the overestimation of record frequencies. However, the categories of environmental events are not organised in this source.

In order to classify the type of environmental impact for each record, it is necessary to extract the environmentally relevant description from the content based on the analysis of historical context. Word frequency analysis was employed to collect the environmental description, and relevant words are considered if their frequency is over 5. In total, there were 724 words recognised as containing some description of meteorological related content. Due to constraints of meteorological knowledge in the historical context, the categories were limited to events of flood, drought, wind, and cold. For example, from the list of 724 meteorological description words, 233 words were recognised as the descriptions for flood or water logging. The top 30 frequency of these words are listed in Table 2.2.

Therefore, any record containing flood description words that were recognised from the word frequency analysis are considered as the record for flood events. The flood records and drought records can overlap within a year and location since the individual records could contain references for both types. For example, in the record of 1911 for Yunnan province, “Kunming city experienced drought in the summer, and then flood······”, drought and flood occurred within one year in the same county.

As for the descriptive words of other categories, 59 words were recognised as drought, 62 words were considered as cold, and 42 words were classified as wind. The rest of the description words are for other abnormal events that could not be clearly identified.

Table 2.2: Word Frequency for Flood Description Words

Word	Frequency
大水 (Heavy Water lodge)	14226
大雨 (Heavy Rain)	5770
水 (Water lodge)	3554
霪雨 (Continuous Rain)	2308
雨 (Rain)	1818
水灾 (Water lodge disaster)	1411
溢 (Overflow)	1169
淹没 (Submerge)	1009
大风雨 (Heavy rain with wind)	991
水深 (Deep water)	959
雨水 (Rainy water)	916
河决 (River dyke breaching)	852
雷雨 (Rain with lightning)	772
漂没 (Submerge)	733
决 (Dyke breaching)	670
暴涨 (Skyrocketing)	587
潦 (Humidity)	557
洪水 (Flood)	553
风雨 (Rain and wind)	550
涨 (River rises)	525
霖雨 (continuous heavy rain)	519
水涨 (Water rises)	422
海溢 (Overflow of sea)	379
冲决 (Breaching the dyke)	369
泛溢 (General overflow)	365
水溢 (Water overflow)	361
漂流 (Floating)	304
大雨如注 (Severe heavy rain)	303
暴雨 (Rainstorm)	283
河溢 (River overflow)	250

Data sources and notes

* Words were collected from "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*" (Zhang, 2004b).

Process for "*China Meteorological Disaster Dictionary (in Chinese)*"

The "*China Meteorological Disaster Dictionary (in Chinese)*" focuses more on the boundaries of provinces and environmental events. However, the "*China Meteorological Disaster Dictionary (in Chinese)*" does not identify the county-level distribution, and the categories of disasters are not consistent across provinces. There are 31 volumes of this source. The categories of disasters for each volume are shown in Table 2.3.

Xizang province is dropped since the records in the Ming and Qing dynasties are not available. Among the remaining provinces, a number of categories of events are consistent, and thus can be used to construct a databases covering the entire China. More precisely, records were classified into four types of environmental events - drought, flood, cold, and wind. Specifically, a drought event involves a number of categories, including the words "drought" or "high temperature", while a flood is identified with the words "rainstorm", "flood" or "water logging". Cold involves all the categories including the word "cold" or "snow". Wind involves all the categories including the word "wind", "hurricane", "typhoon" or "tornado".

According to the sources of "*China Meteorological Disaster Dictionary (in Chinese)*", records are listed in each year for each category in each volume. Therefore, in order to code from these texts, each category of each province was separated into an individual text file. Then, it is possible to extract the time and location from the digital text by using AI techniques of lexical analysis. Each location was identified with the standardised county based on the county boundaries of 1911.

Table 2.3: Categories of Disasters for Every Province

Anhui	Drought, frost, hailstone, lightning strike, rainstorm and flood, rainy, storm and mudslide, snow, wind and typhoon
Beijing	Drought, hailstone, lightning strike, rainstorm and flood, cold and snow, fog, high temperature, mudslide, wind and sandstorm
Chongqing	Rainstorm and flood, rainy, mudslide, drought and high temperature, hailstorm and wind, lightning, snow and cold
Fujian	Drought, frost, hailstone, hurricane, rainstorm, storm surge, typhoon
Gansu	Drought, rainstorm and flood, mudslide, wind and sandstorm, abnormal climate and frost, hailstone and lightning
Guangdong	Drought, storm surge, flood, frost and cold, strong convection,
Guangxi	Drought, rainstorm and flood, frost could and snow, hailstone wind and lightning, typhoon and hurricane
Guizhou	Drought, rainstorm and flood, mudslide, hailstone wind and lightning, frost and snow
Hainan	Hailstone, rainstorm and flood, Hurricane, cold, drought disaster, tornado
Hebei	Drought, hailstone, rainstorm and flood, fog, high temperature, lightning, storm surge, frost and cold, wind sandstorm and tornado
Heilongjiang	Drought, hailstone, rainstorm and flood, snow, cold frost, wind and tornado
He'nan	Drought, frost, hailstone, snow, lightning, glaze, water lodge, wind
Shanghai	Drought, hailstone, rainstorm and flood, cold and snow, fog, lightning, typhoon, tornado, wind
Hubei	Drought, rainy, strong convection, cold, rainstorm and flood,
Hu'nan	Drought, frost, rainstorm and flood, fog, lightning, cold and rainy, wind and hailstone,
Jiangsu	Rainstorm and flood, rainy, drought and high temperature, hurricane, strong convection, wind, cold snow and frost, plum rains,
Jiangxi	Rainstorm and flood, fog, drought and high temperature, lightning, cold, wind and hailstone, glaze snow and frost, tornado and hurricane

Jilin	Drought, hailstone, Rainstorm and flood, frost and cold, wind,
Shanxi	Frost, hailstone, rainstorm and flood, rainy, snow, fog, high temperature, lightning, cold, tornado, glaze, wind, sandstorm
Liaoning	Drought, hailstone, rainstorm and flood, wind, cold snow and frost
Shandong	Drought, hailstone, rainstorm and flood, snow, fog, wind and sandstorm, lightning, storm surge, cold, tornado,
Neimenggu	Drought, frost, hailstone, snow, water lodge, wind,
Ningxia	Drought, hailstone, wind and sandstorm, frost and cold, water lodge,
Qinghai	Drought, hailstone, snow, cold, wind sandstorm and tornado, flood and mudslide
Shaanxi (Shan)	Drought, rainstorm and flood, cold and frost, wind and hailstone, snow fog and lightning
Sichuan	Drought, hailstone, rainstorm and flood, rainy, lightning, cold, wind
Tianjin	Drought, hailstone, rainstorm and flood, wind tornado and sandstorm
Xinjiang	Drought, frost, hailstone, snow, flood, wind

Data sources and notes

* Categories were summarised from "*China Meteorological Disaster Dictionary (in Chinese)*" (Wen, 2006).

There are several issues for county identification. The first is that records are collected from multiple historical documents in sources, and the name of each county varies. Therefore, every county name in the records was edited to correspond with the county name of 1911. Tracing history names of each county, it is possible to establish the connection between the name in records with the name of 1911 for most of the cases. However, some records contain confusing location names. To be specific, for some names it is hard to tell whether they are the names of administrative units or the names of some large areas. For instance, "陇西" (the Pinyin is "long xi") in Chinese history could represent the west of Gansu province and the south of Gansu province or the name of Longxi county. To deal with such confusion, it is necessary to return to the original records to determine a certain meaning. If the name is mentioned alongside other county names, it is likely also a county name, such as in the drought record in 1759 of the Gansu province:

"······August, relief released on Gansu Gaolan, Jin county, Jinyuan, Hezhou, Didao, Weiyuan, Longxi, Ningyuan, ······"

Meanwhile, if the name is used in parallel with other names of a higher spatial level, it is also possible to it is similarly referring to that spatial level. Fortunately, in all records "long xi" could always be recognised as a county name. Besides, name of the prefecture and name of the prefecture capital are usually the same, and sometimes records would use the prefecture name to refer to the prefecture capital. For example, in the record of Gansu province "······two prefectures of Lin, Gong (Longxi) were deteriorated ······", since the capital of Gongchang prefecture (the abbreviation form is "Gong") is Longxi county, it is possible to identify this record "Gong" to correspond with the Longxi county. While in another case, in the record of Gansu province in 1562, “November, drought in the counties of six prefectures of Shaanxi province, Xi, Yan, Ping, Lin, Gong (the area on the east of the river), Han······”, the “Gong” should be recognised as the area on the east of the river within the boundary of the whole prefecture. Unfortunately, not all the contexts provide enough information for such identification. Therefore, to reduce the overestimation of the number of recorded counties, the prefecture name is assumed to refer to the prefecture capital, since at least the prefecture capital would maintain such record. The rest of the possible affected counties are omitted if there is no other supplementary information.

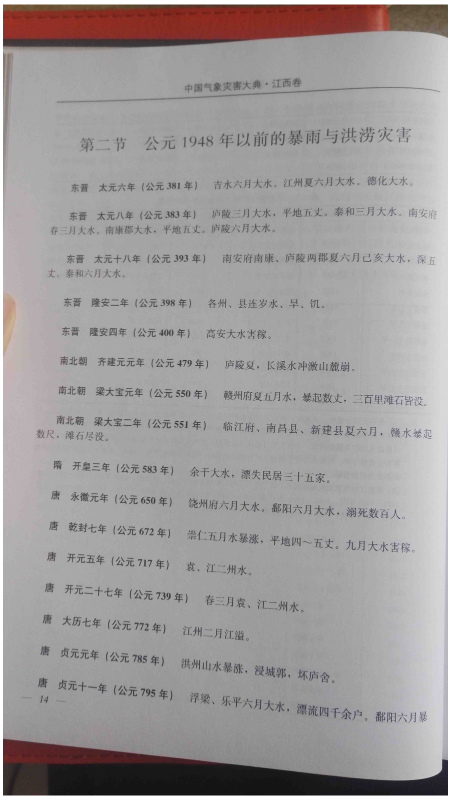
Apart from when the name possibly represents multiple levels of the locations (county, prefecture, and area), the most complicated task is to identify those names without clear connections with current knowledge. Because of the constraints of historical knowledge and skills, the names which could not be identified as validated locations were omitted in this work, such as “bin hai” (which was mentioned along with other county names, for example in 1703, “······Wuding, Binhai and other counties were severely flooding······”) in Shandong province, Guangdong province, and also Fujian province. Additionally, town names were counted as the counts of counties where these towns are located only if the county name is not referenced elsewhere. For instance, in the flood record of Zhejiang province in 1898, “10th May, in Shouchang, from Rendu to Yuejiazhuang was raining heavily···.”, the name Yuejiazhuang is a town name but the county name Shouchang is already referenced before. Thus the town name was not counted in this case. As for the duplicate town names across different counties, if the county name is not referenced and there is no other supplementary sources to identify the location, the records of these towns were omitted. Therefore, based on the maps of the Qing dynasty in 1911 and modern China, all names can be related to current coordinates.

Following the process just outlined, one can generate annual county-level environmental data sets based on the two different sources mentioned.

Guideline of the Operation

The original sources are paper version, for which the content is shown in Figure 2.1. Each page is scanned as an individual picture, as shown in Figure 2.2, and all these pictures are arranged into the correct order. The Application Programming Interface (API) of optical character recognition (Baidu, 2019) provided by Baidu is employed to convert these pictures into editable text files automatically. Since each text file represents an individual page of original sources, these files can merged according to different requirements for the next operation.

The structures of the two sources are different. According to Figure 2.2, "*China Meteorological Disaster Dictionary (in Chinese)*" arranges historical records in temporal order for each category of environmental disasters of every province. For example, from page 14 to page 78 in "*China Meteorological Disaster Dictionary: Volume Jiangxi (in Chinese)*", flood records of Jiangxi province before 1948 are listed in annual order. Therefore, text files conversions from those pages can be merged into individual files to represent the flood records of the Jiangxi province. The structure of each record can described as "the name of dynasty" + "the name of year" + "(the numeric year in solar calendar)" + "the content of the record". Basically, the content of the record includes the record of the location, category



中国气象灾害大典·江西卷

第二节 公元 1948 年以前的暴雨与洪涝灾害

东晋 太元六年（公元 381 年） 吉水六月大水。江州夏六月大水。德化大水。

东晋 太元八年（公元 383 年） 庐陵三月大水，平地五丈。泰和三月大水。南安府春三月大水。南康郡大水，平地五丈。庐陵六月大水。

东晋 太元十八年（公元 393 年） 南安府南康、庐陵两郡夏六月己亥大水，深五丈。泰和六月大水。

东晋 隆安二年（公元 398 年） 各州、县连岁水、旱、饥。

东晋 隆安四年（公元 400 年） 高安大水害稼。

南北朝 齐建元元年（公元 479 年） 庐陵夏，长溪水冲激山麓崩。

南北朝 梁大宝元年（公元 550 年） 赣州府夏五月水，暴起数丈，三百里滩石皆没。

南北朝 梁大宝二年（公元 551 年） 临江府、南昌县、新建县夏六月，赣水暴起数尺，滩石尽没。

隋 开皇三年（公元 583 年） 余干大水，漂失民居三十五家。

唐 永徽元年（公元 650 年） 饶州府六月大水。鄱阳六月大水，溺死数百人。

唐 乾封七年（公元 672 年） 崇仁五月水暴涨，平地四～五丈。九月大水害稼。

唐 开元五年（公元 717 年） 袁、江二州水。

唐 开元二十七年（公元 739 年） 春三月袁、江二州水。

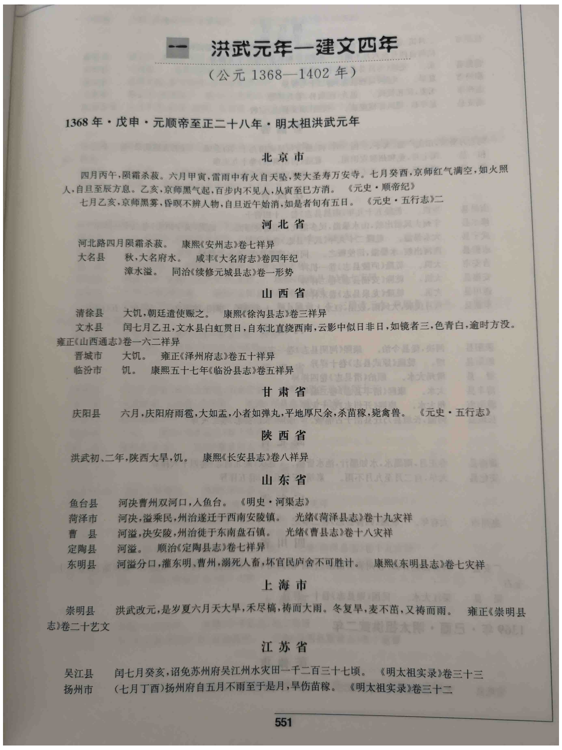
唐 大历七年（公元 772 年） 江州二月江溢。

唐 贞元元年（公元 785 年） 洪州山水暴涨，浸城郭，坏庐舍。

唐 贞元十一年（公元 795 年） 浮梁、乐平六月大水，漂流四千余户。鄱阳六月暴

— 14 —

(a) China Meteorological Disaster Dictionary



一 洪武元年—建文四年

（公元 1368—1402 年）

1368 年·戊申·元顺帝至正二十八年·明太祖洪武元年

北京市

四月丙午，陨霜杀菽。六月甲寅，雷雨中有火自天坠，焚大圣寿万安寺。七月癸酉，京师红气满空，如火照人，自旦至辰方息。乙亥，京师黑气起，百步内不见人，从寅至巳方消。《元史·顺帝纪》

七月乙亥，京师黑雾，昏暝不辨人物，自旦近午始消，如是者旬有五日。《元史·五行志》二

河北省

河北路四月陨霜杀菽。 康熙《安州志》卷七祥异

大名县 秋，大名府水。 咸丰《大名府志》卷四年纪

漳水溢。 同治《续修元城县志》卷一形势

山西省

清徐县 大饥，朝廷遣使赈之。 康熙《徐沟县志》卷三祥异

文水县 闰七月乙丑，文水县白虹贯日，自东北直绕西南，云影中似日非日，如镜者三，色青白，逾时方没。雍正《山西通志》卷一六二祥异

晋城市 大饥。 雍正《泽州府志》卷五十祥异

临汾市 饥。 康熙五十七年《临汾县志》卷五祥异

甘肃省

庆阳县 六月，庆阳府雨雹，大如盂，小者如弹丸，平地厚尺余，杀苗稼，毙禽兽。《元史·五行志》

陕西省

洪武初、二年，陕西大旱，饥。 康熙《长安县志》卷八祥异

山东省

鱼台县 河决曹州双河口，入鱼台。《明史·河渠志》

菏泽市 河决，溢乘民，州治遂迁于西南安陵镇。 光绪《菏泽县志》卷十九灾祥

曹县 河溢，决安陵，州治徙于东南盘石镇。 光绪《曹县志》卷十八灾祥

定陶县 河溢。 顺治《定陶县志》卷七祥异

东明县 河溢分口，灌东明、曹州，溺死人畜，坏官民庐舍不可胜计。 康熙《东明县志》卷七灾祥

上海市

崇明县 洪武改元，是岁夏六月天大旱，禾尽槁，祷而大雨。冬复旱，麦不苗，又祷而雨。 雍正《崇明县志》卷二十艺文

江苏省

吴江县 闰七月癸亥，诏免苏州府吴江州水灾田一千二百三十七顷。《明太祖实录》卷三十三

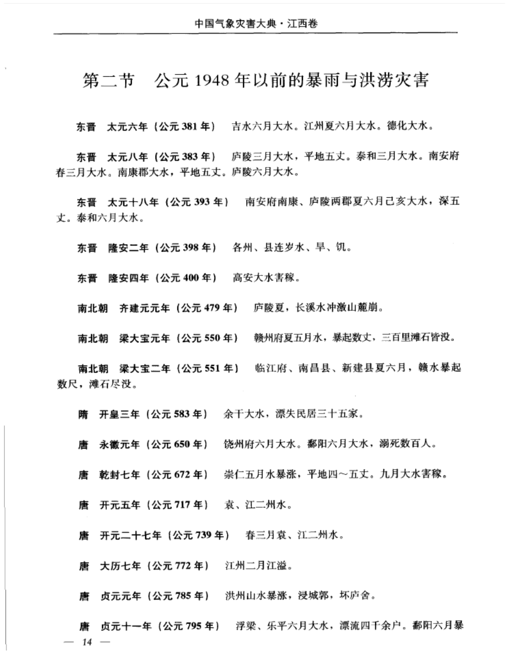
扬州市 （七月丁酉）扬州府自五月不雨至于是月，旱伤苗稼。《明太祖实录》卷三十二

551

(b) Chinese Three Thousand years Meteorological Records Collection

Figure 2.1: Photograph of Source

of event, and precise time (which is not always available). Each record was re-arranged into a single line in the text file, to be used for further lexical analysis. In order to facilitate the operation, data in text files are cleaned by coding in Python (version 3.7.2). As a result, only the numeric year according to the solar calendar and the content of the record remain. Thus, the style of these records is converted into this form: “year” + “the content of record”. Each record accounts for one line in the text file. After the preparation above, it is possible to apply the lexical analysis provided by the Baidu Cloud (Baidu, 2019). According to the Baidu



中国气象灾害大典・江西卷

第二节 公元 1948 年以前的暴雨与洪涝灾害

东晋 太元六年（公元 381 年） 吉水六月大水。江州夏六月大水。德化大水。

东晋 太元八年（公元 383 年） 庐陵三月大水，平地五丈。泰和三月大水。南安府春三月大水。南康郡大水，平地五丈。庐陵六月大水。

东晋 太元十八年（公元 393 年） 南安府南康、庐陵两郡夏六月己亥大水，深五丈。泰和六月大水。

东晋 隆安二年（公元 398 年） 各州、县连岁水、旱、饥。

东晋 隆安四年（公元 400 年） 高安大水害稼。

南北朝 齐建元元年（公元 479 年） 庐陵夏，长溪水冲激山麓崩。

南北朝 梁大宝元年（公元 550 年） 赣州府夏五月水，暴起数丈，三百里滩石皆没。

南北朝 梁大宝二年（公元 551 年） 临江府、南昌县、新建县夏六月，赣水暴起数尺，滩石尽没。

隋 开皇三年（公元 583 年） 余干大水，漂失民居三十五家。

唐 永徽元年（公元 650 年） 饶州府六月大水。鄱阳六月大水，溺死数百人。

唐 乾封七年（公元 672 年） 崇仁五月水暴涨，平地四～五丈。九月大水害稼。

唐 开元五年（公元 717 年） 袁、江二州水。

唐 开元二十七年（公元 739 年） 春三月袁、江二州水。

唐 大历七年（公元 772 年） 江州二月江溢。

唐 贞元元年（公元 785 年） 洪州山水暴涨，浸城郭，坏庐舍。

唐 贞元十一年（公元 795 年） 浮梁、乐平六月大水，漂流四千余户。鄱阳六月暴

— 14 —

(a) China Meteorological Disaster Dictionary

明代气象记录 洪武元年—建文四年

一 洪武元年—建文四年

（公元 1368—1402 年）

1368 年・戊申・元顺帝至正二十八年・明太祖洪武元年

北 京 市

四月丙午，陨霜杀菽。六月甲寅，雷雨中有火自天坠，焚大圣寿万安寺。七月癸酉，京师红气满空，如火照人，自旦至辰方息。乙亥，京师黑气起，百步内不见人，从寅至巳方消。《元史・顺帝纪》十

六月甲寅，大都大圣寿万安寺灾。是日未时，雷雨中有火自空而下，其殿脊东鳌鱼口火焰出，佛身上亦火起。帝闻之泣下，亟命百官救护，惟东西二影堂神主及宝玩器物得免，余皆焚毁。《元史・五行志》二

七月乙亥，京师黑雾，昏暝不辨人物，自旦近午始消，如是者旬有五日。《元史・五行志》二

河 北 省

河北路四月陨霜杀菽。 康熙《安州志》卷七祥异

大名县 秋，大名府水。 咸丰《大名府志》卷四年纪

漳水溢。 同治《续修元城县志》卷一形势

山 西 省

清徐县 大饥，朝廷遣使赈之。 康熙《徐沟县志》卷三祥异

文水县 闰七月乙丑，文水县白虹贯日，自东北直绕西南，云影中似日非日，如镜者三，色青白，逾时方没。雍正《山西通志》卷一六二祥异

晋城市 大饥。 雍正《泽州府志》卷五十祥异

临汾市 饥。 康熙五十七年《临汾县志》卷五祥异

陕 西 省

洪武初、二年，陕西大旱，饥。 康熙《长安县志》卷八祥异

甘 肃 省

庆阳县 六月，庆阳府雨雹，大如盂，小者如弹丸，平地厚尺余，杀苗稼，毙禽兽。《元史・五行志》二

山 东 省

鱼台县 河决曹州双河口，入鱼台。《明史・河渠志》一

曹 县 河溢，曹州徙治安陵镇。 嘉靖《山东通志》卷三十九灾祥

定陶县 河溢。 顺治《定陶县志》卷七祥异

东明县 河溢分口，灌东明、曹州，溺死人畜，坏官民庐舍不可胜计。 康熙《东明县志》卷七灾祥

上 海 市

崇明县 洪武改元，是岁夏六月，天大旱，禾尽槁，祷而大雨。冬复旱，麦不苗，又祷而雨。 雍正《崇明县志》卷二十艺文

江 苏 省

南京市 正月四日，于钟山之阳设坛备仪，昭告上帝。自壬戌以来，连日雨雪阴沍，至正月朔旦雪霁，粤三

(b) Chinese Three Thousand years Meteorological Records Collection

Figure 2.2: Scan of Source

API of lexical analysis, content in the text file can be analysed line by line. The programme inputs one sentence automatically and returns the attribute of each word in the sentence. The results of lexical analysis contain the characteristic or property of a certain word, which could indicate whether a word represents the time or location of the event. By coding these results, only words representing location will be extracted for each line. Accordingly, the final record can reconstructed into the new form: “year” + “locations”. Therefore, each individual file contains all available records of a specific category in a specific province in this new form.

At the same time, in “Chinese Three Thousand years Meteorological Records Collection”, records are collected in this manner: “modern county name” + “content of record" + "source reference". For example, from page 551 to page 552, all records in 1368 are listed, including different kinds of environmental events and different locations. Therefore, text files converted from pages 551 to 552 can be merged into one file to represent all these records within the boundary of the whole country in 1368. Moreover, the data in the text file needs to be cleaned as well. Since only the content of the record includes the information of the category of environmental events, the source reference can be removed. By matching the description extracted from word frequency analysis with the content of the record, records of flood, wind, cold, and drought events were recognised separately and then re-arranged into different files. Therefore, each file contains all available records of a specific category in a specific year in the form "modern county name" + "content of record". Furthermore, since each file is classified into a specific category already, this information in the content of a record is no longer necessary for the next step of the analysis. The final file includes all available county names alone of a given category in a given year.

All these text files were converted into the number of records of each county and transferred into tables. The definition of the number of records in this study can be understood as the counts of each county in a given year. For example in the final text file, the content of a line is "1911 Xiaoshan Yuyao Xiaoshan" for flood, then in the table the number of Xiaoshan in 1911 would add 2 and that of Yuyao in 1911 would add 1. The final data set contains fields of "NAME_CH" which are the Chinese characters of county name, "NAME_PY" which is Pinyin of county name, "LEV2_CH" which are the Chinese characters of prefecture name, "LEV2_PY" which is Pinyin of prefecture name, "LEV1_CH" which are the Chinese characters of province name, "LEV1_PY" which is Pinyin of province name, "Year" which is the time of record, "Category" which is the category of environmental events, and "Frequency" which is the number of records for a given county. Therefore, the example record for Xiaoshan and Yuyan would be enrolled in the data set as shown below in Table 2.4

Table 2.4: Record in the Dataset

NAME_CH	NAME_PY	LEV2_CH	LEV2_PY	LEV1_CH	LEV1_PY	Year	Category	Frequency
萧山县	Xiaoshan Xian	绍兴府	Shaoxing Fu	浙江	Zhejiang	1911	Flood	2
余姚县	Yuyao Xian	绍兴府	Shaoxing Fu	浙江	Zhejiang	1911	Flood	1

Notes

* This is just an example to show the structure of dataset.

** It is supposed that only one record "1911 Xiaoshan Yuyao Xiaoshan" for flood existed in this example.

According to the coordinates of locations in records, every location name can be pinned on the map as a point. If the data set is constructed at the county level of 1911, all points are expected to join the counts to county boundaries of 1911. The counts of every county can be considered as the number of environmental records, which indicates the count of environmental events of every county every year. The county names and boundaries of 1911 are taken from the source of "*China Historical Geographic Information System*" including 1986 counties, as shown in Figure 2.3. For each source, data is recorded along with the combination of the year and the county, and is classified into flood, drought, cold and wind files of the period from 1368 to 1911, which covers the whole of the Ming and Qing dynasties. In "*China Meteorological Disaster Dictionary (in Chinese)*", there are 91199 counts of flood counties, 38630 counts of drought counties, 52034 counts of cold counties, and 129070 counts of wind counties. However, in "Chinese Three Thousand years Meteorological Records Collection", there are only 48162 counts of flood counties, 22357 counts of drought counties, 8450 counts of cold counties, and 8187 counts of wind counties. The differences between the two sources are large, especially in the cold and wind categories.

Figure 2.3: Counties of China in 1911 (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Some boundaries contain multiple counties (Zhou, 2007).

2.3.3 Combination of Dataset

As mentioned before, records from "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*" are conservative, but the uniformity of this conservative source is of good quality. Therefore, the records from the conservative source are used as the baseline data set. Records from "*China Meteorological Disaster Dictionary (in Chinese)*" were taken as the supplementary sources to complement possibly missing records from documents with different stylistic rules and layouts. Since the quality of supplementary source varied across different volumes, and this source has been criticised for neglecting uniformity (Xia, 2015), records from supplementary source were modified before being merged into the baseline data set.

Naturally, records from conservative sources should fully overlap with records from supplementary sources in this case, since the compilation groups of supplementary sources claimed that they had gathered as many materials as they could to fill the volumes. However, there are many records missing in the additional sources compared with conservative sources. Besides, records of these two sources deviate with each other significantly across the time and counties.

One can interpret this massive difference in several ways. Firstly, there must be some missing information in the conservative source, since records outside official history are massive (Zhang, 1996). Secondly, the categorisation of different environmental events varies. Classification of supplementary sources is based on the existing categories of the original compilation, while the categorisation of the conservative source is based on the collection of relevant descriptions. There might be some mistakes in the categorisation criteria, which would entail mis-recording from both of these two sources. Thirdly, due to the diverse attitude and ability of compilers of supplementary sources, records may be missed, incorrectly recorded or repeatedly recorded for a given historical fact. Fourthly, due to the non-uniformity of supplementary sources, records may not represent the historical facts (extreme environmental events in this case) under the same standard. In this regard, a county can involve a record for a disaster if this county was affected but was not where it occurred.

In order to eliminate the issues mentioned above, this work estimated the quality of each volume of supplementary sources. It came up with a couple of weights to represent the credibility of interrelated records. The rules to evaluate the quality of the compilation work of supplementary sources are made based on following aspects:

- 1. Records that are not included in the supplementary sources but exist in the conservative sources are considered the result of inadequate research. Therefore, $\lambda = 1 - \frac{the\,number\,of\,missing\,records\,in\,supplementary\,source}{the\,total\,number\,of\,records\,in\,conservative\,source}$ is adopted as the essential reliability of each volume.
- 2. The number of references, the specific description of original sources ,and any evidence for precise verification represent the rigorousness of each work.
- 3. The number of pages can denote the richness of environmental materials to some extent. The larger number of years to complete a volume might indicate a greater carefulness of the work and might suggest the insufficient

ability of the compilation group. Therefore, this study only reduced the reliability if the value of $\rho = \frac{the\, number\, of\, pages}{the\, number\, of\, years}$ is extremely high or extremely low.

The basic credibility (λ) was calculated for each province based on the administrative boundaries in 1911. Considering the probable errors in the categorisation process, it is possible that the missing records were due to misclassification. However, the conservative source is regarded as the baseline data set, which means that the classification criteria of environmental category should follow the standard of the conservative source as well. Since the original sources of supplementary sources are not cited, it is unlikely that one can identify the details of previous classification criteria. It is also complicated to reclassify the records based on criteria of conservative sources. Consequently, the ratio of basic reliability is expected to adjust for the deviation from the mismatch of categorisation standard to some extent. As shown in Table 2.5, Qinghai province does not have records, and the reliability of records from the Xinjiang Province is zero. Records of Xizang province in supplementary sources were omitted, which means there is no necessity to calculate the ratio of reliability.

For the next step, overlapping records of the two sources were eliminated before employing the ratio of reliability. Consequently, we determined the number of the remaining records in additional sources by the interrelated reliability to generate a new annual data set at the county level (we call it data set A for convenience sake). Since there are obviously other factors affecting the quality of compilation according to the rules mentioned previously, data set A requires further modifica-

Table 2.5: Ratio of Basic Reliability of each Province (λ)

Province	Basic Reliability
Anhui	0.71
Fengtian	0.35
Fujian	0.13
Gansu	0.66
Guangdong	0.52
Guangxi	0.68
Guizhou	0.56
He'nan	0.19
Heilongjiang	0.82
Hubei	0.57
Hu'nan	0.55
Jilin	0.72
Jiangsu	0.29
Jiangxi	0.64
Neimenggu	0.46
Qinghai	No Data
Shandong	0.62
Shanxi	0.38
Shaanxi	0.19
Sichuan	0.65
Xinjiang	0.00
Yunnan	0.50
Zhejiang	0.57
Zhili	0.47

Data sources and notes

* Province boundaries are based on the administrative units in 1911 (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

** The ratio retains two decimal places.

tion. For rule 2, rigorousness was divided into two levels, denoted as an acceptable level and a doubtful level. If a volume is estimated to be doubtful, data from an associated province is omitted in data set A if below a given threshold since lower rigorousness represents a higher probability of incorrect records. Rule 3 was to determine the level of the threshold that is adopted to judge whether the data should be omitted for doubtful sources. Volumes in supplementary sources can be divided into two levels under rule 3. One is inadequate if the number of pages over the number of years is extremely high or extremely low, and adequate otherwise. Data at a doubtful level is omitted in data set A if the value is less than 1 when volume is considered being at an inadequate level, otherwise data is omitted if the value equals the value of correlated λ since in this case there is only 1 record at the corresponding county in a given year. The levels corresponding to each rule are provided in Table 2.6.

According to Table 2.6, the volumes of Anhui, Chongqing, Gansu, Guangxi, Heilongjiang, Hubei, Hu'nan, Jiangsu, Jiangxi, Ningxia, Shanxi and Sichuan are estimated to be at an acceptable level based on the conditions of rigorousness given above. The field "Reference" denotes the number of references for each volume, while the field "Evidence" represents whether there is any evidence or description of the proofreading work and verification. A large number of references for historical documents, notification for uncertain records, and detailed citations for each record are viewed as strong evidence. Volumes which contain at least 27 references or have strong evidence or have references and evidence at the same time were regarded to maintain acceptable rigorousness. As for the level of rule 3, if the value of ρ is within the range from 68 to 162, the volume which attains this value

Table 2.6: Levels of Rule 2 and Rule 3 for Each Volume

Province	Reference	Evidence	Level_1	ρ	Level_2	Omit Criteria
Anhui	10	Strong	Acceptable	76.40	Normal	None
Beijing	13		Doubtable	95.00	Normal	λ
Chongqing	27		Acceptable	63.67	Inadequate	None
Fujian	15		Doubtable	65.33	Inadequate	<1
Gansu	39		Acceptable	225.00	Inadequate	None
Guangdong	7		Doubtable	70.20	Normal	λ
Guangxi	27		Acceptable	68.83	Normal	None
Guizhou			Doubtable	99.00	Normal	λ
Hainan	13		Doubtable		Inadequate	<1
Hebei			Doubtable	93.00	Normal	λ
He'nan		Weak	Doubtable	199.00	Inadequate	<1
Heilongjiang	38	Weak	Acceptable		Inadequate	None
Hubei	26	Strong	Acceptable	71.00	Normal	None
Hu'nan	70	Weak	Acceptable	127.25	Normal	None
Jilin	11	Weak	Doubtable	56.29	Inadequate	<1
Jiangsu		Strong	Acceptable		Inadequate	None
Jiangxi	27	Strong	Acceptable	470.00	Inadequate	None
Liaoning			Doubtable	107.33	Normal	λ
Neimenggu			Doubtable		Inadequate	<1
Ningxia	16	Strong	Acceptable		Inadequate	None
Shandong	29	Strong	Acceptable	162.00	Normal	None
Shanxi		Strong	Acceptable	229.75	Inadequate	None
Shaanxi			Doubtable	42.20	Inadequate	<1
Sichuan	27		Acceptable	150.00	Normal	None
Tianjin	24		Doubtable	52.40	Inadequate	<1
Xinjiang	8		Doubtable	68.00	Normal	λ
Yunnan	22		Doubtable	180.00	Inadequate	<1
Zhejiang			Doubtable	143.50	Normal	λ

Data sources and notes

* Province boundaries are based on the modern administrative. (Resource and Environment Data Center of Chinese Academy of Sciences, 2015).

** λ denotes the associated ratio of basic reliability calculated in Table 2.5. $\rho = \frac{the\,number\,of\,pages}{the\,number\,of\,years}$. Blank represents no information.

is considered to be at a normal level. Criteria to classify the number of references and values of ρ were based on the distribution of these numbers. 27 is the upper quartile of the numbers of references, while 68 and 162 are the lower quartile and upper quartile of the values of ρ. For volumes without the number of references and the value of ρ, levels were used to downgrade the reliability.

When applying the omit criteria for dataset A, what should be noticed is that the boundaries of administrative units for the Qing dynasty and modern day are different, as shown in Figure 2.4. Therefore, locations inside current boundaries followed omit criteria, but locations inside boundaries in 1911 were assigned λ.

After the modifications based on rule 2 and rule 3, data set A becomes data set B where some data were omitted due to insufficient reliability. Therefore, data set B can be merged to the data set from conservative sources to obtain the final data set. In the new final data set, the value of the environmental parameter represents the frequency level of a given type of disaster but may not be an integer since λ is less than 1. In this regard, parameters in the new data set could be interpreted as the expected true frequency.

2.3.4 Data Management in GIS Database

The application of GIS can benefit the visualisation of historical environmental disasters (Berman, 2009). To link in the environmental attributes of the spatial characteristics, the data sets were formatted to fit the requirements of ArcGIS. There are 6 feature classes in the Geodatabase, which is a point feature class

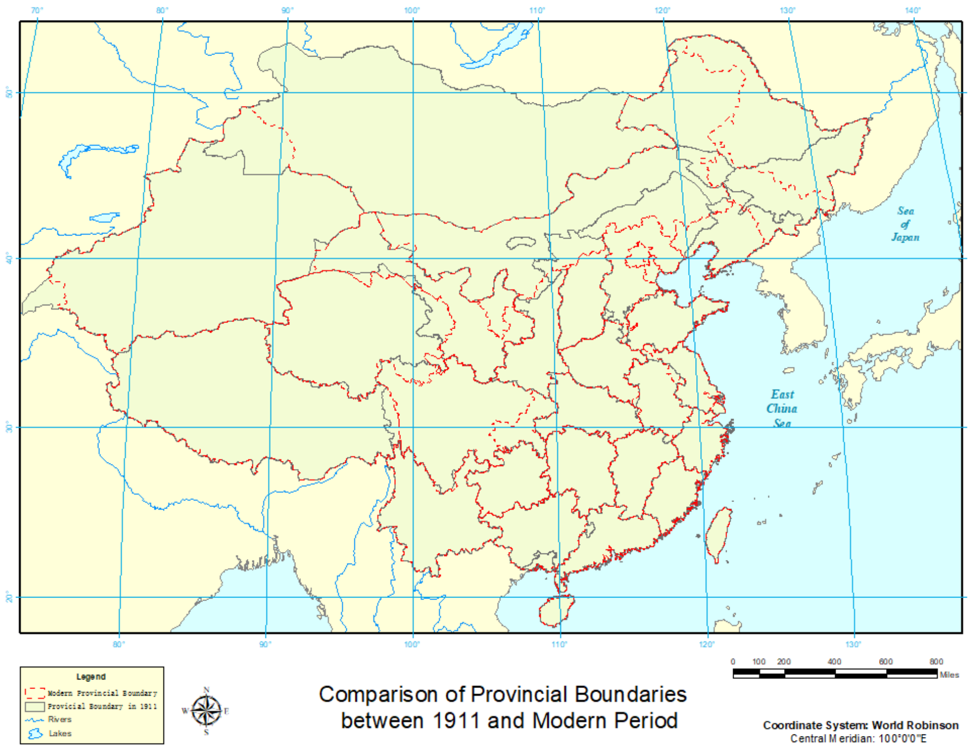


Figure 2.4: Comparison of Provincial Boundaries between 1911 (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016) and Modern Period (Resource and Environment Data Center of Chinese Academy of Sciences, 2015)
. Red dash denotes the boundary for modern period.

which contains all the locations of counties in 1911. These 5 polygon classes include the boundaries of modern provinces, provinces in 1911, prefectures in 1911, counties in 1911, and physiographic macro-regions of China (Cartier, 2002). Data of modern provincial boundaries comes from the Resource and Environment Data

Cloud Platform (Resource and Environment Data Center of Chinese Academy of Sciences, 2015). Data of other boundaries comes from CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). The main structures of attribute tables for layers are shown in Table 2.7, Table 2.8, Table 2.9, Table 2.10, Table 2.11, and Table 2.12.

Table 2.7: Structure of Attribute Table for County Point

Name of Feild	Meaning of Feild
NAME_CH	Name of county
X_COORD	Longitude of county
Y_COORD	Latitude of county
LEV1_CH	Name of province
LEV2_CH	Name of Prefecture

Data sources and notes

* Coordinates of counties are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

** Some fields such as the Pinyin of county name are omitted since they are not necessary information.

Attributes in the point class can be joined with the polygon classes by connecting the attribute table according to the primary keys and spatial locations. Specifically, fields "LEV1_CH", "LEV2_CH" of county points can be related to the "NAME_CH" of province polygon and prefecture polygons, respectively. Polygons of modern provinces, county boundaries in 1911, and physiographic macro-

Table 2.8: Structure of Attribute Table for Modern Province Polygon

Name of Feild	Meaning of Feild
NAME_CH	Name of modern province

Data sources and notes

* Polygons of provinces are obtained from Resource and Environment Data Cloud Platform (Resource and Environment Data Center of Chinese Academy of Sciences, 2015).

** Some fields such as the Pinyin of province name are omitted since they are not necessary information.

Table 2.9: Structure of Attribute Table for 1911 Province Polygon

Name of Feild	Meaning of Feild
NAME_CH	Name of province in 1911

Data sources and notes

* Polygons of provinces are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

** Some fields such as the Pinyin of province name are omitted since they are not necessary information.

Table 2.10: Structure of Attribute Table for 1911 Prefecture Polygon

Name of Feild	Meaning of Feild
NAME_CH	Name of prefecture in 1911

Data sources and notes

* Polygons of prefectures are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

** Some fields such as the Pinyin of prefecture name are omitted since they are not necessary information.

Table 2.11: Structure of Attribute Table for 1911 County Polygon

Name of Feild	Meaning of Feild
NAME_CH	Name of county in 1911

Data sources and notes

* Polygons of counties are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

** Some fields such as the Pinyin of county name are omitted since they are not necessary information.

Table 2.12: Structure of Attribute Table for Physiographic Macro-region Polygon

Name of Feild	Meaning of Feild
SYSTEM	Code of subsystem
MR	Code of physiographic macro-region

Data sources and notes

* Polygons of physiographic macro-regions are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

** Some fields such as the features class of physiographic macro-region name are omitted since they are not necessary information.

regions can be joined to the county points if the point falls inside the polygon. Based on these classes, relevant attributes can be presented in maps at the county level, prefecture level, province level, and physiographic macro-regional level.

According to the process of combining data in this study, the sorted environmental attributes were stored in the new data set, which is the main environmental data set of this research and follows the structure as shown in Table 2.13. By joining the attributes in the main environmental data set with county points according to the primary key "NAME_CH", environmental attributes can be used for geographic analysis and visualisation. Data in "Parameter" represent the parameterisation results processed from approaches in previous sections of this chapter.

Table 2.13: Structure of Main Environmental Dataset

Name of Feild	Meaning of Feild
NAME_CH	Name of county
LEV1_CH	Name of province in 1911
LEV2_CH	Name of Prefecture in 1911
Year	Time of the occurrence of extreme environmental events
Category	Types of extreme environmental events
Parameter	Parametrisation result of environmental record

Data sources and notes

* List of counties are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

** Some fields such as the Pinyin of county name are omitted since they are not necessary information.

*** Fields “NAME_CH”, “LEV1_CH” and “LEV2_CH” coincide with fields in attribute table of county points.

2.4 Result

2.4.1 Datasets

From the methodology in this work, we generated two data sets for future application. The first one is the baseline data set. This data set contains the frequency of four types of environmental events for each county annually from conservative sources "*Chinese Three Thousand years Meteorological Records Collection (in Chinese)*". As mentioned before, the four environmental categories are flood, drought, cold, and wind. All records can be joined to county administrative boundaries of 1911. There are 1986 counties in total from 1368 to 1911 across the Ming and Qing dynasties. The frequency of each province is shown in Table 2.14.

The second one is the main environmental data set. The main environmental data set consists of county-level weighted frequency of each disaster category from two different sources each year. The frequency is the number of disaster records but was revised according to the combination process demonstrated in the previous section. This weighted frequency approach is expected to map the environmental variation of the real world at the county level more accurately compared with the baseline data set. Data were joined to county administrative boundaries of 1911 as well. The weighted frequency of each province is shown in Table 2.15. The modified frequencies in the data set are not necessarily integers. Apparently, the main data set contains more frequencies than the baseline data set. To be specific, the frequencies in Xinjiang, Xizang and Neimenggu are low partially due to the

Table 2.14: The Total Frequency for each Province by Disaster Category From 1368 to 1911 in the baseline dataset

Province	Cold	Drought	Flood	Wind
Anhui	500	1491	2886	257
Fengtian	38	59	277	33
Fujian	278	691	1793	416
Gansu	192	450	587	76
Guangdong	396	1133	3362	1320
Guangxi	195	508	1008	150
Guizhou	92	195	439	49
He'nan	555	1588	3310	580
Heilongjiang	9	2	14	2
Hubei	361	1207	2584	179
Hu'nan	319	1227	2263	211
Jilin	2	0	9	0
Jiangsu	959	2249	5763	1174
Jiangxi	472	1384	2890	207
Neimenggu	3	3	5	0
Shandong	869	2077	4122	863
Shanxi	632	1252	1768	297
Shaanxi	345	923	1360	217
Sichuan	197	754	1383	150
Xizang	54	4	55	1
Xinjiang	2	0	2	0
Yunnan	171	273	957	82
Zhejiang	842	2032	3892	712
Zhili	828	2424	5685	983

Data sources and notes

* List of Province are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Qinghai and Wuliyasutai are omitted since there is no data in the baseline dataset.

small number of credible records.

Compared with conservative sources, the main data set is expected to contain more information from supplementary sources. After modification, there are still many excess records supplemented into baseline data set (data set from conservative sources) as described in Table 2.16. The comparison implies that more counties are recorded to have disasters or more periods are identified to have disasters for a given county. In Fujian, Xizang and Xinjiang, there is no county supplemented since records of Xizang and Xinjiang in supplementary sources were not collected due to none-credibility of records, and most of the records in Fujian in supplementary sources were omitted due to low reliability of records. Comparatively, the reliability of He'nan province is also relatively low, but several records remained in the supplementary sources after modification. The amount of excess records in He'nan province is also noticeable. The majority of provinces have obtained a large number of excess records, which indicates a considerable improvement over the conservative sources. Additionally, Table 2.17 illustrates more significant extensions in the cold and wind categories. It is rational to suppose a better performance for the main data set at a higher resolution, such as the county level. More information contained in the main data set suggests higher accuracy. Therefore, the main data set can be illustrated at multiple levels, including the county level, to generate relevant maps for different themes.

Table 2.15: The Total Modified Frequency for each Province by Disaster Category From 1368 to 1911 for the main dataset

Province	Cold	Drought	Flood	Wind
Anhui	10481.18	2521.03	6098.65	5621.68
Fengtian	124.92	247.55	793.79	589.30
Fujian	283.00	691.00	1793.00	423.00
Gansu	1601.41	4081.80	9833.93	9818.62
Guangdong	1111.37	1256.83	3966.59	2267.57
Guangxi	942.32	961.17	2272.26	2855.75
Guizhou	126.91	218.09	572.46	129.53
He'nan	565.76	1597.45	3313.79	597.26
Heilongjiang	589.75	215.57	559.43	569.61
Hubei	418.40	1828.99	4115.62	199.23
Hu'nan	408.91	1793.87	3975.80	2838.89
Jilin	303.09	115.63	445.36	464.00
Jiangsu	1240.36	2454.62	6422.49	3216.91
Jiangxi	1521.08	1730.95	4558.98	5995.90
Neimenggu	16.46	12.69	45.15	42.92
Shandong	7701.07	4296.24	9336.77	15071.83
Shanxi	1757.56	1709.29	2437.72	2739.82
Shaanxi	360.76	925.50	1363.07	229.52
Sichuan	990.04	1508.05	3280.86	2458.20
Xizang	72.00	4.00	55.00	1.00
Xinjiang	2.00	0.00	2.00	0.00
Yunnan	930.97	379.05	1485.26	959.59
Zhejiang	912.98	2126.58	4228.16	6621.93
Zhili	1041.03	3837.06	6714.16	2635.22

Data sources and notes

* List of Province are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Qinghai and Wuliyasutai are omitted since there is no data in the main dataset, since relavent records were deleted due to low credibility.

** The decimal places for the data in the assignment are two.

Table 2.16: The Total Number of Excess Yearly Disaster Counties Where Have Disaster Records for each Province by Disaster Category from 1368 to 1911

Province	Cold	Drought	Flood	Wind
Anhui	5116	1228	3251	5479
Fengtian	82	218	500	553
Fujian	0	0	0	0
Gansu	897	3085	4517	4988
Guangdong	584	107	462	505
Guangxi	931	537	1459	2036
Guizhou	28	19	101	64
He'nan	2	23	4	17
Heilongjiang	368	170	369	382
Hubei	78	743	1870	29
Hu'nan	141	813	2152	3591
Jilin	211	104	270	283
Jiangsu	350	517	1621	3823
Jiangxi	1480	462	2023	3950
Neimenggu	6	5	16	18
Shangdong	7690	2613	5527	8152
Shanxi	2266	1049	1381	2688
Shaanxi	17	13	15	42
Sichuan	904	895	1998	2698
Xizang	0	0	0	0
Xinjiang	0	0	0	0
Yunnan	646	97	426	746
Zhejiang	7	67	235	2873
Zhili	170	1213	760	1341

Data sources and notes

* List of Province are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Qinghai and Wuliyasutai are omitted since there is no data in conservative sources.

** The number of excess counties with disaster records is defined as the number of disaster counties in the main dataset minus the number of disaster counties in the baseline dataset (Wen, 2006; Zhang, 2004b).

Table 2.17: The Percentage of Excess Disaster Records for each Province to the Number of Disaster Counties in the baseline dataset by Disaster Category from 1368 to 1911

Province	Cold	Drought	Flood	Wind
Anhui	1023.20%	82.36%	112.65%	2131.91%
Fengtian	215.79%	369.49%	180.51%	1675.76%
Fujian	0.00%	0.00%	0.00%	0.00%
Gansu	467.19%	685.56%	769.51%	6563.16%
Guangdong	147.47%	9.44%	13.74%	38.26%
Guangxi	477.44%	105.71%	144.74%	1357.33%
Guizhou	30.43%	9.74%	23.01%	130.61%
He'nan	0.36%	1.45%	0.12%	2.93%
Heilongjiang	4088.89%	8500.00%	2635.71%	19100.00%
Hubei	21.61%	61.56%	72.37%	16.20%
Hu'nan	44.20%	66.26%	95.10%	1701.90%
Jilin	10550.00%		3000.00%	
Jiangsu	36.50%	22.99%	28.13%	325.64%
Jiangxi	313.56%	33.38%	70.00%	1908.21%
Neimenggu	200.00%	166.67%	320.00%	
Shandong	884.93%	125.81%	134.09%	944.61%
Shanxi	358.54%	83.79%	78.11%	905.05%
Shaanxi	4.93%	1.41%	1.10%	19.35%
Sichuan	458.88%	118.70%	144.47%	1798.67%
Xizang	0.00%	0.00%	0.00%	0.00%
Xinjiang	0.00%		0.00%	
Yunnan	377.78%	35.53%	44.51%	909.76%
Zhejiang	0.83%	3.30%	6.04%	403.51%
Zhili	20.53%	50.04%	13.37%	136.42%

Data sources and notes

* List of Province are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Qinghai and Wuliyasutai are omitted since there is no data in conservative sources.

** The number of excess counties with disaster records is defined as the number of disaster counties in the main dataset minus the number of disaster counties in the baseline dataset (Wen, 2006; Zhang, 2004b). Blank represents that there is no record in the baseline dataset, so the persentage cannot be calculated.

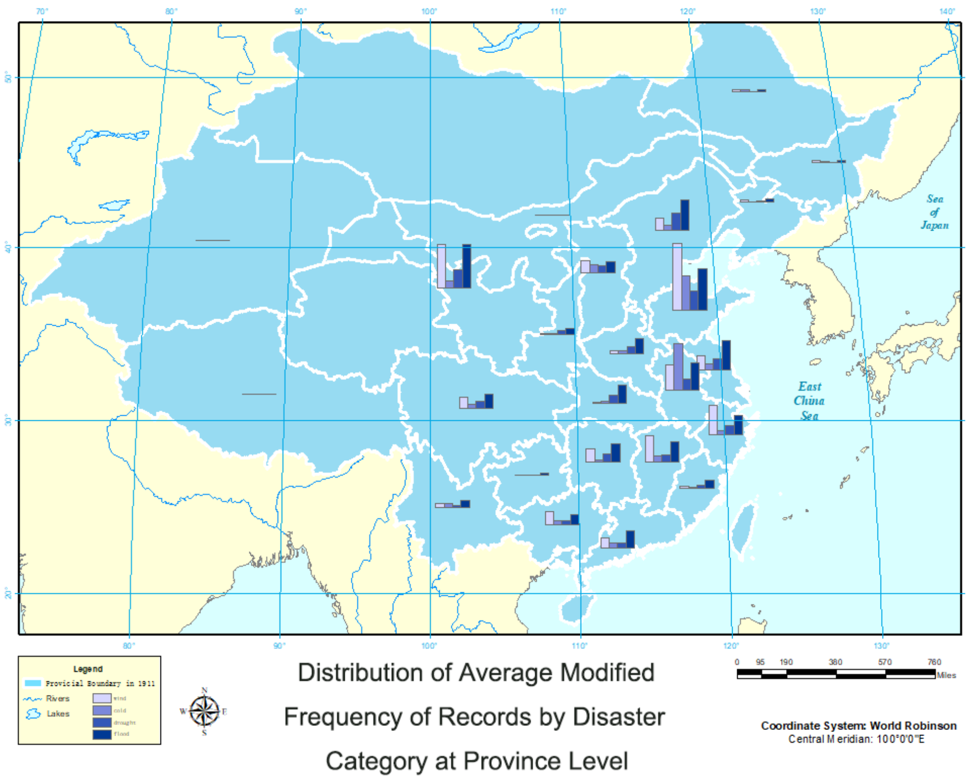


Figure 2.5: Distribution of Average Modified Frequency of Records by Disaster Category at the province level (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

2.4.2 Theme Maps of the Main Dataset

The distribution of the average frequency of records by disaster category at the province level is shown in Figure 2.5. Except for Gansu province, the wind records and cold records are concentrated on the east coast of China. The number of wind records in Shandong province is significantly high. Areas around the east coast of China may be more sensitive to wind and cold. However, this requires further validation in the field of meteorology. For the validated categories (drought and flood in this case), county-level frequency for any given period could be presented.

According to Figure 2.6 and Figure 2.7, the frequency of flood records was higher in the Qing dynasty, and more counties were recorded in the Ming dynasty. The highest average flood records in the Ming dynasty is around 1.04 per year, but in the Qing dynasty is 2.41 per year. Except for some extra counties in the Qing dynasty, the spatial concentration of flood records does not differ significantly from the Ming dynasty to the Qing dynasty. Most records are for East China and Southeast China. The distribution matches the map in Figure 2.5. Consequently, the distribution of the records of drought performs in the same way with floods. In the Qing dynasty more counties are discovered to maintain records. However, the highest average frequency in the Ming dynasty is 1.14 per year, but in the Qing dynasty is 0.81. The spatial concentration of drought in the Ming and Qing dynasty did not vary massively. It is rational to consider that from the Ming dynasty to the Qing dynasty, floods occurred more frequently than droughts.

It seems that records are particularly concentrated in some regions, such as Shandong province. No matter which data set, Shandong province is characterized by a relatively high number of records in all four categories. This concentration can be observed in the distributions of cold and wind more easily, as shown in Figure 2.8 and Figure 2.9. As the origin region of Chinese education, Shandong province may have had a better tradition to keep everything recorded. However, previous validation work on official histories did not specify a significant deviation of the recording quality between Shandong province and other regions. Therefore, it is reasonable to believe that the uniformity is not violated before any new evidence is discovered. Under this premise, we can infer that environmental features are spatially stable in the long run.

2.4.3 Other Issues

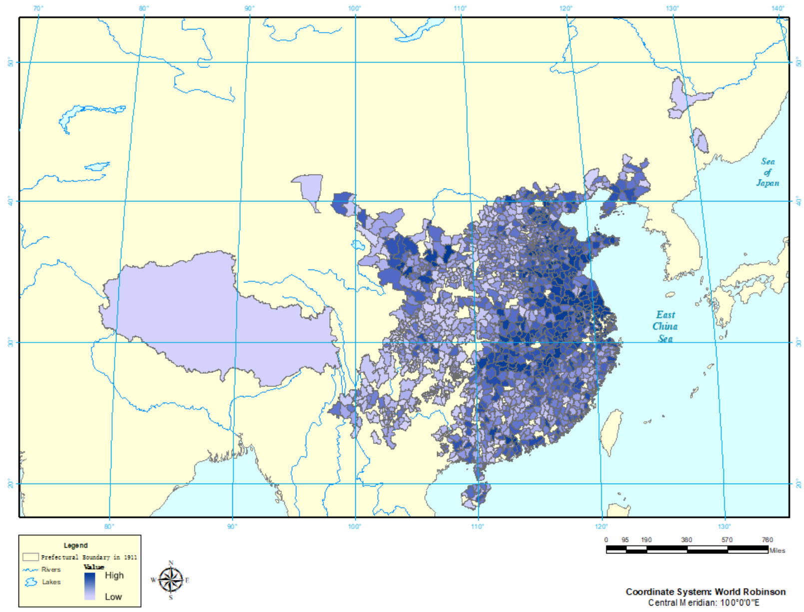
A concern is that the increase of flood or drought records was because of more complete historical documents. However, according to the validation for records in historical documents, the Ming and Qing dynasty are likely to have maintained relatively complete environmental records since continuous compilation for local history existed for both periods (Zhang, 1996). What is more, if the increase of records was due to an increase of complete historical documents, the number of drought records for a given region should rise simultaneously (the actual situation was that drought records for the individual counties did not increase significantly). In this regard, it might be possible for prefecture-level or province-level records to be aggregated due to some expansion of historical documents (for example, more counties acquired the ability to compile more detailed local history) especially if there was no record before the Qing dynasty, but the sufficiency of environmental records for a given county should be considered stationary if the records were continuous. Therefore, refining data to the county-level is necessary to some extent to further eliminate the bias from the insufficiency of historical records. This research intends to inspire considering the possibility of combining data from various sources and find a capable approach to parameterise historical environmental information at the county level. The final data set shows adequate consistency with the data set from conservative sources. Additionally, more records were supplemented to make the new data set more complete. At the same time, extreme environmental events of wind and cold were also involved in this data set, which would help to reconstruct the weather condition for wind and cold in the Ming and Qing dynasty.

Overall, the digitalisation process from the paper source in this study was considered appropriate to convert previous collections into digital information, especially when the paper source was not shared digitally. Newly developed techniques based on Artificial Intelligence (AI) should thereby improve the efficiency in the collection work of historical documents. Although there will be some errors when employing such techniques, the number of these errors are assumed to be acceptable enough to satisfy most of the relevant research using the data. What is more, the proportion of errors from AI techniques has already been likely much smaller than man-made mistakes, particularly for some specific fields (character recognition and lexical analysis in this study), which makes it easier for further proofreading. It is necessary to engage in an effort to develop specialised techniques to facilitate data collection in historical fields. What is more, the Geographic Information System is appropriate to manage historical data systematically and can benefit the presentation for relevant data sets. Accordingly, the spatial linkage of historical information can be illustrated in maps for different themes (Xu, Xu, and Dong, 2012), as well as across different time periods.

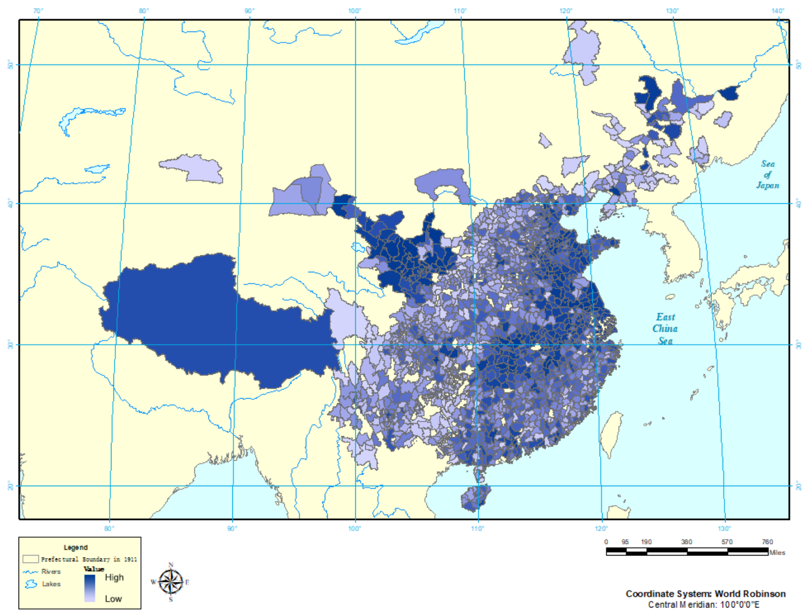
2.5 Conclusion and Limitation of this Research

In conclusion, historical documents in China are deemed to be sufficient to reconstruct historical environmental variation, especially for the Ming and Qing dynasties. However, the current collection of historical documents were not digitalized, which hindered researchers from other fields to employ these documents on a large scale. There were some unshared databases for the historical environmental information, but only some outputs based on such databases are publicly available (Xia, 2015; Zhang, 2005). What is more, previous environmental data sets were usually at the prefecture level or higher, which was not enough for more precise spatial analysis. Therefore, this study verified the application of newly developed AI techniques, including character recognition and lexical analysis (from Baidu cloud), to digitalise historical information from paper sources. Environmental information was extracted from different sources. It then was combined and parameterised to construct a new county-level environmental data set, which covers the environmental event categories of cold, drought, flood, and wind. AI techniques can facilitate the data collection from paper sources, especially for historical documents. Data combination based on the quality evaluation of the separate sources was proposed to supplement records into official history, and such an approach did not exaggerate the historical environmental records to a significant scale. Thus, the frequency of disaster records was modified to correct for possible non-uniformity and miss-recordings and to generate county-level parameters. The final data set matched the conclusions from previous work but with large numbers of records supplemented, and was validated to have an acceptable effect on mapping the real world historically (Zheng, Zhang, and Zhou, 1993; Wei, 2007; Zhang and Liu, 2002). Moreover, the use of Geographic Information System was shown to benefit the management and presentation of historical spatial data. According to relevant maps of different themes, the spatial distribution of environmental records did not vary significantly across the Ming and Qing dynasties. The number of flood records increased significantly from the Ming dynasty to the Qing dynasty in that the highest average frequency of flood records in a county increased from 1.04 to 2.41. This new data set was considered to be sufficient to supplement data set from official history and is expected to benefit further research at the county level.

There are some limitations in this study that are anticipated to be addressed in the future. Firstly, supplementary sources were only from "*China Meteorological Disaster Dictionary (in Chinese)*" which might not be complete enough to fulfil county-level information for every county. Additionally the criteria employed for quality evaluation were relatively simple, which means bias from the modification could be considerable. The traditional way to correct historical records is to compare the record from different sources to identify whether these records are referring to the same historical facts according to historical descriptions, and then to select the record from the most reliable source. However, this would violate the application for environmental records at the county level. To explain further, the severity information can only be captured by the description of the content in context or by the number of disaster counties introduced in previous works (Zheng, Zhang, and Zhou, 1993; Hao, Ge, and Zheng, 2010; China Meteorological Administration Institute of Meteorology, 1981). While this study tried to incorporate different sources available while assigning separate weights based on the reliability from the quality evaluation of sources, the approach to establish the reliability should be considered experimental, requiring further validation. The potential improvement of this approach is to rank all the possible available sources and to find the missing rate for each class in the ranking as the criteria to determine the weights. Unfortunately, due to time limitations, the collection of complete sources and further improvement in quality evaluation was not possible. Moreover, the classification of different environmental categories was fairly broad in that it was not strictly determined by the meteorology definition but by the description and personal awareness of different disasters, which could bias the records for each category. Preferable would have been to categorise each record based on more meteorological based knowledge and this might be a further venue for future research . Moreover, the linkage of different records required a sufficient level of comparison across records from different sources. The validation of this data set could also be expanded to a larger scale, although unfortunately, data availability for records from equipment or other reliable proxies is not always plentiful, which means historical work is constrained by the progress from the meteorology field or the palaeoclimatology field. In summary, data reconstruction of the historical environment necessitates larger coordination across different fields, including history and meteorology.

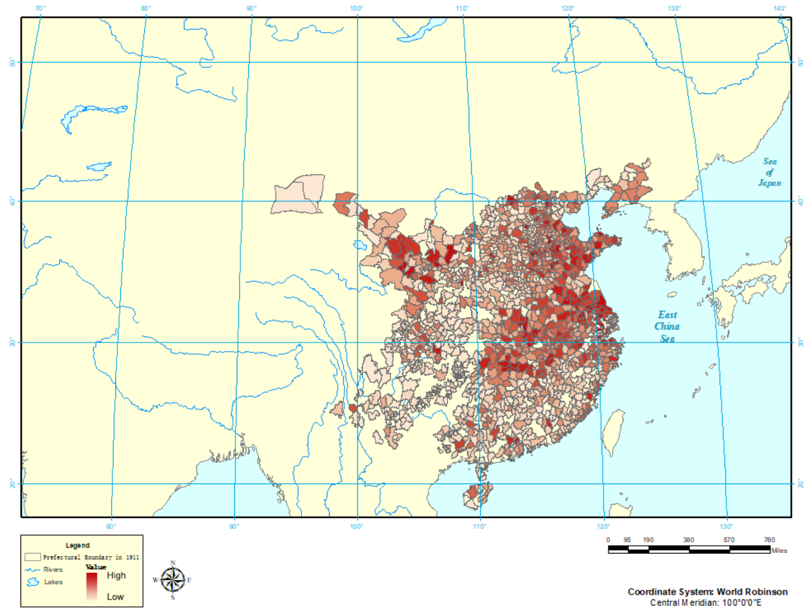


(a) Ming

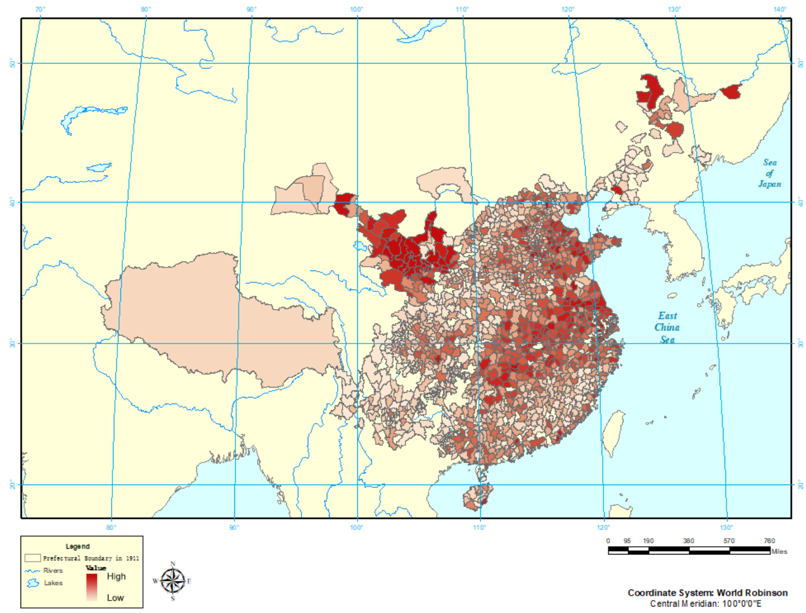


(b) Qing

Figure 2.6: Distribution of Average Modified Frequency of Flood Records in the Ming and Qing Dynasty (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). County boundaries are based on the administrative boundaries in 1911.

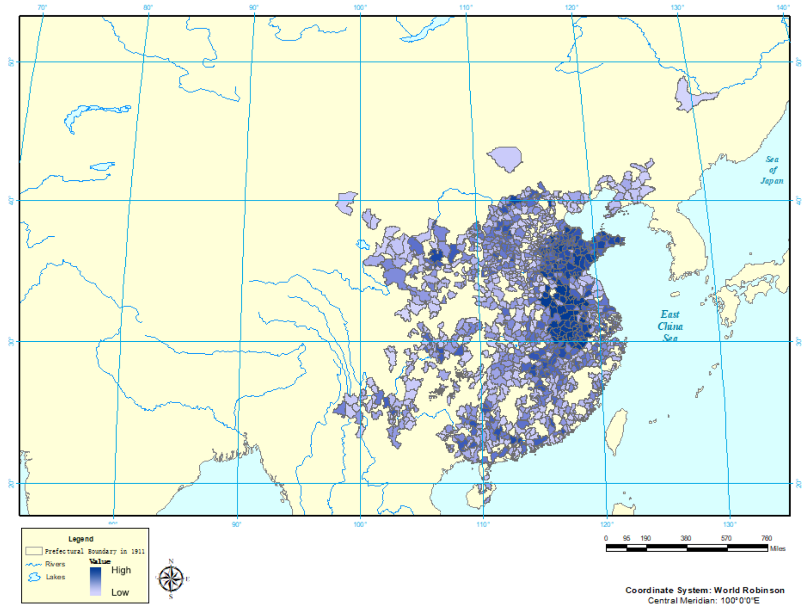


(a) Ming

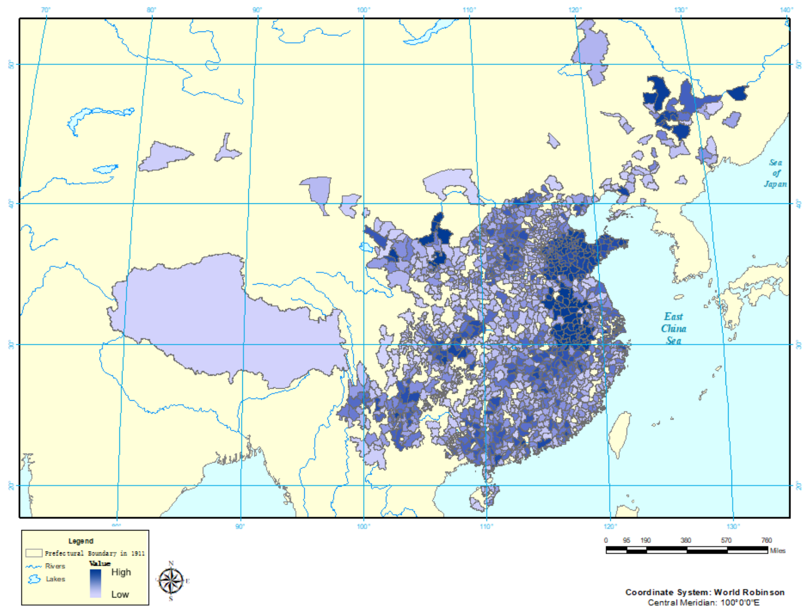


(b) Qing

Figure 2.7: Distribution of Average Modified Frequency of Drought Records in the Ming and Qing Dynasty (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). County boundaries are based on the administrative boundaries in 1911.

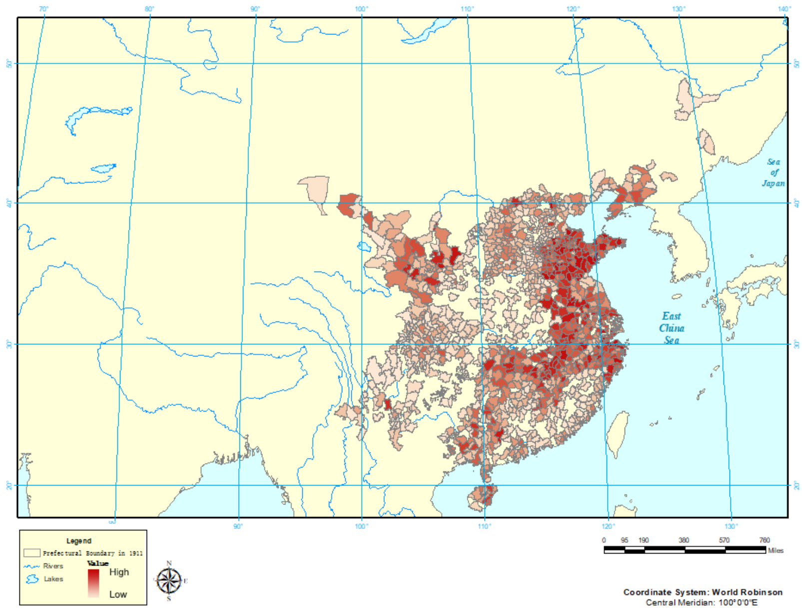


(a) Ming

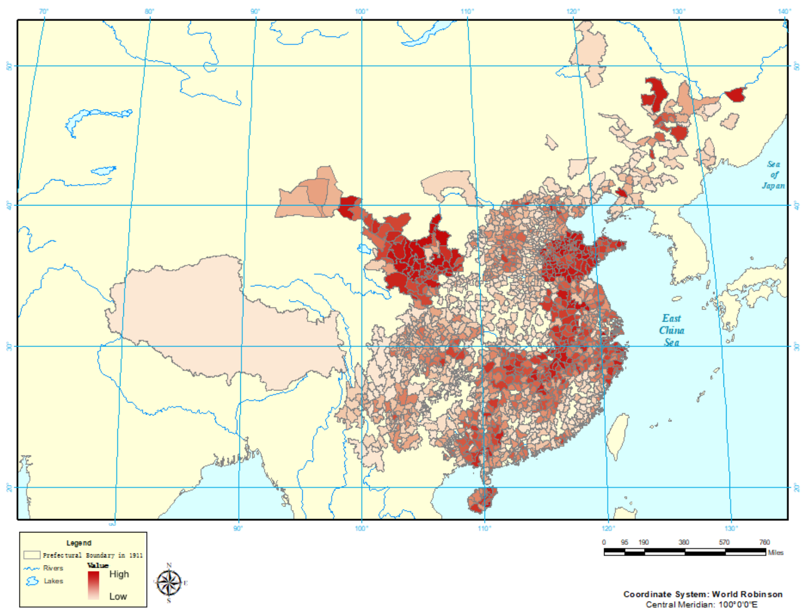


(b) Qing

Figure 2.8: Distribution of Average Modified Frequency of Cold Records in the Ming and Qing Dynasty (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). County boundaries are based on the administrative boundaries in 1911.



(a) Ming



(b) Qing

Figure 2.9: Distribution of Average Modified Frequency of Wind Records in the Ming and Qing Dynasty (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). County boundaries are based on the administrative boundaries in 1911.

Chapter Three

Comparisons among Countrywide Environmental Datasets From 1368 to 1911

3.1 Introduction

Research in economic history has long recognised the potential importance of environmental factors (Roberts, 2004; Webb, 1931; Turner, 1921), where several studies have indeed found that human systems have strong interactions with natural systems (Nash, 1972). As a matter of fact, environmental variation, such as natural disasters and climate change, has been considered as key variable in shedding further light on old topics (Fenske, 2013), as well as inspiring some research venues (Alesina, Giuliano, and Nunn, 2013). These trends have been coupled with the extensive attention from the international academic community towards data construction of environmental factors.

Chinese academics in particular have conducted a lot of research work on historical environmental data. Based on historical documents and current data, paleoclimatologists and researchers of other fields in China have developed multiple approaches to parameterise historical environmental information, and have collated several prominent data sets which have been employed in different economic studies (Bai and Kung, 2011; Lee, Zhang, et al., 2017). Among these data, an early one is "*The Atlas of Wetness and Dryness Distribution in China in the Past Five Hundred Years (in Chinese)*"(China Meteorological Administration Institute of Meteorology, 1981) published in the 1970s. This atlas contains the grade of wetness and dryness parameterised from around 2100 local gazetteers across 120 sites. However, the total number of existing local gazetteers and official histories is over 10 thousand (Xia, 2015). Moreover, the grade is constructed from description ranking and thus likely to be at least partially subjective, as well as potentially lose quantitative information of the severity of events. As a result, Zheng, Zhang, and Zhou (1993) constructed annual flood/drought grades with regard to the number of flood/drought affected counties for each year for each province. The approach has been validated and employed in other regional studies (Zheng, Hao, and Ge, 2005). In this chapter we instead construct a county level environmental data set for the entire China, primarily based on "*Chinese Three Thousand years Meteorological Record Collection*" (Wen, 2006)- the most sufficient compilation of official histories. Importantly however, while this data set is a rich information source for the broader regional level, such as prefectures, the number of disaster counties is insufficient on its own to identify county level environmental disaster events. We thus supplemented it with records in "*China Meteorological Disaster Dictionary (in Chinese)*" (Wen, 2006) and "*Chinese Three Thousand years Meteorological Record Collection*", allowing us to annual construct county level environmental disaster series for the period 1368-1911.

Data sets constructed from different sources and approaches can be characterised by substantial differences, possibly due to critical errors. It is thus crucial to assess the quality of these data sets and verify their potential in capturing environmental events at least as good as the sources do separately. There are several issues that require further validation. The first one is whether the new data set could reveal similar trends and characteristics with data sets constructed from a single source by other approaches. The second one is whether the new approach to supplement the source would fill the missing information appropriately. The last one is that whether the new parametrisation method would alter the quantitative features of historical records. This study thus compares the newly constructed data set to a data set derived from official histories alone, in which the latter one represents the validation data set. We show that the new data set demonstrates acceptable performance at broader regional levels compared with previous data sets. Apart from temperature, previous data sets have usually only quantified flood and drought events. Since the new data set also captures other types of natural disasters, which have been largely ignored by other data, we also validated its ability to map actual environmental variations by comparing it with actual precipitations of the Beijing after 1764. Overall, the new data set appears to be more accurate and provide more information at a more spatially dis-aggregated level (i.e., county) than previously constructed data, and thus can provide a rich source for future quantitative historical research of China related to environmental disaster events.

The remainder of the chapter is organised as follows. Section I introduces the relevant background and the features of three different data sets used in the analysis. Section II introduces the data sources of relevant data and the methods used to assess the new data set. Section III presents the results of comparisons and validations of the new data set. The last section contains the conclusion and a discussion of the limitations of this research.

3.2 Data Sources and Methodology

3.2.1 Data Sources

The data set derived from "*The Atlas of Wetness and Dryness Distribution in China in the Past Five Hundred Years (in Chinese)*"(China Meteorological Administration Institute of Meteorology, 1981) consists of 120 climate series of the grade of wetness and dryness from 1470 to 1979 and is constructed from some 2100 local gazetteers and actual observed precipitation. Each series represents the wetness and dryness situation of around 2 to 3 prefectures. The index of wetness and dryness grade is parameterised by the description of wetness and dryness in historical records when there is no record for accurate precipitation. For the current chapter data before 1911 were extracted. Records are classified into five grades, where grade 1 represents wet, grade 2 subwet, grade 3 normal, grade 4 subdry, and grade 5 dry conditions. Considering the time, range, severity and frequency of rainfall across seasons in different regions, the distribution of each grade is set to 10% for grade 1 and 5, 20% - 30% for grade 2 and 4, and 30% - 40% for grade 3. The criteria to identify the grade from historical documents is as follows:

- Grade 1: continuous and heavy rainfall across seasons with descriptions such as "heavy rain across spring and summer".
- Grade 2: continuous and heavy rainfall during one season and in conservative areas with descriptions such as "flood in April".
- Grade 3: no relevant record.
- Grade 4: drought in one season at lower severity with descriptions such as "drought in summer".
- Grade 5: continuous and severe drought across time in large areas with descriptions such as "drought in summer and autumn".

If precipitation records exist, the grade is defined as below to coincide with the grade from historical documents:

- Grade 1: $R_i > (\bar{R} + 1.17\sigma)$
- Grade 2: $(\bar{R} + 0.33\sigma) < R_i \leq (\bar{R} + 1.17\sigma)$
- Grade 3: $(\bar{R} - 0.33\sigma) < R_i \leq (\bar{R} + 0.33\sigma)$
- Grade 4: $(\bar{R} - 1.17\sigma) < R_i \leq (\bar{R} - 0.33\sigma)$

- Grade 5: $R_i < (\bar{R} - 1.17\sigma)$

Generally, the precipitation data from May to September is employed to reclassify a given station, so that $\bar{R}$ is the precipitation mean of May to September across several years. R_i is the precipitation of May to September for each year, and σ is the standard error. One should note that employing the sources of historical documents to undertake this classification is not viable since the total number of existing local gazetteers is over 10 thousand. Additionally, most of the actual records of precipitation only existed after 1911, which means only the grade from historical documents can provide a reference value to cover the majority of areas in the Ming and Qing dynasty. Moreover, since the historical grade of precipitation is based on an analysis of the description for environmental events in records, it is possible to contain perspective error for the severity of environmental events.

The second data set is used is "*Chinese Three Thousand years Meteorological Record Collection (in Chinese)*". This source is accepted as the most credible official historical collection of meteorogical information. The derived data covers the periods from 1368 to 1911 and is reconstructed at the annual prefecture level using the number of disaster struck counties each year. Disasters are classified as flood, drought, cold, and wind. Since prefecture boundaries changed over time, records were joined to coincide with the 381 prefectures classification of 1911. Additionally, data can be converted to the physiographic macro-regional. Since this data set was constructed from official histories alone, it is a data set which is more conservative.

The third data set used in the analysis consists of using the conservative data set but supplementing it with records from "*China Meteorological Disaster Dictionary (in Chinese)*". In similar manner to the conservative data set, this new data set consists of annual data for 381 prefectures according 1911 administrative boundaries for four disaster event categories. Importantly we also assign probability weights on the supplementary records, resulting in a weighted frequency for each county. Even though the new data set itself is at the county level, to facilitate comparisons the data is aggregated to the prefecture level using these weighted frequency weights and the number of disaster counties for each year. Since the new data set combines data from two sources, it would be regarded as the combined data set.

Finally, annual actual observed precipitations for Beijing from 1764 to 1911 (Zheng, Zhang, and Zhou, 1993) are used for further validation.

3.2.2 Approach for Validation and Comparison

To assess the new data set, a comparison will be made between it and the second data set to validate the sufficiency. Additionally, the new data set will be compared with actual precipitation to validate the ability to map the environmental features. It is crucial to note that the approach of the percentage change to the mean of the number of disaster counties has been accepted in several studies (Zheng, Zhang, and Zhou, 1993; Hao, Ge, and Zheng, 2010; Xia, 2015). Therefore, this research regards the approach of the percentage change to the mean of the number of disaster counties as the validated criteria. Consequently, the data set, which is based on the number of disaster counties from the conservative source (official histories alone), can be considered the validation data set.

Firstly, the combined data set and the validation data set are compared to identify the spatial deviation of the distribution of credible records. To this end the two data sets were converted to the number of disaster counties at the prefecture and province level in order to have these at the same spatial level. If the record in a county is greater than 0 for a given year in the new data set, this county would be counted as a disaster county for that year. Therefore, if there are flood records in 10 counties for a prefecture in a given year, the number of flood counties for this prefecture in this given year is 10. In this regard this research expanded the application of the number of disaster counties to all regions where records exist. The number of counties from the new data set and the conservative source were compared at the prefecture level, the province level, and the physiographic macro-regional level.

Secondly, the combined data set and the validation data set are compared in order to identify any temporal deviation of credible records. An index for the number of disaster counties at each level was established to eliminate the impacts from large magnitudes of values. Accordingly, an index is be constructed as:

$$IF_i = \frac{F_i - \bar{F}}{\bar{F}} \times 100\%, (i = 1368, 1369, ..., 1911) \tag{3.1}$$

$$ID_i = \frac{D_i - \bar{D}}{\bar{D}} \times 100\%, (i = 1368, 1369, ..., 1911) \tag{3.2}$$

$$IC_i = \frac{C_i - \bar{C}}{\bar{C}} \times 100\%, (i = 1368, 1369, ..., 1911) \tag{3.3}$$

$$IW_i = \frac{W_i - \bar{W}}{\bar{W}} \times 100\%, (i = 1368, 1369, ..., 1911) \tag{3.4}$$

where F_i is the number of flood counties of each region, and $\bar{F}$ is the mean number of flood counties from 1368 to 1911. D_i is the number of drought counties of each region, and $\bar{D}$ is the mean number of drought counties from 1368 to 1911. C_i is the number of cold event counties of each region, and $\bar{C}$ is the mean number of cold event counties from 1368 to 1911. W_i is the number of drought counties of each region and $\bar{W}$ is the mean number of drought counties from 1368 to 1911. IF, ID, IC, and IW are the flood index, drought index, cold index, and wind index, respectively. While the mean number of disaster counties is assumed to represent the mean incidence of the corresponding disaster, the index is the percentage deviation from the mean level of the number of disaster counties. Therefore, if the index obtains a high positive value, this would indicate large exceedances from the average level of relevant disaster (Zheng, Zhang, and Zhou, 1993).

The boundaries of prefectures, provinces, and physiographic macro-region are illustrated in Figure 3.1 and Figure 3.2.

The processes above can provide evidence of the similarity between two data sets at the prefecture and upper levels. The reason is that data sets constructed from official histories have been verified to show satisfactory quality at the prefecture level (Zheng, Zhang, and Zhou, 1993; Hao, Ge, and Zheng, 2010) using the approach of the percentage change to the mean of the number of floods/droughts counties. Therefore, if the new data set has similar spatio-temporal variations with the data set from official histories at the prefecture and upper levels, it is acceptable to claim that the new data set has the potential to illustrate the overall picture of the environment of historical China.

Figure 3.1: Administrative Boundaries in 1911 (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016)

Thirdly, an approach based on the number of flood and drought counties (Zheng, Zhang, and Zhou, 1993) is employed to generate flood/drought grades to further qualify the new data set. The grade from the combined data set is compared with the grade from the validation data set to validate the adequacy of excess records or supplementary sources. Since the approach to generate the grade was validated on official histories, the grade from the validation data set is considered reliable in

Figure 3.2: Physiographic Macro-regions of China (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016), 1820. Different polygons with the same colour represent the subsystem of corresponding physiographic macro-region.

capturing actual precipitation. In this study, in order to facilitate the validation work, a dryness/wetness index in Shuntian Fu is constructed:

$$DW_i = ID_i - IF_i (i = 1368, 1367, ..., 1911) \tag{3.5}$$

where DW_i is the dryness/wetness index of Shuntian Fu for each year, ID_i is the drought index as shown in equation (2), and IF_i is the flood index as shown in equation (3). DW_i will be 0 if there was no record of a drought or flood. When the magnitude of DW_i is large, the drought index exceeds the flood index, thus implying lower precipitation. Employing similar criteria as previous studies (Zheng, Zhang, and Zhou, 1993; Zhang, 1996), the grade is divided into seven levels denoted by I to illustrate the annual situation of dryness/wetness condition. In this regard level 1 and level 7 should account for around 4% to 5%, Level 2 and level 6 for around 15%-20%, Level 3 and level 5 for around 35-40%, and Level 4 for around 35%-40% (Zheng, Zhang, and Zhou, 1993; Zhang, 1996; China Meteorological Administration Institute of Meteorology, 1981). The actual distributions may vary across different sources and locations, which means that the range of the proportion may not be strictly identical across different sources and different areas. As a result, the proportion from every grade may vary for the combined data set and the validation data set. More precisely, the relationship between I and DW_i for Shuntian Fu in the combined data set can be described as:

$$\begin{cases} DW_i \geq 370\%, & I = 7(Severe\,Dought) \\ 370\% > DW_i \geq 150\%, & I = 6(Dought) \\ 150\% > DW_i \geq 50\%, & I = 5(Subdought) \\ 50\% > DW_i \geq -50\%, & I = 4(Normal) \\ -50\% > DW_i \geq -150\%, & I = 3(Subwet) \\ -150\% > DW_i \geq -370\%, & I = 2(Wet) \\ DW_i \leq -370\%, & I = 1(Severe\,Wet) \end{cases} \tag{3.6}$$

Meanwhile, the relationship between I and DW_i for Shuntian Fu in the validation data set can described as:

$$\begin{cases} DW_i \geq 435\%, & I = 7(Severe\, Dought) \\ 435\% > DW_i \geq 200\%, & I = 6(Dought) \\ 200\% > DW_i \geq 53\%, & I = 5(Subdought) \\ 53\% > DW_i \geq -53\%, & I = 4(Normal) \\ -53\% > DW_i \geq -200\%, & I = 3(Subwet) \\ -200\% > DW_i \geq -435\%, & I = 2(Wet) \\ DW_i \leq -435\%, & I = 1(Severe\, Wet) \end{cases} \tag{3.7}$$

The dryness/wetness grades from the combined data set and the validation data set were generated to be compare with each other.

Fourthly, the grade of wetness/dryness based on the weighted frequency in the combined data set is constructed to be compared with the grade from the validation data set based on the number of flood/drought counties. Such a comparison serves to validate the parametrisation approach. If the weighted frequency is a consent parameter, it should have a roughly identical performance with a validated parameter. Again, to facilitate comparison, the index from the weighted frequency of Shuntian Fu in the combined data set are constructed as:

$$IF'_i = \frac{FF_i - \bar{FF}}{\bar{FF}} \times 100\%, (i = 1368, 1369, ..., 1911) \tag{3.8}$$

$$ID'_i = \frac{FD_i - \bar{FD}}{\bar{FD}} \times 100\%, (i = 1368, 1369, ..., 1911) \tag{3.9}$$

$$IC_i' = \frac{FC_i - \bar{FC}}{\bar{FC}} \times 100\%, (i = 1368, 1369, ..., 1911) \tag{3.10}$$

$$IW_i' = \frac{FW_i - \bar{FW}}{\bar{FW}} \times 100\%, (i = 1368, 1369, ..., 1911) \tag{3.11}$$

where FF_i is the frequency of flood records of Shuntian Fu for each year, and $\bar{FF}$ is the average frequency of flood records from 1368 to 1911. FD_i is the frequency of drought records of Shuntian Fu, and $\bar{FD}$ is the average frequency of drought records from 1368 to 1911. FC_i is the frequency of cold records of Shuntian Fu, and $\bar{FC}$ is the average frequency of cold records from 1368 to 1911. FW_i is the frequency of drought records of Shuntian Fu, and $\bar{FW}$ is the average frequency of drought records from 1368 to 1911. IF', ID', IC', and IW' are the flood index, drought index, cold index, and wind index for weighted frequency of records, respectively. Based on these indices, a dryness/wetness index is built as:

$$DW_i' = ID_i' - IF_i'(i = 1368, 1367, ..., 1911) \tag{3.12}$$

where DW_i' is the dryness/wetness index of Shuntian Fu for each year. Dryness/wetness grades can be re-organised into 7 levels, where the relationship between level I and DW_i' is illustrated as follows:

$$\begin{cases} DW_i' \geq 400\%, & I = 7(Severe\,Dought) \\ 400\% > DW_i' \geq 150\%, & I = 6(Dought) \\ 150\% > DW_i' \geq 50\%, & I = 5(Subdought) \\ 50\% > DW_i' \geq -50\%, & I = 4(Normal) \\ -50\% > DW_i' \geq -150\%, & I = 3(Subwet) \\ -150\% > DW_i' \geq -400\%, & I = 2(Wet) \\ DW_i' \leq -400\%, & I = 1(Severe\,Wet) \end{cases} \tag{3.13}$$

where level 1 and level 7 waccount for around 5%, level 2 and level 6 for around 20%, level 3 and level 5 for around 40% and Level 4 for 35%.

According to the comparisons above, it is possible to verify the potential of the new data set to illustrate county-level environmental features. Specifically, the wetness/dryness grade based on the percentage change to the mean of the number of floods/droughts counties can be only adopted at the prefecture or upper levels, and missing data are filled by Chebyshev polynomial interpolation and some historical criteria (Zheng, Zhang, and Zhou, 1993). Therefore, if the combined data set can generate a similar grade with the validation data set, the parametrisation and supplementary approaches are satisfactory enough to construct county-level data.

There are other validation approaches that will be considered as well. The first one is to re-conduct the validation work of Zheng, Zhang, and Zhou (1993) where the grade is be compared with actual records of precipitation in the Beijing. The grade from the weighted frequency is employed for the combined data set in such comparisons.

3.3 Result of data sets Comparison and Validation

3.3.1 Reliability of Historical Documents and Relevant Validation

There is some information systematically recorded by the government, such as the records from disaster report systems in the Qing dynasty. Data from these systems are considered to have higher reliability than historical documents (Zhang, 1996). The reason is that these systems were established at a crucial location where bureaucracy was well developed. Although disaster report system was conducted by the viceroy of each province, officials at every level from military affairs, salt administration, manufacturing, customs, and river affairs were responsible for reporting the situation of disasters. What is more, the emperor of the Qing dynasty would check the correction of reports sometimes and punish fake reports. Therefore, impossible for local officials cannot report too much fake disaster information to ask for tax relief or to fawn the emperor (Li, Xia, and Zhu, 2010).

Aside from government records, environmental records in the personal notes, diaries, travel notes, and other similar sources are undoubtedly reliable, since these kinds of records are not affected by any factors outside personal awareness.

Records in local gazetteers are less reliable than the sources above. Since the compilation of gazetteers was done in order to provide evidence for local officials for decision making and this was done normally by successors of the situations, records may be affected by memory lapses. There is also the possibility that the disaster information was hidden, exaggerated, or copied from other unreliable sources. Additionally, provincial records are less accurate than prefecture records, and prefecture records are less accurate than county annals since compilation was more frequent at the county level and the location of the events is nearer to the source collection (Su, 2018). Nevertheless, it is normally accepted that records at the county level for the Ming and Qing dynasty are credible, since local gazetteers played an important role in these periods. Specifically, the first focus for local officials after taking over former officials' position was to check the compilation of local gazetteers. The regulation to re-compile the local history in Qing dynasty required the restoration of the local gazetteers once every 60 years, but local compilation would have been activated earlier in actual practice (Zhang, 1996). Frequent re-compilations have guaranteed the continuity and integrity of records. Furthermore, the compilation of local history at the county level was usually very rigorous. Compilers would look up multiple reliable sources and do their own research before they enter a record. Also, many compilers were native residents who were familiar with local situations and would have experienced some of what they have recorded.

Another characteristic of environmental records in the historical documents is the possibility of interpretation errors for environmental events. More specifically, due to the ability of precise description and the ambiguity of ancient language, historical records can vary significantly for the same fact (Zheng, Zhang, and Zhou, 1993).

Because of the insufficient data from disaster report systems and scattered personal notes, the number of counties affected by disasters was applied to construct the environmental index on a broader border and time period (Hao, Ge, and Zheng, 2010). In addition, the number of counties affected can avoid interpretation errors in the historical records to some extent, given the belief that authors for historical records would not have miss-identified the type of disaster. For example, the author might recognise heavy rain as light rain, but he would not have recorded rain as drought. Academics have verified the validation of this parametrisation approach by comparing the dryness/wetness index based on the number of drought/flood counties with that based on the actual precipitation records (Hao, Ge, and Zheng, 2010; Zheng, Zhang, and Zhou, 1993; Li, Xia, and Meng, 2012; Zhang and Liu, 2002). To this end the reliability of local gazetteers in the Ming and Qing dynasty has been validated quantitatively.

3.3.2 Comparison of the Spatial Distribution of Records

First of all, aggregations of the number of annual disaster counties from 1368 to 1911 of the combined data set and the validation data set were compared at the physiographic macro-regional level as shown in Figure 3.3. The red bar represents the aggregate magnitude of the number of annual disaster counties and in the legend wind, cold, drought, and flood are represented as ranging from light to

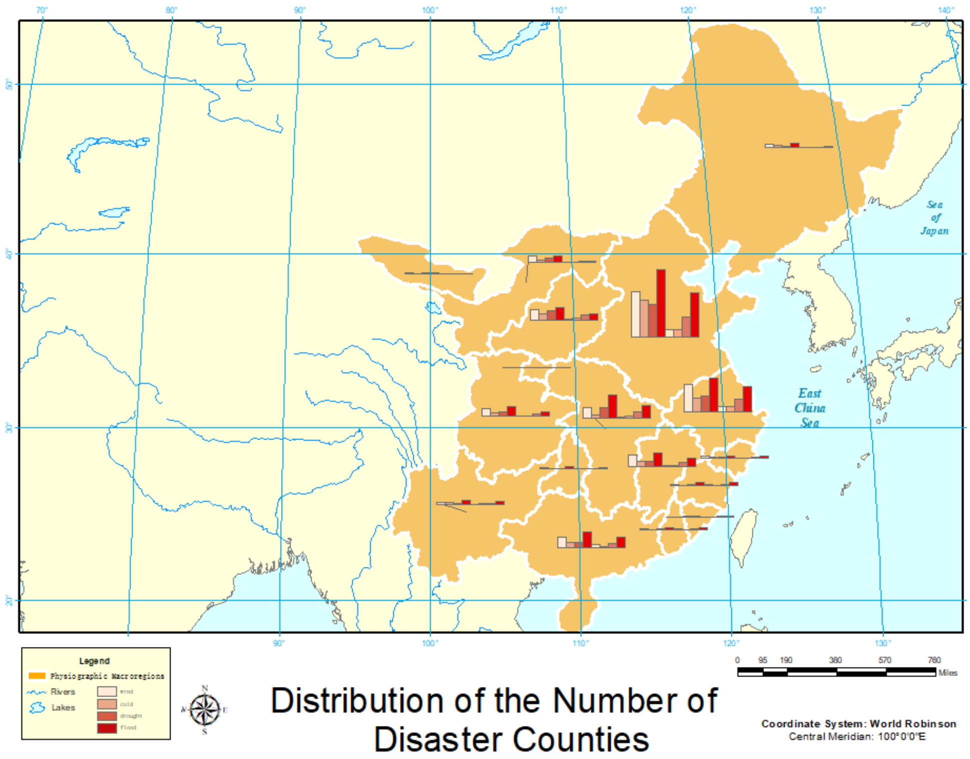


Figure 3.3: Distribution of the total number of annual disaster counties for the conservative dataset (four bars on the right side) and the combined dataset (four bars on the left side) at the Physiographic Macro-region Level (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Data is generated form "Chinese Three Thousand years Meteorological Records Collection" directly.

dark colour according to their value (from top to bottom). The number of flood counties in North China (light green area in Figure 3.2) remains the highest for the combined data set as well as for the validation data set. The distributions of flood and drought records in the combined data set are roughly the same as those in the validation data set. North China obtains the highest value of the number of flood or drought counties. The number of records of Middle Yangtze region (brown area in Figure 3.2) and Lower Yangtze region (Dark purple area in Figure 3.2) are approximately at the same level for floods and droughts following that of North China. For the rest of the areas, differences are not significant enough to question the reliability of the combined data set. The deviations of wind and cold records seem to be large. Especially for the wind, in the combined data set, wind obtains the second-largest quantity of records in every region, but accounts only for a small proportion of records in the validation data set. However, if one only considers wind, the distributions for the combined data set and the validation data set would be comparable. North China and Lower Yangtze again are characterised by the highest quantity of records in both of these two data sets. Interestingly, the number of wind counties in Lingnan area (dark green in Figure 3.2) is greater than that in Southeast Coast (light pink in Figure 3.2) for both of these two data sets. Overall, one can be somewhat confident that environmental features from the combined data set generally coincide with that of the validation data set.

To further confirm the consistency at a lower regional level, the number of disaster counties was aggregated to the province level. The distributions are depicted in Figure 3.4 and Figure 3.5. Figure 3.6 illustrates the differences between the combined data set and the validation data set by showing the aggregate number of disaster counties for each province, where the height of darkest blue bar represents the value of the total number of flood counties across the whole province from 1368 to 1911. In the legend, the boxes represent wind, cold, drought, and flood where

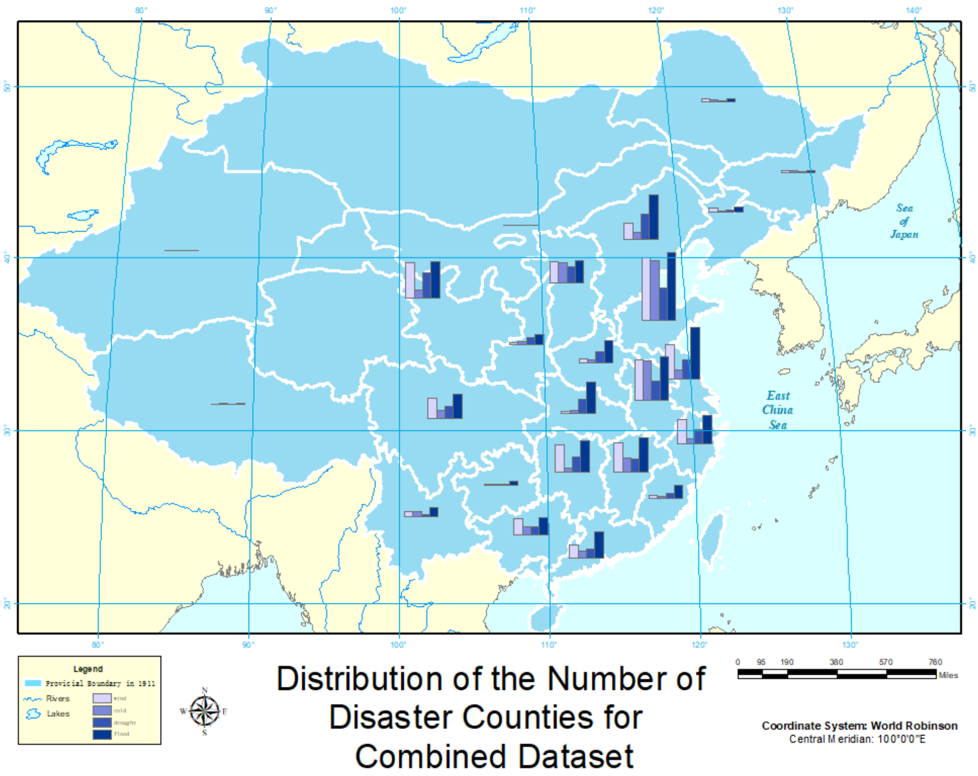


Figure 3.4: Distribution of the Number of Disaster Counties for the combined dataset at the province Level (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Data is generated by combining records from "*China Meteorological Disaster Dictionary (in Chinese)*" and "Chinese Three Thousand years Meteorological Records Collection"

values increase with the darkness of the color scheme (from top to bottom). Except for the Gansu province, which shows massive differences in records of wind and cold, the number of flood counties and drought counties tells a similar story based

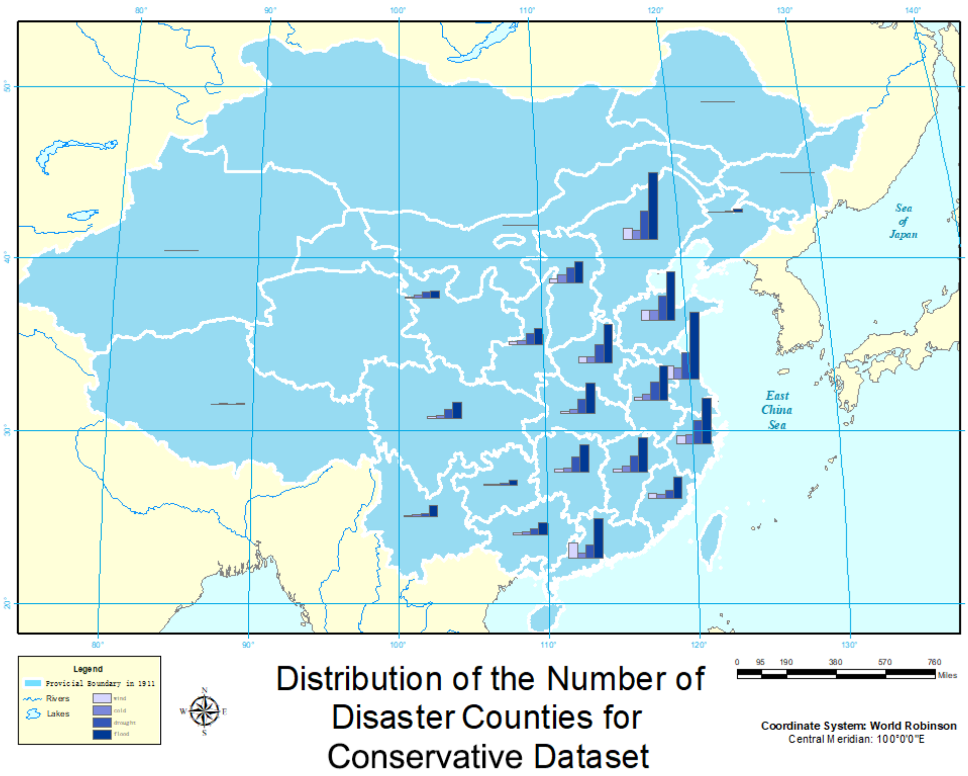


Figure 3.5: Distribution of the Number of Disaster Counties for the Conservative Dataset at the province Level(Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Data is generated form "Chinese Three Thousand years Meteorological Records Collection" directly.

on the two data set. Provinces near to the east and south-east coast have relatively more records on flood. In contrast, the Fujian province, which is located opposite of the Taiwan province (island to the south-east coast), has left a relatively lower quantity of records not only for the combined data set but also for the validation

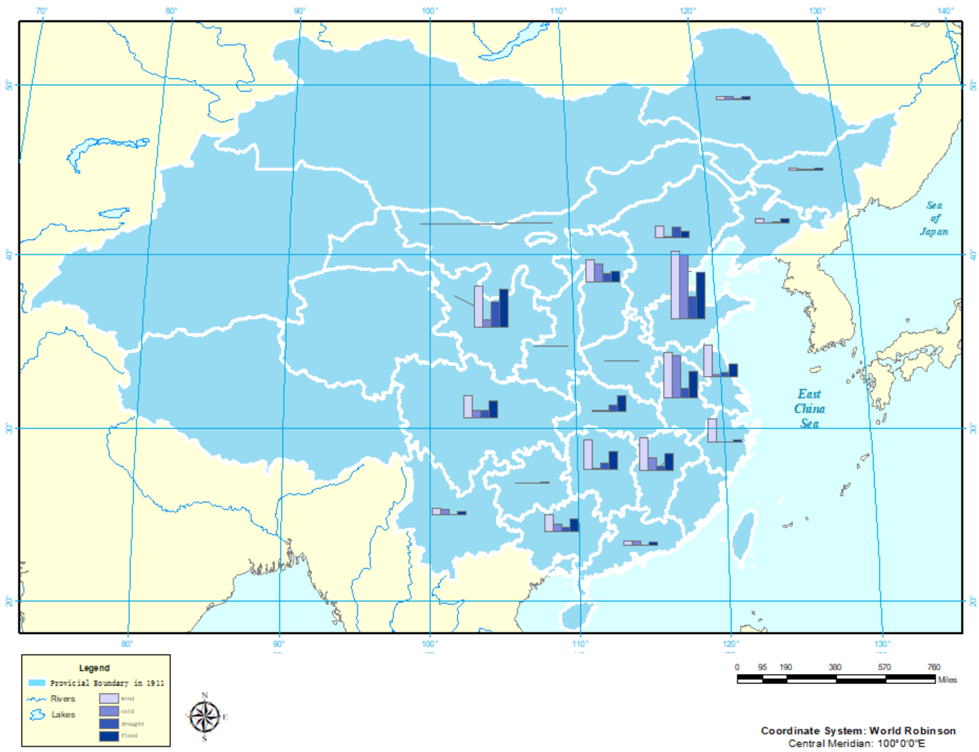


Figure 3.6: Distribution of the Number of Disaster Counties for the Combined Dataset Minues the Conservative Dataset at the province Level(Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Data is generated form "Chinese Three Thousand years Meteorological Records Collection" directly.

data set. There is thus arguably considerable consistency between the two data sets. What is more, both data sets tell us that north China suffered more in terms of droughts in the historical period. In this regard, flood and drought records supplemented from additional sources after the modification process are capable of capturing the features of the historical environment (recall that the index from validation approach is consistent with the index from actual records of precipitation (Zheng, Zhang, and Zhou, 1993) and other relevant studies (Zheng, Hao, and Ge, 2005; Bi et al., 2016)). As a result, much more indirect information of relevant disasters would be included in the combined data set.

To check whether the consistency of the combined data set with the validation data set would be altered when regions were further divided, the total number of disaster counties aggregated to the prefecture level. Since the number of total prefectures is large, prefectures with high values in the number of disaster counties were filled with the dark colour to present the concentration of environmental records. According to the previous comparison, it should be cautious to consider that records of floods and droughts are more comparable between the two data sets. The distribution of flood and drought was generated as shown in Figure 3.7, Figure 3.8, Figure 3.9, and Figure 3.10.

Examining the figures of disaster distributions at the prefecture level one can conclude that the dark areas generally overlap across two data sets except for the prefectures of the Gansu province. What is more, if one compares the figures for the number of flood counties with the figures of the number of drought counties, the distributions display high similarity with each other as well. This can be taken as evidence that the distributions of the locations where the environmental events were recorded are consistent across the combined data set and the validation data set. Considering the comparisons at the physiographic macro-regional level, province level and prefecture level, the spatial distribution of records in the combined data set in general coincides with the validation data set at an accept-

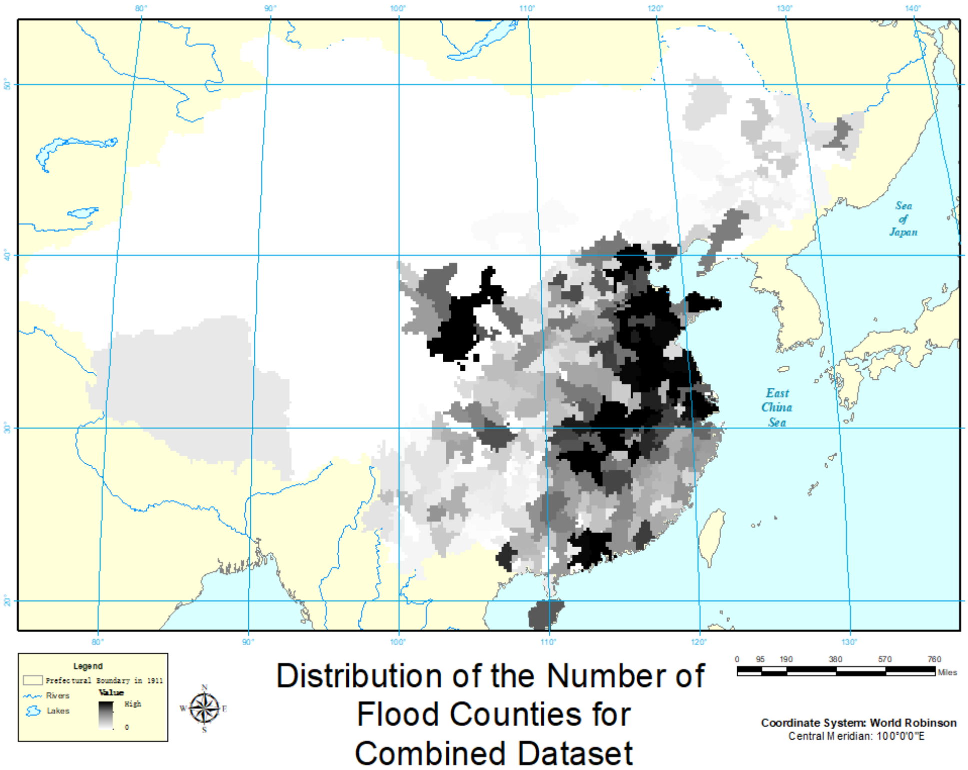


Figure 3.7: Distribution of the Number of Flood Counties for the Combined Dataset at the prefecture Level (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Data is generated by combining records from "*China Meteorological Disaster Dictionary (in Chinese)*" and "Chinese Three Thousand years Meteorological Records Collection".

able level. Nevertheless the occurrence of some significant deviations, such as the excess records for the Gansu province, should be kept in mind.

The consistency across data sets is partially depicted in the scatter plot in Figure

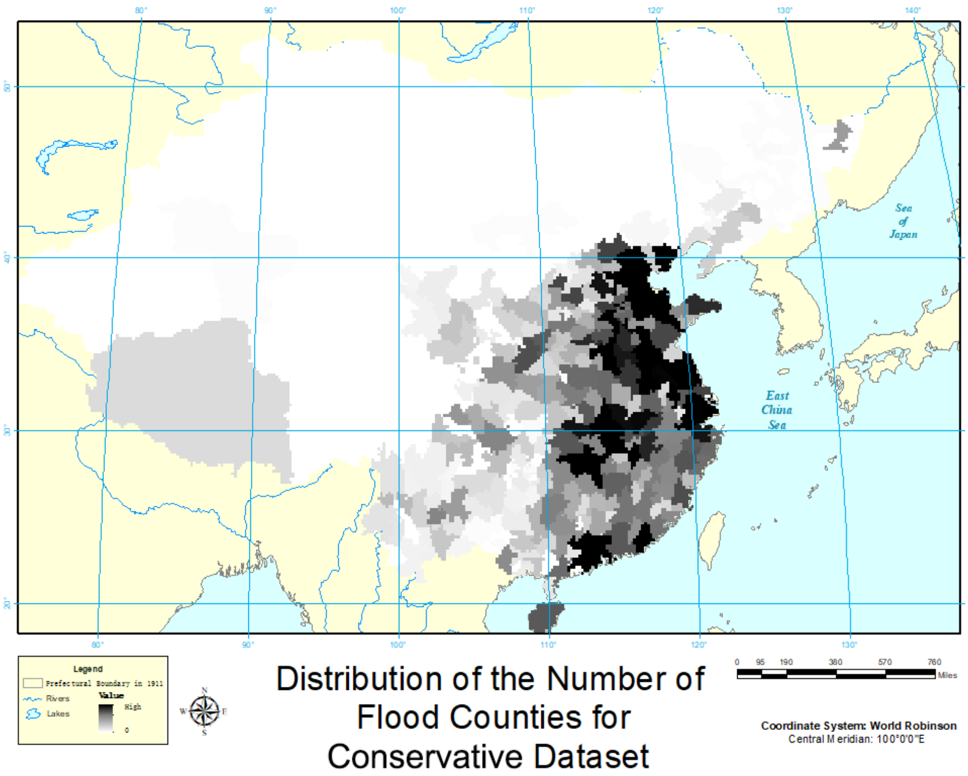


Figure 3.8: Distribution of the Number of Flood Counties for the Conservative Dataset at the prefecture Level(Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Data is generated form “Chinese Three Thousand years Meteorological Records Collection” directly.

3.11. The majority of prefectures show similar spatial relationships in the combined and validation data sets since the points are located near the linear trend. There are nevertheless several prefectures that contain significantly excess records. Data for drought and flood seem to be more concentrated in the plot, but data for

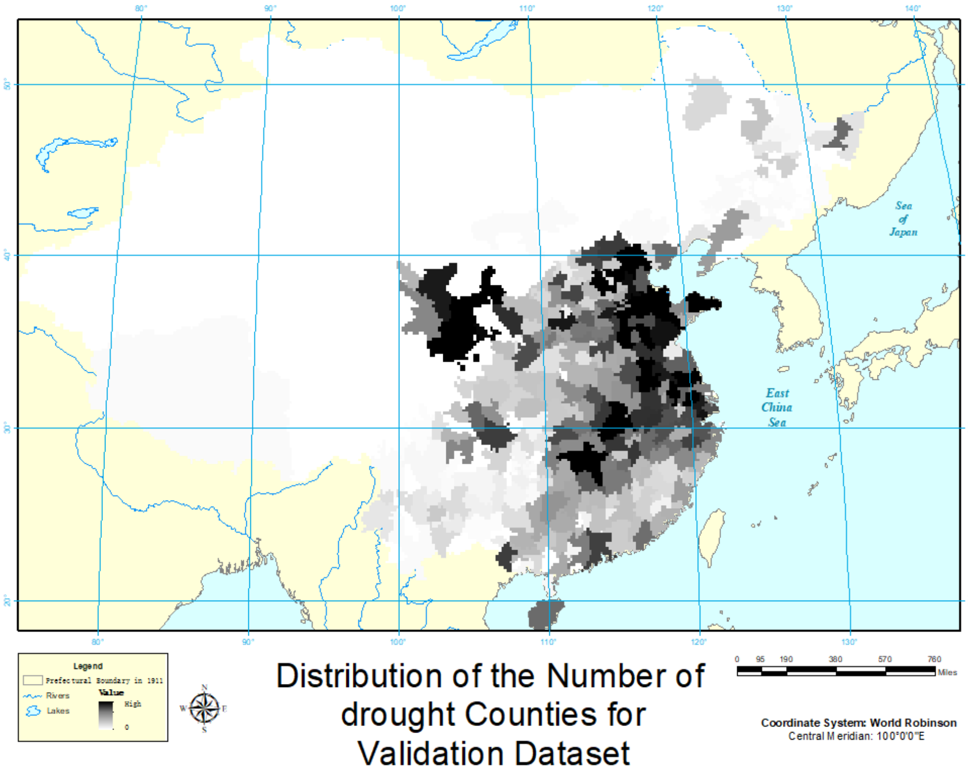


Figure 3.9: Distribution of the Number of Drought Counties for the Combined Dataset at the prefecture Level (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Data is generated by combining records from "*China Meteorological Disaster Dictionary (in Chinese)*" and "Chinese Three Thousand years Meteorological Records Collection"

cold and wind are a bit discrete, while the trend is relatively apparent. Spatial characteristics thus do not deviate a lot in the combined data set at the prefecture level.

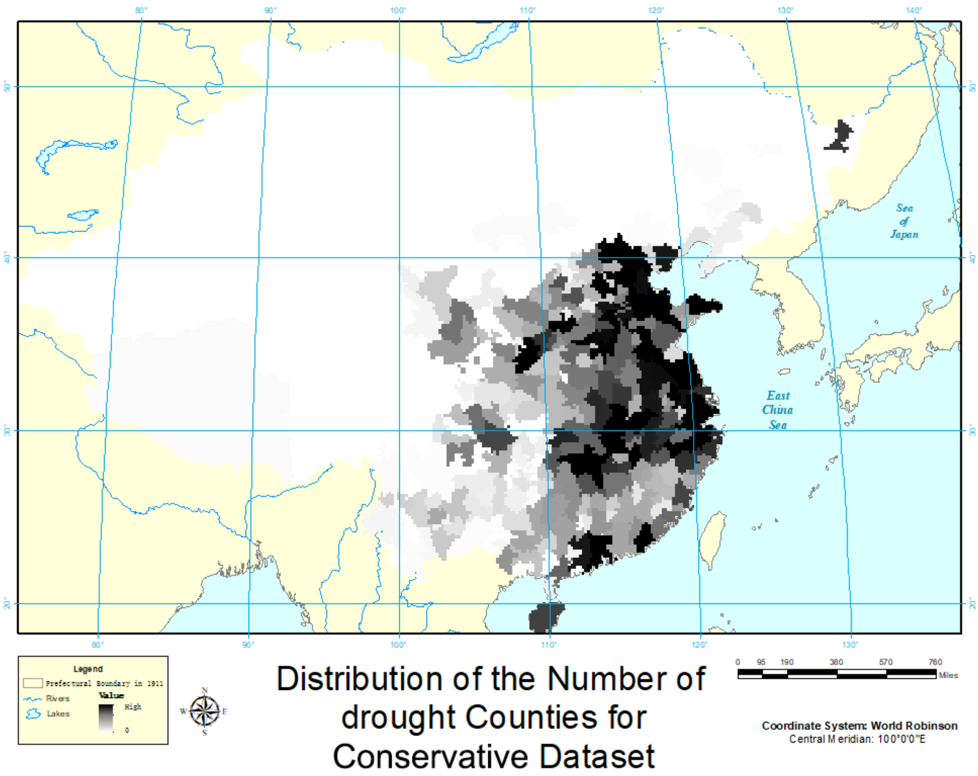


Figure 3.10: Distribution of the Number of Drought Counties for the Conservative Dataset at the prefecture Level (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Data is generated form "Chinese Three Thousand years Meteorological Records Collection" directly.

However, since the combined data set is constituted by records from the conservative source and modified records from supplementary sources, it is natural to obtain considerable similarity between the combined data set and the validation data set. The point is that if the combined data set does not differ significantly

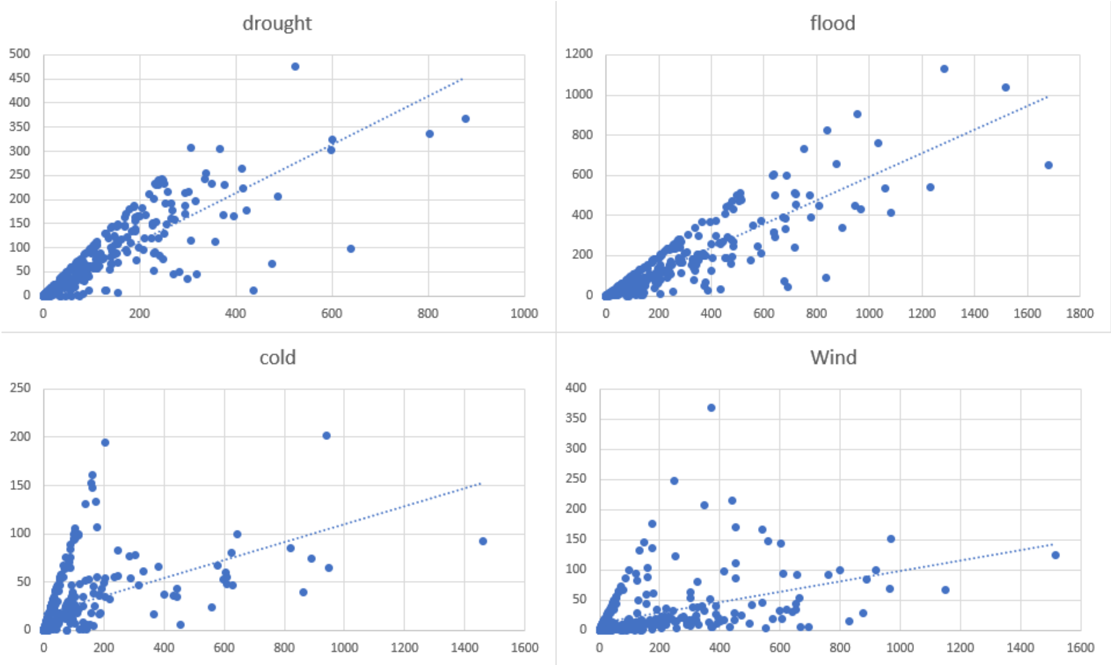


Figure 3.11: Scatter Plot of the Number of Annual Disaster Counties in the Combined Dataset (Horizontal Axis) Versus in Coniservative Dataset (Verticle Axis) (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Data is generated form "Chinese Three Thousand years Meteorological Records Collection" and "*China Meteorological Disaster Dictionary (in Chinese)*" .

from the validation data set, the approach to supplement official history in this paper should be validated as being reliable. Therefore, the modified supplementary records can be accepted as featuring satisfactory reliability. In this regard, even if the supplementary sources were not in the same stylistic rules and layout, it is possible to generate the weighted frequency which eliminates the influences of non-uniformity or low quality of sources. However, validations according to figure comparisons of the spatial distribution are ambiguous and only provide qualitative concepts.

It might be concerned that if the excess records in "*China Meteorological Disaster Dictionary (in Chinese)*" are a result of multiple duplicate records of the same event, the correlation between these two sources should be relatively high, while another argument could be that duplication might not always be in similar proportion to the actual number of events. As shown in Figure 3.11, different regions may have a different thresholds for recording abnormal events. Although it is acceptable to infer that more severe disasters are more likely to be recorded, the high frequency of the event to be recorded not only relates to the actual severity of this event but also implies possible influence (both physically and psychologically) on neighbouring regions and subsequent periods sometimes. Nevertheless, the amount of the number of event records is also dependent on the availability of historical documents. If a county was more developed with a high proportion of educated citizens, it is reasonable to assume that the collection of records in this county would be larger than that in a less developed county, even if the influence is less severe in the more developed county. Accordingly, quantitative validation for spatial distribution in this case does not make much sense. A possible solution to this issue is to calculate the percentage deviation from local means for the records series to retain the variation across time. What is more, the modification for records based on the combination process is expected to reduce the issues just outlined to an acceptable level.

3.3.3 Comparison for the Inter-temporal Distribution of Records

Table 3.1: Correlations of Index across Physiographic Macro-region by Environmental Category

Name of Macro-region	γ_{IC}	γ_{ID}	γ_{IF}	γ_{IW}
ManChuria	0.433	0.365	0.736	0.233
North China	0.616	0.945	0.902	0.576
Northwest China	0.641	0.677	0.582	0.283
Upper Yangzi	0.689	0.871	0.803	0.526
Middle Yangzi	0.721	0.966	0.841	0.559
Lower Yangzi	0.780	0.985	0.886	0.484
Southeast Coast	0.963	0.998	0.995	0.826
Lingnan	0.623	0.925	0.862	0.648
Yungui	0.502	0.855	0.856	0.444

Data sources and notes

* List of the physiographic macro-region are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). .

** γ denotes the correlation of a given index between the combined dataset and the conservative dataset. Correlation retains three decimal places.

*** All these correlations are significant at 0.01 level.

The indices generated from the combined data set were compared to those from the validation data set at the physiographic macro-regional level, province level and prefecture level. Firstly, the larger region was considered. As shown in Table 3.1, the correlation for each index varied across different macro-regions. Except for the flood index, none of these indices performs well in the ManChuria area. Correlations of the cold index, drought index, and wind index are relatively low (0.433, 0.365, 0.233 respectively). For the rest of the regions, the drought and flood indices from the combined data set are highly correlated with those from the validation data set, where the correlations are all over 0.8 except for Northwest China. What is more, regions with higher correlations, such as North China, Lower Yangzi, Southeast Coast, and Lingnan, are relatively more populated and developed (Cao, 2002a). As for the cold and wind events, indices are only have high consistency between the combined data set and the validation data set in the Southeast Coast, where for the rest regions the overall correlation is around 0.6 for the cold and 0.5 for the wind index. From the comparison it can be concluded that records for drought and flood have similar distributions. Nongovernmental records were strongly compatible with the stories illustrated in official history for droughts and floods, but did not completely agree with the description of cold and wind events. Although the general correlations for cold and wind are tolerable, the weaker connection between the combined data set and the validation data set indicates non-negligible divergence for the temporal traits of records in supplementary sources. People might record environmental events based on their knowledge. As a result, some records could refer to events that were not officially considered as natural disasters and some might be missed if the occurrence of extreme environmental events were far away from the populated region. Therefore, the deviation of records for cold and wind may be due to the large differences of identification for extreme events across different authors. What should be noted is that for most of official history it was standard practise to record abnormal deviations rather than normal weather fluctuations (Zhang, 1996). However, records in more personal manners would contain some general weather conditions. In this regard, wind and cold occurrences in the combined data set might represent extreme events and normal weather variation at the same time. Additionally, the mean subtracted in the index can only remove the steady trait of the temporal distribution but cannot eliminate the fluctuations due to climate change.

Following the steps taken in the last section, the reliability of the combined data set is validated at a more disaggregated spatial level. Correlations at the province level are shown in Table 3.2. The conclusion is roughly the same compared with the comparison at the physiographic macro-regional level. For the indices of drought and flood the overall correlations are relatively high across different provinces. Correlations for the drought index are over 0.8 except for the Jilin, Fengtian, Gansu, and Heilongjiang provinces, where Jilin and Xinjiang do not have records for droughts. Correlations for flood are over 0.7 except for the Neimenggu, Jilin, Fengtian, Gansu, and Heilongjiang provinces. Additionally, the Fengtian fu, Jilin, Heilongjiang provinces are in the Manchuria region, and Gansu is in Northwest China. Neimenggu and Xinjiang are not calculated in the physiographic macro-regions. As for the index of cold and that of wind, only 58% of provinces have correlations over 0.6 for the former, and 38% for the latter. Thus, indices of droughts and floods appear to be comparable and consistent between the combined data set and the validation data set. Specifically, provinces with higher correlations should be viewed as having sufficient records officially and privately, since supplementary sources of this paper contain some records with different stylistic rules from var-

Table 3.2: Correlations of Index across Provinces by Environmental Category

Name of Province	γ_{IC}	γ_{ID}	γ_{IF}	γ_{IW}
Yunnan	0.392	0.813	0.807	0.398
Neimenggu	0.519	0.865	0.576	
Jilin	0.180		0.522	
Sichuan	0.650	0.862	0.786	0.506
Fengtian	0.413	0.251	0.516	0.168
Anhui	0.412	0.940	0.722	0.327
Shandong	0.431	0.826	0.719	0.425
Shanxi	0.662	0.901	0.867	0.442
Guangdong	0.749	0.983	0.951	0.915
Guangxi	0.431	0.882	0.712	0.311
Xinjiang	1.000		1.000	
Jiangsu	0.957	0.975	0.874	0.553
Jiangxi	0.491	0.974	0.818	0.360
He'nan	0.999	0.999	1.000	0.996
Zhejiang	0.999	0.996	0.986	0.495
Hubei	0.970	0.911	0.778	0.954
Hu'nan	0.823	0.922	0.770	0.389
Gansu	0.325	0.408	0.355	0.164
Zhili	0.972	0.957	0.984	0.741
Fujian	1.000	1.000	1.000	1.000
Xizang	1.000	1.000	1.000	1.000
Guizhou	0.873	0.976	0.925	0.694
Shaanxi	0.991	0.999	0.998	0.949
Heilongjiang	0.325	0.228	0.477	0.155

Data sources and notes

* List of province are obtained from version 6, CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

** γ denotes the correlation of a given index between the combined dataset and the conservative dataset. Correlation retains three decimal places.

*** All these correlations are significant at 0.01 level.

ious origins, including private records. Accordingly, counties with records from supplementary sources are distributed parallel to records in official history after modifying records according to the quality evaluation for relevant sources, at least for droughts and floods records.

Table 3.3: Correlations Distribution of Index across Prefectures by Environmental Category

Item	cold	drought	flood	wind
$P<0.01$	99.24%	99.62%	100%	93.25%
$P<0.05$	99.24%	99.62%	100%	95.63%
$P<0.1$	99.24%	99.62%	100%	97.22%
$\gamma>0.6(1)$	60.15%	88.51%	81.88%	37.02%
$\gamma>0.8(1)$	50.57%	68.58%	47.83%	31.06%
$\gamma>0.6(2)$	60.15%	88.51%	81.88%	35.51%
$\gamma>0.8(2)$	50.57%	68.58%	47.83%	29.80%

Data sources and notes

* Fields “cold”, “drought”, “flood” and “wind” denote the proportion satisfied the condition in “Item” for the cold index, drought index, flood index, and wind index.

** (1) denotes the proportion based on the prefectures with correlations at 0.1 level. (2) denotes the proportion based on the prefectures with correlations at 0.01 level. P denotes the P-value of correlations.

For the case that the temporal distribution deviates when smaller regions were considered, correlations at the prefecture level were assessed to observe the reliability and spatial variation of disaster index series. Prefectures without significant correlations or without records were omitted for each category (cold, drought, flood,

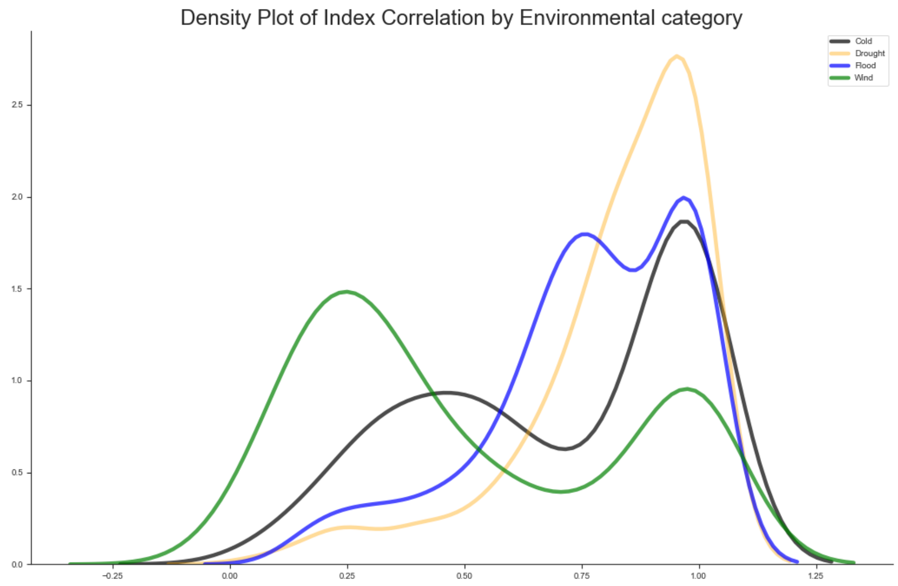


Figure 3.12: Density Plot of Index Correlation between the new dataset and the second dataset at the prefecture Level by Environmental Category.

and wind). According to the correlations at the prefecture level, conclusions from the analysis of the macro-regional level and province level are further substantiated. From Table 3.3, over 99% correlations for cold, drought and flood events are significant at the 0.01 level, but only 93.25% correlations for wind are significant at the same level. The indices of drought and flood from the combined data set are highly correlated with those from the validation data set, where 88.51% γ_{ID}s have correlation coefficients greater 0.6, and 81.88% γ_{IF}s are greater than 0.6. Performance of γ_{IC} is moderate, since 88.51% prefectures have relatively high correlation ($\gamma_{IC} > 0.6$). However, the index of wind from the combined data set does not resemble the temporal distribution of the wind index from the validation data

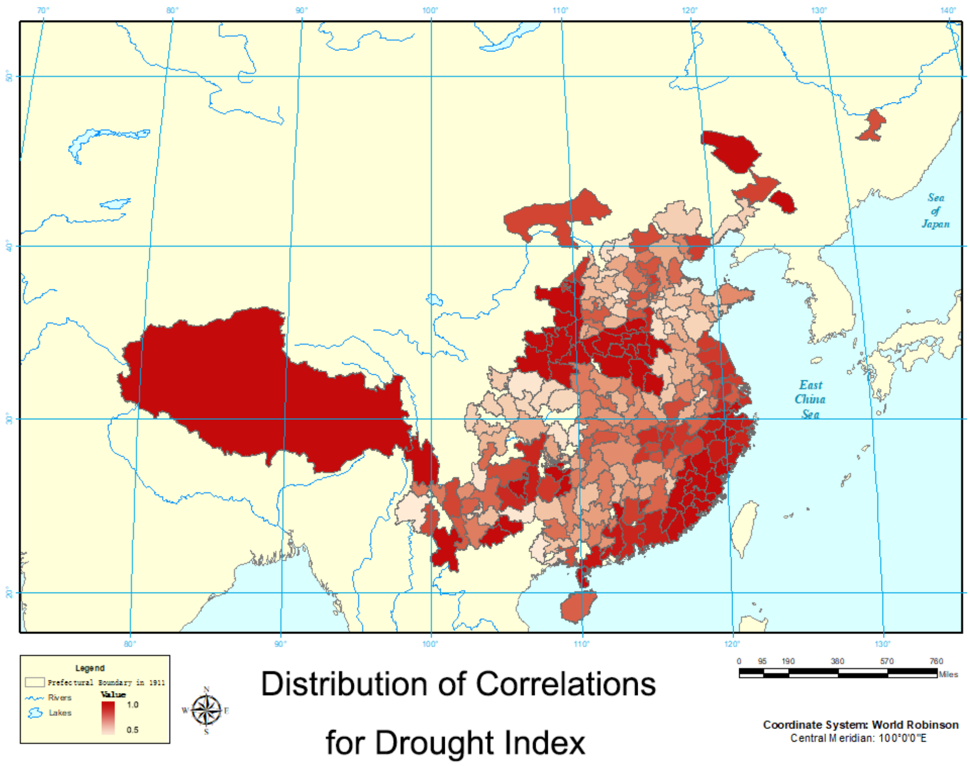


Figure 3.13: Distribution of Correlations for Drought Index at the prefecture Level (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

set, in that over 60% γ_{IW}s are less than 0.6. The specific distributions of these correlations for each category are illustrated in Figure 3.12. To be comparable with previous analysis, Figure 3.13 and Figure 3.14 show spatial distributions of correlations for the drought index and flood index, respectively. Prefectures with correlations over 0.5 are presented, where a darker red denotes a higher correlation. Prefectures near to the coast show a satisfactory linkage between additional

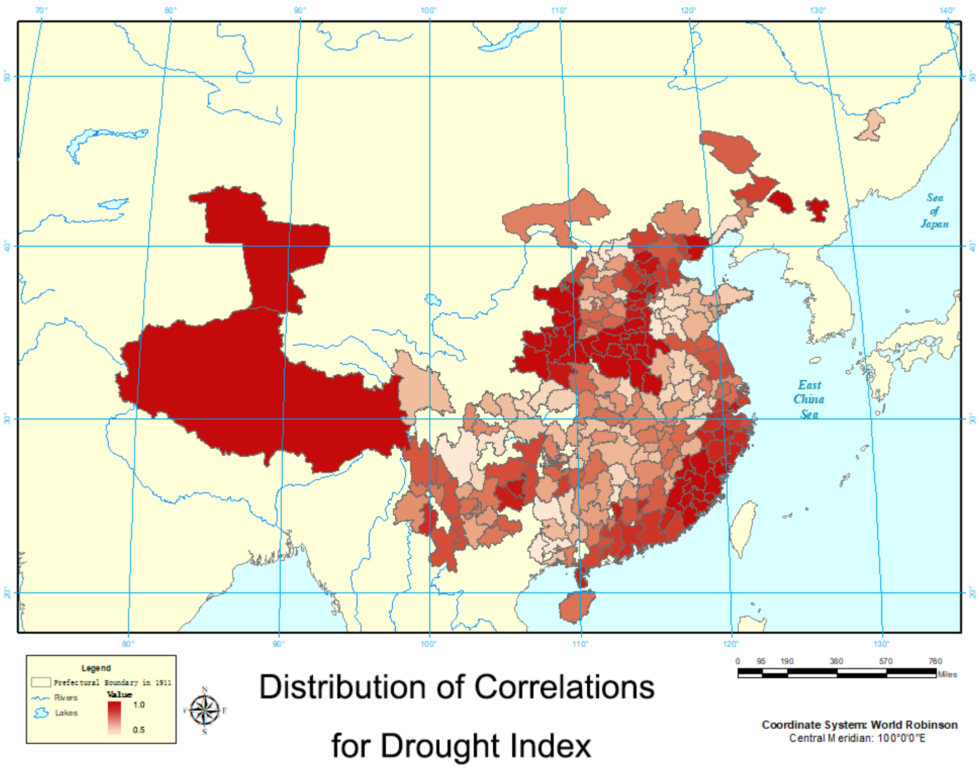


Figure 3.14: Distribution of Correlations for Flood Index at the prefecture Level (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

sources and the conservative source. However, prefectures in North China do not perform equivalently to province-level and the physiographic macro-regional level.

3.3.4 Validation for Supplementary Approach

According to previous comparisons, one should be careful to confirm that the approach of modifying records by estimating the quality of different sources does not deviate from the distribution of recorded counties spatially and temporally to a large scale. Especially for the counties with droughts records and floods records, the spatial and temporal distributions should have higher similarity compared with counties with cold records and wind records. Although there are many low correlations across the entire country, this may not constitute a violation of reliability since there are almost certainly missing records in the conservative sources as mentioned before. Therefore, a low correlation could imply a considerable number of missing records in official sources. As illustrated in Figure 3.4, the number of wind counties and the number of cold counties for the combined data set are significantly larger than those for the validation data set. In this regard, if the flood records and drought records in the combined data set have shown sufficient uniformity in a given source, it is logical to accept that other records for this source should depict a similar level of uniformity, although the correlations of the cold index and the wind index between the combined data set and the validation data set are usually much lower (as can be seen in Table 3.2). Furthermore, if the modification approach for data combination in this study fits one of these sources and can correct the record to eliminate some non-uniformity, it is tolerable to accept that such an approach fits other sources as well, although there must have been considerable space for improvement through this approach.

It is reasonable to expect that the combined data set is able to fill missing records in the validation data set (from conservative source). However, such extensions should not alter the degree to which the real world is mapped from the historical records. Accordingly, reliability could be validated when actual environmental variation is captured by this approach. In this regard, drought and flood indices generated from the percentage deviation from the mean number of drought/flood counties in Beijing should match the actual precipitation from 1724 to 1950 (Zheng, Zhang, and Zhou, 1993). Therefore, if we compare the drought index and flood index of Shuntian Fu (of which Bejeing is part of) between two data sets, it is possible to validate the reliability of the data extension. Since the correlations of the drought index and flood index in Shuntian Fu are over 0.8, one can conclude that these indices from the combined data set fit the actual precipitation to a sufficient extent. Thus, the modification approach using supplementary sources can be considered credible, and thus it is acceptable to view the combined data set as a more sufficient data set than the data set from conservative source. Moreover, since previous studies (Zheng, Ge, Fang, et al., 2007; Zhang, 1996) have suggested that historical records for droughts and floods should be converted into grades of dryness/wetness, it is necessary to build the dryness/wetness grades of the Beijing to validate the supplementary approach further.

One can view the deviation for the combined data set from the validation data set as deviation from actual precipitation. Fortunately, among 544 years, there are 331 years with the same dryness/wetness grades, 188 years with 1 unit difference, and 21 years with two units differences. The only four years with exceptions of over three units differences accounted for less than 1% in the total number of years. Specifically, the correlation between the two data sets for dryness/wetness grades is 0.82 at 0.01 significance level, which strongly suggests that the data extension can present precipitation in Shuntian Fu fairly well. Therefore, the supplementary approach extending the conservative source is validated at least for Shuntian Fu. Unfortunately for other prefectures there is not sufficient data to conduct validation for the application of the dryness/wetness index in official history, and it is beyond our capacity to validate these grades for other prefectures due to the lack of availability of actual precipitation records.

According to the analysis above, the dryness/wetness grade for the combined data set strongly coincides with the grade from the validation data set, at least in the Beijing area (part of Shuntian Fu). It is feasible to consider that the supplementary approach in this study should provide equivalent effectiveness in mapping the real world with the conservative source under the same parameterisation approach if records in the conservative source perform well in this regard. As a result, validation for the Beijing area indicates the reliability of the extension data in other regions. In this regard, if the grade from the validation data set matched actual precipitation, the grade from the combined data set should also do so, while if the grade from the validation data set did not fit the variation of actual environmental events, the grade from the combined data set might also depart significantly from this. Since the performance in mapping the real world with the extension data in the combined data set is roughly equal to that of the validation data set, the supplementary approach (combining records based on quality assessment) is expected to provide more information especially where the correlations for the disaster index are lower. Thus, a comparison of dryness/wetness grades between the two data sets is proposed to illustrate the improvement of supplementary sources over the conservative source.

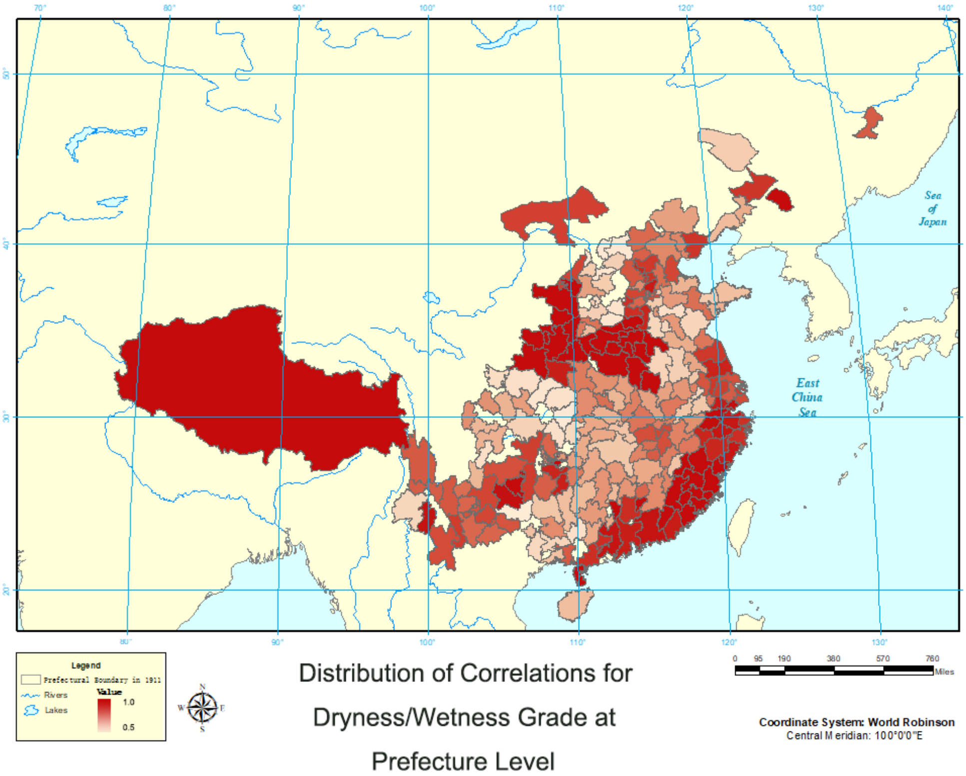


Figure 3.15: Distribution of Correlations for Dryness/Wetness Grade at the prefecture Level (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016).

However, since the criteria to build the grade according to the dryness/wetness index varies across different regions, the grade should be generated separately for each region. It is possible that the differences of dryness/wetness index across the whole periods are relatively small, less than 1% for example. In this case, the grade should be divided into less than seven levels. Differences in the criteria magnitude of dryness/wetness index between neighbouring grades should be over 1%. For instance, if I=4 when $50\% > DW_i \geq -50\%$, the upper bound for I=5 should be greater than 51%. The chosen quality threshold for the grade construction in this study is to have at least 35% of grade 4, around 40% of grade 3 and grade 5 in total, and around 20% of grade 2 and grade 6. To investigate this dryness/wetness grades at each level for the combined data set and the validation data set are constructed and compared. Figure 3.15 shows the spatial distribution of correlations of grade indices between the combined data set and the validation data set. In general the correlation for each prefecture between the combined data set does not vary spatially at a significant level.

3.3.5 Validation for the weighted frequency Approach

Since the number of disaster counties could not be used at the county level, instead the frequency of corresponding disaster records is considered in this study. According to previous research, the frequency can not be employed directly since records from different source do not have uniformity and thus would not be able to map the real world (Chu, 1926). Even if the data attained good uniformity, frequency as a measure is considered to be less reliable than the number of disaster counties (Zhang, 1996). The argument for this is that the number of counties with records within a year should not capture exaggerated personal perceptions of disasters being related. Moreover, records for different severity levels would be counted identically. To be specific, authors of records might record light rains three times within a year for a given region (a county for example) but heavy rains only once for another county, which does not mean that the total precipitation of light rain incidences would have exceeded that of the one event of heavy rain. However, the frequency of recorded events in the former county would exceed that of the latter. However, if the severity level of disaster is higher, a larger area should be affected where records exist for the corresponding disaster. This is why the frequency has not been widely employed in previous studies. Nevertheless, when considering records from multiple sources, more severe disasters should be recorded more. In this case, disaster frequency would provide accurate information to map the real world when sources are sufficient. Therefore, if the non-uniformity of different sources was eliminated and the various sources were available, the frequency of disaster could be adopted without much concern.

From the validations above, the approach to supplement conservative records in this research is deemed acceptable and should provide more information for relevant disaster events. As a result, records modified from supplementary sources were counted to extend the conservative source with approximately equivalent uniformity. In this regard, the frequency of records does not represent the frequency of disaster occurrence, but the number of records within a given region. More severe disasters should be represented by a larger number of disaster records. Moreover, since records in supplementary sources were assigned reliability weights for various volumes, the impact of exaggeration effects of excess records should be corrected at an approved level. In order to validate the performance of this correction for excess records, the weighted frequency of records was compared with an actual environmental variation which was matched by dryness/wetness grades in the Beijing area. Therefore, the dryness/wetness grades at the prefecture level were constructed according to the weighted frequency of records. Shuntian Fu again was considered to possess the best performance for the grade from the conservative source. If the grade from the weighted frequency of records matched the grade from the conservative source at an acceptable level (correlations over 0.8), it is feasible to accept the approach for frequency modification. Furthermore, county-level grades can be constructed.

Among the 544 years in Shuntian Fu, 313 years have the same levels, which accounts for over 57% in all periods. 205 years have deviations in grade by 1 unit, and 22 years have deviations by 2 unit. The correlation between the two grades is 0.80 at a 0.01 significance level. There are 4 exceptions for the years 1436, 1768, 1780 and 1818. For these exceptional years, grades from the conservative source are around 4, but grades from supplementary sources suggest excess floods in 1436 and 1780, and excess droughts in 1768 and 1818. Overall these two grades show similar time trends as illustrated in Figure 3.16.

3.3.6 Additional Validation

It is also helpful to compare the grades in this study with grades generated by other studies. As referenced before, Zheng, Zhang, and Zhou (1993) generated the dryness/wetness grade for the Beijing areas from 1471 to 1950 based on the number of drought/flood counties in official history. Therefore, the grade from the conservative source used here should match Zheng's work. What is more, there are reconstructed series for precipitation in the Beijing from 1724 to 2005 (Zhang and Liu, 2002; Li, Xia, and Meng, 2012; Lan, Hao, and Zheng, 2015; Wei, 2007).

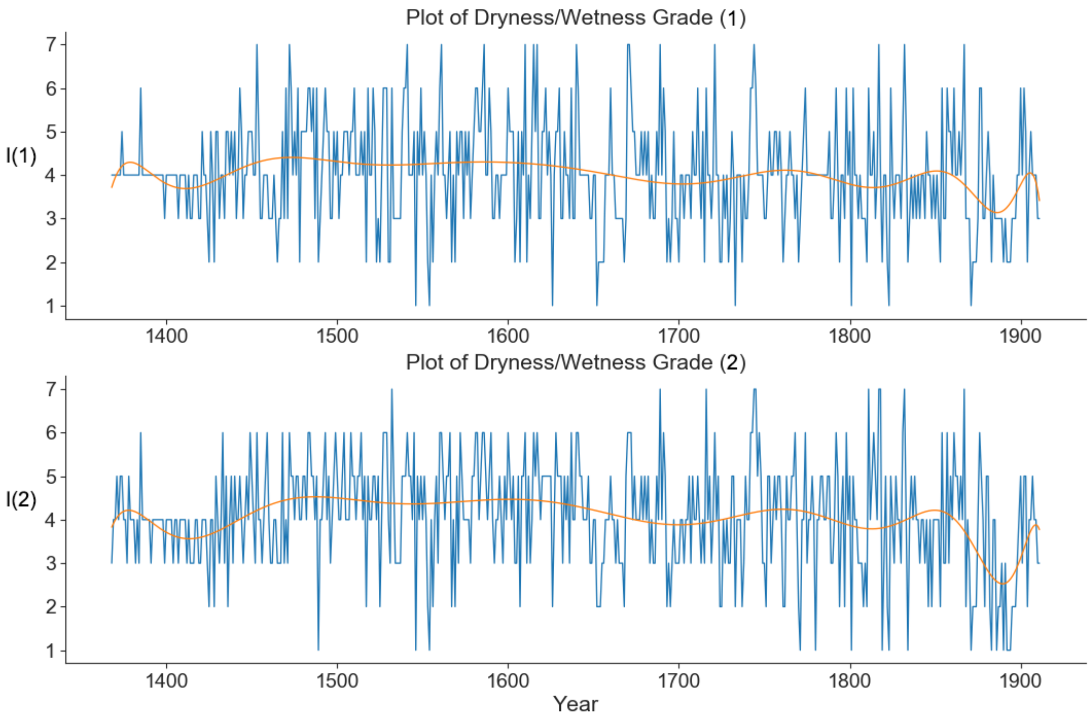


Figure 3.16: Plot of Dryness/Wetness Grades in Shuntian Fu. (1) is the grade generated from conservative source based on the number of drought/flood counties, and (1) is the the grade generated from the combined dataset based on the modified frequency of drought/flood records. Red curves are polynomial fitting from 1368 to 1911

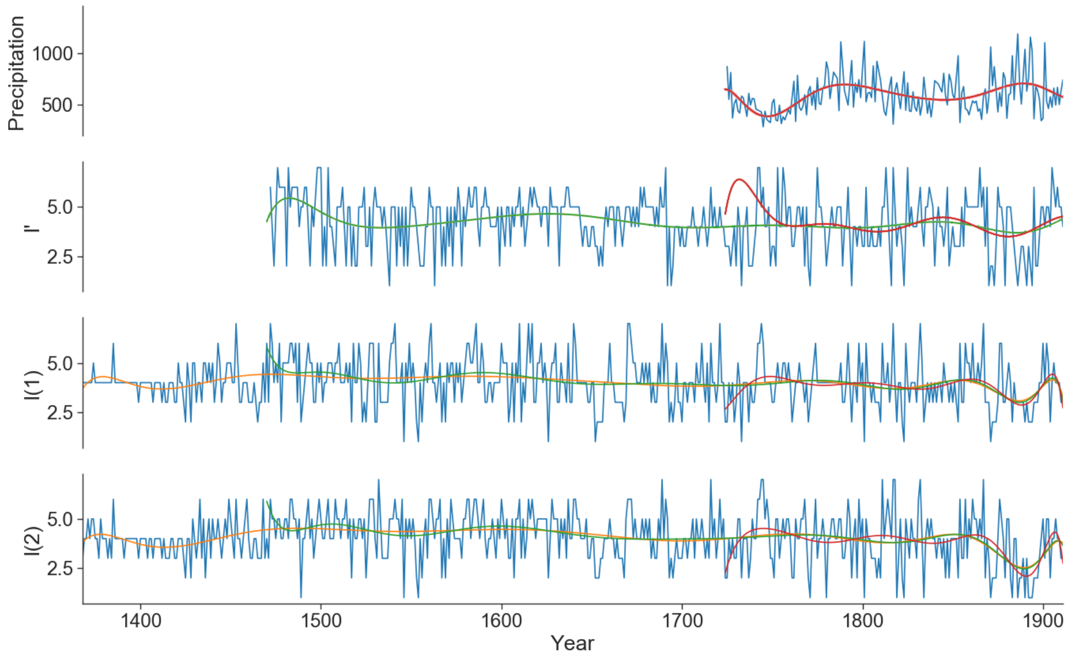


Figure 3.17: Plot of relevant indices in Shuntian Fu. I(1) is the grade generated from conservative source based on the number of drought/flood counties, and I(2) is the grade generated from the combined dataset based on the modified frequency of drought/flood records. I' is dryness/wetness grade constructed by Zheng and Precipitation is reconstructed precipitation in the Beijing. Red curves are polynomial fitting from 1724 to 1911. Green curves are polynomial fitting from 1470 to 1911. Orange curves are polynomial fitting from 1368 to 1911.

Thus, comparisons between these data would provide further evidence for the reliability of the approach designed in this paper. Figure 3.17 shows the plots of precipitation, dryness/wetness grade of Zheng's work, and dryness/wetness grades of this study. Grades generated from this study do not vary too much from previous work, especially from 1724 to 1911. The differences between grades from previous work and grades from the conservative source could arise for two reasons. One is that previous work on the dryness/wetness grade and precipitation reconstruction for the contained area of Beijing did not specify the exact spatial boundaries of the relevant historical data. Therefore, Shuntian Fu selected in this study may not match the areas with previous study one for one, but rather only partly overlapped with the Beijing area. Another possible explanation for the deviation in this study is that the dryness/wetness grade was not constructed in the entirely same way as the previous construction. In Zheng's work, dryness/wetness grade is recognised based on criteria as follow:

$$\begin{cases} PD_i \geq 200\%, & I = 7(Severe\,Dought) \\ 200\% > PD_i \geq 0, & I = 6(Dought) \\ 0 > PD_i \geq -90\%, & I = 5(Subdought) \\ PD_i < 90\% or PF_i < -90\%, & I = 4(Normal) \\ 0 > PF_i \geq -90\%, & I = 3(Subwet) \\ -200\% PF_i \geq 0, & I = 2(Wet) \\ PF_i \leq -200\%, & I = 1(Severe\,Wet) \end{cases} \tag{3.14}$$

where PD_i is the percentage deviation from the mean of the number of drought counties from 1471 to 1950 for each year in the Beijing area, and PF_i is the percentage deviation from the mean of the number of flood counties from 1471 to 1950 for each year in the Beijing area. Only records for summer and autumn were counted, as precipitation in summer and autumn would account for the majority of North China. Since there would be years with drought records and floods records simultaneously, dryness or wetness was determined according to the larger magnitude of percentage change from the mean. If the percentage deviation of the flood is greater, the grade would be confirmed as wet, and degraded by 1 level to include the effect of drought records in the same year. This could be the cause of some of the observed differences with respect to the previous study. What is more, the time periods are different. Thus, the different mean would imply varied normal traits for records in the Beijing area, where it has been noted that the precipitation for the Beijing has different time trends within various periods (Wei, 2007). Apart from the comparison with the previous dryness/wetness grade, comparison with reconstructed precipitation provides adequate validation for the grade in this study, if one considers the trend from polynomial fitting within 1724 to 1911 when the upgrade of wet level responded to precipitation increases. Therefore, the weighted frequency in this research would have a similar effect as the proxy of precipitation records compared with the approach of the number of disaster counties and should be able to map the real world at an acceptable level.

3.4 Conclusion and Limitation of this Research

This study compared a newly constructed data set with other data sets and actual records of precipitation and in general validated the reliability of the combined data set. The combined data set reveals similar spatial and temporal distributions as the data set from official history at the prefecture level or upper. The approach to construct this new data set was verified to be acceptable. Thus this study shows that supplementing existing data with alternative sources and the method used to parameterise such historical information can be a fruitful approach to constructing historical climatic event data.

However, the validations were all undertaken at the prefecture level, and thus there is still a possibility to violate the application of such an approach at the county level for the weighted frequency for each county every year. More specifically, due to the constraints of personal abilities, validation approaches in this study were not critically evaluated. As a result, it is necessary to improve the validation approach with the help of knowledge and skills from other fields (Tan et al., 2014). For example, previous studies have proposed a formula to assess the accuracy of historical documents (Yang, Wang, and Man, 2009), which can be further developed to merge data from different sources.

Chapter Four

The Impact of the Environment on Urbanisation During the late Qing Dynasty

4.1 Introduction

The origins of the Industrial Revolution have been a topic of interest in economic history for many years. Part of the discussion is related to the Great Divergence between western and eastern economies. The economies of China and the western world exhibited manifestly different growth after the start of 18th century - known as the great divergence. Countries in the west, such as England, experienced a sustained economic transformation from an agricultural society to an industrial society. However, countries in the east, including China, experienced economic stagnation and failed to develop competing industrial activities. As a result, a number of studies (Allen et al., 2011; Bernhofen et al., 2017; Li and Zanden, 2012; Pomeranz, 2001) have investigated various aspects of the global diffusion of the non-agricultural economy and why China failed to keep up the with growth of countries in the west.

Considerable attention has been given to failure of the late Qing emperor as a cause of China being left behind. In a recent paper, Ma and Rubin (2017) established a perfect refined sub-game equilibrium to prove that unlimited monarchical power reduces the ability of the central government to raise taxes. The conclusion is that the central tax system is closely aligned with social transformation. Another widely employed interpretation is from involution theory (Huang, 1990) which demonstrates that when China became overpopulated after the 17th century (Chen, 2016; Zhang, 2017), it imposed tremendous pressure on the agriculture sector so that more people chose to remain in the agricultural sector rather than find work in the cities and in the early industrial and proto-industrial sectors. Marks (2011) also suggests that from the mid-Qing dynasty investment tended to focus on the agricultural sector meaning there was no qualitative change in the economic structure of the country.

However, other academics argue that involution theory is not compatible with the overall picture of historical China. Several studies posit that the Yangtze Delta was comparable with England and the Netherlands from the scale of population and size (Allen, 2009; Allen et al., 2011; Baten et al., 2010), and had the same level of market integration as England and the Netherlands, at least in the early 18th century (Bernhofen et al., 2017; Bernhofen et al., 2015; Li and Zanden, 2012). For example, Li (1998) mentions that agriculture in the Jiangnan region was more effective for the development of relevant techniques and management skills. Therefore, involution may not be an appropriate description for historical China and the economic stagnation is not absolute (Li, 2001). For example, the number of towns and country fairs increased continuously from the Ming to Qing dynasties leading to growing urbanisation although this the trend slowed in the late Qing dynasty (Li, 2000). Overall, the transition (speed and timing) of historical China from agriculture to non-agriculture is still subject to considerable debate.

More recently there has been a small but growing literature that investigates the impact of environmental factors on China's historical development. It is well documented that environmental conditions deteriorated after the 14th century with the Little Ice Age occurring during that period (Ge, Zheng, Hao, Shao, et al., 2010; He, Li, and Liu, 2010). The frequency of natural disasters increased significantly between the Ming and Qing dynasties. Meanwhile, Marks (2011) claimed that land reclamation exacerbated the dangers from extreme environmental events. For example, the Yellow River severely burst its banks in the late Qing dynasty leading to widespread damage. The course of the Yellow River changed from Henan province to Shandong province in 1855. In concequence, there were at least 32 counties that reported zero crop gain in summer and autumn. There were three significant breaks of Yellow River from 1841 to 1843, causing more than one million deaths (Zhang, 2007). Part of this research agenda Chinese academics have compiled an extensive data set of historical events concerning the environment that has encouraged research in this area.

Motivated by the interaction between environmental shocks and the economic transition from agriculture to non-agriculture of China in history, this paper focuses on the late Qing dynasty and tests a number of theoretical predictions. The theory simplifies the diffusion of the non-agricultural economy as an issue of rural-urban migration and adds possible environmental impacts into the two-sector Harris-Todaro model (Harris and Todaro, 1970). The analysis suggests that an equilibrium in rural-urban migration exists when expected wages in rural and urban areas are equalised. According to the theoretical model, unfavourable environmental conditions can result in erosion of both the agriculture and manufacture sectors (Zhang, 2004a; Ashraf and Michalopoulos, 2015), which affects the expected wages of these two sectors.

More specifically, environmental shocks and abnormal weather events reduce the marginal product of labour and affect the relative prices of agricultural and manufactured goods. Hence, economic agents decide whether to work in the agriculture or non-agricultural sector based on expected wages. Land reclamation was also taken into account as it can mitigate the environmental impact on the agricultural sector to some extent assuming population remains constant.

To test the theoretical predictions, cross-sectional regressions have been used with the increase in market towns taken as a proxy for the increase in population in urban areas. Controlling for population density and territory area, this proxy represents the migration from rural to urban areas. This is useful as there is no accurate estimation of the urbanisation rate for historical China. What does exist from academics studying population history are estimations of province or regional urbanisation rates for some specific periods, such as the mid-Ming dynasty, the early Qing dynasty, and the late Qing dynasty (Li, 2000; Xu, Bas van Leeuwen, and Jan Luiten van Zanden, 2018; Cao, 2002b).

However, previous studies have tended to be at a quite aggregate level. This research focuses on the linkages between the variations in rural-urban migrations and environmental events in the late Qing dynasty from 1820 to 1911. The key variables are environmental factors that are identified as the mean of the annual modified frequency of environmental events from 1820 to 1911 where these frequencies are parametrised from the compilation of two historical documents. The documents contain four basic categories: flood, drought, cold, and wind. The geographical coverage is 17 provinces that cover the traditional agriculture regions (Pei, Zhang, and Lee, 2016). Where the population data exists it is based on the administrative boundaries in 1911.

In this paper we identify that different types of environmental events affect rural-urban migration separately. Since previous research on parametrisation for historical environment records typically considers floods and droughts, we also focus primarily on testing the the impact of these two events. We then include parameters for cold and wind. In addition, we are able to include interaction terms between different parameters and geographic features to investigate the role of possible mitigation effects that act to reduce the magnitude of any environmental shock.

Thus, in this paper we examine the relationship between environmental factors and urbanisation both theoretically and empirically. The aim is to provide additional inputs into the explanation for the Great Divergence. Compared with previous studies, we apply a rural-urban migration framework to the late Qing dynasty to establish a link between urbanisation and environmental factors.

Similar to the research on agriculture diffusion in the Neolithic Revolution (Ashraf and Michalopoulos, 2015), the framework we employ in this study allows us to investigate the determinants or mechanisms of social transition during this period, especially for the transition from an agricultural to a non-agricultural society. One of the main contributions is that we introduce a new proxy for rural-urban migration. To our best knowledge, the historic linkages between urbanisation and environmental impacts have seldom been tested quantitatively for any period prior to 1911 in China due in part to data issues. Perhaps more importantly, those previous studies that do exist are all at the prefecture level or above. Inspired by previous studies, we contribute to the growing historic environment information using county level data. This research also contributes to the market town research literature.

The structure of this paper afterwards is organised as follows. The first section reviews the literature on the development of town and county fairs in historical China and then introduces relevant analyses for rural-urban migration theoretically and empirically. The following section describes a theoretical framework that includes a simple two-sector theoretical model connecting the environmental shocks with rural-urban migration. The following section provides describes our data sources, outlines our identification strategy and presents the empirical results. The final section concludes and makes some suggestions on how the work can be taken forward.

4.2 Literature Review

The economic divergence between the West and East is usually studied from a comparative economic perspective so the advantages of western countries are compared to the disadvantages of eastern countries. For example, it has been suggested that China failed to avoid Malthusian positive checks on the population which resulted in a growing population that was absorbed by the agricultural sector which slowed improvements in human capital accumulation and innovation in more capital intensive agricultural techniques. Huang (1990) examined the role of the Malthusian check in a Chinese context and claimed that "involution" (which means the growth rate of output is below the growth rate of population when the total land areas were constrained in the pre-industrial agriculture society) within agricultural reproduction in traditional China is one of the main reasons for economic stagnation meaning that population growth did not lead to increases in productivity or more innovation. However, Li (2001) argues that that Huang's work is based on a number of misleading assumptions of traditional China, which would violate the "involution" narrative.

In contrast, research on the demography of historical China partially denied the positive check theory during Ming and Qing dynasty (Lee, Campbell, and Feng, 2002; Feng, Kugler, and Zak, 2002). A genealogy study (Shiue, 2017) based on historical documents illustrated that people from the Ming dynasty had an awareness of population control and the importance of investing in education. However, this awareness declined after the middle of the nineteenth century, implying that positive checks were a consequence rather than a determinant of economic growth.

Through an increasing use of historical documents, academics reconsidering the stagnation of historical China found a unique form of urbanisation in the Ming and Qing dynasties. In recent years, Chinese scholars have started to focus on the "market town" ("Shi Zhen"), which is thought to represent the bloom of urbanisation that started in the Ming dynasty. Initially, almost all studies related to the urbanisation of China were developed from the work of Skinner(Skinner, 1964) who established a rural-urban structure, which fits with China's urabn development after the fourteenth century. Unfortunately, Skinner's work cannot be reproduced as the data sources were not identified. Later research tried to estimate the rate of urbanisation from population records (Xu, Bas van Leeuwen, and Jan Luiten van Zanden, 2018; Cao, 2002a; Cao, 2002b). However, estimations of the urban population and total population are not compatible with more disaggregated geographical levels in China, such as prefecture or county level.

At the same time, after a closer examination of the historical record, academics realised that urbanisation in historical China performed in the manner consistent with the development of the market town. Therefore, the rise and fall of the market town has been the focus of recent work in this area. Using a case study of Shaobo town in the Yangzhou area, Liu (1987) claimed that disruption to canal transportation, warfare, and competition from foreign trade were the main reasons why the market town of Jiangnan region went downhill in late imperial China. The water network in historical China is commonly accepted to be one of the dominant factors that allowed market towns to flourish (Sun, 2017; Duan, 2013; Pan and Man, 2013).

The development of market towns was aided by simple and convenient transport via an extensive and growing water network. Merchants delivered goods along the rivers and canals to trade with local people. Although previous research has tended to focus on the dangers associated with maintaining the water network, especially floods, the channels by which the environment impacts urbanisation are now being studied (Liu, 2013). For example, a study of the northern part of Jiangnan region shows how rural peasants relocate when subject to famine and extreme disasters (Lin, 2011). In such cases, migrants tend to be absorbed into existing cities and then become part of the urban population.

The Ming dynasty also witnessed the birth of the Huizhou merchant who were renowned for the economic prowess (and formed a strong political force regionally and nationally). The rise of the Huizhou merchant class was due, in part, to the lack of farmland and poor environmental conditions (Wang, 1995). However, extreme disasters such as earthquakes or tsunamis were able to destroy whole cities and force people back into the countryside (Ren, 2003).

The challenge for researchers is that rural-urban migration triggered by environmental shocks are hard to capture due to data issues and the difficulty in isolating the impact from other factors happening at the same time. The majority of the previous studies have tended to concentrate on case studies which are hard to generalise across the whole of China. What is missing in the literature is a study of urbanisation in historical China regionally and with a higher resolution so that one can include other controls that may also impact urbanisation.

Theories explaining rural-urban migration has been around since the 1950s. The first framework was proposed by Lewis (1954) who argued that over population in the rural areas resulted in the excess population moving to urban areas. This approach was formalised as a two sectors model that became known as the Harris-Todaro model (Harris and Todaro, 1970) where rural-urban migration is explained by comparing expected wages in the agricultural and manufacturing sectors. More recently models and rural-urban migration have included environmental factors (Li and Zhou, 2015; Bahns, 2005) and estimated the impact of environmental issues empirically (Zhang and Song, 2003; Barrios, Bertinelli, and Strobl, 2006). Specifically, Barrios, Bertinelli, and Strobl (2006) employed a panel model to test the relationship between rainfall and urbanisation in Africa.

In other empirical research, Ashraf and Michalopoulos (2015) examine the relationship between temperature volatility and agricultural diffusion during the Neolithic Revolution and showed that an unpredictable climate delayed societies transition to an agricultural based economy.

Building on this literature this paper investigates whether environmental disasters has an impact on rural-urban migration in traditional agriculture regions during the late Qing dynasty.

4.3 Theoretical Framework

4.3.1 Concepts and Main Elements

The urbanisation process in historical China was different to many other countries at the time because of the importance of the market town. Following Skinner

(1964), during any specific period, different levels of urban unit have different carrying capacities due to their existing provision of services, transport, government funded police force and so on. When people in the countryside wanted to trade, they would move to the nearest market towns(Skinner, 1964). If there was no such settlement nearby, people would find a gathering place and form a market. Once the number of residents reached a certain level, people would ask the government for police/guardian forces or the government would send such forces (such an action was costly to the government). As a result these market places were "upgraded" to towns (Skinner, 1964; Cao, 2002b).

If a nearby settlement already existed, the pressure from population inflows would often lead to that settlement being upgraded to the next level of urban unit (Cheng, 2007). Thus, central or local government has the power to determine whether a town moved up (or down the scale) depending on the extent of rural-to-urban migration flows and and population density over long periods of time. We argue that one important factor in this decision is how a region experienced environmental shocks distinct from warfare shocks (Cheng, 2007). While there were a number of towns that did not have any direct governance from local government, the majority were business centres of multiple regions (Zou, 2013). Overall, it is possible to argue that the development of the market town represents patterns of urbanisation in historical China.

As Chinese historical studies have shown (Chu, 1973; Zhang, 2004b; Zhang, Zhan, et al., 2004), at the end of the Ming dynasty and at the beginning of the Qing dynasty, was a period known as the little ice age, when the temperature reached its lowest level for 500 years and environmental concerns reached their highest level.

Even considering relatively sparse data before the Ming dynasty, this conclusion holds (Fan, 2010; Zhang and Crowley, 1989; Cenler and Slale, 1995). Around this time, the centre of the China moved from north to south because of the growth in paddy rice in the south of China, especially in the Yangtze Delta (Pei, Lee, and Zhang, 2018). Paddy rice in the south of China had two to four times the output (Li, 2016b; Shi, 2015) of wheat in the north during the Song dynasty in the 12th century (Shi, 2015; Allen, 2009; Li, 2016c). The transfer of the economic centre and the main grain type led to a rapid expansion of the population (Deng, 2015) and resulted in greater dependency on water control systems due to the nature of paddy rice cultivation. The increase of grain outputs also helped with the distribution of labour and the formation of an integrated grain market over the whole country (Hu, 2017).

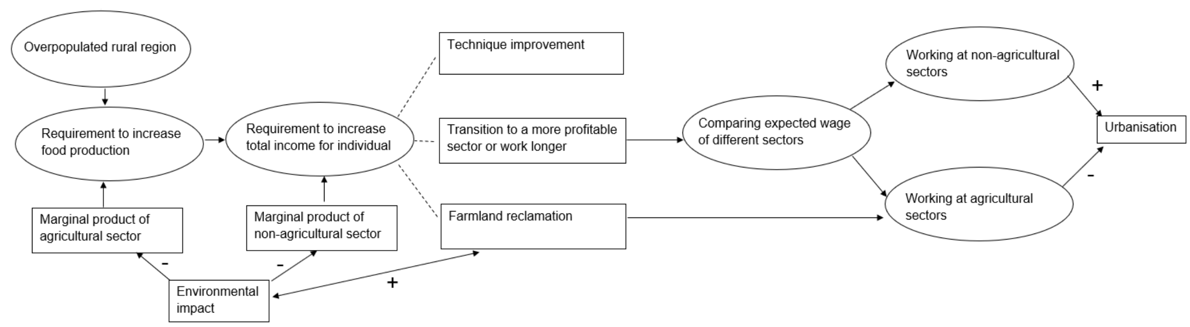


Figure 4.1: Schematic illustration of the proposed theory

Considering the context of historical China of the Ming and Qing dynasties, as illustrated in Figure 4.1, overpopulation and various environmental impacts drove the demand for increased income (Pomeranz, 2001). First, a larger population requires more food to be produced or purchased. Any environmental impact is likely to lead to a decline in agriculture production and manufacturing production. Therefore, individuals have an incentive to raise their general income to purchase essential goods. In addition, the supply of agricultural and manufacture production would affect the relative price of these two goods, reflecting the welfare or income of people from the two sectors.

In order to satisfy the demand for greater income, there are four ways that the agricultural sector can respond (1) technological improvements (advances in farming methods), (2) transition to a more profitable sector or (3) working longer, (4) and farmland reclamation (Deng, 2015). Specifically, improving farming methods would increase the production per mu; transferring to another profitable sector depends on the marginal product of the different sectors; farmland reclamation, such as building embankment around lakes to generate polder, would improve the agricultural sector's marginal product.

As for the non-agricultural sector, individuals have similar choices, (1) technological improvements and (2) transition to a more profitable sector or (3) work longer hours, and (4) farmland reclamation in rural regions (which means moving to the rural area). Therefore, to increase income, assuming technology stays constant, individuals need to choose between migration to a new sector and remaining to work longer or reclaiming farmland.

Individuals in both sectors are assumed to be price takers, which suggests that the relative price level should determine the equilibrium distribution of the urban to rural populations but should not be affected by any individual decision. Thus, the welfare impact of the environment will be through the effect of rural-urban migration on the two sectors and people's decision on how best they can increase income.

Typically, a natural disaster would have a negative impact on the marginal product of both agriculture and non-agriculture sectors, which would impact the relative prices of agricultural and non-agricultural goods. In general, production in the non-agriculture sector should be less dependent on environmental conditions (e.g. extreme weather damaging crops). Being less dependent on weather conditions should reduce future income uncertainty. For the agriculture sector, environmental shocks can reduce farmland productivity significantly. For example, floods may result in the erosion of farmland topsoil leading to reduced land fertility, that in turn reduces the marginal product of the land (Marks, 2011). Similarly, heavy rainstorms prevent peasants from working the land and can lead to crop failure (Gazetteer Museum of Zhejiang Province, 1900). In this regard, climate shocks are likely to encourage migration of people from rural areas to urban areas as the agriculture sector becomes less profitable after an environmental shock.

In addition climate shocks can destroy property directly, which triggers a greater demand for income. As a result, China has always been aware of the need to protect its people from the devastating effects of natural disasters (Wang, 2014; Zhang, 2017; Zhang, 2004a).

According to historical documents, natural disasters are classified into four categories: (1) water logging or flood, (2) drought, (3) wind, and (4) cold. To be specific, the components of water logging or flooding contain all of the over humidity meteorological events in which a flood is representative. Drought includes all the extremely high-temperature disasters. Wind includes all the events of big wind, typhoons, etc. Cold contains all the low-temperature damage in the meteorology categories. Intuitively, each environmental event could affect production differently. The historical records show strong evidence that natural disasters had highly negative outcomes. The record from "Qianlong Jishui County Annals" states:

"In April, great floods occurred in Jiangxi province from Qianda to the city's outer wall. Public and private workplaces and buildings were totally destroyed...... From April to May, floods occurred in Wanan. The city was entirely soaked for several days. Farmlands were collapsed, and buildings were submerged...... (in ancient Chinese writings)." (Huang and Mi, 1711)

Such destruction is also severe in the central part around the eastern Jiangnan region, as mentioned in the governor memorial:

"Recently, although the wheat in Shangxiang was ripe, heavy rain continued for seven days......Cultivated seedlings were swept, and uncut wheat became dirt...... (in ancient Chinese writings)." (Wang, 1988)

Drought and freezing temperatures, which are related to extreme temperature variation, are only serious in that they effect crop yields (Marks, 1998). According to the annual record, drought events usually coincide with locust swarms:

"In 1679....... Severe drought occurred in the autumn. locusts covered the sky, and even grass disappeared in the field (in ancient Chinese writings)." (Li and Wu, 1872).

As for freezing temperatures, in the historical record, there is usually a description of the freezing to death of crops such as:

"In 1892, freezing and snowing in the Gaoyou lead to the death of trees (in ancient Chinese writings)." (Hu and Lu, 2015).

High winds and typhoons are also associated with risk to welfare and human life and is hard to prevent. As described in the documents:

"In February 1654, strong wind destroyed houses and trees. Memorial arches in the city were entirely ruined" (Zhao and Li, 2008).

Accordingly, from the historical records, it can be confirmed that environmental shocks had a greater impact on the marginal product of the agricultural sector in historic China with manufacturing less impacted (Chen, 2016). Although there are several industries closely connected to agriculture (Feng and Yan, 2019) that will have been impacted by climatic conditions, other parts of the non-agricultural sector should be less impacted.

We turn now to farmland reclamation. From the literature, it is found that farmland reclamation is fairly straightforward in most cases (Sun, 2017). For example, there is some evidence to show that the unemployed migrated from the lower reaches of the Yangtze River to the upper reaches from the Ming dynasty to the Qing dynasty as a result of population pressure. A lot of the reclaimed lands were in the mountain areas and enabled the migrants to survive without coming into conflict with those already farming the less mountainous land. In addition, changes in environmental conditions leads individuals to choose what best increases total income and hence whether to migrate or reclaim farmland locally.

When considering non-agriculture areas, urban districts except existing cities, were initially formed as regular country fairs where people gathered to exchange essential goods. With the development of canal transportation, large numbers of market towns grew up alongside the water network (Liu, 2013; Pan and Man, 2013). Advanced goods were traded in the “upper” markets with the level of market being determined by the capacity of the population in the urban area (Cheng, 2007). Primary country fairs usually facilitated the trade of mulberry leaf, silk, crops and other essential goods (Feng and Yan, 2019).

Governments used fiscal policy to build and maintain the water control system to guarantee the flexibility of the transportation system. After the early Qing dynasty, the government invested in the water control system and the cost was paid for by the local people which meant they had to earn more to enable them to pay for the water control system (Chen, 2016). As a result, farmland reclamation was encouraged by the need to cover the cost burden of building and maintaining the water network.

Farmland reclamation was commonly done by converting existing lakes into polders surrounded by embankments and dikes. The lake area is then reduced. The reduction in the size of the lake reduces its ability to absorb water during heavy rains and increases the risk and frequency of future natural disasters, especially flooding, which in turn puts more pressure on water control system (Marks, 2011). The longer the process continues the greater the danger. For example, the canal system failed after a large number of dikes and dams broke in the north part of China (Zhang, 2004b) during the late Qing dynasty resulting in a number of towns and even cities being overwhelmed. The water network was therefore important urban development in historical China (Liu, 2013). A well maintained water network enhanced market integration of neighbouring regions (Duan, 2013; Yoshinobu, 1998; Pan and Man, 2013). Of interest is that a well maintained water network meant that urban regions were partially insulated from the impact of the flooding.

Another reminder of this framework concerns the water conservancy community (Sun, 2017), in which the local community undertakes the building and maintaining of dikes and dams. Thus, profits would be assigned by the usage of water (Sun, 2017). To ensure that maintenance happened, the water conservancy community developed a precise management system (Sun, 2017). To encourage maintenance work, taxation and compulsory services could be mitigated or even avoided. The formation of a conservancy community inspired a high level of agricultural production stability since such an organisation would gather people to fight against natural disasters and to help each other within the community.

People who participated in local water conservancy communities would have less incentive to leave since they would lose the benefits from reduced taxation and avoiding compulsory services. This discouraged migration to urban areas as it was almost impossible to re-join the community after leaving (Sun, 2017). Water conservancy community explains why farmland reclamation was the main choice when it came to attempting to increase income. Agricultural income would be relatively less uncertain in those communities but would set barriers for people to move. For example, when individuals in rural areas want to raise income, their prior choice might be farmland reclamation. However, when urban areas want to raise income, they might find other ways in urban regions even if the marginal production was relatively low. If individuals in rural regions decided to move to urban areas, they might fear future unemployment. In this regard, reclamation might only represent the natural growth of the agricultural population rather than the migration from urban regions to rural regions.

In addition to previous elements that may affect people's preferences, a series of measures aiming at stable agricultural production was also widely recognised. A general narrative given by historians was that officials would level off the fluctuation of agricultural production through grain granaries at different levels, although the effectiveness was commonly doubted (Perdue, 1982). Measure to establish grain granaries is a long-lasted tradition in historical Chinese politics, which has been claimed to exist around two millennia (Wong et al., 1991). There were three types of the granary developed till the Qing dynasty, including the ever-normal (changpingcang), charity (yicang), and community (shecang), of which state officials managed the former two for price stabilisation and relief releasing. The source of the grain for these state-owned granaries was mainly from taxation, which required peasants to hand over the grain as their taxes. Officials were responsible for selling the grain from granaries when the prices were too high and buying in the grain to supplement granaries when the prices were too low. Therefore, it is reasonable to expect more stable grain prices during the Ming and Qing dynasties.

There are several alternative concerns needing acknowledgement prior to further analyses. One primary is that the exogeneity of environmental factors, especially for the flood, is challenged. Previous studies have mentioned that population-induced land reclamation in highlands and upper streams, and excessive building of waterworks, broke the ecological environment for agriculture and increased the risk of floods with respect to higher frequencies and severer consequences (Perdue, 1982; Bernhofen et al., 2018; Muscolino, 2016; Osborne, 1994). Specifically, Polder reclamation reduced the water storage capacity and flood control of lakes, leading to higher costs on polder maintenance and more potential severe damage from the flood. Then, repeated polder reclamation weakened the flood resistance of the embankment and raised the riskiness of flood by raising river bed (Pomeranz, 2001; Marks, 2011). More specifically, Perdue (1982) provided two cases from the Dongting Lake area to show how officials failed to control illegal dike buildings and land reclamation by powerful local lineages, which related to severer floods. Therefore, it is possible that floods were generated endogenously rather than exogenous triggers. Another one is the effectiveness of granaries, as mentioned before. Many arguments discussed the possibilities of weak management, poor funding and so on, which violated the initial goal.

However, there are still some spaces left for us to argue for the concerns mentioned above. For the possible endogeneity of floods, we decided to follow assumptions from previous research where floods and droughts were considered mostly exogenous to fulfil empirical estimates (Jia, 2014). What is more, the causal connections between human activities and floods were usually assumed on a large spatio-temporal scale (across provinces and over decades), which is difficult to assign the contribution to individual incidents. Therefore, incidents at lower levels and shorter periods (yearly for example) can still be viewed as random events. It is also acceptable that floods were still random triggers even if the excessive land reclamation and dike buildings exacerbated the destruction of floods. Additionally, floods records were not clear enough to isolate every human-induced incident, and there was inadequate direct evidence to support the potentially inverse causal direction. As a result, we consider that all our environmental factors have remained satisfactory exogeneity on average. As for the ineffectiveness of granaries, we expect more stable grain prices at least than the cases when there was no granary at all. Other concerns, including the impacts from warfare, conflicts between state and locals, and transportation, are less straightforward in our stories and beyond the range of discussions.

4.3.2 Simple Model of Two Sectors Migration

Building on the standard Harris-Todaro two-sector model, Wei and Yabuuchi (2003) present a prefecture-level model that is similar to a small economy consisting of urban manufacturing and rural agriculture sectors. Market integration is high enough to be assumed as a competitive equilibrium. Surplus from the agricultural sector and demand for non-agricultural goods (Brueckner and Fansler, 1983; Liu, 1987) are exchanged through regular markets that act as trading platforms for agriculture and non-agricultural goods (Cai, 2012). Such markets usually exist in an urban region surrounded by several rural regions. Urban areas, including markets, mainly produce non-agricultural goods, and rural areas mainly produce agricultural goods. Thus, there is a relative price of agricultural goods and non-agricultural goods and people in both sectors are price takers. Households determine the labour allocation between urban and rural areas by comparing the expected real wage.

The expected real wages of urban and rural areas depends on the marginal product of labour (Harris and Todaro, 1970). Capital and labour cannot be substituted for land efficiently in the agriculture sector in traditional China (Ho, 1959; Pomeranz, 2001), which implies there is an important role of farmland in the agriculture (Chao and Yu, 1993).

Environmental shocks can reduce production that would also be affected by the stock of current capital (investment in infrastructure etc.), which may reduce the risks of an environmental crises (e.g. flood defences). In this research, to simplify the analysis, the accumulation of technology is related to the level of capital stock but it is assumed to be constant since there is little evidence of a broad adoption of new technology. Therefore, the mitigation effect of capital on environmental damage on productivity remains unchanged.

Equilibrium exists when the expected wage in the agriculture sector equals the expected wage in the manufacturing sector. Like other research using the Harris-Todaro framework, we assume full employment in rural areas initially. However, we also assume that the lowest wage in the urban area is not constant since there is no such policy to constrain the lowest wage in the urban area, and the urban region is modelled more carefully, including regular markets without government authority as a town or city. Therefore, unemployment and manufacturing productivity is assumed to be mainly affected by the environment, especially when the total population is assumed to be fixed.

Suppose there is no negative effect from a climatic shock on productivity and no relationship between unemployment and environmental events. In this case, the production function of the agriculture sector would be:

$$X_A = q(N_A, \bar{L}, \bar{K}_A), q' > 0, q'' < 0 \tag{4.1}$$

where X_A is the output of the agriculture sector, N_A is the agricultural labour input, $\bar{L}$ is the stock of farmland which is supposed to be constant in the short term, and $\bar{K}_A$ is the stock of capital in the agriculture sector, which is also considered to be constant. q' is the deviation of agriculture production with respect to the agricultural labour, which represents the marginal product.

The production function for the non-agriculture sector is:

$$X_N = f(N_N, \bar{K}_N), f' > 0, f'' < 0 \tag{4.2}$$

where X_N is the output of the non-agriculture sector, N_N is the non-agricultural labour input, and $\bar{K}_N$ is the stock of capital in the non-agricultural sector which is considered to be constant. Land is not included as a factor in the non-agricultural sector.

Initially, if the environmental shock only imposed damage on the agriculture sector and has no relationship with unemployment, the equilibrium price in terms of non-agricultural goods obtained from the competitive market is given by:

$$P = \rho(\frac{X_N}{e^{-\eta}X_A}) = \rho(e^{\eta}\frac{X_N}{X_A}), \eta > 0, \rho' > 0, \rho_{N_N} > 0, \rho_{N_A} < 0 \tag{4.3}$$

where the relative price is determined by the output of agricultural and non-agricultural goods. $\rho_{N_N} > 0$ denotes that a larger scale of non-agricultural output would lead to the relative scarcity of agricultural goods. Thus, the price of the agricultural good would increase. In contrast, $\rho_{N_A} > 0$ denotes that more agriculture output would lead to the relative scarcity of non-agricultural goods. e is the index of the environmental impact and $e \geq 1$. Thus, a higher value of e the higher the frequency of environmental disasters where there are no disasters when $e = 1$. If the mitigation effects of land and capital are excluded, the magnitude of the damage is determined only by the severity of the environmental event. Thus, $0 < e^{-\eta} \leq 1$ represents the impact of an environmental shock on productivity. As a result, if $e > 1$, agricultural production decreases, which means there is a relative scarcity of agricultural goods leading to an increase in the price of agricultural goods, which means $\frac{\partial \rho}{\partial e} > 0$.

Since households are price takers, the real wage of the agriculture sector is:

$$W_A = e^{-\eta} P q' \tag{4.4}$$

and real wage of non-agriculture sector

$$W_N = f' \tag{4.5}$$

Assume that the agriculture sector is at full employment, but unemployment exists in the non-agriculture sector. Assume the unemployment rate is u. Therefore, the total population of the urban region should be $N_n = N_N/(1-u)$, and the expected wage of the non-agricultural sector is $E(W_N) = (1-u)W_N = (1-u)f'$. Hence, an equilibrium exists when the expected wages of the agriculture sector and non-agriculture sector are equalised since there would be migration between the sectors if there are differences in expected wages. The equilibrium condition is given by:

$$E(W_A) = E(W_N) \Rightarrow e^{-\eta} P q' = (1-u) f' \tag{4.6}$$

Over a long period of time (1820 to 1910) it is reasonable to assume the market reaches equilibrium. Hence, in equilibrium, rural-to-urban migration will be affected by the magnitude of long-term environmental damage. Hence, the condition to determine the level if migration is given by:

$$\Delta N_n = \psi[(1-u)f' - e^{-\eta} P q'] = \psi(\frac{N_N f'}{N_n} - e^{-\eta} P q'), \psi' > 0, \psi(0) = 0 \tag{4.7}$$

where ΔN_n is the time derivative and represents migration from the rural to urban region. Migration will stop when the difference in expected wages between urban and rural wages is zero. If the expected wage in the rural region is greater than that of the urban region, people in the urban area will return to the rural area in pursuit of a higher income. Subsequently, more labour in the agriculture sector would decrease the scarcity of agricultural goods and hence the marginal product. The loss of the urban population would increase the marginal product of labour in the non-agriculture sector. Then, P and q' would fall, but f' would increase. Therefore, environmental degradation would change the expected wage of the agriculture sector until a new equilibrium is reached. The reduction in the urban population will then shift the economy from the previous equilibrium to a new equilibrium. Specifically, in equation (4.7), rural-urban migration shifts the economy from the equilibrium without environmental degradation to the equilibrium with environmental degradation. Hence, ΔN_n can be thought of as expected migration.

In order to further recognise the migration response to environmental degradation, we assume that variation in urban and rural populations are not changed by environmental events directly (or endogenously) especially in the long-term. In other words, $\frac{\partial N_N}{\partial e} = \frac{\partial N_u}{\partial e} = \frac{\partial N_A}{\partial e} = 0$ in equation (4.7). As a result, the relationship between rural-urban migration and the environmental degradation can be derived as follows:

$$\begin{aligned} \frac{\partial \Delta N_n}{\partial e} &= \frac{\partial \psi(\frac{N_N f'}{N_n} - e^{-\eta} P q')}{\partial e} \\ &= \psi' \frac{\partial (\frac{N_N f'}{N_n} - e^{-\eta} P q')}{\partial e} \\ &= \psi'(\eta e^{-\eta-1} P q' - e^{-\eta} q' \frac{\partial \rho(e^{\eta} \frac{X_N}{X_A})}{\partial e}) \end{aligned} \tag{4.8}$$

As indicated in equation (4.8), the impact of environmental degradation on the expected wage in the agricultural sector is not certain, since the degradation would reduce the marginal product (which is represented by the part $\psi' \eta e^{-\eta-1} P q'$) but would increase the relative price of agricultural goods in terms of non-agricultural goods (which is represented by the part $\psi' e^{-\eta} q' \frac{\partial \rho(e^{\eta} \frac{X_N}{X_A})}{\partial e}$). The sign of $\frac{\partial \Delta N_N}{\partial e}$ is determined by the magnitude of $\frac{\partial \rho(e^{\eta} \frac{X_N}{X_A})}{\partial e}$. Hence, whether environmental degradation increases or reduces the expected wage of the rural region depends on the sensitivity of price in response to the scarcity of agricultural goods. If the price function is homogeneous of degree one as expressed by:

$$P = \rho(e^{\eta}\frac{X_N}{X_A}) = e^{\eta}\rho(\frac{X_N}{X_A}), \rho' > 0 \tag{4.9}$$

thence the equation (4.8) can be derived as

$$\begin{aligned}\psi'(\eta e^{-\eta-1}Pq' - e^{-\eta}q'\frac{\partial\rho(e^{\eta}\frac{X_N}{X_A})}{\partial e}) &= \psi'(\eta e^{-1}\rho(\frac{X_N}{X_A})q' - e^{-\eta}q'(e^{\eta})'\rho(\frac{X_N}{X_A})) \\ &= \eta\rho\psi' q'(e^{-1} - e^{-1})\end{aligned} \tag{4.10}$$

where $\rho = \rho(\frac{X_N}{X_A})$ represents the relative price without any environmental degradation. Equation (4.10) indicates that if the price function is homogeneous of degree one, the impact on the marginal products and the relative price can be totally offset by each other. Therefore, environmental degradation would have no impact on rural-urban migration.

Furthermore, if the degree of homogeneity is not one, the environmental degradation would alter the direction of migration between rural and urban areas. Suppose the degree of homogeneity is ε, where:

$$P = \rho(e^{\eta}\frac{X_N}{X_A}) = e^{\varepsilon\eta}\rho(\frac{X_N}{X_A}), \rho' > 0 \tag{4.11}$$

thus, ε can be defined as the price sensitivity to the relative supply (or relative scarcity of agriculture goods). Hence, in this case, equation (4.8) can be derived as:

$$\psi'(\eta e^{-\eta-1}Pq' - e^{-\eta}q'\frac{\partial\rho(e^{\eta}\frac{X_N}{X_A})}{\partial e}) = \psi'(\eta e^{(\varepsilon-1)\eta-1}\rho(\frac{X_N}{X_A})q' - e^{-\eta}q'(e^{\varepsilon\eta})'\rho(\frac{X_N}{X_A}))$$
$$= (1-\varepsilon)\eta\rho\psi' q' e^{(\varepsilon-1)\eta-1} \tag{4.12}$$

If $\varepsilon > 1$, which represents a higher sensitivity of the relative price to the relative supply, $\frac{\partial \Delta N_n}{\partial e} < 0$, while if $0 \leq \varepsilon < 1$, $\frac{\partial \Delta N_n}{\partial e} > 0$.

In this context of this study, there were several measures put in place to ensure price stability in the agricultural sector including aid relief (Chen, 1986) and grain storage (Liang, 1980). It is rational to assume that price is not sensitive enough to offset the income reduction caused by the decrease in marginal product. However, it is also possible that this assumption is violated, especially when environmental disasters destroyed the water network which would also impact market integration which was shown to fall during the late Qing dynasty (Bernhofen et al., 2017; Bernhofen et al., 2015).

Next, if we drop the previous constraints by assuming that environmental degradation would also impact the marginal product of the non-agriculture sector, the degradation across sectors could be different. Intuitively, degradation in the non-agriculture sector would be less significant, as only part of non-agricultural goods would be affected compared with the majority of handicrafts, such as weaving (Li, 2000). Thus, any migration decision is determined according to the expression:

$$\Delta N_n = \psi(e^{-\xi}\frac{N_N f'}{N_n} - e^{-\eta}Pq'), \psi' > 0, \psi(0) = 0 \tag{4.13}$$

where $0 < \xi < \eta$, $e^{-\xi}$ is the environmental degration on non-agriculture products and

$$P = \rho(\frac{e^{-\xi}X_N}{e^{-\eta}X_A}) = \rho(e^{\eta-\xi}\frac{X_N}{X_A}), \rho' > 0, \rho_{N_N} > 0, \rho_{N_A} < 0 \tag{4.14}$$

Similarly, if the price function is homogeneous of degree one, equation (4.13) can be written as

$$\Delta N_n = \psi(e^{-\xi}\frac{N_N f'}{N_n} - e^{-\xi}\rho q'), \psi' > 0, \psi(0) = 0 \tag{4.15}$$

where $\rho = \rho(\frac{X_N}{X_A})$. Since $\frac{N_N f'}{N_n} - \rho q' = 0$ if there is no environmental impact, $\frac{\partial \Delta N_n}{\partial e} = 0$, which indicates that the price sensitivity to relative scarcity of agriculture goods always determines the migration directions in the same way. Lower price sensitivity ($0 < \varepsilon < 1$) encourages rural to urban migration, while higher price sensitivity ($\varepsilon > 1$) encourages migration in the other direction. Specifically, the expression of derivative in terms of ε is given by:

$$\begin{aligned}
\frac{\partial \Delta N_N}{\partial e} &= \frac{\partial \psi(\frac{e^{-\xi}N_N f'}{N_n} - e^{-\eta}Pq')}{\partial e} \\
&= \psi' \frac{\partial(\frac{e^{-\xi}N_N f'}{N_n} - e^{-\eta}Pq')}{\partial e} \\
&= \psi'(\eta e^{-\eta-1}Pq' - e^{-\eta}q' \frac{\partial \rho(e^{\eta-\xi}\frac{X_N}{X_A})}{\partial e} - \xi \frac{e^{-\xi-1}N_N f'}{N_n}) \\
&= (\eta - \varepsilon\eta + \varepsilon\xi)\rho\psi' q' e^{(\varepsilon-1)\eta-\varepsilon\xi-1} - \xi\psi' \frac{e^{-\xi-1}N_N f'}{N_n}
\end{aligned} \tag{4.16}$$

where $\eta e^{-\eta-1}Pq'$ denotes the shift from the environmental erosion on marginal products of agriculture, $e^{-\eta}q'\frac{\partial\rho(e^{\eta-\xi}\frac{X_N}{X_A})}{\partial e}$ denotes the shift from the environmental degradation on relative prices and $\xi\frac{e^{-\xi-1}N_N f'}{N_n}$ denotes the shift from the impact of environmental degradation on the marginal product of manufacturing (the non-agriculture sector). Remember that if the price sensitivity to relative scarcity was one ($\varepsilon = 1$), an environmental shock should have zero impact on rural-urban migration.

To clarify, when the price is relatively stable, the benefits from a higher relative price would be lower. On the one hand, a reduction in total non-agriculture production would weaken the negative impact of increasing agricultural goods scarcity. However, the reduction in the marginal non-agriculture products can offset this weakness. On the other hand, a lower price sensitivity indicates that the loss from the marginal products of agriculture may not be offset by the income gain from the increase in the relative price. On the contrary, when the price fluctuates, environmental degradation would still potentially discourage rural-to-urban migration.

Alternatively, if we assume that the environment also affects unemployment in the non-agriculture sector, an environmental impact $e^{-\delta}$ can be imposed on the employee population. The population is considered to be exogenous. Hence, migration is given by:

$$\Delta N_n = \psi(e^{-\xi}f'\frac{e^{-\delta}N_N}{N_n} - e^{-\eta}Pq'), \psi' > 0, \psi(0) = 0 \tag{4.17}$$

Thus, the deviation of rural-to-urban migration in terms of ε can be expressed as:

$$\begin{aligned}
\frac{\partial \Delta N_n}{\partial e} &= \frac{\partial \psi(\frac{e^{-\xi-\delta}N_N f'}{N_n} - e^{-\eta}Pq')}{\partial e} \\
&= \psi' \frac{\partial(\frac{e^{-\xi-\delta}N_N f'}{N_n} - e^{-\eta}Pq')}{\partial e} \\
&= \psi'[\eta e^{-\eta-1}Pq' - e^{-\eta}q' \frac{\partial \rho(e^{\eta-\xi}\frac{X_N}{X_A})}{\partial e} - (\xi+\delta)\frac{e^{-\xi-\delta-1}N_N f'}{N_n}] \\
&= (\eta - \varepsilon\eta + \varepsilon\xi)\rho\psi' q' e^{(\varepsilon-1)\eta-\varepsilon\xi-1} - (\xi+\delta)\psi' \frac{e^{-\xi-\delta-1}N_N f'}{N_n}
\end{aligned} \tag{4.18}$$

If we assume that $\varepsilon = 1$, equation (4.18) can be written as $\psi'[\xi e^{-\xi-1}\rho q' - (\xi + \delta)e^{-\xi-\delta-1}f'\frac{N_N}{N_n}]$ and $\Delta N_n < 0$ for all $e > 1$. Therefore, if $\xi e^{-\xi-1} = (\xi+\delta)e^{-\xi-\delta-1}$, $\frac{\partial \Delta N_n}{\partial e} = 0$. Hence, the condition for an environment shock to approach its maximum effect is $e^{-\delta} = \frac{\xi}{\xi+\delta} \Rightarrow e = (1+\frac{\delta}{\xi})^{\frac{1}{\delta}}$, which indicates that the magnitude of e would affect the relationship between the migration and the environmental shock. As a result, an unfavourable environment would potentially encourage migration from the urban area to the rural area but such effect would be reduced when the environment index e is over $(1+\frac{\delta}{\xi})^{\frac{1}{\delta}}$, which indicates a non-linear relationship. Although the non-linear prediction depends on the construction of the environmental index, it is intuitive that extreme disasters would destroy both the agriculture and non-agricultural sectors and then there would also be no migration. In addition, if $0 < \varepsilon < 1$ (or the price is stable), the maximum point for the shock to have an effect would be through a shift leftwards and even do change the sign of ΔN_n.

However, the previous analysis does not consider the mitigation effect of other factors since land and capital are regarded as fixed. Generally, when facing with the reduction of expected wage, people would decide on the distribution of labour force on the agriculture sector and non-agriculture sector. In contrast, except for the migration, people would also choose to invest more in capital and land reclamation. Specifically, land reclamation would significantly increase the marginal product in the agriculture sector to mitigate the income reduction resulting from environmental erosion. Suppose individuals of the agriculture sector and non-agriculture sector are price taker, and the price is constant. The expected wage of the agriculture sector is:

$$E(W_A) = e^{-\eta} P q' \tag{4.19}$$

$$\Rightarrow \frac{\partial E(W_A)}{\partial L} = e^{-\eta} P \frac{\partial q'}{\partial L} + e^{-\eta} q' \frac{\partial \rho(e^{\eta} \frac{X_N}{X_A})}{\partial (e^{\eta} \frac{X_N}{X_A})} \frac{\partial (e^{\eta} \frac{X_N}{X_A})}{\partial L} \tag{4.20}$$

where $\frac{\partial q'}{\partial L} > 0, \frac{\partial \rho(e^{\eta} \frac{X_N}{X_A})}{\partial (e^{\eta} \frac{X_N}{X_A})} > 0$ but $\frac{\partial (e^{\eta} \frac{X_N}{X_A})}{\partial L} < 0$. Thus, the sign of $\frac{\partial E(W_A)}{\partial L}$ is not certain.

Hence, land reclamation would increase the marginal product of labour of the agricultural area, which would possibly prevent migration from the rural region to the urban region. However, if land reclamation impacts prices, the direction of any migration is ambiguous since higher agricultural production would reduce the relative price of agriculture goods. In this regard, the mitigation effect of the land reclamation might vary by reclamation level. Marginal changes in the land used for production and prices matter. If the marginal impact of a change in land use on price is relatively small, land reclamation may potentially increase the expected wage of the agriculture sector. However, if the marginal deviation in price is relatively large, the expected wage may be reduced. As noted by (Marks, 2011), land is the factor that cannot be substituted efficiently in historical China (the Qing dynasty in this context) for the agriculture sector, which implies that the increase in land can be considered the main source of any production increase. Therefore, it is suggested that if the change in land used for production or the marginal product of labour of the agriculture sector is relatively large, then the expected wage will be increased by land reclamation and would impact rural-urban migration.

Overall, it is expected that an environmental shock will reduce the marginal product in both sectors, but the impact will be greater on the agriculture sector. It is not easy to evaluate the environmental impact on prices. Therefore, the environmental impact on the expected wage of the agriculture sector is hard to specify. Typically, the price varies slightly with the relative output ratio of non-agricultural goods to agricultural goods. According to the analysis above, one needs to be careful when concluding that the effect of an environmental shock is determined by the sensitivity of the relative price to the relative scarcity of agriculture goods ε. A stable price ($0 < \varepsilon < 1$) would encourage rural-to-urban migration when the environment gets worse. Nevertheless, urban-to-rural migration would be encouraged if price sensitivity is high ($\varepsilon > 1$). What is more, migration is not the only way to deal with the reduction of the expected wage of agricultural goods. Land reclamation is widely employed in the Ming and Qing dynasties to mitigate the increase in income demand.

Additionally, different environmental events would probably have different impacts on the marginal products of the agriculture and non-agriculture sectors. Therefore, it is possible that $\eta = \sum_{i=1}^{n} \eta_i$, $\xi = \sum_{i=1}^{n} \xi_i$ and $\delta = \sum_{i=1}^{n} \delta_i$ where n is the number of categories of environmental damage. What is more, previous assumptions simplify the situation that all kinds of environmental damage share the same price sensitivity. However, it is possible that different disasters have a different impact on ε. Hence, estimations for a single environmental factor would probably be biased due to different local conditions.

4.4 Methodology

4.4.1 Data Sources

This research tests the theoretical framework outlined in Section 3.3 concentrating on the the traditional agriculture regions of China which includes 17 provinces where population data is available over a long time period (Gansu province was excluded due to poor data reliability). The regions chosen are based on the administrative boundaries of 1911 as shown in Figure 4.2. In total there are 1,634 counties. Relevant data are obtained from the "China Historical Geographic Information System" (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). The CHGIS is a free database covering Chinese dynasties and includes place names, historical administrative units, rivers and lakes, and other geographic related data. It was established under the leadership of Fudan University and Harvard University. All the maps in this research are generated through the software ArcGIS version 10.6.

Since the original data in the CHGIS contain some self-intersecting polygons that cannot be easily combined with the data for other variables, the administrative

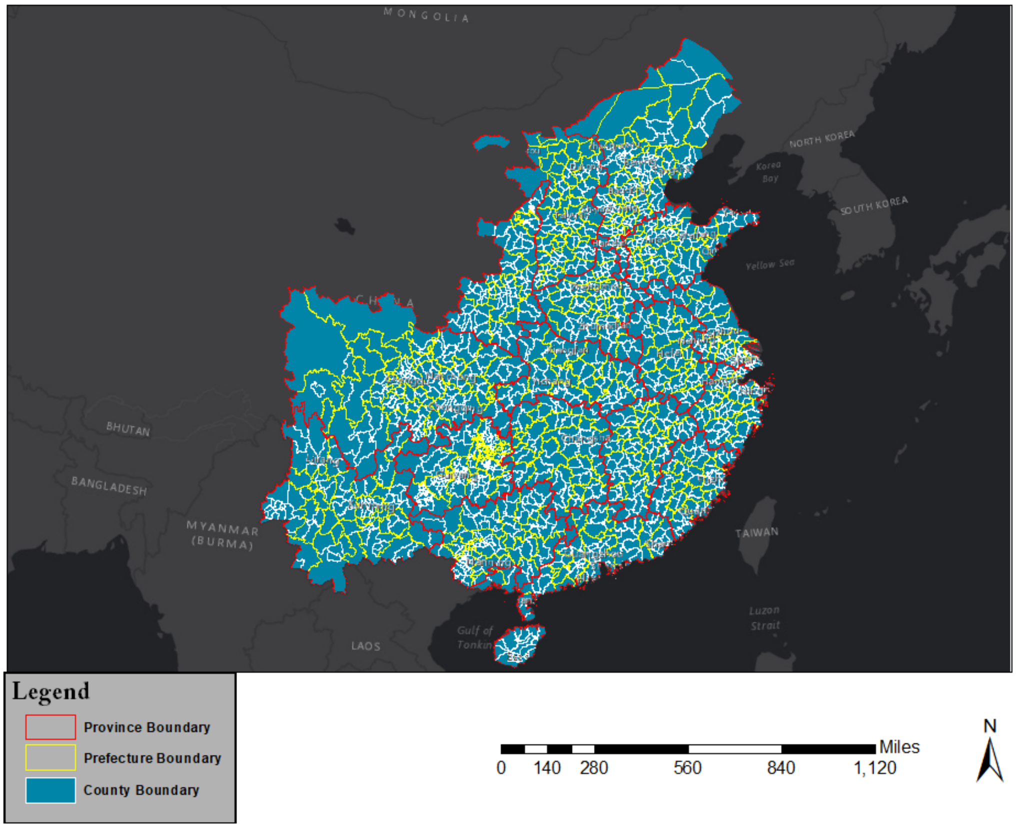


Figure 4.2: The Map of Administrative Boundaries of the Research Regions

Source: China Historical Geographic Information System, Version 4. Cambridge: Harvard Yenching Institute and Fudan Center for Historical Geography, January 2007

boundaries were re-organised into simple polygons at the prefecture and county levels. Hence, each simple polygon is regarded as an individual geographic unit. The area of each geographic unit was calculated directly in ArcGIS based on the Xian'80 geodetic coordinates systems. In addition, data on rivers are obtained from CHGIS as illustrated in Figure 4.3. The length of the river is also calculated in ArcGIS based on the Xian'80 geodetic coordinates system. We mainly focus on

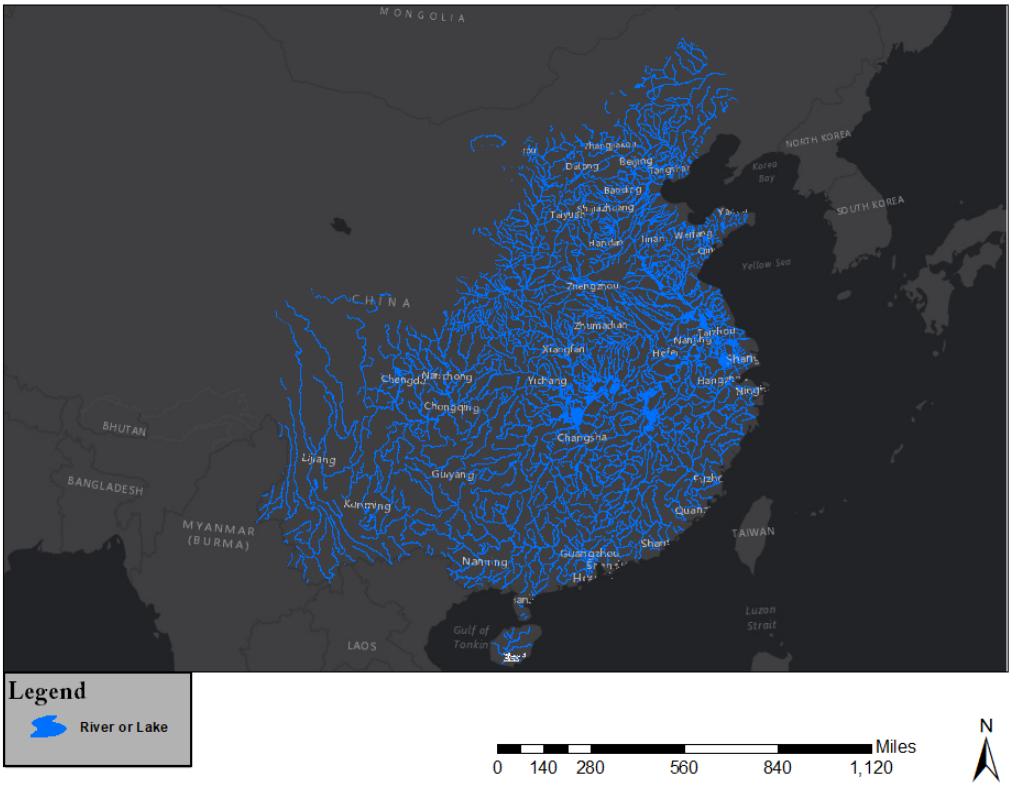


Figure 4.3: The Map of Rivers in Research Region

Source: China Historical Geographic Information System, Version 6. Cambridge: Harvard Yenching Institute and Fudan Center for Historical Geography, January 2016

the late Qing dynasty from 1820 to 1911.

Our indicator of urbanisation is the increase in the number of towns of each geographic unit. There are data for the town's location in 1820 and 1911 which we obtain from the CHGIS. The number of the towns was calculated for each geographic unit, and then the difference between the number in 1911 and 1820 was calculated. This parameter is our measure of the degree of rural-to-urban migration. The advantages of this parameter is that it was obtained from a reliable source, and the spatial resolution is at the county level which is much more detailed that any previous research into urbanisation rates in historic China.

The environmental data were generated from two compilations of historical records. Thanks to the rich historical records, most meteorological records exist in local gazetteers, poems, letters, books, official documents, lyrics and personal dairies. Previous work has collected those records from multiple sources (Pan, 2012). The first source we use is “China’s three-thousand-year Meteorological Records (in Chinese)” (Zhang, 2004b), which is complied from 8,228 documents, including about 220,000 records over thousands of years. This source is often considered to be the most complete and reliable collection of official histories (Zhang, 1998). The records collected in this book are verified and include detailed references.

Another dataset we use is from the collection of meteorological disasters called “China meteorological disaster dictionary (in Chinese)” (Wen, 2006; Zeng, 2006). This collection includes torrential rains, droughts, cold weather, humid weather, freezing damage, strong winds and hail, lightning strikes, thick fog, geological disasters caused by rainfall, forest fires and so on (Wen, 2006). The data we use comes from different volumes for difference provinces to match our regional sample over the same time period. This collection of data was complied more recently and contains an official history and records from other informal documents. This source has more environmental information that we use to supplement official records at the county level.

This study adopts an approach of weighted frequency to combine the two sources of data described above. In order to facilitate the classification, disasters were categorised into flood, drought, cold, and wind. Data were constructed annually at the county level (so we can aggregate later if needed). From 1820 to 1910, the mean annual frequency was calculated to represent the average impact of natural disasters across the whole period. Alternatively, the means for the periods 1729 to 1819, 1638 to 1728 and 1368 to 1819 were also calculated to capture the impact of past disasters.

We use the following categories. (1) Flood which contains those records from the categories of torrential rains, thick fog, flood, and humid weather to denote the degree of wetness and any description related to severe water logging. (2) Drought includes the records of droughts, high-temperature damage, and any description that relates to a shortage of water. These two categories are the ones most often used in previous studies of the environmental history of China (Chen, 1986; Xia, 2010; Li, 2007). Thus, data from historical documents for floods and droughts have been employed in numbers of studies and have been shown to be compatible with historical research during this period of Chinese history. For example, figure 4.4 presents the county-level distribution of floods. As can be seen, the east of China experienced more floods according to the historical records. Most flooding occurs around the main streams and rivers.

The third category we include is (3) Cold that contains damage from freezing temperatures, hail and cold weather, and any records containing any description of cold related phenomena. Finally, (4) Wind includes strong winds, hurricanes/typhoons, and records containing any description of wind. These two categories have not

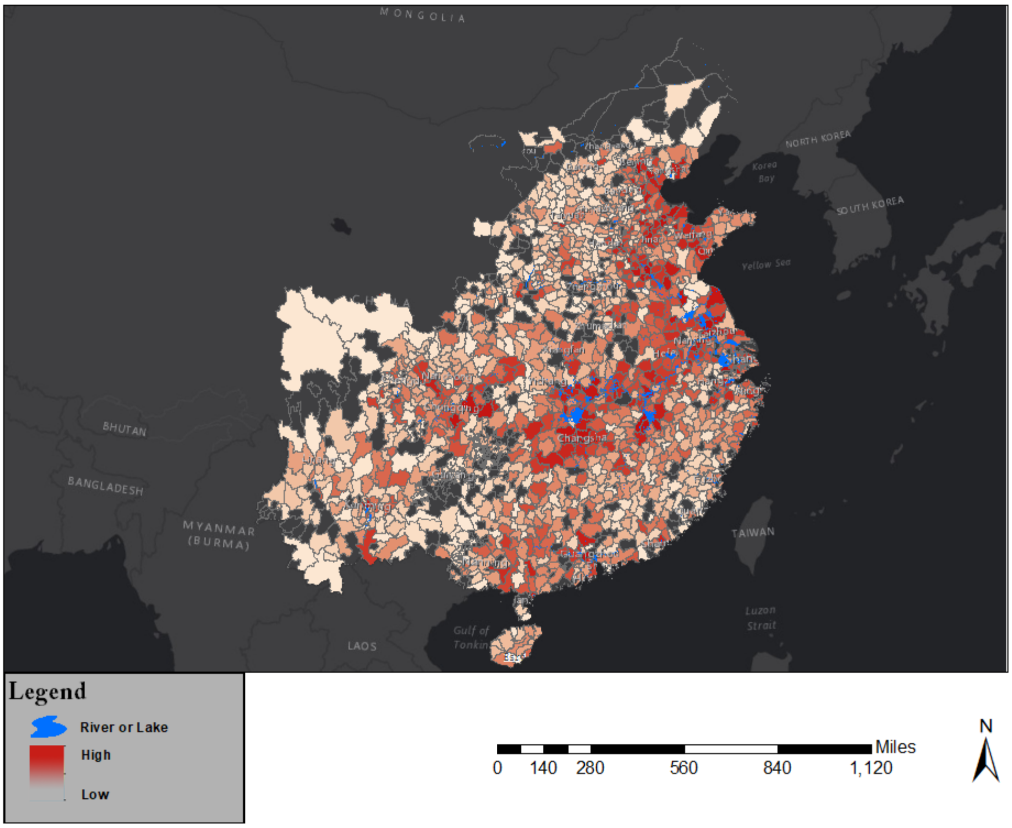


Figure 4.4: Distribution of the Frequency of Floods from 1820 to 1911

Source: China Historical Geographic Information System, Version 6. Cambridge: Harvard Yenching Institute and Fudan Center for Historical Geography, January 2016

“China’s three-thousand-year Meteorological Records (in Chinese)”

“China meteorological disaster dictionary (in Chinese)”

Note: The frequency is modified frequency every year from the source. Aggregation of frequencies from 1820 to 1911 was calculated. Symbol of “high” represents high number of aggregated frequencies.

commonly been used in previous studies although studies have reconstructed a series of typhoon events.

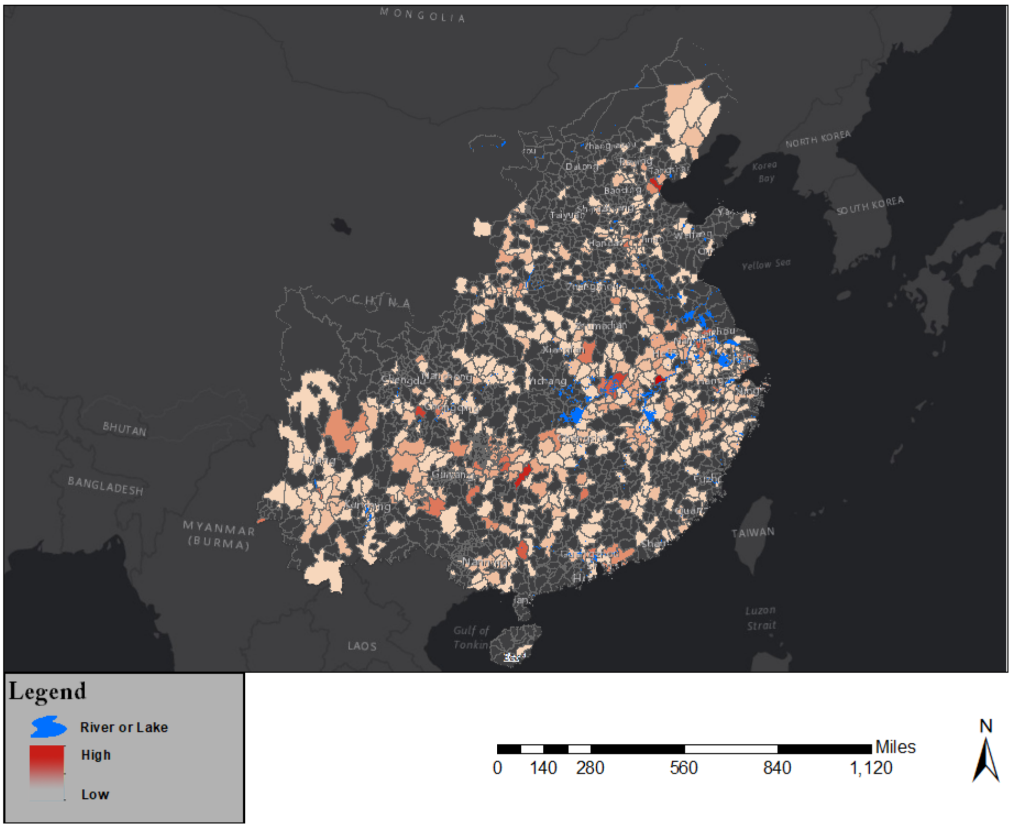


Figure 4.5: Distribution of War from 1820 to 1911

Source: China Historical Geographic Information System, Version 6. Cambridge: Harvard Yenching Institute and Fudan Center for Historical Geography, January 2016

"Chronology of Wars in China"

Note: War is counted as the number of war years for each county. Symbol "high" represents a high number of aggregated events. The highest value is 9.

Although we are focusing on the impact of environmental shocks on urbanisation we also want to control for other external shocks such as war and civil unrest. Hence, we collect data on war related events from the "Chronology of Wars in

China" (Compilation Group of Chinese military History, 2003). We collect data on the the number of years when warfare occurred across the Ming and Qing dynasty from 1368 to 1911 at the county level. For Jiangnan area during the period the first Opium War (1840-1842) and the Taiping Rebellion (1851-1864) are two prominent wars and involved millions of combatants. The population of Jiangnan region fell by around 40% from 1840 to 1880. Especially for the Taiping Rebellion, the rebel armies had conquered over 600 cities across 18 provinces (Compilation Group of Chinese military History, 2003), which caused around 20 to 70 million deaths (Cao, 2002a). In addition to these large events there were also a small number of uprisings but they tended to be suppressed within a short period. The distribution of warfare is shown in figure 4.5 and indicates that the records tended to concentrate in the same locations as the first Opium War and the Taiping Rebellion.

The population data is from "Population history of China (in Chinese)" (Cao, 2002a; Cao, 2002b). This is considered to be the most complete data set of China's historical population from 581 to 1948 following the research of Ho (1959) and has been employed by multiple studies (Deng, 2015; Cao, 1999; Li, An, et al., 2015). There are data at the prefecture-level for six base periods during the Ming-Qing dynasty (1391, 1776, 1820, 1851, 1880, 1910). Data exist for 18 provinces, but only 17 were included in this study as Gansu province does not have credible records on environmental events Xia (2015). Hence we omit this region from our analysis.

To generate our variable of interest, we took prefecture-level and county-level population density in 1820 and 1910 and converted them to county level based on a geostatistical method called areal interpolation introduced by Rosenshein (2010).

It is a general approach to implement polygon-to-polygon predictions. The population distribution of 1820 and 1910 is shown in 4.6. It can be observed that the distribution density varied to some extent from 1820 to 1910. The deviation of population density is calculated by taking the population density in 1910 minus population density in 1820. As illustrated in Figure 4.7, the population density in north China and southwest China increased while the population density in southeast China and east China fell.

The farmland data was obtained from the newest version of the "History Database of the Global Environment" (HYDE 3.2.1). This database contains 5° gridded worldwide cropland from 10,000 B.C. - 2005 A.D with ten year intervals after 1700. The total area of cropland is constructed in square kilometre per grid. Cropland areas for 1820 and 1910 in China were selected and are shown in Figure 4.8 and were then merged at the county level and prefecture level. Farmland reclamation was calculated as the cropland area in 1910 minus the area in 1820. The distribution of the increase in the cropland area as shown in Figure 4.9 implies that cropland around the main streams of rivers fell significantly.

In addition to the main variables we also include elevation data based on the Digital Elevation Model (DEM) from Shuttle Radar Topography Mission (SRTM) with 250-meter resolution. SRTM is a joint survey completed by NASA, the National Surveying and Mapping Agency of the Ministry of Defense, and the German and Italian space agencies. The SRTM system was carried on the space shuttle Endeavour launched by the United States. This mapping mission started from February 11 to 22, 2000. A total of 222 hours and 23 minutes of data collection was carried out for 11 days, and the total areas between 60 degrees north latitude and 56 degrees south latitude exceeded 119 million square meters. Kilometres of radar image data cover more than 80% of the earth' s land surface. The data volume of radar images acquired by the SRTM system is about 9.8 trillion bytes. After more than two years of data processing, a digital terrain elevation model (DEM) was created. The geographic surface of the regions in our study is illustrated in Figure 4.10. Average elevation and the Terrain Ruggedness Index (Nunn and Puga, 2012) of each geographic unit were calculated to capture the geographic characteristics of each county.

4.4.2 Empirical Strategy

In light of the theoretical predictions form Section 3.3, if the technology level has been fixed to a certain extent, the index for urbanisation should be related to the environmental impact index. Inspired (Ashraf and Michalopoulos, 2015; Michalopoulos, 2012), a baseline specification for cross-sectional data is estimated:

$$\begin{aligned} DTWN_i =& \beta_0 + \boldsymbol{\beta}_1' \boldsymbol{X}_i + \beta_2 AREA_i + \beta_3 AVGPOP_i \\ & + \beta_4 XCOORD_i + \beta_5 YCOORD_i + \boldsymbol{\gamma}' \boldsymbol{\Delta}_i + \epsilon_i \end{aligned} \tag{4.21}$$

where $DTWN_i$ is the change in the number of towns within a given geographic unit from 1820 to 1910. The geographic unit is a prefecture at the prefecture level and a county at the county level. $\boldsymbol{X}_i$ is the vector of environmental indices, which include our flood, drought, cold, and wind variables. Specifically, since the environmental indices represent the average deviation from the previous equilibrium according to the theoretical analysis, we use the mean frequency from 1820 to 1910 minus the mean frequency from 1729 to 1819 for each geographic unit and each factor. In this baseline specification, the environmental factors and dependent variable are originally time-variant. The other variables have some time-invariant geographic components. $AREA_i$ is the total area of the corresponding geographic unit. $AVGPOP_i$ is the average population of the geographic unit from 1820 to 1910 (average in years), which can partially represent economic scale. $XCOORD_i$ and $YCOORD_i$ are the longitude and the altitude of the centre point of the geographic unit. $\boldsymbol{\Delta}_i$ is the vector of a number of geographic dummies such as dummies for coastline, agriculture region (Institute of Geographic Sciences and Natural Resources of CAS, 2017), the physiographic macroregion (Skinner, Henderson, and Yue, 2007), the province and the prefecture. ϵ_i is the geographic unit specific error term.

It is normal to worry about whether the change in the number of towns really captures rural-to-urban migration since the natural growth of the urban population would also lead to an increase in the formation of market towns. Supposing that the increase in the number of towns (ΔN) comes from two factors, one is from rural-to-urban migration (ΔN_m), and the other is from the natural growth of the urban population (ΔN_g). Specifically, sources of rural-to-urban migration may be environmentally induced (ΔN_n in equation (4.7)) and non-environmentally induced (ΔN_{n2}). In this case, the increase in the number of town can be expressed as:

$$\Delta N = \Delta N_m + \Delta N_g \tag{4.22}$$

$$\Rightarrow \Delta N = \Delta N_n + \Delta N_{n2} + \Delta N_g$$

Suppose $\boldsymbol{x}$ is the vector of variables and

$$E(\Delta N|\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\beta} \tag{4.23}$$

where $\boldsymbol{\beta}$ is the vector of coefficients. Hence, the derivative of a given variable x_k is given by:

$$\frac{\partial \Delta N}{\partial x_k} = \frac{\partial \Delta N_n}{\partial x_k} + \frac{\partial \Delta N_{n2}}{\partial x_k} + \frac{\partial \Delta N_g}{\partial x_k} = \beta_k \tag{4.24}$$

If x_k is the natural disaster index, it is unlikely to have a different impact on the natural growth of population between the agriculture and non-agriculture sectors. Therefore, the impact should be entirely captured by the variable which represents the increase in population. In this regard, $\frac{\partial \Delta N_g}{\partial x_k}$ can be regarded as 0. The coefficient of key variables should represent the influences mainly on rural-to-urban migration. Additionally, since ΔN_{n2} is the non-environmental induced migration, $\frac{\partial \Delta N_{n2}}{\partial x_k} = 0$.

To be cautious, the increase induced by non-environmental factors is also estimated:

$$\begin{aligned} DTWN_i = & \beta_0 + \boldsymbol{\beta}_1' \boldsymbol{X}_i + \beta_2 AREA_i + \beta_3 AVGPOP_i \\ & + \beta_4 XCOORD_i + \beta_5 YCOORD_i \\ & + \beta_6 DWAR_i + \beta_7 DLAND_i + \beta_8 DPOP_i + \boldsymbol{\gamma}' \boldsymbol{\Delta}_i + \epsilon_i \end{aligned} \tag{4.25}$$

where $DWAR_i$ is calculated from the number of war years for each geographic unit, and also denotes the deviation of the mean like our environmental variables. $DLAND_i$ represents land reclamation, and $DPOP_i$ represents the deviation of the population in unit i.

A range of other possible geographic features can also be included to control for geographic differences. Thus, following specification is estimated:

$$
\begin{aligned}
DTWN_i = &\beta_0 + \boldsymbol{\beta}_1' \boldsymbol{X}_i + \beta_2 AREA_i + \beta_3 AVGPOP_i \\
&+ \beta_4 XCOORD_i + \beta_5 YCOORD_i \\
&+ \beta_6 DWAR_i + \beta_7 DLAND_i + \beta_8 DPOP_i \\
&+ \beta_9 DEM_i + \beta_{10} TRI_i + \beta_{11} RVRLEN_i + \boldsymbol{\gamma}' \boldsymbol{\Delta}_i + \epsilon_i
\end{aligned}
\tag{4.26}
$$

where DEM_i is the average elevation of unit i. TRI_i is the Terrain Ruggedness Index, and $RVRLEN_i$ denotes the total length of rivers within unit i.

Previous studies tended to concentrate on floods and droughts when looking at issues related to the historical climate. Hence, we look first at floods and droughts and then include cold and wind as supplementary variables to investigate the impact of different disasters. Previous studies, when employing historic environment records, usually ran their estimations at the prefecture level. Our empirical approach is to first run our regressions using cross-sectional data at the prefecture level to allow us to compare our results with previous studies and then to run them at the county-level to obtain more precise estimations.

One concern is that whether it is necessary to work at the county level. Studies at the prefecture level have so far presented a compelling narrative. However, there was some doubt about the reliability of the results given the research area is relatively small (Adams et al., 2018) and the time series data was often not long enough. County-level estimations allow us to capture more detailed spatial differences and expand the empirical approach.

We are not worried about endogeneity since this research uses environmental factors as key variables as illustrated in previous section, at least for the main result. As for other variables, we include several geographic dummies to capture as many as possible features to correct for possible biases.

4.5 Empirical Evidence

4.5.1 Prefecture Level Analysis

In this section we look at the results at the prefecture level. Table 4.1 presents the descriptive statistics for the prefecture level dataset. There are 283 prefectures in our research area and 15 variables that are included in the regression analysis.

The environmental indices are given by $DFLOOD$, $DDROUGHT$, $DCOLD$ and $DWIND$ and are the variables included in vector $\boldsymbol{X}$ in specifications (4.21), (4.25) and (4.26). To help with the presentation, some variables are in units. For example, $AREA$ is 10,000 km^2, $AVGPOP$ and $DPOP$ are denoted in ten thousand people, $DLAND$ is 100 km^2, $RVRLEN$ is 100 km, DEM is 1km, $DWAR$ and WAR are timed by 100, and TRI is divided by 100. Specifically, TWN, $FLOOD$,

Table 4.1: Descriptive Statistics for the Prefecture Level

Variable	n	Mean	S.D.	Min	0.250	Mdn	0.750	Max
DTWN	283	106.7	86.93	-5	45	88	147	609
DFLOOD	283	0.200	0.390	-1.540	0	0.110	0.300	2.340
DDROUGHT	283	0.0400	0.150	-0.420	-0.0200	0.0100	0.110	0.660
DCOLD	283	0.100	0.280	-2.030	0	0.0400	0.140	1.530
DWIND	283	0.190	0.410	-1.010	0	0.0500	0.260	2.190
AREA	283	1.360	1.190	0.0100	0.610	1.110	1.760	10.13
AVGPOP	283	140.2	127.9	-23.04	47.80	97.21	205.1	692.8
XCOORD	283	112.0	5.560	98.76	108.3	112.5	116.2	122.2
YCOORD	283	30.89	5.480	19.19	26.38	30.35	35.20	43.67
DWAR*100	283	3.020	4.330	-6.590	0	1.100	5.490	25.27
DLAND	283	1.830	4.530	-13.63	-0.250	1.340	3.850	21.12
DPOP	283	15.94	83.05	-326.0	-7.430	9.700	47.50	324.7
RVRLEN	283	6.390	4.930	0	2.830	5.180	8.520	27.05
DEM	283	0.710	0.770	0	0.150	0.450	1.080	4.280
TRI/100	283	0.980	0.640	0.0400	0.470	0.970	1.420	3.170
TWN	283	28.86	22.87	0	13	24	39	171
FLOOD	283	0.510	0.630	0	0.0800	0.300	0.700	3.770
DROUGHT	283	0.210	0.270	0	0.0200	0.120	0.310	1.820
COLD	283	0.240	0.590	0	0	0.0500	0.180	4.630
WIND	283	0.430	0.800	0	0.0100	0.120	0.490	6.220
WAR*100	283	0.710	1.480	0	0	0	1.100	9.890

Notes: $AREA$ is 10,000 km^2, $AVGPOP$ and $DPOP$ are denoted in ten thousand people, $DLAND$ is 100 km^2, $RVRLEN$ is 100 km, DEM is $1km$.

Table 4.2: Correlation Matrices of the Variables for the Prefecture level

	DTWN	DFLOOD	DDROUGHT	DCOLD	DWIND	AREA	AVGPOP
DTWN	1						
DFLOOD	0.456***	1					
DDROUGHT	0.241***	0.572***	1				
DCOLD	0.262***	0.747***	0.544***	1			
DWIND	0.396***	0.834***	0.588***	0.713***	1		
AREA	0.293***	0.123**	0.0220	0.0560	0.0810	1	
AVGPOP	0.506***	0.350***	0.0460	0.180***	0.252***	0.314***	1
XCOORD	0.00600	-0.00700	-0.0940	-0.0550	0.0460	-0.245***	0.398***
YCOORD	0.00800	0.00100	0.00300	-0.0630	-0.0630	-0.0200	0.0890
DWAR	0.286***	0.362***	0.111*	0.290***	0.288***	0.0720	0.420***
DLAND	0.146**	-0.0390	0.0410	0.0480	-0.0530	0.151**	0.0910
DPOP	0.391***	0.161***	0.100*	-0.0430	0.0310	0.282***	0.215***
RVRLEN	0.471***	0.290***	0.00300	0.117*	0.157***	0.669***	0.683***
DEM	-0.144**	-0.109*	0.00700	-0.0220	-0.125**	0.395***	-0.419***
TRI	-0.0450	-0.0360	0.0430	0.00600	-0.0420	0.244***	-0.418***

	XCOORD	YCOORD	DWAR	DLAND	DPOP	RVRLEN	DEM
XCOORD	1						
YCOORD	0.327***	1					
DWAR	0.197***	-0.104*	1				
DLAND	-0.00500	0.241***	-0.113*	1			
DPOP	-0.243***	0.0420	-0.142**	0.175***	1		
RVRLEN	0.120**	0.0810	0.306***	0.110*	0.245***	1	
DEM	-0.700***	0.0390	-0.259***	0.0780	0.0360	-0.129**	1
TRI	-0.603***	-0.432***	-0.186***	-0.0560	-0.0140	-0.110*	0.662***

	TRI
TRI	1

$DROUGHT$, $COLD$, $WIND$ and WAR represent levels of last period.

Note that the change in the number of towns covers a large range. Given the distribution shows that the first quartile is positive (45), it is clear that the increase in the number of towns was a universal phenomenon during the period from 1820 to 1910. This seems to be in contrast with the narrative that the development of market towns stagnated in the late Qing dynasty (Li, 2000; Cao, 2002a; Xu, Bas van Leeuwen, and Jan Luiten van Zanden, 2018). Regarding the data on the changes in population $DPOP$, the first quartile is negative, which suggests that some places experienced population decline but also experience an increase in the number of towns simultaneously. Therefore, there must be an increase in the proportion of the urban population induced by other factors.

In terms of the environmental variables, the highest average frequency change was for floods where the index was 0.200 higher between 1820 and 1910 in every year for each prefecture compared to 1729 to 1819. The average change in droughts was the lowest (0.04 higher in every year for each prefecture), which illustrates our research region experienced more severe and frequent flood disasters compared with the previous period which was a pattern also highlighted by Li (2007).

A correlation matrix is provided in Table 4.2 and shows that more severe floods are positively related with an increase in the number of towns (significant at the 1% level). The correlation is the highest (0.456) among all other environmental indices. However, except for the average elevation, the Terrain Ruggedness Index and the coordinates, all other variables are positively and significantly correlated with the increase in the number of towns. To understand the magnitude of any effect we need to look at the estimation results.

We now present the main results from the regression analysis. In order to facilitate the interpretation, all results are reported in terms of standardised coefficients. We start with Table 4.3 that presents the results from our cross-sectional analyses excluding cold and wind related events at the prefecture level. Column (1) and Column (2) are the results from the baseline specification (4.21). "MacroRegion" in Column (2) represents physiographic macro-region suggested by CHGIS (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University, 2016). Results show a significant correlation between floods and an increase in the number of towns at the 5% level. The coefficient in model (2) is 0.245 (the more flexible model), which indicates that one standard deviation higher in average flood frequency every year from the previous period would accelerate the increase 0.245 extra towns, which in turn encourage the migration from the rural region to the urban region. Since the levels of town and flood at the beginning are 28.86 and 0.510, 196% increase in flood frequency led to 188.53% increase in town number. The town number elasticity of flood frequency is 0.96.

In contrast, changes in drought events do not significantly impact rural-urban migration at the prefecture level. This may be because the number of drought events was of a similar magnitude in the previous one hundred year period. In other words, the drought-induced decision making did not shift significantly from the mid-Qing dynasty to the late Qing dynasty. Therefore, we find that any shift in rural-urban migration was positively impacted by flood events. This conclusion is supported by other historical studies (Li, 2007; He, Li, and Liu, 2010).

Table 4.3: Prefecture level results (without Wind and Cold)

	(1)	(2)	(3)	(4)	(5)	(6)
	DTWN	DTWN	DTWN	DTWN	DTWN	DTWN
DFLOOD	0.246**	0.245**	0.231**	0.217**	0.217**	0.200**
	(2.38)	(2.31)	(2.59)	(2.35)	(2.46)	(2.20)
DDROUGHT	-0.052	-0.093	-0.056	-0.087	-0.054	-0.080
	(-0.72)	(-1.22)	(-0.86)	(-1.30)	(-0.84)	(-1.22)
AREA	0.161*	0.217**	0.127	0.175*	0.083	0.125
	(1.88)	(2.33)	(1.53)	(1.90)	(0.73)	(0.97)
AVGPOP	0.540***	0.535***	0.467***	0.481***	0.383***	0.413***
	(8.12)	(7.76)	(7.17)	(7.24)	(4.51)	(4.64)
XCOORD	0.108	0.290	0.068	0.229	-0.023	0.134
	(0.70)	(1.59)	(0.47)	(1.30)	(-0.15)	(0.71)
YCOORD	-0.293	-0.463*	-0.318	-0.431*	-0.288	-0.369
	(-1.15)	(-1.81)	(-1.29)	(-1.72)	(-1.10)	(-1.37)
DWAR			0.103*	0.098*	0.102*	0.099*
			(1.89)	(1.75)	(1.91)	(1.80)
DLAND			0.068	0.045	0.063	0.046
			(1.16)	(0.78)	(1.08)	(0.78)
DPOP			0.175***	0.158**	0.176***	0.163**
			(2.66)	(2.39)	(2.62)	(2.42)
DEM					-0.173	-0.239
					(-1.15)	(-1.27)
TRI					0.073	0.091
					(0.81)	(0.87)
RVRLEN					0.140	0.126
					(1.20)	(0.97)
Coastline	Yes	Yes	Yes	Yes	Yes	Yes
Agriculture Zone	Yes	Yes	Yes	Yes	Yes	Yes
MacroRegion	No	Yes	No	Yes	No	Yes
Province	Yes	Yes	Yes	Yes	Yes	Yes
Observations	283	283	283	283	283	283
R^2	0.683	0.731	0.702	0.744	0.708	0.749
Adjusted R^2	0.631	0.666	0.649	0.678	0.651	0.679
F	281.6	.	.	.	.	.

Standardised coefficients are reported
t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.4: Prefecture level results (with Wind and Cold)

	(1) DTWN	(2) DTWN	(3) DTWN	(4) DTWN	(5) DTWN	(6) DTWN
DFLOOD	0.291**	0.305**	0.228*	0.232*	0.189	0.190
	(2.08)	(2.06)	(1.77)	(1.68)	(1.45)	(1.36)
DDROUGHT	-0.040	-0.076	-0.054	-0.079	-0.058	-0.076
	(-0.62)	(-1.12)	(-0.89)	(-1.28)	(-0.94)	(-1.23)
DCOLD	-0.156	-0.143	-0.133	-0.130	-0.117	-0.119
	(-1.55)	(-1.48)	(-1.46)	(-1.45)	(-1.28)	(-1.32)
DWIND	0.086	0.047	0.124	0.093	0.143*	0.115
	(0.94)	(0.53)	(1.41)	(1.10)	(1.66)	(1.34)
AREA	0.160*	0.217**	0.126	0.176*	0.080	0.124
	(1.91)	(2.33)	(1.57)	(1.95)	(0.73)	(0.99)
AVGPOP	0.529***	0.529***	0.461***	0.477***	0.381***	0.412***
	(8.31)	(7.96)	(7.37)	(7.40)	(4.72)	(4.79)
XCOORD	0.143	0.302*	0.094	0.241	0.008	0.147
	(0.95)	(1.69)	(0.66)	(1.39)	(0.05)	(0.79)
YCOORD	-0.392	-0.535**	-0.387	-0.492**	-0.368	-0.439*
	(-1.61)	(-2.18)	(-1.59)	(-2.00)	(-1.42)	(-1.66)
DWAR			0.114**	0.108*	0.114**	0.109*
			(2.04)	(1.88)	(2.07)	(1.94)
DLAND			0.077	0.056	0.070	0.056
			(1.30)	(0.97)	(1.17)	(0.94)
DPOP			0.152**	0.138**	0.162***	0.149**
			(2.59)	(2.25)	(2.73)	(2.42)
DEM					-0.154	-0.226
					(-1.02)	(-1.17)
TRI					0.086	0.099
					(0.94)	(0.93)
RVRLEN					0.136	0.125
					(1.22)	(0.99)
Coastline	Yes	Yes	Yes	Yes	Yes	Yes
Agriculture Zone	Yes	Yes	Yes	Yes	Yes	Yes
MacroRegion	No	Yes	No	Yes	No	Yes
Province	Yes	Yes	Yes	Yes	Yes	Yes
Observations	283	283	283	283	283	283
R^2	0.692	0.737	0.709	0.750	0.714	0.754
Adjusted R^2	0.638	0.671	0.654	0.682	0.656	0.683
F	.	.	.	.	.	.

Standardised coefficients are reported
t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Turning to the other variables we find that the average population is positively correlated with the increase in towns at the 1% level. The coefficient is 0.535 in Column (2), which implies that one standard deviation increase in population would encourage 0.535 more towns in a prefecture from 1820 to 1910. Thus, the number of towns would rises faster in prefectures with higher populations. As denoted in the previous section, the average population may also be capturing economic scale in historical China. In this regard, a higher population indicates higher gross domestic product. As a result, the positive correlation between the average population and increases in towns implies that a larger economies accelerate the urbanisation process.

In Columns (3) and (4) we include additional controls for farmland reclamation, the change in population and the change in the number of wars based on the specification (4.25). Columns (5) and (6) are based on the specification (4.26) and include average elevation, the Terrain Ruggedness Index (TRI), and rivers' total length as geographical controls.None of the geographical controls influence are significant and do not change the magnitude or significance of our main variables of interest.

Estimations of the change in flood frequency and the average population are consistent across specifications. The coefficients of the change in the number of war related events are consistently positive and significant at the 10% level. Coefficients of the change in population are also consistently positive at least at the 5% level across Columns (3) to (6). Specifically, $DWAR$ presents the change in the number of conflict years averaged by year. Thus, the maximum value of $DWAR$ in theory should be one. In this regard, $DWAR$ can be interpreted as the percentage change in years that have conflicts or wars. For example, in Column (6), the coefficient is 0.099. Then, if $DWAR$ increases by one standard deviation, there would be 0.099 extra towns. As described in previous studies (Cao, 2002a; Zou, 2013), people moved to the city or other urban areas to escape from warfare since urban areas had a higher defence capacity.

For population growth, the change in population represents the population growth in a prefecture, of which the coefficient is 0.163 in Column (6). The interpretation is that if a population grows by one standard deviation compared to the previous period this would accelerate the increase in towns by 0.163 units. In other words, 59,000 more people would increase one more town. This is very close to previous estimations for the urbanisation rate in historical China (around 5%-25% at the province level) (Cao, 2002a; Li, 2000; Xu, Bas van Leeuwen, and Jan Luiten van Zanden, 2018). To be more specific, Cao (2002a) estimates that the population of town would be around 1,000 to 20,000. Thus, 5%-25% of 59,000 more people in the population growth is enough to form a new town.

In table 4.4 we include cold and wind to the previous estimations. The results are generally consistent with the previous results excluding cold and wind. The average population, the change in the frequency of war, and population changes remain consistent across all specifications. Coefficients of the deviation of the flood frequency are significant at the 5% level in Columns (1) and (2), at the 10% level in Columns (3) and (4) but this time we lose significance when geographic features are included in Columns (5) and (6). By themselves $DCOLD$ and $DWIND$ are generally insignificant. As reported in Table 4.2, $DCOLD$ and $DWIND$ have high correlations with $DFLOOD$, and $DWIND$ is significantly correlated with

DEM and $RVRLEN$. It is possible that $DFLOOD$ is capturing some artificial variation which is related to other environmental factors. Therefore, we need to run similar estimations at a higher geographical resolution.

4.5.2 County Level Analysis

Table 4.5: Descriptive Statistics at the County Level

Variable	n	Mean	S.D.	Min	0.250	Mdn	0.750	Max
DTWN	1634	18.34	11.79	-5	11	17	24	82
DFLOOD	1634	0.0300	0.0900	-0.620	-0.0100	0.0200	0.0700	0.430
DDROUGHT	1634	0.0100	0.0400	-0.280	-0.0100	0	0.0200	0.430
DCOLD	1634	0.0200	0.0700	-1.200	0	0	0.0300	0.590
DWIND	1634	0.0300	0.100	-0.890	0	0.0100	0.0500	0.750
AREA	1634	0.230	0.390	0	0.100	0.170	0.270	9.550
AVGPOP	1634	24.17	21.91	-25.44	9.910	19.25	32.37	307.7
XCOORD	1634	112.3	5.470	97.78	108.4	113.0	116.5	122.4
YCOORD	1634	30.78	5.260	18.54	26.38	30.47	35.13	44.17
DWAR*100	1634	0.520	1.090	-3.300	0	0	1.100	7.690
DLAND	1634	0.310	1.080	-4.900	-0.120	0.160	0.650	10.21
DPOP	1634	2.800	14.78	-80.60	-1.910	2.510	8.550	173.6
RVRLEN	1634	1.100	1.040	0	0.450	0.850	1.450	10.60
Dem	1634	0.570	0.650	0	0.0800	0.350	0.850	4.510
TRI/100	1634	0.880	0.670	0.0400	0.240	0.850	1.370	3.320
TWN	1634	4.980	4.010	0	2	4	6	41
FLOOD	1634	0.0900	0.110	0	0.0100	0.0500	0.130	0.930
DROUGHT	1634	0.0400	0.0500	0	0	0.0200	0.0500	0.750
COLD	1634	0.0400	0.110	0	0	0.0100	0.0300	1.740
WIND	1634	0.0700	0.140	0	0	0.0200	0.0800	1.430
WAR*100	1634	0.120	0.430	0	0	0	0	4.400

Notes: $AREA$ is 10,000 km^2, $AVGPOP$ and $DPOP$ are denoted in ten thousand people, $DLAND$ is 100 km^2, $RVRLEN$ is 100 km, DEM is $1km$.

Table 4.5 displays the descriptive statistics for the county-level dataset. Observe that the number of observations increases from 283 to 1,634. For consistency we

Table 4.6: Correlation Matrix at the County level

	DTWN	DFLOOD	DDROUGHT	DCOLD	DWIND	AREA	AVGPOP
DTWN	1						
DFLOOD	0.097***	1					
DDROUGHT	0.085***	0.597***	1				
DCOLD	0.0290	0.653***	0.442***	1			
DWIND	0.096***	0.750***	0.506***	0.636***	1		
AREA	0.144***	-0.00800	-0.0160	-0.0190	-0.0170	1	
AVGPOP	0.194***	0.044*	-0.068***	-0.00300	0.0120	0.366***	1
XCOORD	-0.092***	-0.0330	-0.078***	-0.051**	0.00100	-0.168***	0.347***
YCOORD	0.042*	0.0110	0.00800	-0.0260	-0.0360	0.00100	0.120***
DWAR	0.0220	0.071***	-0.0150	0.053**	0.081***	-0.067***	0.102***
DLAND	0.096***	-0.0170	0.0120	0.00500	-0.047*	0.124***	0.0210
DPOP	0.307***	0.042*	0.046*	-0.047*	-0.00200	0.311***	0.270***
RVRLEN	0.207***	0.049**	-0.0380	-0.00700	-0.00400	0.676***	0.544***
Dem	0.096***	-0.0200	0.041*	0.0370	-0.0350	0.385***	-0.272***
TRI	0.090***	-0.00100	0.043*	0.0190	-0.00800	0.214***	-0.281***

	XCOORD	YCOORD	DWAR	DLAND	DPOP	RVRLEN	Dem
XCOORD	1						
YCOORD	0.374***	1					
DWAR	0.112***	-0.062**	1				
DLAND	-0.0250	0.190***	-0.099***	1			
DPOP	-0.256***	0.0390	-0.143***	0.094***	1		
RVRLEN	0.0350	0.093***	0.0350	0.114***	0.250***	1	
Dem	-0.681***	-0.097***	-0.130***	0.126***	0.067***	0.097***	1
TRI	-0.492***	-0.416***	-0.124***	-0.00300	-0.00800	0.0350	0.668***

	TRI
TRI	1

estimate the same specifications at the county-level as we did at the prefecture-level. Not surprisingly the mean values are smaller. For example, the mean change in the number of towns has fallen from 106.7 in Table 4.1 to 18.34, with a maximum value of 82 and a minimum value of -5. Intuitively, the scale should not alter the relationship between the dependent variable and independent variables. However, correlations between the increase of the number of towns and environmental factors also decrease as shown in Table 4.6. Specifically, the correlation of $DFLOOD$ and $DTWN$ is 0.097, indicating a weaker connection at the county level.

As shown in Table 4.7, when wind and cold are not included, coefficients on the change in the number of flood events is significant and positive at the 5% level in Columns (1) to (5), but only the 10% level in Column (6). As expected, the magnitude of these coefficients is much smaller as the average county has considerably fewer towns at the mean. Taken Column (2) as example, the town number elasticity of flood frequency is 0.14, which verified the weaker connection as well. The average population is consistent across all models and all levels (the prefecture and county levels). The change in the number of war related events is consistently significant.

The coefficients on the change in population are significant at the 1% level in Columns (3) and (5) but lose some significance in Columns (4) and (6). Specifically, Columns (1), (3) and (5) report estimations with province dummies, but Columns (2), (4) and (6) include prefecture dummies instead. Accordingly, the prefecture dummy may capture some prefecture traits related to flooding and population growth which reduces the significance.

Turning to the geographic controls we now find that the total length of rivers is positively related to the change in the number of towns at the 1% level in a county. As reported in Column (6), one standard deviation longer in the total river length increases 0.260 extra towns. Thus, more rivers appear to encourage rural-to-urban migration since the foundation of market towns relies heavily on the water network for transportation (Skinner, 1964; Zou, 2013).

Overall, the county level results are broadly in line with the prefecture level results. More severe floods encourage migration from the rural areas to the urban areas but the elasticity is much smaller at the county level. What we do show is that easier access to the water network encouraged the formation of new towns. Therefore, the prefecture estimations would appear to exacerbate the environmental impact and miss the influence from some geographic features such as river length.

Wind and cold are included in Table 4.8. The change in the floods and the change in population coefficients lose significance when the prefecture is controlled in Columns (2), (4) and (6). Now we find that cold and the wind demonstrate some significance with the growth of towns. The significance is lost when the prefecture dummies are included, but the significance is relatively high in Columns (1), (3) and (5). For example, in Column (5), the coefficient of the change in cold events is significant and negative at 1% level, and the coefficient of the change in the number of wind related events is significant and positive at the 1% level. The negative coefficients of cold events can be regarded as positive impact on urban-to-rural migration.

As for prefecture dummies, it may be that these dummies are capturing some

Table 4.7: County level Results (without Wind and Cold)

	(1) DTWN	(2) DTWN	(3) DTWN	(4) DTWN	(5) DTWN	(6) DTWN
DFLOOD	0.067***	0.062**	0.062**	0.054**	0.052**	0.041*
	(2.59)	(2.45)	(2.45)	(2.17)	(2.12)	(1.69)
DDROUGHT	-0.009	-0.011	-0.007	-0.006	-0.005	-0.007
	(-0.36)	(-0.45)	(-0.27)	(-0.24)	(-0.18)	(-0.25)
AREA	0.101	0.153	0.080	0.137	-0.032	0.003
	(1.16)	(0.96)	(0.94)	(0.89)	(-0.38)	(0.02)
AVGPOP	0.371***	0.359***	0.332***	0.338***	0.265***	0.240***
	(8.24)	(5.83)	(7.91)	(5.52)	(5.51)	(4.23)
XCOORD	0.565***	0.212	0.520***	0.188	0.559***	0.217
	(4.39)	(0.39)	(4.06)	(0.34)	(4.37)	(0.44)
YCOORD	-0.316**	-0.097	-0.309**	-0.081	-0.415***	-0.059
	(-2.24)	(-0.75)	(-2.21)	(-0.66)	(-2.95)	(-0.60)
DWAR			0.082***	0.093***	0.079***	0.077***
			(3.94)	(4.57)	(3.90)	(3.98)
DLAND			0.016	0.023	0.008	0.021
			(0.50)	(0.76)	(0.26)	(0.70)
DPOP			0.118***	0.071	0.130***	0.089*
			(2.69)	(1.42)	(3.37)	(1.94)
Dem					0.097	0.103
					(1.23)	(0.92)
TRI					0.034	-0.018
					(0.72)	(-0.27)
RVRLEN					0.208***	0.260***
					(3.54)	(4.68)
Coastline	Yes	Yes	Yes	Yes	Yes	Yes
Agriculture Zone	Yes	Yes	Yes	Yes	Yes	Yes
MacroRegion	Yes	Yes	Yes	Yes	Yes	Yes
Province	Yes	No	Yes	No	Yes	No
Prefecture	No	Yes	No	Yes	No	Yes
Observations	1634	1634	1634	1634	1634	1634
R^2	0.424	0.640	0.434	0.646	0.449	0.659
Adjusted R^2	0.403	0.552	0.413	0.559	0.427	0.574
F	24.90	.	24.38	.	24.12	.

Standardised coefficients are reported
t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.8: County level Results (with Wind and Cold)

	(1) DTWN	(2) DTWN	(3) DTWN	(4) DTWN	(5) DTWN	(6) DTWN
DFLOOD	0.074**	0.052	0.068**	0.050	0.054*	0.037
	(2.21)	(1.62)	(2.08)	(1.56)	(1.70)	(1.25)
DDROUGHT	-0.007	-0.011	-0.005	-0.006	-0.004	-0.006
	(-0.28)	(-0.43)	(-0.21)	(-0.21)	(-0.14)	(-0.22)
DCOLD	-0.087***	-0.037*	-0.078***	-0.031	-0.077***	-0.030
	(-4.02)	(-1.89)	(-3.66)	(-1.61)	(-3.61)	(-1.64)
DWIND	0.065**	0.042*	0.058**	0.030	0.064**	0.029
	(2.37)	(1.72)	(2.13)	(1.24)	(2.34)	(1.20)
AREA	0.100	0.153	0.081	0.138	-0.031	0.005
	(1.16)	(0.97)	(0.97)	(0.90)	(-0.37)	(0.03)
AVGPOP	0.371***	0.357***	0.335***	0.338***	0.296***	0.240***
	(8.38)	(5.81)	(8.14)	(5.52)	(5.70)	(4.23)
XCOORD	0.572***	0.240	0.528***	0.209	0.571***	0.239
	(4.48)	(0.60)	(4.14)	(0.51)	(4.49)	(0.61)
YCOORD	-0.352**	-0.114	-0.341**	-0.094	-0.449***	-0.073
	(-2.50)	(-0.73)	(-2.44)	(-0.63)	(-3.21)	(-0.57)
DWAR			0.082***	0.091***	0.078***	0.076***
			(3.90)	(4.47)	(3.86)	(3.89)
DLAND			0.017	0.024	0.009	0.021
			(0.53)	(0.76)	(0.29)	(0.71)
DPOP			0.108**	0.068	0.120***	0.087*
			(2.47)	(1.36)	(3.15)	(1.87)
Dem					0.102	0.104
					(1.30)	(0.94)
TRI					0.038	-0.016
					(0.80)	(-0.24)
RVRLEN					0.206***	0.260***
					(3.55)	(4.69)
Coastline	Yes	Yes	Yes	Yes	Yes	Yes
Agriculture Zone	Yes	Yes	Yes	Yes	Yes	Yes
MacroRegion	Yes	Yes	Yes	Yes	Yes	Yes
Province	Yes	No	Yes	No	Yes	No
Prefecture	No	Yes	No	Yes	No	Yes
Observations	1634	1634	1634	1634	1634	1634
R^2	0.428	0.641	0.437	0.647	0.453	0.660
Adjusted R^2	0.407	0.552	0.416	0.559	0.430	0.574
F	24.96	.	24.34	.	23.87	.

Standardised coefficients are reported
t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

common environmental features at the prefecture level, since there are comments claiming that the climate followed similar patterns across neighbouring prefectures (Zhang, 1996; Zheng, Ge, Fang, et al., 2007; Wen, 2006). It is also possible that estimations without prefecture controls are biased due to missing variables. For example, the influence of environment changes can be non-linear and may be interacted with other factors. Therefore, we run a number of alternative specifications to investigate this issue further.

4.5.3 Alternative Estimations

The first concern is that certain factors may affect the longer term and have possible common regional features. Therefore, changes from the last period are observed. As reported in Table 4.9, $DFLOOD$,$DDROUGHT$,$DCOLD$,$DWIND$ and $DWAR$ are average excess deviations in the current period (the mean of the frequency from 1820 to 1910 minus the mean of the frequency from 1729 to 1819), variables with $L.$ are average excess deviations of last period (the mean of the frequency from 1729 to 1819 minus the mean of the frequency from 1638 to 1728). Lags of these variables are considered since aggregation of these variables (or the stock of these variables) across periods are usually not considered.

According to Table 4.10, the significance of the flood and the cold variables increase even if when we include prefecture dummies. Meanwhile, the lag of the flood variable is significant in Columns (2) and (4). In addition, the lag of average excess wars is significant and negative rather than positive, which indicates that more frequent conflicts in the previous period discouraged migration from

Table 4.9: Descriptive Statistics of Lags at the County level Level

Variable	n	Mean	S.D.	Min	0.250	Mdn	0.750	Max
DLFOOD	1634	0.0300	0.0900	-0.620	-0.0100	0.0200	0.0700	0.430
DDROUGHT	1634	0.0100	0.0400	-0.280	-0.0100	0	0.0200	0.430
DCOLD	1634	0.0200	0.0700	-1.200	0	0	0.0300	0.590
DWIND	1634	0.0300	0.100	-0.890	0	0.0100	0.0500	0.750
DWAR*100	1634	0.520	1.090	-3.300	0	0	1.100	7.690
L.DLFOOD	1634	0	0.0800	-0.560	-0.0300	0	0.0300	0.610
L.DDROUGHT	1634	-0.0100	0.0500	-0.360	-0.0300	0	0.0100	0.320
L.DCOLD	1634	0	0.0600	-0.560	-0.0100	0	0.0100	0.530
L.DWIND	1634	0	0.100	-0.960	-0.0200	0	0.0100	0.980
L.DWAR	1634	-0.500	1.160	-8.790	-1.100	0	0	4.400

the rural area to the urban area, possibly due to the prolonged influence on the population level. To be specific, more severe warfare decreased the overall population level in the previous period, which discouraged rural-to-urban migration. The situation in Fuzhou prefecture might indicate the impact of the severe warfare from the population perspective. From 1391 to 1781, the average growth rate was only 0.05%, which is significantly lower than the average level, mainly due to war-induced population loss (Cao, 2002b). For example, in 1648 alone, at least 100 thousand people were involved in wars. Another example is Yangzhou prefecture, where experienced a ten-day massacre in 1645 and lost around 800 thousand people. Before the massacre, the population of Yangzhou city was around one million, but during the Qing dynasty after 1645, the population of the Yangzhou city only recovered to 400 to 500 thousand (Cao, 2002b).

A second concern is whether environmental degradation has a non-linear relationship with urban-rural migration, which would match the theoretical predictions.

Table 4.10: County level Results with Lags

	(1) DTWN	(2) DTWN	(3) DTWN	(4) DTWN
DFLOOD	0.055**	0.056**	0.069**	0.059*
	(2.16)	(2.16)	(2.07)	(1.84)
DDROUGHT	0.008	-0.007	0.007	-0.006
	(0.30)	(-0.24)	(0.26)	(-0.21)
DCOLD			-0.084***	-0.041*
			(-3.48)	(-1.93)
DWIND			0.051*	0.026
			(1.78)	(1.04)
L.DLFOOD	0.034	0.055**	0.082**	0.082**
	(1.37)	(2.03)	(2.39)	(2.37)
L.DDROUGHT	0.031	-0.001	0.048**	0.008
	(1.28)	(-0.03)	(1.98)	(0.32)
L.DCOLD			-0.011	-0.018
			(-0.43)	(-0.70)
L.DWIND			-0.070**	-0.032
			(-2.40)	(-1.16)
DWAR	0.054***	0.055***	0.055***	0.054***
	(2.61)	(2.74)	(2.67)	(2.70)
L.DWAR	-0.083***	-0.066***	-0.085***	-0.067***
AREA	-0.022	0.016	-0.023	0.017
	(-0.27)	(0.11)	(-0.28)	(0.12)
AVGPOP	0.261***	0.237***	0.262***	0.237***
	(5.54)	(4.27)	(5.79)	(4.35)
XCOORD	0.533***	0.206	0.551***	0.234
	(4.15)	(0.34)	(4.31)	(0.55)
YCOORD	-0.396***	-0.070	-0.447***	-0.102
	(-2.81)	(-0.68)	(-3.17)	(-0.71)
DLAND	0.015	0.029	0.017	0.031
	(0.51)	(0.96)	(0.59)	(1.01)
DPOP	0.127***	0.083*	0.123***	0.081*
	(3.34)	(1.82)	(3.25)	(1.76)
DEM	0.085	0.078	0.093	0.078
	(1.09)	(0.70)	(1.19)	(0.70)
TRI	0.041	-0.006	0.042	-0.002
	(0.86)	(-0.08)	(0.90)	(-0.02)
RVRLEN	0.192***	0.245***	0.194***	0.244***
	(3.34)	(4.46)	(3.39)	(4.44)
Coastline	Yes	Yes	Yes	Yes
Agriculture Zone	Yes	Yes	Yes	Yes
MacroRegion	Yes	Yes	Yes	Yes
Province	Yes	No	Yes	No
Prefecture	No	Yes	No	Yes
Observations	1634	1634	1634	1634
R^2	0.457	0.664	0.462	0.665
Adjusted R^2	0.434	0.579	0.439	0.579
F	23.94	.	23.35	.

Standardised coefficients are reported
t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Hence, we include quadratic terms of environmental factors in Table 4.11. We only present the results for the environmental factors and their quadratic terms for reasons of space. All other variables remain the same as those in specification (4.26).

The results suggest that there is no non-linear relationship. Specifically, the average effect of the flood is to encourage rural-to-urban migration, and the average effect of the cold is to discourage rural-to-urban migration. The significance of the lag of the flood implies that the flood might impose influences across more prolonged periods.

Finally, we include interaction effects between environmental and non-environmental factors in Table 4.12. Interaction terms between environmental factors and $DWAR$,$DPOP$,$DLAND$ are assumed to capture possible mitigation effects. Similarly, only environmental factors and relevant interaction terms presented.

A general trend of all estimations in Columns (1) to (9) is that the magnitude of the coefficients for the flood and the cold variables increase, especially when we include prefecture dummies in Column 8, which indicates that results without interaction term could potentially be biased. To be specific, effects of the flood and the cold are mitigated by more frequent warfare since interaction terms $DFLOOD \times DWAR$ and $DCOLD \times DWAR$ are significant at the 5% level. As for the land reclamation, there is no significant evidence that land reclamation mitigates the impact from environmental shocks.

Overall, what we show in our analysis is that environmental factors impact rural-urban migration, which suggests that the agriculture sector and the non-agriculture

Table 4.11: County level Results with Quadratic Terms

	(1) DTWN	(2) DTWN	(3) DTWN	(4) DTWN	(5) DTWN	(6) DTWN	(7) DTWN	(8) DTWN
DFLOOD	0.047*	0.038	0.051*	0.058**	0.055	0.028	0.070*	0.052
	(1.92)	(1.54)	(1.96)	(2.13)	(1.46)	(0.76)	(1.81)	(1.35)
DFLOOD × DFLOOD	0.012	0.005	0.004	-0.007	-0.001	0.020	-0.001	0.017
	(0.53)	(0.30)	(0.19)	(-0.23)	(-0.02)	(0.50)	(-0.03)	(0.47)
DDROUGHT	-0.001	-0.002	0.014	-0.002	-0.002	-0.002	0.012	-0.001
	(-0.03)	(-0.05)	(0.55)	(-0.03)	(-0.07)	(-0.05)	(0.46)	(-0.01)
DDROUGHT × DDROUGHT	-0.009	-0.014	-0.018	-0.017	-0.005	-0.014	-0.016	-0.017
	(-0.33)	(-0.76)	(-0.73)	(-0.92)	(-0.19)	(-0.73)	(-0.63)	(-0.97)
L.DLFOOD			0.033	0.056**			0.082**	0.084**
			(1.32)	(2.02)			(2.41)	(2.39)
L.DDROUGHT			0.035	0.002			0.050**	0.009
			(1.44)	(0.11)			(2.05)	(0.38)
DCOLD					-0.080***	-0.031	-0.088***	-0.043**
					(-3.29)	(-1.57)	(-3.45)	(-1.97)
DCOLD × DCOLD					-0.005	-0.006	-0.009	-0.004
					(-0.28)	(-0.31)	(-0.49)	(-0.19)
DWIND					0.063**	0.033	0.050	0.034
					(2.10)	(1.14)	(1.60)	(1.15)
DWIND × DWIND					0.004	-0.014	0.002	-0.024
					(0.13)	(-0.53)	(0.06)	(-0.86)
L.DCOLD							-0.008	-0.023
							(-0.30)	(-0.84)
L.DWIND							-0.071**	-0.024
							(-2.34)	(-0.85)
Coastline	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agriculture Zone	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MacroRegion	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province	Yes	No	Yes	No	Yes	No	Yes	No
Prefecture	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1634	1634	1634	1634	1634	1634	1634	1634
R^2	0.449	0.660	0.457	0.664	0.453	0.660	0.463	0.665
Adjusted R^2	0.427	0.574	0.434	0.578	0.429	0.573	0.438	0.578
F	23.42	.	23.27	.	22.50	.	22.21	.

Standardised coefficients are reported
t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.12: County level Results with Interaction effects

	(1) DTWN	(2) DTWN	(3) DTWN	(4) DTWN	(5) DTWN	(6) DTWN	(7) DTWN	(8) DTWN
DFLOOD	0.049*	0.047*	0.057*	0.064**	0.030	0.046	0.051	0.071**
	(1.69)	(1.80)	(1.92)	(2.34)	(0.78)	(1.34)	(1.30)	(2.00)
DDROUGHT	0.011	0.001	0.020	0.001	0.005	-0.003	0.013	-0.003
	(0.39)	(0.07)	(0.67)	(0.05)	(0.19)	(-0.06)	(0.45)	(-0.05)
DFLOOD × DWAR	-0.036	-0.050**	-0.043*	-0.051**	-0.025	-0.060*	-0.039	-0.066**
	(-1.36)	(-2.12)	(-1.65)	(-2.18)	(-0.67)	(-1.86)	(-1.09)	(-2.08)
DFLOOD × DLAND	0.024	0.019	0.018	0.011	0.066*	0.023	0.064*	0.017
	(0.94)	(0.74)	(0.69)	(0.44)	(1.73)	(0.63)	(1.65)	(0.46)
DFLOOD × DPOP	0.063**	0.039	0.055**	0.035	0.057	0.046	0.047	0.039
	(2.25)	(1.46)	(2.00)	(1.34)	(1.33)	(1.04)	(1.11)	(0.88)
DDROUGHT × DWAR	-0.017	-0.018	-0.015	-0.019	-0.011	-0.018	-0.009	-0.018
	(-0.71)	(-0.89)	(-0.69)	(-0.98)	(-0.44)	(-0.88)	(-0.42)	(-0.89)
DDROUGHT × DLAND	-0.015	0.008	-0.012	0.009	-0.008	0.007	-0.007	0.009
	(-0.66)	(0.33)	(-0.51)	(0.42)	(-0.36)	(0.31)	(-0.29)	(0.38)
DDROUGHT × DPOP	-0.018	-0.011	-0.005	-0.001	-0.020	-0.011	-0.010	-0.003
	(-0.65)	(-0.42)	(-0.20)	(-0.06)	(-0.74)	(-0.43)	(-0.36)	(-0.15)
L.DLFOOD			0.027	0.050*			0.080**	0.081**
			(1.11)	(1.86)			(2.37)	(2.33)
L.DDROUGHT			0.036	0.003			0.059**	0.019
			(1.46)	(0.13)			(2.43)	(0.77)
DCOLD					-0.076**	-0.051*	-0.084**	-0.061**
					(-2.29)	(-1.87)	(-2.42)	(-2.15)
DWIND					0.093***	0.040	0.076**	0.034
					(2.70)	(1.31)	(2.18)	(1.10)
DCOLD × DWAR					0.036	0.056**	0.043	0.058**
					(1.30)	(2.21)	(1.56)	(2.28)
DCOLD × DLAND					-0.026	0.006	-0.027	0.006
					(-0.93)	(0.18)	(-0.95)	(0.16)
DCOLD × DPOP					-0.001	-0.014	-0.002	-0.016
					(-0.04)	(-0.44)	(-0.07)	(-0.50)
DWIND × DWAR					-0.045	-0.027	-0.040	-0.024
					(-1.52)	(-1.07)	(-1.57)	(-0.99)
DWIND × DLAND					-0.027	-0.006	-0.029	-0.007
					(-0.76)	(-0.13)	(-0.80)	(-0.14)
DWIND × DPOP					0.001	-0.001	0.011	0.011
					(0.02)	(-0.01)	(0.27)	(0.25)
L.DCOLD							-0.014	-0.025
							(-0.57)	(-1.01)
L.DWIND							-0.078***	-0.035
							(-2.66)	(-1.26)
Coastline	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agriculture Zone	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MacroRegion	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province	Yes	No	Yes	No	Yes	No	Yes	No
Prefecture	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1634	1634	1634	1634	1634	1634	1634	1634
R^2	0.454	0.663	0.462	0.667	0.458	0.664	0.468	0.669
Adjusted R^2	0.430	0.577	0.437	0.581	0.431	0.575	0.440	0.580
F	23.60	.	23.95	.	21.61	.	21.97	.

Standardised coefficients are reported
t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

sector respond differently to different environmental events. As predicted in theory, if the relative price is stable ($0<\varepsilon<1$), average excess environmental incidents should encourage migration from rural to urban areas. However, cold related events do not appear to follow the theoretical predictions. A possible interpretation based on the theoretical framework is that severe cold events affect labour supply more in the non-agriculture sector , or that cold related events alter price sensitivity so it becomes less stable. For example, since the price-sensitive is referring to people's responses to the relative scarcity of production, cold weather would change the preference of food demand to maintain more energy (Bhattacharya et al., 2003) or to deal with possible famine in the pre-industrial age. In consequence, the price would be less stable. Furthermore, as mentioned by previous studies (Elvin and Liu, 1998; Liu et al., 2006; Xiao et al., 2018), cold weather had a long-run impact on agriculture production in a large area, which suggests a lower capacity of official food stock to release relieves. There was abnormal cold weather even in the Southernmost part of China, Guangdong province during the Ming and Qing dynasty, where domestic animals were frozen to death (Zhang, Zhan, et al., 2004). Not only does the three-cropping system a year no longer exist, but the scope of the two-cropping system is also significantly reduced in the Guangdong area due to the continuous cooling. For instance, the relative production during the colder period (1883-1911) was only 68% but 86% in the warmer period (1730-1749) in terms of harvest index (He, Li, and Liu, 2010). Therefore, it is possible to violate the original mechanism of price stability through governmental food stock if lacking sufficient nongovernmental relief (Zhi-yuan, 2008). As a result, price changes more sensitively in response to the relative scarcity of agricultural goods, encouraging people to move back to the rural area continuously.

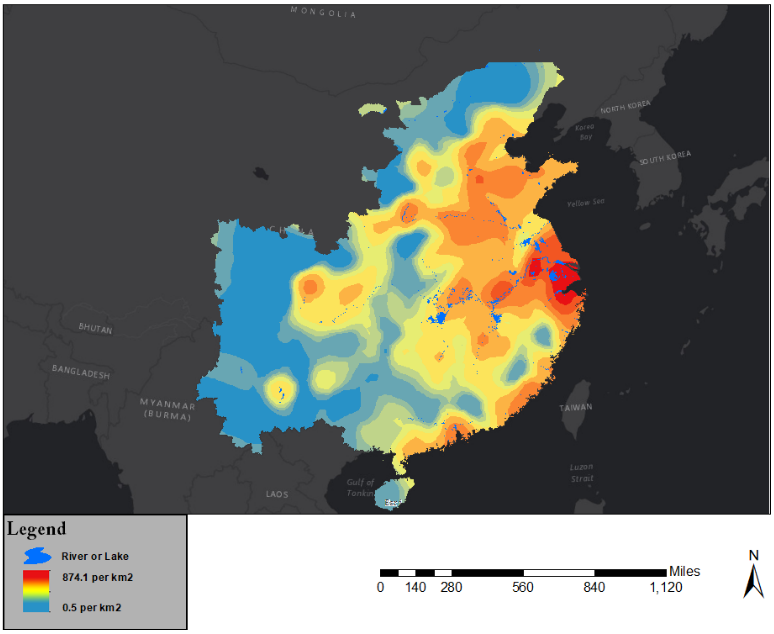
4.6 Conclusions

This research examines the impact of environmental events on rural-urban migration theoretically and empirically. The theory adopts a two-sector Harris-Todaro model to include environmental degradation on the marginal product as an incentive for people to migrate. It argues that unfavourable environmental conditions would encourage people to move to the urban region for higher expected incomes. However, different natural disasters impose different impacts on the expected wage of the agricultural sector and non-agricultural sector. In addition, farmland reclamation was usually considered to be first choice as way to mitigate the reduction of expected income rather than deciding to migrate.

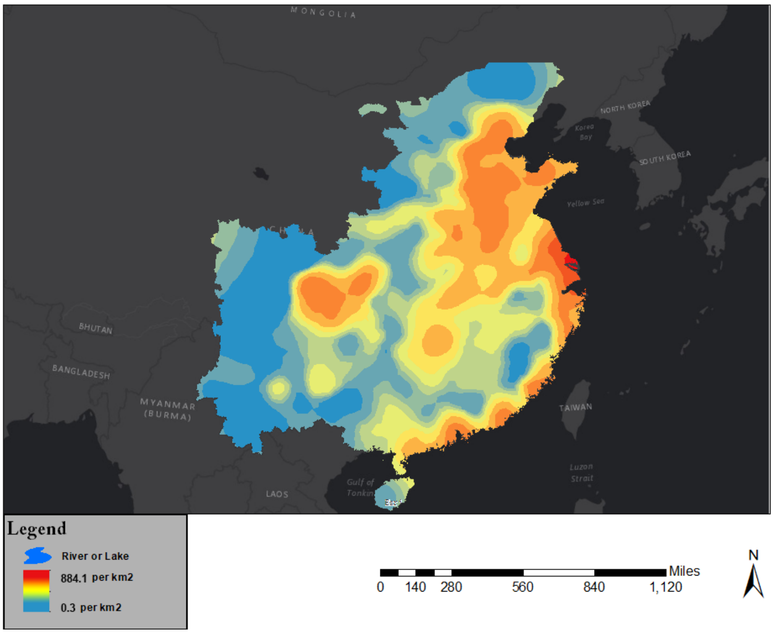
The theoretical predictions regarding a possible positive effect of natural disasters on migration is tested using regression analysis. Cross-sectional regressions at the prefecture level and the county level were estimated. Our main finding is that more frequent flood events encourages rural-to-urban migration, while more frequent cold related events encourages urban-to-rural migration.

There are several weaknesses with the theoretical framework. One is that we omit the cost of migration. Todaro (1969) extended the two-sector model to a two-stage model, which claims that unskilled migrants from the rural area can not find high rewarding jobs immediately during the first stage but remain in the so-called "urban traditional" sector, indicating that unskilled people need to spend time to find better jobs. What is more, it is also possible that unskilled people need to afford extra cost to receive some training (Lucas, 2004). Another study regarded rural-urban migration as a risk avoidance process (Stark and Levhari, 1982). Therefore, risk was considered the major factor in the theoretical model. Possible improvements could come from the difference between the long term and short term, and interactions among different factors (Tacoli, 1998).

Due to data availability, the urbanisation rate could not be estimated at a higher resolution. Furthermore, we do not consider spatial correlation in this analysis which might lead to bias in the results. Further research is needed on the data reconstruction at a lower level and to think carefully about the use of applied spatial models such as SAR (Spatial Auto-regressive model) and SDM (Spatial Durbin Model).



(a) 1820



(b) 1911

Figure 4.6: Distribution of Population Density from 1820 to 1911

Source: China Historical Geographic Information System, Version 6. Cambridge: Harvard Yenching Institute and Fudan Center for Historical Geography, January 2016 "Population history of China (in Chinese)"

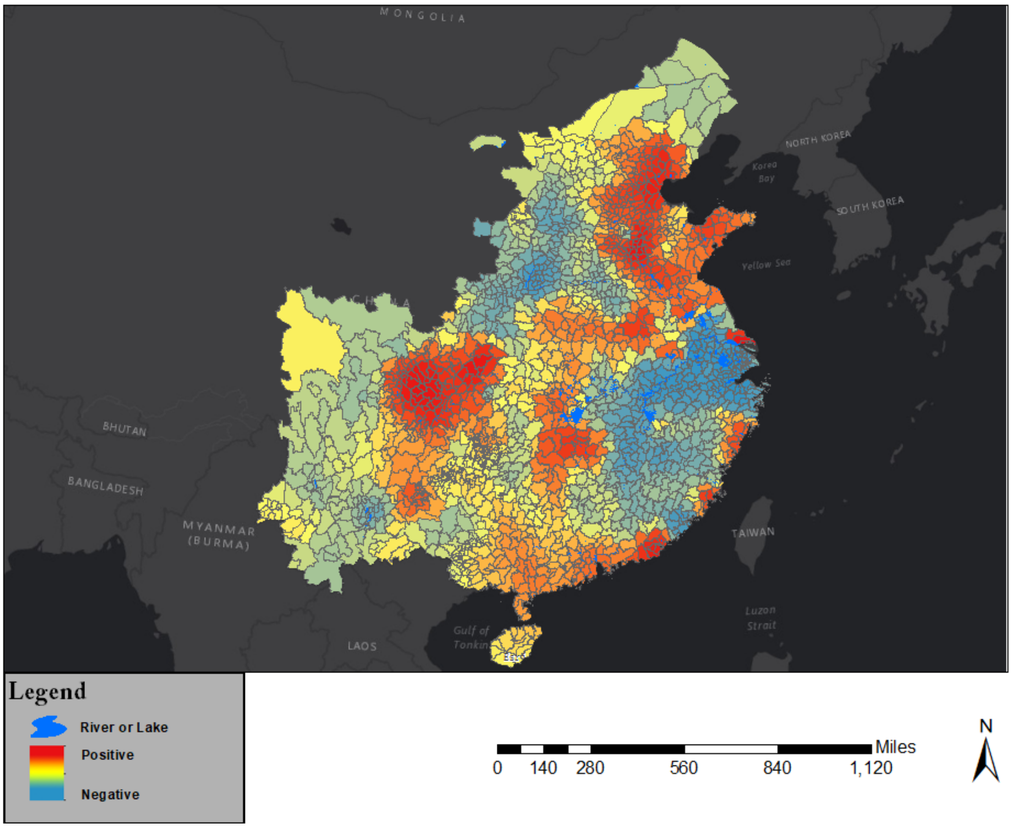
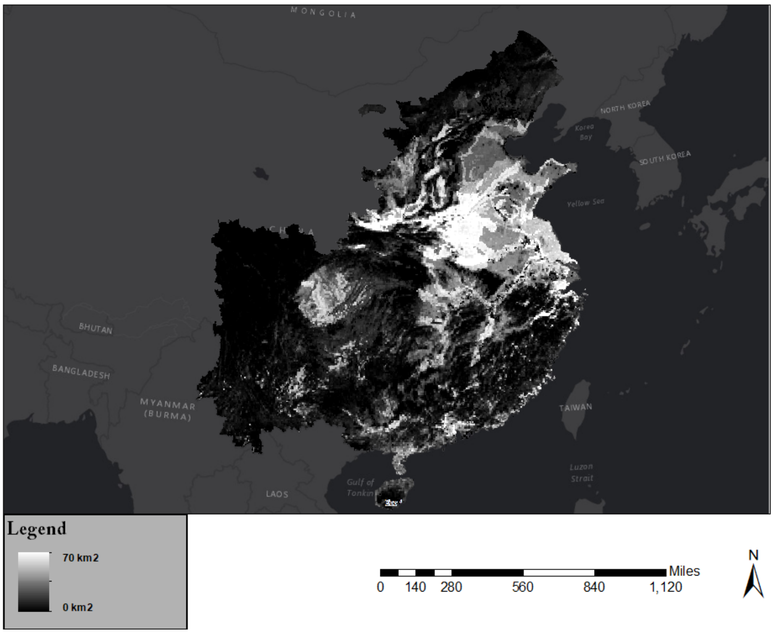


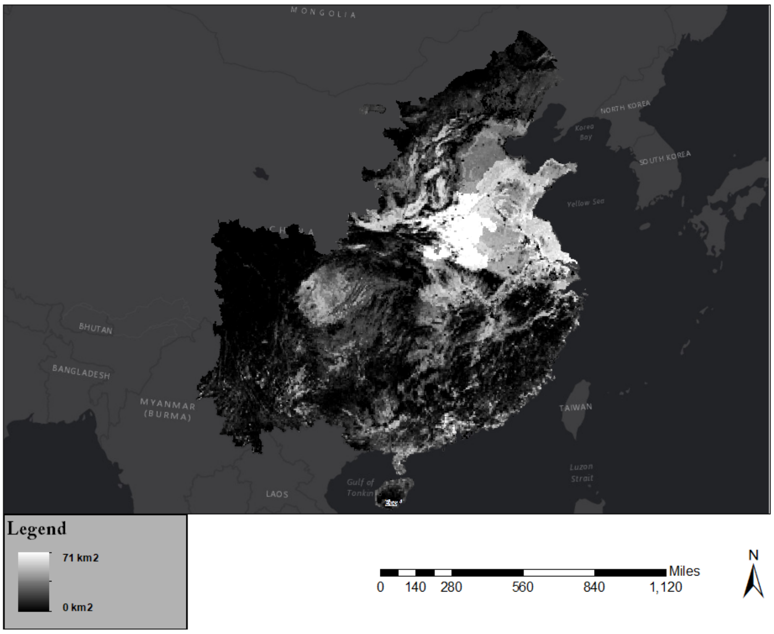
Figure 4.7: Distribution of the Deviation of Population Density from 1820 to 1911

Source: China Historical Geographic Information System, Version 6. Cambridge: Harvard Yenching Institute and Fudan Center for Historical Geography, January 2016

"Population history of China (in Chinese)"



(a) 1820



(b) 1911

Figure 4.8: Distribution of Population Density from 1820 to 1911

Source: HYDE 3.2.1

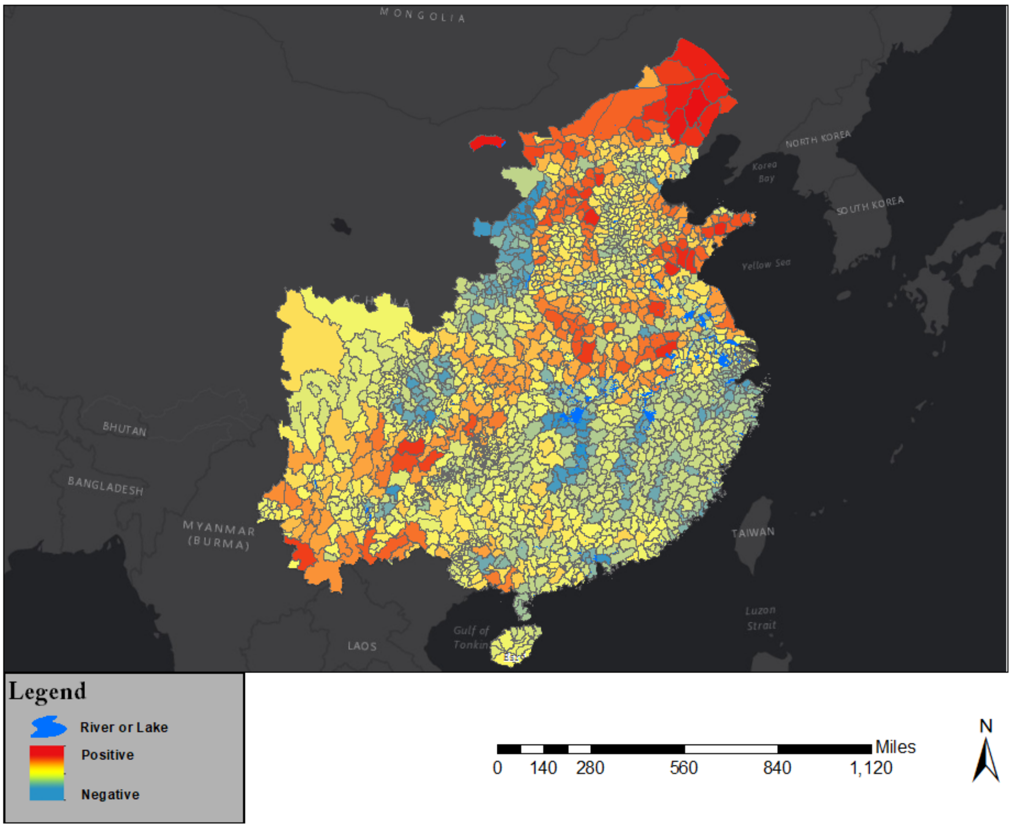


Figure 4.9: Distribution of the Reclamation Scale from 1820 to 1911

Source: HYDE 3.2.1

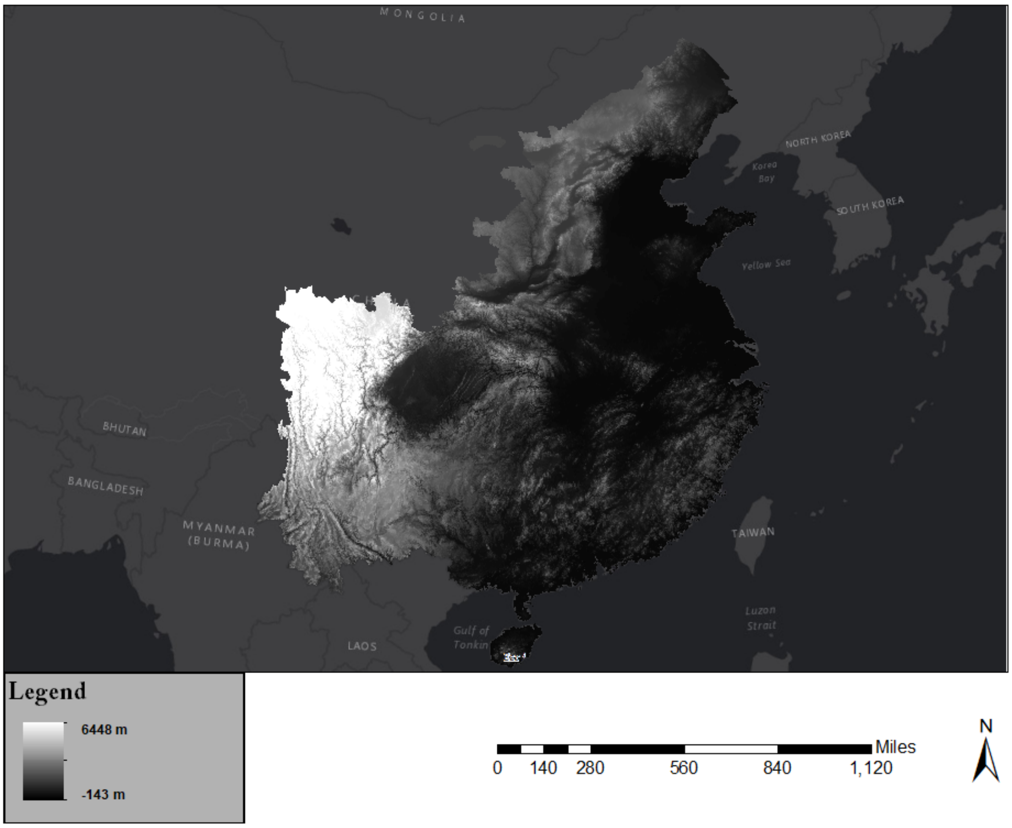


Figure 4.10: DEM of Research Region

Source: SRTM

Chapter Five

Re-estimate the Weather-conflict Relationship of China during the Ming and Qing dynasties

Rising concerns over the environment have motivated the emergence of an abundant literature on the impact of climate changes on human activities. (Carleton and Hsiang, 2016; Li, Ruan, and Ye, 2018; Gray and Mueller, 2012). Dramatic climatic variations have been identified as triggers or factors that exacerbate social instabilities (Dell, Jones, and Olken, 2014; Burke, Hsiang, and Miguel, 2015; Mach et al., 2019). The link between weather and conflict has also been investigated across different contexts, such as in Africa (Miguel, Satyanath, and Sergenti, 2004; Ciccone, 2011; Sandholt Jensen and Skrede Gleditsch, 2009), Europe (Tol and Wagner, 2010; Carleton and Hsiang, 2016; Anderson, Johnson, and Koyama, 2017), and Asia (Bai and Kung, 2011; Jia, 2014). Considering China's unique geographic diversity, long written history (Elvin and Liu, 1998; Elvin, 2006) and the evolution of social structures between the modern age and pre-industrial world, new estimations have been emerging to quantify the significance of adverse environmental shocks on social instabilities across China history(Fan, 2010; Yu, Dong, and Lei, 2015; Zhao, 2012). This research intends to refine previous estimations of the relation between adverse climatic events and the emergence of civil conflicts (Jia, 2014) by using data covering 1634 counties in the traditional agriculture region from 1368 to 1911. What were the exact impacts of different kinds of weather shocks on the occurrence of social instability (defined as conflicts and battles in this study)? Did different weather shocks imply different mechanisms? To what extent did the state capacity mitigate or exacerbate these impacts? To what extent did neighbouring conflicts and weather shocks affect local instability at the county level?

Thanks to China's extensive historical documents, we are able to collect data at the county level to test the connection between weather shocks and conflicts during the Ming and Qing dynasty. Specifically, conflict including specific battle, the assertion of rebellion, and other county records are approved to be the proxy of social instability (Zhao, 2012; Yu, Dong, and Lei, 2015). The geographic area for this research is constrained within the traditional sedentary agriculture regions (Pei, Zhang, and Lee, 2016), which excludes pastoral regions to make sure that the channel that transfers negative impact of weather shock on agriculture output remains similar. The time period analysed extended from 1368 to 1911, covering the whole Ming and Qing dynasty(Zheng, Zhang, and Zhou, 1993; Chen and Chen, 2016).Through our re-estimatation of the relationship between peasant revolts and weather shocks of previous work (Jia, 2014) at the county level, we address the concern of possible sampling bias from sample scales (Adams et al., 2018) affecting former studies. We find that drought increased the probability of conflict occurence but our results temper Jia' s conclusions that revolts were more likely to be triggered by droughts compared to floods. If set aside the deviation of data construction from different sources, one of the possible interpretations for deviations of our results is that prefectural data would merge a year with conflict but without disaster or with disaster but without conflict for some counties within a given prefecture into a single observation. As a result, conflicts occurrence irrelevant of weather conditions or spatial dependent conflicts could be connected with weather shocks, which biases the prefectural estimations.

Because of the unavailability of agriculture output and crop price at the county level data, this research does not strictly follow the previous estimation to test the direct connection between weather shocks and agriculture output (Jia, 2014; Sheng and Xu, 2019). However, it focuses on the individual impact of different weather shocks on revolts and internal conflicts. Specifically, the wetness/dryness grade used in Jia' work was constructed from a comprehensive flood and drought records. For example, if there were flood incidents and drought incidents in the same year, the grade would be converged in the direction to grade 3 (which represents an average level without abnormal precipitation) in Jia's work. However, the impact of flood and drought cannot be offset with each other, and normally drought and flood in the same year should be more disruptive (Elvin, 2006; Marks, 2004). As a result, this research combines weather data from two exhaustive sources and constructs county-level frequencies of droughts and floods to test the individual effect of each disaster. We find that flood and drought raise the occurrence of conflict incidents significantly however they did not affect revolts and internal conflicts differently.

This research includes the frequency of cold and wind as additional weather shocks, which have been seldom employed in previous works. Other factors that may mitigate or exacerbate the impact of weather shocks on the nexus conflict/income has also been considered in the recent literature. Jia (2014) mentioned that the adoption of sweet potatoes mitigated the impact of droughts, and Chen (2015) estimated that weak state capacity exacerbated the impact of abnormal weather. What is more, it is also possible to investigate the joint effects of different weather shocks on conflicts. If we accept the argument from previous work (Chen, 2015), the mitigation effect of sweet potato adoption can be captured by time controls since the effect was estimated according to the adoption timing. Therefore, our panel model, which employs county-specific time trends, is expected to eliminate possible time-related mitigation effects, such as the effect from sweet potato, in the main result. What is more, this research does not view the state capacity as linear increased as the governance strength should be weaker at the beginning and the end of the entire dynasty. Intuitively, the government would start to enhance power from the beginning of the dynasty, experience the peak of state capacity around the middle of the dynasty, and then crash at the end of the dynasty. Therefore, state capacity should have non-linear dynamics, which can be detected through a quadratic term. We confirm this intuition by adding a dynastic year and the quadratic term of the dynastic year.

We discussed several potential concerns arising from this research. One is related to the influence of external conflict. Due to the limited extend of the research region and research period, most of the external conflicts can be considered exogenous shocks, added as a control variable. Another concern comes from the annual analysis since the impact of weather shocks might be lagged to next year, and weather shocks outside cultivation season would impose a lower impact on agriculture output (Harari and Ferrara, 2018). We add the lag terms for weather shocks and external conflicts in order to capture possible lagged effects as well. As for factors introduced by other scholars, such as taxation (Ma and Rubin, 2017), economic development (Besley and Persson, 2009), population (Chen, 2015), cultural norms (Kung and Ma, 2014), market integration (Shiue and Keller, 2007) and land possession (Fang et al., 2019; Zou, 2013), we capture them by adding county fixed effect and time trend, since these factors usually have consistent long run time trends or are time-invariant. We did not concern too much about the possible endogeneity of environmental factors like floods, since we were simply in line with previous studies that set floods and droughts exogenous. It is also acceptable to set aside this potential issue due to inadequate direct evidence.

In addition, to the best of our knownledge, we are the first to consider the spatial spillover effect for the weather shock and conflict in historical China , since previous works have implemented analyses within relatively larger areas (the province or the prefecture level) where climate deviations were aggregated. Accordingly, climatic influences were unlikely to be altered significantly by neighbouring areas (Zheng, Ge, Hao, et al., 2014; Zhang, 1996). Motivated by Harari and Ferrara (2018), this research contributes to the literature on the estimation of spatial dependency of the link between weather of weather changes and conflicts shock and conflict in China. To be specific, Harari and Ferrara (2018) verified the critical spillover effect using grid-level data in Africa, which indicates the necessary consideration of spatial dependency when analysing data at a high resolution. Our specifications for examining spatial dependency are closely related to their work, focusing on the spatial dependency for both dependent and independent variables. In addition, this research can also relate to the literature on the spillover effect of conflict in other fields such as the spillover effect of conflict on the economic growth(De Groot, 2010), the potential conflict contagion in ethnically fractionalized countries (Schleussner et al., 2016), and the reason of conflict contagion (Buhaug and Gleditsch, 2008).

This research mainly contributes to the literature on the historical link between climate and social instability (Tol and Wagner, 2010; Carleton and Hsiang, 2016; Iyigun, Nunn, and Qian, 2017; Bai and Kung, 2011; Jia, 2014; Chen, 2015). Previous studies (Zhang, Jim, et al., 2006; Tol and Wagner, 2010; Zhao, 2012) have focused on time series analysis in the long run or panel analysis for a large area (at the country level or the prefecture level etc.), due to constraints related to data availability. For example, Carleton, Campbell, and Collard (2021) employed a Bayesian approach to estimate the impact of temperature from a time series model. Jia (2014) used a linear probability model using the prefecture year panel to confirm the more significant impact of drought. We attempt to construct a county-level dataset mobilizing a collection of historical documents and provide detailed estimations of the relation between climatic events and conflicts using county-year panel to avoid possible sampling bias (Adams et al., 2018). We control for multiple varieties of trended and time-invariant drivers of conflict.

Inspired from the historical climate-conflict relationship, this research could also contribute to estimations for modern cases. The most prominent study is the work of Miguel, Satyanath, and Sergenti (2004), which used the rise in precipitation and correspondent lags as instruments for GDP of sub-Saharan Africa after 1979. Several following works have re-estimated their results by employing new datasets and approaches (Sandholt Jensen and Skrede Gleditsch, 2009; Ciccone, 2011; Miguel and Satyanath, 2010; Abel et al., 2019). We involve the first order lags for weather shocks to respond to previous estimations' specification but cover a longer period over five centuries. We can also examine potential endogenous influences of conflict by add lag of the dependent variable, which again relates to Harari and Ferrara (2018).

Our analysis also relates to the literature on the determinants of conflict. For example, Besley and Persson (2009) proposed a range of explanation for conflict occurence. They then estimated that several factors, including GDP, parliamentary democracy, exports, weather shocks and import prices, matter in determining the states of peace, repression and conflict. Berman and Couttenier (2015) found that external shocks have played a role in conflict occurrence. (Kung and Ma,

2014) confirmed that Confucius norms would reduce conflict. Our analysis focuses more on weather shocks that are strongly exogenous in historical context (Zhang, Zhan, et al., 2004) and reconfirms the influences of other types of events rather than droughts, specifically in historical China.

The structure of the rest of this article is as follows: the first section reviews the relevant literature to illustrate the historical context and investigate former research on the weather-conflict relationship in historical China. The second section introduces the sources for data. The third section proposes an empirical strategy and presents the associated results. The fourth section employs some robustness checks, and the last section is the conclusion.

5.1 Historical Context

At the juncture of the Ming and Qing dynasties, various natural disasters and abnormal weather conditions were part of the Little Ice Age (Marks, 2004; Elvin and Liu, 1998). Historians have illustrated that such environmental shocks harmed agriculture output and then affected social instability. Meanwhile, economists have proposed two channels, decrease in relative returns of labour vs the return to fighting and reduction of the total available economic pie, through which economic shocks (such as negative shocks on agriculture output) increase the likelihood of conflicts. Consequently, economic agents would face a trade-off between gains from economic activities and benefits from participating conflicts (Harari and Ferrara, 2018; Collier and Hoeffler, 1998). In order to know how the economic theory works in the historical context of China, it is necessary to review the literature touching upon the measurement of conflict across time in the country and on the weather-conflict relationships.

5.1.1 Reconstructing the historical environment in Ming and Qing dynasties

The foundational work to standardise historic environmental records of China was established by Chu (1973), who matched historical records and modern meteorologic classification to look at climate variation (variation in temperature in particular) over the past 5000 years. Several criteria, such as the consistency of the distribution of the records of meteorological events across time (uniformity), were proposed. Following studies (Yang, Braeuning, et al., 2002; Ge, Wang, et al., 2007) focusing on temperature reconstruction confirmed the continuous cooling of the climate after the 14th century detected by Chu (1973).

Summarised in the collection of "The Chronology of Natural Disasters and Human Calamities in Chinese History (in Chinese)" (Chen, 1986), both the natural disasters and human calamities frequencies increased during the Ming and Qing dynasties, this increase being considered as the manifestation of the Little Ice Age. Specifically, human calamities include a variety of social instabilities, such as internal and external conflicts. The increasing trend of both categories might imply some possible connections between each other. However, since this collection did not identify the stylistic rules of the texts used as a source or the layout of the record and only covered selected parts of historical documents, the uniformity of the data-set cannot be ensured. As a result, it is possible that the increase in environmental records did not respond to the increase in weather shocks on average. Thus, frequency statistics might bias the picture of China's historic environment. For example, the geographic unit of disaster could be a county, prefecture, province or regional China, which indicates that one-time occurrence might represent different levels of disasters. Similarly, various description of disaster could represent discrete performance as well. What else is that the rise of disaster number might be the result of more sufficient records during Ming and Qing dynasties. Therefore, to demonstrate the environmental variation more precisely, a number of approaches have been developed to diminish potential bias (China Meteorological Administration Institute of Meteorology, 1981; Zheng, Zhang, and Zhou, 1993).

One approach that is commonly accepted was introduced in ”The Atlas of Wetness and Dryness Grade Distribution in China in the Past Five Hundred Years (in Chinese)” (China Meteorological Administration Institute of Meteorology, 1981). Descriptions of flood and drought were classified into five levels, from which grade 1 represents wet and grade 5 represents dry. This atlas presented yearly wetness/dryness distribution in separated maps from 1470 to 1979 across China based on the grade series from 120 stations. These maps can show the spatial variations of floods and droughts across time clearly, although the time trend cannot be displayed. A comparable approach, which sorts the level of descriptions of precipitation, was employed to reconstruct numerical precipitation series across the traditional agricultural region of historical China (Ge, Zheng, Hao, Zhang, et al., 2005). Ordering the description by analysing the extent of the disaster in the text was widely adopted to reconstruct some historical series (Zheng, Ge, Hao, et al., 2014; Xia, 2015) since historical records are complicated to quantify. As a result, meteorologists and paleometeorologists usually reconstructed historical temperature and precipitation according to actual observations (Zhang and Liu, 2002; Ge, Zheng, Hao, Zhang, et al., 2005) or some proxies like tree rings and ice core (Yang, Braeuning, et al., 2002; Ge, Wang, et al., 2007), which seldom exist for the Ming and Qing dynasty.

In contrast, another approach, which insists on the frequency of records, argued that ordering description would lose information about the severity of disasters implied by the number of records and would be biased by perspective errors (Zheng, Zhang, and Zhou, 1993). This approach was first employed to reconstruct the wetness/dryness grade using the number of flood/drought counties. Constrained at the prefecture level, the number of counties with flood or drought within each year was expected to capture the disaster severity and eliminate issues in uniformity. Validated by other estimations, based on actual observations of average rainfall, Zheng's reconstruction approach was popularised to the whole of eastern historical China, combining with other estimations (Hao, Ge, and Zheng, 2010).

According to Hao's work (Hao, Ge, and Zheng, 2010), precipitation situation was illustrated as wetness/dryness grade, which integrated grade reconstruction based on description ordering before 1470 (China Meteorological Administration Institute of Meteorology, 1981), grade reconstruction based on the percentage deviation from the mean of flood/drought counties after 1470 (Zheng, Zhang, and Zhou, 1993), and grade reclassification based on some actual observations from Qing archive for 1736 1911 (Zhang and Liu, 2002; Ge, Zheng, Hao, Zhang, et al., 2005). Therefore, it is possible to review the environmental characteristics of flood and drought across the Ming and Qing dynasties. Considering the features of historical records in which abnormal events were usually recorded, Hao, Ge, and Zheng (2010) estimated that historical records for flood and drought in the Ming and Qing dynasties cover over 90% of years from 1368 to 1911, which represents a high level of confidence and indicates good narration for weather fluctuation. Consequently, Hao, Ge, and Zheng (2010) identified that in North China, severe droughts were frequently occurred during the early 17th century, the late 19th century and the early 20th century, while severe floods came out during the late 17th century and the late 19th century. Comparably, the Jianghuai area experienced frequent droughts during the early 16th century, the early 17th century, the late 19th century and the early 20th century, and experienced frequent floods during the late 16th century, the early 18th century and the 19th century. What is more, the Jiangnan area saw the concentration of droughts during the late 15th century to the early 16th century, and the early 20th century, while went through frequent floods during the early 15th century, the early 19th century and the early 20th century. Across the whole of eastern China, total occurrences, including floods and droughts, reached the maximum magnitude around the end of the 19th century. In addition, the probability of flood/drought occurrence was identified to be at the highest level, which is over two times higher than the average level among the past 2000 years regardless of few exceptional periods in the first millennium. The conclusions above have provided some specific insights into the historical environment of the Ming and Qing dynasties. However, they have seldom discussed the dynamic distributions of other kinds of weather shocks and have omitted the influences of different disasters within the same year.

To further illustrate the ancient meteorology of China, Zhang (1998) employed unpublished ”Chinese Three Thousand years Meteorological Records Collection (in Chinese)” (Zhang, 2004b), which was evaluated as the most appreciated collection of meteorological records from historical documents (Xia, 2015), to map different historical weather and climate conditions including severe frost, drought, flood, typhoon, etc. Since Zhang's collection carefully diagnosed each disaster at the county level, which suggests persistent uniformity among all records, environmental variation can be detected in higher resolution. Accordingly, our analysis improves previous scales of environmental perspectives from the prefecture level to the county level by employing weather shock frequency from some well-organised historical documents.

5.1.2 Weather-conflict relationships in China

The relevance of the relation between weather changes and conflicts in China was recognised long before modern history. Kangxi, the emperor of the Qing dynasty in the early 18th century, already realised that cold weather declined the agriculture output (Elvin and Liu, 1998) and failure of appropriate relief to disaster would stimulate social instability (Pierre-Etienne Will, 2002; Deng, 2011). Several historians have claimed that weather-induced famine is the trigger of internal conflicts (Meng, 2017; Twitchett and Mote, 1998; Mote and Twitchett, 1988). Inspired by previous knowledge, Zhang, Zhan, et al. (2004) primarily estimated that war frequency was negatively related to temperature. It was claimed that 70% 80% of wars occurred during the cold period. From 850 to the end of the

Qing dynasty, almost 86% 100% of social unrests and around 54% 59% of dynastic changes happened during the cold periods. Additionally, the agriculture output in 1840 1890 (cold period) was 10% 25% less than that in 1730 1770 (warm period), which implies a sensitive reaction of agriculture output to cooling in historical China. However, it was also argued that different regions of China responded separately to cooling due to different crops cultivated, which is an interpretation reaffirmed by other scholars (Carleton, Campbell, and Collard, 2021; Pei, Zhang, and Lee, 2016). For example, the southern part of China maintained a variety of crops, which allowed people to choose different food to mitigate the impact of cooling. In contrast, the centre part of China was affected by the reduction of temperature more critically. Therefore, the correlation between weather and conflict varied across different climate zones. Differentiating for flood and drought frequency, Wang, Chen, et al. (2010) verified that the impact of climate on war differed geographically.

Zhang, Jim, et al. (2006) tested the correlations between ”average temperature anomalies and the number of wars in each decade”. He found a statistical significant relation between cooler temperature and conflict. Zhang, Jim, et al. (2006) supposed that increasing population pressure was exacerbated by temperature decline and then induced more participants for anti-tax rebellions or other types of conflicts. However, the temperature has been argued to have a weak connection with social instability due to the distribution of crops adapted to local climatic conditions (Carleton, Campbell, and Collard, 2021). As a result,Carleton, Campbell, and Collard (2021) has recommended to include other disasters or temperature-induced weather shocks such as flood and drought, which can directly impact agriculture output. Wang, Chen, et al. (2010) reconstructed the series of desertification and biological productivity of northern China and the Mongolian Plateau to identify the relationship between desertification cycles and Chinese dynastic cycles. The expansion of desertification seems to a significant extent to participate to the collapse of Chinese dynasties. Bai and Kung (2011) explore similar linkages focusing on the impact of precipitation on the Sino-Nomadic conflict . According to their results Bai and Kung (2011), adverse rainfall shocks had a significant and positive relationship with nomadic incursions, as the nomadic groups would attack sedentary groups due to the weather-induced lack of food.

Recalling Zhang's comments on the distribution diversity of crops and climate zones, Jia (2014) estimated the link between peasant revolts and flood/drought grade using data covering 267 prefectures from 1470 to 1900 and found that exceptional drought had a positive impact on the initiation of revolts and the effect of the drought was almost two times higher than the flood. Moreover, Jia (2014) identified the effect of the sweet potato adoption.The introduction of sweet potato reduced the probability of starvation when facing environmental shocks and then discouraged the peasant revolts. Following studies have looked at other mitigation effects such as state capacity (Chen, 2015). Government with strong capacity might be able to maintain logistics networks and infrastructure to release appropriate reliefs to communities impacted by adversarial weather shocks and suppress rebellions (Harari and Ferrara, 2018; Meng, 2017). There were charity granaries widely built at the county level in particular to release food to refugees and local people after disasters (Wong et al., 1991), but the effectiveness was argued to depend on officials capabilities or funding availability. While the regional and temporal discrepancy of conflict response to specific weather shocks (typically referring to floods and droughts or extreme temperature) has been widely investigated, Zhao (2012) showed also that different type of social instabilities have been affected diversely by varied types of climate conditions. Specifically, abnormal snow reduced the probability of internal disorders but raised the probabilities of external aggression in the short term, while flood has a positive impact on internal disorders but a negative effect on external aggression.

Therefore, this research combines knowledge from previous works by employing revolts and internal conflicts (contain revolts) as two different dependent variables and estimating the link between separate weather shocks (flood, drought, cold and wind) and conflict at the county level.

5.2 Data

5.2.1 Historical Administrative Unit and Geographic Information

The geographical area of interest in this research is made of 17 provinces based on the 1911 administrative boundaries shown in Figure 5.1. Red lines circle the boundaries of provinces, yellow lines circle the boundaries of prefectures, and white lines circle the boundaries of counties. Data was obtained from "China Historical Geographic Information System" (Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University,

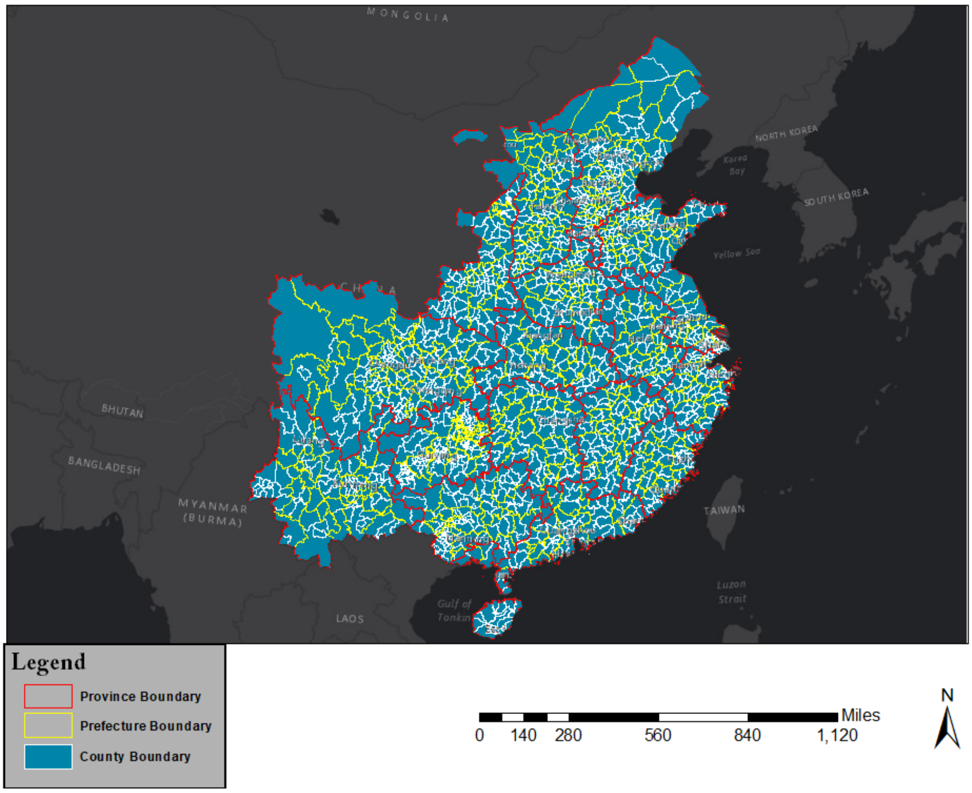


Figure 5.1: The Map of Administrative Boundaries of the Research Regions

Source: China Historical Geographic Information System, Version 6. Cambridge: Harvard Yenching Institute and Fudan Center for Historical Geography, January 2016

2016), which was established by Fudan University and Harvard University and was abbreviated as CHGIS. The CHGIS is open-source, and includes administrative information of historical China, which has been already used in academic research (Jia, 2014; Kung and Ma, 2014). Specifically, data for 1911 includes the complete collection of the names and boundaries of counties, prefectures and provinces.

According to the original data, there are some blanks in the county-level map, since some county-level states were governed as prefecture-level units and these states were not collected in the county-level map. Therefore, these prefecture-level counties were selected to fill the blanks in the county-level map. Coordinates information was generated based on the centre point of each county polygon employing Xi'an80 geodetic coordinates systems through ArcGIS version 10.6. Other historical records were merged if they were identified to locate in the same county boundary (some conflict records were georeferenced through the location of the town, village, etc.).

This research substitutes the agricultural suitability to the division of agricultural region proposed by the Resource and Environment Science and Data Center of Institute of Geographic Sciences and Natural Resources of CAS, as shown in Figure 5.2 (Institute of Geographic Sciences and Natural Resources of CAS, 2017). The division of agricultural regions was classified through different climate zones for which temperature and precipitation vary across different regions. This division can wipe-out possible deviations from different crops in different environmental zones and provide more information about the weather-land conditions.

5.2.2 Historical Conflicts

Conflict data is collected from a multi-volume book entitled "Chronology of Warfare in Dynastic China (in Chinese)" (Compilation Group of Chinese military History, 2003). Records for wars and conflicts are collected by year with the descriptions of time, battle locations and other relevant information such as the spread and result. Existing studies researching historical conflict in China have

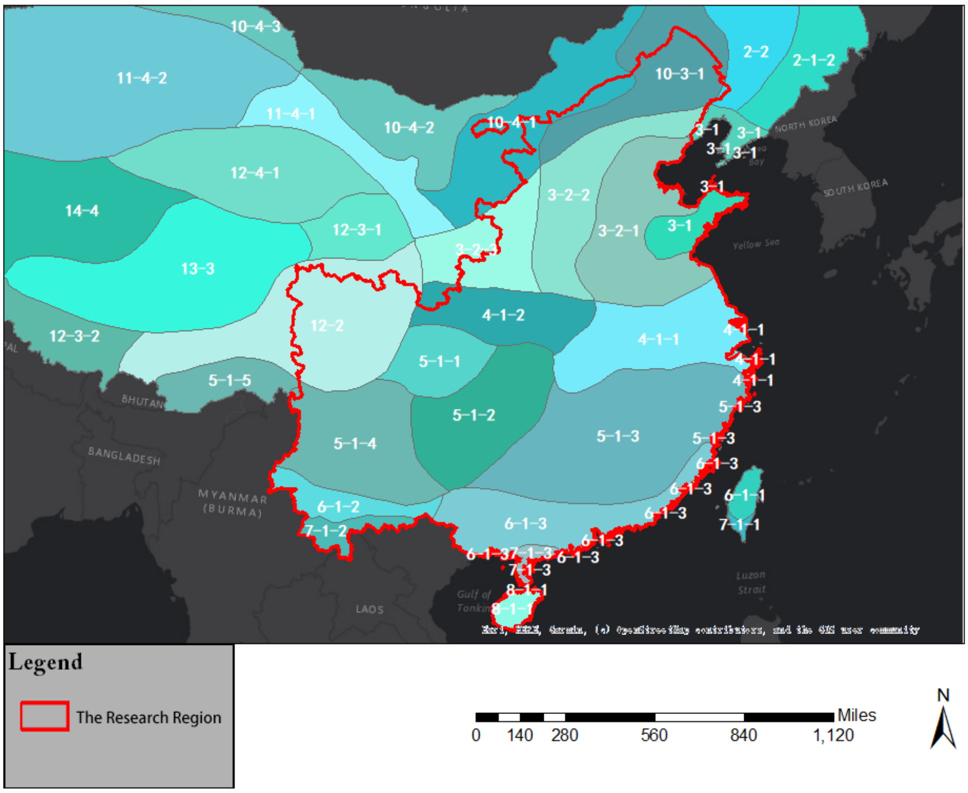


Figure 5.2: The Division of Agricultural Region

Source: The Resource and Environment Science and Data Center of Institute of Geographic Sciences and Natural Resources of CAS

universally employed this source, such as Zhang, Tian, et al. (2010) who looked at violent events or Jia (2014) who focus on revolts. Our research identified all the conflicts within the research region between the government and local forces as internal conflicts and conflicts between government or local people and political entities outside the research region as external conflicts. In addition , every record with the descriptions of ”revolt” or ”rebellion” with peasant participants was also coded as revolt.

There are several issues related to the identification of the locations of violent events, the types of events and their time, especially at the county level. Since the precise level of description of each record varies across the Ming and Qing dynasties,it is inevitable to lose or exaggerate some information for a few counties. However, as accepted by other historians (Zhang, 1996; Chen and Chen, 2016), the influence of a given event should be historically insignificant if there were fewer records. For example, the record for the Rebellion in Lipu county of Guangxi province in 1522 describes that ”···People in Lipu county of Guangxi province rebelled, attacked Guilin, Yangshuo etc., and killed officials. In November, the governor of Guangdong and Guangxi for the Ming government led their army to suppress them. Then, the rebellion failed”. According to the record, it can be known that the counties of Guilin and Yangshuo were attacked, which presents conflict incidents in these two counties. However, the actual situation in Lipu county and other counties around Guilin and Yangshuo is impossible to recognise (although it is simple to identify the prefecture). Therefore, we release the definition of conflict incident to include all the recorded counties affected by battles, unless the county could be easily found irrelevant to the conflict. In this regard, Lipu county, Guilin county and Yangshuo county in the example were counted as conflict incidents while other possible counties (not recorded) around Guilin and Yangshuo were ignored as they were considered unimportant. Accordingly, records can be expected to capture possible negative influence generated by battles nearby.

Other records contain similar information compared with the example, but the contagion and routine of a specific conflict can be more detailed. Most revolts were recorded in short sentences as they were normally suppressed very quickly. Consequently, a peasant revolt has been considered to have a smaller average scale than a modern civil war (*The COW Typology of War* 2021). Although battle-related deaths are not contained in the records, the severity of conflicts can be captured by more recorded counties. It can be confirmed to some extent by reviewing the records for some commonly known wars to check if more counties were recorded. Fortunately, wars, including the Anti-Japanese war during the mid-16th century, the Ming-Qing war during the mid-17th century, Taiping Rebellion during the mid-19th century etc., contain far more locations recorded than the rest of the conflicts. Therefore, it would be cautious to consider that all the recorded counties present the contagion range of conflicts. The range of conflicts can indicate the severity to a certain extent. Within the range, people were more likely to be affected by social instability, such as migrating to flee from the battles or participating in the rebellion (Ge, Wu, and Cao, 1997; Abel et al., 2019). Additionally , neighbouring counties are more likely to be recorded together in the same war according to the descriptions in the records.

Another concern comes from the sometime blurred records for location and time. Generally, conflicts are listed by year in the source, but several records contain descriptions for other years. Specifically, the example for a record in 1524 of the Rebellion of Cai Mengsan in Xinning is as follow: ”Peasants, including Cai Mengsan, from Xinning in Guangdong province, gathered a crowd to resist the brutal rule of the Ming dynasty, and his activities grew to over 10000 people in a few years. In March 1524, ···”. As described in the record, it is hard to know when the rebellion was actually initiated and how many years it took to grow to the scale described. As a result, we count the conflict in the year where the record is listed when descriptions of the time were not certain. Additionally, there are also records containing historical location names at a lower or upper administrative level than the county or specific geographical locations, such as some stockades in Yunnan and Sichuan provinces, rivers, mountains and some historical passes (Shanhaiguan pass, for example). We try our best to match these locations into the nearest county boundaries. Alternatively, we identify a conflict as an external war if one of the participants did not regularly apply for the central government of the Ming or Qing to recognise the legitimacy of local governance. We also identify a peasant revolt if one of the participants was led by peasants or largely constituted of peasants.

5.2.3 Historical Weather Shocks

Historical weather shocks are collected from the"Chinese Three Thousand Years Meteorological Records Collections (in Chinese)" (Zhang, 2004b) and the "China Meteorological Disaster Dictionary (in Chinese)" (Wen and Ding, 2008). The former has been accepted as the best collection of official meteorological history (Xia, 2015) and is the main source of this research. The latter is a more recent collections, and involves a larger number of sources (private journals for exemple). Data of the both sources are combined at the county level. We obtain the frequency of meteorological events after eliminating possible overlapping and checking the potential credibility of the supplementary source. Specifically, frequencies are merged into four categories, including flood, drought, cold and wind.

"Chinese Three Thousand Years Meteorological Record Collections (in Chinese)" is a book published in 2004, which collected records of weather condition, climate status, atmospheric physical phenomenon across entire China covering the period from 1300s B.C. to 1911 A.D. In particular, records for the Ming and Qing dynasties were compiled in volume 2, volume 3 and volume 4. All these records were drawned from official historical documents, including 8228 types of Chinese official history archives (the total number of Chinese official histories was estimated to be around 10 thousand). Around 7713 local gazetteers and 28 local biographies were employed to construct the environmental chronicle. The data were organised provincially and listed by year. Every record was identified with multiple sources to cautiously confirm that the records were not repeatedly collected and were precisely referenced. In addition, the location of every record was carefully matched with the modern county, which facilitates our work to identify the recorded location on the map. This data contains records for flood, drought, rainfall, snow, cold, warm, ice, freeze, frost, hail, wind, haze, storm, surge, lightning, etc., with time and location information. We identify 233 words corresponding to the description of flood or water lodge, 59 words for drought, 62 words for cold and 42 words for wind. Any record which contains one of these words is classified into the correspondent category. Therefore, for example, if a county had two records identified as flood in a given year, the flood occurrence for this county in this given year should be 2.

In addition, concerns for uniformity of data raised in previous works (Chu, 1973; Deng, 2011; Xia, 2015) can be avoided through "Chinese Three Thousand Years Meteorological Record Collections (in Chinese)", since this collection only included records with the same stylistic rules and layout or at least the style can be converted into the same one. Accordingly, records outside official histories were not considered, which may lose information or miss actual weather variations. This research selects ”China Meteorological Disaster Dictionary (in Chinese)” as a supplementary source to capture possible lost information outside official histories.

”China Meteorological Disaster Dictionary (in Chinese)” is a national scale work for which the compilation lasted from 2001 to 2008. The entire data contain 31 volumes (30 provincial volumes and one comprehensive volume). It is claimed that this data collected records from various historical documents that include official history and some informal documents such as diaries. Therefore, we employ this data in order to expect to fill some possible missing. All records included in these volumes were categorised into different meteorological types and then listed in years. Specifically, we merge records from categories including the word ”drought” or ”high temperature” as drought records. Flood records contain all the categories, including the word ”rainstorm”, ”flood”, or ”water lodge”. Cold records contain categories including ”cold” or ”snow”. Wind involves categories including ”wind”, ”hurricane”, ”typhoon” or ”tornado”. Information of location can be identified to the county level for the majority of cases but would contain some confusing location names which are hard to identify. As a result, confusing location names without any references are omitted in a cautious manner. Since each provincial volume was compiled by local, provincial institutions, the quality of this work varies across different volumes. Particularly, the criteria for each category was not claimed to be consistent across different provinces, and the rules of reference diverged either. We have estimated the quality of each volume according to the compiling time, pages, references, and the comparison with our main source and have assigned some credibility weights on the data to construct the full data set. Thus, for each category, the numbers of recorded counties are counted as occurrences for each year, and are aggregated to the main data set after multiplying credibility weights.

Unlike the previous parametrisation to the wetness/dryness grade (Zheng, Zhang, and Zhou, 1993; Hao, Ge, and Zheng, 2010), this research directly applies the weighted frequency mentioned above. Specifically, to construct the wetness/dryness grade, none record in three consecutive years was considered missing for a prefecture. Therefore, previous research (Zheng, Zhang, and Zhou, 1993) found that only a few missing existed at the prefecture level, which could be interpolated through Chebyshev polynomials (Zheng, Zhang, and Zhou, 1993; Hao, Ge, and Zheng, 2010). However, the number of zero records increases massively when the data is separated into the county level, which would violate former interpolation methods. More specifically, prevailed wetness/dryness grade construction approach, which is based on the quantitative records, has employed the number of flood/drought counties per prefecture per year. Apparently, deviations across different counties would be omitted to some extent, which leads to significant biases, especially if the number of flood/drought counties were adopted at the county level.

One possible concern arises from the definition of disaster’ s frequency according to the weighted counts of event counties in a year. As illustrated by historians (Zhang, 1996), the traits of historical records are that the majority of records refer to abnormal events. Therefore, people in history would likely record a weather shock if they were significantly affected (Xia, 2015). In this regard, the weighted count is not identical to the disaster’ s frequency, but represents the count of observed impact. Accordingly, records for different disaster categories were not based on the same standard, suggesting that records for different categories might need to be parametrised separately. As a result, this research employs the weighted frequency at the county level for each disaster category.

5.2.4 Data Description for Main Variables

Table 5.1 reports the descriptive statistics for main variables at the prefecture level. In the prefecture-year data, similar to Jia’ s definition (Jia, 2014), multiple battles within a year are counted as one conflict for a prefecture. Therefore, there were 2767 wars in total within the research region from 1368 to 1911, in which 2193 conflicts were internal wars and 574 were external wars. Specifically, there were 1472 revolt prefecture-years, hence 0.9984% of prefecture-years contain a peasant revolt.

The number of revolts is over five times higher than the data employed in Jia’ s work, which cannot be explained as collection errors. Since the conflict data in this research is from the same source of Jia’ s work but is merged from the county level, Jia’ s work might lose some information. To be specific, the revolt data in Jia’ s indicated the initiation of the revolt more (whether there was conflict or not in a prefecture in a year). What is more, although only a few prefecture-years contained more than one records, there were over 20% prefecture-years for the revolt, as shown in Figure 5.3.

Table 5.1: Descriptive Statistics for Prefecture by Year

	Obs.	Mean	SD	Min	Max
Revolts×100	147968	0.9948	9.92	0	1
Internal Wars×100	147968	1.4821	12.08	0	1
External Wars×100	147968	0.3927	6.25	0	1
All Wars×100	147968	1.8700	13.55	0	1
Exceptional drought	147968	0.0352	0.18	0	1
Exceptional flood	147968	0.0226	0.15	0	1
Limited drought	147968	0.0656	0.25	0	1
Limited flood	147968	0.1375	0.34	0	1
Drought	147968	0.2017	0.74	0	23.1132
Flood	147968	0.4458	1.11	0	37.8173
Cold	147968	0.2082	0.89	0	20.1271
Wind	147968	0.3710	1.31	0	36.2167

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

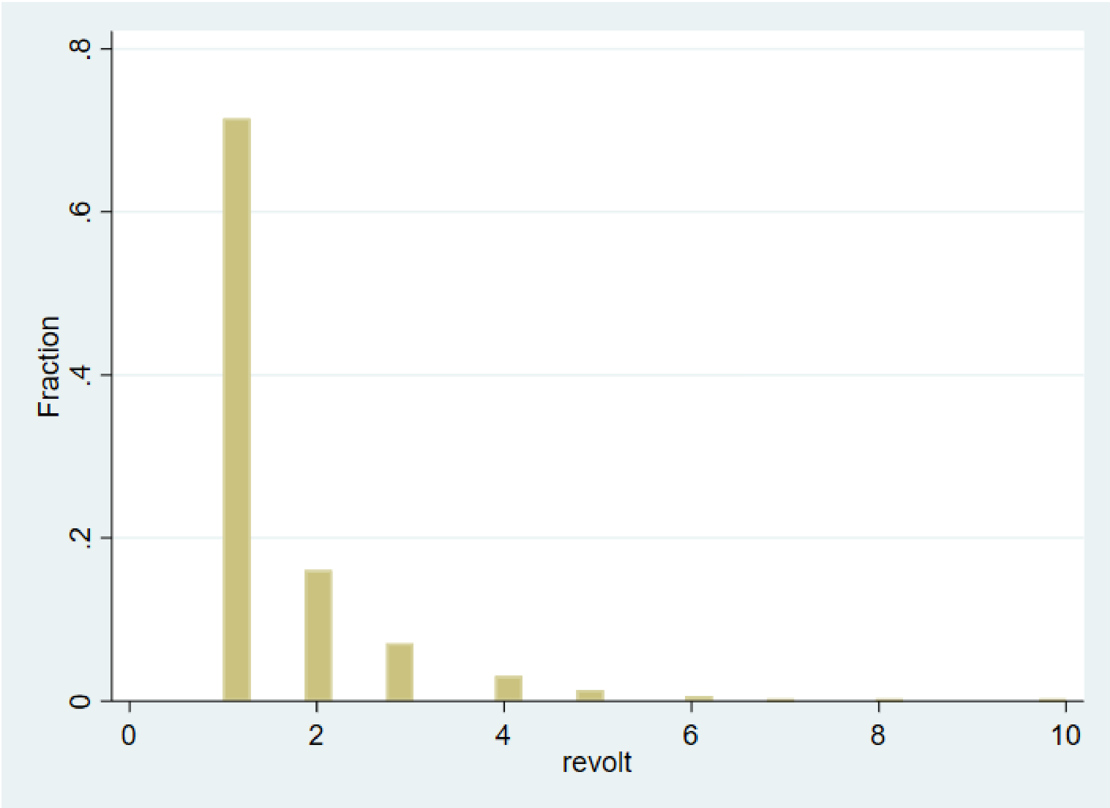


Figure 5.3: Distribution of Revolt Records by Prefecture

The weighted count for prefecture by year for each disaster category is illustrated in Table 5.1 as well. There were 0.4458 floods, 0.2017 droughts, 0.2082 cold and 0.3710 wind on average across all prefecture-years. The magnitude of the weighted county or the weighted frequency for flood is significantly higher than other weather shocks, which coincides with recognitions from historians in which floods dominated East China and Southeast China (Xia, 2010). However, when the weighted frequencies are converted to the wetness/dryness grade (in which grade 1 to 5 represent exceptional floods, limited floods, normal, limited drought and exceptional drought, respectively) through the method introduced by previous studies (Zheng, Zhang, and Zhou, 1993; Hao, Ge, and Zheng, 2010), distributions of floods and droughts are not symmetric as described in previous data set (China Meteorological Administration Institute of Meteorology, 1981; Jia, 2014), which indicates that perspective errors indeed existed in the grade constructed from description ordering (Zheng, Zhang, and Zhou, 1993; Zhang, 1996). The details of distribution in prefecture-year data are the following 3.52% of prefecture-years are identified as exceptional droughts, 6.56% as limited droughts, 2.26% as exceptional floods and 13.75% as limited floods. Accordingly, there were more exceptional droughts but less limited droughts compared with exceptional floods and limited floods, respectively. In addition, there were 1.6393% of drought prefecture-years and 1.6567% of flood prefecture years containing a peasant revolt. Therefore, a peasant revolt would be around two times more likely in a flood year or a drought year.

Table 5.2 presents summaries for the main variables at the county level, where there were 888896 county-years in total. There were 0.4210% of county-years containing a conflict. 0.3409% of county-years included an internal war, and 0.0807% of county-years included an external war. The proportion of county-years that contained a peasant revolt is around one-fourth of that in prefecture-year data (0.2347%). There were more proportion of county-years without any record of the revolt.

Similarly, the mean of the weighted frequency scales down at the county level as well. The average frequencies of the flood, the drought, the cold and the wind are 0.0742, 0.0336, 0.0347, 0.0618, respectively. The most frequent weather shock was still floods, but the mean is only around one-sixth of the statistics in prefecture-year

Table 5.2: Descriptive Statistics for County by Year

	Obs.	Mean	SD	Min	Max
Revolts×100	888896	0.2347	4.84	0	1
Internal Wars×100	888896	0.3409	5.83	0	1
External Wars×100	888896	0.0807	2.84	0	1
All Wars×100	888896	0.4210	6.47	0	1
Exceptional drought	888896	0.0131	0.11	0	1
Exceptional flood	888896	0.0072	0.08	0	1
Limited drought	888896	0.0202	0.14	0	1
Limited flood	888896	0.0472	0.21	0	1
Drought	888896	0.0336	0.19	0	5.6589
Flood	888896	0.0742	0.28	0	9.3836
Cold	888896	0.0347	0.23	0	8.7925
Wind	888896	0.0618	0.32	0	18.0570

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

data. In consequence, it can be realised that some county-years with conflicts but without weather shocks are merged with some county-years with weather shocks at the prefecture level, which would yield incorrect connections between the conflict and the weather shock. Additionally, over 90% of county-years with conflict records only contained one battle, as illustrated in Figure 5.4, which implies that dummies of whether there was a conflict (or a peasant revolt) in a county-year would remain fewer biases at the county level. In short, the relationship between the conflict and the weather shock revealed at the county level should be more precise. Table 5.2 also presents relevant statistics for reconstructed wetness/dryness grade. Specifically, 1.31% of county-years are identified as exceptional droughts, 2.02% as limited droughts, 0.72% as exceptional floods and 4.72% as limited floods. In this regard, 91.23% of county-years are indicated as normal weather, which should account for around 40% theoretically (Zhang, 1996). Therefore, the construction of the wetness/dryness grade may violate the efficiency of data at the county level.

According to Table 5.3 to Table 5.5, the proportions of the revolt and the internal war are 0.4398% and 0.6067% in drought county-years respectively, are 0.4576% and 0.6121% in flood county-years and are 0.4664% and 0.6227% in county-years with multiple weather shocks respectively. The probability of a revolt or an internal conflict is about two times higher when floods or droughts occurred. In addition, the conflict is slightly more likely to be triggered if there were multiple weather shocks in the same county in the same year.

According to Figure 5.5 and Figure 5.6, both conflicts and weather shocks are not homogeneously distributed across counties. When the external wars are excluded,

Table 5.3: Conflict Summaries for County-Years with Drought

	Obs.	Mean	SD	Min	Max
Revolts×100	31151	0.4398	6.62	0	1
Internal Wars×100	31151	0.6067	7.77	0	1
External Wars×100	31151	0.1252	3.54	0	1
All Wars×100	31151	0.7319	8.52	0	1

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5.4: Conflict Summaries for County-Years with Flood

	Obs.	Mean	SD	Min	Max
Revolts×100	67315	0.4576	6.75	0	1
Internal Wars×100	67315	0.6121	7.80	0	1
External Wars×100	67315	0.1189	3.45	0	1
All Wars×100	67315	0.7294	8.51	0	1

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

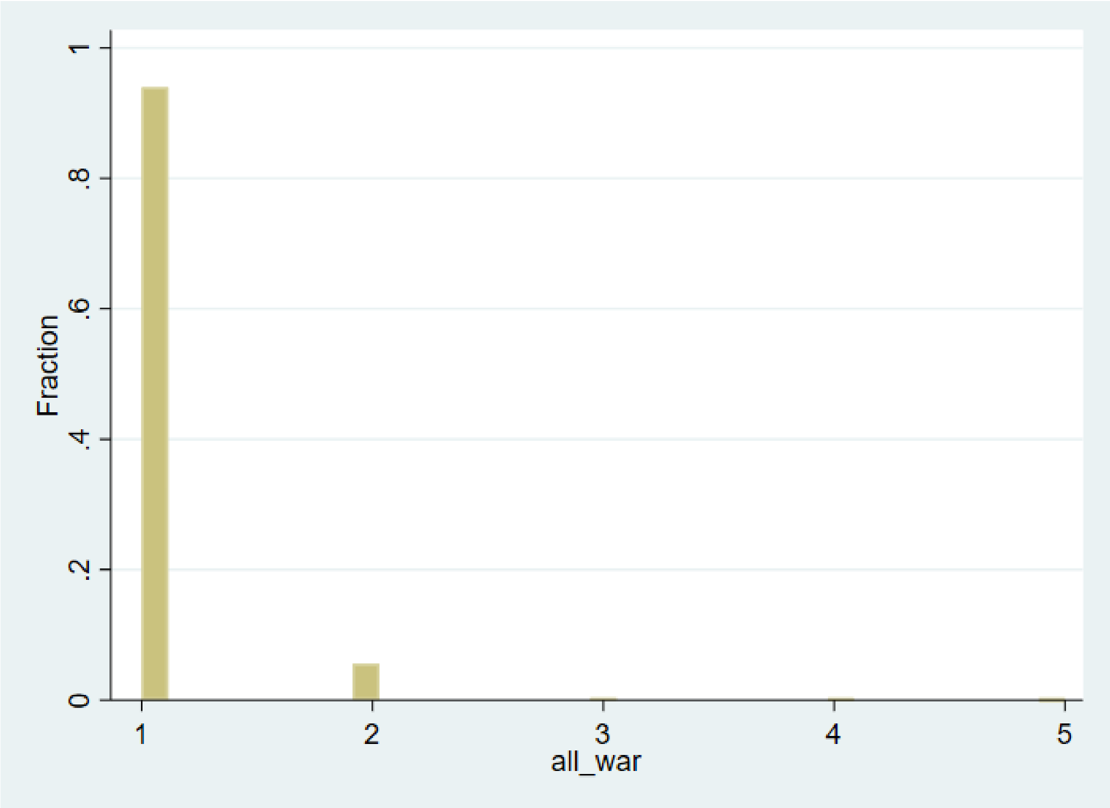


Figure 5.4: Distribution of Conflict Records by County

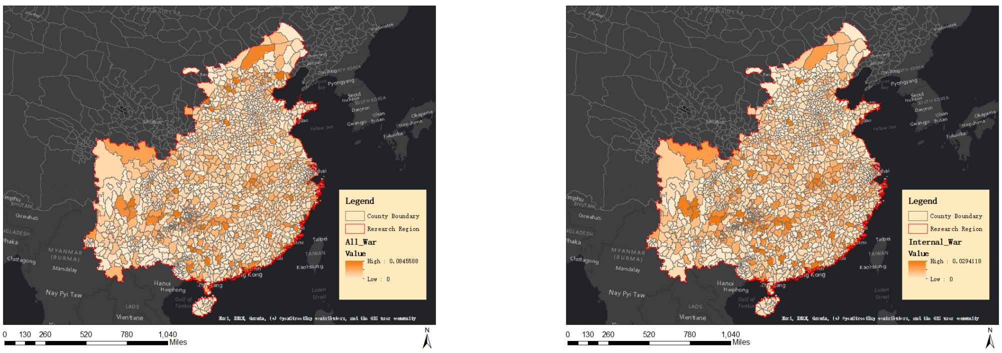
it can be observed that internal wars were clustered in the middle part of East China and Southwest China, where there were outbreaks of several crucial rebellions, including the rebellions of Sichuan chieftains and the Taping Rebellions. Furthermore, most of the flood and drought records were concentrated in the lower branch of the Yangtze River and the Yellow River. Especially between the Yellow River and the Huai River in East China, river courses changed periodically during the Ming and Qing dynasties (Ge, Wu, and Cao, 1997), which are associated with periodic floods and droughts.

Table 5.5: Conflict Summaries for County-Years with Multiple Weather Shocks

	Obs.	Mean	SD	Min	Max
Revolts×100	47378	0.4664	6.81	0	1
Internal Wars×100	47378	0.6227	7.87	0	1
External Wars×100	47378	0.1372	3.70	0	1
All Wars×100	47378	0.7599	8.68	0	1

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$



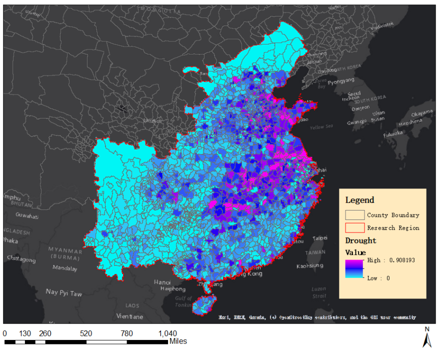
(a) All War

(b) Internal War

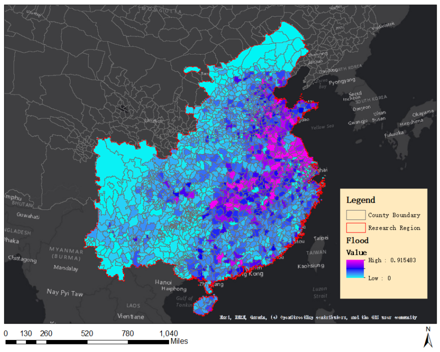
Figure 5.5: Spatial Distribution of Conflicts

5.3 Estimation Strategy and Main Results

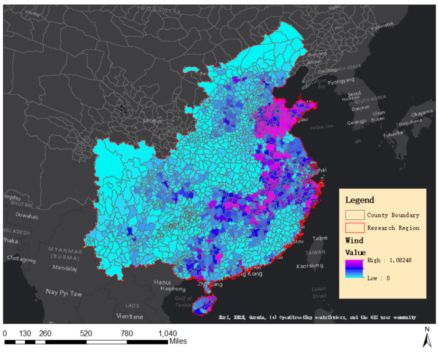
This research reproduces Jia' s estimation for the relationship between conflicts and weather shocks using a similar specification at the county level. We then discuss the role of state capacity and joint impact of multiple types of weather shocks. In the end, we test possible spatial spillover effects of weather shocks on



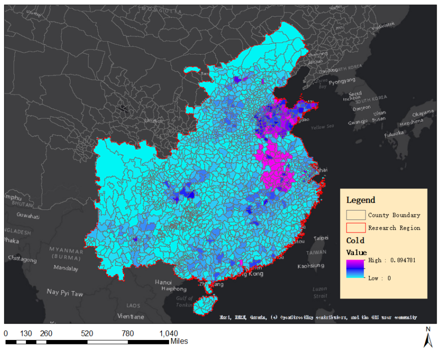
(a) Drought



(b) Flood



(c) Wind



(d) Cold

Figure 5.6: Spatial Distribution of Weather Shocks

conflicts employing spatial panel models.

5.3.1 The Impact of Weather shocks on Conflicts

We employ a similar linear-probability model as Jia at the county level to link the conflict to the weather shocks:

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 F_{it} + \beta_3 LD_{it} + \beta_4 LF_{it} + \delta \boldsymbol{X}_i \tau + \alpha_i + \gamma_t + \epsilon_{it} \tag{5.1}$$

where Y_{it} is a dummy for the occurrence of conflict in county i and year t. D_{it} and F_{it} represent dummies for exceptional droughts and exceptional floods, respectively. Similarly, LD_{it} and LF_{it} represent dummies for limited droughts and limited floods, respectively. $\boldsymbol{X}_i$ indicates a vector of time-invariant controls including longitude and latitude, county area, average value of digital terrain elevation model (DEM), county' s terrain ruggedness indicator (TRI) and dummies for division of the agricultural region (as illustrated in Figure 5.2). Therefore, $\boldsymbol{X}_i \tau$ denotes controls time time trends to exclude the impact from spatially diverse time trends. In addition, it is also possible to substitute the $\boldsymbol{X}_i \tau$ to a county-specific time trend ($County_i \tau$) to exploit more flexible estimations. α_i and γ_t denote county fixed effects and year fixed effects. ϵ_{it} is the error term. All estimations are clustered at the county level as default. In order to facilitate the interpretation, dependent variables are multiplied by 100 to convert coefficients into percentage point.

Table 5.6 shows the results of the estimations for the first specification. Following previous research (Jia, 2014), the revolt is selected to be the dependent variable. Column (1) is the result through ordinary least squares (OLS) estimations without any fixed effects controls. Column (2) shows the result with county and year fixed effects. Column (3) to column (6) attempt to control different levels of spatial varied time trends. Estimations become stable when fixed effects and spatial varied time trend are controlled. Coefficients of exceptional droughts and floods are not significant across column (2) to column (6), which differ from Jia' s work.

However, results for limited droughts and floods supported Jia's conclusions that "the impact of droughts on revolts is more than twice that of floods", and such finding "is consistent with qualitative historical studies" (Jia, 2014). Let us look at the most flexible model in column (6), where the county-specific time trend is controlled. The existence of limited droughts increases the likelihood of revolts by 0.220%. However, the limited floods only increase revolts by 0.0904% on average in a county within a year, at 1% significant level.

Following the same consideration as in Jia's work, this research includes the lag of weather shock in Table 5.7. column (1) and column (3) control interactions between the control variables and time trend, while the rest of the columns contain county-specific time trend. Column (3) and column (4) include interaction terms between weather shocks and their lags. The results show some consistency across different models. Limited droughts and floods are still significantly positive, but none of the lagged weather shocks has a significant coefficient. The interaction between exceptional floods and one order lag of exceptional floods is significantly negative (-0.00565) at 1% level in column (3) and column (4). Therefore, the exceptional flood is not independent of the previous exceptional flood, and the impact of floods on revolts can be underestimated. The differences between our estimations and Jia's work could come from the issues of data mentioned in previous sections.

Coming back to the discussions of the previous sections, there are several sources of discrepancies between Jia's work and ours: slightly different definitions for revolts, possible sampling biases between the prefecture level and county level data and different constructions for wetness/dryness grades. To deal with the concerns

Table 5.6: The Impact of Weather Shocks on Revolts

	(1)	(2)	(3)	(4)	(5)	(6)
	revolt	revolt	revolt	revolt	revolt	revolt
ED	0.0710	0.0287	0.0258	0.0275	0.0256	0.0285
	(1.44)	(0.59)	(0.53)	(0.56)	(0.52)	(0.59)
EF	0.202***	0.0932	0.0899	0.0897	0.0782	0.0861
	(2.69)	(1.25)	(1.20)	(1.20)	(1.04)	(1.15)
LD	0.292***	0.223***	0.222***	0.224***	0.222***	0.220***
	(5.24)	(4.13)	(4.11)	(4.13)	(4.12)	(4.07)
LF	0.192***	0.102***	0.0951***	0.0966***	0.0967***	0.0904***
	(5.66)	(3.07)	(2.88)	(2.92)	(2.94)	(2.75)
Control Year	No	No	Yes	No	No	No
Prefecture Year	No	No	No	No	Yes	No
Province Year	No	No	No	Yes	No	No
County Year	No	No	No	No	No	Yes
County and Year FE	No	Yes	Yes	Yes	Yes	Yes
Observations	888896	888896	888896	888896	888896	888896
R^2	0.000	0.015	0.015	0.015	0.016	0.018
Adjusted R^2	0.000	0.014	0.014	0.014	0.015	0.016
F	15.62	3.583	3.581	3.763	.	.

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5.7: The Impact of Lagged Weather Shocks on Revolts

	(1) revolt	(2) revolt	(3) revolt	(4) revolt
ED	0.0191	0.0218	0.0394	0.0426
	(0.39)	(0.45)	(0.72)	(0.78)
EF	0.0892	0.0859	0.147*	0.145*
	(1.18)	(1.14)	(1.75)	(1.72)
LD	0.219***	0.217***	0.189***	0.188***
	(4.08)	(4.05)	(3.39)	(3.37)
LF	0.0980***	0.0934***	0.113***	0.107***
	(2.97)	(2.85)	(3.13)	(3.00)
L.ED	0.0774	0.0800	0.0585	0.0631
	(1.28)	(1.34)	(0.95)	(1.03)
L.EF	0.00130	-0.00183	0.0687	0.0674
	(0.02)	(-0.02)	(0.85)	(0.85)
L.LD	0.0257	0.0238	0.0223	0.0201
	(0.51)	(0.47)	(0.42)	(0.38)
L.LF	-0.0350	-0.0396	-0.0215	-0.0269
	(-1.19)	(-1.32)	(-0.66)	(-0.82)
EDXlED			-0.101	-0.114
			(-0.68)	(-0.77)
EDXlLD			-0.149	-0.133
			(-0.73)	(-0.65)
LDXlED			0.591	0.582
			(1.29)	(1.27)
LDXlLD			0.0609	0.0587
			(0.41)	(0.40)
EFXlEF			-0.565***	-0.565***
			(-4.12)	(-3.97)
EFXlLF			-0.271	-0.289
			(-0.77)	(-0.82)
LFXlEF			-0.413	-0.453
			(-1.39)	(-1.51)
LFXlLF			-0.0611	-0.0569
			(-0.73)	(-0.68)
Constant	-2.086	-0.957***	-2.067	-0.956***
	(-0.75)	(-69.06)	(-0.74)	(-68.96)
Control Year	Yes	No	Yes	No
County Year	No	Yes	No	Yes
County and Year FE	Yes	Yes	Yes	Yes
Observations	887262	887262	887262	887262
R^2	0.015	0.018	0.015	0.018
Adjusted R^2	0.014	0.016	0.014	0.016
F	3.611	.	3.634	.

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

above, we employ another linear-probability model using the weighted frequency of weather shocks directly, as shown below:

$$Y_{it} = \beta_0 + \beta_1 Drought_{it} + \beta_2 Flood_{it} + \delta \boldsymbol{X}_i \tau + \alpha_i + \gamma_t + \epsilon_{it} \tag{5.2}$$

where Y_{it} is the dummy for conflicts in county i and year t. $Drought_{it}$ and $Flood_{it}$ represent weighted frequencies of droughts and floods. The remaining variables stay the same as the first specification. Revolts, internal wars, and all wars are successively used as dependent variables to verify whether the impact of climatic events depends on the definition of conflict types. Moreover, it is also possible to find a general linkage between weather shocks and conflicts.

Table 5.8 shows our estimations on different types of conflicts. Columns 1 and 2 estimate the weather impact on revolts. Columns 3 and 4 illustrate the results for internal wars, and columns 5 and 6 present the estimations for all wars. For each type of conflict, models are estimated with county-specific time trend and time-invariant controls specific time trend separately, since these two models can control most of possible unobserved trends based on geographical features or administrative units, such as population increase, taxation growth (Ma and Rubin, 2017), economic development (Besley and Persson, 2009), population (Chen, 2015), increase of market integration (Shiue and Keller, 2007) and increase of land reclamation (Fang et al., 2019; Zou, 2013).

All columns indicate that droughts have a higher impact on conflicts, but the difference between droughts and floods is less than 50% of the higher one. Column

Table 5.8: The Impact of Weather Shocks on Conflicts

	(1)	(2)	(3)	(4)	(5)	(6)
	revolt	revolt	inter_war	inter_war	all_war	all_war
drought	0.0785**	0.0807**	0.100**	0.100**	0.118**	0.115**
	(2.08)	(2.14)	(2.26)	(2.26)	(2.05)	(2.04)
flood	0.0578**	0.0528**	0.0728**	0.0716**	0.0864**	0.0822**
	(2.25)	(2.05)	(2.30)	(2.25)	(2.52)	(2.39)
Control Year	Yes	No	Yes	No	Yes	No
County Year	No	Yes	No	Yes	No	Yes
County and Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	888896	888896	888896	888896	888896	888896
R^2	0.015	0.018	0.015	0.017	0.013	0.016
Adjusted R^2	0.014	0.016	0.014	0.015	0.013	0.014
F	3.654	.	4.363	.	4.823	.

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

(2) reports that one more drought record increases the occurrence of revolts by 0.0807%, and one more flood record increases the occurrence of revolts by 0.0528%. Both coefficients are significant at 5% level. The magnitude of effects of floods is around 65.4% of that of droughts. The difference between coefficients of droughts and floods is reduced compared with the results in Table 5.6. The difference is further scaled-down when we look at the impact on internal wars and all wars. In column (4), one more drought record significantly increases the likelihood of internal wars by 0.1% at 5% level. One more flood record increases the number of county-year with internal wars by 0.0716% at 5% significance level. The effect of floods is accounted for 71.6% of that of droughts. It is natural that the magnitude of coefficients increases in model 4 compared with model 2, since the number of internal wars is higher than that of revolts. However, the increase in the coefficient of floods is greater than that of droughts indicating that there should be more flood-related conflicts.

Table 5.9 reports the results for the models with lags. For brevity, only the models with county-specific time trend are included. Columns 2, 4, 6 contain interaction terms of weather shocks and their lags and indicated that weather shocks and their lags are independent. Therefore, results in columns 1, 3, 5 are validated. Previous floods and droughts are not significant across all models if a large number of observations were considered. From columns 1, 3, and 5, floods have on average around 74% of the effect of droughts on conflicts .

In summary, droughts would be more likely to trigger a conflict in a county within a year, but floods have around 65% to 75% of effects of droughts. Our results are in line with Jia's. Our estimations, nonetheless, suggest that the effects of

Table 5.9: The Impact of Lagged Weather Shocks on Conflicts

	(1)	(2)	(3)	(4)	(5)	(6)
	revolt	revolt	inter_war	inter_war	all_war	all_war
drought	0.0756**	0.0815*	0.0999**	0.111**	0.117**	0.0993*
	(2.02)	(1.93)	(2.26)	(2.27)	(2.05)	(1.83)
flood	0.0561**	0.0719**	0.0740**	0.0891***	0.0860**	0.0995***
	(2.17)	(2.53)	(2.31)	(2.67)	(2.49)	(2.69)
L.drought	0.0621	0.0683	0.0129	0.0240	-0.00715	-0.0252
	(1.49)	(1.64)	(0.28)	(0.51)	(-0.14)	(-0.42)
L.flood	-0.0400*	-0.0238	0.00178	0.0174	-0.00799	0.00534
	(-1.65)	(-0.90)	(0.06)	(0.55)	(-0.26)	(0.15)
drought × L.drought		-0.0268		-0.0493		0.0839
		(-0.37)		(-0.68)		(1.03)
flood × L.flood		-0.0535		-0.0512		-0.0443
		(-1.60)		(-1.20)		(-0.76)
Constant	-0.967***	-0.964***	-0.529***	-0.525***	-0.198	-0.197
	(-72.77)	(-72.20)	(-4.33)	(-4.31)	(-1.61)	(-1.60)
County Year	Yes	Yes	Yes	Yes	Yes	Yes
County and Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	887262	887262	887262	887262	887262	887262
R^2	0.018	0.018	0.017	0.017	0.016	0.016
Adjusted R^2	0.016	0.016	0.015	0.015	0.014	0.014
F	.	.	.	.	.	.

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

droughts are not strongly above the one induced by floods, especially when the coefficient can varied easily upon the resolution and construction methods of data elected. Meanwhile, the impact of previous droughts is not confirmed. Additionally, historical records for floods and droughts may not be comparable across time. For example, modern definitions for wetness and dryness for floods and droughts have been developed from the level of precipitations (China Meteorological Administration Institute of Meteorology, 1981). However, historical records were normally based on personal perceptions (Zheng, Zhang, and Zhou, 1993), which cannot be strictly converted into the same standard (such as precipitation for floods and droughts records). There is a concern that such records can capture variations within the same category but may not be appropriate to compare between different categories (Xia, 2015). As a result, the wetness/dryness grade may not be appropriate, since the grade combined floods records and droughts records into unique index. Therefore, The use of the weighted frequency rather than wetness/dryness grade has the advantage to capture individual effects for droughts and floods as validated in previous chapter, especially when there were floods and droughts during the same year in the same county. Hence, it is acceptable to suppose that estimations in previous studies have underestimated the effect of the flood on conflicts, and droughts impact on conflicts maybe not significantly higher than flood impact. Both effects of droughts and floods are significantly positive.

5.3.2 The Role of the State Capacity and Multiple Types of Weather Shocks

As mentioned by Xia (2010), extreme drought is among the most severe natural disaster, leading historically to very high number of deaths. However, the impact of droughts was usually exacerbated by a cluster of different types of weather shocks. Therefore, it is necessary to include weather shocks other than floods and droughts. Furthermore, external wars should be considered as encouraging internal revolts due to the redistribution of military power (Zhao, 2012). At the same time, state capacity has been investigated in the field as an important factor of conflicts. Disaster relief has also been employed in previous studies, but there is no county-level data. This research employs the dynastic age and square of the dynastic age as proxies of state capacity to facilitate the analysis.

Before further estimations, it is necessary to illustrate the possible development of the state capacity briefly. For a county that did not have urged desire to expand the frontier, at the beginning of a dynasty, the state capacity was relatively low since institutions for governance and fiscal systems were not well and widely established, and resources were required to be redistributed. After that, there would be two directions that affect the state capacity. One is that the central government would establish more institutions, improve the legal system and invest in relevant infrastructure to strengthen the state capacity. Another one is that local clans would start developing, growing up, combining and finally competing with the central government to distribute newly generated resources, which would weaken the state capacity. Therefore, the state capacity could increase initially and then approach a peak. After the peak, the dynasty would gradually fall into a situation with the fierce political struggle between the central government and local forces (Meng, 2017). Then, the central state capacity decreases in the end. The relationship between the state capacity and the internal conflict is straightforward. Higher state capacity would allow the government to release more reliefs to mitigate the negative impact of disasters, deploy more military power to terrorise and suppress possible rebellions, etc. Then, Higher state capacity is responded by fewer conflicts.

The previous linear-probability model is employed after adding some variables expressed as follow to test the none-linear relationship between the state capacity and conflicts:

$$\begin{aligned} Y_{it} =& \beta_0 + \beta_1 Drought_{it} + \beta_2 Flood_{it} + \lambda \boldsymbol{K}_{it} + \boldsymbol{\omega} County_i \tau \\ &+ \sigma_1 Dyn_t + \sigma_2 Dyn_t^2 + \alpha_i + \gamma_t + \epsilon_{it} \end{aligned} \tag{5.3}$$

$\boldsymbol{K}_{it}$ represents a vector of time variant and spatial varied control variables, which are the dummy of external wars, weighted frequencies related to cold shocks and wind shocks. Dyn_t is age of the ruling dynasty in year t. For example, in 1400, the dynastic age is 1400 minus 1368, which equals 32, since 1400 is during the Ming dynasty, and 1368 is the first year of the Ming dynasty. Alternatively, in 1700, the dynastic age is of 56 (1700 minus 1644). The rest terms remain the same as previous specifications. If the prediction of relationship between the state capacity and conflicts is correct, σ_1 is expected to be negative and σ_2 to be positive, which indicates a diminishing trend for marginal increase of the state capacity.

Table 5.10 shows the results of estimations for the state capacity. Columns 1 and 2 report impact on revolts, while columns 3 and 4 present impact on internal wars. Specifically, Columns 2 and 4 include estimations with lagged variables for drought, flood, cold, wind and external war. For all models, significance for droughts and floods does not deviate too much. Both effects of droughts and floods on conflicts are significantly positive at 5% level, and the effect of drought is slightly higher. Note that, when other types of weather shocks and external wars are controlled, effects of floods are equivalent to 80.9%, 89.5%, 81.4%, 84.3% of effects of droughts in columns 1, 2, 3, and 4, respectively. One order lagged external wars are significant at 1% level in model 2 and 4, but current terms of external wars did not affect internal wars significantly according to model 3 and 4. The interpretation for external war in model 2 is that external wars in a county in a year increase the probability of the outbreak of the revolt by 0.935%. However, the effect of external war lagged by one year is significant and negative, and correspond to a decrease of 0.337% of the frequency of revolts. In column (4), external wars do not significantly affect current internal wars, external wars in the precedent year would reduce the probability of the outbreaks of internal wars by 0.421%. The possible reason for the reduction of internal wars is that external wars would increase the military power of the government. Therefore, local government would have more preparation to deal with possible conflicts next year. It can also be observed that weather shocks in period t-1 do not alter the significance and magnitude of coefficients of weather shocks in period t significantly.

As predicted, results in Table 5.10 for the dynastic age and the square of dynastic age reveal a non-linear trend for conflicts across the whole dynasty. For all

Table 5.10: The Role of State Capacity on Conflicts

	(1) revolt	(2) revolt	(3) inter_war	(4) inter_war
drought	0.0932**	0.0883**	0.110**	0.112**
	(2.32)	(2.21)	(2.34)	(2.37)
flood	0.0754**	0.0790**	0.0895**	0.0944**
	(2.21)	(2.29)	(2.12)	(2.21)
cold	0.0214	0.0163	0.00409	0.00287
	(0.50)	(0.38)	(0.09)	(0.06)
wind	-0.0483	-0.0469	-0.0308	-0.0344
	(-1.56)	(-1.51)	(-0.82)	(-0.93)
exter_war	0.912**	0.935**	0.0603	0.0929
	(2.15)	(2.21)	(0.21)	(0.33)
dynast_year	-1.710***	-1.722***	-1.725***	-1.724***
	(-5.34)	(-5.33)	(-5.38)	(-5.34)
dysq	0.00643***	0.00645***	0.00645***	0.00645***
	(5.33)	(5.33)	(5.36)	(5.33)
L.drought		0.0575		0.00703
		(1.31)		(0.14)
L.flood		-0.0477*		-0.00854
		(-1.66)		(-0.25)
L.cold		0.0533		-0.00406
		(1.16)		(-0.08)
L.wind		-0.0170		0.0193
		(-0.53)		(0.54)
L.exter_war		-0.337***		-0.421***
		(-6.12)		(-2.77)
Constant	-0.968***	0.745**	1.420***	1.189***
	(-77.45)	(2.30)	(3.99)	(3.35)
County Year	Yes	Yes	Yes	Yes
County and Year FE	Yes	Yes	Yes	Yes
Observations	888896	887262	888896	887262
R^2	0.018	0.018	0.017	0.017
Adjusted R^2	0.016	0.016	0.015	0.015
F	.	.	.	.

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

models, the coefficient of dynastic age is significantly negative at 1% level, but the coefficient of the quadratic term is significantly positive at 1% level. For model 2 to model 4, the marginal increase of conflicts every year can be calculated as $\frac{\partial Y_{it}}{\partial Dyn_t} = -1.722 + 0.0129Dyn_t$. Therefore, the minimum of conflicts should be around 133 years from the beginning of the dynasty when $\frac{\partial Y_{it}}{\partial Dyn_t} = 0$. Similarly, for column (1), the minimum of revolts also arrived at around 133 years from the beginning of a dynasty . In this regard, the peak of state capacity for the Ming and the Qing dynasties can be approached roughly in 1501 and 1777, respectively. It has been claimed that Japanese pirates were more rampant around the 1500s because they colluded with local clans of Jiannan regions, indicating the ongoing corruption and weakness of the central government of the Ming dynasty (Meng, 2017). Later, a long term process called Donglin Partisanship has been widely claimed to be one of the most important reasons for the Ming dynasty's crash (Twitchett and Mote, 1998). As for the Qing dynasty, historians usually believed that the Kangxi emperor (1654-1722) pushed the dynasty to its peak, and then later, one of the most famous corrupt officials, Heshen (1750-1799), handled the central government for decades (Peterson, 2016; Meng, 2017). The time points seem not be precisely at the 133th year from the beginning of a dynasty but still suggest a non-linear trend for the dynasty from rise to collapse.

Moreover, controls for other weather shocks indeed capture individual effects on conflicts, but the joint effects of different disasters cannot be estimated. Therefore, to notify the severity of disasters exacerbated by multiple disasters, we construct a dummy to indicate how many types of weather shocks in a county-year. The results are illustrated in Table 5.11. Since lagged weather shocks do not significantly alter the results from none lagged estimations, new estimations for joint effects of multiple weather shocks exclude lagged terms.

In Table 5.11, estimations for joint dummies remain consistent across column (1) and column (2). Column (1) reports that the likelihood of revolts to be triggered increases significantly by 0.205% at 1% level if there was one type of weather shock in a county-year. Additionally, coefficients for two types of weather shocks, three types of weather shocks and four types of weather shocks are higher than that for one type of weather shock, which are 0.302%, 0.253% and 0.286%, respectively at 1% significance level. Effects for multiple weather shocks are 25%-50% higher than effects for only one type of weather shock. Similarly, column (2) reveals an identical relationship, where effects for multiple weather shocks are 35%-37% higher.

5.3.3 The Spatial Spillover of Weather Shocks

According to the distribution of conflicts and weather shocks in Figure 5.5 and Figure 5.6, spatial dependence should be considered. When we recall previous concern about possible sampling biases at the prefecture level, conflicts and weather shocks in neighbouring counties can be one of the most important sources of the biases. Following previous investigations in the relevant field (Hui and Liang, 2016; Harari and Ferrara, 2018), this research employs a spatial panel model which can be estimated with maximum likelihood or GMM techniques. We modify our specifications in previous sections to include spatial lags. The model is expressed as below and estimated by maximum likelihood, as proposed by LeSage and Pace

Table 5.11: The Role of Joint Weather Shocks on Conflicts

	(1)	(2)
	revolt	inter_war
drought	-0.0145	-0.0115
	(-0.28)	(-0.19)
flood	-0.0918*	-0.101
	(-1.83)	(-1.60)
cold	0.0169	-0.00820
	(0.35)	(-0.15)
wind	-0.0764**	-0.0629
	(-2.24)	(-1.56)
exter_war	0.911**	0.0586
	(2.15)	(0.20)
dynast_year	-1.712***	-1.726***
	(-5.34)	(-5.38)
dysq	0.00643***	0.00645***
	(5.34)	(5.36)
joint=1	0.205***	0.237***
	(4.17)	(4.11)
joint=2	0.302***	0.323***
	(4.41)	(3.98)
joint=3	0.253***	0.321***
	(3.08)	(3.21)
joint=4	0.286***	0.325***
	(2.89)	(2.79)
Constant	-0.950***	1.441***
	(-71.51)	(4.06)
County Year	Yes	Yes
County and Year FE	Yes	Yes
Observations	888896	888896
R^2	0.018	0.017
Adjusted R^2	0.016	0.015
F	.	.

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

(2009):

$$
\begin{aligned}
Y_{it} = & \rho \boldsymbol{W}\boldsymbol{Y}_t + \beta_1 Drought_{it} + \beta_2 Flood_{it} + \lambda \boldsymbol{K}_{it} + \delta \boldsymbol{X}_i \tau + \theta_1 \boldsymbol{W}\boldsymbol{Drought}_t \\
& + \theta_2 \boldsymbol{W}\boldsymbol{Flood}_t + \boldsymbol{\phi}\boldsymbol{W}\boldsymbol{K}_t + \boldsymbol{\mu}\boldsymbol{W}\boldsymbol{X}\tau + \alpha_i + \gamma_t + \epsilon_{it}
\end{aligned}
\tag{5.4}
$$

This is a Spatial Durbin Model where $\boldsymbol{W}$ is a normalised spatial weight matrix. Spatial lags for dependent variables and independent variables are included to capture possible spatial spillover effects. The spatial weight matrix of the main results is an inverse distance matrix and is normalised by dividing the maximum eigenvalue of the inverse distance matrix proposed by Kelejian and Prucha (2010). In addition, coefficients can be interpreted as percentage point except for rho since spatial lags of dependent variable are multiplied by 100 as well.

Table 5.12 reports the results of the maximum likelihood estimations for the Spatial Durbin Model. Columns (1) to (3) employ the revolt as a dependent variable, and columns (4) to (6) adopt the internal war as the dependent variable. The results remain consistent when different controls are considered. Models in column (3) and column (6) are taken as main results since model 3 and model 6 contain all controls to facilitate the analysis.

Column (3) confirms that there were significant spillover effects in revolts, which means that uprisings in neighbouring counties would encourage local revolts. According to the results, the neighbouring revolts increase the likelihood of an outbreak of local revolts by 96.5% on average (rho = 0.965), which implies that coefficients for weather shocks can be underestimated due to possible feedback ef-

Table 5.12: The Results of Spatial Durbin Models

	Revolt			Internal War		
	(1)	(2)	(3)	(4)	(5)	(6)
Flood	0.0579**	0.0579**	0.0558**	0.0779**	0.0779**	0.0770**
	(2.26)	(2.26)	(2.17)	(2.52)	(2.52)	(2.49)
Drought	0.0624*	0.0624*	0.0624*	0.0728*	0.0729*	0.0728*
	(1.94)	(1.94)	(1.94)	(1.88)	(1.88)	(1.88)
Cold	-0.0209	-0.0207	-0.0230	-0.0267	-0.0267	-0.0288
	(-0.68)	(-0.67)	(-0.75)	(-0.72)	(-0.72)	(-0.78)
Wind	-0.0414*	-0.0417*	-0.0423*	-0.0369	-0.0370	-0.0370
	(-1.71)	(-1.73)	(-1.75)	(-1.27)	(-1.27)	(-1.27)
External war		0.906***	0.910***		0.0871	0.0860
		(5.01)	(5.04)		(0.40)	(0.39)
rho	0.965***	0.962***	0.962***	0.966***	0.966***	0.964***
	(912.40)	(844.15)	(837.90)	(926.41)	(925.75)	(891.28)
W*Flood	0.905***	0.899***	0.935***	0.592	0.584	0.514
	(2.71)	(2.69)	(2.76)	(1.47)	(1.45)	(1.26)
W*Drought	0.263	0.260	0.277	0.401	0.394	0.416
	(0.89)	(0.88)	(0.93)	(1.12)	(1.10)	(1.16)
W*Cold	2.480***	2.471***	2.461***	1.759***	1.754***	1.711***
	(5.20)	(5.19)	(5.12)	(3.06)	(3.06)	(2.95)
W*Wind	-1.193***	-1.185***	-1.110***	-0.657	-0.646	-0.553
	(-3.31)	(-3.29)	(-3.06)	(-1.51)	(-1.49)	(-1.27)
W*External war		1.884	1.217		5.026	4.151
		(0.60)	(0.38)		(1.32)	(1.09)
Obs	888896	888896	888896	888896	888896	888896
Control Trend	No	No	Yes	No	No	No
County Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.0477	0.0476	0.0477	0.0468	0.0468	0.0468
log-likelihood	-2643136.4	-2643118.2	-2643100.7	-2808826.9	-2808825.9	-2808798.2

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

fects. Specifically, if all neighbours had outbreaks of revolts for a given year, the likelihood of the outbreak of local revolts would increase by 96.5%. These effects are decayed when the distance to neighbours increases. Neighbours near to local county have higher weights in the aggregated effects (Vega and Elhorst, 2013). Given the maximum eigenvalue (2.76) of inverse the distance-based spatial weight matrix, neighbours within 3.63 kilometres accounted for around 10% of the total effects (which means each of neighbouring revolts within 3.63 kilometres increases the outbreak of local revolts by approximately 9.65%) and neighbours over 363 kilometres only accounted for less than 0.1%. To be specific, coefficients of spatial lags for the flood, the cold, and the wind are significant at 1% level, indicating strong spillover effects for the flood, cold and wind as well. Effects of floods and external wars on revolts are significantly positive at 5% and 1%, respectively.

Similarly, column (6) reports that internal wars are also spatially dependent , denoted by the coefficient of rho (0.964) at 1% significance. The effect of floods on internal wars is estimated to be significantly positive at 5%, but external wars no longer affect internal wars. Coefficient of spatial lags of cold keep significant at 1% level but other spatial lags do not significantly impact internal wars. Estimations for spatial models are comparable to non-spatial models in Table 5.10, where external wars stimulate the outbreak of revolts but have an insignificant impact on internal wars. Effects of floods remain significantly positive while effects of droughts lose significance in spatial models. Estimations in non-spatial models may capture spatially dependent effects as local effects, which biases the results.

To further investigate the spatial spillover effect of every independent variable, direct effects and indirect effects (Elhorst, 2014) are considered. The results are presented in Table 13 (results are based on model 3 and model 6 in Table 5.12). The direct effect represents the response of the local dependent variable to the local independent variable. The indirect effect (or the spillover effect) illustrates the impact of the neighbouring independent variable on the local dependent variable. Intuitively, direct effects include feedback effects representing the independent variable' s effect (weather shocks, for example) passing through the neighbouring dependent variable (conflicts, for example) and back to the local county. In details, neighbouring conflicts can stimulate local conflicts and can be affected by local weather shocks. Therefore, when weather shocks alter neighbouring behaviours for conflict and change the likelihood of a local outbreak of conflicts, the effect from this path is recognised as the feedback effect.

According to Table 5.13, columns 1 to 3 show the results of estimations for revolts and columns 4 to 6 show the results for internal wars. For weather shocks, the direct effect on the revolts of the flood appears to be 0.0702% at 1% significance. The direct effects of the drought and the wind on revolts are 0.0654% and -0.0571% at 5% significance level, respectively. However, the direct effect on internal wars of the drought is 0.0778%, only at 10% significance level. In this regard, droughts are less likely to affect revolts and other types of internal wars. The coefficients of droughts in non-spatial models (0.0892% on revolts and 0.11% on internal wars at 5% significance as shown in Appendix A.1) are generally overestimated. Since the coefficient of floods effects on revolts is 0.0558% in column (3) of Table 5.12, and its direct effect is 0.0702%, the feedback effect of the flood amounts to 0.0144%, which accounts for 20.51% of the direct effect. In other words, the feedback effect of the flood is significantly large. The interpretation can be that one more flood record

Table 5.13: The Estimations of Spatial Spillover Effects

	Revolt			Internal War		
	Direct effect	Indirect effect†	Total effect†	Direct effect	Indirect effect†	Total effect†
Flood	0.0702***	2.5034***	2.5104***	0.0866***	1.5392	1.5479
	(2.75)	(2.89)	(2.90)	(2.81)	(1.40)	(1.41)
Drought	0.0654**	0.8577	0.8642	0.0778*	1.3084	1.3162
	(1.98)	(1.15)	(1.16)	(1.96)	(1.38)	(1.39)
Cold	0.0086	6.3317***	6.3326***	-0.0063	4.6318***	4.6312***
	(0.29)	(5.30)	(5.30)	(-0.18)	(3.06)	(3.06)
Wind	-0.0571**	-2.9154***	-2.9211***	-0.0446	-1.5335	-1.5380
	(-2.50)	(-3.19)	(-3.20)	(-1.62)	(-1.32)	(-1.32)
External war	0.945***	5.1752	5.2697	0.154	11.2598	11.275
	(5.18)	(0.65)	(0.66)	(0.70)	(1.11)	(1.12)
Obs	888896	888896	888896	888896	888896	888896
Control Trend	Yes	Yes	Yes	Yes	Yes	Yes
County Year FE	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

† Coefficients multiply 0.1 to mimic the effects from counties within 3.63km

increases the likelihood of local revolts outbreak by 0.0702%, of which 20.51% is from the feedback effect. The significant feedback effect of the flood can be explained as the enormous magnitude of spatial spillovers. The indirect effect of the flood on revolts is estimated to be 2.5034% at 1% significance for counties within 3.63 kilometres (10% of all effects as illustrated before), accounting for 99.7% of the total effect (2.5104%). Therefore, if the flood of neighbour within 3.63 kilometres increased by 1 unit for a given year, the likelihood of the outbreak of local revolts would increase by 2.5034%. The impact of floods from surrounding counties is significantly higher than that from the local county. Since the spatial weight matrix of the main results is an inverse distance matrix, estimations can capture effects from multiple orders neighbours, which suggests that the flood's spillover effect also remains significant in a larger region. In contrast, the feedback effect of droughts on revolts is only 4.6% of its direct effect, and the indirect effect of droughts is insignificant. Results keep consistent in estimations on internal wars, which suggests that the impact of droughts on conflicts is more likely to be contained within the local area (within a county).

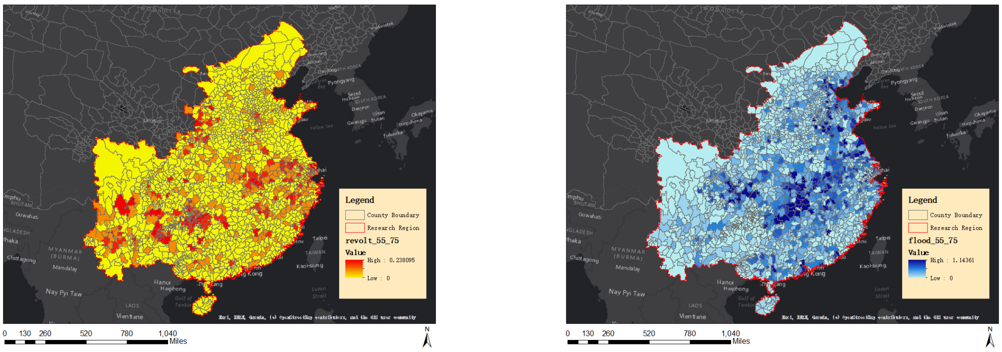
Moreover, the direct effects of cold on revolts and internal wars appear insignificant but the indirect effects are significant at 1% significance level. 1 unit increase in neighbour average cold counts within 3.63 kilometres would raise the probability of local revolts by 6.3317% and of local internal wars by 4.6318%. The consistency of estimations across revolts and internal wars indicates that the cold has strong spatial spillover effects on conflicts. By contrast, the wind shows different influences on varied war types. The direct effect and indirect effect of the wind on revolts are significantly negative at 1% level, while those effects on internal wars are insignificant. Accordingly, the wind would reduce the likelihood of the outbreak of local and neighbouring revolts but have insignificant effects on internal wars.

Possible mechanisms of the spillover effects are complicated and it is challenging to find direct evidence of them, especially for weather shocks. The spillover effect of conflicts has been interpreted as emulations of neighbouring rebellions (Buhaug and Gleditsch, 2008) and intention-based conflicts induced by refugee flows (Gleditsch, 2007). Revolts or rebellions during the Ming and Qing dynasties can provide evidence to support these mechanisms. For example, the record for the refugee related Wang’ er uprising from Shaanxi province in 1628 mentioned : “Disasters in northern Shaanxi have been severe for years. Local officials imposed strict taxes, and farmers have revolted. In 1628, Wang’ er in Chengcheng gathered hundreds of people to revolt, attacking the Xiaotong town in Pucheng, Zichuan Town in Hancheng, killing officials and robbing prisons. Wang Jiayin in Fugu, Wang Zuogui in Yichuan, Gao Yingxiang in Anzhai, etc., successively responded with troops and combated across counties⋯” . (Compilation Group of Chinese military History, 2003) This example indicates that conflicts were spatially connected since peasants in neighbouring areas would be encouraged to join a successful rebelling, and battles would spread to exploit a first streak of successes. In this regard, neighbouring counties, where people had great grievances or where the governance is weak, would be more likely to be a target for the next battle. As for the refugee flows, intuitively, revolts can be strengthened since more people would join in, and refugees from other regions would have a more outstanding grievance on local officials if policies or relieves were inadequate (Ge, Wu, and Cao, 1997;

Meng, 2017).

Channels of spillover effects of weather shocks are far more complicated. The cold had significant and consistent spillover effects on internal conflicts, possibly due to the destructive impact on agriculture production as proposed in previous studies (Ge, Wu, and Cao, 1997; Zhang, Tian, et al., 2010; He, Li, and Liu, 2010). There is evidence showing that the crop maturity period was delayed and the production reduced (rotten seedlings, less crop ears and empty crop batch) due to cold weather (He, Li, and Liu, 2010). Xu Jinzhi summarised that from 1470 to 1520, winter remained cold (Chu, 1973). For example, in 1493, heavy snow had fallen in the Huaihe River Basin and ended in February of the following year. In 1513, Dongting Lake, Poyang Lake and Taihu Lake had frozen at the same time. The spring period favourable to wheat cultivation in southern Jiangsu was delayed by more than 10 days compared with modern cases. From 1859 to 1861, extreme colds were recorded. For example, in 1861, “It snowed heavily in Puyi, and snows accumulated to around 2 metres. Several people and animals died and all rivers froze” . At the same time, Taiping Rebellions and other revolts spread across half of China. The record for Bagua Rebellions of Shandong province denotes potential connections: “…In the winter of 1860, there were revolts across Shandong province against the food tax…” Due to the incomplete records for connections between conflicts and weather shocks, it is hard to find direct evidence for cold-induced battles, which may explain the cold’ s insignificant direct effects on conflicts. In other words, cold weather would accelerate or expand the spread of conflicts but is less likely to initiate a conflict. According to other scholars (Buhaug and Gleditsch, 2008; Turchin et al., 2013), it makes sense since determinants of conflicts are usually complex and interacted with each other. The wind would induce physical destruction and reduce agriculture production as well but may affect the logistics of warfare, which suggests that big wind could hinder the movement of troops (Buhaug and Gleditsch, 2008). Furthermore, results of some famous wars in Chinese histories were altered by the wind. For example, the army of the Yuan dynasty was hindered by gale on its way to Japan. In 1401, a record for the war between the successors of the first Ming Emperor mentioned that "Zhudi (the fourth son of Hongwu emperor) attacked Baoding in February…Lure the defender Wujie of Zhending to strike and initiate the battle in Gaocheng…Zhudi's army attacked along with the wind and won the battle…". Therefore, commanders of conflicts would typically consider the influence of unfavourable wind direction and difficulties of the movement of troops. However, only revolts were significantly affected by wind, according to the estimation. The impact of wind can probably be mitigated by more vital logistics abilities and troops' quality since it is reasonable to assume that peasants were not better trained and better organised as soldiers. Indirect evidence can be identified from a record in 1495 which mentioned that the soldiers "marched 14 days and not hindered by the wind and blizzard". It is natural to consider that the cold may impose similar impact on the movement of troops, while the positive indirect effects can exclude this mechanism.

Droughts and floods usually induced massive deaths and considerable damage to society. From the end of the 15th century to the beginning of the 17th century, droughts generated numbers of refugees, who were previous farmers but lost their lands. Droughts also induced locust plague, which worsened the impact of disasters (Ge, Wu, and Cao, 1997). From 1875 to 1878, extreme droughts were recorded in Shanxi province, Henan province, Shaanxi Province, Hebei province, Shandong province and part of the Jiangnan area (Ge, Wu, and Cao, 1997). Therefore, droughts are expected to capture strong direct and indirect effects on conflicts which coincided with previous studies (Jia, 2014; Xia, 2010; Zhang, 1996). However, it is surprising to find the low significance of direct effects and the insignificant indirect effects of droughts. A possible explanation is that droughts led to more deaths, which mitigated the impact of refugees or other residents to source the revolts. As for floods, the most well-known flood records were the breaks of river embankments and rerouting of river courses. In 1855, there was a dramatic diversion of Yellow River (The estuary changed from Huaihai to Bohai) (Ge, Wu, and Cao, 1997). The River was breached in Henan province and spread to Shandong province without a specific river course in 20 years. The new river's course had destroyed local villages and farmlands, which led to population outflow, whereas lands around the abandoned river courses had attracted migrants to move in. Therefore, if floods frequently occurred over long period, people would decide to move far away from affected local areas, spreading the economic influence of floods to farther regions. Integrating the channels through which floods affect conflicts, previous studies (Harari and Ferrara, 2018), proposed that short term floods would increase local conflicts by reducing the opportunity cost of joining rebellions and weakening the state capacity of local government. However, long-term floods would induce people to move far away and diffuse the effects of floods on a larger area. As shown in Figure 5.7, the dark blue counties in the northeast part represent the areas where floods were induced by frequent rerouting of the river course. However, battles were seldom triggered locally (0 records in correspondence fraction in subfigure a). Similar relationships can be observed in the central part as well. In this regard, since revolts would be affected by refugees, the spillover effects of floods can be significant for estimations on revolts. In contrast, the spillover effects of floods on internal wars are not significant.



(a) Revolt (b) Flood

Figure 5.7: Spatial Distribution of Revolts and Floods from 1855 to 1875

There is the possibility for floods to generate spillover effects through the impact on agricultural production. Li (2007) claimed that the extreme flood of 1823 in Songjiang prefecture was the reason for the recession in the Jiangnan region in the 1820s and then triggered the Taiping Rebellions. The flood destroyed the farmland and corroded the soil fertility, which reduced agricultural production in the long term. Therefore, long-term food scarcity led to long-term recessions, and long-term recessions incited rebellions. In this case, the flood in Songjiang prefecture spread its influences to the whole Jiangnan region and the whole country through economic performance and enhanced the riskiness of the outbreak of conflicts across time and space. However, there were nearly 30 years between the Songjiang floods and the Taiping Rebellions. Many conflicts and floods occurred during this period. Therefore, it is incautious to argue that Songjiang floods causally related to the

Taiping Rebellions. Additionally, there were sufficient water control systems and well-developed water conservancy communities (Sun, 2017) to ensure the agricultural production against floods risk. In this regard, whether the flood could spread its influences through the channel of agriculture reduction can be doubtful. It is also possible that people would get used to frequent floods where waterworks were not well-maintained. As a result, people in these areas might find other channels to mitigate the negative impact on agricultural production. The spillover effect of floods through the agriculture channel might account for only a part of the total spillover effect. Additional research will be required to delve in depth in the mechanism through which floods impact in the short term and long term conflict.

5.4 Robustness Checks

5.4.1 Robustness checks of Choice of the Spatial Weight Matrix

One concern about spatial estimations is that the choice of the spatial weight matrix may affect spatial dependence. This research applied several distance-based matrices to verify the consistency of estimates for the baseline spatial model. Table 5.14 shows results in which revolt is the dependent variable. Model in column (1) employs inverse-distance spatial weight matrix with the normalisation method proposed by Ord (1975). Columns 2 to 4 represent estimations using binary contiguity matrices with distance cutoffs of 270km, 370km, and 600km.

To be specific, 270km represents the maximum distance of the nearest neighbour, and 370km represents a boundary where the spillover effect decays to 0.1% of the average level. Column (5) applies a binary contiguity matrix that contains four nearest neighbours of each county. Matrices of columns 2 to 5 are row normalised. Column (6) employs an inverse distance square matrix which is normalised by the maximum eigenvalue. Similarly, Table 5.15 shows results using internal wars as the dependent variable, and from columns 1 to 6, the matrices are the same as the matrices in Table 5.14.

According to Table 5.14, coefficients for floods and external wars remain consistently significant across different choices of matrices from columns 1 to 6. Coefficients for the wind are negative but at a relatively lower significance level. In addition, coefficients of spatial lags of the cold show strong consistency as well. These results reaffirmed the estimations in the baseline model. Spillover effects of revolts are significantly positive across all models (the coefficient of rho denotes effects), but the magnitude varies. Intuitively, the spillover effect should decay when the distance increases. However, magnitudes of rho in columns 2 to 4 increase from 0.667 to 0.864, which indicates the spillover effect gets more potent when the distance grows. Meanwhile, rho in column (5) is only 0.293. Since matrices in columns 2 to 5 are normalised by row, each unit accounts for the same proportion in the estimation. Therefore, it is likely to overestimate neighbours with different distances. Specifically, distances between neighbours in this research are not identical. The radius of the range of the nearest four neighbours varies from 50 km to over 270 km. In this regard, it is likely for estimations to exaggerate the effect of the nearest neighbours when the cutoff distance is small.

Table 5.14: The Estimations on Revolts across Different Spatial Matrices

	(1)	(2)	(3)	(4)	(5)	(6)
Flood	0.0552**	0.0683***	0.0745***	0.0674***	0.0587**	0.0710***
	(2.15)	(2.68)	(2.93)	(2.65)	(2.33)	(2.79)
Drought	0.0592*	0.0422	0.0450	0.0623**	0.0593*	0.0477
	(1.83)	(1.32)	(1.41)	(1.96)	(1.88)	(1.49)
Cold	-0.0265	-0.0193	-0.0162	0.0126	0.0102	-0.0190
	(-0.86)	(-0.64)	(-0.53)	(0.42)	(0.33)	(-0.63)
Wind	-0.0421*	-0.0450*	-0.0455*	-0.0480**	-0.0467*	-0.0468*
	(-1.74)	(-1.88)	(-1.91)	(-2.01)	(-1.96)	(1.96)
External war	0.906***	0.914***	0.908***	0.956***	0.954***	0.931***
	(4.99)	(5.09)	(5.05)	(5.29)	(5.38)	(5.18)
Rho	0.965***	0.667***	0.759***	0.864***	0.293***	0.965***
	(907.96)	(195.21)	(220.97)	(261.09)	(211.87)	(929.25)
W*Flood	1.005***	0.0702	-0.0772	0.287	0.0762*	0.0301
	(2.78)	(0.56)	(-0.51)	(1.21)	(1.73)	(0.17)
W*Drought	0.344	0.123	0.101	0.0518	0.0620	0.195
	(1.05)	(1.02)	(0.73)	(0.26)	(1.18)	(1.19)
W*Cold	2.856***	0.597***	0.776***	1.016***	0.0918*	0.775***
	(5.74)	(3.77)	(4.01)	(3.30)	(1.74)	(3.19)
W*Wind	-1.192***	-0.169	-0.159	-0.619**	-0.0512	-0.231
	(-3.17)	(1.37)	(-1.04)	(-2.41)	(-1.25)	(-1.22)
W*External war	1.504	-0.396	-0.399	-1.260	-0.291	-0.752
	(0.49)	(-0.40)	(-0.33)	(-0.63)	(-0.86)	(-0.45)
Obs	888896	888896	888896	888896	888896	888896
Control Trend	Yes	Yes	Yes	Yes	Yes	Yes
County Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.0475	0.0554	0.0459	0.0475	0.0856	0.0524
log-likelihood	-2643101.6	-2640656.6	-2644485.4	-2650399.5	-2631290.1	-2641783.3

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In addition, the inverse distance square matrix would suggest stronger distance decay and capture part of the spatial spillovers of the second-order neighbours (Jensen and Lacombe, 2012). Therefore, matrices in columns 2 to 6 may not be accurate but still suggest significant and large spillover effects of revolts. Similar conclusions can be addressed from estimations on internal wars. As shown in Table 5.15, floods are suggested to have significant impact on internal wars directly and the spatial dependence of internal wars is robust.

5.4.2 Randomisation Test of Key Variable

Another concern is driven by the false independence of key variable. This research employs the randomisation test proposed by Kennedy (1995) to check the robustness of significant estimations. In order to facilitate the estimation, records of floods are reshuffled across 1634 counties by 1000 times since coefficients of floods remain high significance in baseline models. Therefore, there would be 1000 t values for the coefficient of the flood recalculated from the test. Each t value would be compared with original t value in the baseline model (suppose to be t_0). In this regard, it is possible to test the robustness of our baseline models according to the probability p of $t > t_0$. If $p < 0.05$, we can reject the null hypothesis that coefficients of baseline models are zeros.

We select column (3) and column (6) in Table 5.12 as baseline models. The p value of variables "Flood" and "W*Flood" are 0.04 and 0.005 for estimation on the revolt (column (3)), and the p value of variables "Flood" and "W*Flood" are 0.012 and 0.026 for estimation on the internal war (column (6)). All reject null

Table 5.15: The Estimations on Internal Wars across Different Spatial Matrices

	(1)	(2)	(3)	(4)	(5)	(6)
Flood	0.0727**	0.0806***	0.0849***	0.0801***	0.0680**	0.0888***
	(2.35)	(2.63)	(2.77)	(2.62)	(2.23)	(2.89)
Drought	0.0692*	0.0478	0.0547	0.0840**	0.0647*	0.0594
	(1.78)	(1.24)	(1.42)	(2.20)	(1.69)	(1.54)
Cold	-0.0343	-0.0306	-0.0247	0.0026	-0.0013	-0.0277
	(-0.93)	(-0.84)	(-0.68)	(0.07)	(-0.04)	(-0.76)
Wind	-0.0346	-0.0336	-0.0325	-0.0387	0.0385	-0.0389
	(-1.18)	(-1.17)	(-1.13)	(-1.35)	(-1.34)	(-1.35)
External war	0.0916	0.147	0.126	0.168	0.124	0.116
	(0.42)	(0.68)	(0.58)	(0.77)	(0.58)	(0.54)
rho	0.964***	0.725***	0.709***	0.855***	0.276***	0.965***
	(892.53)	(241.32)	(180.71)	(246.84)	(197.57)	(934.31)
W*Flood	0.879**	0.0758	0.0495	0.391	0.103*	-0.0746
	(2.02)	(0.51)	(0.27)	(1.37)	(1.93)	(-0.34)
W*Drought	0.565	0.213	0.263	0.0451	0.114*	0.245
	(1.43)	(1.47)	(1.58)	(0.19)	(1.80)	(1.24)
W*Cold	2.295***	0.488**	0.721***	0.718*	0.0791	0.522*
	(3.83)	(2.56)	(3.10)	(1.94)	(1.24)	(1.78)
W*Wind	-0.914**	-0.182	-0.283	-0.503	-0.0531	-0.0749
	(-2.02)	(-1.23)	(-1.54)	(-1.63)	(-1.07)	(-0.33)
W*External war	3.544	-0.478	-0.0526	-1.094	0.0581	0.837
	(0.97)	(-0.41)	(-0.036)	(-0.45)	(0.14)	(0.42)
Obs	888896	888896	888896	888896	888896	888896
Control Trend	Yes	Yes	Yes	Yes	Yes	Yes
County Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.0472	0.0573	0.0474	0.0336	0.0767	0.0524
log-likelihood	-2808591.5	-2805759.4	-2808911.9	-2814342.4	-2800097.7	-2807099.2

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

hypothesis. Therefore, our results are robust.

5.5 Conclusion

In this paper, we reproduce the estimations of Jia' s work using a linear probability model but employing county-level data. The linear probability model results are similar to the prefecture-level estimates in Jia' s work, but the drought does have not a particularly higher coefficient than the flood. We develop some estimations to capture the impact of the state capacity and the joint influence of multiple disasters. As a result, we find that state capacity plays a vital role to reduce internal wars, revolts included. The dynasty would approach its peak of state capacity in the 133rd year of dynastic age on average. What else, multiple types of weather shocks in the same year would have a significantly higher impact on conflicts compared with a single type of disaster.

Additionally, to further investigate the deviation between the prefecture level and the county level, spatial dependency is considered, and the Spatial Durbin Model is employed. We find that the spatial dependence of conflicts is significantly strong, which alters non-spatially corrected estimations. Specifically, non-spatial estimations overestimate the effect of droughts and underestimated the impact of the flood. More specifically, the feedback effect of floods is significantly higher. Furthermore, cold episodes have positive spillover effects on internal wars, but floods only have positive spillover effects on revolts. On the contrary, wind has negative spillover effects on revolts.

Spatial analysis seems to go against previous conclusions from non-spatial studies (Jia, 2014; Hao, Ge, and Zheng, 2010). Previously proposed mechanisms have claimed that droughts and floods would reduce agriculture production, leading to famine and then inducing revolts (Jia, 2014; Xia, 2010). In contrast, cold, which is related to temperature, did not significantly impact conflicts since crop types varied across different regions. However, this paper argues that cold can facilitate the spread and emulations of conflicts through the negative impact on agriculture production. Meanwhile, floods might generate more refugees since the spillover effect is more significant, which would source troops, and then encouraged the spread of revolts specifically. In contrast, the wind had significantly negative spillover effects on revolts, suggesting an assumption that the wind hindered the movement of refugees and troops, which would reduce the spread of revolts. Moreover, the channel between weather and conflict is more likely to be effective through some physical and direct destruction on the economy and society rather than agriculture production, since the latter one might be mitigated by other factors such as trade and relief (Li, Cheng, et al., 2020; Deng, 2011).

Due to the constraints related to the frequency of time period elected and data availability, several limitations were further investigated. The first one was related to data. Since county-level data are not appropriate to construct wetness/dryness grade using the previous method (Zheng, Zhang, and Zhou, 1993), we employ weighted frequencies of historical records based on the estimated credibility of sources. Therefore, the validation of this data construction needs to be checked in a more cautious manner. In addition, connections between different disasters are not considered in this research, and consequently ignores the fact that climatic events interact. Besides, the classification of massive weather shocks is also required to be further investigated. Better estimations can be expected for the ongoing database, which is compiled by the team of Xia (2015). The second caveat is related to the specification which includes our spatial analysis. This research applies a static Spatial Durbin Model due to the limitation of estimation techniques, which means that dynamic spatial model cannot be run successfully with large matrices. However, there are evidences that conflicts have strong temporal dependence (Harari and Ferrara, 2018). Thus, a dynamic Spatial Durbin Model should be considered in future estimations. The third one is connected to the choice of the spatial weight matrix. Although appropriately normalised inverse-distance based spatial weight matrix is better than most binary based spatial weight matrices, boundaries are lost. It is possible to find a better spatial weight matrix such as combined distance-boundary weights (Cliff and Ord, 1967) as follow:

$$w_{ij} = \frac{l_{ij} d_{ij}^{-\alpha}}{\Sigma_{k \neq i} l_{ik} d_{ik}^{-\alpha}} \tag{5.5}$$

where w_{ij} is the weight of unit i on j. l_{ij} represents shared boundaries between unit i and j. d_{ij} represents the distance between unit i and j. The last one is related to the partitioning effects. As introduced by Jensen and Lacombe (2012), the spatial effect varies across different distances, which is not captured in details in the spatial model. Therefore, a partitioning of direct and indirect effects can be proposed to split the impact in higher orders. The portioning techniques can be employed through the transformation from the general spatial model as follow:

$$
\begin{aligned}
\frac{\partial Y}{\partial X} &= (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}(\beta_k + \boldsymbol{W}\theta_k) \\
&= (\boldsymbol{I} + \rho \boldsymbol{W} + \rho^2 \boldsymbol{W}^2 + \rho^3 \boldsymbol{W}^3 + ...)(\beta_k + \boldsymbol{W}\theta_k) \\
&= (\boldsymbol{I}\beta_k + \rho \boldsymbol{W}\theta_k) + (\rho \boldsymbol{W}\beta_k + \rho \boldsymbol{W}^2\theta_k) + (\rho^2 \boldsymbol{W}\beta_k + \rho^2 \boldsymbol{W}^3\theta_k) + ...
\end{aligned}
\tag{5.6}
$$

where $\boldsymbol{I}\beta_k + \rho \boldsymbol{W}\theta_k$ represents the local impact, $\rho \boldsymbol{W}\beta_k + \rho \boldsymbol{W}^2\theta_k$ describes the impact of the first order neighbours etc.

Another constraint is about the causal direction between conflicts and disasters. Although most disasters were not controlled by human power, there were several examples indicating that disasters might be triggered or caused by human activities. For example, there was a comment claimed that Li Zicheng attacked Kaifeng in 1642, causing the Yellow River to burst and flood the city (Mote and Twitchett, 1988). However, there was also a record argued that the flood in Kaifeng in 1642 was a natural disaster rather than a man-made disaster (Li, 2016a). Additionally, Marks (2011) mentioned that the long term farmland reclamation surround lakes and rivers was the main reason for frequent floods (especially the Yellow River burst in 1855) during the late Qing dynasty, but there was no robust evidence to ensure the conclusion. A similar issue about the endogeneity of floods has also been widely ignored in previous studies and this research. For example, Osborne (1994) illustrated a human-environment circle in the late Qing dynasty to show the connection between land reclamation of shack migration and ecological deterioration. Since the land reclamation was induced by universal population pressure across the country, it is reasonable to assume that human-induced disasters could account for considerable proportions. However, there was no direct record to isolate these causes for individual incidents. Thus, the actual causal direction was complicated to be verified from historical records. It requires further investigation to exclude possible endogeneity from this inverse causal relationship.

Chapter Six

Conclusion

This thesis examines various aspects of the environmental economic history of China and provides the data and analysis to allow for more research at the county level. After a brief introduction, Chapters two and three construct and validate a county-level environmental data set for the period 1368 to 1911. The study verifies the use of some newly developed AI based techniques, including character recognition and lexical analysis, to digitise historical information from paper sources. Such techniques can help academics to save time and improve efficiency. Moreover, the combination of data from separate sources does not bias the historical environmental records. Meanwhile, the use of GIS benefits the management and visualisation of spatially-varied historical environmental data.

According to relevant maps generated from the data set, the spatial distribution of environmental records did not vary significantly across the Ming and Qing dynasties. However, the number of flood records increased considerably from the Ming dynasty to the Qing dynasty when the highest average frequency of flood records in a county increasing from 1.04 to 2.41. This new data set can be used alongside data from the official records and can benefit further research at the county level.

Applications of this newly constructed data set are shown in chapters four and five. In chapter four, we theoretically and empirically examine the impact of environmental shocks on urban-rural migration. The theory adopts the two-sector Harris-Todaro model and uses the decline in marginal output caused by the environment as the decision-making basis for urban and rural migration. We believe that unfavourable environmental conditions will encourage people to move to urban areas to obtain higher expected incomes. However, different natural disasters will affect the expected wages in the agricultural and non-agricultural sectors differently. In addition, farmland reclamation is generally considered the preferred way to reduce expected income reduction rather than migration. We test the theoretical prediction that natural disasters may positively impact urban-to-rural migration using cross-sectional regressions at the prefecture and county levels. Our main finding is that more frequent flood events will encourage rural-to-urban migration, while more frequent cold-related events will encourage urban-to-rural migration.

Chapter five mainly focuses on the county-level estimations and finds that the flood and multiple disasters in the same year encourage the outbreak of conflict. In order to further investigate the origin of the bias in the prefecture-level research, we adopt a Spatial Durbin Model and find that the spatial dependence of conflicts is strong. In this regard, the influence of floods are underestimated, and the influence of droughts is overestimated. To be specific, floods had significantly positive direct effect on the outbreak of conflicts, and cold had a significantly positive indirect effect (spillover effects) on the outbreak of conflict. The possible mechanisms for the spillover effect of weather shocks are whether the weather shocks are spatially connected or the weather shocks encourage refugees to move to other places.

There are several limitations in the research presented in this thesis. First, in data construction, supplementary sources are not always sufficient to complete county-level information for every county. In addition, a stronger understanding of history and environmental science would have helped with information extraction from historical documents. Other technique related issues such as the efficiency of AI may also have affected the accuracy of the data set, which calls for more extensive corporation across different fields, including history, meteorology, computer science, etc. Second, in the model selection, the theoretical model in chapter four is fairly dated and does not match closely enough with the empirical estimations. In addition, the spatial model in chapter five does not employ a dynamic model due to software constraints, and the spatial weights can be improved. Finally, data availability. Since the county-level data have to previous been constructed, the county-level estimations are fairly in scope which may introduce bias from the choice of region and period which are constrained within the traditional agriculture regions from 1368 to 1911. There are also a large number of zeros in the linear spatial model in chapter five. All these issues are expected to be solved in the future if a more comprehensive county-level data set is constructed.

Appendix One

First Appendix

A.1 Appendix for Non-spatial Estimations

	(1)	(2)	(3)	(4)
	revolt	revolt	inter_war	inter_war
drought	0.0932**	0.0892**	0.110**	0.110**
	(2.32)	(2.22)	(2.34)	(2.34)
flood	0.0754**	0.0794**	0.0895**	0.0926**
	(2.21)	(2.33)	(2.12)	(2.21)
cold	0.0214	0.0280	0.00409	0.0103
	(0.50)	(0.66)	(0.09)	(0.22)
wind	-0.0483	-0.0477	-0.0308	-0.0353
	(-1.56)	(-1.53)	(-0.82)	(-0.95)
exter_war	0.912**	0.907**	0.0603	0.101
	(2.15)	(2.15)	(0.21)	(0.37)
Constant	-0.968***	-1.383***	1.420***	0.935**
	(-77.44)	(-6.64)	(3.99)	(2.18)
Control Year	No	Yes	No	Yes
County Year	Yes	No	Yes	No
County and Year FE	Yes	Yes	Yes	Yes
Observations	888896	888896	888896	888896
R^2	0.018	0.015	0.017	0.015
Adjusted R^2	0.016	0.014	0.015	0.014
F	.	3.685	.	4.343

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

References

Abel, Guy J. et al. (2019). “Climate, conflict and forced migration”. In: *Global Environmental Change* 54, pp. 239–249.

Adams, Courtland et al. (2018). “Sampling bias in climate–conflict research”. In: *Nature Climate Change* 8.3, pp. 200–203.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (2013). “On the Origins of Gender Roles: Women And The Plough”. In: *Quarterly Journal of Economics* 128.2, pp. 469–530.

Allen, Robert C (2009). “Agricultural Productivity and Rural Incomes in England and the Yangtze Delta, c.1620–c.1820”. In: *Economic History Review* 63.3, pp. 525–550.

Allen, Robert C. et al. (2011). “Wages, Prices, and Living Standards in China, 1738–1925: in Comparison with Europe, Japan, and India”. In: *Economic History Review* 64.1, pp. 8–38.

Alsan, Marcella (2015). “The Effect of the TseTse Fly on African Development”. In: *American Economic Review* 105.1, pp. 382–410.

Andersen, Thomas Barnebeck, Carl-Johan Dalgaard, and Pablo Selaya (2016). "Climate and the Emergence of Global Income Differences". In: *Review of Economic Studies* 83, pp. 1334–1363.

Anderson, Marti J. and John Robinson (2001). "Permutation Tests for Linear Models". In: *Australian & New Zealand Journal of Statistics* 43.1, pp. 75–88.

Anderson, Robert Warren, Noel D. Johnson, and Mark Koyama (2017). "Jewish Persecutions and Weather Shocks: 1100–1800". In: *The Economic Journal* 127.602, pp. 924–958.

Arthi, Vellore (2014). "'The Dust Was Long in Setting': Human Capital and the Lasting Impact of the American Dust Bowl".

Ashraf, Quamrul and Stelios Michalopoulos (2015). "Climatic fluctuations and the diffusion of agriculture". In: *The Review of Economics and Statistics* 97.3, pp. 589–609.

Bahns, Karin (2005). "Rural-to-Urban Migration in Developing Countries - The Applicability of the Harris Todaro Model with a Special Focus on the Chinese Economy". Thesis. Baden-Wurttemberg: University of Konstanz. 237 pp.

Bai, Huzhi (2010). *An Atlas of Drought and Flood Distribution in Northwest China in the Past 500 Years(1470-2008).* Beijing Shi: Meteorological Press.

Bai, Ying and James Kai-sing Kung (2011). "Climate Shocks and Sino-nomadic Conflict". In: *Review of Economics and Statistics* 93.3, pp. 970–981.

— (2014). "The Shaping of an institutional choice: Weather Shocks, the Great Leap Famine, and Agricultural Decollectivization in China". In: *Explorations in Economic History 54* 54, pp. 1–26.

Baidu (2019). *Baidu Cloud.*

Bailey, Adrian J. (2011). "Population geographies and climate change". In: *Progress in Human Geography* 35.5, pp. 686–695.

Barrios, Salvador, Luisito Bertinelli, and Eric Strobl (1, 2006). "Climatic change and rural–urban migration: The case of sub-Saharan Africa". In: *Journal of Urban Economics* 60.3, pp. 357–371.

Baten, Joerg et al. (2010). "Explorations in Economic History Evolution of living standards and human capital in China in the 18–20th centuries: Evidences from real wages, age-heaping, and anthropometrics". In: *Explorations in Economic History* 47.3, pp. 347–359.

Beinart, William (1984). "Soil Erosion , Conservationism and Ideas about Development: A Southern African Exploration, 1900-1960". In: *Journal of Southern African Studies* 11.1, pp. 52–83.

Berman, Merrick Lex (2009). "Modeling and Visualizing Historical GIS Data". In: p. 14.

Berman, Merrick Lex, Fei Carnes, and Zhenyang Hua (2014). "China Air Pollution Study Based on Volunteer Monitoring Stations". In: p. 5.

Berman, Nicolas and Mathieu Couttenier (2015). "External Shocks, Internal Shots: The Geography of Civil Conflicts". In: *Review of Economics and Statistics* 97.4, pp. 758–776.

Bernhofen, Daniel et al. (2015). "Assessing Market (Dis) Integration in Early Modern China and Europe".

— (2017). "The Evolution of Markets in China and Western Europe on the Eve of Industrialisation".

— (2018). "Sources of Market Disintegration in 18th Century China ". In: p. 37.

Bertanha, Marinho and Petra Moser (2016). “Spatial Errors in Count Data Regressions”. In: *Journal of Econometric Methods* 5.1.

Besley, Timothy and Torsten Persson (2009). “Repression or Civil War?” In: *The American Economic Review* 99.2, pp. 292–297.

Bhattacharya, Jayanta et al. (2003). “Heat or eat? Cold-weather shocks and nutrition in poor American families”. In: *American Journal of Public Health* 93.7, pp. 1149–1154.

Bi, Shuoben et al. (2016). “Spatial characteristics analysis of drought disasters in North China during the Ming and Qing Dynasties”. In: *Natural Hazards and Earth System Sciences Discussions*, pp. 1–13.

Bleakley, Hoyt (2007). “Disease and Development: Evidence From Hookworm Eradication in the American South”. In: *The Quarterly Journal of Economics* February, pp. 73–117.

Bol, Peter K (2007). “The China Historical Geographic Information System (CHGIS) Choices Faced, Lessons Learned”. In: *Conference on Historical Maps and GIS, Nagoya University.*

Broadberry, Stephen and Bishnupriya Gupta (2006). “The early modern great divergence: wages, prices and economic development in Europe and Asia, 1500–1800”. In: *The Economic History Review* 59.1, pp. 2–31.

Brooks, C. E. P. (2007). *Climate Through the Ages*. London, England: Brooks Press.

Brueckner, Jan K. and David A. Fansler (1983). “The Economics of Urban Sprawl: Theory and Evidence on the Spatial Sizes of Cities”. In: *The Review of Economics and Statistics* 65.3, p. 479.

Buhaug, Halvard and Kristian Skrede Gleditsch (2008). “Contagion or Confusion? Why Conflicts Cluster in Space”. In: *International Studies Quarterly* 52.2, pp. 215–233.

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel (2015). “Climate and Conflict”. In: *Annual Review of Economics* 7.1, pp. 577–617.

Cai, Jiaxing (2012). “Skinnerian Model in American Sinology (in Chinese)”. In: *Journal of Chifeng College* 12.

Canay, Ivan A., Joseph P. Romano, and Azeem M. Shaikh (2017). “Randomization Tests Under an Approximate Symmetry Assumption”. In: *Econometrica* 85.3, pp. 1013–1030.

Cao, Shuji (1999). “A study on the Population of Southeast Prefectures and Counties in Hongwu Period (in Chinese)”. In: *Researches in Chinese Economic History* 1, pp. 19–33.

— (2002a). *Population History of China (in Chinese)*. Vol. 05 Qing Dynasty. Shanghai: Fudan University Press.

— (2002b). *Population History of China (in Chinese)*. Vol. 04 Ming Dynasty. Shanghai: Fudan University Press.

Carleton, T. A. and S. M. Hsiang (2016). “Social and economic impacts of climate”. In: *Science* 353.6304, aad9837–aad9837.

Carleton, W. Christopher, Dave Campbell, and Mark Collard (2021). “A reassessment of the impact of temperature change on European conflict during the second millennium CE using a bespoke Bayesian time-series model”. In: *Climatic Change* 165.1-2, p. 4.

Cartier, Carolyn (2002). “Origins and Evolution of a Geographical Idea: The Macroregion in China”. In: *Modern China* 28.1, pp. 79–112.

Caruso, Germán Daniel (2017). "The legacy of natural disasters : The intergenerational impact of 100 years of disasters in Latin America". In: *Journal of Development Economics* 127, pp. 209–233.

Cenler, Polar Resfarcll and Tile Ohio Slale (1995). "1000 years of climatic change in China:ice-core δ^{18}O evidence". In: *Annals of Glaciology* 21, pp. 189–195.

Chambru, Cedric (2019). *Do the Right Thing! Leaders, Weather Shocks and Social Conflicts in Pre-Industrial France*. Tech. rep. European Historical Economics Society, pp. 1–75.

Chaney, Eric (2013). "Revolt on the Nile: Economic Shocks, Religion, and Political Power". In: *Econometrica* 81.5, pp. 2033–2053.

Chao, Chi-Chur and Eden S. H. Yu (1993). "Land-releasing policy and welfare in a dual economy". In: *The Journal of Real Estate Finance and Economics* 6.2, pp. 133–145.

Chen, Gaohua and Zhichao Chen (2016). *Science of Historical Materials of Ancient Chinese History (in Chinese)*. Beijing: Zhonghua Book Company.

Chen, Gaoyong (1986). *The Chronology of Natural Disasters and Human Calamities in Chinese History (in Chinese)*. Vol. 1. Shanghai: Shanghai Bookstore.

Chen, Jiaqi (1987). "An approach to the Data Processing of Historical Climate Materials on the Basis of Floods and Droughts of Tai Hu Basin". In: *ACTA Geographica Sinica* 54.3, pp. 231–242.

Chen, Qiang (2015). "Climate Shocks, State Capacity and Peasant Uprisings in North China during 25-1911". In: *Economica* 82.326, pp. 295–318.

Chen, Rao (2016). "Eastern Jianhuai Urban Development History Research". Thesis. Southeast University.

Cheng, Yinong (2007). “City Size and City Administrative Level in the Qing Dynasty (in Chinese)”. In: *Journal of Yangzhou University* 3, pp. 124–128.

China Meteorological Administration Institute of Meteorology (1981). *The Atlas of Wetness and Dryness Distribution in China in the Past Five Hundred Years (in Chinese)*. Vol. 1. Beijing: Map Publishing House.

Christian, Cornelius and Liam Elbourne (2017). “Droughts of Dismay : Rainfall and Assassinations in Ancient Rome”.

Christian, Cornelius and James Fenske (2015). “Economic Shocks and Unrest in French West Africa”.

Chu, Ko-Chen (1926). “Climatic Pulsations during Historic Time in China”. In: *Geographical Research* 16.2, pp. 274–282.

— (1973). “A PRELIMINARY STUDY ON THE CLIMATIC FLUCTUATIONS DURING THE LAST 5,000 YEARS IN CHINA”. In: *Scientia Sinica* 16.2, pp. 226–256.

Ciccone, Antonio (2011). “Economic Shocks and Civil Conflict: A Comment”. In: *American Economic Journal: Applied Economics* 3.4, pp. 215–227.

Cliff, A. D. and Keith Ord (1967). “The problem of Spatial autocorrelation”. In: *London Papers in Regional Science 1, Studies in Regional Science*. Ed. by A. J. Scott. London: Pion, pp. 25–55.

Collier, Paul and Anke Hoeffler (1998). “On Economic Causes of Civil War”. In: *Oxford Economic Papers* 50.4, pp. 563–573.

Compilation Group of Chinese military History, ed. (2003). *Chronology of Wars in China (in Chinese)*. Vol. 1. Beijing: Chinese people’s Liberation Army Press.

Crosby, A W (1972). *The Columbian Exchange: Biological and social consequences of 1492*. Greenwood Press.

De Groot, Olaf J. (2010). "THE SPILLOVER EFFECTS OF CONFLICT ON ECONOMIC GROWTH IN NEIGHBOURING COUNTRIES IN AFRICA". In: *Defence and Peace Economics* 21.2, pp. 149–164.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken (2014). "What Do We Learn from the Weather? The New Climate-Economy Literature". In: *Journal of Economic Literature* 52.3, pp. 740–798.

Deng, Kent (2015). "China's Population Expansion and Its Causes during the Qing Period, 1644–1911". In: pp. 1–55.

Deng, Kent and Patrick Karl O'Brien (2016). "China's GDP Per Capita from the Han Dynasty to Communist Times".

Deng, Yunte (2011). *China's Famine Relief History*. Vol. 1. Beijing: The Commercial Press.

Duan, Wei (2013). "A preliminary study on the Water Environment of towns in the South of the Yangtze River (in Chinese)". In: *Researches on the Geography and Environment of Cities and Towns in the Yangtze River Delta since the Ming and Qing Dynasties (in Chinese)*. Ed. by Yilin Zou. Shanghai: The Commercial Press, pp. 348–373.

Edwards, Ronald A. (2013). "Redefining Industrial Revolution : Song China and England".

Elhorst, J. Paul (2001). "Dynamic Models in Space and Time". In: *Geographical Analysis* 33.2, pp. 119–140.

— (2014). *Spatial Econometrics*. SpringerBriefs in Regional Science. Berlin, Heidelberg: Springer Berlin Heidelberg.

Ellerman, A. Denny and Barbara K. Buchner (2007). “The European Union Emissions Trading Scheme: Origins, Allocation, and Early Results”. In: *Review of Environmental Economics and Policy* 1.1, pp. 66–87.

Elvin, Mark (2006). *The Retreat of the Elephants: An Environmental History of China*. Illustrated edition. New Haven: Yale University Press.

Elvin, Mark and Ts’ui-jung Liu (1998). *Sediments of Time: Environment and Society in Chinese History*. Cambridge ; New York: Cambridge University Press.

Fairbank, John K., ed. (1978). *The Cambridge History of China: Volume 10: Late Ch’ing 1800–1911*. Vol. 10. The Cambridge History of China. Cambridge: Cambridge University Press.

Fairbank, John K. and Kwang-Ching Liu, eds. (1980). *The Cambridge History of China: Volume 11: Late Ch’ing, 1800–1911*. Vol. 11. The Cambridge History of China. Cambridge: Cambridge University Press.

CHGIS, Version: 6 (2016). Tech. rep. Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University.

Fairhead, James and Melissa Leach (1996). *Misreading the African Landscape: Society and Ecology in a Forest-Savanna Mosaic*. Cambridge: Cambridge University Press.

Fan, Ka-wai (2010). “Climatic change and dynastic cycles in Chinese history: a review essay”. In: *Climatic Change* 101.3-4, pp. 565–573.

Fang, Xiuqi et al. (2019). “Cropland cover change and its environmental impacts in the history of China (in Chinese)”. In: *Journal of Palaeogeography* 1, pp. 160–174.

Feng, Yi, Jacek Kugler, and Paul J. Zak (2002). "Population Growth, Urbanisation and the Role of Government in China: A Political Economic Model of Demographic Change". In: *Urban Studies* 39.12, pp. 2329–2343.

Feng, Zhijie and Huoqi Yan (2019). "The Mulberry Leaves Divination at Clear and Bright and Mulberry Leaves' Markets in Jiangnan in Ming and Qing". In: *Agricultural History of China* 3, pp. 96–103.

Fenske, James (2013). "Does Land Abundance Explain African Institutions?" In: *The Economic Journal* 123.December, pp. 1363–1390.

— (2014). "Ecology, Trade, and States in Pre-colonial Africa". In: *Journal European Economic Association* 12.3, pp. 612–640.

Fields, Gary S. (1, 1975). "Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs". In: *Journal of Development Economics* 2.2, pp. 165–187.

Fischer, Manfred M., Thomas Scherngell, and Eva Jansenberger (2006). "The Geography of Knowledge Spillovers Between High-Technology Firms in Europe: Evidence from a Spatial Interaction Modeling Perspective". In: *Geographical Analysis* 38.3, pp. 288–309.

Fischer, Manfred M., Thomas Scherngell, and Martin Reismann (2009). "Knowledge Spillovers and Total Factor Productivity: Evidence Using a Spatial Panel Data Model". In: *Geographical Analysis* 41.2, pp. 204–220.

Fishback, Price V et al. (2011). "Information and the Impact of Climate and Weather on Mortality Rates during the Great Depression". In: *The Economics of Climate Change: Adaptations Past and Present*. Ed. by Gary D. Libecap and Richard H. Steckel. May. Chicago: University of Chicago Press, pp. 131–167.

Fiszbein, Martin (2017). “Agricultural Diversity , Structural Change and Long-run Development : Evidence from the U.S.”

Gazetteer Museum of Zhejiang Province, ed. (1900). *Revision of Zhejian Gazetteer Manuscript: Volume 12 (in Chinese)*. Vol. 12. Local Chronicle Publishing House.

Ge, Jianxiong, Song Wu, and Shuji Cao (1997). *Immigration History of China (in Chinese)*. Fuzhou: Fujian People’s Publishing House.

Ge, Q.-S, Jingyun Zheng, Zhixin Hao, X.-M Shao, et al. (2010). “Temperature variation through 2000 years in China: An uncertainty analysis of reconstruction and regional difference”. In: *Geophysical Research Letters - GEOPHYS RES LETT* 37.

Ge, Q.-S., J.-Y. Zheng, Z.-X. Hao, P.-Y. Zhang, et al. (2005). “Reconstruction of Historical Climate in China: High-Resolution Precipitation Data from Qing Dynasty Archives”. In: *Bulletin of the American Meteorological Society* 86.5, pp. 671–680.

Ge, Quansheng, Shaowu Wang, et al. (2007). “Temperature and precipitation changes in China during the Holocene”. In: *Advances in Atmospheric Sciences* 24.6, pp. 1024–1036.

Ge, Shasha (2012). “A Case Study of Gansu Donation and Supervision of Corruption during Qianlong Period”. In: *Journal of Lanzhou University*.

Gleditsch, Kristian Skrede (2007). “Transnational Dimensions of Civil War”. In: *Journal of Peace Research* 44.3, pp. 293–309.

Gooch, Elizabeth (2017). “Estimating the Long-Term Impact of the Great Chinese Famine (1959–61) on Modern China”. In: *World Development* 89, pp. 140–151.

Gray, Clark and Valerie Mueller (2012). "Drought and Population Mobility in Rural Ethiopia". In: *World Development* 40.1, pp. 134–145.

Griffith, Daniel (2010). "Toward a Theory of Spatial Statistics". In: *Geographical Analysis* 12, pp. 325–339.

Han, Zhaoqing (2016). "Historical geography and environmental history in China". In: *Journal of Chinese Studies* 1.4, pp. 1–8.

Hansen, Zeynep K and Gary D Libecap (2004). "Small Farms , Externalities , and the Dust Bowl of the 1930s". In: *Journal of Political Economy* 112.3, pp. 665–694.

Hao, Zhixin, Quansheng Ge, and Jingyun Zheng (2010). "Variations of extreme drought/ flood events over eastern China during the past 2000 years". In: *Climatic and Environmental Research (in Chinese)* 15.4, pp. 388–394.

Harari, Mariaflavia and Eliana La Ferrara (2018). "Conflict, Climate, and Cells: A Disaggregated Analysis". In: *The Review of Economics and Statistics* 100.4, pp. 594–608.

Harris, John R and Michael P Todaro (1970). "Migration, Unemployment and Development: A Two-Sector Analysis". In: *American Economic Review* 60.1, pp. 126–143.

He, FN, Ke Li, and HL Liu (2010). "The influence of historical climate change on agriculture in ancient China". In: *Geographical Research* 29.12, pp. 2289–2297.

History Group, Earthquake Working Committee, Chinese Academy of Sciences (1956). *Chronological Table of China Earthquake Data*. Science Press.

Ho, Ping-ti (1959). *Studies on the Population of China, 1368-1953*. 1st. Cambridge, Mass: Harvard University Press.

Hornbeck, Richard (2012). “The Enduring Impact of the American Dust Bowl: Short-and Long-Run Adjustments to Environmental Catastrophe”. In: *American Economic Review* 102.4, pp. 1477–1507.

Hornbeck, Richard and Suresh Naidu (2014). “When the Levee Breaks : Black Migration and Economic Development in the American South”. In: *American Economic Review* 104.3, pp. 963–990.

Hsiang, S. M. (2010). “Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America”. In: *Proceedings of the National Academy of Sciences* 107.35, pp. 15367–15372.

Hsiang, S. M., M. Burke, and E. Miguel (2013). “Quantifying the Influence of Climate on Human Conflict”. In: *Science* 341.6151, pp. 1235367–1235367.

Hsiang, Solomon M., Kyle C. Meng, and Mark A. Cane (2011). “Civil conflicts are associated with the global climate”. In: *Nature* 476.7361, pp. 438–441.

Hu, Peng (2017). “Study on Grain Market Integratiion in North China during Mid-Qing Dynasty (1776-1840)”. PhD Thesis. China Agricultural University.

Hu, Weihe and Hongjun Lu, eds. (2015). *Third Version Supplement for Gaoyou State Annals (in Chinese).* Vol. 7. Chinese Annals Series. Guangling Bookstore.

Huang, Philip C. C. (1990). *The Peasant Family and Rural Development in the Yangzi Delta, 1350-1988*. Stanford: Stanford University Press. 440 pp.

Huang, Philip c. c. (2002). “Development or involution in eighteenth-century Britain and China ? A Review of Kenneth Pomeranz’s The Great Divergence: China, Europe, and the Making of the Modern World Economy”. In: *The Journal of Asian Studies* 61.2, pp. 501–538.

Huang, Shicheng and Jiaji Mi (1711). *Qianlong Jishui County Annals (in Chinese).* Vol. 40. Nanchang: Jiangxi Provincial Local Records Compilation Committee Office.

Hui, Eddie C.M. and Cong Liang (2016). "Spatial spillover effect of urban landscape views on property price". In: *Applied Geography* 72, pp. 26–35.

Institute of Geographic Sciences and Natural Resources of CAS (2017). *Division of Agricultural Regions.*

Iyigun, Murat, Nathan Nunn, and Nancy Qian (2017). *Winter is Coming: The Long-Run Effects of Climate Change on Conflict, 1400-1900.* Tech. rep. w23033. Cambridge, MA: National Bureau of Economic Research, w23033.

James, Fenske and Namrata Kala (2017). "Environmental Economic History". In: *The Long Economic and Political Shadow of History.*

Jensen, Christa D. and Donald J. Lacombe (2012). "A note on partitioning effects estimates over space". In: *Letters in Spatial and Resource Sciences* 5.1, pp. 47–53.

Jia, Ruixue (2014). "Weather Shocks, Sweet Potatoes and Peasant Revolts in Historical China". In: *The Economic Journal* 124.575, pp. 92–118.

Kakamu, Kazuhiko (2005). "Bayesian Estimation of A Distance Functional Weight Matrix Model". In: *Economics Bulletin* 3.57, pp. 1–6.

Kasahara, Hiroyuki and Bingjing Li (2017). "The Causes of China's Great Famine, 1959-1961: County-Level Evidence".

Kelejian, Harry H. and Ingmar R. Prucha (2010). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances". In: *Journal of Econometrics* 157.1, pp. 53–67.

Keller, Edmond J. (1992). "Drought, War, and the Politics of Famine in Ethiopia and Eritrea". In: *The Journal of Modern African Studies* 30.4, pp. 609–624.

Kennedy, Peter E. (1995). "Randomization Tests in Econometrics". In: *Journal of Business & Economic Statistics* 13.1, pp. 85–94.

König, Michael D. et al. (2017). "Networks in Conflict: Theory and Evidence From the Great War of Africa". In: *Econometrica* 85.4, pp. 1093–1132.

Kung, James Kai-sing and Chicheng Ma (2014). "Can cultural norms reduce conflicts? Confucianism and peasant rebellions in Qing China". In: *Journal of Development Economics* 111, pp. 132–149.

Lan, Yu, Zhixin Hao, and Jingyun Zheng (2015). "Variations of Monthly Precipitation in Rainy Season of Beijing since AD1724 Derived from the "Clear and Rain Records"". In: *Journal of Chinese Historical Geography* 30.2, pp. 41–55.

Lee, Harry F., David D. Zhang, et al. (2017). "Quantitative analysis of the impact of droughts and floods on internal wars in China over the last 500 years". In: *Science China Earth Sciences* 60.12, pp. 2078–2088.

Lee, James, Cameron Campbell, and Wang Feng (2002). "Positive Check or Chinese Checks ?" In: *The Journal of Asian Studies* 61.2, pp. 591–607.

LeSage, James and R Kelley Pace (2009). "Introduction to Spatial Econometrics". In: p. 331.

Lewis, W. Arthur (1954). "Economic Development with Unlimited Supplies of Labour". In: *The Manchester School* 22.2, pp. 139–191.

Li, Bozhong (1998). "Changes in Climate, Land, and Human Efforts: The Production of Wet-Field Rice in Jiangnan during the Ming and Qing Dynasties". In: *Sediments of Time: Environment and Society in Chineses History*. Ed.

by Mark Elvin and Ts'ui-jung Liu. Cambridge University Press, pp. 447–486.

Li, Bozhong (2000). *Early Industrialization of Jiangnan (1550-1850) (in Chinese)*. version 1. Beijing: Social Sciences Academic Press.

— (2001). "The British Model, the Jiangnan (South of the Yangtze River) Method and the Seeds of Capitalism". In: *Historical Recearch* 1.

— (2007). "The Daoguang Depression and the 1823 Flood——Economic Decline, Climatic Cataclysm and the Nineteenth-century Crisis in Songjiang". In: *Journal of Social Sciences* 6, pp. 173–178.

Li, Bozhong and Jan Luiten Van Zanden (2012). "Before the Great Divergence? Comparing the Yangzi Delta and the Netherlands at the Beginning of the Nineteenth Century". In: *The Journal of Economic History* 72.4, pp. 956–989.

Li, Dan and Nan Li (2017). "Moving to the right place at the right time: Economic effects on migrants of the Manchuria Plague of 1910−11". In: *Explorations in Economic History* 63, pp. 91–106.

Li, Hu, Chengbang An, et al. (2015). "Population history and its relationship with climate change on the Chinese Loess Plateau during the past 10,000 years". In: *The Holocene* 25.7, pp. 1144–1152.

Li, Jingming (2016a). ""Yu Bian Ji Lue" and the Peasant War at the End of Ming Dynasty". In: *Journal of Binzhou University* 32.1, pp. 65–69.

Li, Miao, Jun Xia, and Dejuan Meng (2012). "Long-Term Trend Analysis of Seasonal Precipitation for Beijing, China". In: *Journal of Resources and Ecology* 3.1, pp. 64–72.

Li, Mingkui (2017). "Retrospection and Reflection on the Study on the Study of Historical Data of Environmental History in China". In: *Journal of Poyang Lake* 4, pp. 101–116.

Li, Nan and Chu Lin (2015). "Rebellion on Population Loss in Modern China: An Empirical Analysis Based on Historical Natural Experiment". In: *China Economic Quarterly* 14.4, pp. 1325–1346.

Li, Wei and Kanglin Wu, eds. (1872). *Tonzhi Lu'an State Annals (in Chinese)*. Vol. 55.

Li, Wenhai, Xiao Cheng, et al. (2020). *Ten Famines in Modern China (in Chinese)*. Beijing: People's Publishing House.

Li, Wenhai, Mingfang Xia, and Hu Zhu, eds. (2010). *Chinese Famine Policy Book Integration (in Chinese)*. Tianjin: Tianjin Ancient Books Publishing House.

Li, Wenzhi, ed. (2016b). *Chinese Modern Agricultural History Data: 1st Series (1840-1911)*. China's Modern Economic History Reference Series. Beijing: Science Press.

Li, Xiaochun and Jing Zhou (1, 2015). "Environmental effects of remittance of rural–urban migrant". In: *Economic Modelling* 47, pp. 174–179.

Li, Xiaochun and Yu Zhou (2013). "An economic analysis of remittance of unskilled migration on skilled–unskilled wage inequality in labor host region". In: *Economic Modelling* 33, pp. 428–432.

Li, Yang (2016c). "Analysis on the Data of Cultivated Land Area in China and Britain from 14th to 19th Century". PhD Thesis. Northeast Normal University.

Li, Yao, Jianqing Ruan, and Chunhui Ye (2018). "How natural disasters affect the evolution of grain markets: evidence from 18th-century China". In: *Applied Economics* 50.45, pp. 4901–4911.

Liang, Fangzhong (1980). *China's Previous Households, Lands, Land Tax Statistics (in Chinese)*. Shanghai: Shanghai People's Publishing House.

Library of Congress (2002). *The Evolution of the Conservation Movement, 1850-1920*.

Lin, Qing (2011). "Merchant, Trade Routes and Regional Vicissitude: The Study of Individual Case Oriented to Poyang Since Ming & Qing Dynasties". Dissertation. Jiangxi Normal University.

Liu, Shiji (1987). *Research on Jiangnan Towns in the Ming and Qing Dynasties (in Chinese)*. China Social Sciences Press.

— (2013). "A Brief Discussion on the Rise and Fall of Urban in the Lower Reaches of the Yangtze River in the Qing Dynasty (in Chinese)". In: *Study on Urban Geography and Environment in the Yangtze River Delta since the Ming and Qing Dynasties (in Chineses*. Ed. by Yilin Zou. Vol. 1. Beijing: Commercial Press.

Liu, Wei et al. (2006). "INFLUENCE OF COLD CLIMATE CONDITIONS ON SOCIAL DEVELOPMENT DURING THE MING AND QING DYNASTIES IN GUANGDONG AREA". In: *Journal of South China Normal University (Natural Science Edition)* 3.

Lucas Jr., Robert E. (2004). "Life Earnings and Rural-Urban Migration". In: *Journal of Political Economy* 112 (S1), S29–S59.

Luo, ChenZheng (2006). *The Great Flood of Chinese History (in Chinese)*. China Bookstore.

Ma, Debin (2011). "Rock, Scissors, Paper: the Problem of Incentives and Information in Traditional Chinese State and the Origin of Great Divergence". London.

Ma, Debin and Jared Rubin (2017). "The Paradox of Power: Understanding Fiscal Capacity in Imperial China and Absolutist Regimes". In: *SSRN Electronic Journal*.

Mach, Katharine J. et al. (2019). "Climate as a risk factor for armed conflict". In: *Nature* 571.7764, pp. 193–197.

Malthus, T. R., Donald Winch, and Patricia James (1798). *An essay on the principle of population, or, A view of its past and present effects on human happiness: with an inquiry into our prospects respecting the future removal or mitigation of the evils which it occasions.* Cambridge texts in the history of political thought. Cambridge [England] ; New York, NY, USA: Cambridge University Press.

Marks, Robert B. (1998). "'It Never Used to Snow': Climatic Variability and Harvest Yields in Late-Imperial South China, 1650-1850". In: *Sediments of Time: Environment and Society in Chineses History*. Ed. by Mark Elvin and Ts'ui-jung Liu. Cambridge University Press, pp. 411–446.

— (2004). *Tigers, Rice, Silk and Silt - Environment and Economy in Late Imperial South China Studies in Environment and History*. Cambridge-New York-Port Melbourne-Madrid: Cambridge University Press.

— (2011). *China: Its Environment and History*. Lanham, Md: Rowman & Littlefield Publishers.

McNeill, J.R. (1998). "China's Environmental History in World Perspective". In: *Sediments of Time: Environment and Society in Chineses History*. Ed. by Mark Elvin and Ts'ui-jung Liu. Cambridge University Press, pp. 31–52.

Meng, Senzhu (2017). *Lecture Notes on Ming and Qing History (in Chinese)*. Beijing: Commercial Press.

Meng, Xin, Nancy Qian, and Pierre Yared (2015). "The Institutional Causes of China's Great Famine, 1959-1961". In: *Review of Economic Studies* 82.April, pp. 1568–1611.

Michalopoulos, Stelios (2012). "The Origins of Ethnolinguistic Diversity". In: *American Economic Review* 102.4, pp. 1508–1539.

Michalopoulos, Stelios, Alireza Naghavi, and Giovanni Prarolo (2016). "Islam, inequality and pre-industrial comparative development". In: *Journal of Development Economics* 120, pp. 86–98.

Miguel, Edward and Shanker Satyanath (2010). *Understanding Transitory Rainfall Shocks, Economic Growth and Civil Conflict*. Tech. rep. w16461. Cambridge, MA: National Bureau of Economic Research, w16461.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti (2004). "Economic Shocks and Civil Conflict: An Instrumental Variables Approach". In: *Journal of Political Economy* 112.4, pp. 725–753.

Millo, Giovanni (2017). "A simple randomization test for spatial correlation in the presence of common factors and serial correlation". In: *Regional Science and Urban Economics* 66, pp. 28–38.

Mohan, Preeya and Eric Strobl (2013). "The Economic Impact of Hurricanes in History : Evidence from Sugar Exports in the Caribbean from 1700 to 1960". In: *Weather, Climate, and Society* 5, pp. 5–13.

Mote, Frederick W. and Denis Twitchett, eds. (1988). *The Cambridge History of China: Volume 7: The Ming Dynasty, 1368–1644*. Vol. 7. The Cambridge History of China. Cambridge: Cambridge University Press.

Muscolino, Micah (2016). "Jiayan Zhang. *Coping with Calamity: Environmental Change and Peasant Response in Central China, 1736–1949* ." In: *The American Historical Review* 121.1, pp. 220–221.

Nash, Roderick (1972). "American Environmental History, New Teaching Frontier". In: *Pacific Historical Review* 41, pp. 363–372.

National Office for Philosophy and Social Science (2018). *Mid-term Inspection Report of the "Famine and Famine Chronology of Qing Dynasty and the Construction of Information Integrated Database"*.

Nunn, N. and N. Qian (1, 2011). "The Potato's Contribution to Population and Urbanization: Evidence From A Historical Experiment". In: *The Quarterly Journal of Economics* 126.2, pp. 593–650.

Nunn, Nathan and Diego Puga (2012). "Ruggedness: The Blessing of Bad Geography in Africa". In: *Review of Economics and Statistics* 94.1, pp. 20–36.

Ord, Keith (1975). "Estimation Methods for Models of Spatial Interaction". In: *Journal of the American Statistical Association* 70.349, pp. 120–126.

Osborne, Anne (1994). "The Local Politics of Land Reclamation in the Lower Yangzi Highlands". In: *Late Imperial China* 15.1, pp. 1–46.

Pan, Mingbo (2012). "A Summary of Researches on China's Environmental History in 2010". In: *Chinese History Research Developments* 1, pp. 48–58.

Pan, Wei and Zhimin Man (2013). "Reform of Jiangnan Water Network on the Eve of Urbanization". In: *Study on Urban Geography and Environment in* *the Yangtze River Delta since the Ming and Qing Dynasties*. Ed. by Yilin Zou. Vol. 1. Beijing: Commercial Press.

Pei, Qing, Harry F Lee, and David D Zhang (2018). “Long-term association between climate change and agriculturalists’ migration in historical China”. In: *The Holocene* 28.2, pp. 208–216.

Pei, Qing, Harry F. Lee, David D. Zhang, and Jie Fei (2019). “Climate change, state capacity and nomad–agriculturalist conflicts in Chinese history”. In: *Quaternary International* 508, pp. 36–42.

Pei, Qing, David D. Zhang, and Harry F. Lee (2016). “Contextualizing human migration in different agro-ecological zones in ancient China”. In: *Quaternary International*. Domestication East Asia 426, pp. 65–74.

Perdue, Peter C. (1982). “Official Goals and Local Interests: Water Control in the Dongting Lake Region during the Ming and Qing Periods”. In: *The Journal of Asian Studies* 41.4, pp. 747–765.

Peterson, Willard J., ed. (2016). *The Cambridge History of China: Volume 9: The Ch’ing Dynasty to 1800*. Vol. 9. The Cambridge History of China. Cambridge: Cambridge University Press.

Pierre-Etienne Will (2002). *Bureaucratie et Famine en Chine au 18e Siecle*. Version 1. Vol. 1. Nanjing: Jiangsu People’s Publishing House.

Pomeranz, Kenneth (2001). *The Great Divergence: China, Europe, and the Making of the Modern World Economy: 9*. New e. edition. Princeton and Oxford: Princeton University Press.

Qian, Xiaohong (2014). “Frontiers and Prospects of the Study of Chinese Environmental History (in Chinese)”. In: *Historical Research* 6, pp. 23–27.

Ren, Fang (2003). *Research on the Economy of Towns in the Middle Reaches of the Yangtze River in Ming and Qing Dynasty (in Chinese).* Book of Economic, Social and Cultural Changes in the Yangtze River Basin from the 15th to the 20th Century. Wuhan: Wuhan University Press.

Resource and Environment Data Center of Chinese Academy of Sciences (2015). *Data of Chinese Provincial Boundaries in 2015*.

Roberts, Jane (2004). *Environmental policy.* Routledge introductions to environment. London: Routledge.

Rosenshein, Lauren (2010). "The Local Nature of a National Epidemic: Childhood Overweight and the Accessibility of Healthy Food". PhD thesis. George Mason University.

Sandholt Jensen, Peter and Kristian Skrede Gleditsch (2009). "RAIN, GROWTH, AND CIVIL WAR: THE IMPORTANCE OF LOCATION". In: *Defence and Peace Economics* 20.5, pp. 359–372.

Sandmo, Agnar (2015). "The Early History of Environmental Economics". In: *Review of Environmental Economics and Policy* 9.1, pp. 43–63.

Schleussner, Carl-Friedrich et al. (2016). "Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized countries". In: *Proceedings of the National Academy of Sciences* 113.33, pp. 9216–9221.

Sheng, Yu and Xinpeng Xu (2019). "The productivity impact of climate change: Evidence from Australia's Millennium drought". In: *Economic Modelling* 76, pp. 182–191.

Shi, Zhihong (2015). "Estimation of agricultural production indicators in the Qing Dynasty (in Chinese)". In: *China Economic History Research* 5, pp. 5–30.

Shiue, Carol H (2017). “Human capital and fertility in Chinese clans before modern growth”. In: *Journal of Economic Growth* 22.4, pp. 351–396.

Shiue, Carol H. and Wolfgang Keller (2007). “Markets in China and Europe on the Eve of the Industrial Revolution”. In: *The American Economic Review* 97.4, pp. 1189–1216.

Skinner, G. William (1964). “Marketing and Social Structure in Rural China”. In: *Journal of Asian Studies* 24.2.

Skinner, GW, M Henderson, and ZND Yue (2007). *A note regarding the Physiographic and Socioeconomic Macroregions of China.*”.

Stark, Oded and David Levhari (1982). “On Migration and Risk in LDCs”. In: *Economic Development & Cultural Change* 31.1, p. 191.

State Seismological Bureau Institute of Geophysics and Fudan University Institute of Chinese History and Geography (1990). *Historical Seismological Atlas of China (in Chinese).* Map Publishing House.

Strobl, Eric (2012). “The economic growth impact of natural disasters in developing countries: Evidence from hurricane strikes in the Central American and Caribbean regions”. In: *Journal of Development Economics* 97.1, pp. 130–141.

Su, Lujian (2018). “Reviews of Local History Theory in 2017 (in Chinese)”. In: *Compilation of the 2018 Conference of Local History and Local History Theory (in Chinese).* Shanghai: Fudan University Press.

Sun, Jingchao (2017). “Polder Environment and Regional Society in the South of the Yangtze River-A Discussion Centered on Furongwei Area (in Chinese)”. In: *Agricultural archaeology* 4, p. 7.

Tacoli, Cecilia (1, 1998). “Rural-urban interactions: a guide to the literature”. In: *Environment and Urbanization* 10.1, pp. 147–166.

Tan, Liangcheng et al. (1, 2014). “Past climate studies in China during the last 2000 years from historical documents”. In: *Journal of Earth Environment* 5, pp. 434–440.

The COW Typology of War (2021). *The COW Typology of War: Defining and Categorizing Wars —Correlates of War*. File.

Todaro, Michael P. (1969). “A Model of Labor Migration and Urban Unemployment in Less Developed Countries”. In: *The American Economic Review* 59.1, pp. 138–148.

Tol, Richard S. J. and Sebastian Wagner (2010). “Climate change and violent conflict in Europe over the last millennium”. In: *Climatic Change* 99.1-2, pp. 65–79.

Turchin, P. et al. (2013). “War, space, and the evolution of Old World complex societies”. In: *Proceedings of the National Academy of Sciences* 110.41, pp. 16384–16389.

Turner, Frederick Jackson (1921). *The Frontier in American History*. New York, NY: Henry Holt and Company.

Twitchett, Denis C. and Frederick W. Mote (1998). *The Cambridge History of China: Volume 8: The Ming Dynasty*. Vol. 8. The Cambridge History of China. Cambridge: Cambridge University Press.

Vega, Solmaria Halleck and J Paul Elhorst (2013). “On spatial econometric models, spillover effects, and W”. In: European Regional Science Association, p. 28.

Vries, Jacob J de, Peter Nijkamp, and Piet Rietveld (2009). “Exponential or Power Distance-Decay for Commuting? An Alternative Specification”. In: *Environment and Planning A: Economy and Space* 41.2, pp. 461–480.

Wang, Anhui, Minshou Xie, and Liming Wen (1983). *China Earthquake Historical Data*. Ed. by China Earthquake Historical Data Editorial Board. Science Press.

Wang, Lihua (2013). “Fact Discovery and Fact Identification of Ecological History (in Chinese)”. In: *Historical Research* 3.

Wang, Tingyuan (1995). “On the Formation and Development of Huizhou Business Group (in Chinese)”. In: *Chinese History Research* 3, pp. 39–46.

Wang, Xiaoqing, Jian Liu, and Zhiyuan Wang (2015). “Comparison of Simulated and Reconstructed Temperature in China during the past 2000 Years”. In: *Advances in Earth Science* 30.12, pp. 1318–1327.

Wang, Xunming, Fahu Chen, et al. (2010). “Climate, Desertification, and the Rise and Collapse of China’s Historical Dynasties”. In: *Human Ecology* 38.1, pp. 157–172.

Wang, Yanhong (2014). “Floods and Droughts in Wan-jiang Rural Basin and Corresponding Coping Projects During Ming and Qing Dynasties”. Thesis. Anhui Normal University.

Wang, Yishi (1988). “The Memorial of the Governor of Fengyang Fu Wan Yishi for the Statement of Abnormal Flood in June, 1649 (in Chinese antient writtings)”. In: *Historical Archive* 4.

Webb, Walter Prescott (1931). *The Great Plains: a study in institutions and environment*. Ginn and Company.

Wei, Fengying (2007). "Probability Distribution and Prediction of Extreme Precipitation in Beijing (in Chinese)". In: *Paper collection of climate credit conference of 2007 annual meeting of China Meteorological Society (in Chinese)*. China Academic Journal Electronic Publishing House., pp. 247–256.

Wei, Gu and Shigemi Yabuuchi (2003). "Land and Urban Unemployment". In: *Pacific Economic Review* 8.1, pp. 71–77.

Wen, Kegang, ed. (2006). *China Meteorological Disaster Dictionary (in Chinese)*. Beijing: China Meteorological Press.

Wen, Kegang and Yihui Ding, eds. (2008). *China Meteorological Disaster Dictionary (in Chinese)*. Vol. 1. Beijing: China Meteorological Press.

Williams, Michael (1994). "The relationship of environmental history and historical geography". In: *Journal of Historical Geography* 20.1, pp. 3–21.

Winkler, Anderson M. et al. (2014). "Permutation inference for the general linear model". In: *NeuroImage* 92, pp. 381–397.

Wong, R. Bin et al. (1991). "Chinese Traditions of Grain Storage". In: *Nourish the People*. The State Civilian Granary System in China, 1650–1850. University of Michigan Press, pp. 1–16.

Worster, D (2008). *The Ends of the Earth: Perspectives on Modern Environmental History*. First Edition edition. Cambridge England ; New York: Cambridge University Press.

Xia, Mingfang (2010). "Droughts in Chinese history and their causes". In: *Guangming Daily*.

— (2015). "Big Data and Ecological History: The Compilation and Database Construction of the Historical Sources of Chinese Disaster History in an Information Age". In: *The Qing History Journal* 2, pp. 67–82.

Xiao, Jie et al. (2018). “Climate change and social response during the heyday of the little ice age in the Ming and Qing dynasties”. In: *J. Arid Land Resour. Environ* 32, pp. 79–84.

Xu, Ronghan, Shijin Xu, and Shaochun Dong (2012). “Design and implementation of historical natural disasters database based on GIS”. In: *Science of Surveying and Mapping* 37.1, pp. 85–88.

Xu, Yi, Bas van Leeuwen, and Jan Luiten van Zanden (2018). “Urbanization in China, ca. 1100–1900”. In: *Frontiers of Economics in China* 13.3, p. 322.

Xu, Zhengrong (2014). “Review of Data about the Environmetal History of China”. In: *Journal of Baoshan University* 33.6, pp. 18–25.

Yang, Bao, Achim Braeuning, et al. (2002). “General characteristics of temperature variation in China during the last two millennia: GENERAL CHARACTERISTICS OF TEMPERATURE VARIATION IN CHINA”. In: *Geophysical Research Letters* 29.9, pp. 38–1–38–4.

Yang, Yida, Meisu Wang, and Zhimin Man (2009). “Progress in the Research Methods of China’s Historical Climate in the Past 30 Years—Centering on Documentary Data”. In: *Theories of Chinese History and Geography* 2.

Yang, Yuda, Zhimin Man, and Jingyun Zheng (2006). “The Grade Reconstruction of 1711～1911 Precipitation during the Rainy Season in Kunming and its Preliminary Study: a Exploitation of Archives”. In: *Geographical Research* 25.6, pp. 1041–1049.

Yoshinobu, Shiba (1998). “Environment versus Water Control: The Case of Southern Hangzhou Bay Area from the Mid-Tang Through the Qing.” In: *Sediments of Time: Environment and Society in Chineses History*. Ed. by Mark Elvin and Ts’ui-jung Liu. Cambridge University Press, pp. 108–135.

Yu, Weihua, Xinxing Dong, and Ming Lei (2015). "Climate Changes and Wars, the Rise and Fall of Dynasties (in Chinese)". In: *Dong Yue Tribune* 36.9, pp. 81–86.

Yue, Ricci P.H. and Harry F. Lee (2020). "Examining the Direct and Indirect Effects of Climatic Variables on Plague Dynamics". In: *Atmosphere* 11.4, p. 388.

Zeng, Qinghua (2006). *China Meteorological Disaster Dictionary: Volume of Hunan*. Ed. by Kegang Wen. Beijing: China Meteorological Press.

Zhang, Chunwang (2004a). "Research on Interactive Relation between Natural Disasters, Society and Economy in Jianghuai Region in Ming and Qing Dynasties". Thesis. Xiamen: Xiamen University.

Zhang, David D, Jane Zhang, et al. (2007). "Climate Change and War Frequency in Eastern China over the Last Millennium". In: *Human Ecology* 35, pp. 403–414.

Zhang, David D., C. Y. Jim, et al. (2006). "Climatic Change, Wars and Dynastic Cycles in China Over the Last Millennium". In: *Climatic Change* 76.3-4, pp. 459–477.

Zhang, De'er (1998). "Paleoenvironmental Records from Chinese Historical Documents". In: *Advance in Earth Sciences* 13.3, pp. 273–277.

— ed. (2004b). *Chinese Three Thousand Years Meteorological Records Collections (in Chinese)*. Nanjing: Phoenix Press.

— (2005). "The Collection of Chinese Historical Climate Document Records and its latest Application". In: *Scientific Guide* 23.8, pp. 17–19.

Zhang, De'er and Yuewei Liu (2002). "A New Approach to the Reconstruction of Temporal Rainfall Sequences from 1724~1904 Qing Dynasty Weather Records for Beijing". In: *Quaternary Sciences* 22.3, pp. 199–208.

Zhang, Deer (1984). "Synoptic-climatic Studies of Dust Fall in China since Historic Times". In: *Science in China Series B-Chemistry, Biological, Agricultural, Medical & Earth Sciences* 27.8, pp. 825–836.

Zhang, Dian, Zhiyong Zhan, et al. (2004). "Climate change and China' s wars, social unrest and dynasties （in Chinese）". In: *Chinese Science Bulletin (in Chinese)* 49.23, pp. 2468–2474.

Zhang, Haifang (2007). "Discussion on the relationship between the diversion of the Yellow River in 1855 and the economic and social development of Shandong". In: *Journal of the Graduate School of the Chinese Academy of Social Sciences* 6, pp. 55–61.

Zhang, Jiacheng and Thomas J. Crowley (1989). "Historical Climate Records in China and Reconstruction of Past Climates". In: *Journal of Climate*, pp. 833–849.

Zhang, Kevin Honglin and Shunfeng Song (2003). "Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses". In: *China Economic Review* 14.4, pp. 386–400.

Zhang, Piyuan, ed. (1996). *China's climate and sea level changes and their trends and impacts: Volume 1-China's historical climate change.* Vol. 1. Qingdao: Shandong Science and Technology Press.

Zhang, Piyuan, Quansheng Ge, and Jingyun Zheng (2002). "The Research to Extract Natural Environment Information from Historical Documents and Archives." In: *Proceedings of the Symposium on the Relationship between*

Heaven and Human in Chinese Economic History 1999. Beijing: China Agriculture Press.

Zhang, Quan (2017). "The Study of Historical Changes in term of Water Environment in Shaoxin District from Ming and Qing Dynasty". Thesis. Hangzhou: Zhejiang University.

Zhang, Zhibin, Huidong Tian, et al. (2010). "Periodic climate cooling enhanced natural disasters and wars in China during AD 10–1900". In: *Proceedings of the Royal Society B: Biological Sciences* 277.1701, pp. 3745–3753.

Zhao, Hongjun (2012). "Did Climate Change Affect the Social Stability of Chinese Agrarian Economy in the Past 2000 Years? A Positive Analysis Based on Paleo-climatic Reconstruction Data and Historic Data". In: *China Economic Quarterly* 11.2, pp. 691–722.

Zhao, Jiyuan and Ying Li (2008). *Taihu County Annals (in Chinese)*. Vol. 40. Huangshan: Huangshan Book Club.

Zheng, Jingyun, Quansheng Ge, Xiuqi Fang, et al. (2007). "Comparison on Temperature series reconstruced from Historical Documents in China for the Last 2000 Years". In: *ACTA Meteorologica Sinica* 65.3, pp. 428–439.

Zheng, Jingyun, Quansheng Ge, Zhixin Hao, et al. (2014). "Paleoclimatology proxy recorded in historical documents and method for reconstruction on climate change". In: *Quaternary Sciences* 34.6, pp. 1186–1196.

Zheng, Jingyun, Zhixin Hao, and Xiaochun Di (2002). "A Study on the Establishment and Application of Environmental Change Database during Historical Times". In: *Geographical Research* 21.2, pp. 146–154.

Zheng, Jingyun, Zhixin Hao, and Quansheng Ge (2005). “Precipitation Changes in the Middle and Lower Reaches of the Yellow River in the past 300 Years”. In: *Science in China Series D-Earth Sciences* 35.8, pp. 765–774.

Zheng, Jingyun, Peiyuan Zhang, and Yufu Zhou (1993). “A New Approach of Reconstructing the Dryness/Wetness Index Eries in China during Historical Times by using the Number of Drought/Flood Counties”. In: *Geographical Research* 12.3, pp. 1–9.

Zheng, Jingyun, Piyuan Zhang, and Weimin Jian (1992). “Climate Informationization of Historical Archives in the Republic of China”. In: *Disaster Reduction in China* 3, pp. 15–18.

Zhi-yuan, ZHOU (2008). “The Nongovernmental Famine-relief Measures in Huizhou During the Ming and Qing Dynasties Reflected in the Local Chronicles of Huizhou Townships”. In: *Journal of Anhui University (Philosophy & Social Sciences)*, p. 01.

Zhou, Zhenhe, ed. (2007). *General History of Chinese Administration (in Chinese)*. Shanghai: Fudan University press.

Zou, Yilin, ed. (2013). *Research on Urban Geography and Environment in the Yangtze River Delta Region Since Ming and Qing Dynasties (in Chinese)*. Beijing: The Commercial Press.