README File for

Spatial Clustering of Natural Disasters, Selection in Migration, and Economic Outcomes

Junni Pan

This README outlines the procedures for replicating the project. Below is an overview:

- **Computational Requirements.** See Table 1.
- **Summary of Raw Data.** Table 2 summarizes the raw data and provides instructions for obtaining them.
- **Preliminary Cleaning Sequence.** Table 3 details the inputs, procedures, and outputs, presented in the order to be executed.
- **Construction of Analysis Datasets and Generation of Results.** Table 4 provides instructions on how to generate the analysis datasets and the results.

Table 1: Computational Requirements

Program	**Version last run**	**Update paths in**
Python	Python 3.13.2; run from Bash on a Linux-based operating system. Python routines are invoked within Bash scripts.	`Codes/cleaning_rainfall.sh`, `Codes/cleaning_temperature.sh`.
ArcGIS (with Spatial Analyst)	ArcGIS Pro 3.2.2; Windows 11.	Set paths as indicated in the replication steps.
Stata	Stata SE 19; Windows 11.	`Codes/_master.do`.

Table 2: Summary of Raw Data

Data File	Source	Access	Instructions
Main Data			
`Data/Censuses/Raw/census2000.dta`	National Bureau of Statistics (NBS)	Confidential	
`Data/Censuses/Raw/census2005.dta`	NBS	Confidential	
`Data/Rainfall/Input/FLDAS_NOAH01_C_GL_M.A198201.001.nc.SUB.nc4`, . . . , `FLDAS_NOAH01_C_GL_M.A202212.001.nc.SUB.nc4`	NASA (McNally and NASA/GSFC/HSL, 2018)	Public	Data are available at `https://disc.gsfc.nasa.gov/datasets/FLDAS_NOAH01_C_GL_M_001/summary`. Register and log in. Click *Subset / Get Data*. Download the subset of rainfall data for 01/1982–12/2022 with region coordinates (4°N, 73°E, 55°N, 136°E), broadly covering mainland China.
`Data/NTL/Input/F142000.v4b_web.stable_lights.avg_vis.tif`, `F152000.v4b_web.stable_lights.avg_vis.tif`, `F152005.v4b_web.stable_lights.avg_vis.tif`, `F162005.v4b_web.stable_lights.avg_vis.tif`	NOAA DMSP-OLS (Version 4)	Public	Data are available at `https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html`. Download `F142000,F152000,F152005,F162005`. Unzip the data.
`Data/Temperature/Input/Original/tmn*.nc`, `tmp*.nc`, `tmx*.nc`	Peng et al. (2019)	Public	Data are available at `https://zenodo.org/records/3185722`. Download the files `tmn*.nc`, `tmp*.nc`, `tmx*.nc` for years 1991–2005.

continued on next page

Data File	Source	Access	Instructions
Data/Map/County_2000.shp	Imbert et al. (2022)	Public	Available at Data/Public/Raw/Administrative in their replication package.
Data/Pop/pop2000(*.adf), Data/Pop/pop2005(*.adf)	Chinese Academy of Sciences	Proprietary	Data are available at https://www.resdc.cn/doi/doi.aspx?DOIid=32. Register and log in to download.
Data/GDP/gdp2005.dta, Data/GDP/gdp2000.dta	China County Statistical Yearbook (2001, 2006) and provincial yearbooks	Proprietary	County-level GDP data are accessible through the libraries of most Chinese universities.
Data Used Only in the Online Appendix			
LandUse/LandUse1990_hk.dta	Chinese Academy of Sciences	Proprietary	Data are available at https://www.resdc.cn/doi/doi.aspx?DOIid=54
census64.shp, census82.shp, census90.shp	NBS	Confidential	

Table 3: Preliminary Cleaning Sequence

Item	**Details**
Step 1: Preliminary cleaning of rainfall data — Part 1	
Program	Python, run from Bash.
Input	`Data/Rainfall/Input/FLDAS_NOAH01_C_GL_M.A198201.001.nc.SUB.nc4`, ... , `FLDAS_NOAH01_C_GL_M.A202212.001.nc.SUB.nc4`.
Instructions	(1) Change directory to the `Codes` folder. (2) Run `chmod +x cleaning_rainfall.sh` to make the script executable. (3) Run `./cleaning_rainfall.sh`. For details on the package used to produce SPI, see Adams (2017).
Purpose	Generate SPI and mask wet and dry disasters using a 1/0 indicator.
Output	`Data/Rainfall/Output/FLDAS_ExtremeWet_`x`sd.nc`, `Data/Rainfall/Output/FLDAS_ExtremeDry_`x`sd.nc`, where $x \in \{14, 15, 16, 165, 17, 18\}$.
Step 2: Preliminary cleaning of rainfall data — Part 2	
Program	ArcGIS.
Input	`Data/Rainfall/Output/FLDAS_ExtremeWet_`x`sd.nc`, `Data/Rainfall/Output/FLDAS_ExtremeDry_`x`sd.nc`, where $x \in \{14, 15, 16, 165, 17, 18\}$.
Instructions	(1) Use Make NetCDF Raster; set **Band Dimension** = `time`. (2) Download NetCDF_time_slice_to_Raster from here; connect to the folder containing the tool, run it, and save outputs to the corresponding folders `Data/Rainfall/`x`sd/FLDAS_`d`_monthly_`x`sd/`d`1_288_`x`sd`, where $d \in \{Wet, Dry\}$, and $x \in \{14, 15, 16, 165, 17, 18\}$.
Purpose	Slice the inputs into monthly rasters.

continued on next page

Item	Details
Output (provided)	`Data/Rainfall/`x`sd/FLDAS_Wet_monthly_`x`sd/Wet1_288_`x`sd/Band_`y`.tif`, `Data/Rainfall/`x`sd/FLDAS_Dry_monthly_`x`sd/Dry1_288_`x`sd/Band_`y`.tif`, where $x \in \{14, 15, 16, 165, 17, 18\}$ and $y \in \{1, 2, \ldots, 288\}$.
Step 3: Preliminary cleaning of rainfall data — Part 3	
Program	ArcGIS.
Input (provided)	`Data/Rainfall/`x`sd/FLDAS_Wet_monthly_`x`sd/Wet1_288_`x`sd/Band_`y`.tif`, `Data/Rainfall/`x`sd/FLDAS_Dry_monthly_`x`sd/Dry1_288_`x`sd/Band_`y`.tif`, where $x \in \{14, 15, 16, 165, 17, 18\}$ and $y \in \{1, 2, \ldots, 288\}$.
Instructions	(1) In the Catalog pane: right-click Toolboxes →Add Toolbox; add `Codes/cleaning_CountyStat_tool.atbx`. (Corresponding Python codes are in `Codes/cleaning_CountyStat_tool.py`). (2) Right-click the toolbox and choose Edit. Update the input path to the input data location and set the output path to the output folder `Data/Rainfall/`x`sd/FLDAS_`d`_monthly_csv_`x`sd`, where $d \in \{$Wet, Dry$\}$ and $x \in \{14, 15, 16, 165, 17, 18\}$.
Purpose	Generate county-level average disasters.
Output (provided)	`Data/Rainfall/`x`sd/FLDAS_Wet_monthly_csv_`x`sd/WetBand_`y`_`x`sd.csv`, `Data/Rainfall/`x`sd/FLDAS_Dry_monthly_csv_`x`sd/DryBand_`y`_`x`sd.csv`, where $x \in \{14, 15, 16, 165, 17, 18\}$ and $y \in \{1, 2, \ldots, 288\}$.
Step 4: Preliminary cleaning of temperature data — Part 1	
Program	Python, run from Bash.
Input	`Data/Temperature/Input/Original/tmn*.nc`, `tmp*.nc`, `tmx*.nc`.
Instructions	(1) Change directory to the `Codes` folder. (2) Run `chmod +x cleaning_temperature.sh` to make the script executable. (3) Run `./cleaning_temperature.sh`.

continued on next page

Item	Details
Output	`Data/Temperature/Output/tmn_mean_`x`.nc`, `Data/Temperature/Output/tmp_mean_`x`.nc`, `Data/Temperature/Output/tmx_mean_`x`.nc`, where $x \in \{1992\text{–}1996, \ldots, 2001\text{–}2005\}$.
Step 5: Preliminary cleaning of temperature data — Part 2	
Program	ArcGIS.
Input	`Data/Temperature/Output/tmn_mean_`x`.nc`, `Data/Temperature/Output/tmp_mean_`x`.nc`, `Data/Temperature/Output/tmx_mean_`x`.nc`, where $x \in \{1992\text{–}1996, \ldots, 2001\text{–}2005\}$.
Instructions	Same as Step 3, use the tool `Codes/cleaning_CountyStat_tool.atbx`.
Purpose	Generate county-level average temperature.
Output (provided)	`Data/Temperature/Temperature_csv/tmn_mean_`x`.csv`, `Data/Temperature/Temperature_csv/tmp_mean_`x`.csv`, `Data/Temperature/Temperature_csv/tmx_mean_`x`.csv`, where $x \in \{1992_1996, \ldots, 2001_2005\}$.
Step 6: Preliminary cleaning of nighttime lights data	
Program	ArcGIS.
Input	`Data/NTL/Input/F142000.v4b_web.stable_lights.avg_vis.tif`, `F152000.v4b_web.stable_lights.avg_vis.tif`, `F152005.v4b_web.stable_lights.avg_vis.tif`, `F162005.v4b_web.stable_lights.avg_vis.tif`.
Instructions	(1) Map →Layer →Add Data to add the input. (2) Run Zonal Statistics as Table (Spatial Analyst Tools) with: **Input Raster or Feature Zone Data** = `County_2000`; **Zone Field** = `CNTYGB`; **Input Value Raster** = (the input); **Statistics Type** = `Sum`. (3) Data →Export Table. Export as `.csv` and save to the output path `Data/NTL/Output`.

continued on next page

Item	Details
Output (provided)	`Data/NTL/Output/F142000.csv`, `Data/NTL/Output/F152000.csv`, `Data/NTL/Output/F152005.csv`, `Data/NTL/Output/F162005.csv`.
Step 7: Preliminary cleaning of population data	
Program	ArcGIS.
Input	`Data/Pop/pop2000(*.adf)`, `Data/Pop/pop2005(*.adf)`.
Instructions	(1) Follow the same procedure as Step 6.
Output	`Data/Pop/Output/pop2000_sum.csv`, `Data/Pop/Output/pop2005_sum.csv`.
Step 8: Preliminary cleaning of distance data	
Program	ArcGIS.
Input	`Data/Map/County_2000.shp`.
Instructions	(1) Map →Layer →Add Data to add the input. (2) Right-click to open the Attribute Table. (3) Add fields `cen_x` and `cen_y`. For `cen_x`: Calculate Geometry →Centroid x-coordinate; for `cen_y`: Calculate Geometry →Centroid y-coordinate. (4) Data →Export Table. Export as `.csv` to the output path `Data/Distance`.
Output (provided)	`Data/Distance/cnty2000_cenxy.csv`

Table 4: Construction of Analysis Datasets and Generation of Results

Item	**Details**
Program	Stata
Input	Outputs from Steps 3, 5, 6, 7, and 8 in Table 3.
Instructions	Run `Codes/_master.do`.
Output	
Main data	`Data/1.Panel/panel0005.dta`, `Data/1.Panel/panel0005_age_hete.dta`, `Data/1.Panel/panel0005_sex_hete.dta`, `Data/1.Panel/panel0005_edu_hete.dta`, `Data/1.Pooled/pooled0005.dta`.
Data used only in the online appendix	`Data/1.Panel/panel0005_stu.dta`, `Data/1.Panel/panel0005_ruralhk.dta`, `Data/1.Panel/panel0005_differentlags_`x`sd`, where $x \in \{14, 15, 16, 165, 17, 18\}$.
Tables	Section 3: Table 1. Section 4: Tables 2–3; Table B1. Section 5: Table 4; Tables A1–A2, B2–B8. Section 7: Tables B9–B11.
Figures	Section 5: Figure 1; Figure B3. Section 7: Figures 3–4; Figure B4. Section 3 Figure B1 and Section 4 Figure B2 are produced based on `Data/Map/County_2000.shp` as well as data included in the folders `Data/Rainfall/DataforFigureB1B2` (provided). Details on reproducing the exact display formats are available upon request.

References

Adams, J. (2017, May). climate_indices: An Open Source Python Library Providing Reference Implementations of Commonly Used Climate Indices. https://github.com/monocongo/climate_indices.

Imbert, C., M. Seror, Y. Zhang, and Z. Yanos (2022). Migrants and Firms: Evidence from China. *American Economic Review 112*(6), 1885–1914.

McNally, A. and NASA/GSFC/HSL (2018). FLDAS Noah Land Surface Model L4 Global Monthly 0.1 × 0.1 Degree (MERRA-2 and CHIRPS). Greenbelt, MD, USA: Goddard Earth Sciences Data and Information Services Center (GES DISC). https://disc.gsfc.nasa.gov/datasets/FLDAS_NOAH01_C_GL_M_001/summary. Accessed June 18, 2023.

Peng, S., Y. Ding, W. Liu, and Z. Li (2019). 1 km Monthly Temperature and Precipitation Dataset for China from 1901 to 2017. *Earth System Science Data 11*(4), 1931–1946.