Dry Spells, Urban Swells: Analyzing the Drought-Induced Expansion of Cities in Sub-Saharan Africa

Reproducibility Package README

Overview

This reproducibility package reproduces the figures and tables of the paper "*Dry Spells, Urban Swells: Analyzing the Drought-Induced Expansion of Cities in Sub-Saharan Africa*".

All data files required to run the analysis are located under the folder "data". Scripts required to reproduce the results in the paper using the raw datasets are located in the root directory and generate outputs that are saved under the folder "output". Scripts should be run in the following order:

1. dataset_creation.ipynb
2. Regressions.R
3. figures_1_3_5_6_7.R
4. figures_2_4.ipynb
5. Main plots.R

Data Availability

All data listed below required to run the code are publicly available:

Data publicly available:

- **SPEI:**
  - Filename: Global 12-month 1901-2021 SPEI
  - Source: DIGITAL.CSIC
  - URL: https://digital.csic.es/handle/10261/288226
  - Access year: 2023

- **Africapolis 2020:**
  - Filename: GIS – Download - 2020
  - Source: Africapolis
  - URL: https://africapolis.org/en/data
  - Access year: 2023

- **World Settlement Footprint Evolution:**
  - Filename: Dynamics of built-up expansion (yearly built-up extent) > WB-SMSUA_DLR_WSF-evolution_v01_tif.zip
  - Source: World Bank Data Catalog - World Settlement Footprint
  - URL: https://datacatalog.worldbank.org/search/dataset/0060310/Satellite%20Monitoring%20Service%20of%20Urbanization%20in%20Africa%20%28World%20Settlement%20Footprint%29?version=2

- Access year: 2024

- **GHS population grid:**
  - Filename 1: GHS-POP - Product: GHS-POP, epoch: 2000, resolution: 1 km, coordinate system: Mollweide
  - Filename 2: GHS-POP - Product: GHS-POP, epoch: 2000, resolution: 3 arcsec, coordinate system: WGS84
  - Source: Human Settlement Emergency Copernicus
  - URL: https://human-settlement.emergency.copernicus.eu/download.php?ds=pop
  - Access year: 2023

- **GHS settlement model grid:**
  - Filename 1: GHS-SMOD (Product: GHS-SMOD, epoch: 2000, resolution: 1 km, coordinate system: Mollweide)
  - Source: Human Settlement Emergency Copernicus
  - URL: https://human-settlement.emergency.copernicus.eu/download.php?ds=smod
  - Access year: 2023

- **GHS built-up surface:**
  - Filename 1: GHS-BUILT-S - Product: GHS-BUILT-S, epoch: 2000, resolution: 1 km, coordinate system: Mollweide
  - Filename 2: GHS-BUILT-S - Product: GHS-BUILT-S, epoch: 2000, resolution: 3 arcsec, coordinate system: WGS84
  - Source: Human Settlement Emergency Copernicus
  - URL: https://human-settlement.emergency.copernicus.eu/download.php?ds=bu
  - Access year: 2023

- **World Development Indicators - Agriculture value added (% of GDP):**
  - Filename: DataBank | World Development Indicators
  - Source: DataBank World Bank
  - URL: https://databank.worldbank.org/source/world-development-indicators
  - Access year: 2023
  - Variable Name: NV.AGR.TOTL.ZS

- **World Development Indicators – GDP per Capita – Countries:**
  - Filename: DataBank | World Development Indicators
  - Source: DataBank World Bank
  - URL: https://databank.worldbank.org/source/world-development-indicators
  - Access year: 2023
  - Variable Name: NY.GDP.PCAP.CD

- **World Development Indicators – GDP per Capita – Regions:**
  - Filename: DataBank | World Development Indicators
  - Source: DataBank World Bank

- URL: https://databank.worldbank.org/source/world-development-indicators
- Access year: 2023
- Variable Name: NY.GDP.PCAP.PP.KD

- **Climatic Research Unit - Rainfall:**
  - Filename: crucy.2407032054.v4.08 (-> countries -> pre)
  - Source: University of East Anglia CRU data
  - URL: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.08/
  - Access year: 2023
  - Variable Name: pre

Instructions for Replicators

1. Install python environment using the "environment.yml" file (run conda env create -f environment.yml).
2. Open the Jupyter notebook file "1. dataset_creation.ipynb" located in the root directory. Update the directory path in cell 2 to match the location of the reproducibility package (e.g.: to "C:/Users/User/OneDrive/Downloads/reproducibility_package/", then run the code to generate a .csv file in the "output" folder.
3. Install R version 4.4.1 or above. Open the file "2. Regressions.R" and install the packages required to run the script (through the GUI or by running install.packages(*{package_name}* in line 1). Update the directory path in line 10 to match the location of the reproducibility package. Run all the file.
4. Once the two above routines have been successfully run, users can reproduce figures and tables of the analysis using the following files:
   - 3. figures_1_3_5_6_7.R
   - 4. figures_2_4.ipynb
   - 5. Main plots.R

   The working directory needs to be modified before running each of these files.

List of exhibits

The provided code reproduces all tables and figures in the paper (saved under the "output" folder and named as per their position in the manuscript), from which are derived most numbers provided in text in the paper.

There is one exception: this code does not reproduce the numbers related to flood risk (presented in section 1. Context, second-to-last paragraph, and in section 2.2), which were produced using the Fathom Global Flood Map, a dataset that is not publicly available.

Important remark: Some intermediary outputs which serve as inputs for final outputs in the "output" folder will be created in the "data" folder next to raw inputs.
Namely, these are:

- in "WSF_EVO": 56 country rasters with their ISO_A3 code as file name

- in "GHS_POP / GHS_POP_country_rasters": 44 country rasters with their ISO_A3 and "_pop_2000" as file names
- in "GHS_BU / GHS_BU_country_rasters": 44 country rasters with their ISO_A3 and "_built_s_2000" as file names

Software and memory requirements

A file "environment.yml" details the python library version requirements.

R version 4.4.1 or above is required.

Memory requirements: approximately 18 Gb.