

An Empire of Their Own? A Longitudinal Dataset of Cotton Production and Industry in Republican China, 1919-1937

Yan Duan (corresponding author)

Nanjing Agricultural University, Nanjing, 18120168826@163.com 

Adam K. Frost

Southern Denmark University, adamf@sam.sdu.dk 

Spencer Dean Stewart

Purdue University, stewa443@purdue.edu 

Received 10 January 2025, revised 22 May 2025, accepted 1 September 2025

Abstract

This article introduces a new, hand-collected dataset documenting cotton production and industry in Republican China from 1919 to 1937. The dataset provides granular, county-level data on cotton acreage, yield, and the cultivation of both “native” and “foreign” (American upland) cotton varieties, offering a detailed perspective on the spatial and temporal dynamics of agricultural change during this critical period. This foundational data is augmented by three additional geospatial datasets documenting the locations and attributes of key economic and scientific nodes: (1) modern cotton mills, including their industrial capacity; (2) agricultural experiment stations; and (3) treaty ports and commercial centers. The dataset was constructed through a meticulous process of transcription, validation, error correction, and georeferencing, drawing from a variety of primary sources, including the publications of the Chinese Cotton Millowners Association and other contemporary industry reports. By detailing the dataset’s construction and acknowledging its limitations, this paper makes the case for its use as a unique tool to investigate the core mechanisms of agricultural modernization. Its multi-layered structure enables researchers to test hypotheses about biological innovation and institutional support, and to place China’s experience within the global history of cotton. Ultimately, it offers a granular case study of a nation striving to build its own industrial base, thereby complicating classic core-periphery models of development.

Keywords

cotton production – economic history – republican china – agricultural modernization – technology transfer – spatial analysis

Duan, Y.; Frost, A. K.; Stewart, S. D. (2025). Cotton Production in Republican China: A Longitudinal Dataset (1919-1937). Purdue University Research Repository. DOI: 10.4231/S460-H323.

1. Introduction

China's Republican era (1912–1949) was a period defined by ambitious yet fraught attempts at state-building and industrial modernization. The Nanjing Decade (1927–1937), in particular, represented a concentrated effort by the Kuomintang government to forge a modern, centralized nation-state and a cohesive national economy (Zanasi, 2006). However, these top-down modernization projects were continually challenged by internal political fragmentation, limited state capacity in the countryside, and external pressures (Eastman, 1983; Duara 1988). It is within this context of state-led development and its on-the-ground limitations that a concerted effort to modernize the cotton industry emerged, motivated by the growing demand from China's burgeoning textile mills (Fong, 1932; Feuerwerker, 1983). This modernization drive led to the introduction and adoption of American cotton varieties, which offered fiber qualities more suitable for mechanized production (Pomeranz, 1993; Köll, 2003). This article introduces a novel, hand-collected dataset on cotton production, designed to facilitate new research on the socio-economic and spatial dynamics of China's agricultural landscape during this pivotal period.

To enable a holistic analysis, the dataset is constructed as a multi-layered, spatially-integrated resource designed to capture key links in the cotton commodity chain. The foundational layer consists of granular, county-level data on cotton acreage, yield, and the cultivation of both “native” (primarily *Gossypium arboreum*) and “foreign” (*Gossypium hirsutum*, or American upland) varieties. This is then augmented by three additional geospatial datasets documenting the locations of critical economic and scientific infrastructure: modern cotton mills with data on their industrial capacity, agricultural experiment stations, and treaty ports that served as hubs for trade and information. By providing these distinct but linkable data layers in an accessible format, the project offers a tool for untangling the complex interplay of agriculture, industry, and technology in Republican China.

Methodologically, this project engages the “spatial turn” in historical research while contributing to a recent shift in economic history that leverages granular, georeferenced data to move beyond national aggregates and analyze the local drivers of development (e.g. Bodenhamer, Corrigan, & Harris, 2010; Gregory & Geddes, 2014; Donaldson, 2018). Substantively, the dataset's structure is specifically designed to allow researchers to investigate the core *mechanisms* of agricultural modernization at the center of current scholarly debate. For instance, the multi-layered data on scientific institutions, market infrastructure, and agricultural outputs allows for a direct test of the “modernization package” hypothesis, which posits that growth is driven by a bundle of interlocking factors (Panza & Karakoç, 2021). Furthermore, the dataset's explicit tracking of native versus foreign cultivars makes it possible to precisely

estimate the productivity effects of biological innovation and to explore whether market structures incentivized quality improvements (Olmstead & Rhode, 2003; Olmstead & Rhode, 2008).

The dataset also allows for a reassessment of the factors driving agricultural change in early 20th-century China. While existing scholarship has often emphasized macroeconomic factors, market integration, and the limitations of state capacity in explaining agricultural development (Myers, 1970; Feuerwerker, 1983; Brandt, 1989; Kueh, 1995), this dataset enables a more nuanced analysis of the role of industrial firms and local scientific institutions in the spatial diffusion of technology. Recent work has increasingly used granular data to re-examine the drivers of historical development in China, focusing on everything from the role of railroads to the impact of missions (Chen, Wang, & Qi, 2022; Chen, Wang, & Yan, 2022). Our data speaks directly to this trend. The data on experiment stations, in particular, facilitate an investigation into how scientific knowledge and new agricultural practices, such as the cultivation of American cotton, spread across the Chinese countryside, contributing to debates on the role of science and technology in development (Stross, 1986; Zanasi, 2006; Stewart, 2015; Lavelle, 2020; Wu, 2024).

Beyond China, this dataset provides a vital case study for the global history of capitalism. The story of cotton, after all, is central to the making of the modern world (Riello, 2013). Historians of capitalism have detailed how state power, colonial expansion, and coercive labor regimes were marshalled to construct a global production system that funneled raw cotton from peripheries, such as the American South and colonial India, to the industrial cores of Europe (Beckert, 2014). While this powerful body of work often illuminates the dynamics between imperial cores and colonial peripheries, or the trajectories of successful early industrializers, the case of Republican China complicates these familiar narratives. Our data documents a nation that was neither a passive raw-material supplier, nor a rapidly industrializing global competitor. Instead, the Chinese industrialists and officials documented in this dataset were striving to build a sovereign, vertically-integrated textile industry within a fragmented, semi-colonial state. The county-level detail on the adoption of American cotton reveals a conscious strategy of import substitution: using foreign technology to supply domestic mills in a bid to achieve economic self-sufficiency (Zanasi 2006). This dataset, therefore, offers scholars an essential resource to de-center the classic core-periphery narrative, providing a granular view of how a major non-Western nation navigated the pressures and opportunities of the industrial age on its own terms.

In the following sections, we provide a description of the dataset's construction, including the transcription, validation, error correction, and standardization procedures employed to ensure accuracy and consistency. We also give a brief historical overview of the institutions involved

in data collection during the Republican era, including the Chinese Cotton Millowners Association (CCMA) and the Chinese Cotton Improvement Association. These institutions played an important role in the development of China's early cotton industry and the collection of vital statistics; understanding their history provides valuable context for the dataset.

2. Data Creation and Collection

This section provides an overview of the creation and collection of the four main components of this dataset: cotton production data; cotton mill location data; experiment station location data; and treaty port location data. Given that these data come from historical surveys and publications, the subsections below provide an introduction to how the data were initially collected during the 1910s to 1930s.

2.1. Cotton Production Data Production

In an effort to accelerate the expansion of American cotton production, industrialists began laying the foundations for the first national cotton information system in the early twentieth century. Their primary objective was import substitution: to improve the quality and quantity of domestic raw cotton to supply China's own burgeoning textile mills, thereby reducing the industry's dependence on costly foreign imports. In 1918, a newly founded industrial organization known as the Chinese Cotton Millowners Association (CCMA; 華商紗廠聯合會) began building an organizational framework for the production and dissemination of cotton data. The CCMA recruited a small army of data surveyors, often from the ranks of local colleges and sometimes in collaboration with local governments (Huashang shachang lianhehui, 1922, pp. 1–2; Huashang shachang lianhehui, 1923, notes; Jiang, 1929, p. 9). These surveyors carried out observations of local cotton production, which, together with cadastral surveys, were used to generate county and provincial-level cotton production statistics. Such data was viewed as pivotal to assessing and promoting the progress of cotton improvement work. The following year, the CCMA established the *China Cotton Journal* (華商紗廠聯合會季刊), a platform for disseminating this data, along with market reports and scientific research articles.

While the CCMA's industry reports initially only provided aggregate figures on acreage and overall yields, beginning in 1921, they actively differentiated between native and foreign cotton cultivars. Over time, the production and publication of cotton data became increasingly institutionalized and involved a broader array of stakeholders. During the nearly thirty years of cotton surveys, the formal organization that compiled and published production statistics changed several times as the network of individuals and institutions involved

河北省

民國二十四年棉產統計

河北省 (一)

河北省 (接前頁)

		二十四年		二十三年	
		棉田面積	皮棉產額	棉田面積	皮棉產額
正定	中	265,440	60	221,200	70,784
	洋	7,500	120	2,200	792
新樂	中	4,800	70	5,690	2,731
	洋	180	100	160	62
定縣	中	120,000	100	118,050	66,108
	洋	80,600	150	80,200	38,496
唐縣	中	35,000	70	9,800	12,000
	洋	1,980	140	380	159
博野	中	20,000	80	20,620	9,897
	洋	350	100	230	96
安國	中	13,000	120	6,240	8,296
	洋	350	140	10	4
安平	中	3,400	70	3,200	896
	洋	500	70	105	96
深澤	中	52,000	120	24,960	18,900
	洋	200	120	72	
無極	中	135,000	120	64,900	42,000
	洋	400	140	168	54
饒鹿	中	38,400	100	15,360	19,824
	洋	6,580	140	2,763	96
藁城	中	12,000	80	3,840	25,704
	洋	7,000	120	2,520	8,000
元氏	中	22,500	80	7,200	115,205
	洋	3,000	130	1,170	900
高邑	中	52,000	100	20,800	30,914
	洋	28,000	140	11,760	10,608
趙縣	中	390,000	160	249,600	390,181
	洋	8,000	180	4,320	2,700
甯晉	中	80,000	140	44,800	30,225
	洋	97,000	180	52,380	28,800
東鹿	中	350,000	120	168,000	224,000
	洋	26,000	140	10,920	14,580

		二十二年		二十一年		二十年	
		棉田面積	皮棉產額	棉田面積	皮棉產額	棉田面積	皮棉產額
	中	220,900	37,111	209,000	37,620	210,000	58,800
	洋	2,100	378	1,300	218	1,500	336
	中	5,830	665	6,800	1,387	6,000	1,680
	洋	180	25	150	25	100	25
	中	117,450	32,888	126,000	24,192	130,000	41,600
	洋	79,300	17,763	72,000	12,096	70,000	17,360
	中	27,000	5,400	39,000	8,190	18,000	4,788
	洋	250	38	400	89	500	124
	中	20,400	4,080	17,000	4,624	17,000	4,522
	洋	198	36	800	179	1,500	372
	中	14,210	1,535	44,100	12,343	14,500	3,850
	洋	20	3	60	11	1,000	248
	中	4,000	640	6,500	1,404	21,000	6,550
	洋	2,700	313	3,500	735	400	99
	中	38,000	4,560	70,000	17,150	42,000	13,441
	洋	50	6	100	21	500	135
	中	98,500	30,338	100,000	24,000	90,000	28,800
	洋	100	19	100	21	4,000	576
	中	49,500	3,960	95,000	25,840	115,000	32,200
	洋	500	45	1,000	240	3,000	744
	中	125,000	13,125	98,545	24,143	64,000	20,480
	洋	7,000	1,260	10,000	1,680	1,500	384
	中	123,500	14,079	37,200	10,044	125,000	35,000
	洋	380	68	24,800	5,555	3,000	672
	中	58,220	8,849	56,000	17,024	52,000	14,190
	洋	18,000	3,780	22,000	5,100	1,500	336
	中	210,000	30,240	280,000	68,600	210,000	67,200
	洋	2,000	380	2,000	448	18,000	4,600
	中	118,000	15,576	119,000	29,155	24,000	7,290
	洋	58,500	11,466	51,000	10,200	28,000	7,390
	中	250,000	50,000	295,000	70,800	110,000	30,030
	洋	1,200	230	5,000	1,125	2,000	510

40

41

FIGURE 1: Sample Image of Published Data (Source: Zhonghua mianye tongji hui, 1935, p. 40–41).

expanded. The annual data was published by the CCMA from 1919 to 1930, the China Cotton Production Statistics Association (中華棉業統計會) from 1931 to 1936, and the National Cotton Improvement Institute (中央棉產改進所) in 1937. Changes in who published the statistics did not lead to immediate shifts in how the data was collected. Rather, survey strategies and capacity gradually evolved over time.

The data collected are ultimately estimations, statistical models based on similar surveys conducted by the United States Department of Agriculture. Total agricultural acreage was typically collected in cadastral surveys by local government agencies, from which cotton acreage for both native and foreign cotton was estimated through on-the-ground field surveys. Once surveyors had an estimate for total cotton acreage, they could then estimate average yields through a sample survey (Jiang, 1934: 232–44). Responsibility for verifying accuracy evolved over time, but a network of provincial-level cotton improvement stations emerged in the 1920s to help manage the collection and reliability of data for their respective provinces (Huashang shachang lianhehui, 1932). The number of rural surveyors and supervisors also increased. By 1937, there were over 190,000 rural surveyors (mainly local elites recommended by local extension agencies) that were overseen by around 200 supervisors (employees

TABLE 1: Yearly Coverage and Yield Share of Pre-1926 Surveyed Counties

Year	Total Counties Surveyed	Baseline Counties (pre-1926)	Native Yield Share (%)	Foreign Yield Share (%)	Total Yield Share (%)
1919	87	87	–	–	100.0
1920	111	111	–	–	100.0
1921	153	153	100.0	100.0	100.0
1922	213	213	–	–	100.0
1923	227	227	100.0	100.0	100.0
1926	247	247	100.0	100.0	100.0
1927	250	250	100.0	100.0	100.0
1928	245	244	99.54	100.0	99.90
1929	251	214	88.55	90.25	91.52
1930	309	243	93.11	89.53	92.42
1931	324	231	92.25	93.59	93.53
1932	381	258	87.74	88.00	87.85
1933	424	267	85.32	87.22	86.31
1934	467	282	84.57	84.25	84.40
1935	465	276	89.64	83.11	86.79
1936	570	273	86.99	80.71	83.47
1937	622	285	87.76	87.46	80.83

at cotton improvement stations) (Zhongyang mianchan gaijinchi, 1937, p. 480). To ensure a degree of commensurability, surveyors were given instruction manuals that contained basic scientific knowledge of cotton, including how to recognize the difference between “native” and “foreign” cotton (Nongcun mianchan tongxin diaochayuan xuzhi, 1941). This method of estimating total production based on sampling techniques would continue to be used into the 1950s, when the Chinese Communist Party denounced mathematical statistics as a bourgeois science and sought instead to return to comprehensive and periodic reporting (Ghosh 2018).

One of the most notable changes that occurred in the survey process was an increase in the number of counties surveyed, which more than doubled from around 250 in 1926 to over 600 by 1937 (see Table 1). Constraints of funding, social instability, and regional economic significance all likely shaped which counties were surveyed. However, it should be noted that many counties that would be added in subsequent years were marginal cotton producers. Using 1926 as a baseline, the roughly 300 unique counties that had been surveyed up

to this year still accounted for over 80 percent of total output by 1937 (Table 1). Although our dataset relies on some estimated figures and is not fully comprehensive, it nevertheless provides a robust macro-level portrait of how cotton production geography evolved from the 1920s to the 1930s.

We obtained a complete run of publications containing this data from the online repository CNBKSY-Shanghai Library. We first entered the data into an Excel spreadsheet, with a separate file for each province. In the process of transcription, we identified 144 errors in the original publications which can be classified as calculation mistakes (137) and clerical errors (7) that we corrected for our analysis. County names and boundaries changed considerably during the twentieth century. To link this historical data to geographic units, we classified counties according to their administrative divisions at the time and linked them to the China Historical Geographic Information System (CHGIS) Version 6 dataset (Chinese Historical Geographic Information System, 2016). The 1911 CHGIS county layer was chosen as a stable administrative baseline to ensure spatial consistency across the entire 1919–1937 period. This was deemed necessary because the era was characterized by frequent and inconsistent administrative changes, and using a single baseline allows for reliable longitudinal analysis and data visualization. Because county boundaries shifted over time, we georeferenced each record to the latitude-longitude coordinates of its county seat (or principal city) recorded in the CHGIS shapefile. When a county seat was missing from that dataset, we manually determined its coordinates, added them to our database, and then linked the record to the matching 1911 county polygon in CHGIS to preserve spatial consistency. Users can utilize this point data to project county production data onto different spatial units as additional historical GIS shapefiles become available.

To maintain a degree of consistency for economic comparative analysis, our dataset contains an entry for every county each year, even if it wasn't included in the survey. Missing data has been flagged with the code -9999. Finally, we only include entries for provinces that were surveyed. These include the main cotton producing regions of China: Anhui, Zhili (later changed to Hebei), Henan, Hubei, Hunan, Jiangsu, Jiangxi, Shaanxi, Shandong, Shanxi, and Zhejiang. We extracted all the files into a single CSV file containing 18,967 data points tracking county-level cotton production from 1919 to 1937 that includes acreage and yield of foreign and native cotton.

It is important for users of this dataset to recognize that these statistics (like all statistics) are a product of the specific political and institutional context in which they were created. Data collection by the CCMA and its successors was profoundly shaped by the era's chronic instability. The political fragmentation of the Warlord Era (McCord, 1993) meant that surveyors often had to navigate territories controlled by mutually hostile militarists, severely hampering efforts to conduct a uniform, national survey. Moreover, increasing Japanese imperialist pressure and encroachment in key northern

cotton-producing provinces such as Shandong and Hebei throughout the late 1920s and 1930s undoubtedly compromised data collection in those critical regions (Myers, 1970; Coble, 1991). These on-the-ground realities meant that coverage could vary significantly from year to year based on factors entirely outside the surveyors' control. Furthermore, the CCMA was an organization rooted in the treaty-port industrial economy of the lower Yangzi, representing the interests of a modern, Shanghai-centered bourgeoisie (Chao, 1977; Kraus, 1980). Its institutional focus, therefore, may have inherently favored the commercialized agricultural regions that directly supplied its member mills. This could lead to the relative under-sampling or mischaracterization of more peripheral cotton-growing areas less integrated into the national textile economy.

While a critical awareness of these politically and institutional contingencies is crucial for any analysis, this dataset nonetheless represents the most comprehensive quantitative record of local-level agricultural production available for the Republican era. As such, it affords a unique and invaluable window into the on-the-ground realities of modernization during this transformative period.

2.2. Cotton Mill Dataset

The rapid expansion of cotton mills in early twentieth-century China was a defining feature of the Republican era's industrial landscape. Driven by both foreign and domestic investment, the number of mills grew dramatically, particularly in coastal cities such as Shanghai, Tianjin, and Qingdao, as well as in emerging inland industrial centers (Chao, 1977). The decisions of both foreign and domestic investors were shaped by a complex set of factors, including proximity to raw materials, access to transportation networks, and the availability of labor. These mills were primarily reliant on American long-staple cotton, creating a significant demand that spurred efforts to increase domestic production of this variety. The expansion of cotton mills during this period was a key feature of China's industrial development, with significant growth in both the number of mills and their production capacity (Fong, 1932; Feuerwerker, 1983; Chao, 1977). As such, understanding the spatial distribution of these mills is crucial to investigating regional variations in cotton production.

The dataset includes a comprehensive file on 127 modern cotton mills operating in China up to the early 1930s. This list, originally compiled from H.D. Fong's seminal *Cotton Industry and Trade in China* (1934/2011), represents the vast majority of the mechanized spinning sector of the era. While numerous smaller, traditional workshops existed, these 127 mills constitute the core of the modern industrial sector that this dataset aims to capture. To quantify the scale and capacity of these enterprises, our dataset includes key operational variables such as the number of installed spindles, the size of the

workforce, and the annual consumption of raw cotton (measured in *dan*, a traditional Chinese unit of weight, approx. 50 kg). This rich data allows for detailed analysis of the spatial concentration of industrial demand, economies of scale in production, and the relationship between mill capacity and the adoption of different cotton varieties in surrounding agricultural regions.

2.3. Experiment Station Data

The emergence of agricultural experiment stations in China was closely tied to the broader Republican-era movement for scientific modernization and national strengthening (Stross, 1986; Lavelle, 2015; Wu, 2024). These institutions were tasked with adapting foreign agricultural technologies, particularly American cotton, to local conditions and promoting their adoption among farmers. They conducted research on seed varieties, cultivation techniques, and pest control, and played a crucial role in extension work, disseminating knowledge and providing technical assistance to rural communities (Stewart, 2021; Zhang, 2025).

However, the mere presence of an experiment station did not guarantee its effectiveness. These institutions were often underfunded, poorly staffed, and struggled to translate laboratory findings into practical techniques suitable for the diverse ecologies and socio-economic conditions of the Chinese countryside. Their extension efforts often faced skepticism from farmers, and their success could be highly dependent on the support of local elites and the dedication of individual agronomists (Stross, 1986; Pomeranz, 1993; Stewart, 2021). Therefore, while these stations represented crucial nodes in a nascent network of agricultural science, their actual impact on the ground was uneven.

Our dataset consists of a curated subset of 47 stations, selected based on their documented, active involvement in cotton-specific research and extension, as evidenced in publications from the Chinese Cotton Improvement Association (CCIA; 中華棉產改進會) and other agricultural journals. While this represents a fraction of all agricultural stations of the era, it captures the key national and provincial institutions dedicated to cotton modernization.

2.4. Treaty Port and Commercial Center Data

The treaty ports, open cities, and industrial centers of Republican China were critical nexuses of global trade and domestic commerce, profoundly shaping the country's economic geography (Bergère, 1989). The precise economic impact of these centers has been a subject of long-standing debate. While earlier scholarship often emphasized their role as exploitative colonial enclaves, more recent quantitative work has highlighted their significant, and often positive, long-term effects on local economic development and structural transformation (Jia, 2014). This dataset provides a new tool to engage this debate at a specific commodity level.

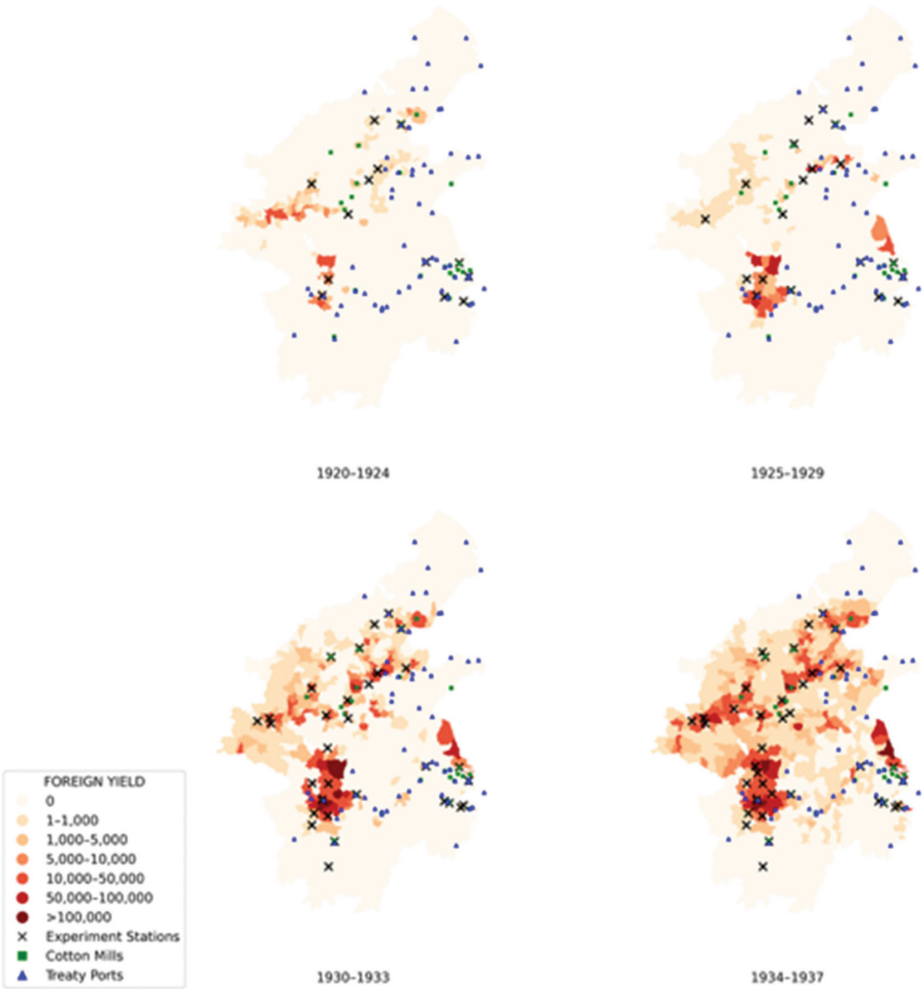


FIGURE 2: Map of Foreign Cotton Production with Treaty Ports, Cotton Stations, and Mills, 1920-1937.

To facilitate this research, we have assembled a dataset of 217 such commercial and industrial centers as compiled by Robert Nield (2015). This file provides the name, location (including coordinates), type of center (e.g. treaty port, open city, industrial center), and year of opening for each entry. This data layer enables spatial analyses that can link the commercialization of agriculture—and the adoption of American cotton in particular—to proximity to the major hubs of trade, finance, and information that were critical conduits for the flow of goods, capital, and technology in Republican China. By offering granular evidence, it allows researchers to move beyond aggregate debates and measure the on-the-ground impact of these complex nodes of the global and local economy.

3. Description of Dataset

- **Citation:** Duan, Y.; Frost, A. K.; Stewart, S. D. (2025). Cotton Production in Republican China: A Longitudinal Dataset (1919-1937). DOI: 10.4231/S460-H323
- **Temporal Coverage:** 1919-1937
- **Deposited:** Purdue University Research Repository (PURR)
- **Files:**
 - **README.md:** A codebook providing detailed descriptions of the four CSV files, including descriptions of the variables, data structure, and encoding formats.
 - **Cotton_production_data.csv:** A CSV file (UTF-8) containing 18,967 rows of annual county-level data on cotton production. Includes variables on county names, geographical coordinates, and the cultivated area and yield of native and foreign cotton varieties.
 - **Cotton_mills.csv:** A CSV file (UTF-8) with data on 127 cotton mills established prior to the 1930s. Variables include mill name, location and geographical coordinates, ownership, year of establishment, and key operational data such as number of spindles, number of workers, and annual cotton consumption and production.
 - **Cotton_stations.csv:** A CSV file (UTF-8) with 46 entries representing active cotton experiment stations during the 1920s and 1930s. Variables include station name, location and geographical coordinates, and the year of establishment.
 - **Treaty_ports.csv:** A CSV file (UTF-8) with 217 entries documenting treaty ports, open cities, and other historical commercial centers. Variables include location name, geographical coordinates, type of center (e.g., treaty port, customs station), and year of opening.

4. Concluding Remarks and Potential for Reuse

This paper has introduced a new, hand-collected dataset on cotton production in Republican China (1919–1937), a multi-layered resource designed to facilitate new quantitative and spatial research. By making this dataset publicly available, we hope to stimulate future scholarship on agricultural and industrial change. The data's true value lies in its potential for reuse, allowing researchers to move beyond broad narratives to rigorously test specific hypotheses about the mechanisms of economic development.

A primary avenue for future research is to conduct a formal test of the Olmstead and Rhode (2008) biological innovation hypothesis in a non-Western context. By modeling county-level yields as a function of the acreage share devoted to American varieties, while using the geospatial data on experiment stations and cotton mills to control for institutional support and market pull,

researchers could better understand the productivity gains from new technology. Such an analysis would allow for a nuanced investigation of the “modernization package” debate, testing whether the interlocking factors of technology, institutions, and credit, identified by Panza & Karakoç (2021) in Egypt, had similar synergistic effects in the Chinese countryside.

Furthermore, the dataset provides a crucial tool for comparative and global history. By analyzing the spatial relationship between treaty ports, mill locations, and the adoption of American cotton, scholars can investigate the specific channels through which global economic forces influenced local agricultural practices. This enables granular, sub-national comparisons with colonial contexts, comparing China’s developmental path with both extractive systems (Roberts, 1996; Dell & Olken, 2020) and constrained imperial projects (De Haas, 2021). Such work would help to clarify the unique features of China’s state-led, nationally-oriented path to industrialization within the broader ‘empire of cotton.’

Finally, the dataset’s georeferenced structure invites new research at the intersection of economic and environmental history. By merging our production data with historical climate, soil, and hydrological datasets, scholars could explore the role of ecological constraints in either facilitating or hindering the spread of new crop varieties. This would offer a powerful way to test the arguments of De Haas (2021) regarding the importance of the physical environment in shaping the success of top-down agricultural initiatives, explaining why the seeds of modernization took root in some Chinese counties but failed in others. We encourage researchers to build upon this dataset, and we look forward to the new questions and answers it will help unearth.

Acknowledgements

Data transcription was made possible by an Exploratory Data and Travel Grant from the Economic History Association.

References

- Beckert, S. (2014). *Empire of cotton: A global history*. Alfred A. Knopf.
- Bergère, M. C. (1989). *The golden age of the Chinese bourgeoisie, 1911–1937* (J. Lloyd, Trans.). Cambridge University Press.
- Bodenhamer, D. J., Corrigan, J., & Harris, T. M. (Eds.). (2010). *The spatial humanities: GIS and the future of humanities scholarship*. Indiana University Press.
- Brandt, L. (1989). *Commercialization and agricultural development: Central and eastern China, 1870–1937*. Cambridge University Press.
- Chao, K. (1977). *The development of cotton textile production in China*. East Asian Research Center, Harvard University.
- Chen, Y., Wang, H., & Yan, S. (2022). The long-term effects of Protestant activities in China. *Journal of Comparative Economics*, 50(2), 394–414.

- Chen, T., Wang, J., & Qi, H. (2022). Railways, telegraph and technology adoption: The introduction of American cotton in early 20th century China. Unpublished manuscript available on SSRN 4868857.
- Chinese Historical Geographic Information System (CHGIS). (2016). *1911 layers UTF8 encoding (V1)*. Harvard Dataverse. <https://doi.org/10.7910/DVN/HHVVHX>
- Coble, P. M. (1991). *Facing Japan: Chinese Politics and Japanese Imperialism, 1931–1937*. Council on East Asian Studies, Harvard University.
- De Haas, M. (2021). The failure of cotton imperialism in Africa: seasonal constraints and contrasting outcomes in French West Africa and British Uganda. *The Journal of Economic History*, 81(4), 1098–1136.
- Dell, M., & Olken, B. A. (2020). The development effects of the extractive colonial economy: The dutch cultivation system in java. *The Review of Economic Studies*, 87(1), 164–203.
- Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4), 899–934.
- Duara, P. (1988). *Culture, power, and the state: Rural North China, 1900–1942*. Stanford University Press.
- Eastman, L. E. (1983). Nationalist China during the Nanking Decade, 1927–1937. In J. K. Fairbank & A. Feuerwerker (Eds.), *The Cambridge history of China: Republican China 1912–1949* (Part 1, Vol. 12, pp. 116–167). Cambridge University Press.
- Fang, X. [H. D. Fong]. (1934/2011). *Zhongguo zhi mianfangzhi ye* [Cotton industry and trade in China] (Reprint ed.). Zhonghua shuju.
- Feuerwerker, A. (1983). Economic trends, 1912–49. In J. K. Fairbank (Ed.), *The Cambridge history of China: Republican China 1912–1949* (Part 1, Vol. 12, pp. 28–128). Cambridge University Press.
- Fong, H. D. (1932). Cotton industry and trade in China. *Chinese Social and Political Science Review*, 16(1).
- Ghosh, A. (2018). Lies, damned lies, and (bourgeois) statistics: Ascertaining social fact in midcentury China and the Soviet Union. *Osiris*, 33, 149–168.
- Gregory, I. N., & Geddes, A. (Eds.). (2014). *Toward spatial humanities: Historical GIS and spatial history*. Indiana University Press.
- Huashang shachang lianhehui [Chinese Cotton Millowners Association]. (1922). *Mianchan diaocha baogao* [Report on cotton production].
- Huashang shachang lianhehui [Chinese Cotton Millowners Association]. (1923). *Mianchan diaocha baogao* [Report on cotton production].
- Huashang shachang lianhehui [Chinese Cotton Millowners Association]. (1932). *Mianchan diaocha baogao* [Report on cotton production].
- Jia, R. (2014). The legacies of forced freedom: China's treaty ports. *The Review of Economics and Statistics*, 96(4), 596–608.
- Jiang, D. (1929). Zhongguo mianchan tongji zhi guoqu ji jianglai [Past and future of China's cotton production statistics]. *Zhongguo Mianchan Tongji*, (18), 1–42.
- Jiang, D. (1934). Mianchan tongji zhi bianzhi ji qi yingyong [The compilation and application of cotton production statistics]. *Zhongguo Mianchan Tongji*, (6–7), 232–244.
- Köll, E. (2003). *From cotton mill to business empire: The emergence of regional enterprises in modern China*. Harvard University Asia Center.
- Kraus, R. A. (1980). *Cotton and cotton goods in China 1918–1936*. Garland Publishing, Inc.
- Kueh, Y. Y. (1995). *Agricultural instability in China, 1931–1990: Weather, technology, and institutions*. Oxford University Press.

- Lavelle, P. (2015). Agricultural improvement at China's first agricultural experiment stations. In *New perspectives on the history of life sciences and agriculture* (pp. 323–344). Springer International Publishing.
- Lavelle, P. B. (2020). Tools for overcoming crisis: Agriculture, scarcity, and ideas of rural mechanization in late Qing China. *Agricultural History*, 94(3), 386–412.
- Lin, J. (2015). Sowing seeds and knowledge: Agricultural development in Taiwan and the world, 1925–1975. *East Asian Science, Technology and Society: An International Journal*, 9(2), 127–149.
- McCord, E. A. (1993). *The Power of the Gun: The Emergence of Modern Chinese Warlordism*. University of California Press.
- Myers, R. H. (1970). *The Chinese peasant economy: Agricultural development in Hopei and Shantung, 1890–1949*. Harvard University Press.
- Nield, R. (2015). *China's foreign places: The foreign presence in China in the treaty port era, 1840–1943*. Hong Kong University Press.
- Nongcun mianchan tongxin diaochayuan xuzhi [Instructions for rural cotton production correspondents]. (1941). Huabei mianchan gaijinhui diaochahe.
- Olmstead, A. L., & Rhode, P. W. (2003). Hog-Round Marketing, Seed Quality, and Government Policy: Institutional Change in US Cotton Production, 1920–1960. *The Journal of Economic History*, 63(2), 447–488.
- Olmstead, A. L., & Rhode, P. W. (2008). Biological innovation and productivity growth in the antebellum cotton economy. *The Journal of Economic History*, 68(4), 1123–1171.
- Panza, L., & Karakoç, U. (2021). Overcoming the Egyptian cotton crisis in the interwar period: the role of irrigation, drainage, new seeds, and access to credit. *The Economic History Review*, 74(1), 60–86.
- Pomeranz, K. (1993). *The making of a hinterland: State, society, and economy in inland North China, 1853–1937*. University of California Press.
- Riello, G. (2013). *Cotton: The fabric that made the modern world*. Cambridge University Press.
- Roberts, R. L. (1996). *Two worlds of cotton: Colonialism and the regional economy in the French Soudan, 1800–1946*. Stanford University Press.
- Schmalzer, S. (2019). *Red revolution, green revolution: Scientific farming in socialist China*. University of Chicago Press.
- Shiroyama, T. (2008). *China during the great depression: Market, state, and the world economy, 1929–1937*. Harvard University Asia Center.
- Stewart, S. (2021). A community crop: Cotton, science, and extension in interwar Shandong, 1918–1937. *Twentieth-Century China*, 46(2), 199–219.
- Stross, R. E. (1986). *The stubborn earth: American agriculturalists on Chinese soil, 1898–1937*. University of California Press.
- Wu, S. (2024). Planting the seeds of the nation: Agricultural experimental stations and the making of the Chinese geo-body. In E. Kaske & E. Köll (Eds.), *Age of exploration: How Chinese scientists and administrators discovered China* (pp. 195–216). De Gruyter Oldenbourg.
- Zanasi, M. (2006). *Saving the nation: Economic modernity in Republican China*. University of Chicago Press.
- Zhang, R. (2025). *A study of China–US transnational agricultural history, 1925–1949*. Chung Hwa Book Co.
- Zhonghua mianye tongji hui [China Cotton Production Statistics Association]. (1928–1936, 1946). *Zhongguo mianchan tongji* [Statistics on China's cotton production] (Nos. 17–25, 35).
- Zhongyang mianchan gaijinchu [National Cotton Improvement Institute]. (1937). Quanguo mianchan diaocha tongji banfa (Fubiao) [National cotton production survey statistical methods (Appendix table)]. *Nongye Jianshe*, (4), 479–485.