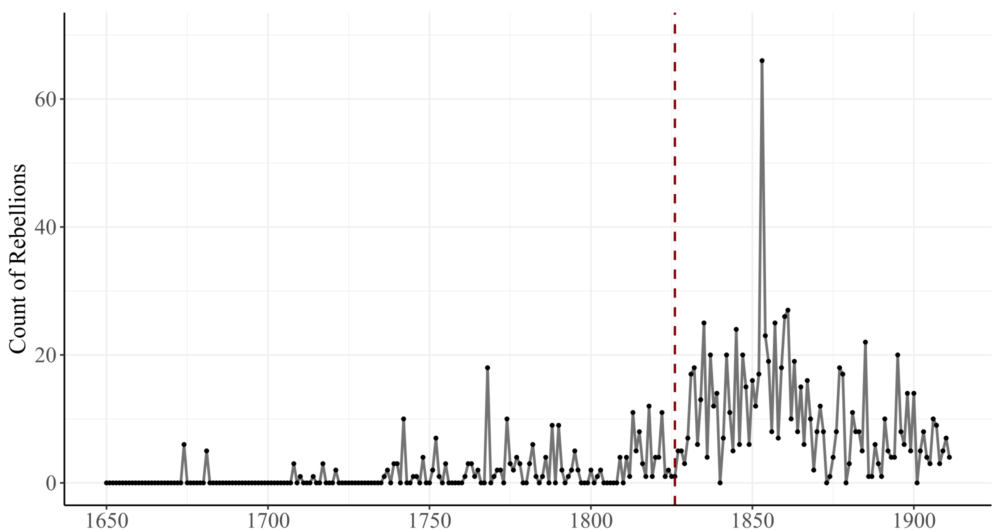
Online Appendix for
Rebel on the Canal: Disrupted Trade Access and Social Conflict in China, 1650–1911

Yiming Cao and Shuo Chen

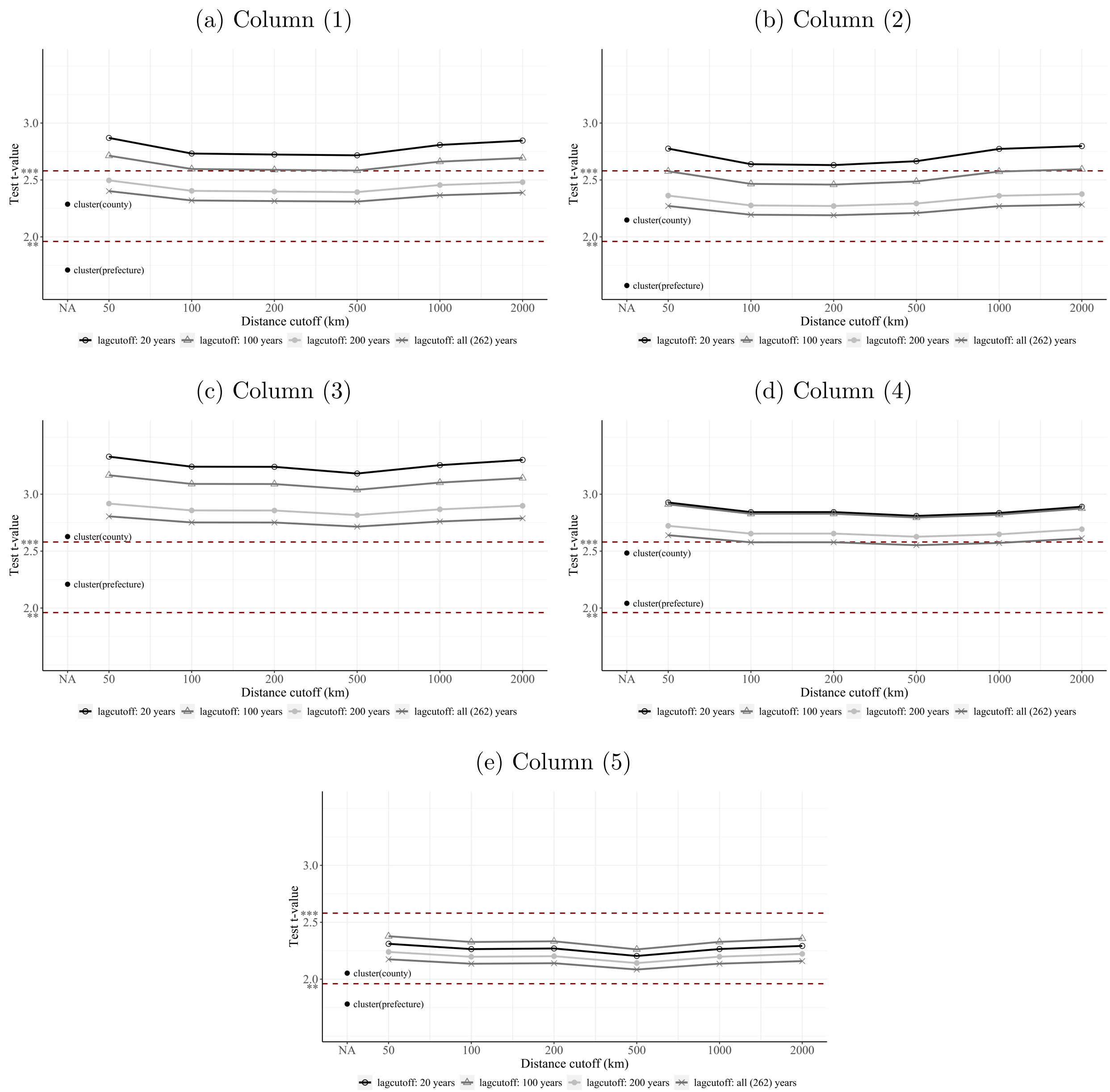
A Supplementary Figures and Tables

Figure A1: Number of rebellion onsets overtime



Note. The figure depicts the yearly number of rebellions recorded in our dataset. The vertical line represents the year in which the closure of the canal started.

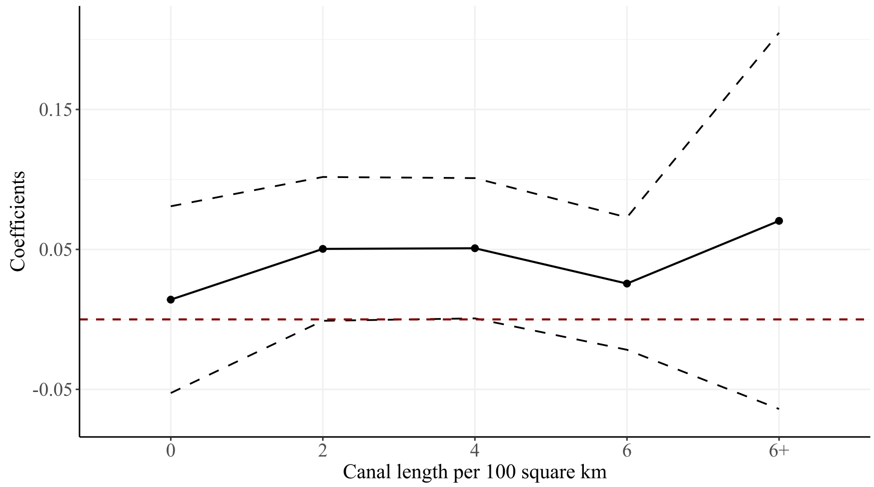
Figure A2: Test statistics of Table 3 varying standard error adjustments



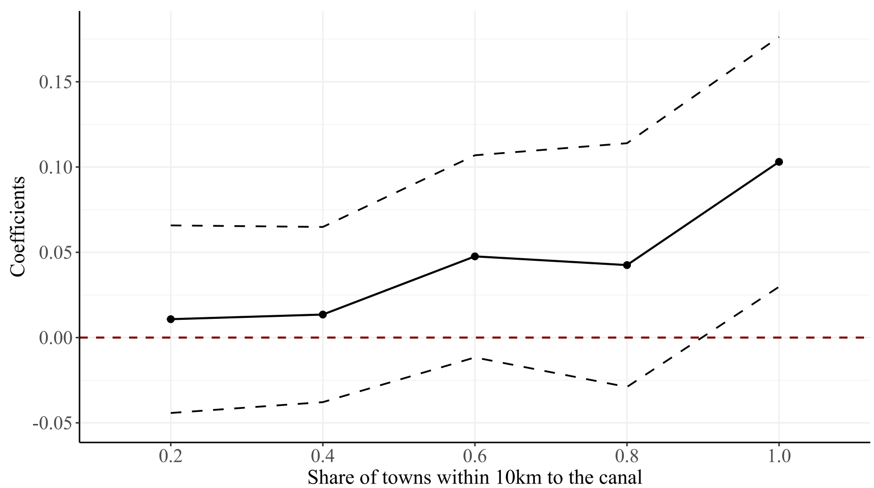
Note. The figure depicts the test t-values calculated using different methods and parameters for standard errors. The five panels correspond to the five columns in baseline Table 3. For each panel, the two isolated dots represent t-values calculated using standard errors clustered at the county level (which is reported in Table 3) and the prefecture level, respectively. The connected lines represent the t-values calculated using Conley standard errors with different combinations of distance and time cutoffs. Specifically, the x-axis displays different distance cutoffs (50, 100, 200, 500, 1000, 2000km), whereas the markers' shapes identify different time cutoffs (20, 100, 200, and 262 years). The dashed horizontal lines mark the t-values of conventional levels of significance (* 0.10 ** 0.05 *** 0.01).

Figure A3: Canal closure and rebellions: flexible treatment intensity

(a) Geographic intensity: length of canal

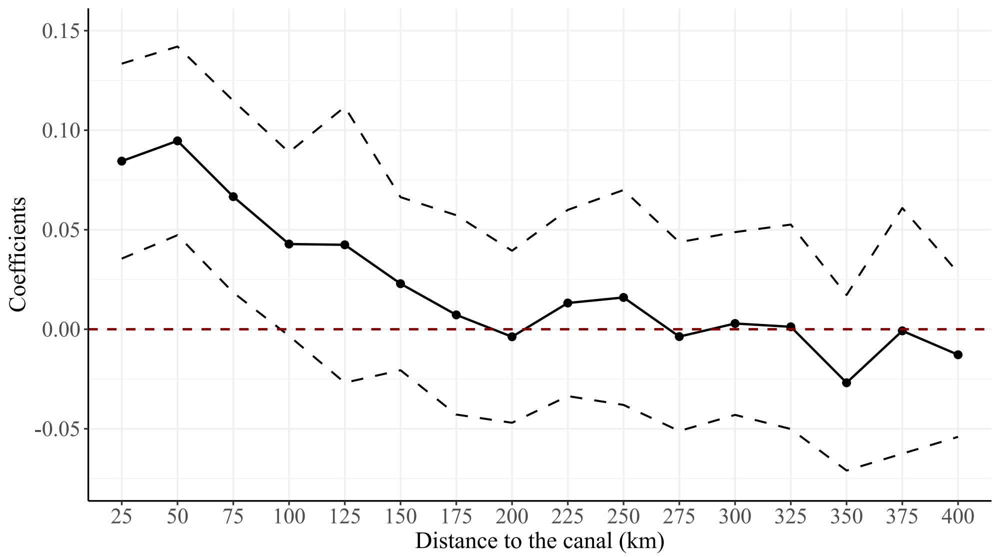


(b) Economic intensity: dependence on canal



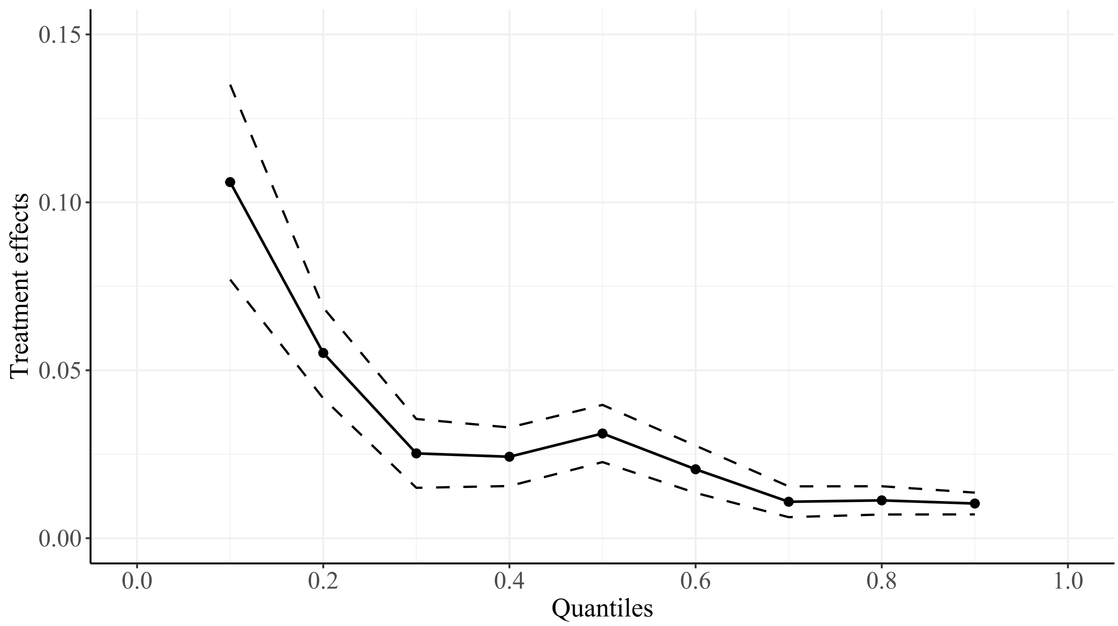
Note. The figure depicts the changes in rebellions before and after the 1826 reform by the canal's geographic and economic impact on the county. The solid lines represent the point estimates, whereas the dashed lines represent the 95% confidence intervals based on standard errors clustered at the county level. The dependent variable is the inverse hyperbolic sine of the number of rebellions normalized by 1600 population. The independent variables for panel A are the length of canal per 100 square kilometers. The independent variables for panel B are the share of 1820 towns within 10 kilometers from the canal. The regression considers county fixed effects, year fixed effects, pre-treatment rebelliousness × year fixed effects, and province × year fixed effects .

Figure A4: Canal closure and rebellions: flexible distance to the canal



Note. The figure depicts the changes in rebellions before and after the 1826 reform by a county's distance to the canal. The solid line represents the point estimates, and the dashed lines represent the 95% confidence intervals based on standard errors clustered at the county level. The dependent variable is the inverse hyperbolic sine of the number of rebellions normalized by 1600 population. The independent variables are the distance from the county's geological center to the canal. The regression considers county fixed effects, year fixed effects, pre-treatment rebelliousness × year fixed effects, and province × year fixed effects .

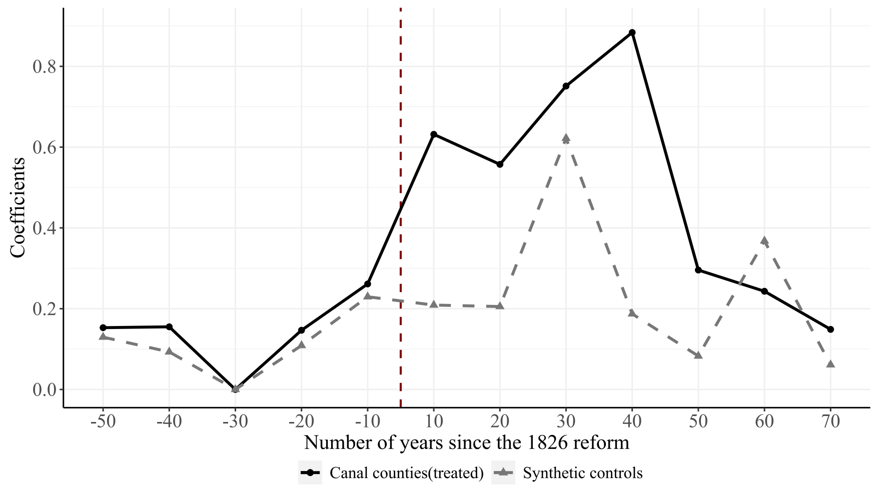
Figure A5: Changes-in-changes estimations



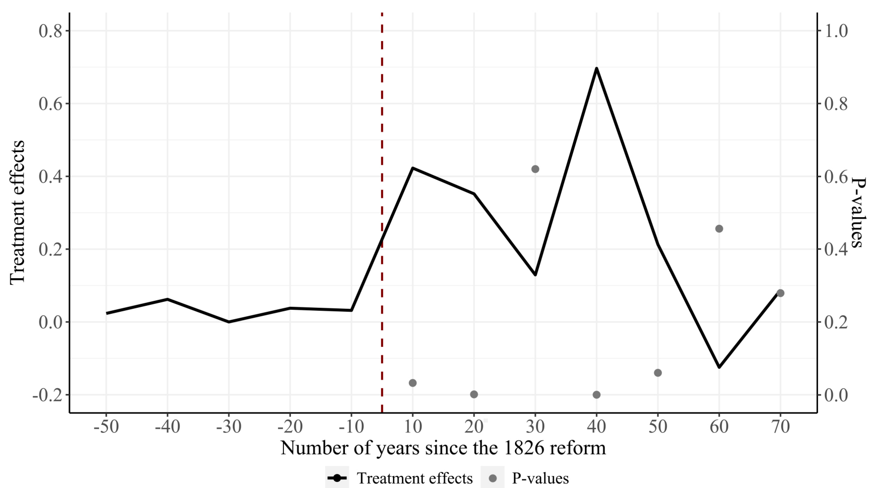
Note. The figure depicts the quantile treatment effects on the distribution estimated using the changes-in-changes method. The solid line represents the point estimates, whereas the dashed lines represent the bootstrapped 95% confidence intervals. The estimation partials out county fixed effects, year fixed effects, pre-treatment rebelliousness × year fixed effects, province × year fixed effects, and prefecture-specific time trends.

Figure A6: Synthetic control estimation

(a) Treatment v.s. synthetic control

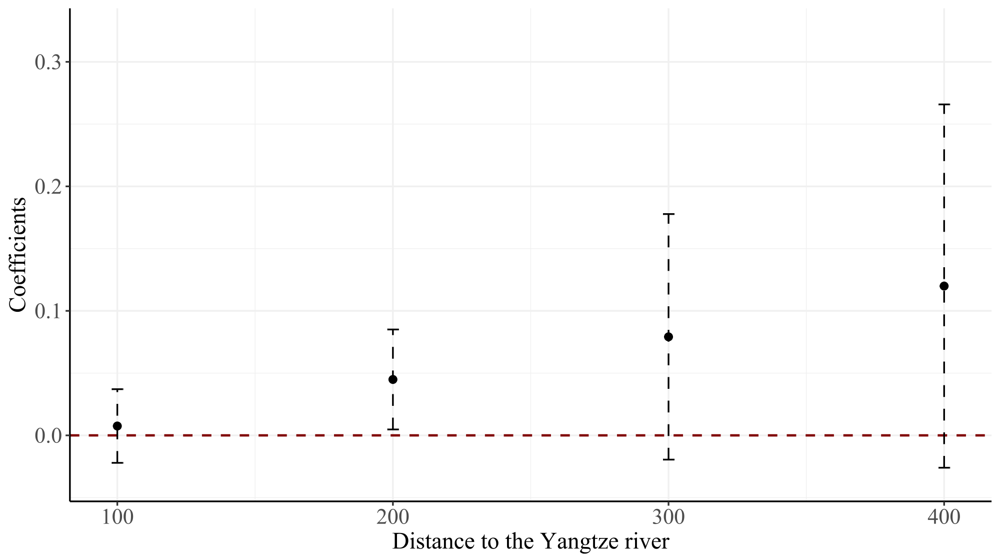


(b) Treatment effects



Note. The figure depicts the estimates using the synthetic control method. The dependent variable is the inverse hyperbolic sine transformation of the number of rebellions normalized by 1600 population. Each synthetic control group is constructed in such a way that it matches the outcome value of the treated unit in each decade before the treatment. Panel A presents the evolution of the treated (solid line) and the synthetic control groups (dashed line). Panel B presents the differences between the treated and the synthetic control groups (referring to the axis on the left), along with the p-values calculated based on a randomization inference (referring to the axis on the right). For both panels, the solid vertical line in both panels represent the 1826 treatment date, and the periods are grouped every 10 years relative to 1826 (i.e., represents the 1821–1825 period, represents the 1826–1830 period, etc..)

Figure A7: Heterogeneous effects by distance to the Yangtze River



Note. The figure depicts the changes in rebellions before and after the 1826 reform by a county's distance to the Yangtze River. The markers and capped spikes represent the OLS estimators and 95% confidence intervals. The dependent variable is the inverse hyperbolic sine of the number of rebellions normalized by population. The regression considers county and year fixed effects, pretreatment rebellions times year dummies, province-year fixed effects, and prefecture-specific year trends. Standard errors are clustered at the county level.

Table A1: Comparisons between canal and non-canal counties

	Canal counties	Non-canal counties	Difference
	(1)	(2)	(1)-(2)
Land area	6.95	7.10	-0.15
	(0.52)	(0.69)	[0.086]
Ruggedness Index	17.2	80.8	-63.6
	(19.8)	(103.2)	[12.3]
Temperature Anomaly	0.29	0.35	-0.054
	(0.29)	(0.31)	[0.039]
Frequency of flooding	0.073	0.077	-0.0040
	(0.030)	(0.025)	[0.0032]
Frequency of droughts	0.10	0.10	0.0028
	(0.037)	(0.035)	[0.0045]
Population density in 1600	3.65	2.97	0.68
	(1.47)	(1.07)	[0.14]
Maize introduction	0.66	0.68	-0.028
	(0.26)	(0.29)	[0.037]
Sweet potato introduction	0.21	0.24	-0.032
	(0.28)	(0.31)	[0.040]
Suitable for wheat	0.46	0.42	0.041
	(0.50)	(0.49)	[0.063]
Suitable for wetland rice	0.13	0.060	0.067
	(0.34)	(0.24)	[0.032]
Observations	71	465	536

Note. Standard deviations in parenthesis. Standard errors in brackets.

Table A2: Canal closure and rebellions: alternative sampling methods

	Dependent Variable: Rebellions					
	County Sample within:					Prefecture
	Prefecture	100km	150km	200km	All	Sample
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: 50-year window (1800 – 1850)					
Along Canal × Post	0.0550	0.0601	0.0660	0.0650	0.0612	0.0973
	(0.0279)	(0.0260)	(0.0259)	(0.0256)	(0.0253)	(0.0857)
	[0.0277]	[0.0264]	[0.0271]	[0.0274]	[0.0271]	[0.0759]
Observations	9150	14900	18750	21200	26800	3750
	Panel B: 100-year window (1775 – 1875)					
Along Canal × Post	0.0461	0.0494	0.0553	0.0556	0.0554	0.1487
	(0.0234)	(0.0225)	(0.0224)	(0.0223)	(0.0220)	(0.0714)
	[0.0251]	[0.0245]	[0.0248]	[0.0248]	[0.0245]	[0.0619]
Observations	18300	29800	37500	42400	53600	7500
	Panel C: 150-year window (1750 – 1900)					
Along Canal × Post	0.0416	0.0435	0.0501	0.0510	0.0510	0.1037
	(0.0211)	(0.0205)	(0.0202)	(0.0201)	(0.0199)	(0.0583)
	[0.0209]	[0.0208]	[0.0207]	[0.0207]	[0.0205]	[0.0496]
Observations	27450	44700	56250	63600	80400	11250
	Panel D: 200-year window (1711 – 1911)					
Along Canal × Post	0.0323	0.0338	0.0401	0.0417	0.0422	0.0869
	(0.0184)	(0.0180)	(0.0179)	(0.0177)	(0.0175)	(0.0542)
	[0.0180]	[0.0180]	[0.0180]	[0.0179]	[0.0177]	[0.0447]
Observations	36600	59600	75000	84800	107200	15000
	Panel E: all years (1650 – 1911)					
Along Canal × Post	0.0332	0.0337	0.0399	0.0419	0.0427	0.0825
	(0.0183)	(0.0178)	(0.0176)	(0.0174)	(0.0172)	(0.0555)
	[0.0173]	[0.0172]	[0.0171]	[0.0170]	[0.0168]	[0.0439]
Observations	47946	78076	98250	111088	140432	19650
County FE	Yes	Yes	Yes	Yes	Yes	
Prefecture FE						Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Pre-reform rebellion × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province × Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Note. The sample consists of 536 counties in the six provinces around the canal from 1650 to 1911. The dependent variable is the inverse hyperbolic sine transformation of the number of rebellions normalized by 1600 population. Each panel presents estimates using a different selection of time window. Each panel presents estimates using a different selection of comparison groups. *AlongCanal* is an indicator that equals one if the county is adjacent to the canal. *Post* is an indicator that equals one in and after 1826. Standard errors in parentheses are clustered at the county level. Standard errors in square brackets are Conley standard errors robust for spatial correlation, assuming a cut-off window of 500 km and a serial correlation of 262 years.

Table A3: Canal closure and rebellions: alternative transformations of outcome values

	Dependent Variable: Rebellions			
	(1)	(2)	(3)	(4)
	A: by 1820 population			
Along Canal × Post	0.0189	0.0190	0.0217	0.0196
	(0.0080)	(0.0083)	(0.0084)	(0.0085)
	[0.0079]	[0.0081]	[0.0083]	[0.0084]
	B: by yearly population			
Along Canal × Post	0.0183	0.0191	0.0214	0.0191
	(0.0074)	(0.0077)	(0.0079)	(0.0080)
	[0.0074]	[0.0076]	[0.0078]	[0.0079]
	C: by land area			
Along Canal × Post	0.0311	0.0312	0.0320	0.0258
	(0.0094)	(0.0098)	(0.0101)	(0.0103)
	[0.0095]	[0.0097]	[0.0099]	[0.0101]
	D: without normalization			
Along Canal × Post	0.0094	0.0092	0.0099	0.0092
	(0.0032)	(0.0033)	(0.0034)	(0.0033)
	[0.0032]	[0.0032]	[0.0033]	[0.0032]
	E: without arcsinh transformation			
Along Canal × Post	0.6154	0.6313	0.8658	0.8266
	(0.3459)	(0.3628)	(0.3577)	(0.3446)
	[0.3517]	[0.3695]	[0.3697]	[0.3610]
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Pre-reform rebellion × Year FE	No	Yes	Yes	Yes
Province × Year FE	No	No	Yes	Yes
Prefecture Year Trend	No	No	No	Yes
Observations	140432	140432	140432	140432
Counties	536	536	536	536

Note. The table presents estimates of the effect of the canal's closure using alternative approaches to constructing the outcome variable. The sample consists of 536 counties in the six provinces around the canal from 1650 to 1911. The depend variable is the number of rebellions with each of the following transformations. The outcome variable in Panel A is normalized by county population imputed for 1820. The one in Panel B is normalized by county population imputed for each year (by assuming linear changes in years without data). The one in Panel C is normalized by the size of the county. The one in Panel D is the count of rebellions without normalization. We apply the inverse hyperbolic sine transformation (arcsinh) to all outcome variables in Panel A–D. The outcome variable in Panel E is normalized by 1600 population as in the baseline, but without the arcsinh transformation. *AlongCanal* is an indicator that equals one if the county is adjacent to the canal. *Post* is an indicator that equals one in and after 1826. Standard errors in parentheses are clustered at the county level. Standard errors in square brackets are Conley standard errors robust for spatial correlation, assuming a cut-off window of 500 km and a serial correlation of 262 years.

Table A4: Canal closure and rebellions: state capacity channel

	Dependent Variable:			
	Rebellions		Attacking	Retreating
	(1)	(2)	(3)	(4)
Canal × Post	0.0142	0.0299	-0.0069	0.0034
	(0.0082)	(0.0121)	(0.0045)	(0.0032)
	[0.0090]	[0.0114]	[0.0044]	[0.0045]
Soldiers × Post	-0.0024			
	(0.0041)			
	[0.0035]			
Soldiers × Canal × Post	0.0100			
	(0.0083)			
	[0.0074]			
Prefecture Capital × Post		0.0132		
		(0.0132)		
		[0.0129]		
Prefecture Capital × Canal × Post		-0.0295		
		(0.0176)		
		[0.0166]		
Constant	0.0317	0.0306	0.0230	0.0155
	(0.0010)	(0.0010)	(0.0003)	(0.0002)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Pre-reform rebellion × Year FE	Yes	Yes	Yes	Yes
Province × Year FE	Yes	Yes	Yes	Yes
Mean of the Dependent Variable	0.0077	0.0077	0.0056	0.0038
No. of Observations	140,432	140,432	140,432	140,432
No. of Counties	536	536	536	536
Adjusted R-squared	0.0472	0.0472	0.1527	0.1550

Note. The sample consists of 536 counties in the six provinces around the canal from 1650 to 1911. The dependent variable for the first two columns, *Rebellions* , is the inverse hyperbolic sine transformation of the number of rebellions normalized by 1600 population. The dependent variable for the third column, *Attacking* , is the number of documented events in which an existing rebel group attacked the county; we normalized the value by 1600 population and applied the inverse hyperbolic sine transformation to it. The dependent variable for the fourth column, *Retreating* , is the number of documented events in which an existing rebel group retreated into the county; we normalized the value by 1600 population and applied the inverse hyperbolic sine transformation to it. *Canal* is the inverse hyperbolic sine transformation of canal length per 100 square kilometers. *Soldiers* is the size of military establishment pre-assigned during the 1750s. *Prefecture Capital* is an indicator that equals one if the county served as the administrative center of the prefecture it belonged to. *Post* is an indicator that equals one in and after 1826. Standard errors in parentheses are clustered at the county level. Standard errors in square brackets are Conley standard errors robust for spatial correlation, assuming a cut-off window of 500 km and a serial correlation of 262 years.

Table A5: Canal closure and rebellions: trade access channel

	Dependent Variable:		
	Town Number	Rebellions	
	(1)	(2)	(3)
Canal × Post	-0.2683	0.0466	0.0130
	(0.0638)	(0.0189)	(0.0095)
	[0.0604]	[0.0165]	[0.0094]
Courier × Post		0.0006	
		(0.0013)	
		[0.0011]	
Canal × Courier × Post		-0.0070	
		(0.0029)	
		[0.0025]	
Temperature Anomaly			-0.0022
			(0.0033)
			[0.0046]
Canal × Temperature Anomaly			-0.0135
			(0.0055)
			[0.0057]
Temperature Anomaly × Post			-0.0082
			(0.0075)
			[0.0085]
Canal × Temperature Anomaly × Post			0.0251
			(0.0151)
			[0.0154]
Constant	2.1284	0.0309	0.0340
	(0.0072)	(0.0015)	(0.0013)
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Mean of the Dependent Variable	2.0982	0.0330	0.0330
No. of Observations	1,058	140,432	140,432
No. of Counties	529	536	536
Adjusted R-squared	0.6588	0.0256	0.0254

Note. The sample consists of 536 counties in the six provinces around the canal from 1650 to 1911. The sample for the first column consists of 560 counties in the six provinces around the canal, for 1820 and 1911. The dependent variable for the first column, *Towns*, is the logarithm of the number of towns within the county. The sample for the next two columns consists of 536 counties in the six provinces around the canal from 1650 to 1911. The dependent variable for the second and the third column, *Rebellions* , is the inverse hyperbolic sine transformation of the number of rebellions normalized by 1600 population. *Canal* is the inverse hyperbolic sine transformation of canal length per 100 square kilometers.. *Courier* is the inverse hyperbolic sine transformation of courier route length per 100 square kilometers. *Temperature Anomaly* is an indicator that equals one if the temperature deviated from the mean by more than one standard deviation. *Post* is an indicator that equals one in and after 1826. Standard errors in parentheses are clustered at the county level. Standard errors in square brackets are Conley standard errors robust for spatial correlation, assuming a cut-off window of 500 km and a serial correlation of 262 years.

Table A6: Canal closure and rebellions: agricultural productivity

	Dependent Variable:		
	Grain Price	Rebellions	
	(1)	(2)	(3)
Canal × Post	0.0058	0.0240	0.0045
	(0.0037)	(0.0116)	(0.0064)
	[0.0044]	[0.0106]	[0.0067]
Rice suitability × Post		-0.0154	
		(0.0173)	
		[0.0155]	
Canal × Rice suitability × Post		-0.0209	
		(0.0160)	
		[0.0147]	
Wheat suitability × Post			0.0405
			(0.0092)
			[0.0103]
Canal × Wheat suitability × Post			0.0286
			(0.0187)
			[0.0170]
Constant	0.5974	0.0318	0.0259
	(0.0004)	(0.0008)	(0.0013)
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Mean of the Dependent Variable	0.5980	0.0330	0.0330
No. of Observations	91,110	140,432	140,432
No. of Counties	560	536	536
Adjusted R-squared	0.7061	0.0254	0.0262

Note. The sample consists of 536 counties in the six provinces around the canal from 1650 to 1911. The dependent variable is the inverse hyperbolic sine transformation of the number of rebellions normalized by 1600 population. *Canal* is the inverse hyperbolic sine transformation of canal length per 100 square kilometers. *Post* is an indicator that equals one in and after 1826. Standard errors in parentheses are clustered at the county level. Standard errors in square brackets are Conley standard errors robust for spatial correlation, assuming a cut-off window of 500 km and a serial correlation of 262 years.

Table A7: Canal closure and the Green Gang

	Dependent Variables:
	Green Gang senior members (early 20^{th} century)
	(1)
Along Canal	1.1736
	(0.2792)
Prefecture FE	Yes
Mean of the Dependent Variable	0.2435
No. of Observations	575
Adjusted R-squared	0.2076

Note. The sample consists of 575 counties in the six provinces around the canal. The dependent variable is the number of Green Gang senior members in the early 20th century. *AlongCanal* is an indicator that equals one if the county is adjacent to the canal. Standard errors in parentheses are robust for heteroskedasticity.

B More on the Historical Background

This section offers additional historical background on sea transportation reform and the abandonment of the Grand Canal in 1826. We start by introducing the tribute grain system in the Qing Dynasty, when the Grand Canal played a central role. This is followed by a brief discussion of the alternative sea route. We conclude the section with a discussion of the potential motivations for the reform.

B.1 The Tribute Grain System and the Grand Canal

The tribute grain system had been operated in ancient China from as early as the first century AD to address the spatial mismatch between the production and consumption of rice. For most of China's history, the political and economic center was located in the north part of the country. Many of its residents were members of the court, official personnel, scholars, and imperial soldiers and their families (Morse, 1913; Chi, 1936), and did not produce their own rice. As shown in Figure B1, however, most of the rice-producing regions were concentrated in the south, particularly in the middle-lower Yangtze River plain. This required the government in each era to collect rice taxes from the south and transport them to the capital in the north. Most of China's natural rivers run west-to-east, though, and the cost of land transportation (via humans or animals) was at least tenfold the cost of water transportation (Shiue, 2002). Therefore, governments throughout China's history invested massive resources in constructing and maintaining the Grand Canal — the only artificial waterway that linked the south to the north.

The Qing Dynasty inherited the Ming Dynasty's tribute grain system. The government collected rice taxes from eight provinces (plotted in Figure B2) in the central and southern parts of China. The circled area in Figure B2 highlights the intersection between Jiangsu and Zhejiang, the most productive area in the region, which contributed more than 50% of the rice collected. The grains collected from Hubei, Jiangxi and Anhui were first transported via the Yangtze River to Huai'an, where the Yangtze River and the Grand Canal intersect. They were then delivered to Beijing via the canal together with the Jiangsu and Zhejiang grains. As a result, the northern part of the canal played a more crucial role in the canal-centered tribute grain system.

B.2 The Alternative Sea Route

While the government had used the Grand Canal as the exclusive route for tribute grain transportation, there was technically an alternative sea route that could have been considered. In fact, maritime sailing technologies had been developed long before the first sea-shipping experiment, and the specific route used for the sea-shipping of tribute rice, which connected Shanghai to Tianjin through the Yellow Sea and the Bohai Sea, was first explored in the Yuan Dynasty (1271 – 1368) (Figure B3 shows the route). The readily available technology allowed seven maritime treasure voyages to take place between 1405 and 1433, which reached as far as the Arabian Peninsula and East Africa. In addition, the same chartered ships that were later used in sea transportation had already been used by private agents at the beginning of the Qing Dynasty. By the end of the eighteenth century, there were 3,000 such ships, twice as many as were used in the first sea shipping experiment.

Despite the readily-available technology, the early Qing emperors generally strongly objected to the sea transportation. In 1656, the Shunzhi Emperor issued the sea ban that prohibited any private maritime trading. This ban was reinforced by the "Great Clearance" in 1662 that required all coastal residents to destroy their property and move 20 km inland. While the ban was lifted in 1684, subsequent emperors continued imposing strong restrictions on maritime trade and rejected all proposals to transport grain by sea. In 1816, the Jiaqing Emperor enacted an order that strictly prohibited any discussion of sea transportation (Ni, 2005).

Historians have highlighted four reasons behind the emperors' conservative attitudes towards the sea. The first motivation was related to national security, to protect against threats from overseas. This was the strongest justification for the sea ban enacted in the Ming and early Qing dynasties. Second, after centuries of maritime isolationism, the emperors were generally ignorant about the ocean. This prompted them to avoid all dealings with the sea as a result of ambiguity (uncertainty) aversion (Epstein, 1999). Third, Confucian culture highly valued obedience on the part of emperors to the time-honored rules of their fathers. In particular, emperors were required to be the role models of filial piety, and one of the criteria of filial piety is not to alter the conventions of their deceased parents. Thus, any emperor who wanted to alter the established convention without a sufficiently strong justification would incur a high reputation cost. Fourth, such a would-be reformer would have to overcome resistance from strong vested interests and entrenched bureaucrats who benefited from maintaining the tribute grain transportation along the Grand Canal. These personal and political constraints kept sea transportation off the agenda until the early nineteenth century.

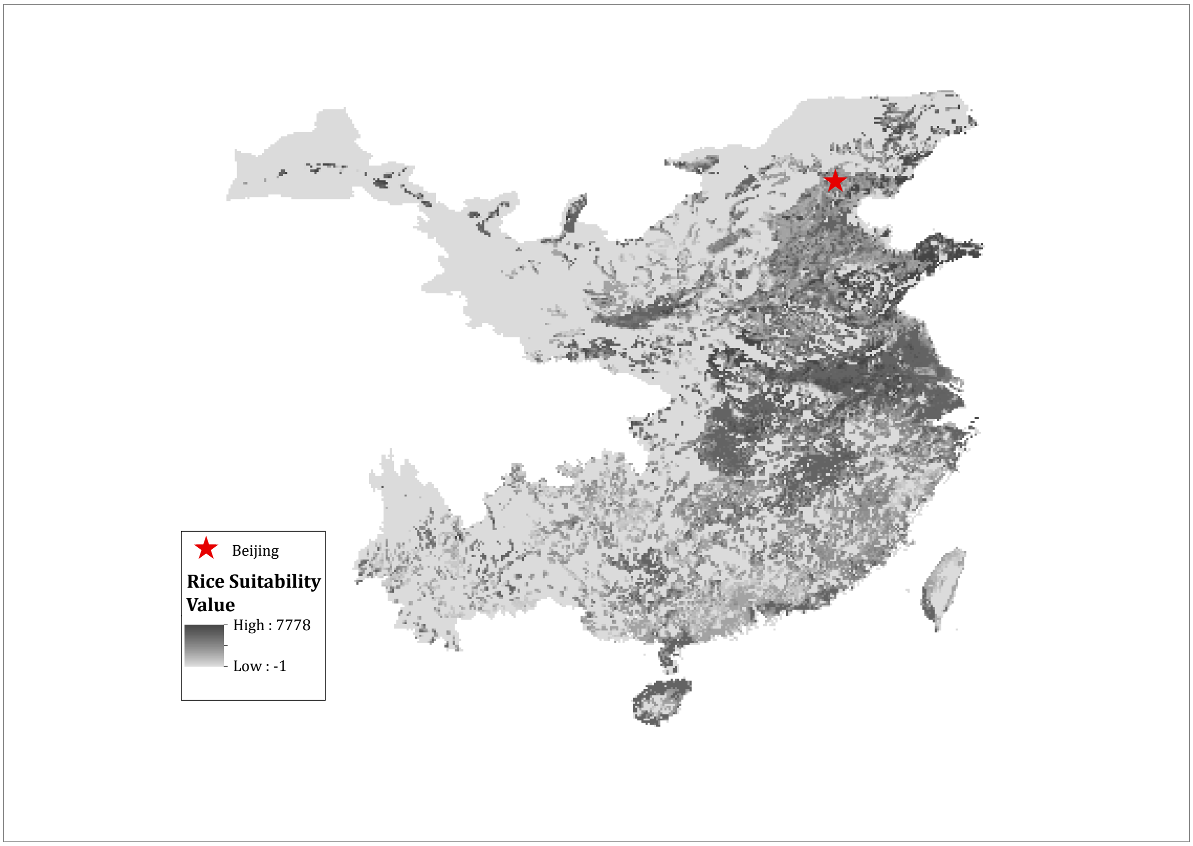
B.3 From the Canal to the Sea

The first sea transportation was implemented in 1826 following a natural disaster and a turnover of emperors that altered the associated constraints. The trigger of the reform process was the breach of the Gaojia Dam, the embankment dam at Hongze Lake near the intersection of the Yellow River and the canal, which made the nearby part of the canal too clogged to navigate. Since the Gaojia Dam was designed to store water in Hongze Lake to flush out the sediments carried into the canal by the Yellow River, the breach of the dam caused a lack of water flow, making the canal silt-clogged and unnavigable. While the breach was temporary, it prompted Daoguang — the newly-enthroned middle-aged emperor — to seriously consider the alternative that had been previously rejected by his father (the Jiaqing Emperor). Launching the sea transportation reform allowed the new emperor to dismantle the old patronage networks and bring in more of his own trusted aides. Although the sea-shipping experiment was resisted by vested interest groups and suspended one year later, officials who played significant roles in implementing the experiment remained in office, and many of their careers flourished during the Daoguang's reign (see Table B1for details).

Importantly, there is no historical evidence to suggest that the reform was inspired or advanced by any actual or anticipated rebellion along the canal. We surveyed the variety of reasons adduced by officials in support of the reform throughout the process. The main argument for sea transportation was its efficiency: it was faster, less expensive, and required less labor. Supporters never mentioned concerns about social instability as a motivation for the reform. In fact, opponents frequently raised concerns about potential disorder against the reform. In one memorial to the Daoguang Emperor, the opponent pointed out that "the sailors would definitely cause trouble" if transportation via the canal was abandoned.

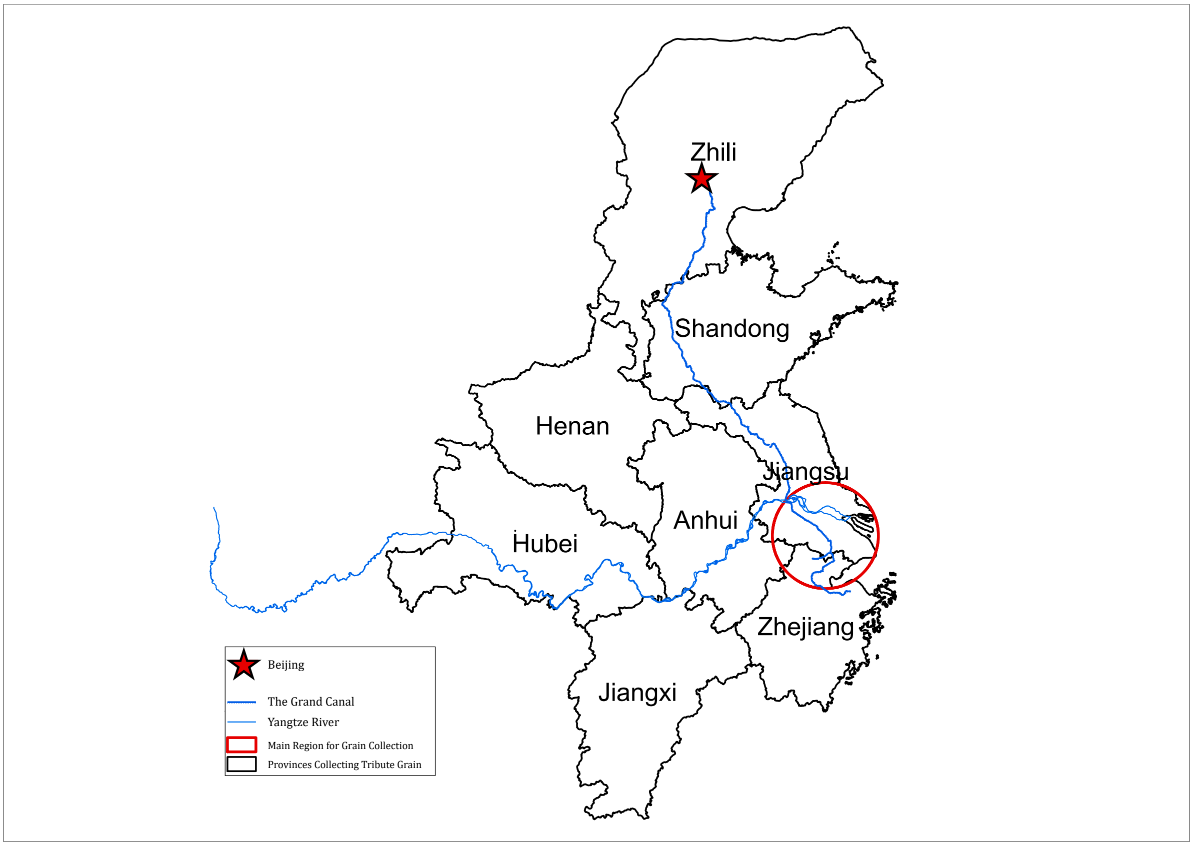
Figures and Tables for Appendix B

Figure B1: Suitability index for wetland rice (irrigation, medium input)



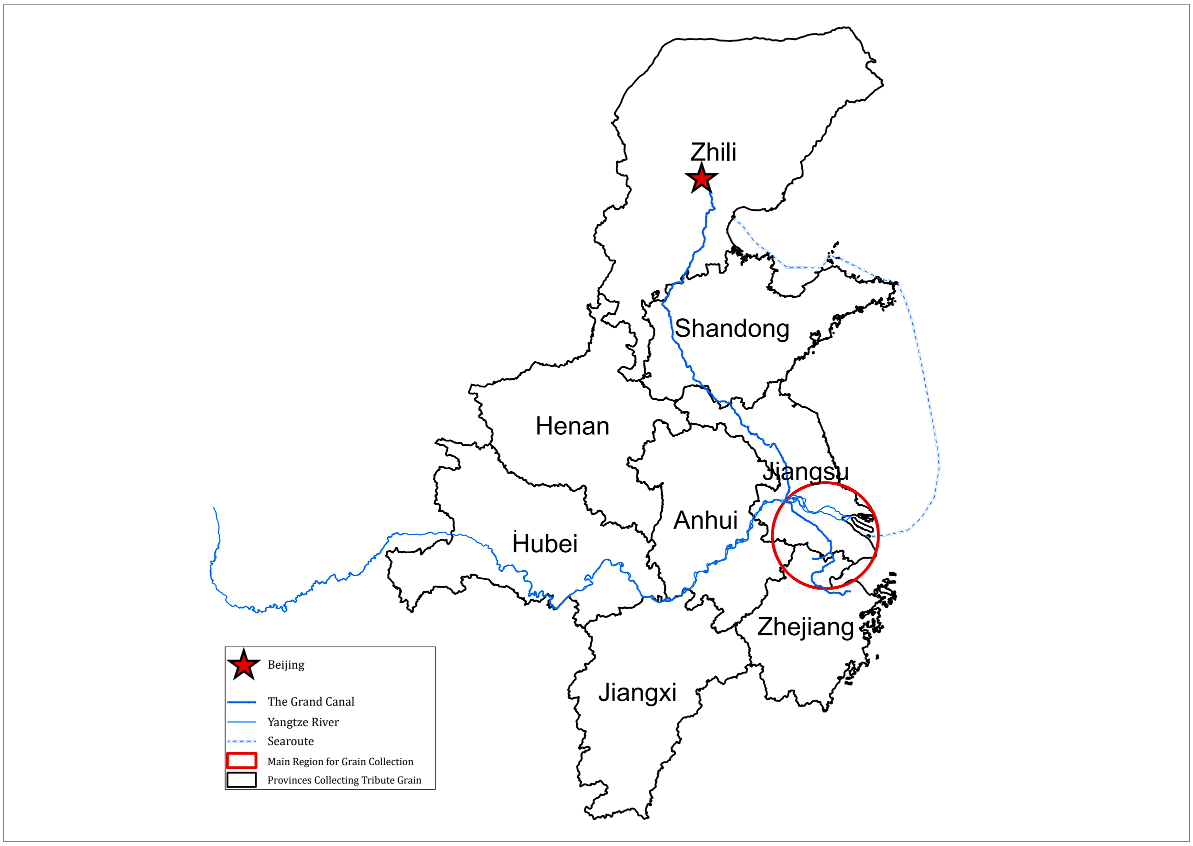
Note. The figure depicts the suitability index for wetland rice plantation across the country under irrigation and medium input.

Figure B2: Sources and shipping routes of tribute rice in the Qing Dynasty



Note. This figure depicts the provinces extensively involved in the tribute grain system. In particular, the circled area produced more than 5% of the rice collected.

Figure B3: The sea route for grain transportation



Note. This figure additionally depicts the sea route for the transportation of tribute grain post the reform.

Table B1: The leading officials for and against sea transportation

Attitude	Name	Birth and Death	Age in 1825	Career after Reform
For	Yinghe	1771–1840	55	Demoted
	Qishan	1776–1854	50	Promoted
	Shu Tao	1779–1839	47	Promoted
	Changling He	1785–1848	41	Promoted
Against	Yuanyu Wei	1767–1825	59	Dead
	Yuting Sun	1752–1834	74	Demoted
	Jian Yan	1757–1832	69	Demoted
	Shicheng Zhang	1762–1830	64	Demoted

Note. Yinghe was demoted in 1827 for misconduct irrelevant to the grain transportation.

C Additional Data Description

C.1 Coding Method

This section summarizes the coding method of rebellions. We start by describing the structure and content of *Qing Shilu* (The Veritable Records of Qing Emperors). We then describe the detailed steps taken to locate and code the relevant records and provide illustrative examples.

Qing Shilu is a collection of 13 books, each corresponding to one of the 13 emperors in Qing China. These books are compiled by Qing officials from the emperors' utterances and edicts into daily records of imperial affairs, and from officials' memorials on matters of government, including all events of national significance. In particular, it has been recognized by historians as the most reliable and comprehensive source that is available for information on rebellions throughout the Qing Dynasty (see Yang (1975) for an evaluation of the completeness and accuracy of the information on rebellions in *Qing Shilu*).

Collecting and extracting data from *Qing Shilu* is challenging because of the nature of these historical records. The original books of *Qing Shilu* are in their traditional format (right-to-left, vertical writing) and traditional Chinese language, which is hard to read and, until now, cannot be reliably processed by automated computer programs. Moreover, since they are not statistical books in a standard format, it often requires a thorough reading of the text to identify and extract the relevant information. To make the data collection process transparent to the readers, we document below the detailed steps in compiling the dataset, the complications that have arisen, and decisions that have been made:

Step 0 We obtain digitized versions of the books available at *Chinese Text Project* (`https://ctext.org`). See Sturgeon (2021) for a description of the project.

Step 1 The first step is to identify items in the books that may be related to rebellions. We do so by searching for the keyword "fei" (literally, "bandits"). We choose this keyword because it is the most common term used by the Qing government to refer to the rebels.[1] A typical record would start by describing the incidents and actions of the rebels, followed by the emperors' edicts on how to address the issue.

Step 2 For each record identified through the keyword-searching, we conduct a thorough reading of the raw text to refine the results returned in the first step and to verify their validity and relevance for the study. Specifically, we discard records in the following situations:

- The word "fei" has a different meaning than "bandits" given the context. For example, in phrases such as "zhouye feixie" (literally, "untiring day and night"), "fei" has another meaning of "not", which is not relevant for rebellions.

[1]The Qing government often referred to rebel groups according to their identity (usually the location, the leader's surname, or a defining feature of the incident), followed by the keyword "fei". For example, "yue fei" refers to rebel groups from Guangdong and Guangxi (also named "yue"); "cuan fei" refers to rebels moving around ("cuan").

- The record is relevant for rebellions, but the incident occurred outside our sample region (Zhili, Henan, Shandong, Anhui, Jiangsu, Zhejiang). This excludes most records of rebellions in other parts of the country (e.g., those in Guangdong, Sichuan, Gansu, etc.).

Step 3 We exclude duplicated records of the same incident. We identify duplications based on the following criteria:

- There are indicators such as "as reported/discussed earlier", and the elements (esp. places involved) match with a previously documented report.
- There are observable connections to a previously recorded incident (e.g., one is a continuation of the other, or initiated by the same leader, or there is some sort of collusion between the rebels, etc.)

Step 4 We identify the year of each incident reported. This step is straightforward for most cases, as these incidents were generally reported in the same year as they took place. However, complications may arise when an incident was reported one or more years later, especially when it took place at the end of a year and was reported at the beginning of the next year. Therefore, while we rely primarily on the year of reporting to identify time of the incident, we also make corrections accordingly when phrases such as "last year" and "back in *some specific year*" are spotted.

Step 5 We pinpoint the location of the incident by matching the names mentioned in the reports to county administrations as of 1911. In the rare cases in which the names did not match easily (often due to name changes, merges, and splits), we rely on online searches of local history to link the county names mentioned in the reports to 1911 counties.[2]

Step 6 We classify the actions of the rebels in each county into five categories, which represent the nature of the incident: *onset*, *attacking*, *defending*, *stationing*, and *retreating*. The definitions and criteria of these categories are summarized below:

- Onset: a single incident or the initiation of a rebel group at a local place; often identified by phrases such as "huyou" (literally, "suddenly there occurred"), "shu qi" (raised their flags), "qi shi" (initiated a rebellion), etc.
- Attacking: a rebel group (already existed) tried to attack a county (which is not yet occupied by that group).
- Defending: the government attempted to use military force to repress a rebel group that already existed.
- Retreating: the rebel group (that already existed) retreated into a different place after being defeated by the governmental military force.

[2]Such occasions represent a very small portion of our sample, as county boundaries were relatively consistent throughout the Qing Dynasty (Ge, 1997).

- Stationing: the longstanding existence of a rebelling group in a place, but when it first started was not explicitly documented in the reports.

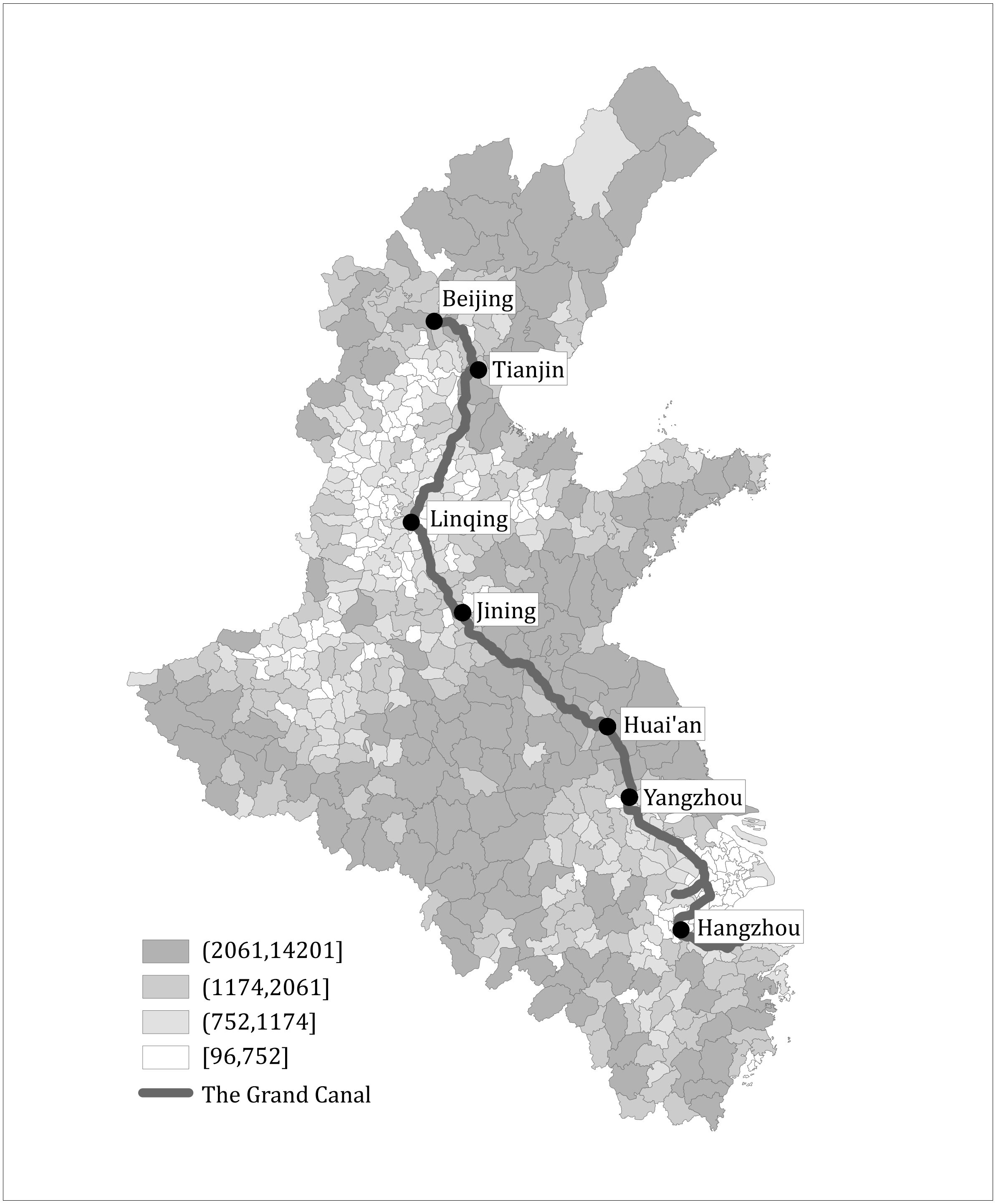
In cases where the rebels spread across multiple counties, we code their actions in each county separately. For example, consider a group of rebels that started in county A, attacked counties B and C, and retreated into county D after being repressed. In our data set, county A will receive 1 count of *onset*, counties B and C will each receive 1 count of *attacking*, and county D will receive 1 count of *retreating*. If there are more than one actions associated with one county in a single incident, we prioritize *onset* over other categories (e.g., if a rebel group started in county A and repressed in the same county, we code it as an *onset* in county A).

Step 7 Finally, for each county-year, we count the number of events by action type and construct a balanced panel where the value of 0 is assigned to county-year pairs with no reports of a specific type of action.

The reports of rebellions that we have identified contain both single, small-scale incidents and those larger events that spread a wide range of areas and lasted for years (e.g., the Taiping Rebellions, the Nian Rebellion, etc.). According to the definitions above, single, small-scale incidents at the local level would be categorized as *onset*, whereas those long-lasting and widespread events would be mostly captured by the other four categories (*attacking*, *defending*, *retreating*, and *stationing*). For this study, we focus primarily on those local incidents (*onset*). We do not use the other four categories because they often reflect complicated military strategies.

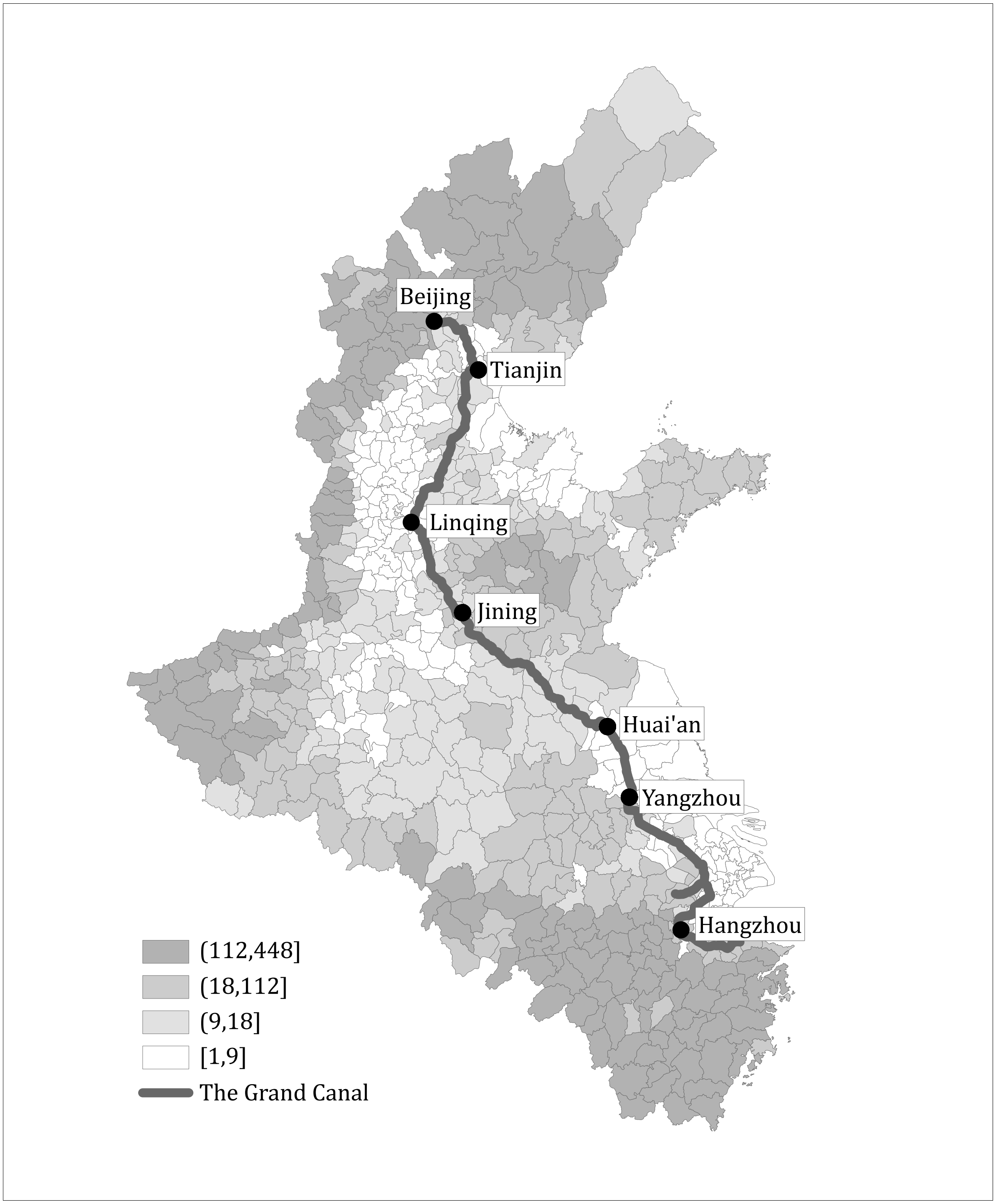
Figures and Tables for Appendix C

Figure C1: Spatial distribution of land size



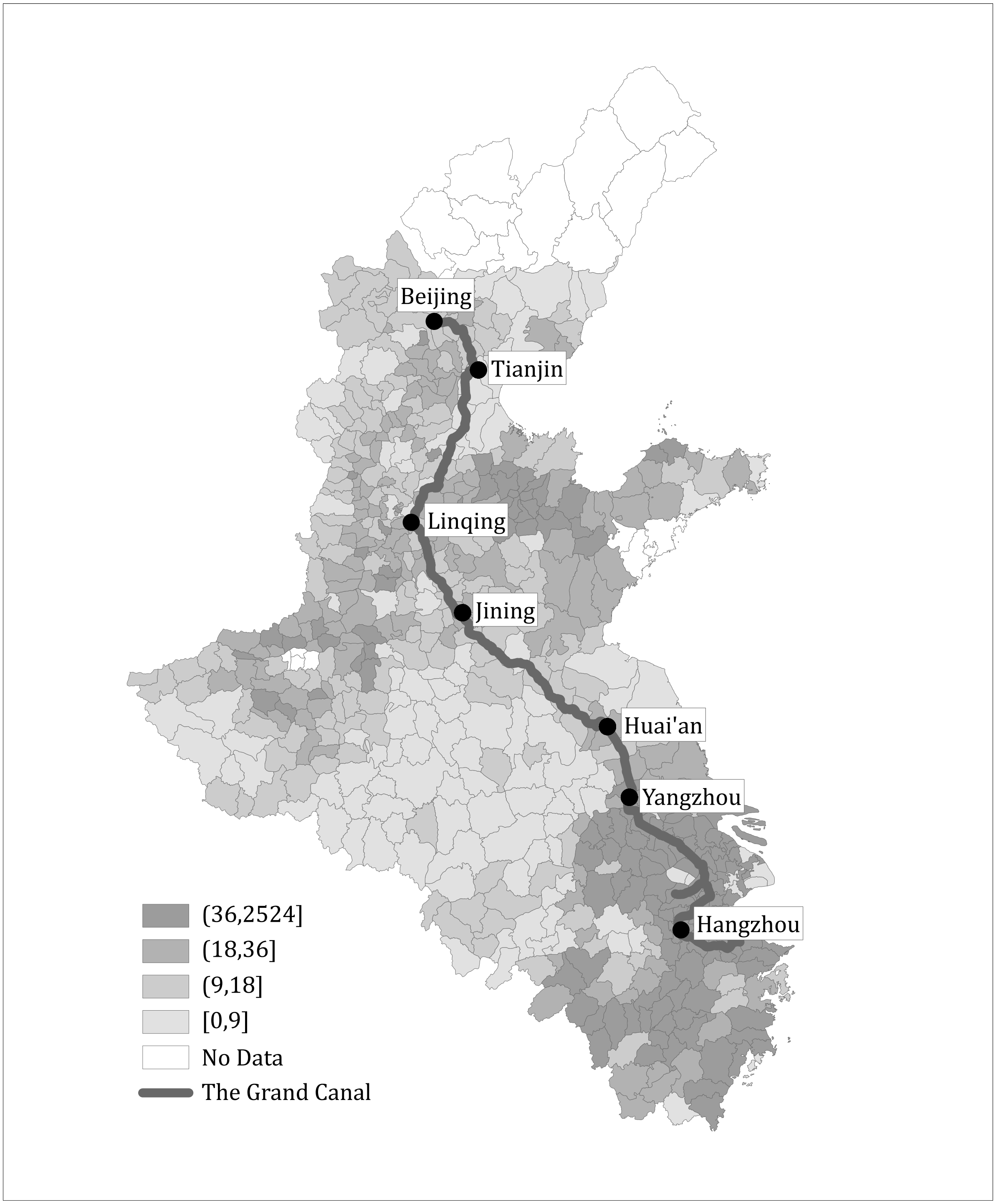
Note. The figure depicts the spatial distribution of a county's land size, measured in kilometers.

Figure C2: Spatial distribution of terrain ruggedness



Note. The figure depicts the spatial distribution of a county's average terrain ruggedness index.

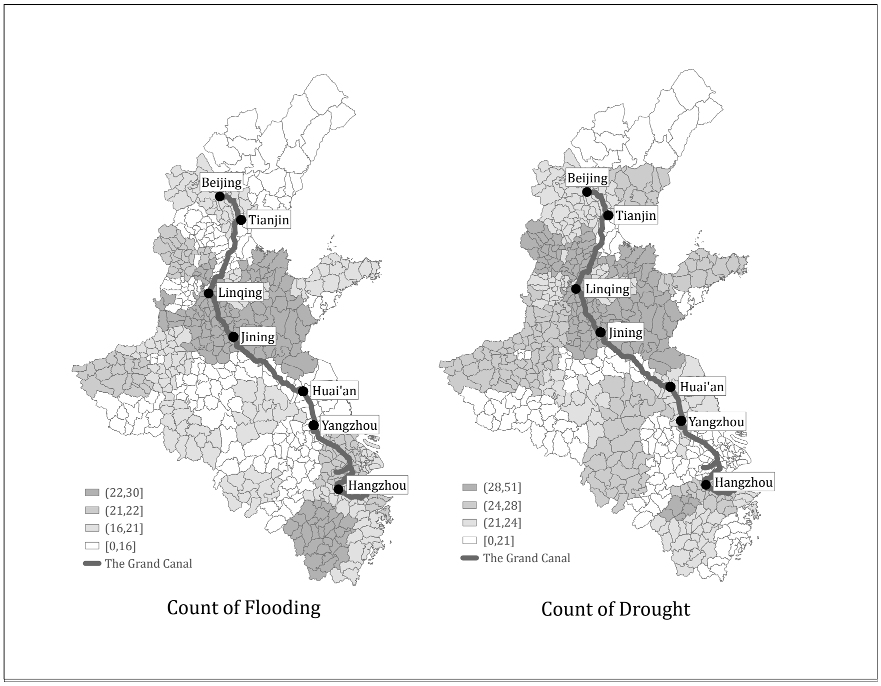
Figure C3: Spatial distribution of initial population density



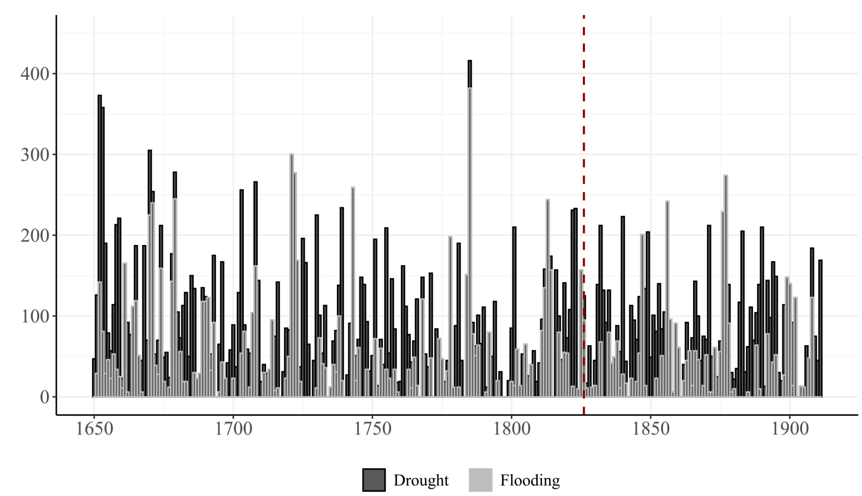
Note. The figure depects the spatial distribution of a county's population density in 1600, measured in people per km^2.

Figure C4: Spatial and chronological distribution of droughts and floodings

(a) Spatial distribution

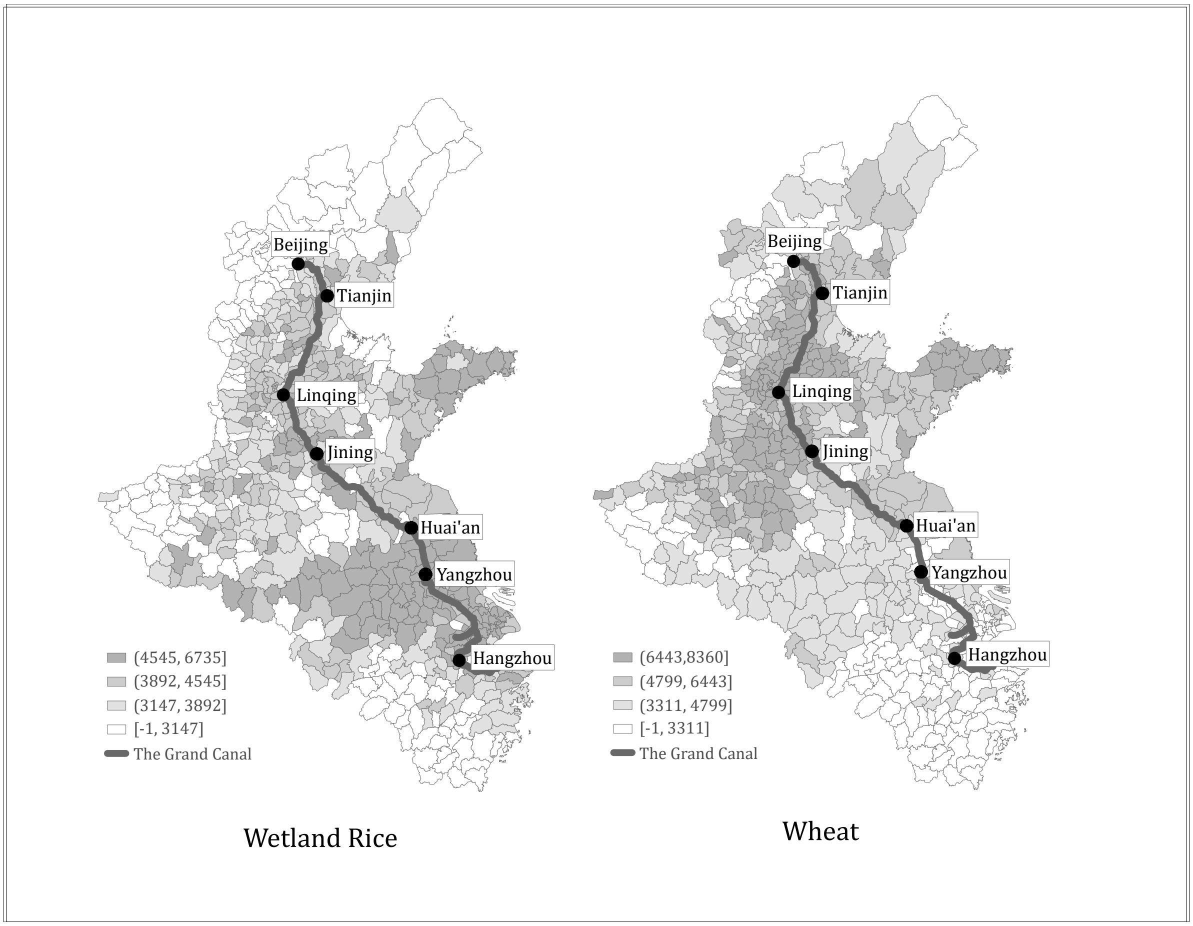


(b) Chronological distribution



Note. This figure depicts the spatial and chronological distribution of droughts and floodings documented in history. Panel (a) shows the total number of drought and flooding records in each county throughout the sample period. Panel (b) shows the total number of each event by year.

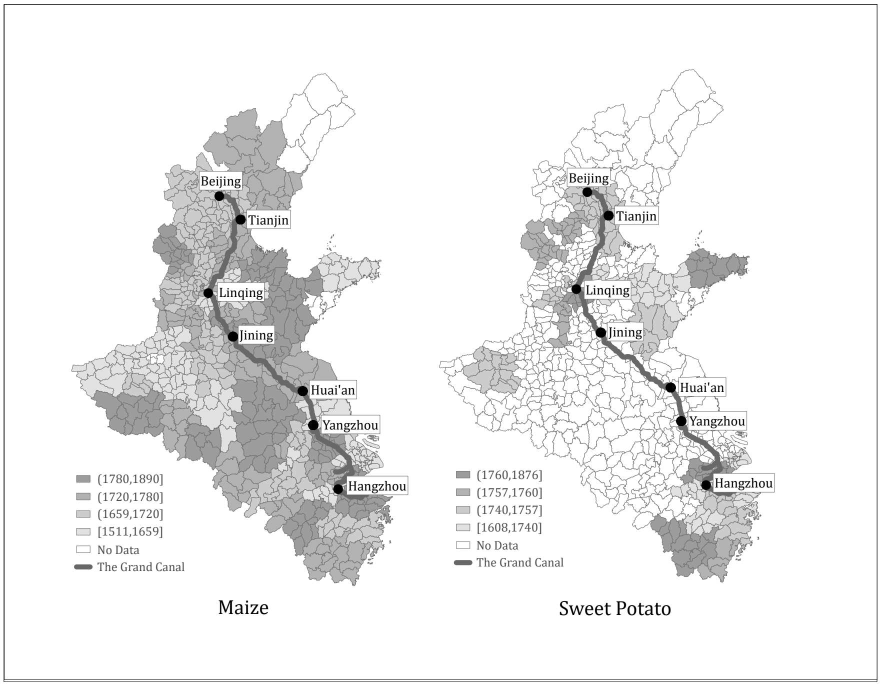
Figure C5: Spatial distribution of crop suitability for wheat and wetland rice



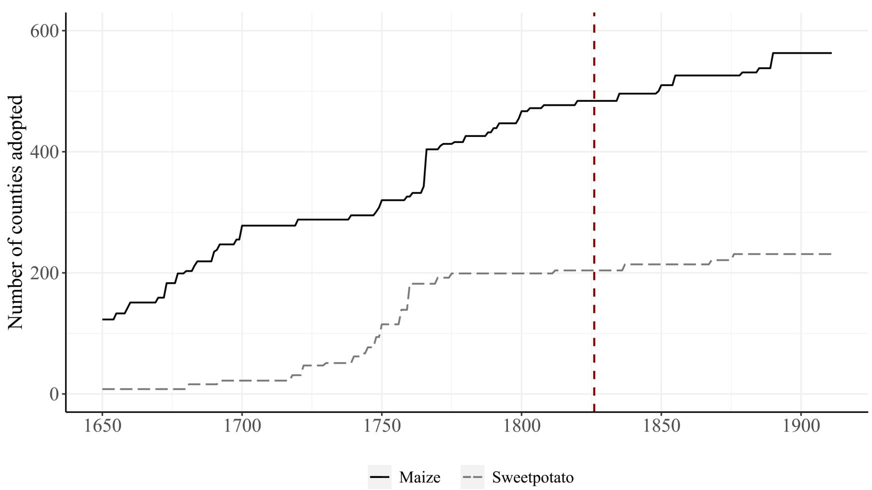
Note. This figure depicts the spatial distribution of crop suitability index for wetland rice and wheat.

Figure C6: Spatial and chronological distribution of new world crop adoption

(a) Spatial distribution



(b) Chronological distribution



Note. This figure depicts the spatial and chronological distribution of the introduction of new world crops (maize and sweet potato). Panel (a) shows the year in which each crop was first introduced. Panel (b) shows the cumulative number of counties that had adopted each crop every year.

References

Cao, Yiming, and Shuo Chen. 2022. “Data and code for: Rebel on the Canal: Disrupted Trade Access and Social Conflict in China, 1650–1911.” *American Economic Association [publisher]*, Inter-university Consortium for Political and Social Research [distributor], http://doi.org/10.3886/E157781V1.

Chi, Ch’ao-Ting. 1936. *Key Economic Areas in Chinese History: As Revealed in the Development of Public Works for Water-control.* London:Routledge.

Epstein, Larry G. 1999. “A Definition of Uncertainty Aversion.” *Review of Economic Studies*, 66(3): 579–608.

Ge, Jianxiong. 1997. *Changes in Boundaries and Administrative Divisions in Chinese History (Zhongguo Lidai Jiangyu Bianqian, in Chinese).* Beijing:Commercial Press.

Morse, H.B. 1913. *The Trade and Administration of China.* London:Longmans, Green and Company.

Ni, Yuping. 2005. *Grain Transportation in the Qing Dynasty and Social Change (Qingdai Caoliang Haiyun Yu Shehui Bianqian, in Chinese).* Shanghai:Shanghai Bookstore Publishing House.

Shiue, Carol H. 2002. “Transport Costs and the Geography of Arbitrage in Eighteenth-Century China.” *American Economic Review*, 92(5): 1406–1419.

Sturgeon, Donald. 2021. “Chinese Text Project: A Dynamic Digital Library of Premodern Chinese.” *Digital Scholarship in the Humanities*, 36(Supplement_1): i101–112.

Yang, Ch’ing-k’un. 1975. “Conflict and control in late imperial China.” , ed. Carolyn Wakeman, Frederic E.; Grant, Chapter Some Preliminary Statistical Patterns of Mass Actions in Nineteenth-Century China. University of California Press.