Environmental Displacement Dataset (EnDis) Codebook

Angela Chesler*

October 10, 2022

1 Overview

The Environmental Displacement Dataset quantifies human movement in response to sudden-onset natural hazards, including floods, storms, wildfires, landslides, earthquakes, and volcanic activity. EnDis currently covers all 54 independent countries on the African continent from 1989-2017.

2 Definitions

Environmental displacement: Movement of people away from their area of residence as a result of a meteorological, hydrological, climatological, or geological event.

Flood: Overflow of the normal boundaries of a body of water or excessive water accumulation over areas that are not normally submerged.

Storm: Meteorological disturbances characterized by low barometric pressure, precipitation, and strong winds (includes convective and tropical storms).

Wildfire: Uncontrolled and non-prescribed burning of plants in a natural setting which spreads by consuming natural fuels including trees, brush, grass, and dead plant material.

Earthquake: Seismic tremors caused by movement of a block of the Earth's crust along a geological fault.

Mass movement (landslide): Downward movement of earth materials due to excessively steeped slope, typically caused by erosion or seismic activity.

Volcanic activity: Eruptions of an opening or vent in the Earth's surface and associated spewing of materials including lava, ash, hot vapor, and gas.

3 Variables

The following variables are included in the Environmental Displacement Dataset (event version):

Variable	Description
Dis.No	Unique disaster identifier (EM-DAT)
Start.Date	Disaster start date (EM-DAT)
End.Date	Disaster end date (EM-DAT)
ISO	3-letter country code
Country	Country struck by hazard

*Perry World House, University of Pennsylvania

Disaster.Type	Primary disaster type (flood/storm/wildfire/earthquake/mass movement/volcanic activity)
Disaster.Subgroup	Disaster subgroup type (hydrological/geological/meteorological)
Displaced.Lower	Lowest estimate of quantity displaced by the observed disaster
Displaced.Best	Best estimate of quantity displaced by the observed disaster
Displaced.Upper	Highest estimate of quantity displaced by the observed disaster

4 Data collection procedures

The Environmental Displacement Dataset codes quantities of human movement in response to sudden-onset environmental shocks. This section details the methodologies for identifying candidate events, searching source material, and estimating displacement.

Step 1: Identify candidate events

Candidate events are all sudden-onset natural hazards. A list of all floods, storms, wildfires, earthquakes, landslides, and volcanic eruptions in Africa was provided by the Emergency Events Database (EMDAT). EMDAT is an international disaster database created by the Centre for Research on the Epidemiology of Disasters (CRED) to facilitate emergency humanitarian action worldwide, and provides the most comprehensive source of natural hazards available. This process identified 1,156 sudden-onset natural hazards occurring across the African continent from 1989-2017.

Step 2: Identify source materials for each event

Once all sudden onset natural hazards were identified, news articles and humanitarian reports for each individual event were compiled and scrutinized for information regarding displacement. This involved comprehensive searches through two databases: NexisUni and ReliefWeb. NexisUni, a widely-used academic search engine, was used to collect news articles from international, regional, and local sources. ReliefWeb, a humanitarian information service provided by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA), provided a comprehensive repository of humanitarian reports by IGOs and NGOs. Examining news articles and IGO/NGO reports from a wide range of reputable sources facilitates data triangulation and improves confidence in displacement estimates.

In both databases, a keyword search was used to identify articles relevant to each natural hazard. For each candidate event identified in EMDAT, research assistants searched NexisUni and ReliefWeb using the keyword formula "country name AND event type" filtering for articles published between the month of the event date and extending 6 months after. In NexisUni, the search focused on 20 reputable news outlets:

- BBC Monitoring
- Associated Press Online/Associated Press International/The Associated Press
- Xinhua International News/Xinhua General News
- dpa international
- The Canadian Press
- New York Times
- Africa News
- Agence France Presse-English
- United Press International (UPI)

- UPI Arabic News Service
- South East Asian News (India)
- The Guardian (London)
- Al Jazeera English
- Jerusalem Post
- EFE Newswire - Relevantes de LatinoAmerica
- EFE Newswire - Relevantes de Central America
- El Tiempo (Colombia)
- El Universal (Mexico)
- El Universal (Venezuela)
- EFE Newswire- Americas in Focus

Step 3: Record displacement estimates

Research assistants read through all relevant articles and reports to identify quantities of environmental displacement for each candidate event. Whenever possible, displacement estimates were triangulated across different types of sources. When multiple natural hazards occurred in the same country at the same time, research assistants used geographic details (available in EMDAT) to ensure that the quantity of environmental displacement is matched to the appropriate event. In cases where sources generated conflicting estimates, coders recorded the lowest and highest estimates. When sources converged on a single displacement quantity, it was recorded as the lowest, best, and highest estimate. In determining the "best" estimate among multiple conflicting accounts, research assistants selected the median value.

5 Intercoder reliability

Intercoder reliability refers to the extent to which two or more independent coders agree on the classification of content when applying the same coding procedure. In the Environmental Displacement Dataset, intercoder tests were conducted to ensure that the displacement data are reliable, i.e., that the coding procedures generate consistent and reproducible quantities of displacement. To achieve this, 20% of all candidate events were randomly selected for double-coding and evaluated for intercoder reliability. In other words, for 232 events, two different research assistants followed Steps 2 and 3 outline above, and coded the quantity of displacement for each event. The displacement estimates were then compared across coders, and the results show an intercoder agreement rate of approximately 91%. In essence, this means that when two research assistants independently followed the protocol outlined above, they reported the same quantity of displacement in 91% of cases (Krippendorffs $\alpha = 0.838$). The high rate of intercoder agreement and Krippendorffs $\alpha > 0.80$ confirms the displacement estimates available in the Environmental Displacement Dataset exhibit a high degree of reliability.