Migration in the Face of Climate Change: Assessing the Potential of UPG Programs: Climate Data Work Documentation

Till Meissner, t.a.meissner@lse.ac.uk

Last Updated: January 13, 2026

Contents

1 Overview

This document summarises the work conducted for the STEG research project *"Migration in the Face of Climate Change: Assessing the Potential of Ultra-Poor Graduation Programs"* focusing on measuring climate stress in Upper Egypt. The work encompasses data acquisition, processing, and aggregation of climate variables at both governorate and household levels to enable a rigorous analysis of climate vulnerability and its relationship to migration decisions.

The work performed can be divided into four main components:

1. **Literature review and metric identification** to capture climate stress phenomena relevant to Upper Egypt, particularly heat stress and drought as identified in household surveys.

2. **Data acquisition and processing** of climate data from multiple sources including the Copernicus Climate Data Store, Google Earth Engine, and National Oceanic and Atmospheric Administration (NOAA), along with auxiliary geographic data.

3. **Climate stress measurement and documentation** through construction of multiple indicators capturing temperature extremes, heat stress indices, and agricultural drought at daily to seasonal resolution. Documentation of increased climate stress in Upper Egypt through time series graphs and maps.

4. **Dataset construction** creating panel datasets at governorate and household levels, including household climate vulnerability indices computed through inverse distance weighting.

2 Study Area and Data Collection

The study focuses on the governorates of Asyut and Suhag in Upper Egypt, with household survey data from 3,001 households. A critical preprocessing step involved imputing missing GPS coordinates for 52.9% of female respondent records and 68% of male respondent records using a two-stage methodology: (1) cross-gender GPS pooling within households, and (2) village-level mean imputation based on governorate, district, and village identifiers. This imputation preserved spatial accuracy while maximising analytical coverage.

Geographic reference data use the HDX Egypt administrative boundaries (Level 1) from the Humanitarian Data Exchange (see **here**), which provide standardised governorate-level polygons.

3 Data Sources and Measurement Strategy

3.1 Rationale for Data Source Selection

ERA5 Reanalysis Data was selected as the primary climate data source for temperature and humidity variables. Produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) as part of the Copernicus Climate Change Service (C3S), ERA5 provides hourly data at approximately 11km resolution from 1960 to present (C3S, 2018). The selection was motivated by three factors: (1) the need for daily resolution data to construct heat stress metrics, (2) long temporal coverage enabling historical baseline construction, and (3) comprehensive variable availability including temperature, humidity, wind, and radiation needed for physiological heat indices. Raster data was accessed through Google Earth Engine.

Drought Measurement required alternative approaches due to Upper Egypt's hyper-arid climate. Precipitation-based indices (Standardized Precipitation Index, Standardized Precipitation-Evapotranspiration Index), which are commonly used in the literature, are undefined for regions with near-zero historical rainfall. Instead, agricultural drought is captured through the Evaporative Stress Index (ESI) from NOAA (see **here**) and vegetation health indices (NDVI, EVI) from NASA's MODIS satellites (see **here**) via Google Earth Engine. These capture crop stress directly rather than inferring it from precipitation deficits.

Auxiliary Geographic Data includes Digital Elevation Models (DEMs) for slope and elevation retrieved through the elevatr R package (Hollister et al., 2025) , OpenStreetMap data for distance to the Nile and other waterways (see **here**), and Google Distance Matrix API for travel distances to the governorate capitals and nearest market towns.

3.2 Climate Stress Metrics

Heat stress is measured through three complementary approaches, each capturing different aspects of thermal exposure:

1. **Measured Air Temperature**: Daily mean, maximum, and minimum temperatures provide the foundation for all thermal metrics. Variables include average temperatures by season and period, temperature shock days (defined as days exceeding 2 standard deviations from historical monthly means), and temperature variability within seasons.

2. **Heat Index**: Combines temperature and relative humidity to approximate perceived temperature. This captures the reduced effectiveness of evaporative cooling under

humid conditions, relevant for the Nile valley's microclimate. I use ERA5 temperature and humidity data and adopt the methodology developed by (Anderson et al., 2013) and implemented with the weathermetrics R package by the same authors to calculate the heat index.

3. **Universal Thermal Climate Index (UTCI)**: A bioclimatic index incorporating not only temperature and humidity, but also wind speed and radiation to assess physiological thermal stress (Bröde et al., 2012; Napoli, 2020). See **here** for a good explainer. Raster data was downloaded from Copernicus Climate Data Store (see **here**). Currently, this data is only available at a relatively coarse horizontal resolution of ∼27km (see more under Section 7.3).

Drought is captured through:

1. **Evaporative Stress Index (ESI)**: A satellite-derived measure of actual evapotranspiration relative to potential evapotranspiration, indicating crop water stress. ESI values below -1 indicate moderate drought, below -2 indicate severe drought. The 12-week composite provides seasonal agricultural drought conditions.

2. **Normalized Difference Vegetation Index (NDVI)**: A measure of vegetation greenness and health derived from red and near-infrared reflectance. Lower NDVI indicates vegetation stress.

3. **Enhanced Vegetation Index (EVI)**: Similar to NDVI but with improved sensitivity in high biomass regions and reduced atmospheric interference. Both NDVI and EVI were extracted for cropland pixels only within household buffers of 1km.

4 Temporal Structure and Reference Periods

All climate variables are constructed for multiple time periods to enable pre-post and temporal trend analysis:

- **Historical baseline**: 2000–2015 serves as the reference period for computing climatological means and standard deviations used in shock identification.

- **Pretreatment period**: 2000–2019 captures climate conditions before program intervention.

- **Treatment periods**: 2015–2019 (pre-intervention), 2020–2024 (post-intervention), and 2015–2024 (full treatment window).

Seasonal definitions follow Upper Egypt's agricultural calendar:

- **Summer**: May–October (months 5–10), the hot season with peak heat stress.
- **Winter**: November–April (months 11, 12, 1–4), the cooler season with agricultural activity. November-December are assigned to the following year's winter to maintain seasonal continuity.

Heat stress variables include several literature-based measures: heatwave days (consecutive days exceeding the 85th percentile of historical July-August temperatures), winter days above 30°C (agronomically stressful warm winter days), and extreme UTCI days (physiologically dangerous conditions).

5 Data Processing Workflow

The analysis follows a sequential numbered workflow:

1. **Data Download**: Climate data acquisition via Google Earth Engine (ERA5, ESI, NDVI/EVI) and Copernicus Climate Data Store (ERA5, UTCI).
2. **Raster Processing**: Multi-year daily rasters are combined, cropped to study area, and reprojected to consistent coordinate reference systems. Temperature data spans 1960–2024 (23,741 daily layers), requiring careful memory management through the terra package (Hijmans et al., 2026).
3. **Spatial Extraction**: Climate values are extracted to governorate polygons and household point buffers using exactextractr and terra, with options for unweighted, population-weighted, and cropland-weighted aggregation (in the case of governorate polygons) (Hijmans et al., 2026; Baston, 2024).
4. **Variable Construction**: Raw extracted values are transformed into climate stress indicators through z-score calculations, threshold exceedance counting (degree days), temporal aggregation, and growth rate computation.
5. **Dataset Integration**: Individual climate variable files are joined into master datasets at governorate and household levels, with consistent temporal periods and variable naming conventions. Final datasets are exported to both R (.rds) and Stata (.dta) formats with comprehensive variable labels and uploaded in the project Dropbox folder.

6 Key Deliverables

1. **Governorate-level panel datasets**: Time series of climate stress indicators for Asyut, Suhag, Alexandria, and Cairo from 1960–2024, enabling long-term trend analysis and cross-governorate comparison.

2. **Household-level cross-sectional datasets**: Climate exposure measures for each household location across multiple time periods, supporting regression analysis of climate vulnerability and migration decisions. Current version (`hh_data_final`) in Dropbox folder contains 2,793 households with 140+ climate variables.

3. **Vulnerability indices**: Based on the climate exposure measures, I create climate vulnerability indices using Inverse Distance Weighting to capture spatial variation in climate stress exposure at household level. Current version (`hh_indices`) in Dropbox folder contains 2,793 households with 9 vulnerability indices measured for the various treatment periods.

4. **Visualisation outputs**: Climate measures and vulnerability indices are visualised through various graphs and maps, documenting mounting climate stress in the study region at both, governorate and household level.

5. **Documentation**: This document and a variable dictionary report key background on the data work done in this project. Access to a Github repository containing the entire code base is available upon request. I am also available for any further queries or questions.

7 Data Considerations and Limitations

Several limitations warrant consideration:

1. **ESI coverage**: The Evaporative Stress Index is only defined for cropland pixels. Consequently, 177 households (5.9%) lack complete ESI time series and are excluded from drought analyses.

2. **GPS imputation**: As described above, missing household coordinates are imputed using village-level mean imputation.

3. **Resolution of UTCI data**: CDS makes UTCI data only available at a relatively coarse horizontal resolution of 0.25° x 0.25° (∼27km). Since households in Upper

Egypt are very clustered around the Nile, this resolution is too coarse to sufficiently capture variation in exposure to climate shocks at the household level. While there is an alternative dataset (Yang et al., 2024) with a finer resolution (∼1km), its temporal resolution (monthly) and coverage (ends in October 2022) rule it out for this analysis.

References

Anderson, G. B., Bell, M. L. and Peng, R. D. (2013), 'Methods to Calculate the Heat Index as an Exposure Metric in Environmental Health Research', **121**(10), 1111–1119.
URL: *https://ehp.niehs.nih.gov/doi/10.1289/ehp.1206273*

Baston, D. (2024), *exactextractr: Fast Extraction from Raster Datasets using Polygons*. R package version 0.10.0, https://github.com/isciences/exactextractr.
URL: *https://isciences.gitlab.io/exactextractr/*

Bröde, P., Fiala, D., Błażejczyk, K., Holmér, I., Jendritzky, G., Kampmann, B., Tinz, B. and Havenith, G. (2012), 'Deriving the operational procedure for the Universal Thermal Climate Index (UTCI)', **56**(3), 481–494.
URL: *https://doi.org/10.1007/s00484-011-0454-1*

C3S (2018), 'ERA5 hourly data on single levels from 1940 to present'.
URL: *https://cds.climate.copernicus.eu/doi/10.24381/cds.adbb2d47*

Hijmans, R. J., Barbosa, M. and Brown, A. (2026), *terra: Spatial Data Analysis*. R package version 1.8-94.
URL: *https://github.com/rspatial/terra*

Hollister, J., Shah, T., Nowosad, J., Robitaille, A. L., Beck, M. W. and Johnson, M. (2025), *elevatr: Access Elevation Data from Various APIs*. R package version 0.99.1.
URL: *https://github.com/usepa/elevatr/*

Napoli, C. D. (2020), 'Thermal comfort indices derived from ERA5 reanalysis'.
URL: *https://cds.climate.copernicus.eu/doi/10.24381/cds.553b7518*

Yang, Z., Peng, J., Liu, Y., Jiang, S., Cheng, X., Liu, X., Dong, J., Hua, T. and Yu, X. (2024), 'GloUTCI-m: A global monthly 1 km universal thermal climate index dataset from 2000 to 2022', **16**(5), 2407–2424.
URL: *https://essd.copernicus.org/articles/16/2407/2024/*