The Emperor's Geography - City Locations, Nature, and Institutional Optimization

Christian Düben[*]   Melanie Krause[†]

November 21, 2020

Abstract

The emergence and growth of cities are shaped by both geographical features and institutional factors. We are able to analyze their interplay at different levels of the urban hierarchy by exploiting a unique data set on cities in imperial China from 221 BCE to 1911 CE, a geographically diverse empire with a long history of centralized rule. Developing a stylized theoretical model, we combine econometrics with machine learning techniques. Our results suggest that the higher a city is in the urban hierarchy, the less important are geographical compared to institutional factors. At the other end of the scale, market towns without government responsibilities are most strongly shaped by geographical characteristics. We also find evidence that many cities of political importance in imperial times still enjoy a special status nowadays, underlining the modern relevance of these historical factors.

JEL Classification: R11, R12, N95, O18
Keywords: Urban Areas, China, Geography, Economic History, Institutions

1   Introduction

What determines in which locations cities are founded and how they develop over time? The economic geography literature highlights the importance of natural characteristics, such as climatic conditions, river access, and soil quality (Krugman, 1991; Fujita et al., 1999; Henderson et al., 2018). In addition, it is obvious that political and institutional factors foster a city's status and growth, for example via (de-)centralization of governance or strategic military and inner-country trade considerations (Ades and Glaeser, 1995; Davis and Henderson, 2003; Bosker et al., 2013).

Yet how exactly geographical and institutional factors act together to explain cities' development is difficult to ascertain. It is conceivable that geographical characteristics play a larger role in determining urban location, whereas institutions later on are more decisive for their growth by conferring them political status and resources (Henderson and Wang, 2007). Alternatively, natural characteristics might be valued differently in different institutional settings; for example, it has been shown that coastal access gained precedence over road access after the demise of the Roman empire (Michaels and Rauch, 2018). The strong history dependence of city location and size, together with data availability and measurement issues, complicates the study of institutions and geography.

In this paper, we study the interplay of geographical and institutional factors at different levels of the urban hierarchy. In many regions of the world, the institutional background and the resulting urban hierarchy present too many changes over time in order to intricate the effects. By contrast, we are able to gain new insights by relying on the unique case of imperial China from the rise of the Qin dynasty in 221 BCE until the end of the Qing dynasty in 1911 CE. China is not only the world's most populous country but has one of the longest urban development histories in the world, with primitive cities first occurring along the Yellow River in northern China more than 4000 years ago (Wu et al., 2014). Compared to traditionally fractionalized Europe, China spent much of its history unified and centrally governed. When dynasties came to an end, they sometimes temporarily disintegrated into smaller rivaling regions, but later merged again into a large, centrally governed unit with the rise of the subsequent dynasty (Ko et al., 2018; Lewis, 2009). It is imperial China's centralized spatial organization controlling a vast administrative urban landscape that allows us to track urbanization in a well defined institutional framework over millennia. While the vestiges of colonialism still shape the urban structure in much of Sub-Saharan Africa (Bonfatti and Poelhekke, 2017), China experienced foreign influence of such kind to a much smaller extent. Europe engaged with China through centuries of trade, coastal concessions in the 19th and 20th centuries as well as two opium wars in the 19th century. Japan invaded China in two

Sino-Japanese wars in the late 19th and the 20th centuries. Multiple dynasties were founded by neighboring peoples, such as the Mongols (Mostern, 2011; von Glahn, 2016). Nonetheless, foreign control remained moderate compared to many other world regions. Urbanization in China can therefore largely be seen as an outcome of its own dynasties and empires. Moreover, the Chinese case allows for a persistent categorization of cities along the urban hierarchy. This provides a unique setting for examining the varying influence of geography on the location of cities of different institutional status.

In our study, we exploit data on the location of administrative cities throughout imperial Chinese history from 221 BCE until 1911 CE (Fairbank Center for Chinese Studies of Harvard University and Center for Historical Geographical Studies at Fudan University, 2016). In our theoretical framework we sketch the utility maximization problem of an emperor who has to place cities of different administrative ranks, namely county and prefecture seats, given constraints posed by local geography as well as military threats. We develop predictions about how county seats and higher-ranked prefecture seats are influenced by the interplay of geographical and institutional factors, which we then test empirically. Using both econometric regressions and random forest techniques from machine learning, we arrive at our main findings: Lower-ranked county seats are more strongly shaped by local geographical factors than higher-ranked prefecture seats. For the latter, a centrality argument to minimize transport costs and to facilitate communication across the empire is more important. We subject these results not only to numerous robustness checks in terms of model specifications, but also elaborate on them with an additional data set. For the years 1820 and 1911 CE, we also have data on market towns, which are non-administrative urban locations. In line with our theory and other findings, market towns are most strongly shaped by local geography, even more so than county seats. Finally, we show the modern relevance of these historical factors. Modern cities that once were a county or prefecture seat are on average more populous nowadays and have a higher local economic activity as proxied for by nighttime lights.

Located at the interplay between geography, institutions, and economic history, our paper relates to various strands of the literature. With our investigation on the relation between geography and urbanization, we anchor our paper in the New Economic Geography (NEG) literature, see Krugman (1991, 1993); Fujita et al. (1999) as well as Proost and Thisse (2019) and Redding (2020) for recent theoretical overviews. The predictions of the NEG literature on how first- and second-nature geography shapes the emergence of cities and the distribution of economic activity have been tested empirically by a growing number of studies. On a global level, Henderson et al. (2018), Mitton (2016) and Motamed et al. (2014), inter alia, show that favorable geographical

conditions have lead to earlier globalization and persistently higher economic activity. As continent-specific studies, Black and Henderson (2003) find that over the 20th century, American cities in warmer and drier regions - first nature - but also in regions with larger market potential in terms of big neighboring cities - second nature - grew faster than others. Bosker and Buringh (2017) compare 250,000 potential European city locations to actual settlements, finding evidence of an increasing importance of sea and river locations over time. We build on these insights and link them to the Chinese context.

In addition, we consider the political determinants of city locations, thereby relating to the institutional strand of the urban economics literature. Various political determinants of city size and status are well-known, ranging from autocracy (Ades and Glaeser, 1995; Henderson and Wang, 2007) to fiscal federalism (Davis and Henderson, 2003). Global studies such as Soo (2005) and Düben and Krause (2020) highlight the importance of political institutions in explaining the cross-country variation in the size distribution of cities.

In contrast to studies focusing either on geography or institutions, our paper adds to the small but growing number of studies that combine both perspectives. Often, these are historical case studies focusing on a particular world region. For example, Barjamovic et al. (2019) use commercial records from Assyrian merchants to estimate the location of Bronze Age cities. With the collapse of the Roman empire as a defining episode for Europe (Scheidel, 2019), Michaels and Rauch (2018) compare the diverging trajectory of French and English cities based on different geographical and institutional factors. Bosker et al. (2013) use historic city locations to investigate why the urban center of gravity moved from the Islamic world to Western Europe during the period from 800 to 1800, while Jedwab et al. (2020) analyze city growth in medieval Europe in terms of their physical geography, finding that institutions played a larger role later on. Schönhölzer and Weese (2019) analyze how administrative boundary switches affected the growth of European cities in the short and the long run, with border location and their effects on the governance of European nation states also the focus of the study by Kitamura and Lagerlöf (2020). Yet, outside of Europe - and in particular in China - much less is known on the nexus between geography and institutions in the emergence of cities. Many economic history papers on China concentrate on individual aspects, such as the threat of nomadic invasion in the North (Bai and Kung, 2011), the flooding of the Yellow River (Chen et al., 2012), tax collection across the empire (Ko et al., 2018), the effects of malaria on Chinese urbanization (Flückinger and Ludwig, 2017), and the economic rationale behind city walls (Du and Zhang, 2019; Ioannides and Zhang, 2017).[1]

---

[1]For ongoing work on the effect of provincial capital status on historic population size in imperial China, see Bai and Jia (2020).

By contrast, we exploit variation across the overall urban hierarchy, thereby investigating the effects of geography conditional on the institutional status. The unique setting of imperial China with its centralized spatial organization, exceptional institutional path dependence, and vast network of administrative cities allows us to answer this question to an extent that would not be feasible with any other part of the world.

Following the discussion of historical mechanisms, we link to the literature on contemporaneous China. By showing that the interplay of geographic and institutional factors shaped the distribution of Chinese cities in ways which still matter today, we add to the vast and increasing number of papers studying modern Chinese cities. Recent contributions have focused, inter alia, on their geography (Christensen and McCord, 2016), railway access (Li et al., 2018), and seaports (Funke and Yu, 2011). Case studies of individual cities (e.g. Zhang et al., 2013; Tan et al., 2014) or comparative studies of several large cities (e.g. Baum-Snow et al., 2017; You and Yang, 2017; Lin and Song, 2002) have provided different contemporaneous perspectives on their growth and structure. However, to the best of our knowledge, there are no papers which have turned to historical Chinese urban and geographical data to shed light on how the urban distribution came about.

This paper clearly relates to various strands of the economic literature that it draws on and extends. Nonetheless, the central foundation upon which the paper is constructed is the vast evidence compiled by historians. Our theoretical framework, the empirical strategy, and the interpretation of the results rely on an extensive collection of historical details that we primarily assemble from publications by Mostern (2011), von Glahn (2016), Major and Cook (2017), Lewis (2009), Mote (1999), Wu and Gaubatz (2013), Elvin (2004), and Wilkinson (2013). Those insights reach from a flood in 11 CE that proved to be a major contributor to the Xin dynasty's downfall (Major and Cook, 2017), over the kaizhong system of provision delivery to the frontier in exchange for salt trading privileges (von Glahn, 2016), to the link between the size of administrative regions and the ease of tax collection (Mostern, 2011), and many others.

The remainder of this paper is structured as follows. Section 2 develops a stylized model on the interaction of geography and institutions in the historical Chinese context. This will serve as the framework for the subsequent empirical analysis. Section 3 presents our data set, descriptive statistics, and the estimation strategy. Section 4 shows our empirical results using historical data. Section 5 links our historical insights to modern-day city size and status. Section 6 concludes. The Appendix contains proofs, detailed data sources as well as numerous supplementary results and robustness checks.

2 Conceptual Framework

We start by laying out the conceptual framework for our empirical analysis. To this goal, we formalize the mechanisms on the interplay of geography and institutions in the context of ancient China. While keeping the exposition stylized, we allow our simple model to capture recurrent features of Chinese history, such as the nomadic military threats in the North (Bai and Kung, 2011) and the flooding of the Yellow River (Chen et al., 2012; Elvin, 2004; Lewis, 2009).

At the core of the conceptual framework, there is the Chinese Emperor setting up administration units to maximize control over his empire given the geographical constraints. In particular, the emperor has to choose the number of county seats N_c and the number of higher-ranked prefecture seats N_p as well as their precise locations.[2] While the role of county seats is local administration and tax collection, prefecture seats are armed with military power and can counter foreign invasions as well as local upheavals (Mostern, 2011).[3]

For our formal analysis, we will consider an empire of a fixed area, consisting of N_i individual cells (pixels). P_i is the population density in grid cell i. We will assume that it depends positively on local geographical conditions A_i, comprising inter alia agricultural productivity, and negatively on a threat factor T_i. Total threats are a function of various threats, both military threats and natural hazards, for example, floods.[4]

Assumptions 1. *Population P_i in cell i is a positive function of local geographical conditions A_i and a negative function of threats T_i:*

$$\frac{\partial P_i(A_i, T_i)}{\partial A_i} > 0 \quad and \quad \frac{\partial P_i(A_i, T_i)}{\partial T_i} < 0 \tag{1}$$

where total threats are a positive function of both military threats M_i and natural hazards

---

[2]Some dynasties installed additional administrative layers on top of the county and prefecture seats, which were of a supplementary nature and did not replace the system of counties and prefectures. The Song dynasty (960 - 1279 CE) e.g. set up provincial circuits that assisted in resource transportation but did not cut the direct link between prefectures and the imperial court (Mostern, 2011).

[3]Spatial policies followed a *Persistence and Transformation* approach that justified frequent adjustments to administrative units (Mostern, 2011). The entire administrative system was optimized towards centralized control by the imperial court. County magistrates and prefects, the officials governing counties and prefectures respectively, were directly installed by the court and posted outside their home regions. This sometimes required the assistance of translators but curtailed the influence of local elites (Major and Cook, 2017; Mostern, 2011; Wilkinson, 2013). Centralized control remained an important characteristic, although it encountered some temporary and regional limitations from time to time. Especially in times of conflict, the imperial court did not always manage to maintain the system of perfect centralization and had to grant more rights to local leaders - e.g. following a rebellion during the Tang dynasty in the 870s (Mote, 1999).

[4]Nomadic invasions in the north, the in many dynasties dominant military threat, was strongly linked to changes in climatic conditions (Elvin, 2004; Bai and Kung, 2011).

H_i:

$$\frac{\partial T_i(M_i, H_i)}{\partial M_i} > 0 \quad and \quad \frac{\partial T_i(M_i, H_i)}{\partial H_i} > 0 \tag{2}$$

When maximizing his control over individual pixels in his empire, the emperor has to place both the N_c county seats and the N_P prefecture seats strategically. Each cell has to pay a lump-sum tax λ proportional to its population, but as tax collection works through the county seats, there is a transport cost associated with it from the cell to the nearest county seat. This can be modeled as an iceberg transport cost τ times the travel-time distance D_{ci} between cell i and county seat c.[5] Denoting the number of cells which are nearest to county seat c with N_{ic}, we obtain taxes collected by county seat c as

$$\sum_{i=1}^{N_{ic}} (1 - \tau D_{ci}) \lambda P_i \tag{3}$$

Summing eq. (3) over all N_c county seats gives the total taxes in the empire collected at county seats as

$$\sum_{c=1}^{N_c} \sum_{i=1}^{N_{ic}} (1 - \tau D_{ci}) \lambda P_i \tag{4}$$

In addition to total tax revenues, the emperor's control can be measured by his ability to ward off military threats. The N_p prefecture seats should thus be located so that threats in all cells can be thwarted quickly and effectively, with the reaction to a threat M_i in cell i depending on the travel-time distance D_{pi} to its nearest prefecture seat. Denoting the number of cells which are nearest to prefecture seat p with N_{ip}, we obtain the impact of military threats occurring in the realm of prefecture p as

$$\sum_{i=1}^{N_{ip}} D_{pi} M_i \tag{5}$$

Summing eq. (5) over all N_p prefecture seats gives the total impact of threats in the empire

$$\sum_{p=1}^{N_p} \sum_{i=1}^{N_{ip}} D_{pi} M_i \tag{6}$$

This sum of threats will enter the emperor's utility function with a negative sign and we will allow for a non-linearity parameter $m > 1$, as (i) responding to military threats is more time sensitive than the regular tax-based correspondence with county seats and (ii) it is primarily the major military threats and large invasions that threaten

---

[5]The transport costs capture the direct cost in terms of collecting the taxes from various cells as well as the likelihood that taxes might not be paid, which can also be assumed to increase with remoteness as measured by travel-time distance (Mostern, 2011).

the stability of the empire.

Apart from countering military threats, prefecture seats coordinate counties within their territory. The emperor places prefecture seats in positions that minimize the distance, and thereby facilitate communication, between prefecture seats and the respective counties seats:

$$\sum_{p=1}^{N_p} \sum_{c=1}^{N_{cp}} D_{pc} \tag{7}$$

We can now write down the emperor's utility function: Choose the numbers N_c and N_p of county and prefecture seats as well as their location so that the travel-time weighted distance from each cell to the nearest county, each county seat to the nearest prefecture seat and each threat-weighted cell to the nearest prefecture seat is as small as possible.[6] This will maximize the emperor's utility:

$$\max_{N_c, N_p, locations} U\Big[\underbrace{\sum_{c=1}^{N_c} \sum_{i=1}^{N_{ic}} (1-\tau D_{ci})\lambda P_i}_{+}, \underbrace{\sum_{p=1}^{N_p} \sum_{c=1}^{N_{cp}} D_{pc}}_{-}, \underbrace{\sum_{p=1}^{N_p} \sum_{i=1}^{N_{ip}} (D_{pi} M_i)^m}_{-}\Big] \tag{8}$$

Minimizing the distance between population-weighted cells and county seats, $\sum_{c=1}^{N_c} \sum_{i=1}^{N_{ic}} D_{ci} P_i$, is equivalent to maximizing the sum of taxes collected at county seats, $\sum_{c=1}^{N_c} \sum_{i=1}^{N_{ic}} (1-\tau D_{ci})\lambda P_i$. To avoid the corner solution of placing a county and a prefecture seat in each cell, so that $N_p = N_c = N_i$, we have to place some constraints on the maximization problem:

Assumptions 2. *Each county seat costs an amount G_c to administer. It also has to export an amount E_c for the administration of the higher-level prefecture seat. County seats have to balance their budgets and finance themselves from the taxes of the N_{ic} cells under its administration so that[7]*

$$G_c + E_c \leq \sum_{i=1}^{N_{ic}} (1-\tau D_{ci})\lambda P_i \qquad \forall \ c \in \{1, ..., N_c\} \tag{9}$$

$$E_c \geq 0 \qquad \forall \ c \in \{1, ..., N_c\} \tag{10}$$

---

[6]Centralized control over the vast empire had, apart from probably contributing to the emperor's happiness, important benefits for the empire. Thanks to centralized control at such a large scale, the imperial state could provide very resource-demanding, yet important public goods such as flood control and protection from northern nomadic invasions (Lewis, 2009). The local population was usually unable to maintain these public goods alone. When the Qing dynasty (1644 - 1911 CE) put water control in private hands, the Middle Yangzi suffered devastating floods (von Glahn, 2016; Perdue, 1987; Marks, 2012; Will, 1985).

[7]Historical evidence suggests that spatial planners changed county borders or merged counties rather than transferring taxes between county seats when a county's tax revenues were to small to support its own administration (Mostern, 2011).

Summing up eq. (9) at the aggregate level gives

$$\sum_{c=1}^{N_c}(G_c + E_c) \leq \sum_{c=1}^{N_c}\sum_{i=1}^{N_{ic}}(1 - \tau D_{ci})\lambda P_i \qquad (11)$$

Assumptions 3. *Each prefecture seat costs an amount G_p to administer. This money can come from the nearest county seats (minus transport costs), while also allowing cross-subsidies from other prefecture seats (minus transport costs)*[8]:

$$G_p \leq \sum_{c=1}^{N_{pc}}(E_c - \tau D_{pc}E_c) + \sum_{q \neq p}^{N_p}(E_{pq} - \tau D_{pq}E_{pq}) \qquad \forall \ p \in \{1, ..., N_p\} \qquad (12)$$

Summing up eq. (12) at the aggregate level gives

$$\sum_{p=1}^{N_p}G_p \leq \sum_{i=1}^{N_p}\sum_{i=1}^{N_{pc}}(E_c - \tau D_{pc}E_c) + \sum_{p=1}^{N_p}\sum_{q \neq p}^{N_p}(E_{pq} - \tau D_{pq}E_{pq}) \qquad (13)$$

Hence, we can formulate the emperor's task as the maximization of eq. (8) subject to the constraints eq. (9) and eq. (10) at the county level as well as eq. (12) at the prefecture level. While no analytical closed-form solution can be obtained, we can simplify the expressions under special conditions imposed on G_c and E_c, see Section A.1.

With a view to our empirical analysis, let us look now derive some predictions of the theoretical framework:

Theorem 1. *(a) In areas with favorable geography, there are more county seats.*

(b) The location of prefecture seats is determined to a lesser extent and only indirectly by the presence of geographical features.

(c) There are more prefecture seats per county seat in regions that are prone to military invasion.

(d) If the population grows at a higher rate than the costs of maintaining county seats, the optimal number of county seats will increase.
Proof: See Online Appendix A

It is the focus of our empirical analysis to take these predictions to the data.

---

[8]Resources were transported from richer, often interior to poorer, often frontier regions. Our model's inter-prefecture transfers capture this mechanism with which prefectures can balance each other's budgets. Despite high transport costs, especially the northern frontier, home an extensive military force, required substantial transfers. Many dynasties motivated merchants to transport border supplies in exchange for salt trading privileges, a system with a limited capacity. When the Ming dynasty (1368 - 1644 CE) issued to many salt trading privileges, the scheme broke down (von Glahn, 2016).

3 Data and Empirical Strategy

3.1 Data

We employ a geo-spatial data set on historic Chinese prefectures, spanning from the rise of the Qin dynasty in 221 BCE until the end of the Qing dynasty in 1911 CE (Fairbank Center for Chinese Studies of Harvard University and Center for Historical Geographical Studies at Fudan University, 2016). The China Historical Geographic Information System (CHGIS) data tracks the shape of these administrative regions and the location of politically relevant cities within them, i.e. prefecture and county seats, at a yearly frequency.[9] County seats handle local administration, including tax collection and legal disputes. Multiple counties make up a prefecture. The prefecture seat manages counties, coordinates their tax collection efforts, holds military power, handles major court cases such as capital crimes and is, unlike county seats, in direct contact with the imperial court. Some dynasties installed additional layers for various purposes, but counties and prefectures remained the foundation of imperial administration (Mostern, 2011; Major and Cook, 2017). Given that until around the 12th century cities primarily served administrative functions (Wu and Gaubatz, 2013), our data of administrative cities paints a rather complete picture of urban settlements throughout the earlier dynasties.[10] Figure 1 shows the data for the year 1350 CE. Zooming in on prefecture Jingzhou Fu in Figure 2, we can get a better idea of the location of the 14 county seats within this prefecture in addition to the prefecture seat.

While our main analysis will focus on the prefecture and county seats, CHGIS also provides information on market towns, i.e. non-administrative urban localities that increasingly played a role in the second millennium CE, cross-sectionally for 1820 and 1911 CE. We employ this data in Section 4.3, comparing the different interplay of geography and institutions for administrative and non-administrative cities.

The set of geographic variables collected for this study includes temperature (Matsuura and Willmott, 2018a), precipitation (Matsuura and Willmott, 2018b), the distance from major rivers (Natural Earth, 2019), the distance from the coast (Wessel and Smith, 1996, 2017), distance from the equator, elevation (Danielson and Gesch, 2011), ruggedness (Nunn and Puga, 2012), dominant soil type, landform, and lithology (Dijkshoorn et al., 2008). Details can be found in Online Appendix B.

---

[9]The CHGIS data covers the entire territory of Chinese empires except for the outlying provinces of Neimeng, Qinghai, Xinjiang, and Xizang.

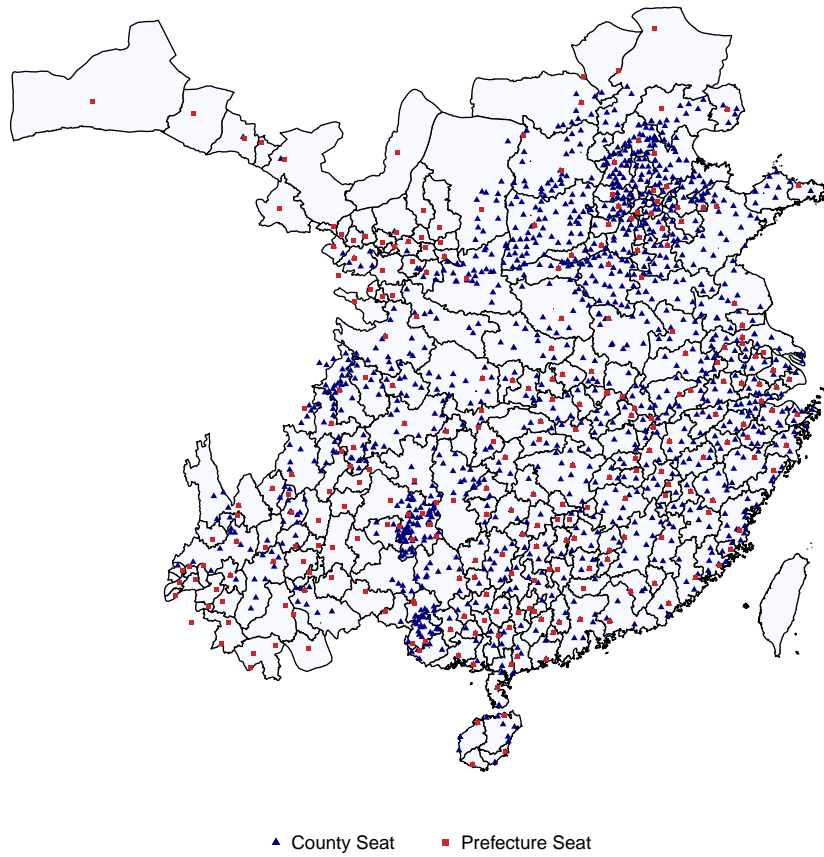[10]As Wu and Gaubatz (2013) describe, trade-oriented cities began to emerge following the Tang dynasty (618 - 907 CE).
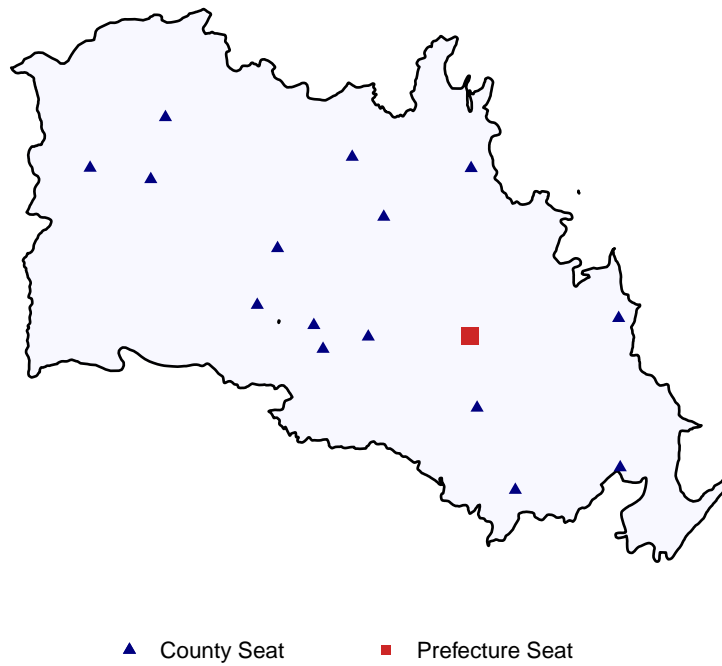
Figure 1: CHGIS: China in 1350 CE

Figure 2: CHGIS: Prefecture Jingzhou Fu in 1500 CE

For the path dependence analysis in Section 5, we supplement our data set with modern settlement locations and shapefiles derived from the GHS Urban Centre Database (Florczyk et al., 2019).[11]

3.2   Descriptive statistics

Table 1: Geography Summary Statistics

	Mean	St. Dev.	Min	Max
Distance from Coast	692.507	499.438	-1.708	2,302.794
Distance from River	129.588	121.051	0.472	701.471
Distance from Equator	3,835.773	872.772	2,012.333	5,934.338
Elevation	1,161.858	1,266.958	-0.139	5,467.496
Ruggedness	198,578.889	197,356.492	0.000	1,373,285.008
Temperature	9.834	7.335	-10.283	25.267
Precipitation	848.889	541.154	59.872	4,968.969

Notes: Distances in km, temperature in °C, precipitation in mm per year, elevation in meters, ruggedness index in millimeters as defined by Nunn and Puga (2012). Values refer to the Chinese empire's baseline shape with 86,257 pixels 7.33 x 9.51 km in size. Landform, dominant soil type, and lithology are categorical variables and summarized in Online Appendix B. See Table B-1 for details on variable generation. Variables are differently scaled in subsequent chapters to facilitate readability.

The summary statistics of the geographical variables at the cell level in Table 1 underline the large variety of local conditions. Average temperature varies from an average of -10 degrees Celsius in some regions to 25 degrees in others; some cells are at the sea level, while others at an altitude of more than 5,000 m. Ruggedness, which is known to strongly determine travel costs (Hirte et al., 2020), varies profoundly across the empire. Apart from the continuous geography variables in Table 1 we also use three categorical variables on landform, dominant soil type, and lithology.[12]

---

[11]Florczyk et al. (2019) combine satellite images on built-up structures and census data to create agglomeration shapes that are economically more meaningful than administrative city borders, see Düben and Krause (2020).

[12]The heartland of Chinese empires is delimited by natural barriers on all sides: in the east by the sea, in the north by nomadic grasslands that were unsuitable to Chinese sedentary agriculture, in the north-west by the desert, in the west by the Himalayas, and in the south by a tropical climate with a disease environment unsuitable to the Chinese. In southern regions such as Guangxi, mortality rates among government officials were high and the Song dynasty settled convicts in these hazardous territories (Mostern, 2011; Marks, 1998).

Table 2: Correlation between County Size and Geography in 1500 CE

	County Size	Elev.	Rugg.	Temp.	Prec.	Dist. C.	Dist. R.
Elevation	0.37						
Ruggedness	0.20	0.77					
Temperature	-0.17	-0.46	-0.10				
Precipitation	-0.11	-0.21	0.22	0.73			
Dist. Coast	0.30	0.69	0.46	-0.45	-0.31		
Dist. River	-0.08	-0.27	-0.10	0.04	0.10	-0.50	
Dist. Equator	-0.04	-0.08	-0.37	-0.80	-0.81	0.14	0.05

Notes: We do not observe the actual county size as county borders are unobserved. Instead, we use the average county size per prefecture, dividing the prefecture's area by the number of counties within it. The geography variables refer to average pixel values within prefectures. The table omits categorical variables.

Table 2 shows the correlation of average county size with the geographical features. We can already see some of the predictions from our theoretical framework represented in the descriptive statistics. Note that a higher density of county seats implies a smaller county size. This means that the prediction of a higher (lower) county seat density in areas of more (less) favorable geography (Prediction 1a of Theorem 1) is reflected in the strongly positive correlation between county size and mean ruggedness as a negative geographical feature.
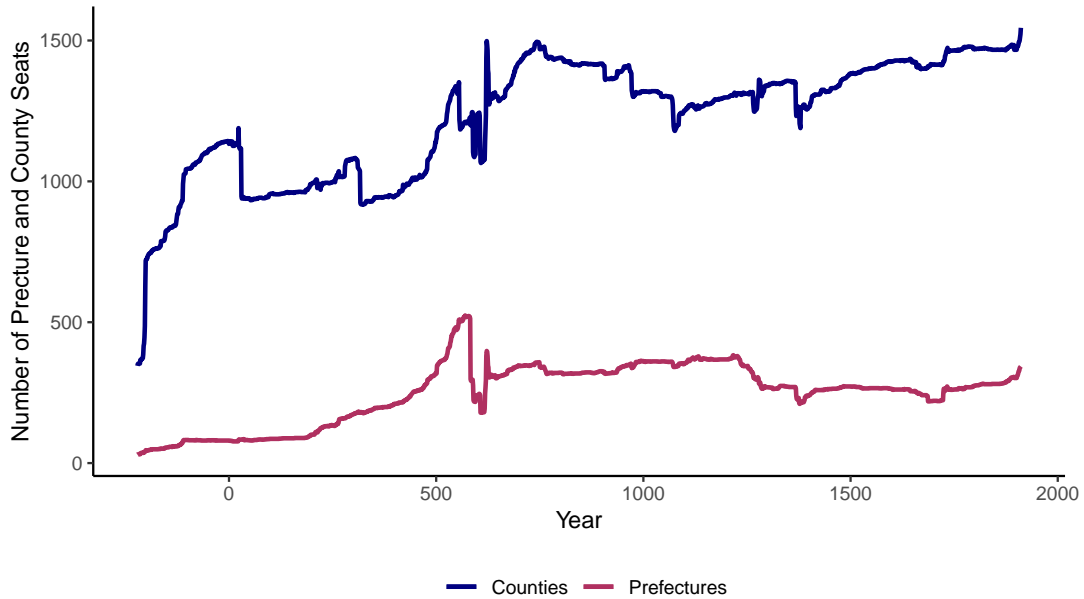
Figure 3: Number of Prefecture Seats and County Seats

Also, the map in Figure 1 shows that there were relatively more prefecture seats

per county seats near the Northern border of the empire, where the threat of nomadic invasions was highest (Bai and Kung, 2011). This is in line with prediction 1c of Theorem 1.

Figure 3 plots the quantity of county and prefecture seats over time. This evolution could be linked to a multitude of factors ranging from the empire's spatial extent over population growth to an evolution of the link between geography and institutions.

3.3 Empirical Strategy

3.3.1 Direct Effects of Geography

We test predictions 1a and 1b about the varying impact of geography on cities of different political statues by running the following regression at the pixel-level:

$$Urban_{it} = \alpha_t + \gamma Geo_i + \varepsilon_{it} \tag{14}$$

$Urban_{it}$ equals one for pixel i if it is home to a county or prefecture seat in year t and zero otherwise. The geography vector Geo entails *temperature*, *temperature*^{2}, *precipitation*, *precipitation*^{2}, *elevation*, *ruggedness*, *distance from the coast*, *distance from the equator*, *distance from major rivers*, *landform*, *dominant soil type*, and *lithology*. Our unit of analysis, the pixels, are 7.33 x 9.51 km in size. They are grid cells in an equal-area Mollweide projection. In our main specification, the geographical variables are assumed to be time-constant, while Section C.9 considers paleo data as far as available as a robustness check.

In our first specification, we will follow the literature in estimating the linear probability model eq. (14) with OLS. This allows us to test predictions 1a and 1b based on the coefficient estimates and the goodness of fit of the model. If geography plays a larger role in determining the location of county rather than prefecture seats, we should see larger and more significant γ coefficients as well as a higher R^2 for county seats rather than prefecture seats.

While being the standard benchmark, OLS comes with two main drawbacks in our setting: (i) It does not capture a potentially highly nonlinear data generating process. Settlement locations might be determined by a complex interplay of various geographic factors which would violate the linearity assumption. (ii) The high correlations between different geographic determinants (see Table 1) might yield questionable ceteris paribus interpretations.[13]

---

[13]For example, a rise in elevation by one kilometer comes with changes in other variables, such as

In our second specification, we add these issues by resorting to random forest algorithms from machine learning (Breiman, 2001). These prediction-oriented methods do not produce coefficient estimates like parametric econometrics does but explore patterns in the data in much larger complexity. Random forests are ensemble methods of individual decision trees, where each tree generates predictions by splitting the sample along its explanatory variables into increasingly homogeneous subgroups. In classification random forests this splitting is based on Gini coefficients; in regression random forests it is done based on the variance. Compared to other tree-based methods, the random forest only considers a randomly drawn subsets of explanatory variables at each split. This way the entire set of trees, or forest, comes up with a larger variety of solutions, a better bias-variance trade-off and an adequate solution in our application (Kuhn and Johnson, 2013). Given the pattern recognized in the data set, the algorithm splits the sample of all pixels based on their geographic characteristics, thereby predicting which pixels are to be classified as hosting a county (or prefecture) seat.[14]

After identifying where cities locate in the first place, we investigate to what extent the mechanism choosing the initial location differs from the one driving persistence. Does local geography explain how long administrative seats persist in their location?

$$Y_i = \alpha + \gamma Geo_i + \varepsilon_i \tag{15}$$

eq. (15) regresses the number of years Y that pixel i hosts an administrative city on the geography vector. For simplicity we subset the years to those considered in the 214 baseline cross-sections and the pixels to those grid cells that ever hosted a city in any of the examined years.

3.3.2 Indirect Effects of Geography

According to Prediction 1b of Theorem 1, first-nature geography should have a much smaller effect on the location of prefecture seats than on that of county seats. Instead, the network position, in particular expressed as centrality within the prefecture, should play a large role. Based on the theoretical framework, we expect prefecture seats to minimize the travel distance to the county seats in their territory (see eq. (5)) and overall pixels in the prefecture weighted by military relevance (see eq. (5)).

---

temperature, ruggedness etc. Estimating and interpreting the influence of one geographical variable while keeping the rest constant might be driven by outliers and might potentially lead to unrealistic results.

[14]Our classification random forests consist of 500 trees each. Within the trees each node employs three out of the ten geographic determinants for splitting. Observations are sampled with replacement. To generate an R^2 we also run regression random forests using the same data which splits the sample based on the variance instead of the Gini coefficient.

In the first estimations we focus on the network between cities. For each city j in prefecture p we compute the total distance D to all other cities in that prefecture. Distance is measured in terms of travel time derived via Tobler's (1993)'s hiking function which takes the topography into account.

$$\ln D_{jpt} = \alpha_t + \beta\, PrefectureSeat_{jpt} + \gamma_{pt} + \varepsilon_{jpt} \tag{16}$$

As eq. (16) illustrates, we regress the natural logarithm of that distance on a prefecture seat indicator and prefecture-fixed effects.

The second stage then considers prefecture seat locations within prefectures overall, irrespective of where the county seats are.

$$\ln C_{ipt} = \alpha_t + \beta\, PrefectureSeat_{ipt} + \gamma_{pt} + \varepsilon_{it} \tag{17}$$

C_{it} is the centrality of a pixel measured as the travel time according to Tobler's (1993) hiking function between pixel i and the respective prefecture's centroid in year t.[15] $PrefectureSeat_{ipt}$ is an indicator that equals one if the pixel hosts the prefecture seat and zero otherwise.

In a third stage we set up an unsupervised machine learning algorithm that imitates the theoretical framework's assumptions on how prefecture seats are placed throughout the empire. It begins with an initial random allocation of prefecture seats and then moves them around based on the location of county seats and military threat levels. Once the iterations reach a stability threshold, the algorithm stops and we compare artificial and actual prefecture seats.

The entire estimation strategy on direct and indirect effects comes as a series of cross-sectional rather than panel estimations.[16] With those independent regressions we introduce a level of flexibility in the identified causal mechanism over time that would otherwise be restricted to some degree.

---

[15]Robustness checks in Section C.11 and Section C.12 repeat eq. (16) and eq. (17) with Euclidean distances and travel time derived via a modified version of Tobler's (1993) hiking function by Márquez-Pérez et al. (2017).

[16]The baseline random forest approach runs on 214 cross-sections, i.e. going from 220 BCE until 1910 CE in ten year steps.

4 Results

4.1 Regression Results for Direct Effects of Geography

Table 3 presents the OLS regression results of eq. (14), investigating the impact of local geography on the location of county seats.[17] We see that geography indeed plays a role, confirming Prediction 1a. Throughout Chinese history, places with flatter terrain and rivers nearby are more likely to host county seats. The impact of temperature and distance from the coast seems to be time-varying, while higher elevation is associated with a lower likelihood to host a county seat throughout the time period covered.[18]

When re-running the regression using prefecture seats rather than county seats, we make three observations (see Table 4): (i) The signs of the geographical variables are typically the same. (ii) But the coefficient estimates in the prefecture seat regressions are much smaller and less often statistically significant. (iii) And the explanatory power in terms of the adjusted R^2 is markedly lower. This is in line with Prediction 1b that prefecture seats are less strongly determined by local geographical factors than county seats.

We take a closer look at the relative importance of the various geographical features with the random forest classification. While these machine learning methods do not yield coefficient estimates, they produce variable importance measures according to how much the variables contribute to the reduction of the gini coefficient. This allows us to rank geographic determinants in their importance for predicting settlement locations.

Figure 4 displays the variable importance ranks for the determination of the county seats, with the most important variable receiving a rank of one and the least important one receiving a rank of ten. Elevation, a variable that also remained highly statistically significant in our OLS for most of the dynastic period, receives the highest importance and landform the lowest importance throughout most of imperial China.[19] Some variables change importance ranks from time to time but often maintain similar orders for long periods.

The results for prefecture seat locations in Figure 5 are more volatile, only agreeing

---

[17]Throughout the paper, we use the term *local geography regression* as a synonym for regressions estimating the direct effects of geography.

[18]Many of the changes in coefficient estimates over time reflect the evolution of the parameters in our model and the degree to which institutional developments alter the causal framework between geography and city locations (see Section C.2).

[19]Dominant soil type, landform, and lithology occupying the lowest ranks throughout all years might be related to them being categorical whereas the other variables are continuous.

Table 3: Local Geography Regressions (County Seats)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.04***	0.07***	0.07***	0.09***	0.12***	0.15***	0.09***	0.08***	0.10***	0.12***	0.12***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Dist. Equator	−0.07***	−0.10***	−0.10***	−0.14***	−0.20***	−0.26***	−0.19***	−0.17***	−0.18***	−0.23***	−0.21***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Dist. Coast	−0.00	−0.05***	0.01	0.07***	0.13***	0.13***	0.12***	0.11***	0.04**	0.07***	0.05***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Dist. River	0.01	−0.05	−0.05	−0.14***	−0.26***	−0.14***	−0.06	−0.03	−0.05	−0.06	−0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Ruggedness	−0.07***	−0.08***	−0.08***	−0.12***	−0.09***	−0.16***	−0.15***	−0.14***	−0.12***	−0.14***	−0.17***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Temperature	0.05***	0.09***	0.05***	−0.01	−0.04*	−0.15***	−0.08***	−0.06***	−0.06***	−0.08***	−0.08***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Temperature^2	−0.09**	−0.02	−0.09	−0.16**	0.32***	0.53***	0.28***	0.20**	0.10	−0.06	−0.02
	(0.05)	(0.06)	(0.06)	(0.07)	(0.09)	(0.11)	(0.10)	(0.10)	(0.09)	(0.11)	(0.11)
Precipitation	−0.19***	−0.35***	−0.19***	−0.07**	−0.25***	−0.10**	0.01	0.03	−0.05	−0.02	−0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Precipitation^2	0.34***	0.64***	0.33***	0.13*	0.44***	0.14	−0.06	−0.11	0.04	−0.02	−0.00
	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
Elevation	−0.38***	−0.32***	−0.38***	−0.71***	−1.19***	−1.32***	−1.10***	−0.94***	−0.82***	−1.09***	−0.93***
	(0.07)	(0.10)	(0.09)	(0.09)	(0.10)	(0.12)	(0.11)	(0.11)	(0.11)	(0.13)	(0.12)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.03	0.04	0.03	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.02
Num. obs.	86,257	86,257	86,257	86,257	86,257	86,257	86,257	86,257	86,257	86,257	86,257

Notes: The table reports the regression results of eq. (14) using the county seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

Table 4: Local Geography Regressions (Prefecture Seats)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.01**	0.01***	0.01*	0.03***	0.03***	0.05***	0.04***	0.06***	0.03***	0.03***	0.03***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Dist. Equator	−0.01**	−0.01**	−0.01	−0.04***	−0.06***	−0.08***	−0.08***	−0.10***	−0.05***	−0.05***	−0.05***
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Dist. Coast	−0.00	−0.00	0.01	0.03***	0.04***	0.02**	0.03***	0.04***	0.02**	0.02**	0.01
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Dist. River	0.00	0.01	0.00	−0.03*	−0.04**	−0.06***	−0.04*	−0.05**	−0.03*	−0.04**	−0.04*
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Ruggedness	−0.01**	−0.02***	−0.02***	−0.03***	−0.02	−0.06***	−0.04**	−0.04**	−0.07***	−0.06***	−0.04***
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Temperature	−0.00	0.01	0.01**	−0.00	−0.01	−0.06***	−0.04***	−0.06***	−0.02	−0.01	−0.02*
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Temperature^2	0.00	−0.03	−0.03	−0.04	−0.07*	0.18***	0.02	−0.01	−0.07	−0.07	−0.02
	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)
Precipitation	−0.02**	−0.04***	−0.03**	−0.07***	−0.04*	−0.05**	−0.04*	−0.04	−0.02	−0.04**	−0.04*
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Precipitation^2	0.03**	0.09***	0.05**	0.14***	0.07*	0.09**	0.06	0.07	0.02	0.07	0.06
	(0.01)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
Elevation	−0.04**	−0.03	−0.00	−0.12**	−0.30***	−0.30***	−0.34***	−0.45***	−0.06	−0.07	−0.15***
	(0.02)	(0.03)	(0.03)	(0.05)	(0.05)	(0.06)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Num. obs.	86,257	86,257	86,257	86,257	86,257	86,257	86,257	86,257	86,257	86,257	86,257

Notes: The table reports the regression results of eq. (14) using the prefecture seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

in attributing the lowest value to landform. There appears to be no clear pattern, with
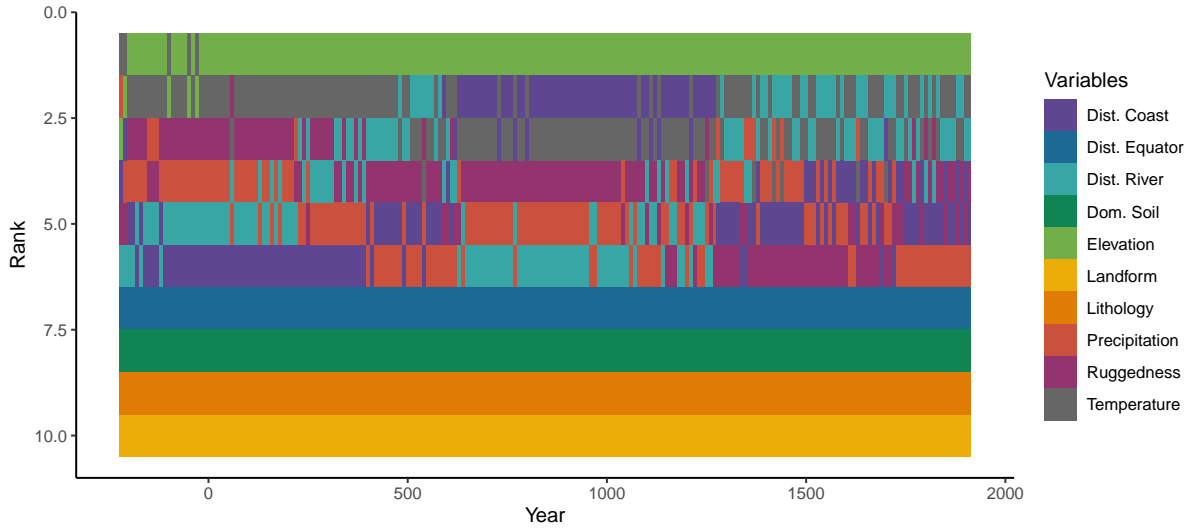
Figure 4: Variable Importance in Classification Random Forests Linking Geography and County Seat Location

the ordering of the top six variables reshuffled at high frequencies. We see that although the random forest allows for much more complex interactions between the various geographical features, they cannot link local geography and prefecture seat locations through any recognizable pattern. In this way, the findings from random forests confirm the OLS results about a more immediate impact of local geography on county rather than prefecture seat locations.

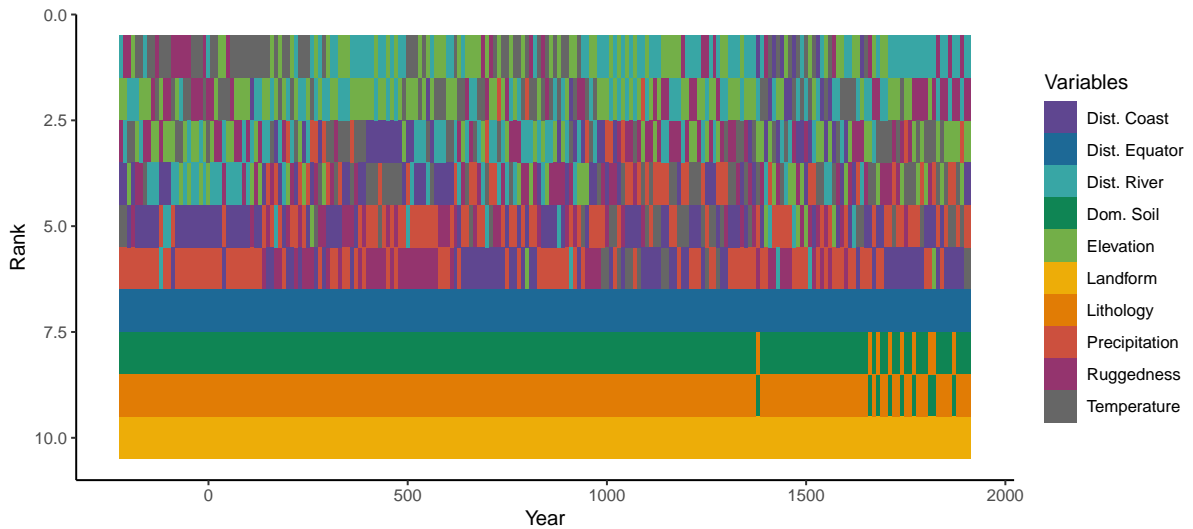**Figure 5:** Variable Importance in Classification Random Forests Linking Geography and Prefecture Seat Location

Let us proceed to look at the goodness of fit. The R^2 in the OLS regression is very low, with geography explaining very little of the variation in county seats (R^2 of ≤ 0.04 in Table 3) and hardly anything of the variation in preference seats (R^2 of ≤ 0.01 in

Table 4). Yet, this low magnitude can to some extent be explained by the Modifiable Areal Unit Problem (Fotheringham and Wong, 1991; Bailey and Gatrell, 1995). It is known that in geo-spatial research the result may vary with the unit of analysis. This is what happens in our case. We consider alternative resolutions, aggregating our 7.33 x 9.51 km baseline pixels by factors of 2, 3, 4 and 5, referred to as *small*, *medium*, *large*, and *very large* respectively. Figure 6 illustrates the fraction of urban pixels that are correctly classified as urban.[20] We find again that (i) the predictive performance increases with the pixel size, and that (ii) the share of correctly predicted county seats always strongly exceeds that of the correctly predicted prefecture seats.[21] Evaluating the results based on R^2 instead (see Figure C-2 in Appendix Section C.6) confirms these conclusions.[22]
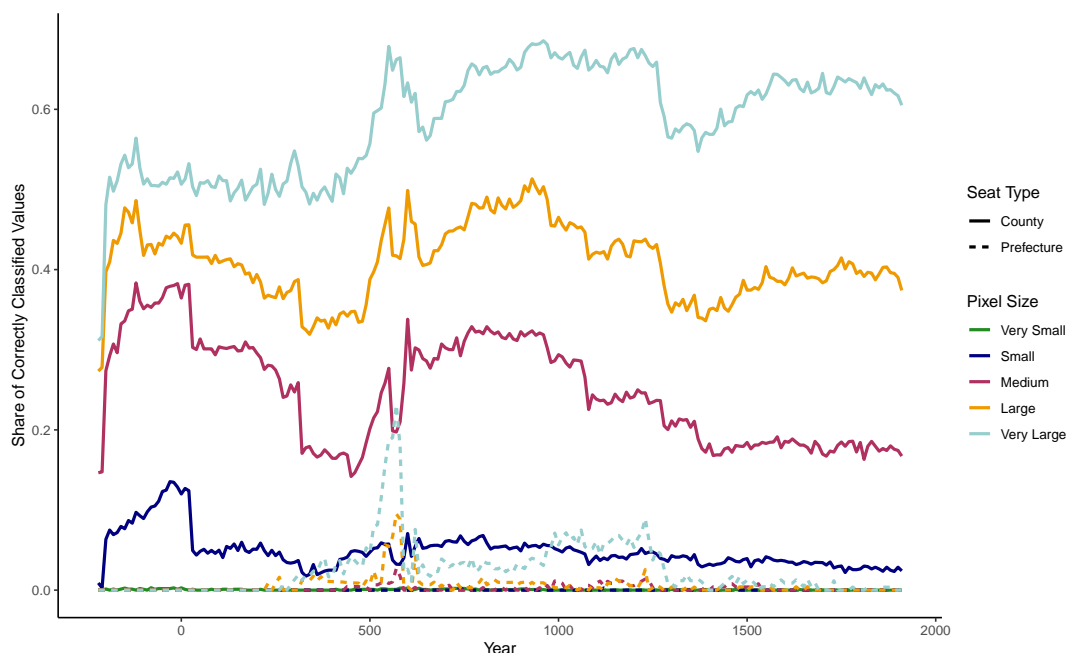
Figure 6: Correctly Classified Urban Pixels in Classification Random Forests Linking Geography and Settlement Location

Section C.5 confirms these insights by comparing the share of correctly classified county and prefecture seats of the random forest classifier, using different resolutions. Coarser resolutions also bring out a characteristic pattern along the time dimension in the data. The pattern with which the share of correctly classified urban grid cells sharply rises and falls is aligned with institutional change. Figure 7 marks the rise of

---

[20]Figure C-1 depicts the performance on rural pixels and on urban and rural pixels jointly.

[21]According to the results displayed in Section C.1, this pattern also holds when estimated via OLS.

[22]This is a classification problem, which is why we use classification random forest's confusion matrices to evaluate the algorithm's ability to predict city locations. Deriving R^2 values from regression random forests, which split the sample based on the variance rather than the gini coefficient, is therefore technically not correct.
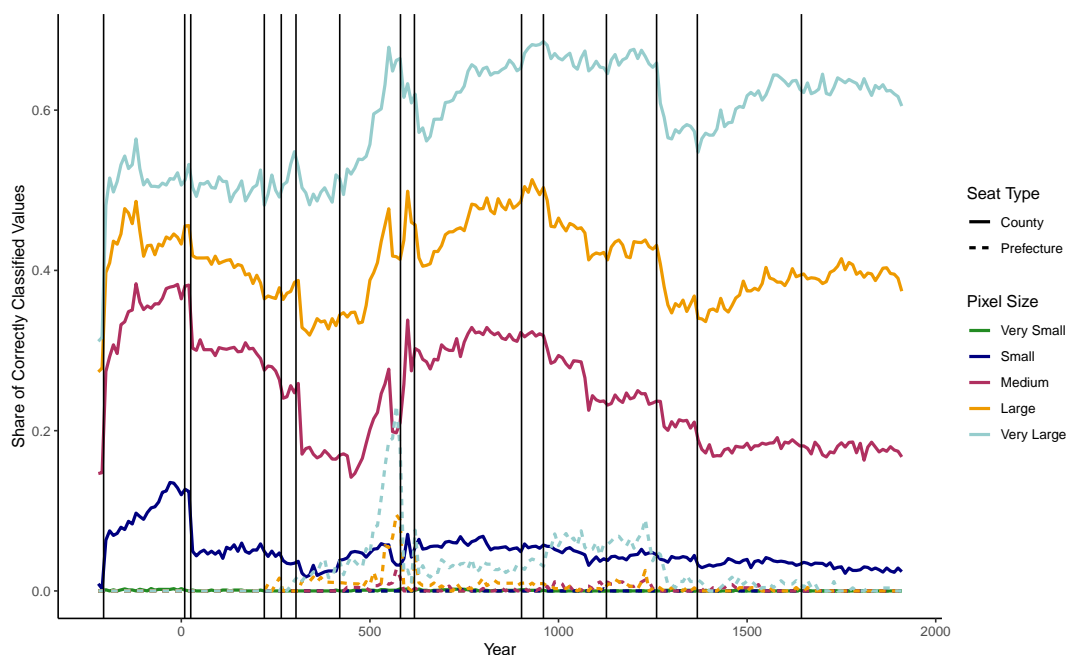
Figure 7: Correctly Classified Pixels in Classification Random Forests and the Rise of Dynasties

new dynasties using vertical lines.[23] Hence, the association between geography and county seat locations changed with the institutions. Even though dynasties often kept many of their predecessors' institutional structures and spatial organization, institutional conditions did not remain constant.[24] And this is what these results reflect. Institutions affect the causal framework between geography and urban settlement locations.[25]

Concerning changes over time, we also investigate the link between urban persistence and geography. This means that we count in how many of the 214 baseline cross-sections a pixel hosts a city and regress it on the geography vector (see eq. (15)).[26] Judging from the resulting R^2 in Figure 8, not just the initial location but also the persistence of county seats appears to be related to geography. The opposite holds for prefecture seats. Local geography neither explains where they arise in the first place nor where they persist over time.[27]

---

[23]Figure C-3 in Appendix Section C.6 is the corresponding plot using the R^2.

[24]Institutional adjustments were often inspired by past dynasties, as contemporary neighbors were too different to serve as inspiration, e.g. by being nomadic rather than sedentary (Mote, 1999).

[25]The pace of adjustment depends on the institutional change, which is more complex than a binary distinction between dynasties founded by Han Chinese and those founded by non-Han peoples. The Yuan dynasty founded by the Mongols in 1260 CE implemented around 71 percent of their spatial modifications within the first 20 years, coinciding with a sharp drop in Figure 7. The Qing dynasty founded in 1644 CE by the Manchu - also non-Han - introduced only around two percent of their spatial adjustments in the first 20 years (Mostern, 2011).

[26]We subset the sample to pixels that are urban in at least one of the years. Dropping zeros circumvents the need for a Tobit model.

[27]Section C.10 confirms these findings via OLS regressions.
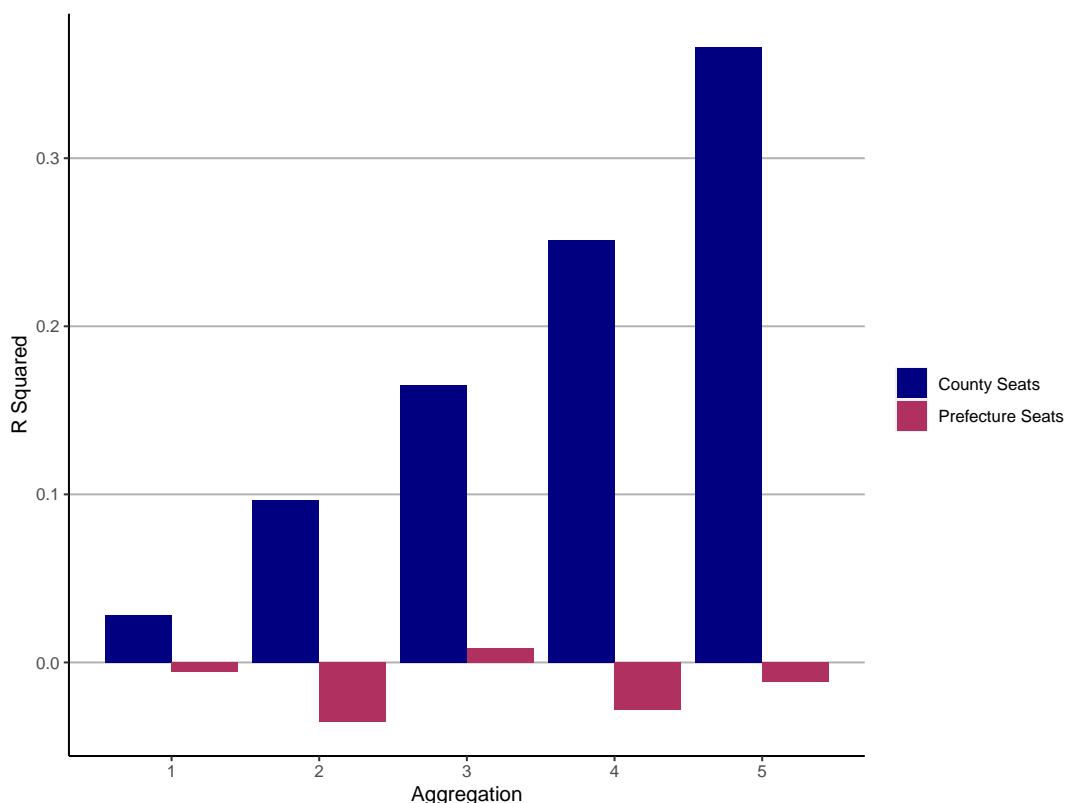
Figure 8: R^2 in Regression Random Forests on Urban Persistence and Geography

Our baseline results about a higher importance of local geographical factors for county rather than prefecture seats are robust to further robustness checks. Section C.3 confirms the findings with logit and probit models and Section C.4 with spatial econometrics. In Section C.8 we discuss alternative ways of dealing with the changing extent of the empire over time, showing that the main results hold for alternative empire shapes. In addition, in Section C.9 we supplement the analysis with historical geographical data, taking inter alia into account how the Yellow river changed its course.

4.2   Regression Results for Indirect Effects of Geography

Given that prefecture seats are to a lesser extent determined by local geography, we investigate the centrality argument to explain their location within a prefecture.

First focusing on the urban network, we estimate eq. (16). Figure 9 plots the coefficient estimates and 95 percent confidence intervals from these cross-sectional regressions. In line with our predictions, prefecture seats are significantly closer to other cities within a prefecture than county seats are.[28]

---

[28]In Section C.11 we repeat the estimation with Márquez-Pérez et al.'s (2017) modified Tobler hiking
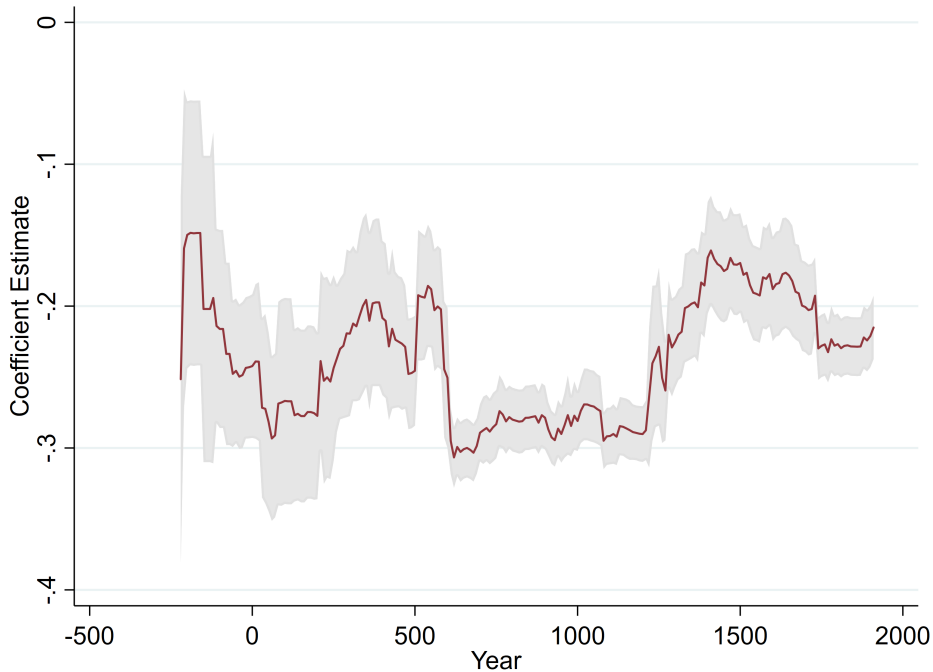
Figure 9: Inter-City Distance and Administrative Status

Looking at the overall within-prefecture space, Figure 10 presents the regression results of eq. (17).[29] For all pixels in a prefecture, the natural logarithm of the distance between pixel centroid and the centroid of the central pixel are regressed on a constant, a prefecture seat dummy (which equals one if the pixel hosts the prefecture seat), and prefecture-fixed effects. The results show that prefecture seats are on average significantly closer to the centroid than other pixels. Prefecture seats are in many years around 50 percent closer to the centroid than the average other pixel within the prefecture is.[30] Some prefecture borders are missing from the data in earlier years which at least partly drives the estimates' stronger fluctuations and lower precision in that period.

Figure 11 illustrates the locations of geographic prefecture centroids, county and prefecture seats for a subset of prefectures in the year 1500 CE. We can see that centrality certainly plays a role. Many prefecture seats are rather close to the centroid marked by the green cross. Yet, there are also exceptions and other strategic considerations might

---

function and in Section C.12 with Euclidean distances. The results are virtually the same.

[29]As in Figure 9, the red line denotes the coefficient estimate of interest and the grey area the respective 95 percent confidence intervals.

[30]In Section C.11 we repeat the estimation with Márquez-Pérez et al.'s (2017) modified Tobler hiking function and in Section C.12 with Euclidean distances. The results are virtually the same.

[31]The selected prefectures are: Pingyang Fu, Ze Zhou, Huaiqing Fu, Weihui Fu, Lu Zhou, Qin Zhou, Zhangde Fu, Liao Zhou, Guangping Fu, Shunde Fu, Fen Zhou and Daming Fu.
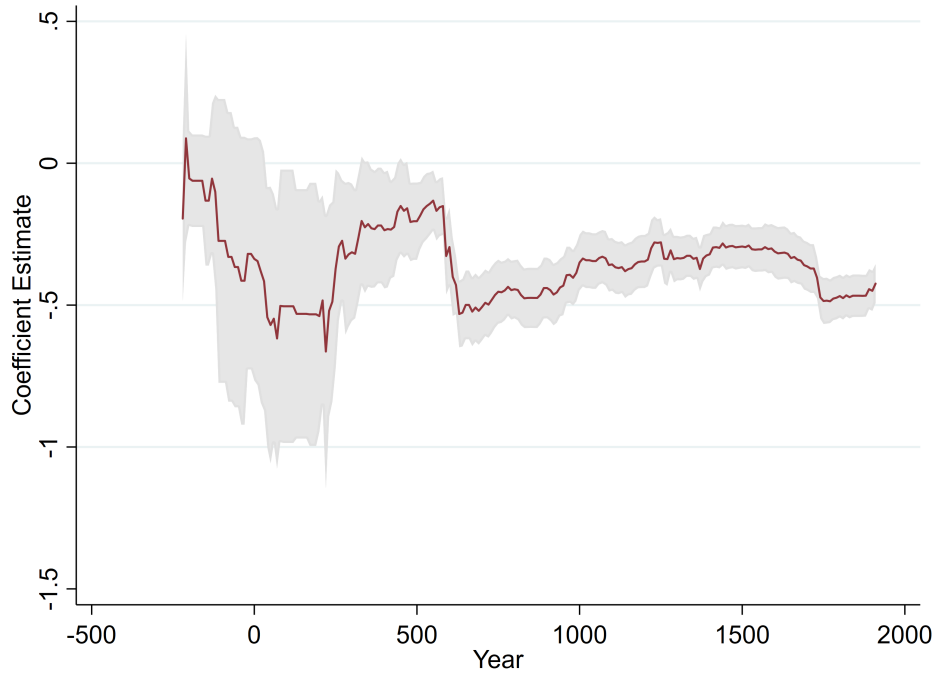
Figure 10: Pixel Centrality and Administrative Status

be at work as well, potentially related to the locally varying threat levels.

Linking the estimations more closely to the theoretical framework, we design an algorithm that simulates the hypothesized prefecture seat location optimization. The

× Centroid   ▲ County Seat   ■ Prefecture Seat
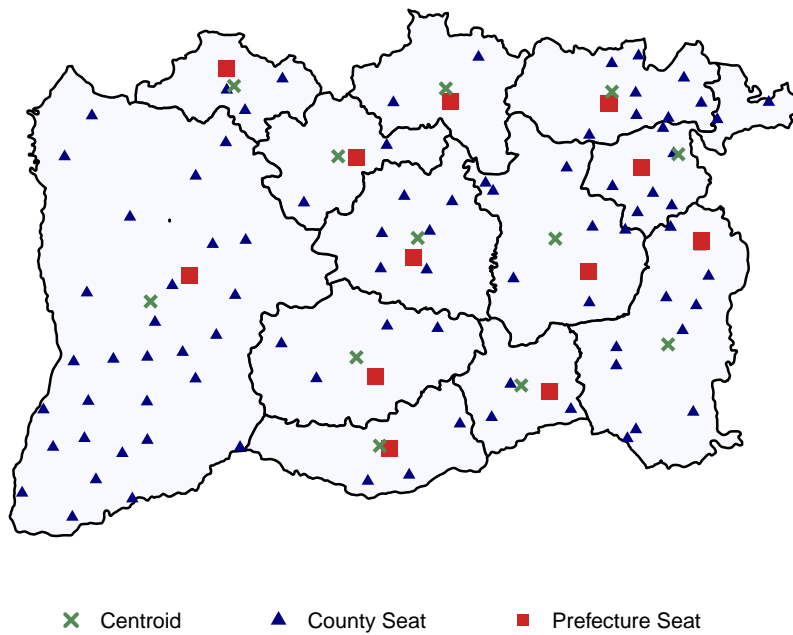
Figure 11: Selected Prefectures in 1500 CE [31]

24

mechanism is an unsupervised machine learning technique applying a concept similar to k-means, but in a spatial setting. It begins by placing prefecture seats randomly throughout the empire and then iteratively optimizes their position until it reaches a stability threshold. The algorithm produces a spatial distribution similar to the one we observe in the data, supporting the hypothesized theoretical mechanisms (see Section C.13).

We conclude that prefecture seats are affected by geography only in an indirect way, as predicted by our model. While geography determines both the location of county seats, prefecture seats are chosen with the centrality argument in mind to minimize the travel-time distance.

4.3 Results for Market Towns

Our main analysis considers the impact of geographical factors on administrative cities of different rank, county and prefecture seats, for which we have location data throughout imperial China. An appealing extension is to look at cities without administrative or military status, so-called market towns which emerge during the second millennium CE (Wu and Gaubatz, 2013). CHGIS provides cross-sectional data on market towns for 1820 CE when the Qing Dynasty ruled China. This allows us to conduct a comparison of the three types of cities at that time, which was a decisive period of Chinese urbanization and saw the establishment of more cities than any other era since Han (Wu and Gaubatz, 2013). Yet, the following results come with the caveat that it is a snapshot from one period and that the spatial distribution of market towns may have looked different under earlier dynasties.[32]

Figure 12 illustrates how in 1820 CE market towns far outnumber administrative settlements. It is obvious that in that year China had long past the times in which administrative cities represented roughly the entire urban landscape. There are 8,659 market towns in that year of which many are based in the heartland of imperial China. In comparison, there are only 301 prefecture and 1859 county seats.

We re-run our random forest models concerning the direct effects of geography on city locations, applied separately to the three types of cities using the formerly discussed five different pixel sizes.[33] Figure 13 presents the fraction of correctly classified

---

[32]The empire's population size tripled between 1680 and 1850 CE (von Glahn, 2016).

[33]The imperial borders used in this section are different from the ones in the previous sections. The here presented results on county and prefecture seat locations, therefore, slightly differ from the above discussed output.
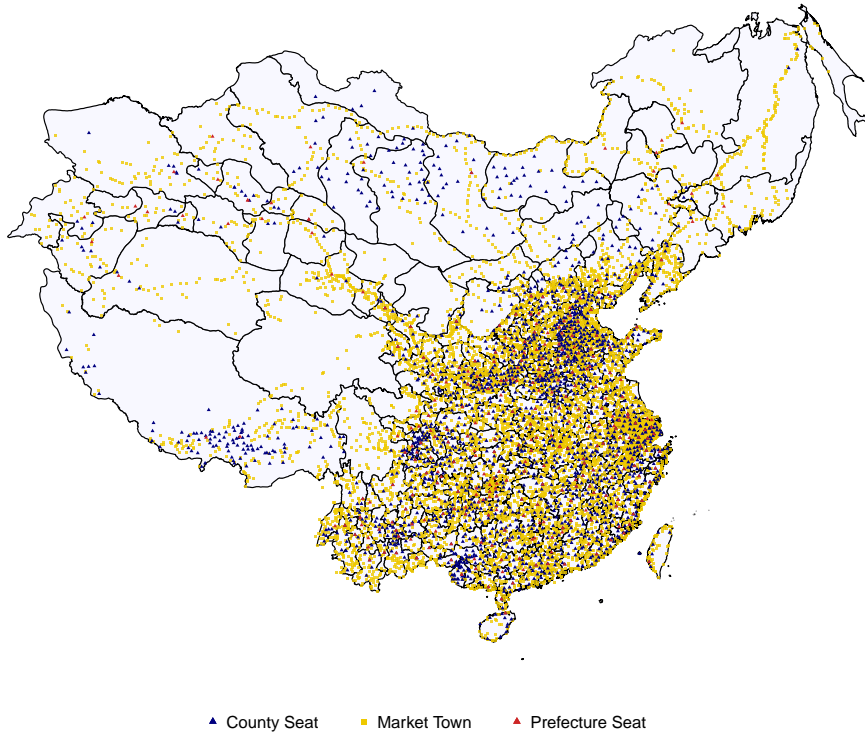
Figure 12: Administrative and Market Towns in 1820 CE

urban grid cells and shows a very clear pattern: Market towns are even more strongly determined by local geography than county seats, which in turn are more strongly
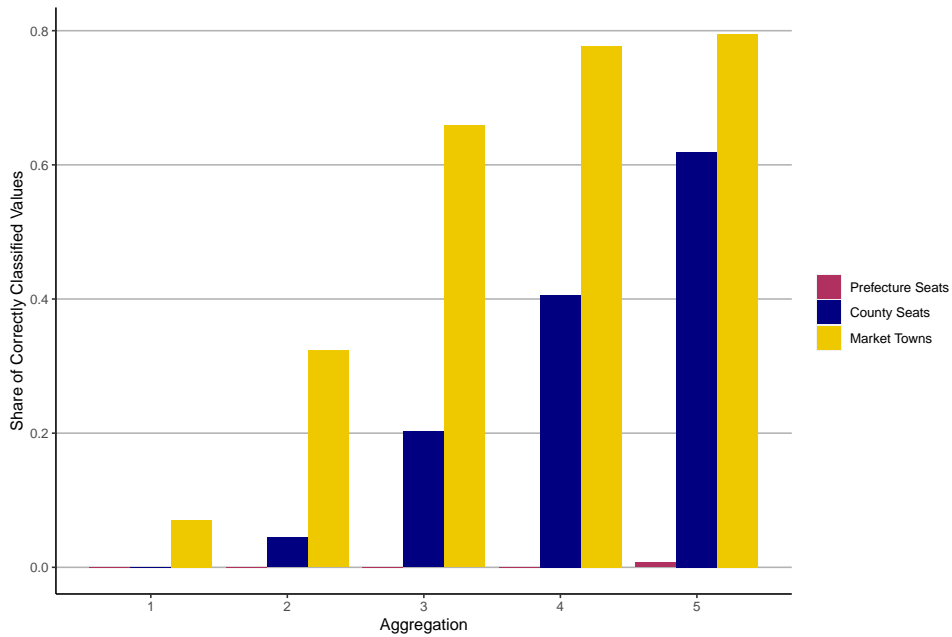
Figure 13: Correctly Classified Urban Pixels in 1820 CE

influenced by geography than prefecture seats.[34] This ranking persists at all levels of data aggregation, with the explanatory power of the geographical factors in all three cases higher at coarser resolutions. In fact, at the coarsest resolution, we can correctly classify around 80 percent of market towns pixels using local geography, compared to slightly over 60 percent of county seat locations and quasi no prefecture seat grid cells.[35]

We conclude that the inclusion of non-administrative market allows us to extend our previous insights to a third hierarchical level: The higher the administrative status of a city, the lower the influence of local geographical conditions.

4.4 Comparison to Europe

We complete our historical analysis by a comparison of Chinese and European cities. Figure 14 illustrates the European cities documented by Bairoch et al. (1988) based on a minimum population of 5,000 inhabitants. This data set is typically used in studies of European urban history, see Jedwab et al. (2020). The European cities are available in the years 800, 900, 1000, 1200, 1300, 1400, 1500, 1600, 1700, 1750, 1800, and 1850. To facilitate our comparison with Europe, we transform the Chinese data and create a corresponding grid in which a cell is set to one if it hosted at least one county seat or prefecture seat in any of these years.

Figure 15 shows the fraction of correctly classified urban pixels in classification random forests of the location of a European or Chinese city on the geographical features, using various levels of pixel aggregation.[36] We see that geographical features can explain the location of historical European cities to a lesser extent than that of Chinese county seats, yet more than that of Chinese prefecture seats.[37]

In contrast to China, Europe was politically highly fragmented at that time. Institutions were much more localized than in China, with independent city states constraining the formation of centrally-governed nations (Jedwab et al., 2020; Acemoglu and Robinson, 2019). Geographical factors such as access to the sea and to rivers have been shown to have contributed to the formation of European cities (Bosker and Buringh, 2017; Bosker et al., 2008). Institutions certainly played a role, but at a more local level and with a larger variance in regional institutional settings and cities'

---

[34]Figure C-15 presents the corresponding R^2 and Figure C-16 the variable importance.

[35]Note that we estimate the locations of different city types in separate estimations, as in the baseline approach.

[36]Figure C-17 in Appendix Section C.15 plots the respective R^2.

[37]The geography differs from the baseline estimations in that it omits the three soil-related variables landform, dominant soil type, and lithology which are derived from an Asia-specific data set.
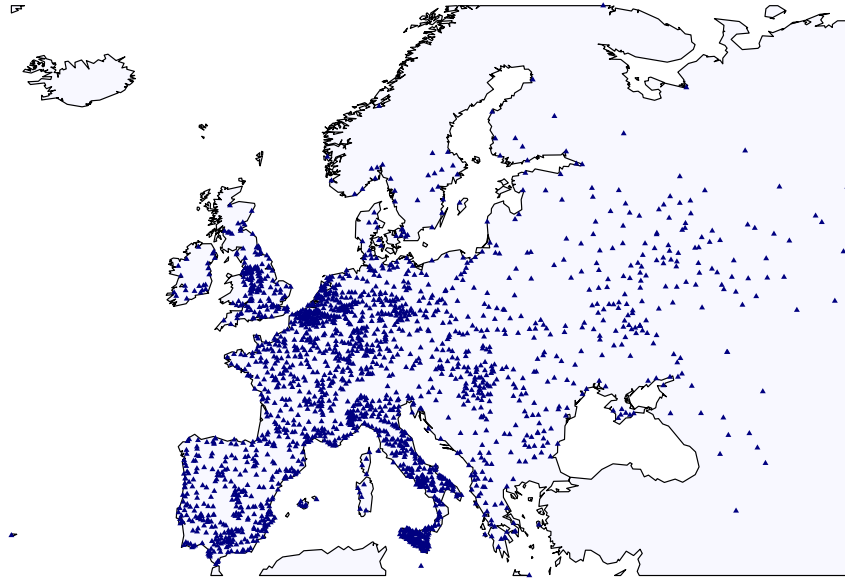
Figure 14: European Cities (800 CE - 1850 CE)

administrative tasks, which might explain why the magnitude of our European results lies between that of Chinese county and top-down prefecture seats.
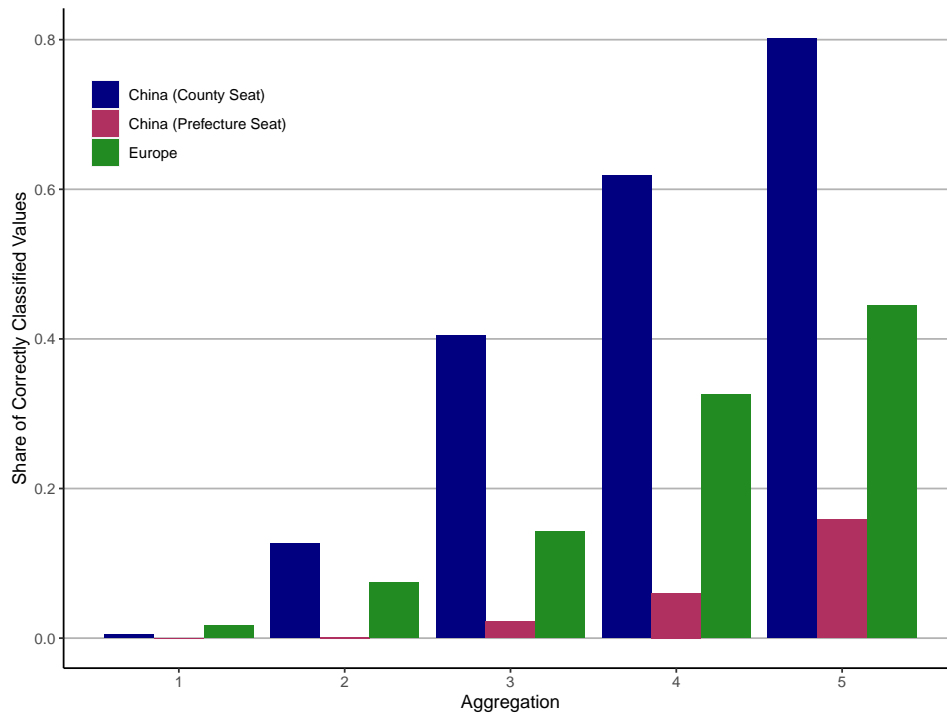
Figure 15: Correctly Classified Urban Pixels in Regression Random Forests for China and Europe

5 History dependence

Our results so far suggest an important role for geography and institutional factors in explaining the location and status of historical cities. This leads to the question of how relevant these insights are today. Our historical coverage period ends with the downfall of the Chinese empire in 1911. Since then, Chinese urban development has undergone enormous changes; in particular, China's integration into world markets in the last decades fueled the rise of coastal cities (Ma, 2002; Anderson and Ge, 2005).

First, we take a look at political path dependence. We examine the 29 current province and autonomous region capitals that fall into the territory of former Chinese empires observed in CHGIS. 28 ouf of these 29 current regional capitals acted as a prefecture seat at some point in historical times.[38] As Figure 16 shows that more than half of them were already prefecture seats in the fifth century CE.
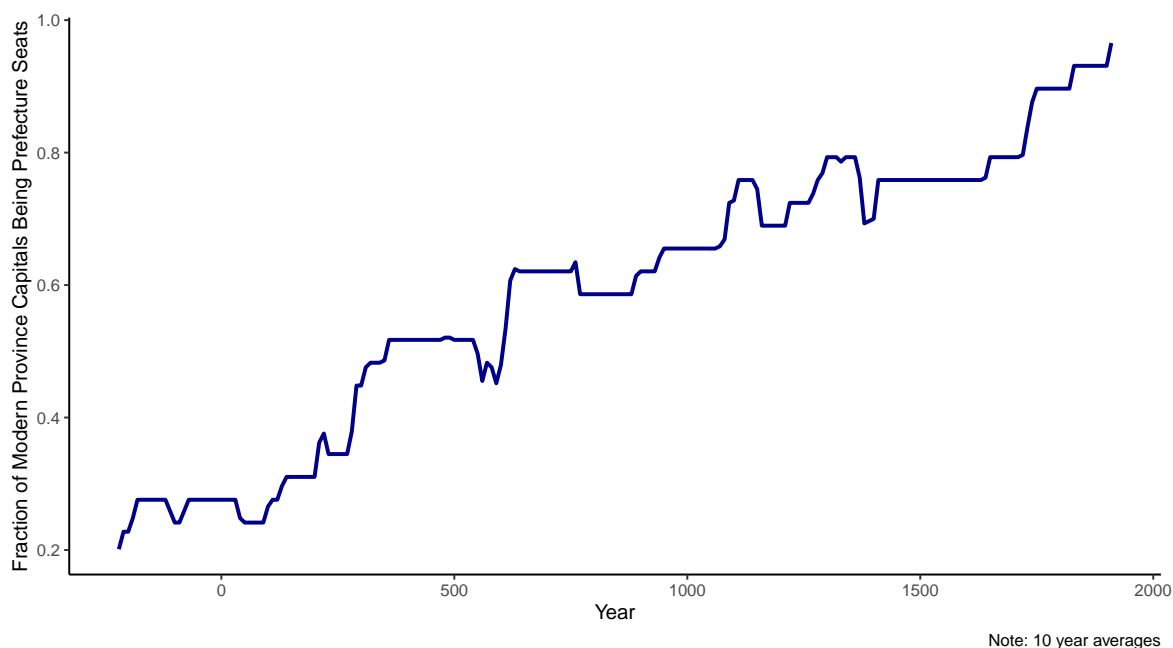
Figure 16: Importance of Modern Provincial Capitals in Chinese History

We then turn to look at economic path dependence. Do modern cities which hosted a county and prefecture seat during imperial times show higher population and economic activity than their peers? In the following, we work with 1,841 Chinese agglomerations with more than 50,000 inhabitants in 2015 as defined by the Global Human Settlement Layers (Florczyk et al., 2019).

---

[38]More precisely, the pixels of the location of the historical prefecture seats fall within the agglomeration of the modern city as defined by the GHS Urban Centre Database (Florczyk et al., 2019). The location of the 29 modern regional capitals are illustrated in Figure C-18 in Section C.16.

Table 5: Imperial History and Population Size

	OLS		Median Regressions	
	Pop.	Log(Pop.)	Pop.	Log(Pop.)
Intercept	$145,172.74$***	11.55***	$86,167.41$***	11.36***
	$(10,232.14)$	(0.02)	$(2,208.90)$	(0.03)
Historic Prefecture Seat	$502,237.18$***	0.71***	$70,499.65$***	0.60***
	$(103,508.56)$	(0.06)	$(8,604.41)$	(0.06)
Historic County Seat	$56,115.67$***	0.24***	$23,772.08$***	0.24***
	$(18,161.79)$	(0.04)	$(4,877.45)$	(0.05)
Adj. R^2	0.02	0.10		
Num. obs.	$1,841$	$1,841$	$1,841$	$1,841$

Notes: Heteroskedasticity-robust standard errors in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Agglomerations' shapes and population refer to the year 2015. Omits four agglomerations with zero nighttime light emissions in either of the three data sets.

In Table 5 we regress the (log) population size of these cities on a historic prefecture seat and a county seat indicator. The prefecture (county) seat indicator equals one for all agglomerations that held prefecture (county) seat status at any point in the past.[39] We see that cities which were a prefecture seat at one time in the past are now 71 percent more populous than others which never had that status. In absolute terms the population bonus is on average around 502,000 people. For cities which were county seats once, the corresponding to difference to non-county seats is 24 percent. These figures are more subdued for median rather than OLS regressions because they are driven to a lesser extent by the largest agglomerations.[40] But the effects remain highly statistically significant and sizable.[41]

Looking at economic outcomes, we work with satellite data of nighttime lights as a proxy for local economic activity (Henderson et al., 2012; Donaldson and Storeygard, 2016). In Table 6, we regress the log total nighttime light emissions of the 1,841 Chinese agglomerations on a historic prefecture seat and a county seat indicator.[42] We can see

---

[39]We define the two indicators to be mutually exclusive. County seats that became prefecture seats at some point are only counted as prefecture seats in these regressions.

[40]Most countries' city size distributions are characterized by a lognormal body and a Pareto distributed upper tail (Düben and Krause, 2020), which explains the difference between the mean and the median effects.

[41]We do not investigate the population size of cities in imperial times because of severe measure error in the historical censuses. Imperial administrations started collecting data on its subjects at an early stage, with the Qin and Han dynasties (221 BCE - 220 CE) maintaining detailed population registers and even listing property. However, the population had a strong incentive the misreport, as the administration used the data to draw people for military service. The reliability of censuses also changed over time. Population registers e.g. became a local local responsibility during the Qing dynasty (1644 - 1911 CE), leading to severe declines in quality (von Glahn, 2016; Major and Cook, 2017).

[42]The nighttime lights refer to the 2013 satellite F18 DMSP OLS stable lights image (Elvidge et al.,

Table 6: Imperial History and Nighttime Light Emissions

	OLS			Median Regressions		
	Stable	Corrected	VIIRS	Stable	Corrected	VIIRS
Intercept	7.11***	7.18***	6.57***	7.18***	7.19***	6.66***
	(0.04)	(0.05)	(0.06)	(0.04)	(0.04)	(0.06)
Historic Prefecture Seat	0.80***	0.90***	1.12***	0.51***	0.53***	0.74***
	(0.07)	(0.08)	(0.09)	(0.07)	(0.08)	(0.09)
Historic County Seat	0.41***	0.40***	0.50***	0.27***	0.26***	0.33***
	(0.06)	(0.06)	(0.08)	(0.06)	(0.07)	(0.08)
Adj. R^2	0.08	0.08	0.09			
Num. obs.	1,841	1,841	1,841	1,841	1,841	1,841

Notes: Heteroskedasticity-robust standard errors in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Agglomerations' shapes refer to the year 2015, the stable and corrected light to the last available year (2013) and the VIIRS light data to 2016. Omits four agglomerations with zero nighttime light emissions in either of the three data sets.

that modern cities that used to be prefecture (county) seats are on average 80 (40) to 112 (50) percent brighter than cities without a historic political status. These results decrease in magnitude when using median rather than OLS regressions.

Table 7: Imperial History, Nighttime Light Emissions, and Population Size

	OLS			Median Regressions		
	Stable	Corrected	VIIRS	Stable	Corrected	VIIRS
Intercept	−5.06***	−7.15***	−8.24***	−4.24***	−6.56***	−7.34***
	(0.17)	(0.18)	(0.24)	(0.13)	(0.13)	(0.21)
Historic Pref. Seat	0.06	0.02	0.21***	−0.07*	−0.08*	0.06
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)
Historic County Seat	0.16***	0.11***	0.19***	0.01	−0.01	0.06
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)
Log(Population)	1.05***	1.24***	1.28***	1.00***	1.20***	1.22***
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
Adj. R^2	0.71	0.74	0.64			
Num. obs.	1,841	1,841	1,841	1,841	1,841	1,841

Notes: Heteroskedasticity-robust standard errors in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Agglomerations' shapes refer to the year 2015, the stable and corrected light to the last available year (2013) and the VIIRS light data to 2016. Omits four agglomerations with zero nighttime light emissions in either of the three data sets.

In Table 7 we repeat the regressions from Table 6 but control for population. This reveals that the higher total luminosity of historical prefecture and county seats is to a large extent driven by their higher population without robust evidence of higher

2014), its top coding corrected version by Bluhm and Krause (2018), and the 2016 VIIRS nighttime lights (Earth Observation Group, NOAA National Centers for Environmental Information, 2016; Elvidge et al., 2017). For a comparison of the different light images, see Bluhm and Krause (2018) and Düben and Krause (2020).

productivity.[43] We summarize our findings on economic history dependence as follows: Historical county and prefecture seats are significantly more populous than other cities, which again makes them more economically active.

6   Conclusion

In this paper we shed light on the interplay between geography and institutions for the emergence of cities. We exploit a unique historical data set, which not only contains the location but also the hierarchical status of cities in imperial China from 221 BCE to 1911 CE. The Chinese empire forms a particularly compelling setting for investigating this research question as it is geographically diverse, has a long history of centralized governance, experienced only moderate foreign interference compared to many other parts of the world, and displays a strong path dependence in institutions and spatial organization. To elucidate the interplay between geography, institutions and city location, we build a stylized model and conduct an empirical analysis using both econometric inference and machine learning-based prediction techniques.

Our results suggest that the direct effect of geography on a city's location decreases with that settlement's status in the institutional hierarchy. Market towns, places without government responsibilities, exhibit the closest association with geographic features within their immediate surroundings. A less dominant but still considerable effect can be found for county seats, low-ranking local administrations. Prefecture seats, important regional political centers, show no sizeable connection to local geography. Instead we find prefecture seats to locate at systematically central and strategic positions within the administrative regions they govern. In addition to this cross-sectional variation between cities of different political status, we see institutional change along the time dimension, i.e. the rise and fall of dynasties, reflected in the role geography had in placing urban settlements throughout the empire. And we observe geography to not just determine the initial location but also persistence of lower ranking cities.

Despite major institutional developments over the past century our findings bear high relevance for China's urban landscape nowadays. We find evidence both of political and economic path dependence, with former prefecture exhibiting much higher levels of both population and nighttime light emissions than other cities. In addition to providing new insights into the interplay of geography and institutions in the hierarchy of cities in general, our paper also helps to shed light on the origins of the modern urban landscape.

---

[43]Figure C-19 provides scatter plots illustrating further insights regarding this association.

References

Acemoglu, D. and J. Robinson (2019). *The Narrow Corridor: States, Societies, and the Fate of Liberty.* Penguin Press.

Ades, A. F. and E. L. Glaeser (1995). Trade and Circuses: Explaining Urban Giants. *Quarterly Journal of Economics 110*(1), 195–227.

Anderson, G. and Y. Ge (2005). The Size Distribution of Chinese Cities. *Regional Science and Urban Economics 35*, 756–776.

Atwell, W. S. (1986). Some Observations on the "Seventeenth-Century Crisis" in China and Japan. *Journal of Asian Studies 45*(2), 223–244.

Atwell, W. S. (1990). A Seventeenth-Century 'General Crisis' in East Asia? *Modern Asian Studies 24*(4), 661–682.

Bai, Y. and R. Jia (2020). The Economic Consequences of Political Hierarchy: Evidence from Regime Changes in China, AD1000-2000. Mimeo.

Bai, Y. and J. K. Kung (2011). Climate Shocks and Sino-Nomadic Conflict. *Review of Economics and Statistics 93*(3), 970–981.

Bailey, T. and A. Gatrell (1995). *Interactive Spatial Data Analysis.* Harlow: Longman Scientific and Technical.

Bairoch, P., J. Batou, and P. Chèvre (1988). *La Population des villes européennes de 800 à 1850: banque de données et analyse sommaire des résultats.* Genève: Droz.

Barjamovic, G., T. Chaney, K. Cosar, and A. Hortacsu (2019). Trade, Merchants, and the Lost Cities of the Bronze Age. *Quarterly Journal of Economics*, 1455–1503.

Baum-Snow, N., L. Brandt, J. Henderson, M. Turner, and Q. Zhang (2017). Roads, Railroads, and Decentralization of Chinese Cities. *The Review of Economics and Statistics 99*, 435–448.

Black, D. and V. Henderson (2003). Urban Evolution in the USA. *Journal of Economic Geography 3*, 343–372.

Bluhm, R. and M. Krause (2018). Top lights — Bright cities and their contribution to economic development. Cesifo working paper 7411.

Bonfatti, R. and S. Poelhekke (2017). From Mine to Coast: Transport Infrastructure and the Direction of Trade in Developing Countries. *Journal of Development Economics 127*, 91–108.

Bosker, M. and E. Buringh (2017). City Seeds: Geography and the Origins of the European City System. *Journal of Urban Economics 98*, 139–157.

Bosker, M., E. Buringh, and J. van Zanden (2013). From Baghdad to London: Unraveling Urban Development in Europe, the Middle East, and North Africa. *Review of Economics and Statistics 95*, 1418–1437.

Bosker, M., H. Garretson, H. De Jogn, and M. Schramm (2008). "Ports, Plagues and Politics: Explaining Italian City Growth 1300-1861. *European Review of Economic History 12*, 97–131.

Breiman, L. (2001). Random Forests. *Machine Learning 45*(1), 5–32.

Chen, Y., I. Overeem, A. J. Kettner, S. Gao, and J. P. M. Syvitski (2015). Modeling flood dynamics along the superelevated channel belt of the Yellow River over the last 3000 years. *Journal of Geophysical Research: Earth Surface 120*(7), 1321–1351.

Chen, Y., J. P. M. Syvitski, S. Gao, I. Overeem, and A. J. Kettner (2012). Socio-economic Impacts on Flooding: A 4000-Year History of the Yellow River, China. *AMBIO 41*(7), 682–698.

Christensen, P. and G. McCord (2016). Geographic Determinants of China's Urbanization. *Regional Science and Urban Economics 59*, 90–102.

Danielson, J. and D. Gesch (2011). Global Multi-Resolution Terrain Elevation Data 2010

(GMTED2010). U.S. Geological Survey Open-File Report 1073.

Davis, J. and J. Henderson (2003). Evidence on the Political Economy of the Urbanization Process. *Journal of Urban Economics 53*(1), 98–125.

Dijkshoorn, K., V. van Engelen, and J. Huting (2008). Global Assessment of Land Degradation - Soil and Landform Properties for LADA Partner Countries. ISRIC report, GLADA report 2008/06, 2008/03.

Donaldson, D. and A. Storeygard (2016). The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives 30*(4), 171–198.

Du, R. and J. Zhang (2019). Walled Cities and Urban Density in China. *Papers in Regional Science 98*, 1517–1539.

Düben, C. and M. Krause (2020). Population, Light, and the Size Distribution of Cities. *Journal of Regional Science* (Forthcoming).

Earth Observation Group, NOAA National Centers for Environmental Information (2016). Version 1 VIIRS Day/Night Band Nighttime Lights. https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html.

Elvidge, C. D., K. Baugh, M. Zhizhin, F. C. Hsu, and T. Ghosh (2017). VIIRS night-time lights. *International Journal of Remote Sensing 38*(21), 5860–5879.

Elvidge, C. D., F.-C. Hsu, K. Baugh, and T. Ghosh (2014). National Trends in Satellite-Observed Lighting: 1992-2012. In Q. Weng (Ed.), *Global Urban Monitoring and Assessment through Earth Observation*, Taylor & Francis Series in Remote Sensing Applications, Chapter 6, pp. 97–120. CRC Press.

Elvin, M. (2004). *The Retreat of the Elephants: An Environmental History of China*. New Haven: Yale University Press.

Fairbank Center for Chinese Studies of Harvard University and Center for Historical Geographical Studies at Fudan University (2016). CHGIS, Version: 6. https://sites.fas.harvard.edu/~chgis/data/chgis/v6/.

Florczyk, A., C. Corbane, M. Schiavina, M. Pesaresi, L. Maffenini, M. Melchiorri, P. Politis, F. Sabo, S. Freire, D. Ehrlich, T. Kemper, P. Tommasi, D. Airaghi, and L. Zanchetta (2019). GHS Urban Centre Database 2015, Multitemporal and Multidimensional Attributes, R2019A. https://data.jrc.ec.europa.eu/dataset/53473144-b88c-44bc-b4a3-4583ed1f547e.

Flückinger, M. and M. Ludwig (2017). Malaria suitability, urbanization and persistence: Evidence from China over more than 2000 years. *European Economic Review 92*, 146–160.

Fotheringham, A. and D. Wong (1991). The Modifiable Areal Unit Problem in Multivariate Analysis. *Environment and Planning 23*, 1025–1044.

Fujita, M., P. Krugman, and T. Mori (1999). On the Evolution of Hierarchical Urban Systems. *European Economic Review 43*(2), 290–251.

Funke, M. and H. Yu (2011). The Emergence and Spatial Distribution of Chinese Seaport Cities. *China Economic Review 22*, 196–209.

Georganos, S., T. Grippa, A. N. Gadiaga, C. Linard, M. Lennert, S. Vanhuysse, N. Mboga, E. Wolff, and S. Kalogirou (2019). Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International* (Forthcoming).

Hall, K. R. (2011). *A History of Early Southeast Asia: Maritime Trade and Societal Development, 100-1500*. Lanham: Rowman & Littlefield.

Henderson, J., T. Squires, A. Storeygard, and D. Weil (2018). The Global Distribution of Economic Activity: Nature, History, and the Role of Trade. *Quarterly Journal of Economics 133*(1), 357–406.

Henderson, J. and H. Wang (2007). Urbanization and City Growth: The Role of Institutions. *Regional Science and Urban Economics 37*(3), 283–313.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012). Measuring Economic Growth from Outer Space. *American Economic Review 102*(2), 994–1028.

Hirte, G., C. Lessmann, and A. Seidel (2020). International Trade, Geographic Heterogeneity and Interregional Inequality. *European Economic Review 127*(August), 103427.

Ioannides, Y. M. and J. Zhang (2017). Walled cities in late imperial China. *Journal of Urban Economics 97*, 71–88.

Jedwab, R., N. Johnson, and M. Koyama (2020). Medieval Cities Through the Lens of Urban Economic Theories. IIEP Working Paper 2020-9.

Kitamura, S. and N.-P. Lagerlöf (2020). Geography and State Fragmentation. *Journal of the European Economic Association 18*, 1726–1769.

Ko, C. Y., M. Koyama, and T. Sng (2018). Unified China and Divided Europe. *International Economic Review 59*(1), 285–327.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy 99*(3), 483–499.

Krugman, P. (1993). First Nature, Second Nature and Metropolitan Location. *Journal of Regional Science 33*(2), 129–144.

Kuhn, M. and K. Johnson (2013). *Applied Predictive Modeling.* Springer Science and Business Media New York.

Lewis, M. E. (2009). *China Between Empires: The Northern and Southern Dynasties.* Cambridge: Belknap Press of Harvard University Press.

Li, C., J. Li, and J. Wu (2018). What Drives Urban Growth in China? A Multi-Scale Comparative Analysis. *Applied Geography 98*, 43–51.

Lin, S. and S. Song (2002). Urban Economic Growth in China: Theory and Evidence. *Urban Studies 39*, 2251–2266.

Ma, L. (2002). Urban Transformation in China, 1949 – 2000: A Review and Research Agenda. *Environment and Planning A 34*, 1545–1569.

Major, J. S. and C. A. Cook (2017). *Ancient China: A History.* Oxon: Routledge.

Marks, R. B. (1998). *Tigers, Rice, Silk, and Silt: Environment and Economy in Late Imperial South China.* Cambridge: Cambridge University Press.

Marks, R. B. (2012). *China: Its Environment and History.* Lanham: Rowman & Littlefield.

Matsuura, K. and C. J. Willmott (2018a). Terrestrial Air Temperature: 1900-2017 Gridded Monthly Time Series, Version 5.01. http://climate.geog.udel.edu/~climate/html_pages/download.html#T2017.

Matsuura, K. and C. J. Willmott (2018b). Terrestrial Precipitation: 1900-2017 Gridded Monthly Time Series, Version 5.01. http://climate.geog.udel.edu/~climate/html_pages/download.html#T2017.

Michaels, G. and F. Rauch (2018). Resetting the Urban Network: 117-2012. *The Economic Journal 128*, 378–412.

Mitton, T. (2016). The Wealth of Subnations: Geography, Institutions, and Within-Country Development. *Journal of Development Economics 118*, 88–111.

Mostern, R. (2011). *Dividing the Realm in Order to Govern: The Spatial Organization of the Song State (960 - 1276 CE).* Cambridge: Harvard University Asia Center.

Motamed, M., R. Florax, and W. Masters (2014). Agriculture, Transportation and the Timing of Urbanization: Global Analysis at the Grid Cell Level. *Journal of Economic Growth 19*(3),

339–368.

Mote, F. W. (1999). *Imperial China: 900 - 1800*. Cambridge: Harvard University Press.

Márquez-Pérez, J., I. Vallejo-Villalta, and J. I. Álvarez-Francoso (2017). Estimated travel time for walking trails in natural areas. *Geografisk Tidsskrift-Danish Journal of Geography 117*(1), 53–62.

Natural Earth (2019). Rivers, Lake Centerlines, Version 4.1.0. https://www.naturalearthdata.com/downloads/50m-physical-vectors/50m-rivers-lake-centerlines/.

Nunn, N. and D. Puga (2012). Ruggedness: The Blessing of Bad Geography in Africa. *Review of Economics and Statistics 94*(1), 20–36.

Perdue, P. C. (1987). *Exhausting the Earth: State and Peasant in Hunan, 1500–1850*. Cambridge: Harvard University Asia Center, Harvard University.

Proost, S. and J.-F. Thisse (2019). What Can Be Learned from Spatial Economics? *Journal of Economic Literature 57*(3), 575–643.

Redding, S. (2020). Trade and Geography. NBER Working Paper No. 27821.

Scheidel, W. (2019). *Escape from Rome (The Princeton Economic History of the Western World)*. Princeton: Princeton University Press.

Schönhölzer, D. and E. Weese (2019). Creative Destruction in the European State System: 1000-1850. Working paper stockholm university.

Shi, H., B. Wang, E. R. Cook, J. Liu, and F. Liu (2018). Asian summer precipitation over the past 544 years reconstructed by merging tree rings and historical documentary records. *Journal of Climate 31*(19), 7845–7861.

Skinner, W. G. (1977a). Cities and the Hierarchy of Local Systems. In W. G. Skinner (Ed.), *The City in Late Imperial China*, pp. 275–351. Stanford: Stanford University Press.

Skinner, W. G. (1977b). Regional Urbanization in Nineteenth-Century China. In W. G. Skinner (Ed.), *The City in Late Imperial China*, pp. 211–249. Stanford: Stanford University Press.

Soo, K. T. (2005). Zipf's Law for cities: a cross-country investigation. *Regional Science and Urban Economics 35*(3), 239–263.

Tan, R., Y. Liu, Y. Liu, Q. He, L. Ming, and S. Tang (2014). Urban Growth and its Determinants Across the Wuhan Urban Agglomeration, Central China. *Habitat International 44*, 268–281.

Tobler, W. (1993). Three Presentations on Geographical Analysis and Modeling: Non- Isotropic Geographic Modeling, Speculations on the Geometry of Geography, Global Spatial Analysis. Technical Report 93-1, National Center for Geographic Information and Analysis.

Tuotuo (1346). *Song Shi*. Original: Hangzhou, Yuan (1346), Reprint: Beijing, Zhong hua shu ju (1977).

von Glahn, R. (2016). *The Economic History of China: From Antiquity to the Nineteenth Century*. Cambridge: Cambridge University Press.

Wessel, P. and W. H. F. Smith (1996). A Global Self-consistent, Hierarchical, High-resolution Shoreline Database. *Journal of Geophysical Research 101*, 8741–8743.

Wessel, P. and W. H. F. Smith (2017). Global Self-consistent Hierarchical High-resolution Geography, GSHHG, Version 2.3.7. https://www.ngdc.noaa.gov/mgg/shorelines/.

Wilkinson, E. (2013). *Chinese History: A New Manual*. Cambridge: Harvard University Asia Center for the Harvard-Yenching Institute.

Will, P.-E. (1985). State Intervention in the Administration of a Hydraulic Infrastructure. In S. R. Schram (Ed.), *The Scope of State Power in China*, pp. 295–347. London: School of Oriental and African Studies, University of London.

Will, P.-E. (1990). *Bureaucracy and Famine in Eighteenth-Century China*. Stanford: Stanford University Press.

Will, P.-E. and R. B. Wong (1991). *Nourish the People: The State Civilian Granary System in China, 1650-1850*. Ann Arbor: Center for Chinese Studies, University of Michigan.

Williams, T. (2020). Silk Roads - Main Corridors. http://worldmap.harvard.edu/maps/7547/info/.

Wu, J., W.-N. Xiang, and J. Zhao (2014). Urban Ecology in China: Historical Developments and Future Direction. *Landscape and Urban Planning 125*, 222–233.

Wu, W. and P. Gaubatz (2013). *The Chinese City*. Oxon: Routledge.

You, H. and X. Yang (2017). Urban Expansion in 30 Megacities of China: Categorizing the Driving Force Profiles to Inform the Urbanization Policy. *Land Use Policy 68*, 531–551.

Zhang, H., J. P. Werner, E. García-Bustamante, F. González-Rouco, S. Wagner, E. Zorita, K. Fraedrich, J. H. Jungclaus, F. C. Ljungqvist, X. Zhu, E. Xoplaki, F. Chen, J. Duan, Q. Ge, Z. Hao, M. Ivanov, L. Schneider, S. Talento, J. Wang, B. Yang, and J. Luterbacher (2018). East Asian warm season temperature variations over the past two millennia. *Scientific Reports 8*(7702).

Zhang, Z., S. Su, R. Xiao, D. Jiang, and J. Wu (2013). Identifying Determinants of Urban Growth from a Multi-Scale Perspective: A Case Study of the Urban Agglomeration Around Hangzhou Bay, China. *Applied Geography 45*, 193–202.

Zhao, G. (2013). *The Qing opening to the ocean: Chinese maritime policies, 1684 - 1757*. Honolulu: University of Hawai'i Press.

A Theoretical Framework: Proofs

A.1 Special Cases

The main paper lays out the emperor's utility maximization problem as

$$\max_{N_c,N_p,locations} U\left[\underbrace{\sum_{c=1}^{N_c}\sum_{i=1}^{N_{ic}}(1-\tau D_{ci})\lambda P_i}_{+}, \underbrace{\sum_{p=1}^{N_p}\sum_{c=1}^{N_{cp}}D_{pc}}_{-}, \underbrace{\sum_{p=1}^{N_p}\sum_{i=1}^{N_{ip}}(D_{pi}M_i)^m}_{-}\right] \quad ((\text{A-1}))$$

subject to the county seat constraints

$$G_c + E_c \leq \sum_{i=1}^{N_{ic}}(1-\tau D_{ci})\lambda P_i \quad \text{for} \quad 1 \leq c \leq N_c \quad ((\text{A-2}))$$

$$E_c \geq 0 \quad \text{for} \quad 1 \leq c \leq N_c \quad ((\text{A-3}))$$

and prefecture seat constraints

$$G_p \leq \sum_{c=1}^{N_{pc}}(E_c - \tau D_{pc}E_c) + \sum_{q\neq p}^{N_p}(E_{pq} - \tau D_{pq}E_{pq}) \quad \text{for} \quad 1 \leq p \leq N_p \quad ((\text{A-4}))$$

No closed-form solution for the number of location of the county and prefecture seats is possible. But it is insightful to consider special cases which allow us to further simplify the expressions involved.

Assumptions 4. *All county seats carry the same cost G_c and they all export the same amount E_c to the prefecture seats*

$$G_c = \bar{G} \quad \text{for} \quad 1 \leq c \leq N_c \quad ((\text{A-5}))$$

$$E_c = \bar{E} \quad \text{for} \quad 1 \leq c \leq N_c \quad ((\text{A-6}))$$

Imposing this additional assumption allows us to write the aggregate-level resources constraint

$$\sum_{c=1}^{N_c}(G_c + E_c) \leq \sum_{c=1}^{N_c}\sum_{i=1}^{N_{ic}}(1-\tau D_{ci})\lambda P_i \quad ((\text{A-7}))$$

as

$$N_c \cdot (\bar{G} + \bar{E}) \leq \sum_{i=1}^{N_i} (1 - \tau D_{ci})\lambda P_i \qquad \text{((A-8))}$$

$$N_c \leq \frac{\sum_{i=1}^{N_i}(1 - \tau D_{ci})\lambda P_i}{\bar{G} + \bar{E}} \qquad \text{((A-9))}$$

Note that this is not a closed-form solution for N_c as the distance D_{ci} between cell i and prefecture seat c depends on it. Yet, it is independent from considerations at the prefecture seat level and illustrates the trade-off: A higher population P_i will make it worthwhile to have more county seats, by boosting tax revenues, while higher expenses in the form of $\bar{G}$ and $\bar{E}$ will lead to fewer county seats.

At the prefecture seat level, Assumption 4 allows to simplify

$$\sum_{p=1}^{N_p} G_p \leq \sum_{i=1}^{N_p} \sum_{i=1}^{N_{pc}} (E_c - \tau D_{pc}E_c) + \sum_{p=1}^{N_p} \sum_{q \neq p}^{N_p}(E_{pq} - \tau D_{pq}E_{pq}) \qquad \text{((A-10))}$$

to

$$N_p \bar{G} \leq N_p \bar{E} \sum_{i=1}^{N_{pc}}(1 - \tau D_{pc}) + N_p \bar{E} \sum_{q \neq p}^{N_p}(1 - \tau D_{pq}) \qquad \text{((A-11))}$$

$$\bar{G} \leq \bar{E}\Big(\sum_{i=1}^{N_{pc}}(1 - \tau D_{pc}) + \sum_{q \neq p}^{N_p}(1 - \tau D_{pq})\Big) \qquad \text{((A-12))}$$

The number N_p of prefecture seats can be determined implicitly from this equation. It balances the costs of each prefecture seat with the receipts, net of all transport costs.

A.2 Proof of Theorem 1

We can show the effects described in Theorem 1 based on the equations laid out before.

(a) In areas with favorable geography, there are more county seats.

Proof: By Assumption 1, favorable geography A_i in cell i increases population P_i in this cell. The maximizing problem eq. (8) depends negatively on the population-weighted travel-time distance $D_{ci}P_i$ between cell i and the nearest county seat. Decreasing the distance to populous cells has the strongest effect on utility, leading to more county seats in areas with favorable geography.

(b) The location of prefecture seats is determined to a lesser extent and only indirectly by the presence of geographical features.

Proof: In contrast to the direct geographical effect of geography on county seat density via population, the prefecture seat locations are chosen primarily on military considerations. While favorable geography A_i increases P_i according to Assumption 1, the latter term in eq. (8) does not include P_i. There is only an indirect channel through which geography affects the location of prefecture seats: County seat density is determined by favorable geography according to (a), and according to eq. (13), it is advantageous for prefecture seats to locate close to county seats to minimize the transport costs of the transfers E_c.

(c) There are more prefecture seats per county seat in regions that are prone to military invasion.

Proof: Military threats affect the location of prefecture seats directly and that of county seats only indirectly. A military threat M_i in cell i has a negative effect on utility in eq. (8), which is reinforced by the non-linearity of $m > 1$. It can only be compensated by decreasing the distance D_{ip} to the nearest prefecture seat. This leads to an increase in the number of prefecture seats, while the number of county seats will either stay the same or might decrease through an indirect effect: According to 1, the total threat level increases in response the increase in M_i, decreasing population P_i. Following the mechanism discussed in (a), a decreased population leads to fewer county seats, reinforcing the result that there are more prefecture seats per county seat in regions that are prone to military invasion.

(d) If the population grows at a higher rate than the costs of maintaining county seats, the optimal number of county seats will increase.

Proof: According to eq. (8), the optimal number N_c of county seats will always be N_i, but it is constrained by the financing restrictions eq. (9) and eq. (10) for each county seat. If P_i increases at a higher rate than G_c, the aggregate budget constraint eq. ((A-7)) does not bind anymore. Resources for establishing additional county seats are then available.

B   Additional Information on Geographical Data

In this part of the Appendix, we present additional information on the data used in our empirical analysis. Table B-1 provides the detailed data sources of the geographical variables. Table B-2, Table B-3 and Table B-4 contain the summary statistics of, respectively, the dominant soil type, landform, and lithology.

Table B-1: Geography Data Sources

Variable	Source	Original Format	Derivations
Distance from equator	Own computations	Raster (5 arc minutes)	Distance calculation between pixel centroid and equator using latitude
Distance from coast	Wessel and Smith (1996, 2017)	Polygon	Distance calculation between land pixel centroid and centroid of nearest ocean pixel
Distance from river	Natural Earth (2019)	Spatial lines	Distance calculation between pixel centroid and nearest river's spatial line
Ruggedness	Nunn and Puga (2012)	Raster (30 arc seconds)	Grid cells containing index measuring elevation difference between grid cells in mm aggregated to 5 arc minute level using averages
Temperature	Matsuura and Willmott (2018a)	Raster (30 arc minutes)	Averages of monthly temperature data from 1900 - 1950 disaggregated to 5 arc minute level using bilinear interpolation
Precipitation	Matsuura and Willmott (2018b)	Raster (30 arc minutes)	Averages of annual precipitation in mm from 1900 - 1950 disaggregated to 5 arc minute level using bilinear interpolation
Elevation	Danielson and Gesch (2011)	Raster (7.5 arc seconds)	Grid cells aggregated to 5 arc minute level using averages
Dominant soil type	Dijkshoorn et al. (2008)	Polygon	Polygons converted to grid cells
Landform	Dijkshoorn et al. (2008)	Polygon	Polygons converted to grid cells
Lithology	Dijkshoorn et al. (2008)	Polygon	Polygons converted to grid cells

Notes: Regressions use this data converted to an equal-area Mollweide projection. The average monthly temperature does not entail more information on the time dimension than average annual precipitation does. Both variables sum up observations from 612 months between 1900 and 1951. In the case of temperature, we then divide that sum by 612 and in the case of precipitation by 51. Cumulating precipitation measured in mm over a year makes sense, while cumulating temperature measured in degrees Celsius does not.

Table B-2: Dominant Soil Summary Statistics

Abbreviation	Full Name	Frequency	Abbreviation	Full Name	Frequency
ACf	ferric acrisols	2682	HSs	terric histosols	42
ACh	haplic acrisols	5189	KSh	hapic kastanozems	993
ACp	plinthic acrisols	104	KSk	caleic kastanozems	289
ACu	humic acrisols	3649	KSl	luvic kastanozems	866
ALf	ferric alisols	201	LP	leptosols	53
ALh	haplic alisols	3271	LPd	dystric leptosols	69
ALp	plinthic alisols	78	LPd	dystric leptosols	69
ANh	haplic andosols	46	LPe	eutric leptosols	1807
ANu	umbric andosols	4	LPi	gelic leptosols	3573
ARb	cambic arenosols	830	LPk	rendzic leptosols	678
ARc	calcaric arenosols	790	LPm	mollic leptosols	1651
ARh	haplic arenosols	1213	LVa	albic luvisols	696
ATa	aric anthrosols	46	LVg	gleyic luvisols	1084
ATc	cumulic anthrosols	5691	LVh	haplic luvisols	12405
ATf	fimic anthrosols	57	LVj	stagnic luvisols	22
CHg	gleyic chernozems	193	LVk	calcic luvisols	975
CHh	haplic chernozems	517	LVx	chromic luvisols	815
CHk	calcic chernozems	749	LXa	albic lixisols	8
CHl	luvic chernozems	227	LXf	ferric lisols	167
CLh	haplic calcisols	481	NTu	humic nitosols	89
CLl	luvic calcisols	424	PDd	dystric podzoluvisols	91
CLp	petric calcisols	53	PDj	stagnic podzoluvisols	2
CMc	calcaric cambisols	5682	PHc	calcaric phaeozems	599
CMd	dystric cambisols	3308	PHg	gleyic phaeozems	955
CMe	eutric cambisols	1318	PHh	haplic phaeozems	2673
CMg	gleyic cambisols	451	PHj	stagnic phaeozems	91
CMi	gelic cambisols	301	PLd	dystric planosols	76
CMo	ferralic cambisols	469	PLe	eutric planosols	552
CMu	humic cambisols	20	RGc	calcaric regosols	1559
CMx	chromic cambisols	611	RGd	dystric regosols	626
FLc	calcaric fluvisols	3987	RGe	eutric regosols	695
FLe	eutric fluvisols	350	RK	rock, rock outcrop	177
FLs	salic fluvisols	695	SC	solonchaks	1
FP	fishpond	18	SCg	gleyic solonchaks	46
FRh	heplc ferralsols	252	SCh	haplic solonchaks	192
FRx	xanthic ferralsols	146	SCk	calcic solonchaks	107
GG	glaciers, ice	139	SCm	mollic solonchaks	233
GLe	eutric gleysols	125	SCn	sodic solonchaks	7
GLi	geic gleysols	28	SCy	gypsic solonchaks	55
GLk	caleic gleysols	219	SNg	gleyic solonetz	103
GLm	molic gleysols	1413	SNh	haplic solonetz	3
GLt	thionic gleysols	12	SNk	calcic solonetz	2
GRh	hapic greyzems	618	ST	salt flats	323
GYh	heplc gypsisols	36	UR	urban areas	28
GYk	calcic gypsisols	801	VRd	distric vertsols	71
GYl	luvic gypsisols	765	VRe	eutric vertsols	302
GYp	petric gypsisols	1415	VRk	calcic vertsols	28
HSf	fibric histosols	3	WR	inland water, lakes	701

Notes: The frequency refers to the number of pixels in the baseline setting, which are 7.33 x 9.51 km in size.

Table B-3: Landform Summary Statistics

Abbreviation	Full Name	Frequency	Abbreviation	Full Name	Frequency
LP	plain	37626	SP	dissected plain	556
LP wet	plain wet	104	TH	high-gradient hill	2988
SH	medium-gradient hill	28235	TM	high-gradient mountain	14745
SM	medium-gradient mountain	1302	WR	inland water, lakes	701

Notes: The frequency refers to the number of pixels in the baseline setting, which are 7.33 x 9.51 km in size.

Table B-4: Lithology Summary Statistics

Abbr.	Full Name	Freq.	Abbr.	Full Name	Freq.
FP	fishpond	18	SC4	shale	6198
GG	glaciers, ice	139	SC5	ironstone	347
IA	acid igneous rock	45	SC7		142
IA1	granite	13301	SO1	limestone, other carbonate rocks	6195
IA2	grano-diorite	37	SO2	marl and other mixtures	661
IA4	rhyolite	32	SO3	coals, bitumen and related rocks	238
IB2	basalt	938	ST	salt flats	323
IB3	dolerite	4	UA1	redeposited natural material	11
II1	andesite, trachyte, phonolite	760	UE	eolean	14
II2	diorite-syenite	51	UE1	loess	12898
IP2	volcanic scoria/ breccia	137	UE1/UR1	loess/ bauxite, laterite	51
IP4	ignimbrite	1500	UE2	sand	2711
MA	acid metamorphic rock	20	UF	fluvial	15037
MA1	quartzite	336	UF/UL	fluvial/ lacustrine unconsolidated rock	511
MA1/SC2	quartzite/ sandstone, greywacke, arkose	71	UF/UM	fluvial/ marine unconsolidated rock	155
MA2	gneiss, migmatite	1309	UF1	sand and gravel	150
MA3/MB1	slate, phyllite (pelitic rocks)	1885	UF2	clay, silt and loam	161
MA4/MB2	schist	677	UG	glacial	243
MB1	slate, phyllite (pelitic rocks)	1877	UL	lacustrine	1670
MB1/MB2	slate, phyllite (pelitic rocks)/ schist	2242	UL/UM	lacustrine unconsolidated rock/ marine unconsolidated rock	26
RK	rock outcrop	177			
SC	clastic sedimentary rock	885	UL2	silt and clay	6
SC1	conglomerate, breccia	792	UM	marine unconsolidated rock	202
SC10		1	UM/UF	marine unconsolidated rock/ fluvial	5
SC16	glacial sedimentary environments	459	UO	organic	4
SC2	sandstone, greywacke, arkose	6300	UR	urban areas	28
SC2/SC4	sandstone, greywacke, arkose/ shale	243	UR1	bauxite, laterite	1775
SC3	siltstone, mudstone, claystone	1558	WR	lakes, permanent water	701

Notes: The frequency refers to the number of pixels in the baseline setting, which are 7.33 x 9.51 km in size.

C Supplementary Empirical Results

Here we provide additional results and robustness checks to supplement the empirical analysis in the paper.

C.1 Supplementary OLS Results

In Section 4.1, we discuss the effect of using increasingly larger pixels based on random forest results. The following Table C-1 to Table C-8 confirm those findings with OLS, comparing the effect on county and prefecture seats when increasingly larger pixels are used. While county seats show a strong connection to local geography, such a clear link does not exist for prefecture seats.

Table C-1: Local Geography Regressions (County Seats, Small Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.15***	0.23***	0.18***	0.29***	0.36***	0.46***	0.22***	0.21***	0.25***	0.30***	0.28***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Dist. Equator	-0.22***	-0.30***	-0.28***	-0.46***	-0.61***	-0.79***	-0.49***	-0.48***	-0.53***	-0.61***	-0.56***
	(0.06)	(0.08)	(0.08)	(0.08)	(0.09)	(0.10)	(0.10)	(0.10)	(0.09)	(0.10)	(0.11)
Dist. Coast	0.01	-0.12**	0.07	0.25***	0.40***	0.42***	0.42***	0.38***	0.14**	0.23***	0.16**
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Dist. River	-0.03	-0.21*	-0.24**	-0.51***	-0.96***	-0.52***	-0.27*	-0.17	-0.20	-0.22	-0.10
	(0.10)	(0.12)	(0.12)	(0.13)	(0.12)	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)	(0.16)
Ruggedness	-0.27***	-0.34***	-0.33***	-0.57***	-0.29***	-0.73***	-0.61***	-0.59***	-0.54***	-0.58***	-0.63***
	(0.07)	(0.09)	(0.09)	(0.09)	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)	(0.13)
Temperature	0.25***	0.36***	0.28***	0.07	-0.09	-0.36***	-0.07	-0.05	-0.08	-0.07	-0.03
	(0.05)	(0.07)	(0.07)	(0.07)	(0.08)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Temperature^2	-0.38**	0.04	-0.23	-0.76***	1.36***	1.52***	0.76**	0.47	0.15	-0.36	-0.36
	(0.18)	(0.25)	(0.25)	(0.29)	(0.32)	(0.40)	(0.37)	(0.37)	(0.36)	(0.40)	(0.41)
Precipitation	-0.68***	-1.10***	-0.58***	-0.21	-0.79***	-0.28*	0.17	0.21	-0.03	0.12	0.15
	(0.11)	(0.13)	(0.12)	(0.13)	(0.14)	(0.16)	(0.15)	(0.15)	(0.15)	(0.16)	(0.16)
Precipitation^2	1.14***	1.86***	0.89***	0.31	1.25***	0.20	-0.57*	-0.67**	-0.18	-0.49	-0.52*
	(0.22)	(0.26)	(0.25)	(0.26)	(0.29)	(0.31)	(0.30)	(0.30)	(0.29)	(0.32)	(0.32)
Elevation	-1.26***	-1.05***	-1.19***	-2.36***	-3.82***	-4.21***	-3.27***	-3.12***	-2.66***	-3.16***	-2.82***
	(0.29)	(0.38)	(0.37)	(0.37)	(0.39)	(0.45)	(0.44)	(0.45)	(0.44)	(0.49)	(0.48)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.09	0.12	0.10	0.07	0.10	0.09	0.09	0.08	0.07	0.07	0.07
Num. obs.	$21,597$	$21,597$	$21,597$	$21,597$	$21,597$	$21,597$	$21,597$	$21,597$	$21,597$	$21,597$	$21,597$

Notes: The table reports the regression results of eq. (14) using the county seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

Table C-2: Local Geography Regressions (Prefecture Seats, Small Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.02*	0.03**	0.01	0.07***	0.09***	0.16***	0.14***	0.18***	0.09***	0.09***	0.09***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Dist. Equator	−0.03*	−0.06**	−0.03	−0.14***	−0.20***	−0.25***	−0.25***	−0.33***	−0.14***	−0.14***	−0.15***
	(0.02)	(0.02)	(0.03)	(0.04)	(0.05)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
Dist. Coast	−0.00	0.01	0.03	0.12***	0.17***	0.08**	0.07*	0.09**	0.04	0.03	0.01
	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Dist. River	0.00	−0.03	−0.03	−0.11*	−0.20***	−0.22***	−0.14*	−0.19**	−0.08	−0.08	−0.09
	(0.03)	(0.04)	(0.05)	(0.07)	(0.06)	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)	(0.08)
Ruggedness	−0.04***	−0.07***	−0.06**	−0.13***	−0.12**	−0.29***	−0.20***	−0.17**	−0.31***	−0.28***	−0.20***
	(0.01)	(0.02)	(0.02)	(0.04)	(0.06)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)
Temperature	−0.01	0.01	0.04*	−0.00	−0.03	−0.16***	−0.11**	−0.17***	−0.03	−0.02	−0.05
	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.05)
Temperature^2	0.02	−0.03	−0.05	−0.16	−0.18	0.61***	0.20	−0.00	−0.13	−0.06	−0.08
	(0.06)	(0.07)	(0.08)	(0.14)	(0.14)	(0.22)	(0.21)	(0.22)	(0.19)	(0.19)	(0.18)
Precipitation	−0.06**	−0.12***	−0.06	−0.13**	−0.07	−0.15*	−0.14	−0.16*	−0.02	−0.06	−0.06
	(0.03)	(0.04)	(0.04)	(0.07)	(0.07)	(0.08)	(0.09)	(0.09)	(0.07)	(0.07)	(0.08)
Precipitation^2	0.10*	0.23***	0.10	0.24*	0.08	0.20	0.18	0.22	−0.04	0.05	0.05
	(0.05)	(0.08)	(0.09)	(0.14)	(0.15)	(0.16)	(0.18)	(0.17)	(0.14)	(0.15)	(0.15)
Elevation	−0.12	−0.14	−0.07	−0.56***	−1.17***	−0.96***	−1.06***	−1.44***	−0.09	−0.11	−0.43*
	(0.08)	(0.11)	(0.13)	(0.20)	(0.21)	(0.26)	(0.27)	(0.28)	(0.23)	(0.24)	(0.24)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.00	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.02
Num. obs.	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597

Notes: The table reports the regression results of eq. (14) using the prefecture seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

Table C-3: Local Geography Regressions (County Seats, Medium Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.17**	0.29***	0.22**	0.45***	0.59***	0.65***	0.30**	0.30**	0.40***	0.51***	0.48***
	(0.08)	(0.10)	(0.10)	(0.10)	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)	(0.13)
Dist. Equator	−0.23*	−0.37**	−0.37**	−0.70***	−0.93***	−1.22***	−0.80***	−0.76***	−0.85***	−0.99***	−0.87***
	(0.13)	(0.17)	(0.17)	(0.17)	(0.18)	(0.21)	(0.20)	(0.20)	(0.20)	(0.22)	(0.22)
Dist. Coast	−0.14	−0.32***	0.07	0.45***	0.80***	0.86***	0.89***	0.75***	0.30**	0.49***	0.31**
	(0.10)	(0.11)	(0.11)	(0.11)	(0.12)	(0.12)	(0.13)	(0.13)	(0.12)	(0.13)	(0.13)
Dist. River	0.07	−0.13	−0.17	−0.66***	−1.49***	−0.79***	−0.25	−0.10	−0.36	−0.58*	−0.26
	(0.20)	(0.24)	(0.24)	(0.25)	(0.25)	(0.29)	(0.29)	(0.30)	(0.30)	(0.31)	(0.33)
Ruggedness	−0.50***	−0.57***	−0.77***	−1.28***	−0.66***	−1.21***	−1.35***	−1.33***	−1.32***	−1.42***	−1.42***
	(0.15)	(0.21)	(0.21)	(0.21)	(0.24)	(0.26)	(0.26)	(0.26)	(0.27)	(0.28)	(0.28)
Temperature	0.58***	0.81***	0.61***	0.27*	0.12	−0.48**	0.00	0.10	0.07	0.07	0.13
	(0.12)	(0.16)	(0.15)	(0.15)	(0.17)	(0.19)	(0.19)	(0.19)	(0.18)	(0.20)	(0.20)
Temperature^2	−0.25	0.35	−0.14	−1.40**	2.42***	3.96***	2.18***	1.36*	0.94	0.42	0.20
	(0.42)	(0.57)	(0.58)	(0.61)	(0.67)	(0.79)	(0.75)	(0.76)	(0.76)	(0.82)	(0.84)
Precipitation	−1.15***	−1.86***	−0.84***	−0.16	−1.31***	−0.15	0.55*	0.59**	0.14	0.40	0.45
	(0.22)	(0.26)	(0.25)	(0.26)	(0.29)	(0.31)	(0.30)	(0.30)	(0.31)	(0.32)	(0.33)
Precipitation^2	1.86***	3.07***	1.15**	0.10	2.04***	−0.32	−1.58***	−1.69***	−0.92	−1.55**	−1.55**
	(0.45)	(0.53)	(0.52)	(0.53)	(0.60)	(0.63)	(0.61)	(0.62)	(0.62)	(0.66)	(0.67)
Elevation	−1.95***	−1.33*	−2.01**	−4.12***	−6.42***	−7.31***	−6.05***	−5.39***	−4.23***	−5.17***	−4.56***
	(0.62)	(0.81)	(0.80)	(0.79)	(0.83)	(0.96)	(0.93)	(0.94)	(0.94)	(1.03)	(1.02)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.17	0.20	0.17	0.12	0.19	0.18	0.18	0.17	0.14	0.15	0.14
Num. obs.	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590

Notes: The table reports the regression results of eq. (14) using the county seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

Table C-4: Local Geography Regressions (Prefecture Seats, Medium Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.05*	0.08**	0.04	0.17***	0.17***	0.32***	0.26***	0.35***	0.16**	0.18**	0.20***
	(0.03)	(0.03)	(0.04)	(0.06)	(0.06)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)
Dist. Equator	−0.08*	−0.10*	−0.04	−0.27***	−0.36***	−0.53***	−0.46***	−0.66***	−0.31***	−0.32***	−0.35***
	(0.04)	(0.06)	(0.06)	(0.10)	(0.10)	(0.13)	(0.14)	(0.14)	(0.12)	(0.12)	(0.12)
Dist. Coast	0.02	0.03	0.05	0.23***	0.40***	0.22***	0.22***	0.33***	0.16**	0.16**	0.10
	(0.03)	(0.04)	(0.05)	(0.07)	(0.07)	(0.08)	(0.08)	(0.09)	(0.07)	(0.07)	(0.07)
Dist. River	−0.03	−0.04	−0.07	−0.21	−0.36***	−0.50***	−0.30*	−0.39**	−0.15	−0.18	−0.09
	(0.07)	(0.09)	(0.10)	(0.15)	(0.14)	(0.16)	(0.17)	(0.18)	(0.16)	(0.17)	(0.17)
Ruggedness	0.00	−0.05	−0.12*	−0.32***	−0.18	−0.49***	−0.46***	−0.39**	−0.51***	−0.47***	−0.41**
	(0.05)	(0.06)	(0.07)	(0.10)	(0.13)	(0.15)	(0.17)	(0.16)	(0.16)	(0.16)	(0.16)
Temperature	−0.03	0.04	0.16***	−0.03	−0.04	−0.31***	−0.17	−0.29**	−0.06	−0.02	−0.08
	(0.04)	(0.05)	(0.06)	(0.09)	(0.09)	(0.12)	(0.13)	(0.13)	(0.11)	(0.11)	(0.11)
Temperature^2	−0.01	−0.22	−0.35*	−0.10	0.00	1.29***	0.63	0.11	−0.06	−0.01	0.06
	(0.16)	(0.19)	(0.21)	(0.32)	(0.33)	(0.49)	(0.48)	(0.50)	(0.44)	(0.45)	(0.44)
Precipitation	−0.08	−0.22***	−0.22**	−0.27**	−0.15	−0.29	−0.25	−0.31*	0.05	−0.15	−0.22
	(0.05)	(0.07)	(0.09)	(0.13)	(0.15)	(0.18)	(0.19)	(0.19)	(0.16)	(0.17)	(0.17)
Precipitation^2	0.13	0.41***	0.39*	0.49*	0.23	0.46	0.35	0.51	−0.24	0.16	0.31
	(0.11)	(0.15)	(0.20)	(0.26)	(0.31)	(0.37)	(0.38)	(0.38)	(0.33)	(0.35)	(0.34)
Elevation	−0.37*	−0.26	0.05	−0.97**	−1.99***	−1.83***	−1.69***	−2.67***	−0.11	−0.10	−0.86
	(0.20)	(0.25)	(0.27)	(0.46)	(0.47)	(0.59)	(0.61)	(0.66)	(0.58)	(0.59)	(0.57)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.01	0.01	0.03	0.03	0.04	0.04	0.05	0.05	0.03	0.03	0.03
Num. obs.	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590

Notes: The table reports the regression results of eq. (14) using the prefecture seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

Table C-5: Local Geography Regressions (County Seats, Large Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.19	0.48***	0.37**	0.75***	0.82***	1.12***	0.52***	0.58***	0.77***	0.86***	0.91***
	(0.13)	(0.17)	(0.17)	(0.17)	(0.18)	(0.20)	(0.19)	(0.19)	(0.20)	(0.21)	(0.22)
Dist. Equator	−0.22	−0.65**	−0.52*	−1.15***	−1.55***	−1.94***	−1.05***	−1.23***	−1.38***	−1.61***	−1.73***
	(0.23)	(0.28)	(0.29)	(0.29)	(0.30)	(0.34)	(0.33)	(0.33)	(0.34)	(0.36)	(0.37)
Dist. Coast	−0.07	−0.27	0.09	0.51***	1.20***	1.06***	1.02***	0.97***	0.39**	0.60***	0.41*
	(0.16)	(0.18)	(0.18)	(0.18)	(0.19)	(0.19)	(0.20)	(0.20)	(0.20)	(0.20)	(0.21)
Dist. River	−0.23	−0.70*	−0.69*	−1.43***	−2.22***	−1.62***	−0.85*	−0.80*	−1.15**	−1.42***	−0.81
	(0.34)	(0.39)	(0.40)	(0.41)	(0.42)	(0.46)	(0.46)	(0.46)	(0.46)	(0.47)	(0.50)
Ruggedness	−0.61**	−0.81**	−1.00***	−1.88***	−0.59	−1.70***	−1.69***	−1.84***	−1.93***	−1.73***	−2.06***
	(0.26)	(0.37)	(0.37)	(0.37)	(0.41)	(0.44)	(0.43)	(0.43)	(0.45)	(0.46)	(0.46)
Temperature	1.11***	1.14***	1.07***	0.38	−0.13	−0.80***	0.13	0.03	−0.05	−0.11	−0.12
	(0.21)	(0.27)	(0.27)	(0.26)	(0.28)	(0.31)	(0.30)	(0.30)	(0.30)	(0.31)	(0.32)
Temperature^2	−1.51**	−0.75	−1.10	−2.05*	3.65***	4.49***	2.63**	1.96	0.66	0.92	0.59
	(0.74)	(0.98)	(0.99)	(1.06)	(1.12)	(1.26)	(1.25)	(1.25)	(1.26)	(1.33)	(1.35)
Precipitation	−1.34***	−2.44***	−1.40***	−0.52	−1.72***	−0.51	0.58	0.67	0.40	0.61	0.31
	(0.35)	(0.41)	(0.41)	(0.42)	(0.44)	(0.48)	(0.47)	(0.47)	(0.47)	(0.50)	(0.51)
Precipitation^2	2.11***	4.06***	2.11**	0.66	2.58***	−0.05	−2.01**	−2.20**	−1.87*	−2.44**	−1.75*
	(0.72)	(0.84)	(0.84)	(0.85)	(0.91)	(0.99)	(0.97)	(0.97)	(0.98)	(1.02)	(1.03)
Elevation	−2.62**	−2.84**	−3.21**	−6.91***	−11.63***	−12.49***	−9.09***	−9.34***	−8.88***	−9.90***	−9.90***
	(1.04)	(1.33)	(1.34)	(1.33)	(1.42)	(1.54)	(1.52)	(1.52)	(1.56)	(1.69)	(1.70)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.21	0.24	0.22	0.17	0.26	0.26	0.25	0.24	0.22	0.23	0.22
Num. obs.	5,402	5,402	5,402	5,402	5,402	5,402	5,402	5,402	5,402	5,402	5,402

Notes: The table reports the regression results of eq. (14) using the county seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

Table C-6: Local Geography Regressions (Prefecture Seats, Large Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.08	0.11	0.09	0.33***	0.29***	0.51***	0.43***	0.64***	0.30**	0.36***	0.34***
	(0.05)	(0.07)	(0.08)	(0.11)	(0.11)	(0.14)	(0.15)	(0.15)	(0.13)	(0.13)	(0.13)
Dist. Equator	−0.08	−0.14	−0.08	−0.53***	−0.67***	−0.93***	−0.78***	−1.16***	−0.54**	−0.62***	−0.56**
	(0.08)	(0.11)	(0.13)	(0.18)	(0.20)	(0.23)	(0.25)	(0.26)	(0.22)	(0.21)	(0.22)
Dist. Coast	0.02	0.03	0.09	0.42***	0.58***	0.33**	0.36**	0.51***	0.18	0.14	−0.02
	(0.05)	(0.07)	(0.09)	(0.12)	(0.13)	(0.13)	(0.15)	(0.15)	(0.12)	(0.12)	(0.13)
Dist. River	0.00	0.05	−0.07	−0.49*	−0.35	−0.58*	−0.34	−0.57*	−0.13	−0.38	−0.17
	(0.12)	(0.17)	(0.19)	(0.27)	(0.25)	(0.31)	(0.34)	(0.33)	(0.29)	(0.29)	(0.31)
Ruggedness	−0.21**	−0.13	−0.38***	−0.87***	−0.50*	−0.93***	−0.53*	−0.65**	−1.06***	−1.10***	−0.87***
	(0.08)	(0.10)	(0.15)	(0.20)	(0.26)	(0.28)	(0.30)	(0.31)	(0.27)	(0.27)	(0.30)
Temperature	0.03	0.08	0.27**	0.10	−0.10	−0.48**	−0.20	−0.43*	−0.05	−0.05	−0.02
	(0.07)	(0.10)	(0.12)	(0.16)	(0.17)	(0.20)	(0.22)	(0.22)	(0.18)	(0.18)	(0.19)
Temperature^2	−0.14	−0.23	−0.62	−0.50	−0.04	1.75**	0.24	−0.76	−0.78	−0.60	−0.59
	(0.26)	(0.35)	(0.39)	(0.59)	(0.62)	(0.84)	(0.88)	(0.88)	(0.75)	(0.75)	(0.73)
Precipitation	−0.21*	−0.40**	−0.47**	−0.91***	−0.41	−0.49	−0.43	−0.66*	−0.12	−0.35	−0.36
	(0.11)	(0.16)	(0.20)	(0.25)	(0.29)	(0.32)	(0.34)	(0.34)	(0.28)	(0.28)	(0.29)
Precipitation^2	0.37	0.72**	0.87**	1.76***	0.69	0.55	0.48	1.02	0.01	0.46	0.43
	(0.23)	(0.33)	(0.41)	(0.52)	(0.59)	(0.66)	(0.70)	(0.69)	(0.58)	(0.58)	(0.60)
Elevation	−0.30	−0.36	−0.07	−1.74**	−3.56***	−3.40***	−3.22***	−5.12***	−0.58	−0.85	−1.59
	(0.32)	(0.45)	(0.56)	(0.79)	(0.85)	(1.04)	(1.15)	(1.19)	(1.02)	(1.03)	(1.05)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.00	0.02	0.03	0.03	0.06	0.06	0.07	0.07	0.03	0.03	0.03
Num. obs.	5,402	5,402	5,402	5,402	5,402	5,402	5,402	5,402	5,402	5,402	5,402

Notes: The table reports the regression results of eq. (14) using the prefecture seats. Heteroskedasticity-robust standard errors are in parentheses ($^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

Table C-7: Local Geography Regressions (County Seats, Very Large Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.10	0.66***	0.52**	0.76***	1.02***	1.36***	0.36	0.48*	0.77***	0.72**	0.79***
	(0.21)	(0.25)	(0.25)	(0.25)	(0.25)	(0.29)	(0.28)	(0.28)	(0.28)	(0.30)	(0.30)
Dist. Equator	0.18	−0.74*	−0.53	−1.00**	−1.81***	−2.43***	−0.85*	−1.09**	−1.45***	−1.51***	−1.39***
	(0.35)	(0.43)	(0.43)	(0.43)	(0.43)	(0.50)	(0.48)	(0.49)	(0.48)	(0.52)	(0.52)
Dist. Coast	−0.30	−0.36	−0.01	0.68***	1.40***	1.29***	1.10***	1.13***	0.59**	0.95***	0.54*
	(0.23)	(0.25)	(0.25)	(0.25)	(0.25)	(0.26)	(0.26)	(0.26)	(0.26)	(0.27)	(0.28)
Dist. River	0.16	−0.73	−1.10*	−1.98***	−2.80***	−2.02***	−0.94	−0.91	−0.92	−1.23*	−0.73
	(0.47)	(0.55)	(0.57)	(0.60)	(0.57)	(0.65)	(0.65)	(0.64)	(0.63)	(0.65)	(0.66)
Ruggedness	−1.01**	−1.03	−1.12*	−2.12***	−0.44	−1.84***	−1.22*	−1.15*	−1.94***	−1.90***	−2.57***
	(0.43)	(0.63)	(0.64)	(0.57)	(0.64)	(0.68)	(0.66)	(0.66)	(0.68)	(0.69)	(0.69)
Temperature	1.93***	1.70***	1.84***	1.43***	0.28	−0.71	0.50	0.56	0.54	0.69	0.95*
	(0.35)	(0.42)	(0.42)	(0.41)	(0.42)	(0.46)	(0.46)	(0.47)	(0.46)	(0.48)	(0.48)
Temperature^2	−2.18*	−1.77	−2.83**	−4.36***	2.77*	4.60***	4.45**	2.31	−0.12	−0.04	−1.21
	(1.12)	(1.43)	(1.43)	(1.50)	(1.54)	(1.74)	(1.77)	(1.82)	(1.80)	(1.88)	(1.91)
Precipitation	−1.56***	−2.86***	−1.82***	−0.64	−1.79***	−0.12	1.56**	1.43**	0.81	1.35*	1.11
	(0.49)	(0.58)	(0.59)	(0.60)	(0.61)	(0.67)	(0.67)	(0.67)	(0.68)	(0.69)	(0.70)
Precipitation^2	2.00*	4.22***	2.35*	0.15	2.07	−1.90	−5.18***	−5.01***	−3.92***	−5.27***	−4.58***
	(1.03)	(1.21)	(1.23)	(1.23)	(1.27)	(1.41)	(1.47)	(1.45)	(1.47)	(1.51)	(1.51)
Elevation	−1.62	−3.45*	−3.42*	−7.91***	−13.50***	−15.98***	−10.20***	−11.02***	−9.97***	−10.99***	−9.61***
	(1.61)	(2.01)	(2.02)	(2.01)	(2.02)	(2.30)	(2.22)	(2.26)	(2.25)	(2.43)	(2.42)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.28	0.29	0.27	0.23	0.33	0.32	0.33	0.32	0.28	0.30	0.31
Num. obs.	3,463	3,463	3,463	3,463	3,463	3,463	3,463	3,463	3,463	3,463	3,463

Notes: The table reports the regression results of eq. (14) using the county seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

Table C-8: Local Geography Regressions (Prefecture Seats, Very Large Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.14*	0.16*	0.07	0.51***	0.19	0.65***	0.57**	0.89***	0.27	0.30	0.28
	(0.08)	(0.10)	(0.11)	(0.16)	(0.16)	(0.22)	(0.23)	(0.24)	(0.20)	(0.20)	(0.20)
Dist. Equator	−0.13	−0.21	0.02	−0.69**	−0.55*	−1.19***	−1.01**	−1.62***	−0.33	−0.40	−0.33
	(0.12)	(0.16)	(0.18)	(0.28)	(0.29)	(0.38)	(0.40)	(0.41)	(0.34)	(0.33)	(0.33)
Dist. Coast	0.00	−0.01	0.02	0.31*	0.80***	0.47**	0.48**	0.75***	0.23	0.19	0.04
	(0.09)	(0.10)	(0.13)	(0.17)	(0.19)	(0.20)	(0.21)	(0.22)	(0.18)	(0.18)	(0.20)
Dist. River	−0.00	0.00	−0.11	−0.83**	−0.80**	−1.17***	−0.70	−0.96**	−0.32	−0.53	−0.12
	(0.18)	(0.23)	(0.26)	(0.39)	(0.37)	(0.45)	(0.48)	(0.46)	(0.41)	(0.44)	(0.46)
Ruggedness	−0.13	−0.27*	−0.35*	−1.11***	−0.48	−1.24***	−0.90*	−1.25**	−1.69***	−1.48***	−1.60***
	(0.12)	(0.15)	(0.19)	(0.32)	(0.40)	(0.46)	(0.50)	(0.50)	(0.44)	(0.42)	(0.47)
Temperature	0.05	0.21	0.58***	0.13	0.22	−0.65*	−0.24	−0.65*	0.31	0.32	0.41
	(0.11)	(0.15)	(0.17)	(0.25)	(0.27)	(0.35)	(0.36)	(0.36)	(0.31)	(0.30)	(0.31)
Temperature^2	0.03	−0.22	−0.66	−0.50	0.03	3.66***	1.34	−0.51	−0.45	0.32	0.08
	(0.37)	(0.47)	(0.55)	(0.78)	(0.89)	(1.27)	(1.22)	(1.19)	(1.15)	(1.16)	(1.12)
Precipitation	−0.43***	−0.78***	−0.69**	−1.37***	−0.32	−0.88*	−0.50	−0.69	0.10	−0.42	−0.50
	(0.15)	(0.21)	(0.28)	(0.38)	(0.40)	(0.48)	(0.51)	(0.50)	(0.43)	(0.43)	(0.45)
Precipitation^2	0.78**	1.35***	1.14**	2.42***	0.25	0.96	0.24	0.71	−0.91	0.04	0.38
	(0.31)	(0.41)	(0.58)	(0.78)	(0.83)	(0.99)	(1.06)	(1.05)	(0.90)	(0.88)	(0.92)
Elevation	−0.40	−0.23	0.85	−2.45*	−3.91***	−4.35**	−4.31**	−7.21***	1.38	1.06	0.50
	(0.53)	(0.72)	(0.80)	(1.36)	(1.36)	(1.75)	(1.78)	(1.86)	(1.63)	(1.59)	(1.59)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.00	0.03	0.04	0.06	0.08	0.09	0.10	0.10	0.04	0.06	0.06
Num. obs.	3,463	3,463	3,463	3,463	3,463	3,463	3,463	3,463	3,463	3,463	3,463

Notes: The table reports the regression results of eq. (14) using the prefecture seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

C.2 Estimated Direct Geography Effects and the Historical Context

In Section 4.1 we estimate to what extent geography directly determines city locations. Following eq. (14), we regress an urban indicator on local geographic conditions. The baseline results and a multitude of robustness checks suggest local geography to determine the location of county seats while the link to prefecture seat locations appears to be much weaker, if existent at all. The effect that geography has on city locations varies over time with coefficient estimates fluctuating in magnitude and some even changing their direction. In this section, we show that many of these changes are related to major historical developments that reflect the evolution of our model's parameters and the impact that institutions have on the causal framework between geography and city locations.

County seats becoming less likely to locate far from the equator coincides with the population's southward shift. Before the An Lushan Rebellion (755 - 763 CE), two thirds of the population lived in the north and one third in the south. By 1100 CE, the distribution was reversed with two thirds living in the south and one third in the north (von Glahn, 2016). By the end of the 13th century the the northern share had dropped to 15 percent (Lewis, 2009).[1] An important contributor to this trend were the disproportional productivity gains in rice-based agriculture along the Yangzi in the south compared to grain-based agriculture along the Yellow River in the north (von Glahn, 2016).[2]

The surge in the coastal distance coefficient estimate over the first millennium CE might be related to a westward extension linked to the role of the silk road. Trade with Central Asia emerged at a larger scale in the second century CE (von Glahn, 2016). A few centuries later, the Tang dynasty (618 - 907 CE) pursued a westward expansion primarily based on military and merchant camps (Mote, 1999).[3]

---

[1]Linked to substantially dissimilar geographic conditions, northern and southern agriculture were rather unalike. In the north, rainfall was unevenly distributed over the year and agriculture centered on e.g. wheat, millet, sorghum, and soybeans. In the south, water was more abundant throughout the year and common agricultural products were e.g. rice, silk, tea, and oils. The different rainfall patterns did not just affect crop choice, but made the north compared to the south more prone to famines caused by floods, droughts, locust invasions, and soil salinization (Lewis, 2009; von Glahn, 2016). These disparities are an example of how geography jointly determines the theoretical framework's A_i and T_i.

[2]See Section C.9 for a discussion on geographic shocks in the form of floods and changing climatic conditions.

[3]Maritime trade also played a role in imperial times but developed differently form silk road trade. In the fifth century, ocean traders began traveling through the street of Melaka (von Glahn, 2016; Hall, 2011). During the Tang dynasty (618 - 907 CE), ocean trade with India and Muslim countries surpassed silk road trade. Maritime trade had marked effects during the Southern Song dynasty (1127 - 1279 CE) with urbanization rates of 25 percent in a maritime trade center, compared three percent a hinterland region, and rapidly developing trade with Japan. Throughout the later centuries ocean trade

We can interpret a lot into the coefficient estimates and their evolution over time. More than two millennia of imperial history offer numerous examples of how institutional developments might have affected the causal framework between city locations and geography. Emperors that invested more in famine relief lowered the threat level T_i coming from geographic variables related to extreme weather events. More attention to flood control reduced the risk associated with living nearby a river. Nonetheless, it is important to keep the motivation of using machine learning methods in this application in mind, implying that there might not be a meaningful interpretation to individual econometric coefficient estimates. Geographic factors co-evolve across space and it may be a complex combination of them determining city locations. In light of that, the linearity and the ceteris paribus assumption appear inappropriate, impeding estimation and interpretation. This paper, therefore, focuses on the role of geography as a whole rather than variable-specific effects.

---

did, however, experience various setbacks that did not eliminate but did complicate maritime trade. The Yuan dynasty (1260 - 1368 CE) fostered trade across Central Asia, but imposed a government monopoly on maritime trade. Similarly, the Ming dynasty (1368 - 1644 CE) imposed a ban on maritime trade for private merchants which lasted until the late 16th century. Merchants circumvented a lasting ban on trade with Japan by trading indirectly, i.e. through Portuguese intermediaries (the Portuguese formally set up the trading base of Macau in 1557 CE), neutral ports, and clandestine traders (von Glahn, 2016). In line with their predecessors, the subsequent Qing dynasty (1644 - 1911 CE) also temporarily banned maritime trade, with large scale ocean trade arising after the ban was lifted (von Glahn, 2016; Zhao, 2013).

C.3 Robustness Check: Probit and Logit Regressions

In this section, we repeat the estimations measuring direct effects of geography on city locations with logit and probit regressions. Observations refer to medium size pixels for which the association between local geography and city locations is clearer than at the baseline resolution (see Section 3.3.1). Table C-3 and Table C-4 in Section C.1 display the respective results when using a linear probability model.

The OLS, probit, and logit coefficient estimates do not allow for comparisons in terms of magnitude but in terms of their sign and statistical significance. And in that regard, the following tables are aligned with our baseline results. We see more significant estimates in county seat than in prefecture seat regressions. Distance from the equator, ruggedness, and elevation have negative effects. Distance from the coast reflects the in Section C.2 discussed westward expansion of the Tang dynasty.

Table C-9: Local Geography (County Seats, Logit, Medium Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	1.36	1.17	−1.25	−0.03	4.31*	3.16*	−1.39	−1.71	−1.22	−1.15	−1.05
	(3.26)	(2.60)	(2.53)	(2.20)	(2.24)	(1.92)	(2.03)	(2.00)	(1.87)	(1.78)	(1.71)
Dist. Equator	−14.33***	−11.61***	−9.15**	−10.82***	−14.22***	−15.60***	−9.37***	−8.66**	−9.74***	−8.91***	−7.23**
	(5.48)	(4.38)	(4.23)	(3.61)	(3.81)	(3.28)	(3.45)	(3.40)	(3.17)	(2.99)	(2.93)
Dist. Coast	2.79	−3.34	1.07	5.12**	14.63***	11.52***	11.28***	8.71***	0.69	2.55	0.42
	(2.51)	(2.05)	(2.03)	(2.13)	(1.90)	(1.75)	(1.85)	(1.80)	(1.80)	(1.69)	(1.67)
Dist. River	13.70**	5.37	4.15	−4.41	−13.16***	1.36	6.91*	7.22*	−0.87	−3.18	−0.60
	(5.62)	(5.12)	(4.92)	(5.16)	(4.81)	(4.17)	(4.15)	(4.16)	(4.27)	(4.05)	(4.10)
Ruggedness	−16.36*	−14.84*	−20.67***	−29.41***	−0.53	−12.92**	−16.53***	−18.16***	−20.92***	−18.77***	−20.27***
	(9.19)	(8.20)	(8.00)	(6.63)	(6.88)	(5.64)	(6.06)	(5.94)	(5.74)	(5.15)	(4.91)
Temperature	65.97***	70.35***	70.92***	74.96***	42.86***	43.01***	55.16***	55.02***	60.84***	57.62***	46.92***
	(11.00)	(10.44)	(10.68)	(10.41)	(7.12)	(6.61)	(7.28)	(6.97)	(7.09)	(6.69)	(5.50)
Temperature^2	−266.92***	−237.78***	−248.93***	−278.14***	−129.36***	−137.77***	−175.67***	−181.42***	−209.56***	−200.36***	−165.66***
	(39.15)	(33.30)	(34.19)	(33.61)	(22.49)	(20.53)	(22.39)	(21.56)	(22.34)	(20.95)	(17.67)
Precipitation	−34.33***	−32.53***	−16.20**	−12.92***	−37.56***	−19.98***	−9.55**	−5.08	−4.60	−0.85	1.31
	(7.35)	(7.43)	(6.96)	(4.83)	(5.06)	(4.47)	(4.58)	(4.85)	(4.84)	(4.55)	(4.54)
Precipitation^2	58.32***	41.93	16.52	27.88***	67.48***	34.88***	16.19	4.16	0.59	−7.84	−11.49
	(17.34)	(26.00)	(22.37)	(10.06)	(10.04)	(9.37)	(10.25)	(12.12)	(11.77)	(11.12)	(11.41)
Elevation	−168.83***	−56.41**	−50.90**	−79.20***	−174.03***	−133.73***	−116.93***	−89.51***	−49.98***	−53.36***	−41.07***
	(33.50)	(24.70)	(24.39)	(21.48)	(22.14)	(18.92)	(20.43)	(19.18)	(16.90)	(15.89)	(15.20)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590

Notes: The table reports the regression results of eq. (14) using the county seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

Table C-10: Local Geography (Prefecture Seats, Logit, Medium Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	−7.03	6.55	−9.97	6.73	4.66	6.55*	4.74	8.87***	1.17	3.05	3.05
	(5.58)	(7.59)	(7.08)	(4.72)	(3.99)	(3.43)	(3.39)	(3.34)	(3.89)	(3.55)	(3.60)
Dist. Equator	−30.12	−22.91*	6.15	−22.31***	−20.82***	−24.31***	−19.53***	−26.88***	−19.84***	−22.15***	−16.80***
	(19.15)	(13.16)	(12.49)	(8.13)	(6.77)	(5.98)	(5.86)	(5.79)	(6.49)	(6.02)	(6.14)
Dist. Coast	5.85	3.12	6.59	12.98***	17.16***	8.98***	6.13**	9.75***	7.10**	7.57**	3.08
	(8.28)	(6.21)	(5.50)	(3.73)	(3.25)	(3.18)	(2.85)	(2.75)	(3.51)	(3.20)	(2.92)
Dist. River	4.39	2.73	−3.01	−6.14	−4.91	−9.92	−0.69	−3.55	6.40	3.77	3.80
	(20.64)	(14.07)	(12.39)	(10.94)	(8.71)	(7.65)	(6.91)	(7.05)	(8.61)	(8.37)	(8.00)
Ruggedness	3.76	−18.58	−27.82	−29.96**	−7.57	−25.35***	−20.22**	−14.49*	−36.95***	−32.27***	−25.00**
	(36.84)	(23.67)	(19.40)	(11.89)	(10.88)	(9.46)	(9.40)	(8.79)	(9.80)	(9.88)	(11.20)
Temperature	−15.99	36.57*	88.54***	11.85	26.19**	26.39***	25.32***	14.74*	34.60***	40.84***	17.40**
	(19.23)	(21.08)	(32.57)	(12.98)	(12.69)	(9.82)	(9.54)	(8.88)	(12.17)	(10.32)	(8.16)
Temperature²	−8.24	−171.40**	−287.38**	−71.36*	−121.79***	−102.89***	−108.89***	−103.90***	−138.30***	−154.70***	−84.27***
	(71.89)	(83.65)	(135.01)	(39.02)	(41.63)	(30.64)	(29.34)	(27.89)	(37.14)	(32.64)	(25.13)
Precipitation	−4.58	−24.67	13.84	−20.63	−21.30**	−21.47**	−15.63*	−20.49**	11.92	−8.80	−7.33
	(32.59)	(25.15)	(25.80)	(13.58)	(8.96)	(8.37)	(8.57)	(8.03)	(13.31)	(10.76)	(10.20)
Precipitation²	−123.08	−8.99	−132.95	19.51	36.19	35.71*	18.91	30.23	−54.98	6.75	−5.17
	(131.10)	(106.88)	(110.19)	(52.22)	(23.03)	(21.12)	(23.92)	(21.40)	(42.42)	(29.73)	(31.45)
Elevation	−133.20	−19.05	90.21	−61.37	−119.37***	−68.47**	−56.81*	−95.81***	42.41	30.21	−19.89
	(110.49)	(66.52)	(56.36)	(39.66)	(37.17)	(30.31)	(29.15)	(29.80)	(30.09)	(29.24)	(31.35)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590

Notes: The table reports the regression results of eq. (14) using the prefecture seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in $100°$C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

Table C-11: Local Geography (County Seats, Probit, Medium Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.33	1.39	−0.21	0.23	2.68**	2.04*	−0.43	−0.56	−0.22	−0.08	−0.13
	(1.57)	(1.44)	(1.39)	(1.18)	(1.21)	(1.07)	(1.15)	(1.12)	(1.04)	(1.01)	(0.95)
Dist. Equator	−6.16**	−7.22***	−5.33**	−5.77***	−8.36***	−9.14***	−5.59***	−5.44***	−6.05***	−5.77***	−4.72***
	(2.73)	(2.41)	(2.32)	(1.91)	(2.06)	(1.81)	(1.93)	(1.90)	(1.75)	(1.68)	(1.62)
Dist. Coast	0.69	−1.27	1.02	3.22***	8.23***	6.65***	6.59***	5.26***	0.75	1.90**	0.67
	(1.34)	(1.16)	(1.15)	(1.16)	(1.06)	(0.98)	(1.05)	(1.01)	(1.02)	(0.97)	(0.96)
Dist. River	7.34**	3.86	3.12	−1.16	−5.72**	0.85	4.01*	4.49*	−0.41	−1.58	0.06
	(2.96)	(2.81)	(2.69)	(2.69)	(2.67)	(2.33)	(2.32)	(2.36)	(2.35)	(2.26)	(2.36)
Ruggedness	−7.85*	−4.33	−7.23	−15.72***	0.60	−5.54*	−7.59**	−8.47**	−10.54***	−9.49***	−10.93***
	(4.53)	(4.73)	(4.62)	(3.38)	(3.86)	(3.24)	(3.55)	(3.45)	(3.32)	(2.98)	(2.67)
Temperature	33.10***	30.65***	31.13***	34.09***	20.13***	20.89***	26.34***	26.87***	29.98***	28.12***	22.30***
	(5.22)	(6.09)	(6.10)	(5.60)	(3.87)	(3.91)	(4.32)	(3.99)	(4.07)	(3.91)	(3.07)
Temperature^2	−128.41***	−107.23***	−112.68***	−128.81***	−62.76***	−70.30***	−86.05***	−91.59***	−106.91***	−102.34***	−82.38***
	(17.96)	(17.81)	(18.13)	(17.56)	(11.80)	(11.92)	(12.91)	(12.08)	(12.58)	(11.96)	(9.76)
Precipitation	−16.86***	−15.95***	−6.44	−6.62***	−19.06***	−9.30***	−3.98	−1.64	−1.97	0.14	1.47
	(4.01)	(4.39)	(4.10)	(2.54)	(2.79)	(2.50)	(2.59)	(2.85)	(2.70)	(2.57)	(2.64)
Precipitation^2	26.40**	13.19	−1.53	13.51**	33.04***	14.57***	4.76	−2.20	−2.13	−7.01	−9.75
	(11.04)	(16.78)	(14.62)	(5.26)	(5.58)	(5.33)	(6.04)	(7.76)	(6.82)	(6.51)	(7.05)
Elevation	−78.10***	−36.16***	−30.83**	−41.80***	−91.44***	−73.84***	−66.82***	−53.53***	−30.69***	−33.05***	−25.07***
	(17.05)	(13.53)	(13.23)	(11.15)	(12.12)	(10.64)	(11.33)	(10.58)	(9.35)	(8.95)	(8.18)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590

Notes: The table reports the regression results of eq. (14) using the county seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

Table C-12: Local Geography (Prefecture Seats, Probit, Medium Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	−0.54	2.74	−4.05	2.60	2.29	2.72*	2.05	4.04**	0.36	1.29	1.17
	(4.50)	(2.97)	(2.80)	(2.08)	(1.79)	(1.61)	(1.63)	(1.60)	(1.72)	(1.56)	(1.59)
Dist. Equator	−12.31	−10.33**	1.55	−9.83***	−10.37***	−11.08***	−9.35***	−12.95***	−9.19***	−10.42***	−7.74***
	(8.08)	(5.15)	(4.96)	(3.61)	(3.05)	(2.80)	(2.82)	(2.78)	(2.89)	(2.67)	(2.70)
Dist. Coast	2.25	1.30	2.53	5.98***	8.33***	3.91**	2.88**	4.97***	3.30**	3.57**	1.66
	(3.43)	(2.76)	(2.56)	(1.68)	(1.50)	(1.54)	(1.42)	(1.32)	(1.57)	(1.44)	(1.31)
Dist. River	2.04	1.17	−1.63	−2.61	−3.05	−4.28	−0.25	−1.74	3.73	2.58	2.03
	(8.85)	(6.14)	(5.60)	(5.03)	(4.06)	(3.53)	(3.34)	(3.42)	(3.95)	(3.90)	(3.71)
Ruggedness	5.92	−7.73	−11.43	−13.42***	−4.36	−10.77**	−9.04**	−6.77	−16.60***	−14.06***	−9.75**
	(15.80)	(10.03)	(7.82)	(4.96)	(4.67)	(4.51)	(4.58)	(4.16)	(4.32)	(4.42)	(4.98)
Temperature	−6.97	14.28*	33.77***	5.06	9.80*	12.09***	11.42**	6.89*	14.76***	17.98***	6.37*
	(8.40)	(7.88)	(11.94)	(5.50)	(5.44)	(4.47)	(4.56)	(4.18)	(5.36)	(4.44)	(3.50)
Temperature^2	1.88	−70.28**	−108.50**	−29.88*	−50.37***	−48.55***	−51.35***	−50.16***	−60.63***	−69.55***	−33.71***
	(29.70)	(29.75)	(47.81)	(16.20)	(17.72)	(14.25)	(13.90)	(13.07)	(16.59)	(14.55)	(10.88)
Precipitation	0.63	−11.59	5.48	−9.72	−9.51**	−8.88**	−6.21	−9.56**	5.87	−4.35	−2.39
	(13.73)	(10.68)	(10.58)	(5.97)	(3.95)	(4.08)	(4.49)	(3.89)	(5.82)	(4.93)	(4.65)
Precipitation^2	−68.60	0.97	−57.75	11.37	16.16*	13.58	4.38	13.62	−28.14	2.71	−6.95
	(55.62)	(45.89)	(44.76)	(22.09)	(9.59)	(10.94)	(14.01)	(10.55)	(18.68)	(14.09)	(15.10)
Elevation	−54.07	−14.32	35.10	−27.41	−57.81***	−32.27**	−28.89**	−47.87***	16.48	10.61	−13.42
	(45.16)	(25.71)	(22.69)	(17.16)	(16.04)	(14.01)	(14.11)	(14.31)	(13.69)	(13.16)	(13.97)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590

Notes: The table reports the regression results of eq. (14) using the prefecture seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table.

C.4 Robustness Check: Spatial Durbin Model

Our baseline methods in Section 4.1 estimate the effect of geography on city locations non-spatially. They do not account for the effect a city or the geographic conditions in surrounding pixels might have on a grid cell.

Our application does not provide a strong motivation for the use of spatial methods. (i) Given that our pixels are not that large, their geography is very similar to that of neighboring pixels. A spatially weighted environment, therefore, barely adds any information. (ii) In terms of the dependent variable, we observe mostly isolated urban pixels in a vast sea of rural pixels. Spatial spillovers in the urban indicator are much less visible than they are in continuous data such as population size. (iii) The three distance variables already account for important determinants along the spatial dimension. (iv) Our analysis of indirect geography effects on city locations in Section 4.2 focuses on the spatial dimension, but does so in a more structural manner shaped by the theoretical framework and the historical context. To illustrate our results' robustness, this section, nonetheless, tests the mechanisms with spatial econometrics.[4]

We estimate a Spatial Durbin Model, i.e. a model of the form

$$y = \rho W y + X\beta + W X\theta + \varepsilon \qquad \text{((C-1))}$$

where y is the urban indicator and X the geography vector. It is estimated with maximum likelihood and accounts for global spatial spillovers. Observations refer to medium size pixels for which the association between local geography and city locations is clearer than at the baseline resolution (see Section 3.3.1). Table C-3 and Table C-4 in Section C.1 display the respective results when using a linear probability model.

The estimates in Table C-13 and Table C-14 support the baseline results. We see more significant coefficient estimates in county seat than in prefecture seat regressions. And cities tend to locate in lower, less rugged terrain, and near rivers. Interesting are the lagged variables' reversed signs, e.g. in the case of elevation and ruggedness. When an administrative city, i.e. a county seat, is set up in region, it is placed in the locally optimal location, making the surrounding non-selected places appear comparatively worse.

---

[4]Spatial models in supervised machine learning are, unlike those in econometrics, still at an early stage. E.g. classifying spatial data is still predominantly done via non-spatial methods. Geographic random forests are a recent innovation that run a set of regional estimations instead of drawing random samples from the full geographic space (Georganos et al., 2019). They can identify regional heterogeneity, but are not spatial in the spatial econometrics sense - they do not address spatial spillovers. And so far, they target continuous outcomes are not suited to binary classification problems. Supervised machine learning algorithms that are able to match spatial econometrics are yet to be developed.

Table C-13: Local Geography Regressions (County Seats, Medium Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	1.23	1.87	2.49	1.60	2.34	0.76	3.34*	1.63	−0.24	−0.22	2.11
	(1.45)	(1.67)	(1.70)	(1.77)	(1.82)	(1.99)	(1.97)	(2.00)	(2.02)	(2.12)	(2.16)
Dist. Equator	−4.14	−6.93	−8.93	−5.87	−7.09	−3.67	−9.96	−4.35	0.20	0.05	−4.76
	(5.31)	(6.09)	(6.21)	(6.47)	(6.64)	(7.27)	(7.18)	(7.28)	(7.38)	(7.74)	(7.88)
Dist. Coast	3.49	4.80	8.01	8.76*	11.06**	8.37	7.30	8.49	11.04*	12.61**	8.36
	(4.35)	(4.98)	(5.09)	(5.30)	(5.44)	(5.96)	(5.88)	(5.97)	(6.05)	(6.34)	(6.46)
Dist. River	−8.24**	−6.43*	−7.63**	−6.95*	−11.64***	−3.23	−7.08	−8.81*	−4.13	−1.19	−6.27
	(3.32)	(3.81)	(3.89)	(4.05)	(4.16)	(4.55)	(4.49)	(4.56)	(4.62)	(4.84)	(4.93)
Ruggedness	−1.57***	−2.19***	−2.66***	−3.37***	−3.14***	−4.12***	−4.86***	−4.95***	−4.87***	−5.21***	−5.38***
	(0.44)	(0.51)	(0.52)	(0.54)	(0.55)	(0.60)	(0.60)	(0.61)	(0.61)	(0.64)	(0.65)
Temperature	−0.58	−0.45	−0.55	−0.45	−1.12	−1.43*	−1.91**	−2.20***	−1.57*	−2.11**	−1.87**
	(0.58)	(0.67)	(0.68)	(0.71)	(0.73)	(0.80)	(0.79)	(0.80)	(0.81)	(0.85)	(0.86)
Temperature^2	4.90*	5.40*	7.58**	8.57***	6.25*	9.28**	7.23**	10.54***	11.83***	12.28***	13.68***
	(2.73)	(3.13)	(3.19)	(3.32)	(3.41)	(3.74)	(3.69)	(3.74)	(3.79)	(3.98)	(4.05)
Precipitation	0.96	0.18	−1.06	−1.19	−0.15	0.68	−4.25**	−2.40	−0.78	−0.87	−0.04
	(1.53)	(1.76)	(1.80)	(1.87)	(1.92)	(2.10)	(2.08)	(2.11)	(2.13)	(2.24)	(2.28)
Precipitation^2	0.63	2.90	5.21	5.76	3.66	4.96	13.28***	10.75**	7.82*	8.53*	6.58
	(3.21)	(3.68)	(3.75)	(3.91)	(4.01)	(4.39)	(4.34)	(4.40)	(4.46)	(4.68)	(4.76)
Elevation	−7.33***	−9.91***	−10.75***	−11.39***	−13.13***	−15.42***	−16.25***	−16.89***	−14.09***	−17.07***	−18.08***
	(1.85)	(2.12)	(2.16)	(2.25)	(2.31)	(2.53)	(2.50)	(2.54)	(2.57)	(2.69)	(2.74)
lag. Dist. E.	3.85	6.45	8.46	5.19	6.23	2.58	9.52	3.88	−0.92	−0.75	4.18
	(5.34)	(6.12)	(6.24)	(6.50)	(6.68)	(7.31)	(7.22)	(7.32)	(7.42)	(7.78)	(7.92)
lag. Dist. C.	−3.45	−4.79	−7.78	−8.31	−10.35*	−7.72	−6.57	−7.76	−10.79*	−12.16*	−8.01
	(4.36)	(5.00)	(5.11)	(5.32)	(5.46)	(5.98)	(5.90)	(5.99)	(6.07)	(6.36)	(6.48)
lag. Dist. R.	8.69**	6.46*	7.59*	6.52	10.83**	2.84	7.27	9.21**	4.02	0.75	5.95
	(3.42)	(3.92)	(4.00)	(4.17)	(4.28)	(4.68)	(4.62)	(4.69)	(4.75)	(4.98)	(5.08)
lag. Rugg.	1.60***	2.45***	3.05***	3.04***	3.38***	3.82***	4.76***	4.89***	4.70***	5.21***	5.30***
	(0.61)	(0.70)	(0.71)	(0.74)	(0.76)	(0.83)	(0.82)	(0.84)	(0.85)	(0.89)	(0.90)
lag. Temp.	0.68	0.38	0.49	0.43	0.86	0.66	1.62*	2.00**	1.16	1.91*	1.83*
	(0.68)	(0.78)	(0.79)	(0.82)	(0.85)	(0.93)	(0.91)	(0.93)	(0.94)	(0.99)	(1.00)
lag. Temp.^{2}	−3.84	−2.75	−5.86*	−9.67***	−3.43	−4.95	−3.03	−7.64*	−9.25**	−9.96**	−11.56***
	(3.00)	(3.44)	(3.52)	(3.66)	(3.76)	(4.12)	(4.06)	(4.12)	(4.18)	(4.38)	(4.46)
lag. Prec.	−1.68	−1.06	0.84	1.30	−0.50	−0.52	5.40**	3.51	1.15	1.88	1.05
	(1.68)	(1.93)	(1.97)	(2.05)	(2.10)	(2.30)	(2.27)	(2.30)	(2.34)	(2.45)	(2.49)
lag. Prec.^{2}	0.47	−1.67	−5.29	−6.30	−2.66	−5.88	−16.38***	−13.83***	−9.82**	−12.12**	−10.06*
	(3.58)	(4.10)	(4.18)	(4.36)	(4.47)	(4.90)	(4.84)	(4.91)	(4.97)	(5.21)	(5.31)
lag. Elev.	6.06***	8.02***	8.31***	5.85**	6.85**	7.58**	10.24***	11.12***	8.96***	11.67***	13.15***
	(2.25)	(2.58)	(2.64)	(2.75)	(2.82)	(3.09)	(3.05)	(3.09)	(3.13)	(3.28)	(3.34)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590

Notes: The table reports the regression results of eq. ((C-1)) using the county seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in 100°C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table. Spatial weights based on queen contiguity.

Table C-14: Local Geography Regressions (Prefecture Seats, Medium Pixels)

	200 BCE	1 CE	200 CE	400 CE	600 CE	800 CE	1000 CE	1200 CE	1400 CE	1600 CE	1800 CE
Intercept	0.33	0.32	0.03	1.17	0.99	−0.32	0.53	0.33	0.74	0.36	0.33
	(0.46)	(0.62)	(0.68)	(0.95)	(1.05)	(1.21)	(1.27)	(1.28)	(1.05)	(1.08)	(1.13)
Dist. Equator	−0.89	−0.91	−0.29	−3.56	−3.01	1.77	−0.88	−0.03	−2.26	−1.04	−0.06
	(1.68)	(2.26)	(2.49)	(3.45)	(3.83)	(4.41)	(4.64)	(4.67)	(3.82)	(3.93)	(4.11)
Dist. Coast	−0.43	−1.23	0.39	2.62	3.05	4.55	5.47	3.24	0.31	0.10	2.30
	(1.38)	(1.85)	(2.04)	(2.83)	(3.14)	(3.62)	(3.80)	(3.82)	(3.13)	(3.22)	(3.37)
Dist. River	−0.84	−2.41*	−2.72*	−5.78***	−5.61**	−4.05	−3.96	−5.47*	−2.70	−4.01	−5.06**
	(1.05)	(1.41)	(1.56)	(2.16)	(2.40)	(2.76)	(2.90)	(2.92)	(2.39)	(2.46)	(2.57)
Ruggedness	−0.13	−0.39**	−0.25	−1.08***	−1.13***	−1.92***	−2.18***	−2.06***	−1.53***	−1.63***	−1.88***
	(0.14)	(0.19)	(0.21)	(0.29)	(0.32)	(0.37)	(0.39)	(0.39)	(0.32)	(0.33)	(0.34)
Temperature	−0.07	0.15	0.23	0.17	−0.45	−0.57	−0.72	−0.45	−0.10	−0.04	−0.36
	(0.18)	(0.25)	(0.27)	(0.38)	(0.42)	(0.48)	(0.51)	(0.51)	(0.42)	(0.43)	(0.45)
Temperature^2	0.54	−0.35	0.18	2.13	1.92	3.19	2.55	1.98	−0.48	0.38	1.84
	(0.86)	(1.16)	(1.28)	(1.77)	(1.97)	(2.27)	(2.38)	(2.40)	(1.96)	(2.02)	(2.11)
Precipitation	−0.70	0.24	1.17	−1.04	0.20	−0.16	−1.24	−1.00	−0.05	0.65	−1.04
	(0.49)	(0.65)	(0.72)	(1.00)	(1.11)	(1.28)	(1.34)	(1.35)	(1.10)	(1.14)	(1.19)
Precipitation^2	1.62	0.18	−1.39	2.79	1.06	2.56	4.26	3.72	1.71	0.79	2.95
	(1.01)	(1.36)	(1.51)	(2.08)	(2.31)	(2.67)	(2.80)	(2.82)	(2.31)	(2.38)	(2.48)
Elevation	−0.65	−1.26	−1.82**	−3.69***	−5.71***	−6.51***	−7.29***	−8.17***	−3.40**	−3.84***	−5.59***
	(0.58)	(0.79)	(0.87)	(1.20)	(1.33)	(1.54)	(1.61)	(1.62)	(1.33)	(1.37)	(1.43)
lag. Dist. E.	0.76	0.72	0.18	3.15	2.59	−2.26	0.55	−0.59	2.21	0.90	−0.06
	(1.69)	(2.27)	(2.51)	(3.47)	(3.85)	(4.44)	(4.66)	(4.69)	(3.83)	(3.95)	(4.13)
lag. Dist. C.	0.45	1.30	−0.26	−2.36	−2.53	−4.39	−5.31	−2.93	−0.19	−0.01	−2.31
	(1.38)	(1.86)	(2.05)	(2.84)	(3.15)	(3.63)	(3.81)	(3.84)	(3.14)	(3.23)	(3.38)
lag. Dist. R.	0.81	2.37	2.71*	5.74***	5.52**	3.68	3.98	5.38*	2.67	4.05	5.16*
	(1.08)	(1.45)	(1.61)	(2.22)	(2.47)	(2.84)	(2.99)	(3.01)	(2.46)	(2.53)	(2.65)
lag. Rugg.	0.09	0.29	0.02	0.81**	0.98**	1.64***	2.13***	2.09***	0.94**	1.15**	1.99***
	(0.19)	(0.26)	(0.29)	(0.40)	(0.44)	(0.51)	(0.53)	(0.54)	(0.44)	(0.45)	(0.47)
lag. Temp.	−0.04	−0.26	−0.23	−0.39	0.41	0.21	0.54	0.05	0.18	0.09	0.31
	(0.21)	(0.29)	(0.32)	(0.44)	(0.49)	(0.56)	(0.59)	(0.59)	(0.49)	(0.50)	(0.52)
lag. Temp.^{2}	−0.16	0.83	0.15	−2.33	−1.70	−1.61	−1.76	−1.63	0.92	0.31	−0.98
	(0.95)	(1.28)	(1.41)	(1.95)	(2.17)	(2.50)	(2.62)	(2.64)	(2.16)	(2.23)	(2.33)
lag. Prec.	0.59	−0.54	−1.36*	0.70	−0.34	−0.27	0.98	0.69	0.04	−0.96	0.85
	(0.53)	(0.71)	(0.79)	(1.09)	(1.21)	(1.40)	(1.47)	(1.48)	(1.21)	(1.24)	(1.30)
lag. Prec. ^{2}	−1.49	0.29	1.67	−2.25	−0.93	−1.93	−4.04	−3.33	−2.10	−0.62	−2.91
	(1.13)	(1.52)	(1.68)	(2.32)	(2.58)	(2.97)	(3.12)	(3.14)	(2.57)	(2.65)	(2.77)
lag. Elev.	0.09	0.60	1.86*	2.29	3.06*	5.03***	6.27***	5.58***	4.37***	4.32***	5.10***
	(0.71)	(0.96)	(1.06)	(1.46)	(1.62)	(1.87)	(1.97)	(1.98)	(1.62)	(1.67)	(1.74)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590	9,590

Notes: The table reports the regression results of eq. ((C-1)) using the prefecture seats. Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 10,000 km, Ruggedness in Ruggedness Index x 10,000,000, Temperature in $100°$C, Precipitation in 10 m, Elevation in 100 km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table. Spatial weights based on queen contiguity.

C.5 Supplementary Results on Random Forest's Correctly Classified Pixels
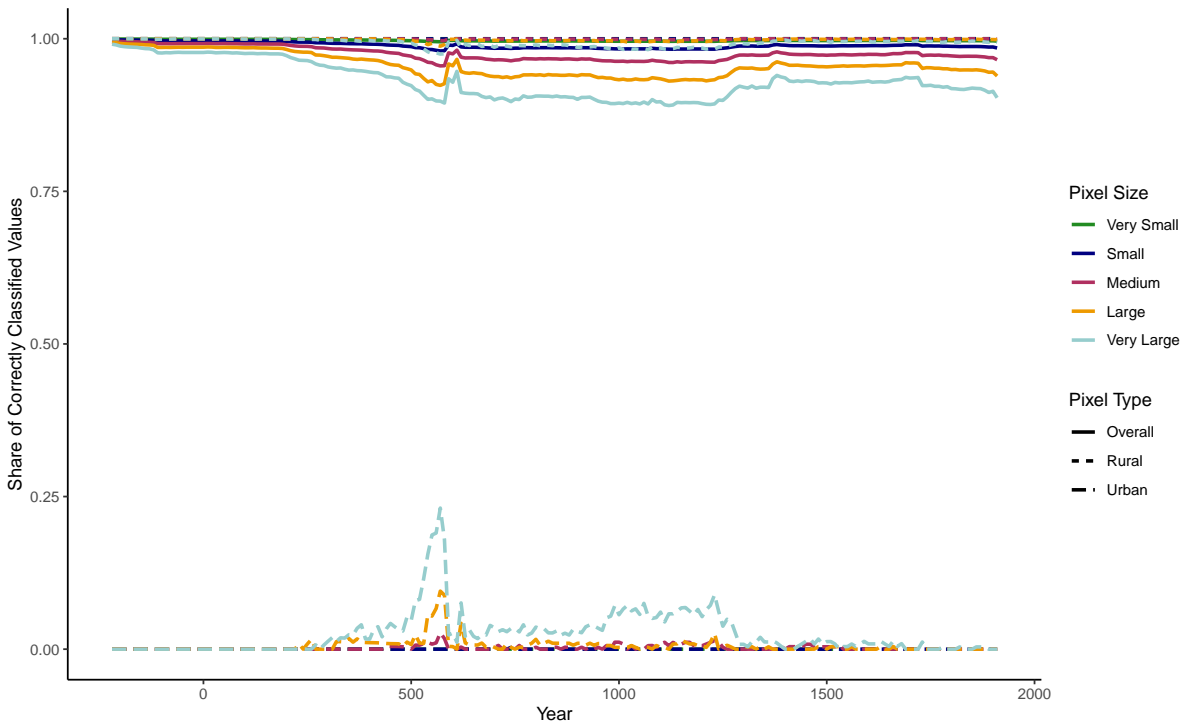
The classification random forests' objective in Section 3.3.1 is to distinguish urban from rural pixels. Most pixels are rural and the difficult part is to correctly spot the few urban grid cells using geographic characteristics. According to Figure 6, the algorithm manages to correctly classify more than 60 percent of county seat locations as urban at the coarsest resolution. In contrast, it marks almost no prefecture seat locations as urban.

Technically, the algorithm could reach a high share of correctly classified urban pixels by generally labeling more pixels urban. In a statistical sense, we would speak of a high sensitivity. This might come at the expense of a low specificity, which would be the case if a lower share of rural pixels were to be correctly classified as rural. In the following, we illustrate that the prediction quality on urban pixels does not undermine the prediction quality on rural pixels.

In Figure C-1 we compare the fraction of urban, rural, and overall pixels that were correctly classified. In the upper figures, the higher classification quality of county seat locations at low resolutions is accompanied by slightly worse predictions of rural pixels. And rural pixels are, apparently, slightly more difficult to identify in later years than in earlier ones. However, that slope is unrelated to the shape of the urban pixel's prediction quality over the years. More correctly classified urban pixels does not imply fewer correctly classified rural pixels. Moreover, we see that the changes over time are never as large as the differences in the prediction performance between county seats (upper figure) and prefecture seats (lower figure).

(a) County Seats

(b) Prefecture Seats

Figure C-1: Classification Random Forest Results

C.6 Supplementary Results on Random Forest's R^2

The paper visualizes the share of correctly classified county and prefecture seats of the random forest classifier using different resolutions in Section C.5. In Figure C-2, we repeat this analysis using the R^2 from random forest regression rather than the correctly classified share from random forest classification. The pattern is very similar. Figure C-3 also links the results to the dynasties.
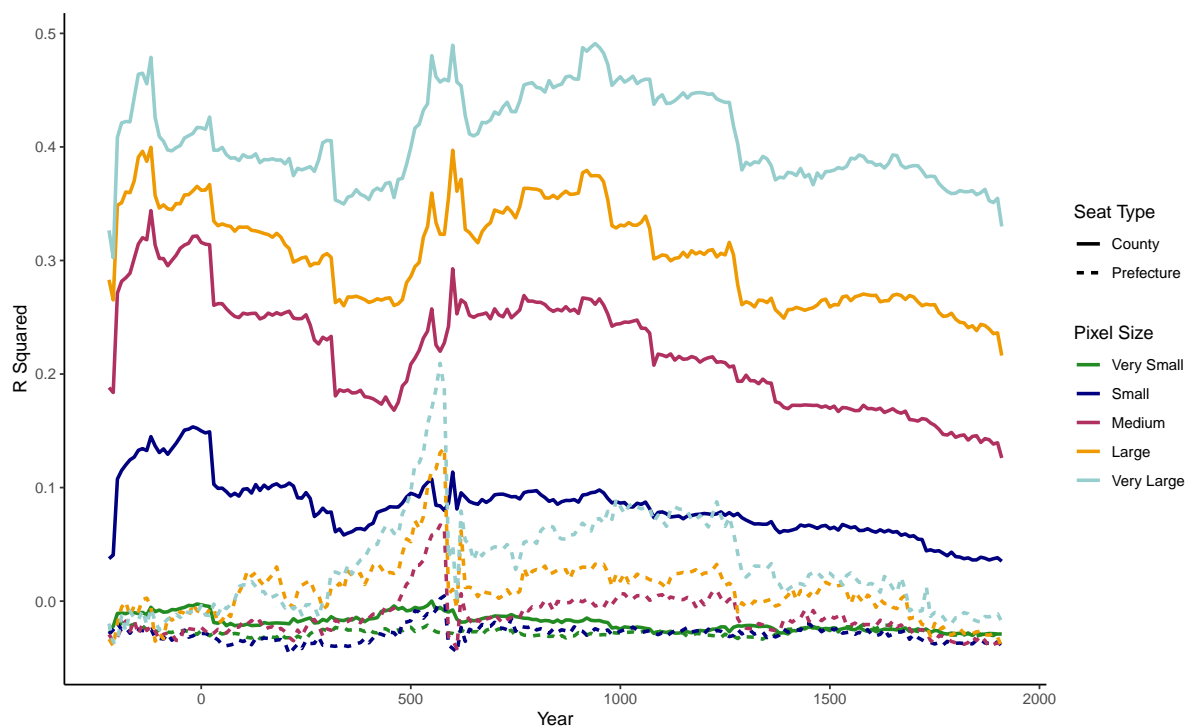
Figure C-2: R^2 in Regression Random Forests Linking Geography and Settlement Location
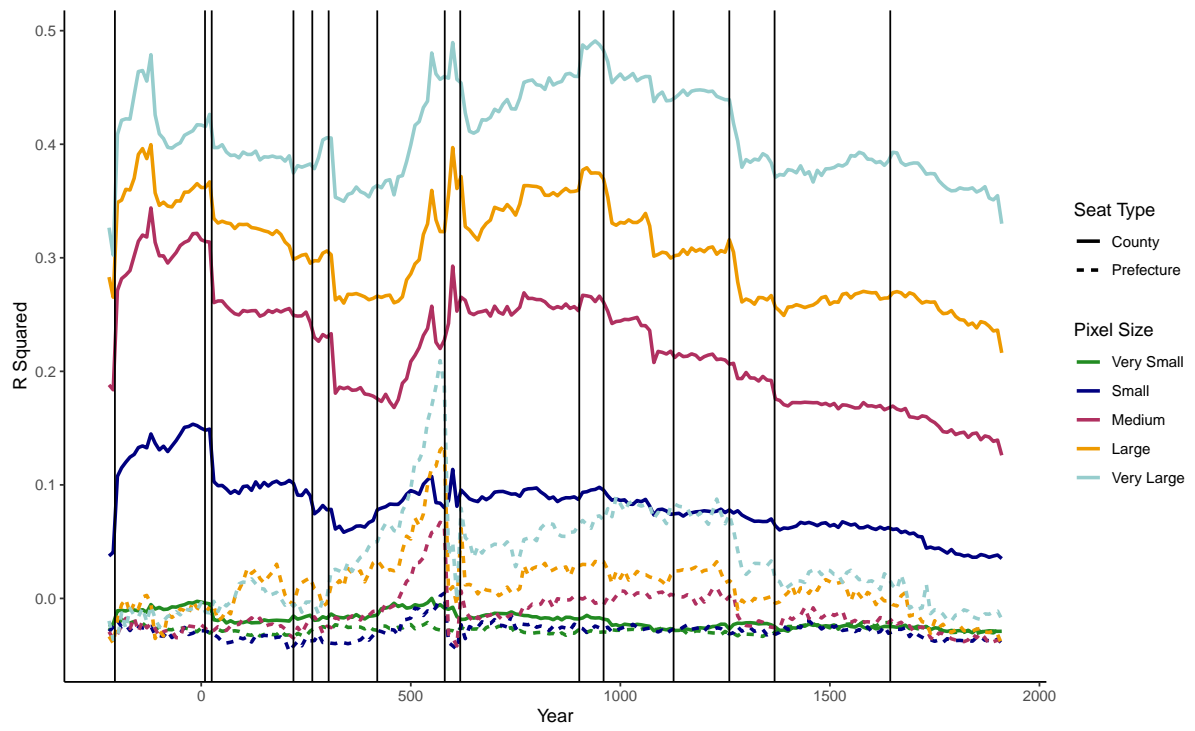
Figure C-3: R^2 in Regression Random Forests and the Rise of Dynasties

C.7 Random Forest: Variable Importance

Figure 4 and Figure 5 illustrate the variable importance of geographic factors in predicting city locations at the baseline resolution. The following Figure C-4 plots the corresponding variable rankings when using larger pixels. It confirms the baseline result of ranks being more stable over time in county seat estimations than in prefecture seat estimations.

(a) County Seats, Small Pixels

(b) Prefecture Seats, Small Pixels

(c) County Seats, Medium Pixels

(d) Prefecture Seats, Medium Pixels

(e) County Seats, Large Pixels

(f) Prefecture Seats, Large Pixels

(g) County Seats, Very Large Pixels

(h) Prefecture Seats, Very Large Pixels
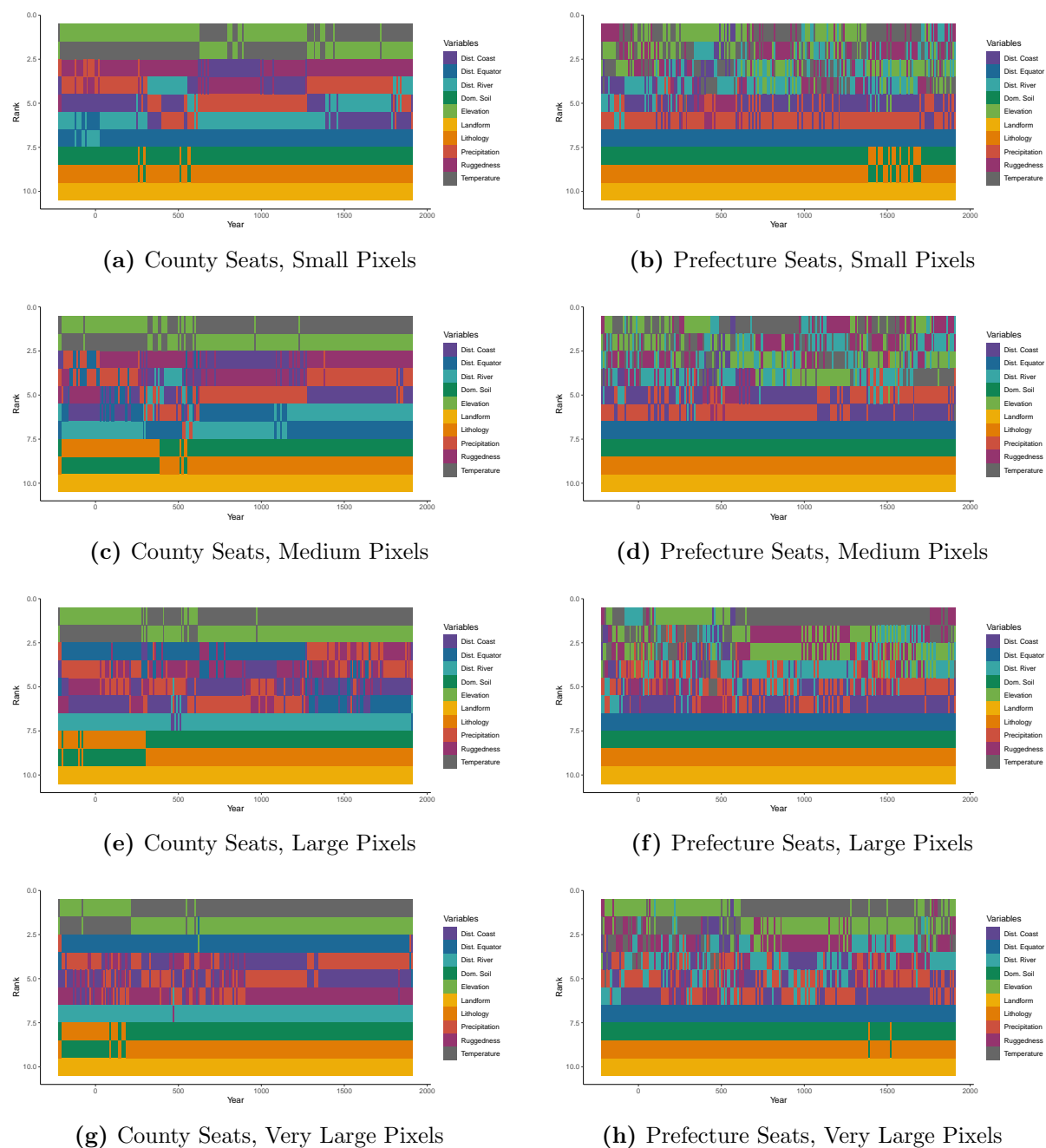
Figure C-4: Variable Importance at Different Pixel Sizes

C.8   Robustness Check: Alternative Empire Size

The baseline strategy relies on the empire's maximum extent, as observed in the CHGIS data. Holding imperial borders constant helps us to overcome missing prefecture borders before 1350 CE. The alternative choice, altering the shape every year and only accounting for pixels that certainly were part of the empire, excludes a lot of cities in earlier years, producing a much smaller, non-representative set of settlements. The maximum extent as a constant shape makes sure that relevant areas are included, at the cost of adding some remote regions that did not belong to the empire in all years. Given the distinctive geographic differences between the periphery and the heartland, adding these additional rural pixels should not pose a major problem to the random forest estimations. Nonetheless, we test alternative shapes in the following paragraphs to evaluate any potential biases.
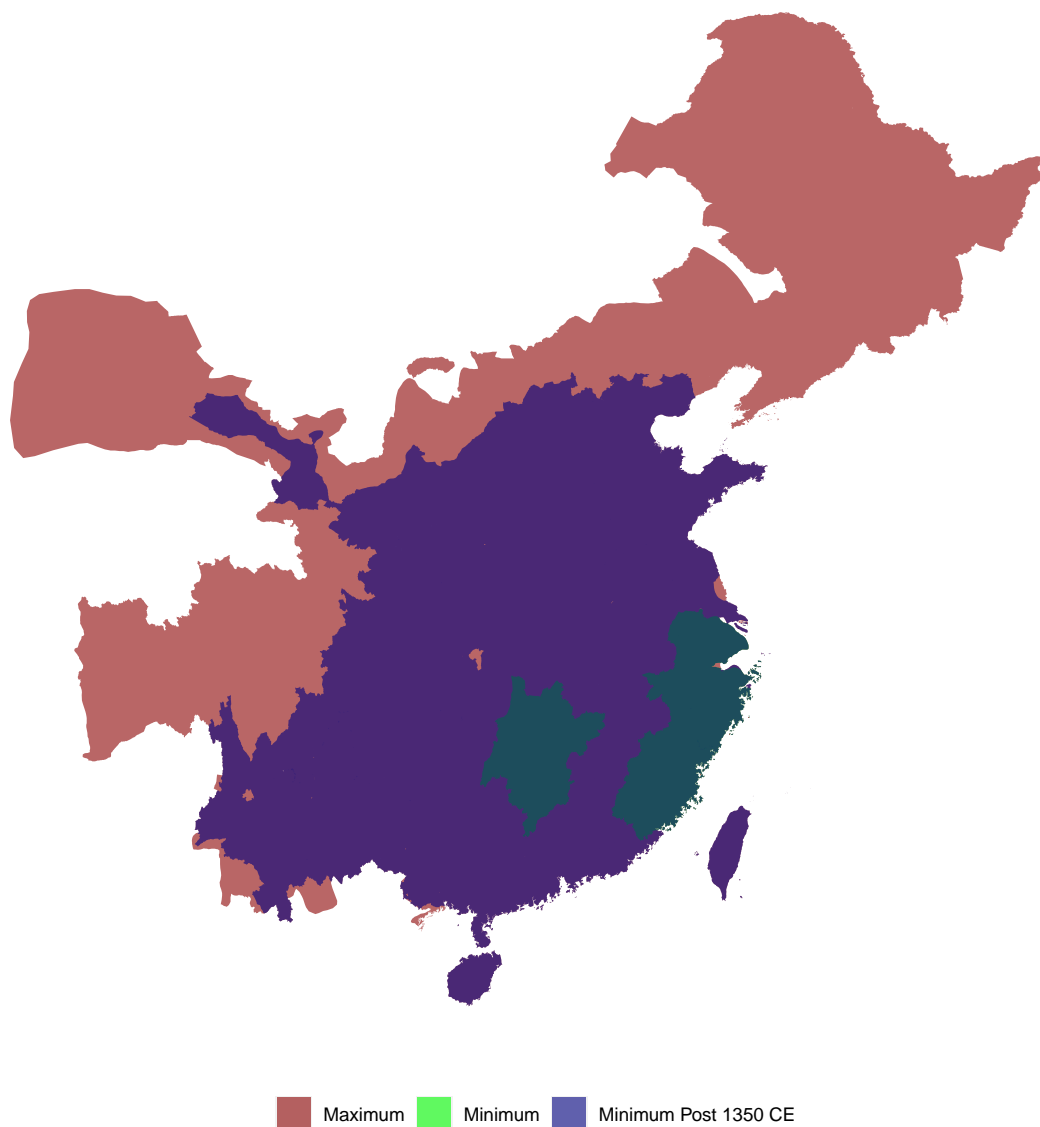
Figure C-5: Constant Extents of the Empire

In Figure C-5 the baseline shape, i.e. the areas that ever belonged to a prefecture observed in CHGIS are denoted in red, those that belonged to a prefecture in all years between 221 BCE and 1911 CE in green, and those that belonged to a prefecture in all years between 1350 CE and 1911 CE in blue. The green territories are primarily small because the data on prefecture borders is incomplete prior to 1350 CE.[5]

---

[5]County and prefecture seat locations are not plagued by these missing data issues.

Table C-15: Geography Summary Statistics, Minimum Empire Extent

	Mean	St. Dev.	Min	Max
Distance from Coast	258.573	210.546	0.208	797.336
Distance from River	191.860	112.223	1.285	475.299
Distance from Equator	3,081.696	208.617	2,668.395	3,572.082
Elevation	334.952	264.877	0.000	1,488.094
Ruggedness	200,837.699	134,018.285	0.000	620,263.812
Temperature	16.564	1.368	11.389	21.197
Precipitation	1,507.894	207.826	969.096	2,046.855

Notes: distances in km, temperature in °C, precipitation in mm per year, elevation in meters, ruggedness index in millimeters as defined by Nunn and Puga (2012). Values refer to the Chinese empire's minimum shape with 6,227 pixels 7.33 x 9.51 km in size. Landform, dominant soil type, and lithology are categorical variables and summarized in Online Appendix B. See Table B-1 for details on variable generation. Variables are differently scaled in subsequent chapters to facilitate readability.

Table C-16: Geography Summary Statistics, Minimum Post 1350 CE Empire Extent

	Mean	St. Dev.	Min	Max
Distance from Coast	519.713	361.411	-1.708	1,748.724
Distance from River	117.039	106.098	0.630	701.471
Distance from Equator	3,319.632	559.178	2,012.333	4,508.966
Elevation	759.549	782.351	-0.139	4,627.913
Ruggedness	200,359.188	170,948.605	0.000	1,147,039.125
Temperature	14.957	4.412	-10.125	25.267
Precipitation	1,085.534	462.630	92.340	3,948.264

Notes: distances in km, temperature in °C, precipitation in mm per year, elevation in meters, ruggedness index in millimeters as defined by Nunn and Puga (2012). Values refer to the Chinese empire's minimum post 1350 CE shape with 48,077 pixels 7.33 x 9.51 km in size. Landform, dominant soil type, and lithology are categorical variables and summarized in Online Appendix B. See Table B-1 for details on variable generation. Variables are differently scaled in subsequent chapters to facilitate readability.
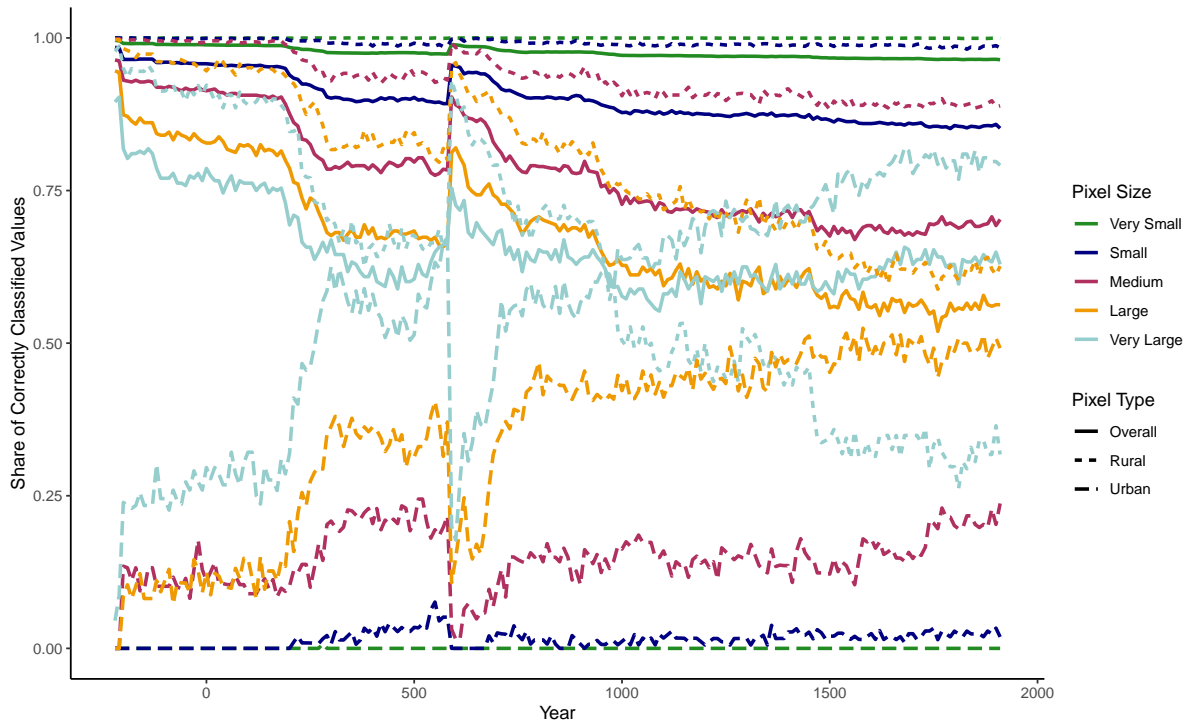
(a) County Seats

(b) Prefecture Seats

Figure C-6: Classification Random Forest Results (Minimum Empire Extent)
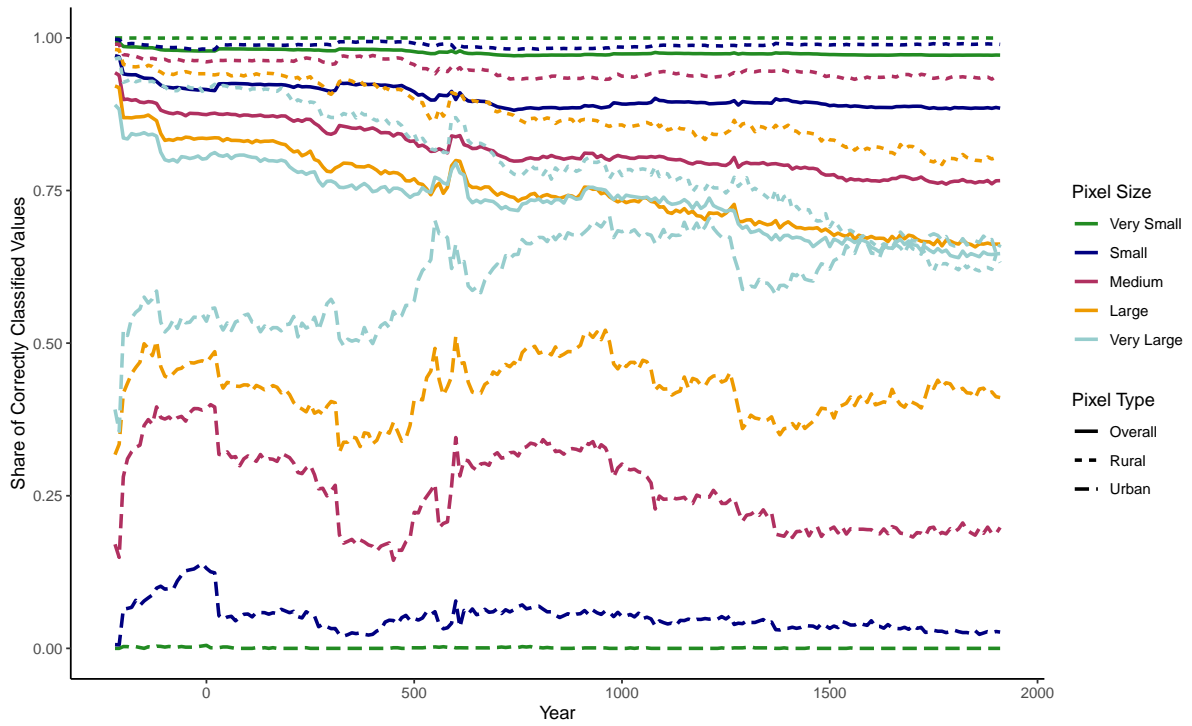
(a) County Seats
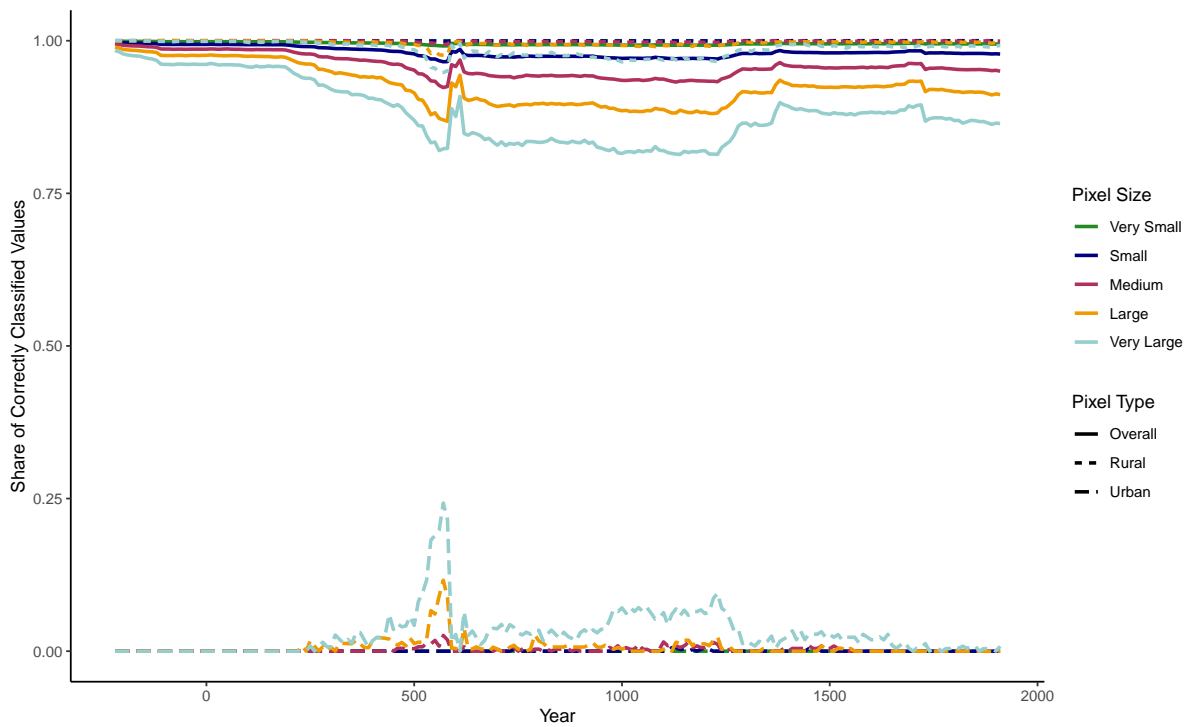
(b) Prefecture Seats

Figure C-7: Classification Random Forest Results (Minimum Post 1350 CE Empire Extent)

Figure C-6 and Figure C-7 replicate Figure C-1 with the two alternative empire shapes. The post 1350 CE minimum extent mostly excludes the little urbanized periphery and, thus, produces results that are very similar to the baseline outcomes -

underlining their robustness. The overall minimum shape, in contrast, omits most cities from the estimations and generates strongly volatile results. These patterns may point to both, the smaller number of observations used in the algorithm and the threat of omitting regions that are crucial to the spatial organization.

We do not split the empire into the physiographic macroregions defined by Skinner (1977a) and Skinner (1977b). As von Glahn (2016) criticizes, those macroregions only began to resemble actual conditions in imperial China after the crises of the 19th century. We, therefore, stick to our model of an empire-wide process that inter alia reflects the need to transport resources from the interior to the frontier and the disparities between the Yellow River and the Yangzi (Mostern, 2011).

C.9 Robustness Check: Paleo Data

Our baseline model uses modern geography data. In contrast to paleoclimatic data, it has the advantages of being available for many variables and of being accurately identified at high resolutions. Nonetheless, some geographic factors changed over time. In this section, we illustrate that using paleoclimatic data does not change the results. The adjustments that we account for are the changes in Yellow River's lower path and developments in temperature and precipitation.[6]

Figure C-8 shows how the Yellow River changed its course over time. The data comes from Chen et al. (2012) and Chen et al. (2015). The many path changes and floods were largely a result of deforestation causing soil erosion and consequently increased sediment uptake in the river (Elvin, 2004). Bursting dikes on 1593 occasions, many floods had a devastating impact on the population and the institutional setting. To name a few examples: a flood in 2 CE directly killed tens of thousands of people and triggered starvation and disease with an even higher death toll (Major and Cook, 2017); a few years later, in 11 CE, again thousands died from a flood that resulted in mass migration, famine, emerging bandit gangs killing county officials and forming armies, and turned out to be a key reason for the fall of the Xin dynasty (Major and Cook, 2017); a flood with a particularly high death tool was the one in 1117 CE that killed over a million people (Elvin, 2004; Tuotuo, 1346).

---

[6]The baseline data on temperature and precipitation refers to observations between 1900 and 1950 and should be less affected by man-made climate change than 21th century data. Nonetheless, it does not account for the historic fluctuations in climate.
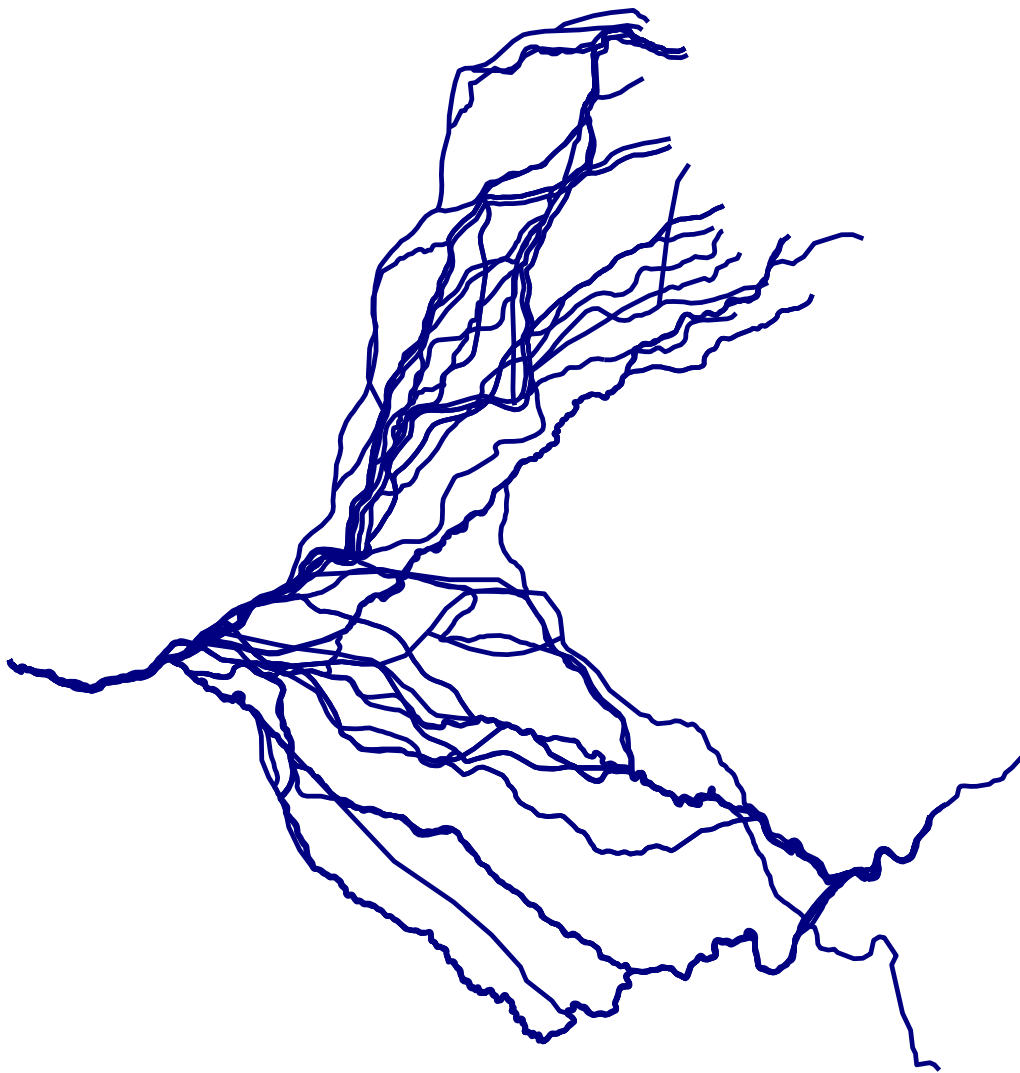
Figure C-8: All Observed Lower Yellow River Paths

We use multiproxy warm season temperature reconstructions by Zhang et al. (2018), which is published at a 5 x 5 degree resolution in decadal intervals, as information on historic temperatures. The precipitation data is reconstructed summer precipitation by Shi et al. (2018) published as 2 x 2 degree pixels at an annual level. With both, a 5 x 5 degree and a 2 x 2 degree resolution, the grid cells are so large that the whole empire just contains a few pixels. To increase the sample size to a level that can meaningfully be used in estimations, we disaggregate cells to the baseline resolution via bilinear interpolation. The severe measurement error this introduces and the fact that the temperature is measured in anomalies rather than levels are the reasons why we add the paleoclimatic data on temperature and precipitation as additional variables, without

replacing the modern counterparts.[7]
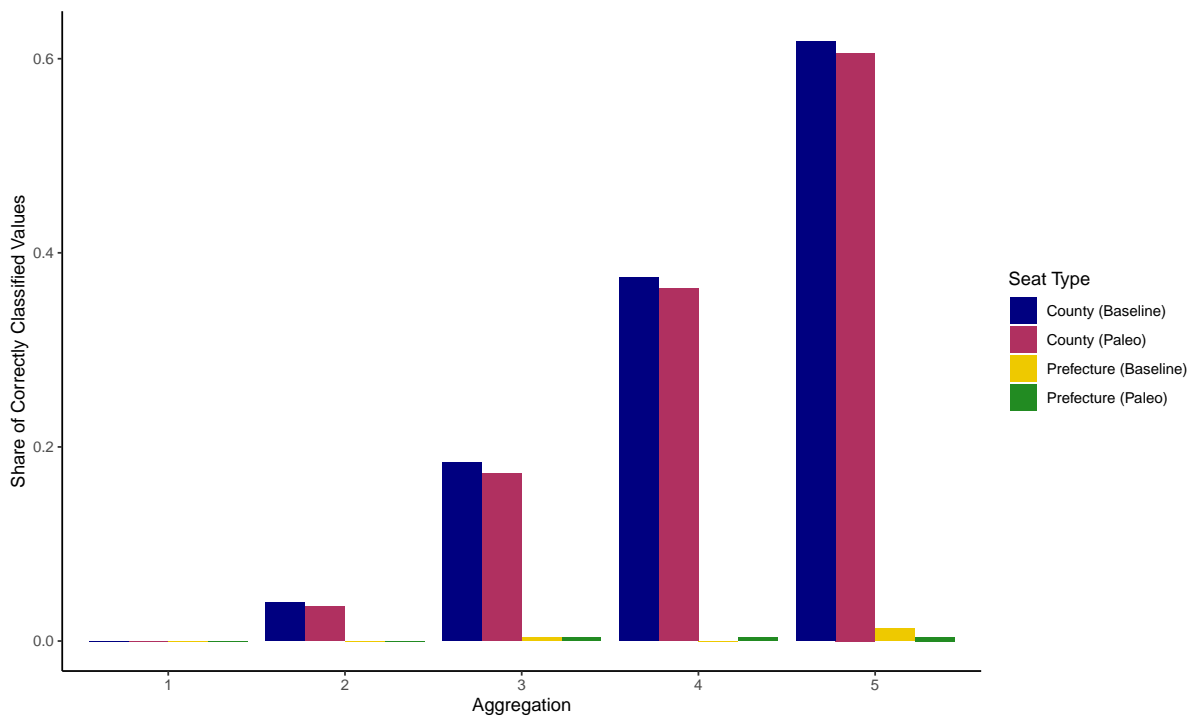
Figure C-9: Correctly Classified Urban Pixels in 1500 CE

Figure C-9 compares results with the baseline outcomes.[8] It turns out that accounting for geographic changes barely affects the estimates, underlining the robustness of our baseline results.

---

[7]Climatic shocks had marked effects on the empire, often setting off a series of events. Climatic disturbances and cold whether led to harvest failures between 1638 and 1642 CE and the Chinese population size shrank by 20 percent in response to the ensuing natural disasters, wars, and epidemics (von Glahn, 2016; Atwell, 1986, 1990; Marks, 2012). Many dynasties set up grain reserves to counter famines resulting from geographic shocks, such as floods and changes in climate, which were such a severe threat that they could lead to the downfall of a dynasty (Mostern, 2011; von Glahn, 2016). Even the Qing dynasty that intervened comparatively little in other regards was heavily invested in a system of grain reserves (von Glahn, 2016; Will, 1990; Will and Wong, 1991). Despite the effort that was put into this system, the Qing's granary reserves were depleted after the Tambora volcanic eruption induced a series of failed harvests at a global scale (von Glahn, 2016). Climatic changes were without doubt important. However, as long as there is only data on large macro regions, the incentives to relocate administrative cities between more fine-grain locations are not traceable.

[8]The results are derived via classification random forests.

C.10 Supplementary Results on Urban Persistence

Table C-17 displays the results of estimating eq. (15) via OLS. County seat regressions produce more statistically significant coefficient estimates and a higher R^2 than prefecture seat regressions do. As concluded from regression random forest outcomes depicted in Figure 8, local geography is a strong predictor of county seat but not prefecture seat persistence.

Table C-17: Urban Persistence

	Very Small		Small		Medium		Large		Very Large	
	C	P	C	P	C	P	C	P	C	P
Intercept	115.39***	95.94*	121.60**	119.62*	134.81**	193.21***	221.31***	205.24**	213.10***	162.69*
	(39.89)	(53.82)	(48.28)	(63.15)	(55.82)	(67.60)	(70.46)	(83.79)	(80.22)	(91.89)
Elevation	−13.81***	−1.36	−15.05***	−3.65	−19.68***	−5.56	−30.41***	−5.02	−34.76***	−4.75
	(3.61)	(4.70)	(4.34)	(5.17)	(5.04)	(5.69)	(6.31)	(7.00)	(7.22)	(8.07)
Ruggedness	−2.43**	−4.04***	−2.87**	−4.32***	−2.86*	−3.51**	−2.18	−5.23**	−0.04	−5.49**
	(1.13)	(1.48)	(1.44)	(1.52)	(1.74)	(1.68)	(2.06)	(2.12)	(2.57)	(2.33)
D. Coast	14.99***	−2.25	16.89***	−2.86	21.69***	4.25	22.50***	2.52	8.91	0.31
	(3.98)	(5.52)	(4.93)	(6.04)	(5.57)	(6.57)	(6.81)	(7.27)	(7.76)	(7.76)
D. River	22.10**	11.56	16.49	7.88	11.81	9.60	−29.98*	38.28*	−53.64***	15.40
	(10.92)	(15.99)	(13.05)	(18.18)	(14.95)	(18.87)	(17.99)	(21.45)	(19.90)	(22.81)
D. Equa.	−17.75**	−18.09*	−19.28**	−25.54**	−24.53**	−38.45***	−33.87***	−38.67***	−30.47**	−34.43**
	(6.99)	(9.68)	(8.39)	(11.22)	(9.69)	(12.06)	(12.10)	(14.51)	(13.64)	(15.76)
Temp.	37.69***	24.87	71.65***	27.49*	105.77***	16.36	96.31***	16.71	129.35***	34.71*
	(12.52)	(15.19)	(14.22)	(14.49)	(16.27)	(14.99)	(23.62)	(17.72)	(24.13)	(20.90)
Precip.	2.90	2.90	−7.40	14.91	−36.56	−21.50	−60.55**	6.98	−58.90*	5.15
	(15.90)	(21.45)	(20.02)	(20.40)	(23.26)	(25.01)	(28.10)	(26.24)	(30.71)	(32.03)
Temp.^{2}	−20.23***	−15.29***	−32.87***	−17.34***	−42.13***	−16.75***	−39.89***	−18.38***	−49.14***	−19.78***
	(4.32)	(5.67)	(4.81)	(5.38)	(5.61)	(5.83)	(7.47)	(6.39)	(7.89)	(7.26)
Precip.^{2}	0.16	−1.40	1.57	−6.77	13.35	9.56	15.76	−5.44	12.67	−6.01
	(6.00)	(8.31)	(7.54)	(6.57)	(8.54)	(9.54)	(10.21)	(7.58)	(10.76)	(10.72)
Soil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.07	0.05	0.10	0.03	0.17	0.07	0.20	0.04	0.29	0.05
Num. obs.	4172	1447	3435	1334	2705	1236	2061	1107	1616	976

Notes: Heteroskedasticity-robust standard errors are in parentheses (***$p < 0.01$, **$p < 0.05$, *$p < 0.1$). Distances in 1,000 km, Ruggedness in Ruggedness Index x 100,000, Temperature in 10°C, Precipitation in m, Elevation in km. Coefficient estimates on categorical soil variables - dominant soil type, landform, lithology - omitted from the table. C and P refer to county seat and prefecture seat estimations respectively. Very small, small, medium etc. denote the pixel size.

C.11 Robustness Check: Alternative Hiking Function

In Section 4.2 we compute the travel time between locations using Tobler's (1993) hiking function. To illustrate the robustness of those results we repeat the estimation with an alternative cost function by Márquez-Pérez et al. (2017) who modified Tobler's specification.

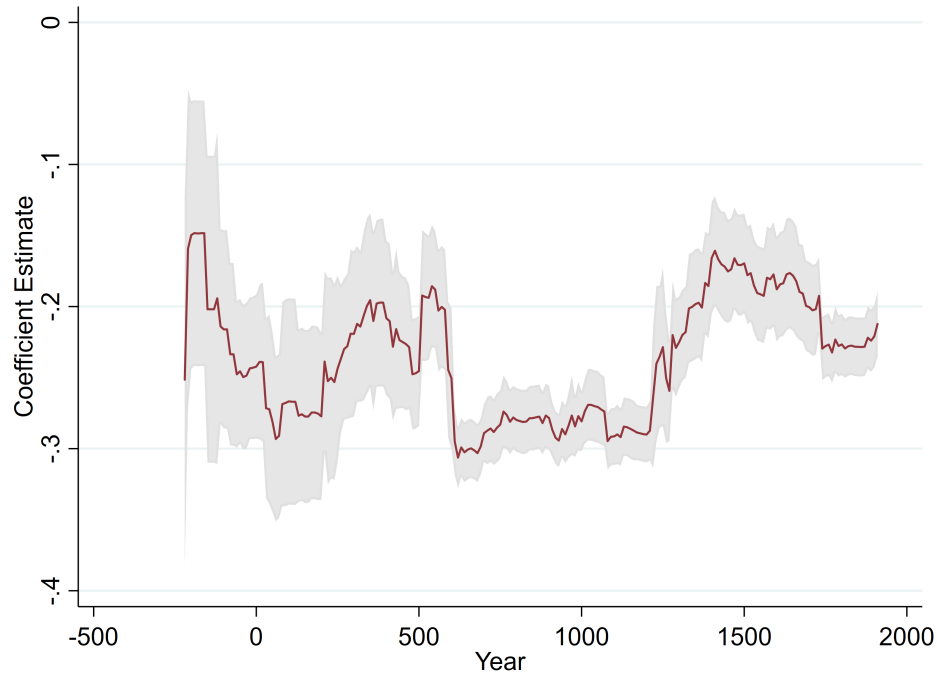**Figure C-10:** Inter-City Distance and Administrative Status (Márquez-Pérez et al., 2017)

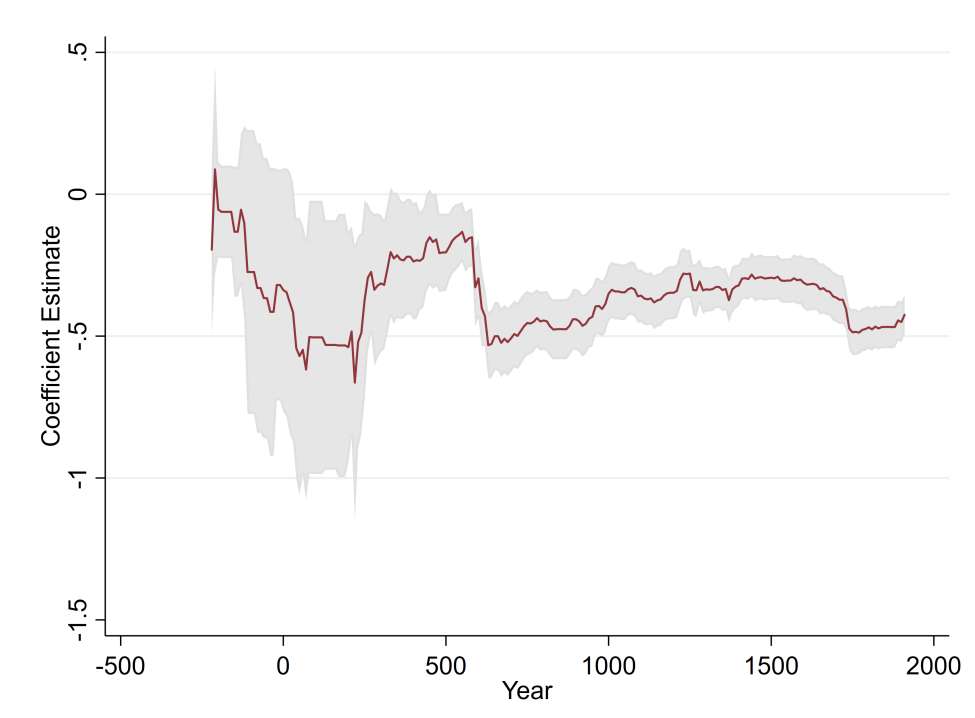The results are so similar that Figure 9 and Figure C-10 look quasi identical.

Figure C-11: Pixel Centrality and Administrative Status (Márquez-Pérez et al., 2017)

The same holds comparing Figure 10 and Figure C-11.

C.12 Robustness Check: Euclidean Distance

The baseline specifications in Section 4.2 use travel time as distance measure. In this section we repeat those estimations with Euclidean distance. Unlike travel time computed via hiking functions, the Euclidean distance measure does not take the topography into account. It is simply the length of a straight line between two locations.

Figure C-12: Inter-City Distance and Administrative Status (Euclidean Distance)

The outcomes depicted in Figure C-12 and Figure C-13 are similar to Figure 9 and Figure 10, confirming the results robustness to alternative distance measures.
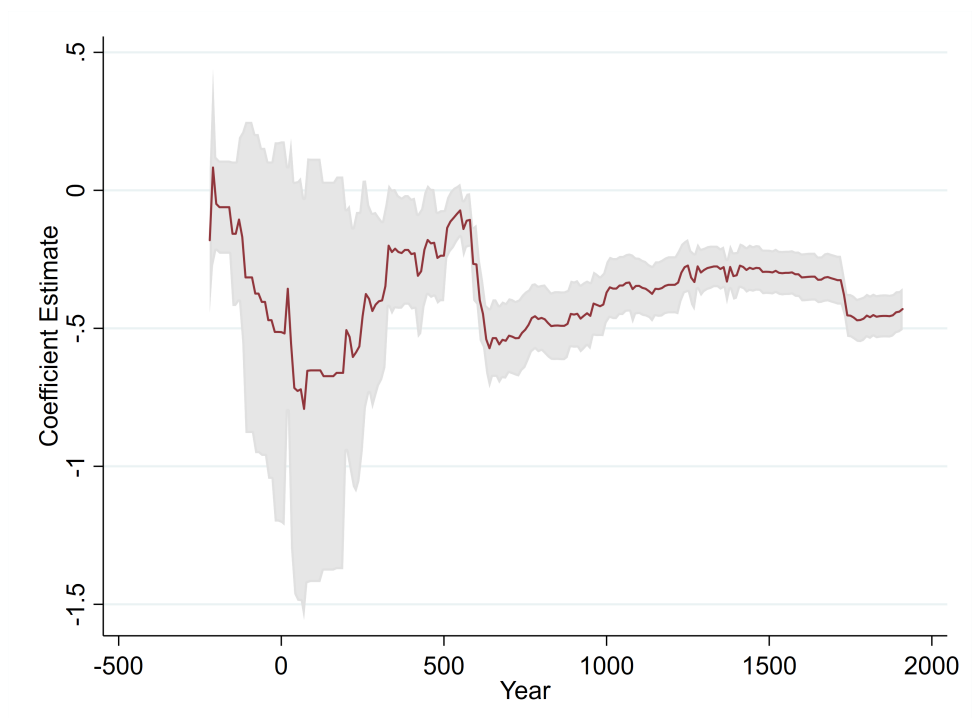
Figure C-13: Pixel Centrality and Administrative Status (Euclidean Distance)

C.13 Supplementary Results on the Prefecture Seat Location Algorithm

The prefecture seat location algorithm begins with randomly drawn prefecture seat locations, where the number of seats is set to the actual number of seats as observed in the data in that year. We then assign county seats to the nearest artificial prefecture seat. If all prefecture seats are assigned at least two county seats, the process enters into the next stage. If not, the prefecture seats with less than two county seats are randomly relocated until all prefecture seats are assigned at least two county seats. In the next stage, the algorithm shifts the prefecture seats to the pixel that minimizes the total distance to the assigned county seats.[9] County seats are then re-assigned to the nearest prefecture seat and prefecture seats moved to the pixel that minimizes the distance. The algorithm stops when at least 90 percent of prefecture seats remain in their location. We run the mechanism eight times per cross-section which should result in eight independent stable prefecture seat allocations.

---

[9]Distances are measured in travel time according to Tobler's (1993) hiking function which takes the topography into account. Using an adjusted version of Tobler's (1993) hiking function by Márquez-Pérez et al. (2017) as a robustnes checks confirms the results.
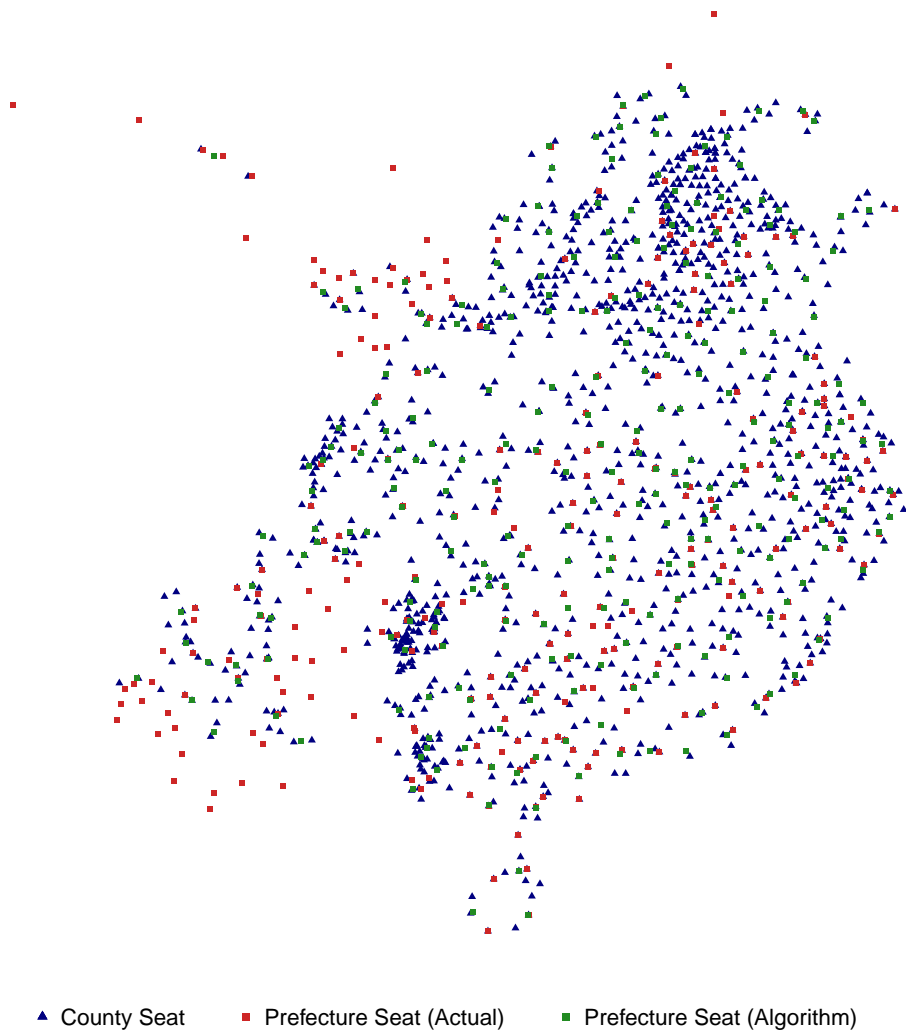
Figure C-14: Prefecture Seat Location Algorithm Outcome in 1350 CE

Figure C-14 plots one of the eight outcomes that we obtain for 1350 CE. The outcomes mostly differ from the actual spatial distribution of prefecture seats in outlying silk road regions. These location were of military importance, but hosted little population which is why the ratio between prefecture and county seats differs from the rest of the empire. Adding the silk road (Williams, 2020) to the algorithm, lifting the two county seat requirement around it, and attributing it a gravitational force, as attributed to the county seats, addresses those systematic discrepancies and improves the resemblance between the actual and the artificial prefecture seat locations. Overall, these results provide evidence in favor of our hypothesized institutional optimization.

C.14 Supplementary Results on China in 1820 CE

Here we present additional results on the extension using market towns in 1820 CE. Figure C-15 shows that the random forest regression R^2 for county seats, prefecture seats, and market towns exhibits the same pattern as the share of correctly classified pixels discussed in the paper. Market towns are determined to a much stronger extent by geographical features than county seats and prefecture seats.

The relative ranks of the individual geographical features are roughly the same across the three types of cities, as Figure C-16 shows. Temperature and elevation are important for all three cities. We can also see that the relative ranks of the geographical features do not change by much when varying the size of the pixel, although the paper shows that it does affect the explanatory power of the model.
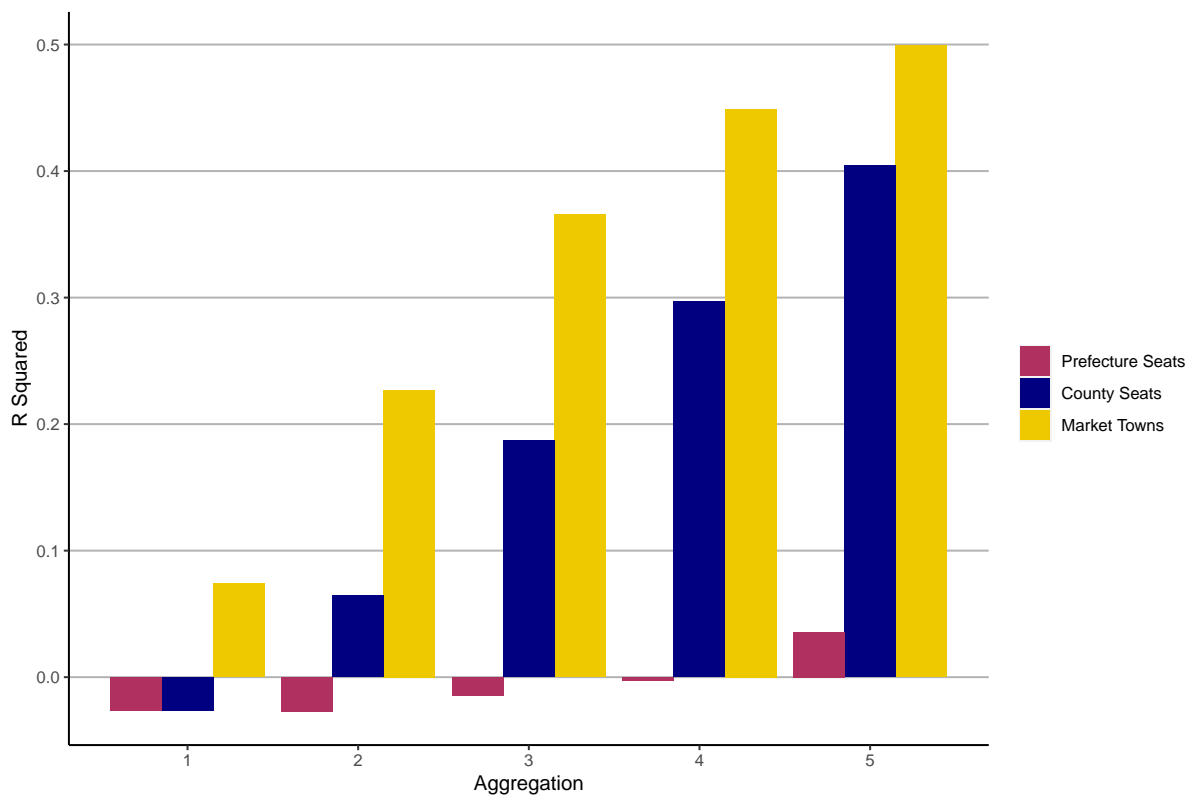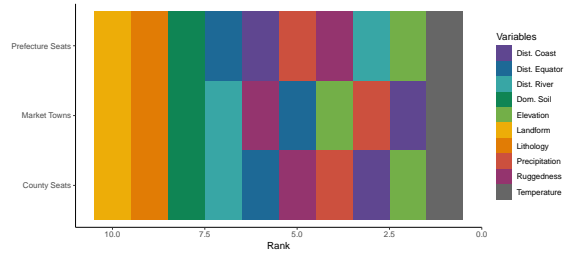

Figure C-15: R^2 for Chinese Cities in 1820 CE
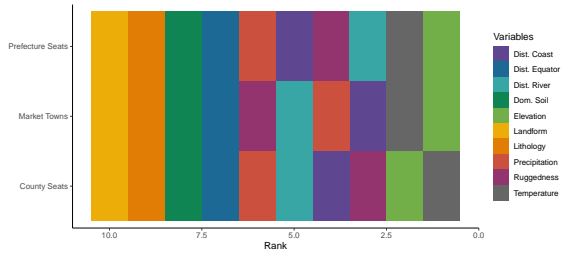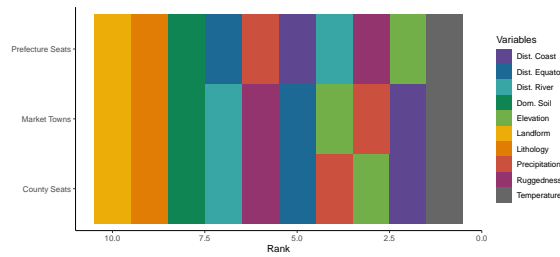
(a) Very Small Pixels
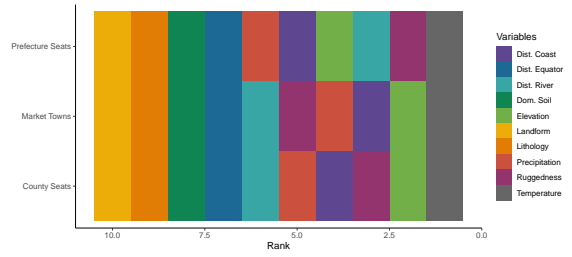
(b) Small Pixels
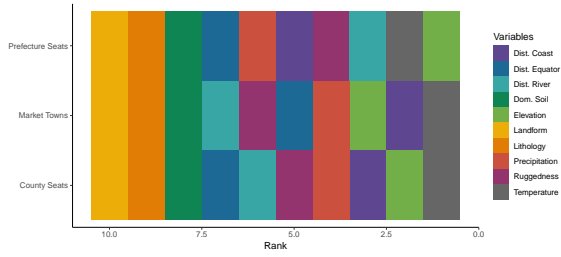
(c) Medium Pixels

(d) Large Pixels

(e) Very Large Pixels

Figure C-16: Variable Importance at Different Pixel Sizes in 1820 CE

C.15 Supplementary Results on Europe

Figure C-17 shows the R^2 of the random forest used in regressing the location of European or Chinese city on the geographical features. The pattern is very similar to that observed in the paper, where the fraction of correctly classified urban pixels is reported. We see that European cities fall in between Chinese county and prefecture seats when it comes to the explanatory power of geographical features for their location. At the coarsest resolution, geographical features can explain nearly 50 percent of the location of Chinese county seats, just above 32 percent of European cities, and 11 percent of Chinese prefecture seats.[10]
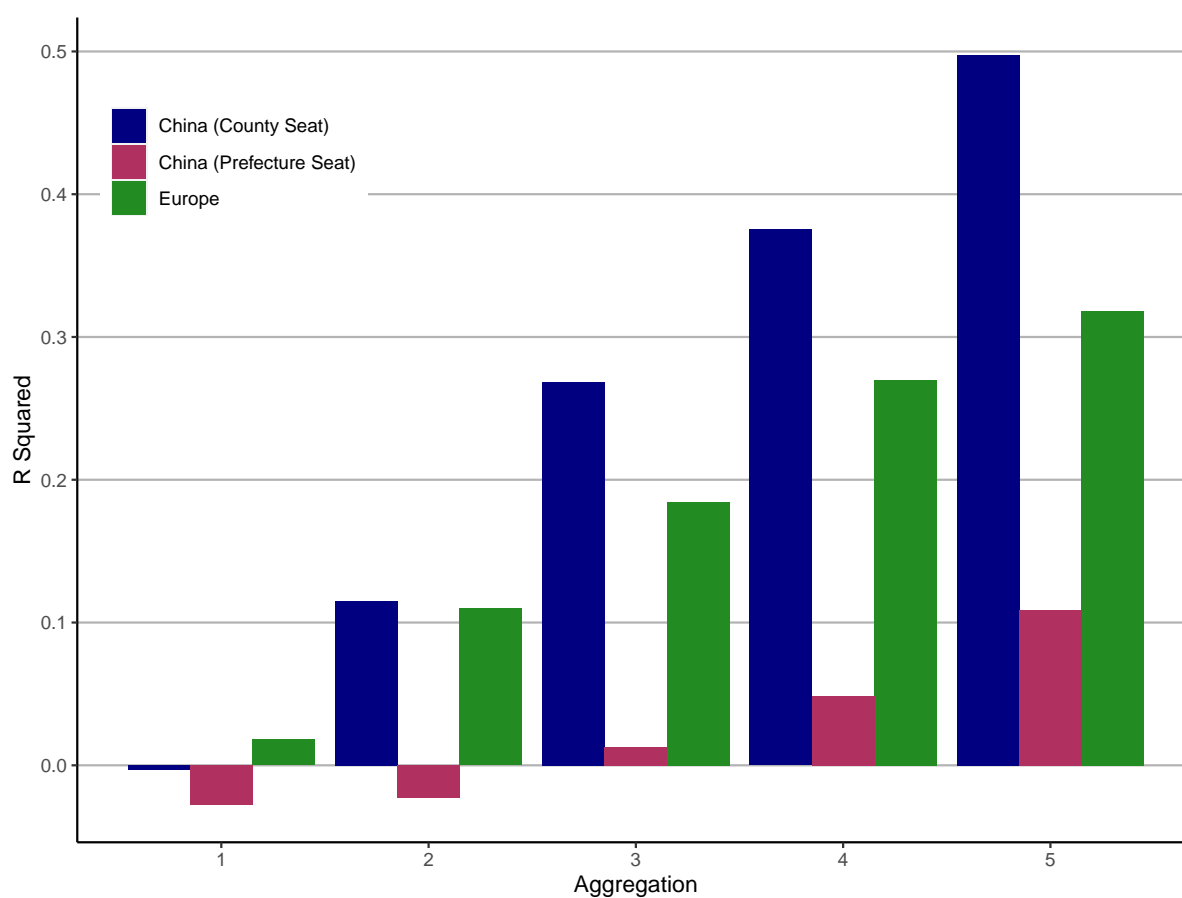

Figure C-17: R^2 in Regression Random Forests for China and Europe

---

[10]The estimations omit the three soil-related variables landform, dominant soil type, and lithology which are derived from an Asia-specific data set.

C.16 Supplementary Results on Modern China

Figure C-18 illustrates the location of the 29 current province and autonomous region capitals that fall into the territory of former Chinese empires observed in CHGIS. As decribed in the main text, 28 of these 29 cities were also a prefecture seat some time in imperial China.
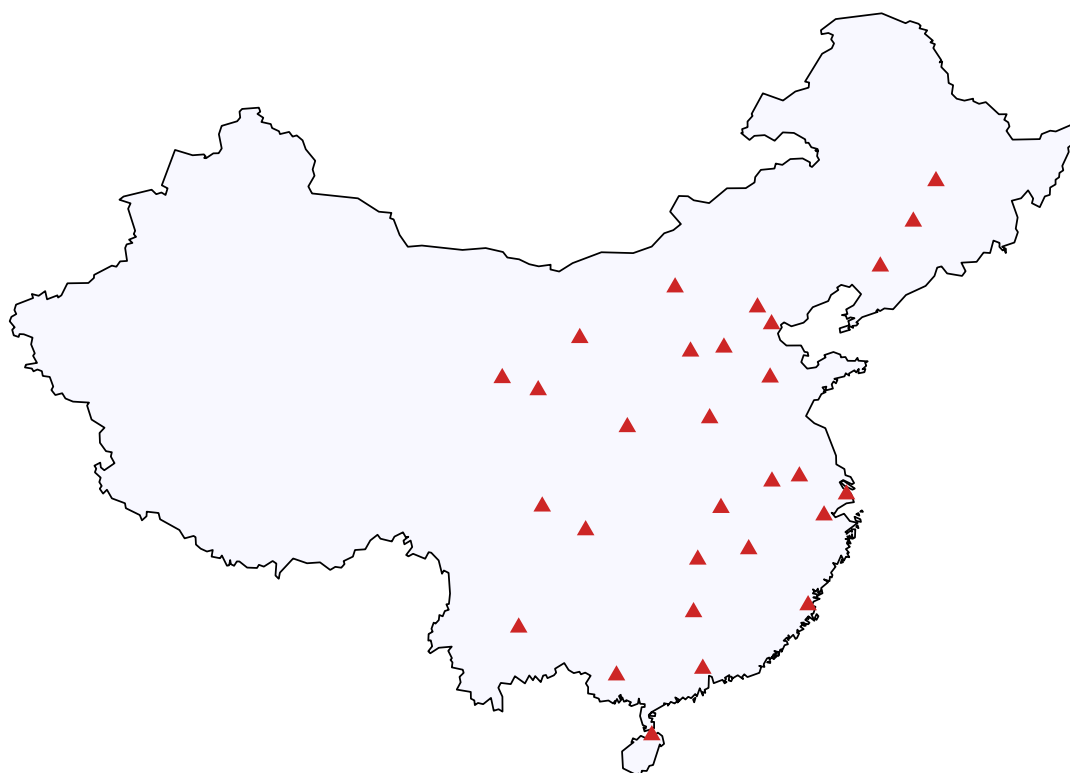

Figure C-18: Selected Modern Provincial Capitals

In Figure C-19 we show scatter plots on the relation between population and satellite data of nighttime light as described in the main paper. The figures illustrate the strong association between the two values, underlining our finding that historical prefecture and county seats, which are more populous, are also more economically active as measured by nighttime lights.
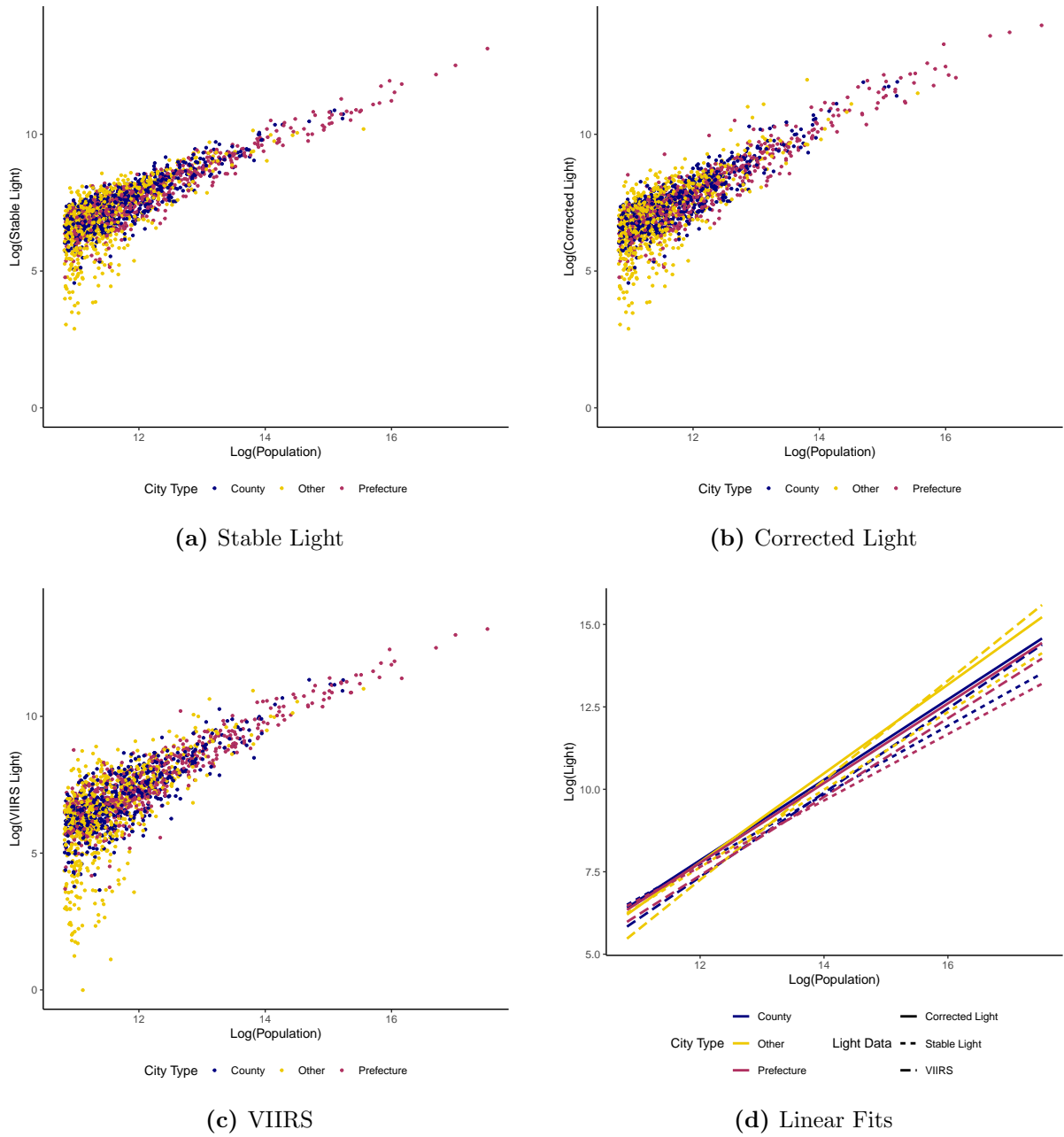
(a) Stable Light

(b) Corrected Light

(c) VIIRS

(d) Linear Fits

Figure C-19: Modern Chinese Cities' Population Size and Nighttime Light Emissions