Received: 29 August 2019 | Revised: 16 March 2020 | Accepted: 27 March 2020

DOI: 10.1002/gdj3.91

DATA PAPER

Geoscience Data Journal RMetS WILEY

Temperature, precipitation and sunshine across China, 1912-51: A new daily instrumental dataset

Ivan P.L. Png[1] | Yehning Chen[2] | Junhong Chu[1] | Yikang Feng[1] | Elaine Kuan-Hui Lin[3] | Wan-ling Tseng[4]

[1]National University of Singapore, Singapore, Singapore
[2]National Taiwan University, Taipei, Taiwan
[3]National Taiwan Normal University, Taipei, Taiwan
[4]Academia Sinica, Taipei, Taiwan

Correspondence
Ivan P.L. Png, National University of Singapore
Email: iplpng@gmail.com

Funding information
Ministry of Education - Singapore, Grant/Award Number: 2013-T2-2-045

Abstract
This project compiled 463,530 instrumental observations of daily temperature, precipitation and sunshine at up to 319 stations in China between 1912 and 1951. The principal sources are monthly reports of the Institute of Meteorology, Nanjing, observatories in Japanese-occupied Manchuria, and the Japanese Army in North China. Gross errors, rounding errors and inhomogeneities are identified. The new dataset is temporally and spatially consistent with existing datasets of monthly temperature, but reports higher precipitation and sunshine.

KEYWORDS
China, data rescue, meteorology

1 | INTRODUCTION

Historical meteorological data at high frequency are important for the study of climate variability, particularly analysis of extreme weather and as input into modelling of reanalyses of changes (Jones and Mann, 2004). Such data are also crucial for understanding the social and economic consequences of weather shocks (Allan *et al.*, 2016). For instance, prediction of agricultural yields requires information on daily temperatures (Ritchie and NeSmith, 1991). To fill gaps in weather records, scholars worldwide are collaborating to recover and digitize information from archival documents that are at risk of permanent loss or deterioration (Brunet and Jones, 2011; Brönnimann *et al.*, 2018).

Despite China's importance, high-frequency data of weather during the first half of the 20th century are limited. The Chinese Academy of Sciences published datasets of instrumental records at 65 and 205 stations at

We gratefully acknowledge superb assistance by Xi Xiong, Yige Duan, Qin Chao, Jin Wang, Jin Miao and many other dedicated research assistants, and institutional support from NUS Libraries. This project was funded by the Singapore Ministry of Education Academic Research Fund grant 2013-T2-2-045.

Dataset

Identifier: https://doi.org/10.5281/zenodo.3380867

Creator: Png I. P. L., Chen Y., Chu J. and Feng Y.

Title: NUS Republican China Weather Database

Publisher: Zenodo

Publication year: 2020

Resource type: Dataset

monthly frequency (Tao *et al.*, 1997). In addition, the China Meteorological Administration is compiling sub-daily data at 19 stations (Williamson *et al.*, 2017). The REACHES project at Academia Sinica, Taiwan, is compiling a database of East Asian climate records from Chinese historical documents (Wang *et al.*, 2018).

This project contributes to the data rescue effort for China with an online repository of historical weather records, and a new dataset of 463,530 daily instrumental records of temperature, precipitation and sunshine at up to 319 stations between 1912 and 1951. The data were compiled from various historical sources, principally, the Institute of Meteorology at Nanjing, observatories in Japanese-occupied Manchuria, and the Japanese Army in North China.

2 | SOURCES

Referring to Table 1, the new dataset was compiled from various historical sources. The largest source, accounting for over 40 per cent of the data, is the Monthly Meteorological Bulletin, published by the Institute of Meteorology of the National Research Institute (now called the Academia Sinica) at Nanjing, then capital of China.

The January 1929 issue published weather reports from 44 stations of the Chinese Maritime Customs Service (which established China's first nationwide meteorological network (Bickers, 2016)) and 10 other stations. The Bulletin published various information on a daily basis, including temperature (mean, minimum and maximum), barometric pressure, absolute and relative humidity, cloud cover, duration of sunshine, precipitation, and wind direction and force, with the coverage differing by station. The geographical coverage of the Bulletin expanded progressively to a peak of 89 stations in January-November 1936. The November 1936 issue is the last that this project has been able to locate.[1] Figure 1 (left) presents a cover of the Bulletin.

Until this project, partial collections of the Bulletin were publicly available, at the China National Library and various university libraries within mainland China, Central Weather Bureau, Taipei, Taiwan, and University of Pennsylvania library. To assist in data rescue, the National University of Singapore Libraries purchased and published an almost complete collection of the Bulletin (Chen *et al.*, 2017).

The next largest source, accounting for around 11 per cent of the data, is the monthly report published by the Manchuria Central Observatory at present-day Changchun from 1935 until 1941, covering 55 stations across Manchuria.[2] Partial collections of Manchuria weather reports are dispersed across the Hiroshima University Library, Japan Center for Asian Historical Records, and the Japan Meteorological Agency.

The third largest source of data is monthly reports of the North China Meteorological Section of the Japanese Army (see Figure 1, right).[3] The first report this project has found reported daily weather in January 1938 at 4 stations across the present-day provinces of Inner Mongolia, Shanxi and Hebei. The hand-written document reports average pressure, temperature (mean, maximum and minimum), wind (average velocity and most frequent direction), humidity, cloud, precipitation and visibility. The issue for March 1944 is the last found. The U.S. Library of Congress holds the largest collections of the reports, with smaller holdings at the Japan Meteorological Agency and Central Weather Bureau, Taipei, Taiwan.

In total, the project compiled weather reports from 64 sources stored at 36 libraries and online archives. The copyright of these publications, if any, has expired. For the convenience of future researchers, the National University of Singapore Libraries published much of the corpus on its portal (URL https://doi.org/10.25541/2nb3-3x47).[4] The records are listed by geographical region and chronologically within region, and the dataset is organized for search through a map.

3 | DATASET

This project aims to compile a dataset of temperature, precipitation and sunshine in China between 1912 and 1951, from historical records.[5] Historical observations that were not recorded according to modern standards and techniques require quality control and homogenization (Brázdil *et al.*, 2010). Due to the quantity and heterogeneity of the data, these processes required considerable manual effort.

Generally, discrepancies and inhomogeneities arise for two reasons. One is peculiarities or changes in the observation procedure, personnel, instruments or monitoring station. The other is typographical errors in the original records or mistakes in the digitization process. This project applied various procedures to identify discrepancies and inhomogeneities, which were noted, with the corresponding observations

[1]Apparently, the China Meteorological Administration archive holds issues of the Bulletin up to January 1937 (Zhongguo Qixiangju Qixiang Danganguan, 2003, p. 34), but the project was not able to access them.

[2]These start and end dates are based on the monthly reports found by this project.

[3]Kobayashi and Yamamoto, 2013, Table 5, detail the various names of the Unit.

[4]This online corpus does not include the bulletins of the North China Meteorological Section of the Japanese Army, which the U.S. Library of Congress will scan and publish.

[5]With the limited available resources, the project focused on temperature, precipitation and sunshine, leaving other parameters for future work.

TABLE 1 Sources

Source title	Compilers	Temporal coverage	Geographical coverage	Stations	Records	Percent	Cumulative percent
Monthly Meteorological Bulletin (Sinica)	China Maritime Customs Service, provincial, and university observatories	1928–1936 (Nov)	Across China	99	191,919	41.4	41.4
Manchuria Meteorological Monthly Bulletin	Manchuria Central Observatory	1935–1940 (Nov), 1941 (Jan–Mar)	Northeast China, Hebei, Inner Mongolia	55	52,397	11.3	52.7
North China Meteorological Report	Japanese army units	1938, 1939 (Feb, Apr, Oct), 1940–1944 (Mar)	Beijing, Hebei, Henan, Inner Mongolia, Jiangsu, Shandong, Shanxi, Tianjin	26	37,483	8.1	60.8
Observatoire de Zi-Ka-Wei Revue Mensuelle	Society of Jesus, Kiangnan Mission, Shanghai	1912–1943 (Mar)	Shanghai	3	26,731	5.8	66.6
Kwantung Meteorological Monthly Bulletin	Kwantung Observatory	1933–1937 (Mar), 1937 (Jul–Dec), 1944–1945 (Feb)	Northeast China, Inner Mongolia	23	26,136	5.6	72.2
Meteorological Report in Manchuria	Dalian Observatory	1924 (Jan), 1925–1928 (Nov), 1929 (Feb–Apr, Jun–Aug, Nov–Dec) 1930 (Jan–Mar, May–Aug, Nov–Dec), 1931 (Jan–Feb, Apr–Dec), 1932, 1937 (Apr–Jun), 1938	Northeast China	12	23,237	5.0	77.2
Fukien Meteorological Monthly Bulletin	Fujian Observatory	1936–1942 (Jul)	Fujian	18	19,687	4.3	81.5
Szechwan Meteorological Monthly Bulletin	National Sichuan University Faculty of Science	1936 (Dec)–1938 (Feb), 1938 (Oct–Dec), 1939 (Jan–Sep), 1940–1941 (Apr), 1941 (Jun–Aug), 1942 (Jan–Feb, Apr–May, Sep–Oct)	Chongqing, Sichuan	22	13,936	3.0	84.5
	Sichuan Observatory	1941–1942, 1948 (Jan–Jun)					
	Sichuan Agricultural Department	1947 (Mar–Apr)					

(Continues)

TABLE 1 (Continued)

Source title	Compilers	Temporal coverage	Geographical coverage	Stations	Records	Percent	Cumulative percent
Beijing Meteorological Data 1841-1980	Beijing Meteorological Bureau	1915 (Apr)–1917 (Nov), 1918–1926 (Nov), 1927 (Jul–Dec), 1928 (Jan–Apr), 1929 (Jul), 1930 (Jan, Mar–Dec), 1931- 1936 (Aug), 1940–1949	Beijing	1	10,565	2.3	86.8
U.S. National Oceanic and Atmospheric Administration (NOAA)	Unknown	1943 (Jun)–1948 (May)	Beijing, Chongqing, Gansu, Guangxi, Guizhou, Hunan, Sichuan, Yunnan	13	5,350	1.2	87.9
Others	Regional, provincial, municipal, or university observatories	1915 (Jun)–1951 (Nov)	Across China	142	56,089	12.1	100

retained. Table 1 reports summary statistics of the errors and inhomogeneities.

3.1 | Gross errors

Two checks were applied to detect gross errors. One checked whether the daily maximum temperature exceeded the minimum. Referring to Table 2, 847 observations failed this condition and were tagged accordingly. The other check compared each observation in a series to the observation for the previous day. If the difference between the observations exceeded two standard deviations of the monthly series, the observation was tagged as a jump. Referring to Table 2, the proportion of jumps ranged from 3 to 4 per cent in minimum and maximum temperature to about 10.5 per cent in precipitation, and about 13.5 per cent in sunshine hours.

3.2 | Rounding

Figure 2 (left) depicts the distributions of the first decimal of the daily minimum and maximum temperatures. Evidently, there are excessive zeros and fives. The main reason for excessive zeros and fives is rounding, due to either recording practices or conversion from Fahrenheit scale to Celsius and re-rounding. The outcome of re-rounding varies according to the precision of the Fahrenheit scale, whether 0.1°F, 1.0°F or 2.0°F.

Rhines *et al.* (2015) propose a precision-decoding algorithm which models the unobserved precision states of each observation in a time series as latent variables, which are then inferred using a Hidden Markov Model. The algorithm decodes globally the time sequence and estimates the most probable states given all of the observations for each station. The original distribution of the first decimal is restored with the application of jitters by generating uniform random values consistent with the precision level and then adding to the sample before rounding to the first decimal.

The present temperature dataset does not seem to exhibit rounding from an original precision of 1.0°F. Such data would exhibit relatively *few* fives (Rhines *et al.*, 2015), which contradicts the pattern of excessive fives depicted in Figure 2 (left). Nevertheless, for completeness, the Rhines *et al.* (2015) procedure was applied with 0.1°F, 1.0°F and 2.0°F precision.

Referring to Table 2, around 22 per cent of minimum and maximum temperatures were identified as being rounded. Figure 2 (right) depicts the distributions of the first decimal of the daily minimum and maximum temperatures after correction for rounding. Evidently, the first decimals are more evenly distributed, consistent with the original data being rounded.

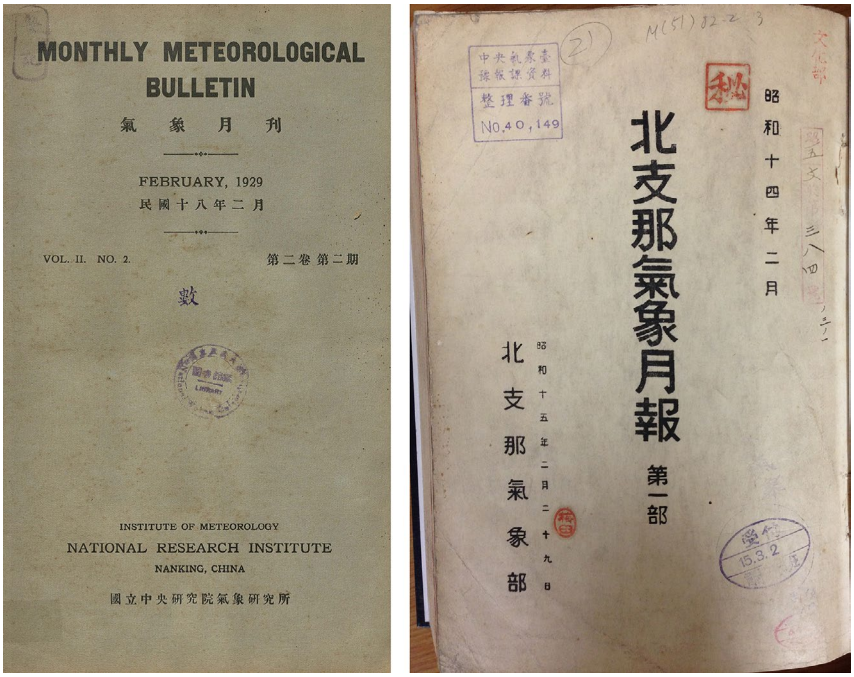


FIGURE 1 Left: Monthly Meteorological Bulletin (Institute of Meteorology, Nanjing), February 1929. Right: North China Meteorological Monthly (Japanese Army), February 1939

In principle, the Rhines *et al.* (2015) method could be applied to correct for possible rounding in records of precipitation and sunshine. However, their code is written to check records of temperature, specifically looking for conversion from Fahrenheit to Celsius. Accordingly, we did not check for rounding in precipitation or sunshine.

3.3 | Inhomogeneities

A standard procedure in climatology is to check time series for inhomogeneities that indicate undocumented change points, which are possibly due to changes in observation procedure, personnel, instruments or station. Figure 3 depicts the frequency of changes of observing organization and Table S1 reports the changes by station. Figure 4 depicts the frequency of movement of station as identified by a change in geo-coordinate or elevation in consecutive issues of the source. Table S2 reports the changes by station. On average, 8.0 per cent of stations changed observer and 6.0 per cent moved. For completeness, Table S3 lists stations whose names were changed, of which there were 13.

Wang (2008a, 2008b) proposes the penalized maximal *F* test (PMFT) to detect change points, or undocumented mean shifts that are *not* associated with any sudden change in the linear trend of a time series. Wang and Feng (2013) implemented the PMFT in the statistical software, R, and published the code as the package, RHtestsV4, to process a single time series. However, the present project involves four series (minimum and maximum temperatures, precipitation and sunshine hours) at up to 319 stations. To simplify the test for inhomogeneity, the RHtestsV4 code was adapted to process multiple time series.

In constructing a monthly dataset, Ashcroft *et al.* (2014) checked inhomogeneities in series of at least 3 years. The present data are daily, and for a similar degree of statistical power, the adapted RHtestsV4 code was applied to series of at least 60 consecutive days. The output of the adapted code was validated with the original RHtestsV4 package for minimum and maximum temperatures at Abaga.

Referring to Table 2, application of the PMFT method to the records of minimum temperature, maximum temperature, precipitation and sunshine yielded less than 0.05 per cent inhomogeneity. Apparently, inhomogeneities were very rare.

4 | VALIDATION

To validate the new dataset, it was compared with three existing datasets by temporal and spatial variation. Two are instrumental datasets (Tao *et al.*, 1997), one of 65 stations (S65),[6] which includes monthly air temperature, precipitation and sunshine duration (SSD), and another of 205 stations (S205), which includes monthly mean temperature and total precipitation. The third dataset is the Climatic Research Unit Time Series (CRU TS) v. 4.03 (Harris *et al.*, 2014), from which daily maximum and minimum temperatures and monthly precipitation were retrieved. S65 and S205 are listed in the World Meteorological Organization historical record, and so, may have been assimilated into the CRU TS.

[6]The original dataset of 60 stations was augmented to 65 stations in 1996 by the Chinese Academy of Sciences.

TABLE 2 Discrepancies and Inhomogeneities

	Total observations	Number of discrepancies	Per cent
A. Minimum temperature			
Minimum greater than maximum	462,498	847	0.18
Jumps	461,422	14,328	3.11
Rounding	462,498	102,867	22.24
Inhomogeneities (break points)	453,888	149	0.03
B. Maximum temperature			
Minimum greater than maximum	462,620	847	0.18
Jumps	461,535	18,263	3.96
Rounding	462,620	102,178	22.09
Inhomogeneities (break points)	454,014	98	0.02
C. Precipitation			
Jumps	199,457	21,028	10.54
Inhomogeneities (break points)	197,696	14	0.01
D. Sunshine hours			
Jumps	183,566	24,743	13.48
Inhomogeneities (break points)	182,471	27	0.01

Notes: Discrepancies for (A) minimum temperature, (B) maximum temperature, (C) precipitation and (D) sunshine hours, comprising minimum greater than maximum, jumps, rounding and inhomogeneities. 'Jump' indicates the difference between consecutive daily observation series exceeding two standard deviations of the monthly series; 'rounding' was detected by precision-decoding algorithm; and inhomogeneities were detected by penalized maximal *F* test.

For completeness and to avoid obvious errors, the dataset subject to validation excluded records with gross errors in temperature (daily minimum temperature exceeding maximum), or sunshine hours less than 0.5 (21.4% of the total) owing to the large number of records with SSD = 0, or missing geo-coordinates.

Referring to Figure 5a,b, both maximum (Tmax) and minimum (Tmin) temperatures in the new dataset correspond

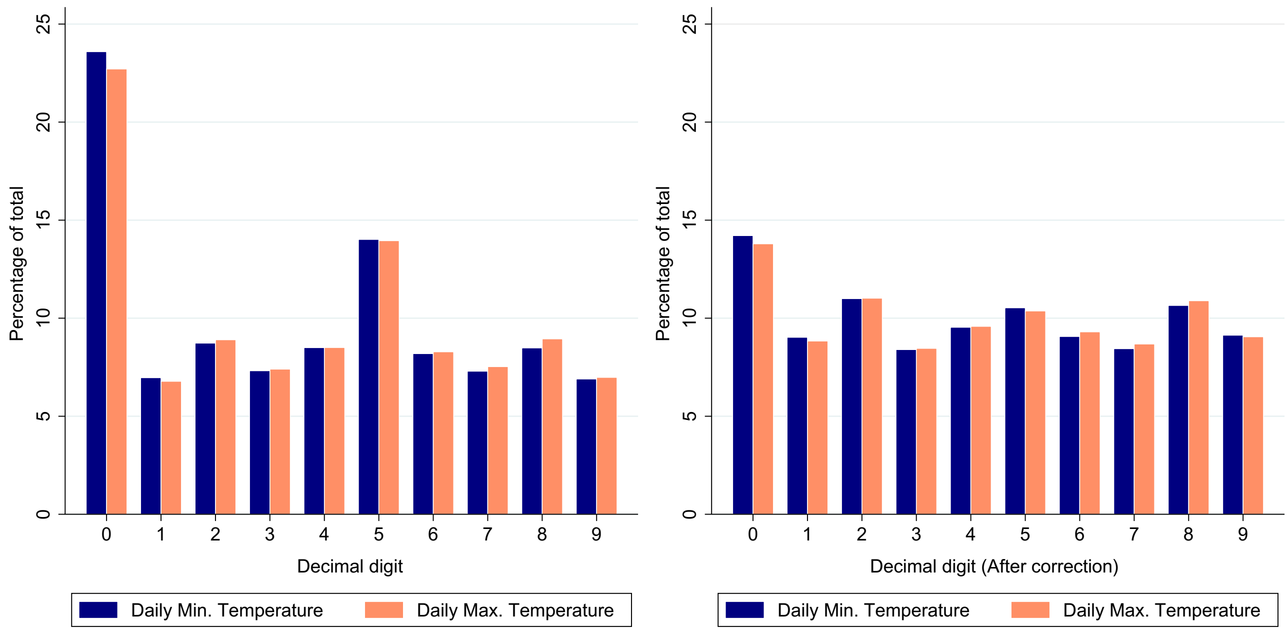


FIGURE 2 Distribution of first decimal of minimum and maximum temperatures. Left: before correction, for both minimum and maximum temperatures, there are excessive zeros and fives. Right: after correction with Hidden Markov Model, the first decimals are more evenly distributed, consistent with the original data being subject to rounding

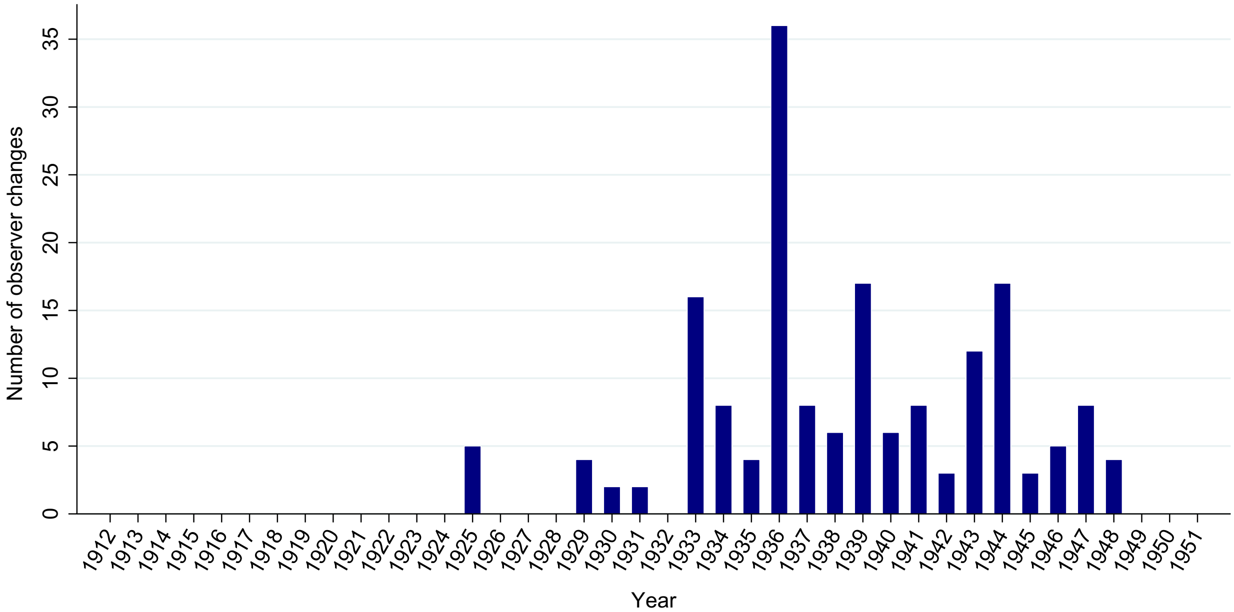


FIGURE 3 Observer changes. Observer change is recorded at the level of observing organization

well to S65 and CRU TS in terms of amplitude and phase. However, referring to Figure 5c, the seasonal cycle of precipitation in the new dataset, while exhibiting a similar shape, seems to over-estimate except in July. Sunshine hours follow a similar seasonal cycle but range between 1.5 and 4 hr higher than S65 (Figure 5d). The disparity might be partly due to the S65 sunshine data being geographically smoothed.

Next, the datasets were compared by region—northern, middle and southern China, and the Sichuan basin, which has a distinct climatology. Figure 6 depicts the daily anomaly temperature, defined as the difference of the daily mean of Tmax and Tmin and the monthly mean, over the climatology period from 1912 to 1951 in CRU TS and the new dataset. The higher frequency of data (indicated by the grey lines) between 1930 and 1945 likely suggests a higher reliability of analysis during this period. Apparently, the new dataset exhibits stronger monthly variation compared to the CRU TS, while the trajectories of the two present a certain degree of similarity over the period in all regions.

Figure 7 depicts the spatial pattern of the datasets. To better compare with CRU TS, the daily mean temperatures of the new dataset and monthly temperatures of S65

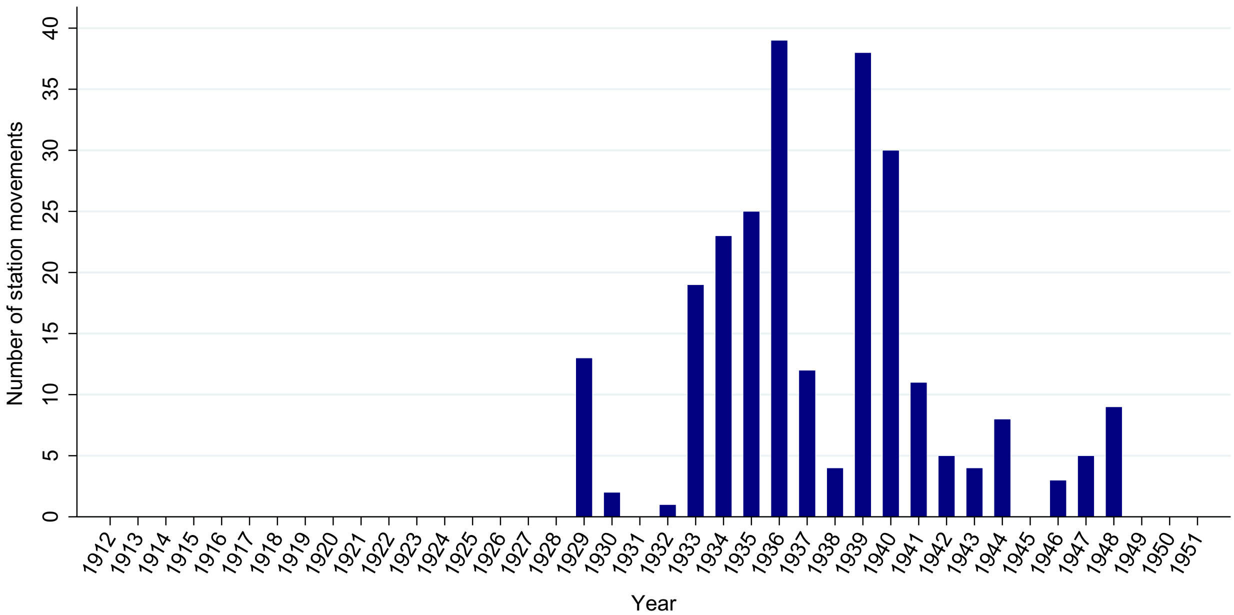
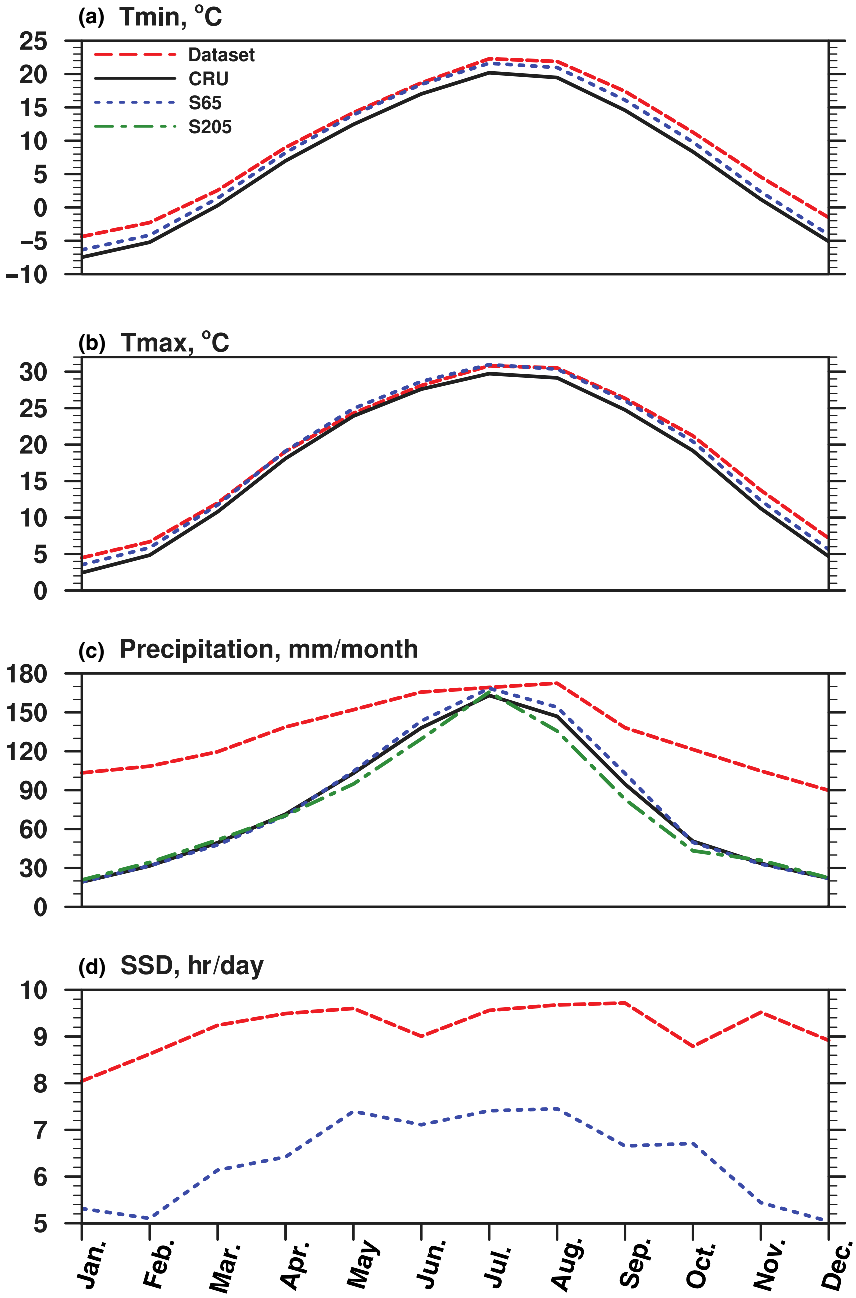


FIGURE 4 Station movements. Movement refers to a recorded change in either latitude, longitude or elevation in consecutive issues of the source. Information on latitude, longitude and elevation is extracted from available digitized primary source files. Any record for which the primary source file does not include the information is excluded from comparison

FIGURE 5 Averaged seasonal cycle over all records in (a) Tmax, unit: °C, (b) Tmin, unit: °C, (c) precipitation, unit: mm/month, and (d) sunshine, unit: hour/day. Black solid line represents CRU TS, blue and green lines represent S65 and S205 (Tao *et al.*, 1997), and red dashed line represents the new dataset



and S205 were interpolated to 0.5 × 0.5 degree grids. The spatial temperature distribution displays strong agreement among the datasets, although the high similarity of the new dataset with S65 and S205 is due in part to interpolation. Nevertheless, the general pattern is similar, with higher temperatures in the south and lower in the north, and the Sichuan Basin being warmer.

Figure 8 presents dot plots of annual precipitation and sunshine hours. Precipitation in northern China is clearly higher than in the CRU TS and S65 and S205 datasets. Sunshine hours are higher overall than S65.

Overall, the new dataset is similar to the S65, S205 and CRU TS datasets in temporal and spatial characteristics of temperature, which suggests its reliability. Yet, the discrepancies between the new dataset and the S65, S205 and CRU TS datasets in precipitation and sunshine suggest that these data should be used with caution.

5 | DISCUSSION

While the past few decades have seen considerable progress in our scientific understanding and assessment of climate change, numerical models of climate simulation and reanalysis still heavily rely on high-resolution quantitative climate data for multiple purposes including training models, cross-check and verification (Brönnimann *et al.*, 2018). Analysis of extreme events, and their behaviour and impact,

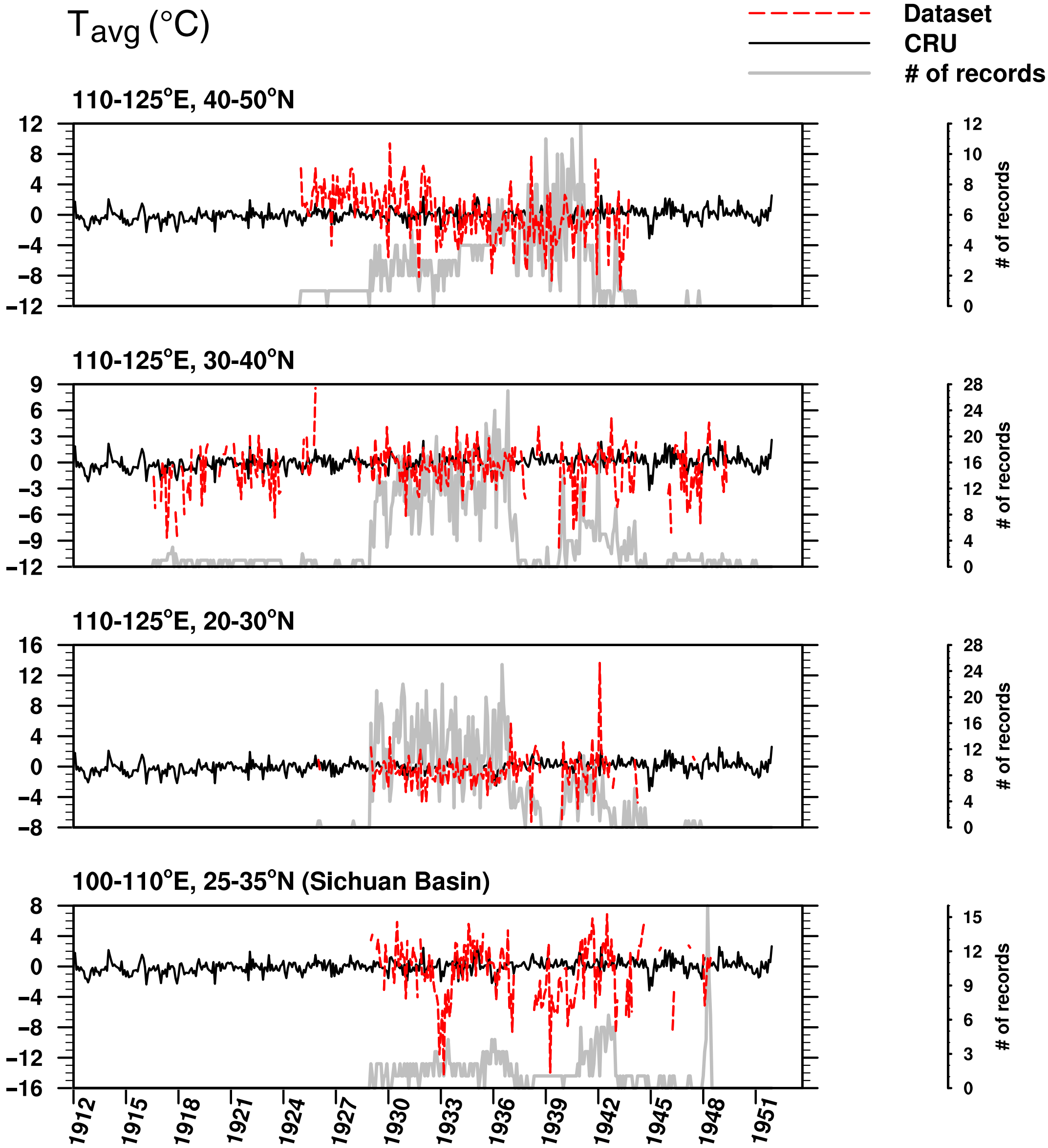


FIGURE 6 Daily mean temperature (unit: °C) of monthly anomaly by region. Black solid line represents CRU TS, and red dashed line represents the new dataset. Grey lines represent the number of records in each time interval

to support future climate risk analysis depends on the availability of more historical data. This motivates further work on historical weather and climate data rescue projects, which have gained recent attention through substantive international efforts (Allan *et al.*, 2011). Data rescue mainly refers to recovery of non-digitized records which pre-date the era of electronic recording. Data of the late nineteenth century and early twentieth century are the most integral for extending our existing instrumental data series to improve and extend reanalysis for a better understanding of past climate variability and change.

Although some long and important terrestrial data series, particularly covering Europe, have been digitized and processed for academic and practical uses, work on a large proportion remains outstanding (Brönnimann *et al.*, 2018). Data rescue in East Asia started at least a decade ago with Japan, and the associated work for mainland China is proceeding in collaboration with ACRE (Atmospheric Circulation Reconstructions over the Earth) and the wider CSSP (Climate Science for Service Partnership) (Williamson *et al.*, 2018).

Contributing to these efforts, this project constructed a dataset of daily temperature, precipitation and sunshine in China between 1912 and 1951. The dataset of 463,530 daily instrumental records of up to 319 stations adds substantially to work by others such as the China Meteorological Administration (Williamson *et al.*, 2017). Figure 9 depicts the geographical coverage of the dataset at its peak in the year 1936, covering 145 stations across 29 present-day provinces of China. As Figure 10 illustrates, the temporal coverage of the dataset varies by year while the functional

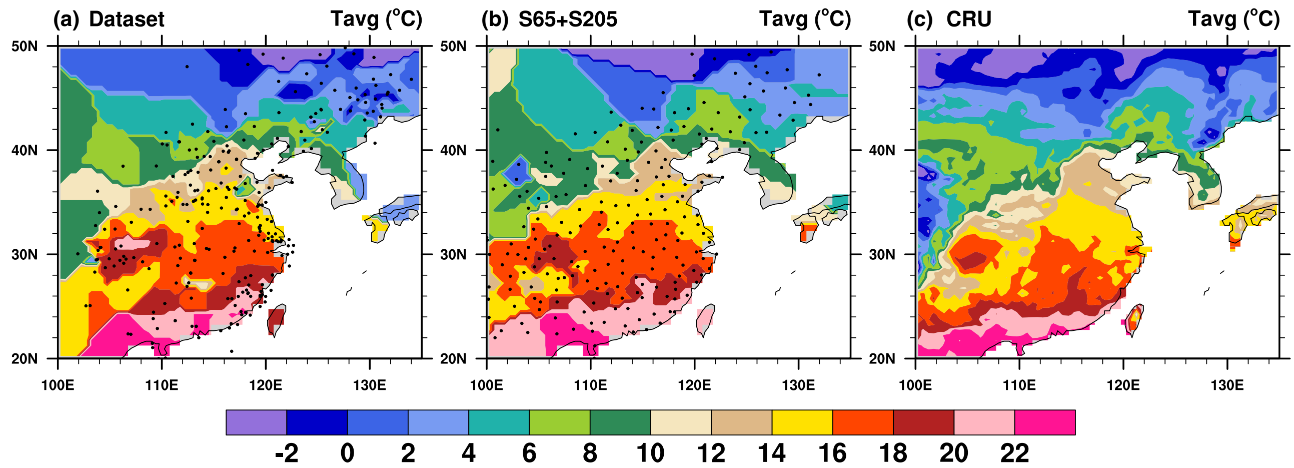


FIGURE 7 Climatology of daily mean temperature (unit: °C) in (a) Dataset, (b) S65 and S205 and (c) CRU. Black dots in the panel mark recording stations

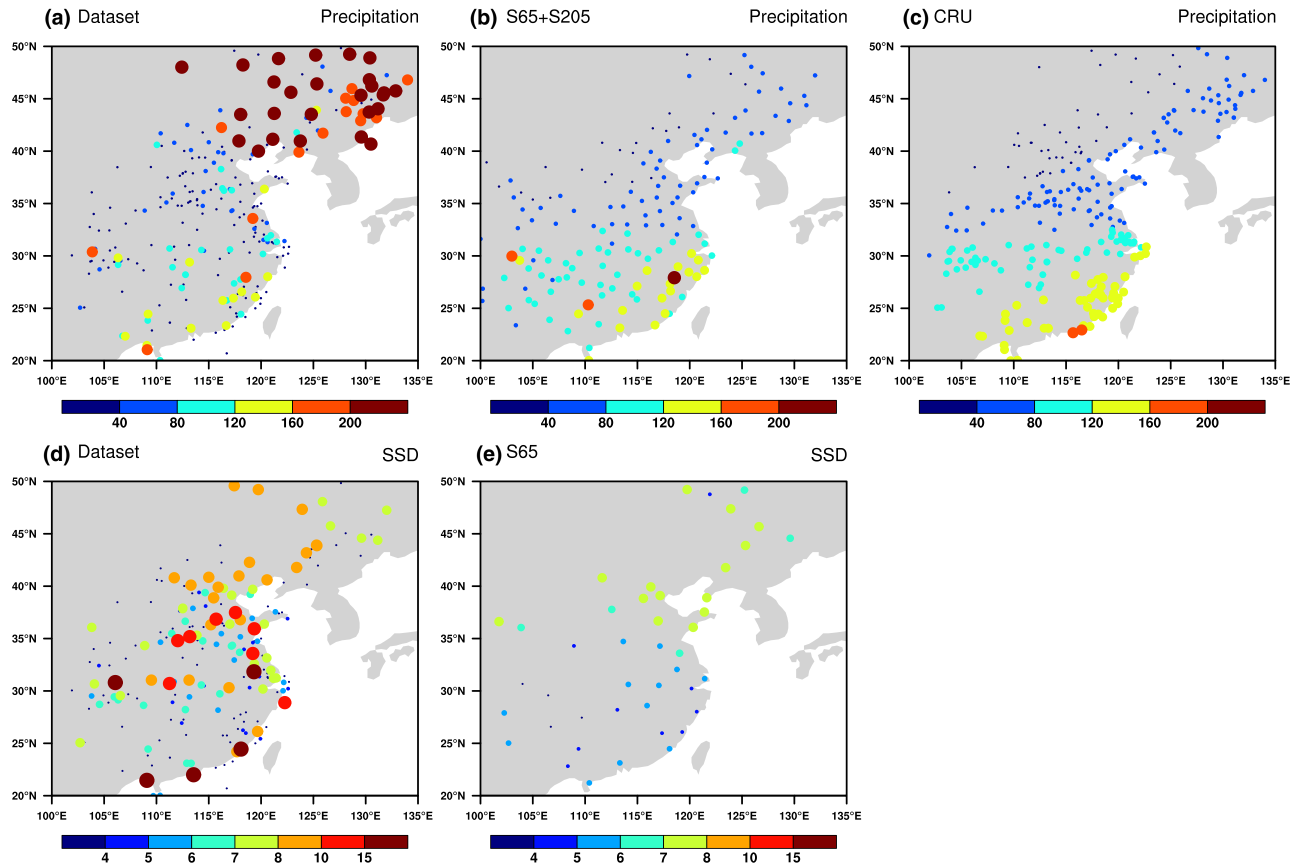


FIGURE 8 Climatology of precipitation, unit: mm/month ((a–c)) and sunshine, unit: hour/day ((d–e)). (a) and (d) represent the new dataset, (b) and (e) represent S65 + S205, and (c) represents CRU TS

coverage is primarily temperature with relatively less data on precipitation and sunshine.

To facilitate further research, a package of the raw data, code in Stata, R, and MATLAB, and processed datasets in CSV and Stata format will be made available through the Zenodo[7] and Scholarbank@NUS, named as 'NUS Republican China Weather Database'. All of the historical records have been scanned and compiled into a repository, made accessible through an intuitive online interface.[8] Future scholars can draw on this unique repository of records of weather in China between 1912 and 1951 to compile data on other dimensions of weather such as air

[7]URL http://doi.org/10.5281/zenodo.3380868

[8]URL https://doi.org/10.25541/2nb3-3x47

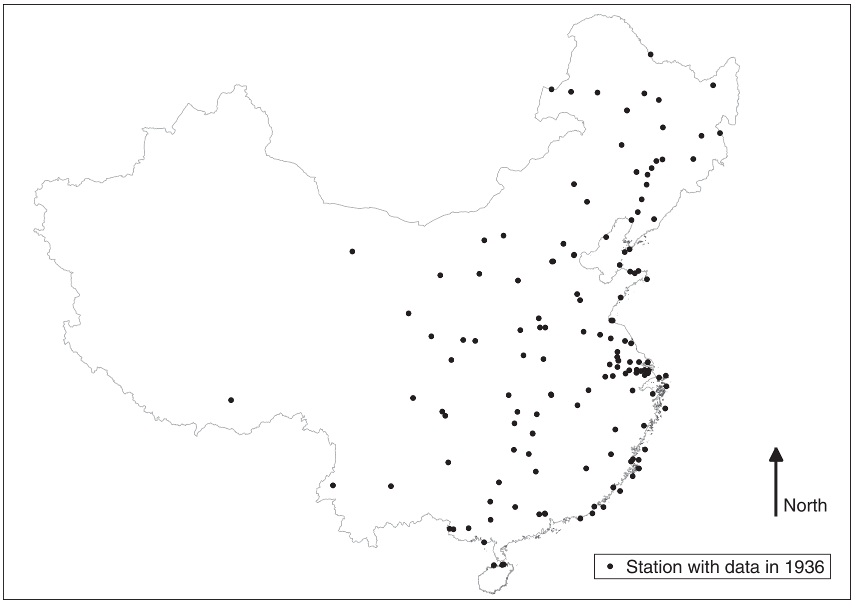


FIGURE 9 Geographical coverage in 1936, the peak year of the Monthly Meteorological Bulletin published by the Institute of Meteorology: 138 stations across 29 present-day provinces of China. Indicative mainland present-day China boundary (source: GADM.org)

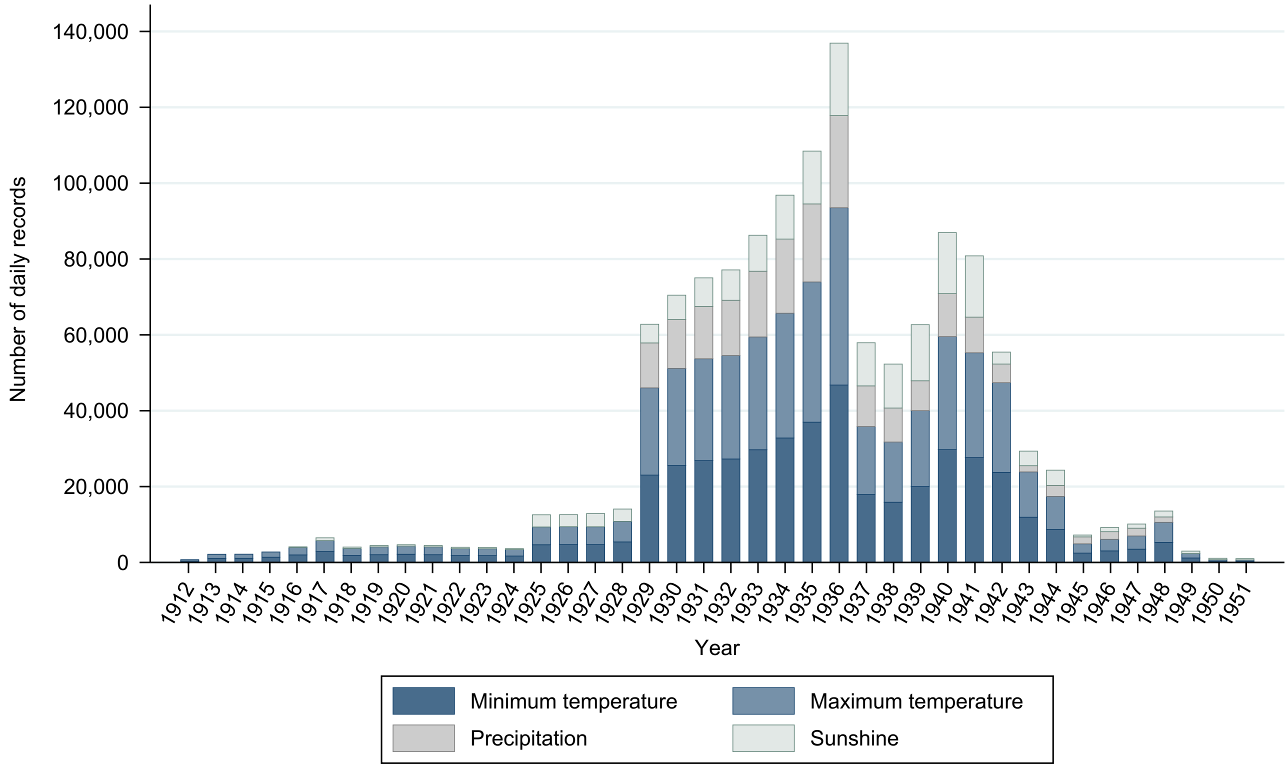


FIGURE 10 Temporal distribution of minimum temperature, maximum temperature, precipitation and sunshine

pressure, wind and humidity. Also, future work can digitize other Nationalist era weather records in the China Meteorological Administration archive (Zhongguo Qixiangju Qixiang Danganguan, 2003).

While large, the dataset presented here is limited in three important ways. The major limitation is the restricted temporal coverage as the Chinese meteorological network took time to develop and was then severely disrupted by the Japanese invasion and even more so, the Civil War (Fei, 2018). The second limitation is variation in the functional coverage as many sources reported only maximum and minimum temperatures. The third limitation is errors in typesetting and digitization, biases in recording, rounding and inhomogeneities, owing to which the data should be interpreted with caution.

Despite the limitations, the new dataset is valuable in several ways. First, compared to other instrumental datasets such as S65 and S205, it contains data at daily frequency which are valuable not only as such but also to the analysis of meso-scale and sub-seasonal climate variation. Second, it contributes to East Asian climate reconstruction studies and climate assimilation projects. Third, although most stations have only two or three decades of records, the temporal scale suffices to analyse decadal climate variability. And importantly, the time span of the dataset extends to 1951, and so, can be readily joined with later modern meteorological data to reveal a broader scale of climate fluctuations such as changing monsoon, cold surges and intra-seasonal oscillations associated with global warming.

OPEN PRACTICES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://doi.org/10.5281/zenodo.3380867 Learn more about the Open Practices badges from the Center for OpenScience: https://osf.io/tvyxz/wiki

REFERENCES

Allan, R., Brohan, P., Compo, G.P., Stone, R., Luterbacher, J. and Brönnimann, S. (2011) The International Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative. *Bulletin of the American Meteorological Society*, 92(11), 1421–1425.

Allan, R., Endfield, G., Damodaran, V., Adamson, G., Hannaford, M., Carroll, F. *et al*. (2016) Toward integrated historical climate research: the example of atmospheric circulation reconstructions over the earth. *Wiley Interdisciplinary Reviews: Climate Change*, 7(2), 164–174.

Ashcroft, L., Gergis, J. and Karoly, D.J. (2014) A historical climate dataset for Southeastern Australia, 1788–1859. *Geoscience Data Journal*, 1(2), 158–178.

Bickers, R. (2016) ‘Throwing light on natural laws’: meteorology on the China coast, 1869–1912. In: Bickers, R. and Jackson, I. (Eds.) *Treaty Ports in Modern China: Law*. Oxford: Land and Power, Routledge, Abingdon, pp. 179–200.

Brázdil, R., Dobrovolnỳ, P., Luterbacher, J., Moberg, A., Pfister, C., Wheeler, D. and *et al*. (2010) European climate of the past 500 years: new challenges for historical climatology. *Climatic Change*, 101(1–2), 7–40.

Brönnimann, S., Brugnara, Y., Allan, R.J., Brunet, M., Compo, G.P., Crouthamel, R.I. *et al*. (2018) A roadmap to climate data rescue services. *Geoscience Data Journal*, 5(1), 28–39.

Brunet, M. and Jones, P. (2011) Data rescue initiatives: Bringing historical climate data into the 21st century. *Climate Research*, 47, 29–40.

Chen, Y., Chu, J. and Png, I.P.L. (2017) *China Meteorological Publications 1872–1951*. ScholarBank@NUS Repository., https://doi.org/10.25540/PYNV-42PB.

Fei, J. (2018). *Meteorological History and Historical Climate of China*. Oxford Research Encyclopedia of Climate Science.

Harris, I.P.D.J., Jones, P.D., Osborn, T.J. and Lister, D.H. (2014) Updated high-resolution grids of monthly climatic observations - the CRU TS3. 10 Dataset. *International Journal of Climatology*, 34(3), 623–642.

Jones, P. and Mann, M. (2004) Climate over past millennia. *Reviews of Geophysics*, 42(2), 1–42.

Kobayashi, S. and Yamamoto, H. (2013) Wartime meteorological observations of the Japanese military in East Asia and the collection of its neglected data. *Historical Geography [in Japanese]*, 55–5(267), 82–98.

Rhines, A., Tingley, M.P., McKinnon, K.A. and Huybers, P. (2015) Decoding the precision of historical temperature observations. *Quarterly Journal of the Royal Meteorological Society*, 141(693), 2923–2933.

Ritchie, J.T. and NeSmith, D.S. (1991). Temperature and crop development. In Hanks, J. and Ritchie, J.T. (Eds.) *Modeling Plant and Soil Systems*. Madison, WI: Crop Science Society of America Inc. pp. 5–29.

Tao, S., Fu, C., Zeng, Z., Zhang, Q. and Kaiser, D.P. (1997). *Two Long-Term Instrumental Climatic Data Bases of the People’s Republic of China*. Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, and Carbon Dioxide Information Analysis Centre, Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A.

Wang, X.L. (2008a) Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or f test. *Journal of Applied Meteorology and Climatology*, 47(9), 2423–2444.

Wang, X.L. (2008b) Penalized maximal f test for detecting undocumented mean shift without trend change. *Journal of Atmospheric and Oceanic Technology*, 25(3), 368–384.

Wang, X.L. and Feng, Y. (2013) *Rhtestsv4 user manual*. Climate Research Division, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada.

Wang, P.K., Lin, K.E., Liao, Y., Liao, H., Lin, Y., Hsu, C. *et al*. (2018) Construction of the REACHES Climate Database based on historical documents of China. *Scientific Data*, 5, 180288.

Williamson, F., Ren, G. and Allan, R. (2017) The Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative ACRE China workshop: Recovery, digitization, and analysis of pre-mid-twentieth century climate observational data in East Asia workshop on 23–24 August, Beijing, China. *Earth and Space Science*, 4(1), 40–43.

Williamson, F., Allan, R., Ren, G., Lee, T., Lui, W., Kubota, H. *et al*. (2018) Collating historic weather observations for the East Asian region: Challenges, solutions, and reanalyses. *Advances in Atmospheric Sciences*, 35(8), 899–904.

Zhongguo Qixiangju Qixiang Danganguan (2003) *Zhongguo Qixiangju Qixiang Danganguan Zhinan [in Chinese]*. Beijing, China: Qixiang Chubanshe.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Png IPL, Chen Y, Chu J, Feng Y, Lin EK-H, Tseng W-L. Temperature, precipitation and sunshine across China, 1912-51: A new daily instrumental dataset. *Geosci. Data J.*. 2020;7:90–101. https://doi.org/10.1002/gdj3.91