

Reproduction Analysis of Malcomb et al 2014

Joseph Holler, Kufre Udoh, Drew An-Pham, Benjamin Cordola, Middlebury Open GIScience Classes

2023-12-15

Abstract

Malcomb, Weaver and Krakowka (2014) published one of the first sub-national geographic climate change vulnerability models for a developing country (1.4). The authors intended for the study to be replicable across space (other African countries with similar data available) (7.1), time (when new survey data is published) (4.5 and 7.1), and vulnerability stimuli (7.1). The study's social impacts are to address extreme vulnerability to climate change (1.3) and assisting in the allocation and evaluation of foreign aid (1.2). The methodology was designed to be “transparent and easily replicable” (2.1) in its use of “locally derived indicators and granular data” (2.1). The study was designed to address critiques of vulnerability models aimed at their uncertainty and sensitivity due to problems of scale and spatial aggregation, normative and subjective modelling decisions, and data availability, and challenges in model comparability (2.1). The model uses household adaptive capacity data from the United States Agency for International Development (USAID) Demographic and Health Surveys (DHS) (1.4 and 4.1) available in 44 African countries (7.1), livelihood sensitivity data from the USAID / Famine Early Warning Systems Network (FEWSnet) livelihood zones baseline surveys available in 23 African countries (3.6), and global physical exposure data from the United Nations Environment Programme (UNEP) Global Risk Data Platform.

This replication study is motivated by three factors. First, there is an urgent need to evaluate the reproducibility of research in human-environment and geographical sciences (HEGS) and to establish protocols and infrastructure for conducting and publishing reproduction/replication studies and reproducible research in HEGS. Second, a fully reproducible publication can be more readily replicated in new geographic, temporal, and thematic contexts, and tested for uncertainty due to data constraints and subjective modelling decisions. Third, climate change is causing increasingly severe in Africa. Improving the reproducibility and replicability of climate vulnerability research will hopefully enhance the potential for research to inform policy and reduce harm caused by climate change.

Malcomb et al (2014) produce two models of interest for Malawi. Figure 4, labelled “Malawi Household Resilience”, visualizes the average adaptive capacity score of households in each traditional authority. Figure 5, labelled “Malawi Composite Vulnerability Index”, visualizes vulnerability scores by locations (cells) in a continuous raster grid. In this study, we will attempt to identically reproduce figure 4 (adaptive capacity by traditional authority) and figure 5 (vulnerability grid) using The R Project for Statistical Computing and the same data sources cited in the original publication. We will visually compare our resulting reproduction figures with the original figures. Comparison will be aided by digitizing and joining the original figure results to the reproduction results for each model, and then calculating any differences between them. Differences will be visualized with thematic maps for both models, a confusion matrix for figure 4 (adaptive capacity by traditional authority), and a scatterplot for figure 5 (vulnerability grid). An exact reproduction should produce exact replicas of the rank order of traditional authorities by adaptive capacity and grid cells by vulnerability. We will test this with the Spearman's Rho Correlation Coefficient, expecting values of 1 for perfect correlation.

The original study is a descriptive geographic multi-criteria analysis based on local expert opinion, and therefore has no testable hypotheses or effects.

The replication study data and code will be made available in a GitHub repository to the greatest extent that

licensing and file sizes permit. The repository will be made public at github.com/HEGSRR/RPr-Malcomb-2014

Malcomb, D. W., E. A. Weaver, and A. R. Krakowka. 2014. Vulnerability modeling for sub-Saharan Africa: An operationalized approach in Malawi. *Applied Geography* 48:17–30. DOI:10.1016/j.apgeog.2014.01.004 (<https://doi.org/10.1016/j.apgeog.2014.01.004>).

Keywords

Reproducibility, Vulnerability, GIS, Climate Change, Africa

Study design

The reproduction study design will first implement the original study as closely as possible to reproduce the 2010 Household Resilience map (F4) and Malawi Vulnerability Map (F5). Our two confirmatory hypotheses are that we will be able to independently reproduce results for both maps.

The working hypotheses are therefore:

H1: There is no perfect positive correlation between Malcomb et al’s ranking of traditional authorities by household resilience and our reproduction study’s ranking of traditional authorities by household resilience.

H2: There is no perfect positive between Malcom et al’s ranking of locations by climate vulnerability and our reproduction study’s ranking of locations by climate vulnerability.

We will evaluate each of these hypotheses using a Spearman’s Rho Correlation. A failure to reject these hypotheses would indicate that our results do not exactly match those of the original authors. A positive correlation approaching 1 would indicate a partial reproduction

Original study design

The original study is **observational** and **descriptive**, with no hypotheses or effect sizes. The study is a multi-criteria analysis using geographic information systems (GIS) to implement a hierarchical geographic model of climate change vulnerability model in Malawi.

The **spatial extent** of the study was the country of Malawi. The **spatial scale** of the study was the third administrative level (traditional authorities) and a raster grid of unknown spatial resolution. The **temporal extent** of the study was explicitly 2004–2010 (4.5), but the contains secondary data collected earlier (3.6 and F5).

The model themes, indicators, and weights were selected based upon 70 interviews and 11 village focus groups from field trips to Malawi in March and August of 2011 (1.4, 4.2 and A1). Themes and indicators were also contextualized in literature (3.3 through 3.7) and adjusted based on redundancy and representativeness across the country (4.3). The model and weights were adjusted through “several iterations of the model using alternative weighting schemes” (4.3) to produce a “final product that reflects Malawi’s contextual and perceptual vulnerability” (4.3). Each theme was constructed of indicators from a single data provider: adaptive capacity is measured with USAID DHS surveys, livelihood sensitivity is measured with FEWSnet/Malawi Vulnerability Assessment Committee (MVAC) livelihood zones baseline data, and physical exposure is measured with UNEP Global Risk Data Platform data (T1 and T2). Although the authors emphasize a grounded local evidence-based selection of indicators and weights (2.1, 4.2, 5.1 and 7.1), other evidence in the publication suggests a model design based on a more pragmatic combination of factors including expert local opinion, deductive theory, and the availability and characteristics of data.

The study did not use any **randomization**.

The original study was conducted using STATA™ (4.4) and ArcGIS™ (4.6, F3 and F4) with unspecified software versions, by 2012 according to creation dates on map figures (F3, F4 and F5).

Computational environment

The study was originally conducted using ArcGIS and unspecified statistical software. This reproduction study uses R, including the rdhs package for DHS survey data, the sf package for vector analysis, the stars package for raster analysis, and the tmap package for cartography.

```
# set up default knitr parameters
knitr::opts_chunk$set(
  echo = FALSE,
  fig.width = 8,
  fig.path = paste0(here("results", "figures"), "/")
)

# these values allow you to access private and public raw data more efficiently
private_r <- here("data", "raw", "private")
public_r <- here("data", "raw", "public")
public_d <- here("data", "derived", "public")
scratch <- here("data", "scratch")
```

Data

Lakes

Major lakes were downloaded from MASDAP, the Malawi Spatial Data Platform.

Lakes data transformations

Dissolve lakes into a single multi-part feature with one field EA containing the value Lake.

Livelihood zones

Livelihood zones geographic data may be downloaded from the FEWS NET Data Center at <https://fews.net/fews-data/335>.

Livelihood sensitivity data is derived from household economic analysis (HEA) baseline surveys of livelihood zones created by MVAC in collaboration with USAID and FEWSnet (3.6). Livelihood zones are distinct from traditional authorities (5.6). They are “geographic areas where populations share characteristics of farming practices, labor, and environmental coping strategies” (3.6). Eleven zones were surveyed in 2003 (3.6). An MVAC 2005 report on livelihood zones appears in the references with an expired URL (R).

Livelihood sensitivity is measured with the following variables from FEWSnet livelihood zone data.

- 6%: percent of food from own farm (T2)
 - ability to meet food needs (T1 theory)
 - % food intake from personal farm (T1 indicator)
 - % of food that poor households receive independently from their own farm, an indication of sustainability of livelihoods (3.6)
- 6%: percent income from wage labor (T2)
 - % of income that poor households receive from wage labor (3.6)
 - income source (T1 theory)
 - % poor income from labor (T1 indicator)
- 4%: percent income from cash crops (T2)
 - % of labor income that is susceptible to market shocks (i.e. tobacco, sugar, tea, & coffee (3.6)
 - cash crop exposure (T1 theory)
 - % non-food crop (cotton, tobacco, tea) (T1 indicator)
- 4%: disaster coping strategy (T2)

- ecological destruction associated with livelihood coping strategies during time of crisis (3.6)
- ecological coping effect (T1 theory)
- access to alternative form of income (T1 indicator)

Livelihood zones attribute data was provided by FEWS NET in the form of one three spreadsheets describing typical livelihood profiles for each zone, with one sheet for **poor** households, one for **middle** income households, and one for **rich** households. This data was based on focus groups with stakeholders in each livelihood zone. The authors have summarized the individual **poor** household spreadsheets into one comprehensive table of variables relevant to the study.

Livelihood zone data transformations

In order to prepare geographic livelihood zone data for analysis, geometry errors are fixed, national parks are removed, and the coordinate reference system is transformed to EPSG:4326 (WGS 1984) geographic coordinates. Livelihood zone attribute data is then joined to the geographic data by livelihood zone code LZCODE.

Traditional authorities

The adaptive capacity analysis is conducted in traditional authorities, which may be provided by the “GADM administrative boundaries for Africa” cited on maps of household resilience (F3 and F4). No date, version, or formal citation for this data is provided in the original study. Traditional authorities (TAs) data can be downloaded from Database of Global Administrative Areas (GADM) version 2.8 at https://gadm.org/download_country_v2.html and unzipped. This data must be downloaded directly from GADM. While the data license permits free use of data for research purposes and publication, it does not permit redistribution.

Traditional authorities (TAs)

Load traditional authorities (TA) data, fix geometry data, and count types of areas.

Type	N
City	4
Headquarter	16
National Park	6
Reserve	8
Sub-chief	66
Town	6
Traditional Authority	134
Urban	3
Water body	13

Visualize Lakes, Livelihood Zones, and TAs

TA data transformations

TA data includes conservation areas (reserves and national parks) and water bodies which do not contain populated villages. Extract conservation areas (forests and parks) to a new **ta_cons_v** layer.

Several of the Lake Malawi water body features in TA data erroneously include populated areas of land. Extract these features as **ta_lake_malawi**. Likoma island is incorrectly labelled as Lake Malawi, so do not include it as an error for extraction.

Remove conservation areas and water bodies from TAs.

```
## [1] "256 features in original traditional authorities"
```

```
## [1] "230 features after removing conservation areas and water bodies"
```

Find areas of Lake Malawi features that are actually land by buffering lakes by 500 meters and clipping the Lake Malawi TA features. Calculate new unique second level ID's as 1000 times the row number. Remove splinter polygons by selecting polygons over 4 km² with centroids intersecting livelihood zones.

Merge fixed TA errors back into TA data and save results as derived `ta_v.gpkg`.

```
## [1] 9 features created by fixing errors on Lake Malawi shore
```

```
## [1] 239 features in final corrected traditional authorities
```

Drought risk and flood risk

Physical exposure data is derived from the United Nations Environment Programme (UNEP) Global Risk Data Platform (1.4) as global (3.7) continuous raster data (5.6). The climate vulnerability map also cites the Dartmouth Flood Observatory (1999-2007) (F5). According to the references to Peduzzi (2011, 2012), the data for flood risk and drought exposure is available from UNEP/DEWA/GRID-Europe at preview.grid.unep.ch/. The drought risk data is based on “a global monthly gridded precipitation dataset obtained from the Climatic Research Unit (University of East Anglia)” and “a global Standardized Precipitation Index based on Brad Lyon (IRI, Columbia University) methodology” (3.7).

Physical exposure is measured with the following two indicators.

- 20%: estimated risk for flood hazard (T2)
 - floods & rain variability (T1 theory)
 - flood events (T1 indicator)
 - risks of flood (3.7)
 - global estimated risk index for flood hazard (R)
- 20%: exposition to drought events (T2)
 - drought & dry spells (T1 theory)
 - drought indices (T1 indicator)
 - (risks of) drought exposure (3.7)
 - physical exposition to drought events 1980 - 2001 (R)

The UNEP Global Risk Data Platform used for this research is no longer available online. The data is provided with the research compendium.

Household DHS data

Household adaptive capacity data is derived from USAID DHS Surveys conducted in 2004 and 2010 (1.4). Readers are referred to the DHS website for an “explanation on using survey data with GPS information” (4.4). The website, www.measuredhs.com, is provided in the references, and forwards to dhsprogram.com. There were 24,850 household surveys in 2010 (5.2), providing data for 203 traditional authorities (F3).

Adaptive capacity is composed of **assets** and **access** with the following DHS survey variables.

Assets

- 6%: Arable land (hectares) (T2)
 - amount of arable land (T1 theory) per household (T1 indicator)
 - larger landholders can diversify crops and sell food (3.4)
- 4%: Number of livestock units (T2)
 - livestock (T1 theory)
 - number of animals per household by type (T1 indicator)
 - animals used as coping strategy (3.4)
- 4%: Wealth index score (T2)
 - money (T1 theory)

- wealth index (based on owned assets) (T1 indicator)
- wealth (disposable capital assets) (3.4)
- income is discussed separately from wealth (3.4) but is not included as an indicator
- 3%: Number in household sick in past 12 months (T2)
 - good health (T1 theory and 3.4)
 - sick in the past 12 months (T1 indicator)
- 3%: Number of orphans in household (T2)
 - orphan care (T1 theory)
 - number of orphans or vulnerable children (T1 indicator)
 - orphans... are a highly socially vulnerable subset of the population (3.4)
 - orphan care adds tremendous burden to families that are... poor and food insecure (3.4)

Access

- 4%: time to water source (T2)
 - basics (T1 theory)
 - water (time to source) (T1 indicator)
 - burden that often falls to women and can consume large amounts of time... in a time of shock or drought, water collection time can be protracted causing even greater hardship and vulnerability (3.5)
- 4%: own a cell phone (T2)
 - media and information (T1 theory)
 - own a cell phone (Y/N) (T1 indicator)
 - households were better prepared, informed and warned about disasters through being well-connected through radio, mobile technology, or tribal networks (3.5)
- 3%: own a radio (T2)
 - technology sharing (T1 theory)
 - own a radio (Y/N) (T1 indicator)
 - Radio programs are powerful tools for reaching previously inaccessible populations (3.5)
- 3%: electricity (T2)
 - basics (T1 theory)
 - electricity (Y/N) (T1 indicator)
 - access to the electrical grid (3.5)
- 2%: type of cooking fuel (T2)
 - basics (T1 theory)
 - cooking fuel type (T1 indicator)
 - selling of charcoal is one of the top coping strategies during periods of food insecurity and market shocks (3.5)
- 2%: house setting (urban/rural) (T2)
 - market access (T1 theory)
 - rural, peri-urban, urban (T1 indicator)
 - nearest vehicle-accessible road can be several kilometers and the nearest paved road for public transportation to urban centers might be a days or more journey by foot (3.5)
- 2%: sex of head of household (T2)
 - power and decision-making (T1 theory)
 - female-headed HH (Y/N) (T1 indicator)
 - households headed by females are more vulnerable based on less access to sources of power, land, and resources (3.5)
 - households headed by one parent or by children (encompassed in the variable family structure) were seen as more vulnerable (3.5)

Geographic USAID Demographic and Health Survey (DHS) data requires pre-approved access clearance and login credentials from the DHS Program. For this reproduction study, the following procedure was used to gain access:

1. Go to <https://dhsprogram.com/Data/>
2. Create an account, ideally with an education or government e-mail address
3. Within the Datasets menu, Create a new project
4. Enter the following information: **Project Title:** Reproducing a Vulnerability Model of Malawi
Description of Study: The purpose of this study is to reproduce the methods of a published research article: Malcomb, D. W., E. A. Weaver, and A. R. Krakowka. 2014. Vulnerability modeling for sub-Saharan Africa: An operationalized approach in Malawi. *Applied Geography* 48:17–30. <https://doi.org/10.1016/j.apgeog.2014.01.004>. The authors of this paper used geocoded DHS surveys for Malawi in 2004 and 2010, in combination with FEWSnet livelihood data and UNEP flood and drought risk data. Following the author’s methodology, we plan to download the data using the `rdhs` package for R and aggregate the data at Malawi’s 2nd administrative level: districts. We will be working with a GitHub repository that stores the raw data locally in a directory ignored by the `.gitignore` file, and only moves the data into a shared and version-controlled directory once it has been aggregated to the District level. This will ensure that the privacy of survey respondents and requirements of data partners are protected, because all of the data will be aggregated into district polygons, as already shown and published in Malcomb et al (2014).
5. Choose Region: Sub-Saharan Africa
6. Click **Show GPS Datasets** at the top-left of the country tables
7. Check **Survey** and **GPS** data for **Malawi**
8. Save selection
9. Read and agree to the conditions of use for the DHS Program datasets and save these conditions for your metadata records.
10. Enter a **Justification for using DHS Program Geographic Datasets:** The research aim is to reproduce Malcomb et al (2014) in which GPS Datasets are used to spatially join DHS Survey data to Malawi’s Districts for the purpose of sub-national climate change vulnerability mapping. Therefore, the research will not be reproducible without the geographic datasets.

The `rdhs` package can be used to download the data, provided a login email and project name via console and password via pop-up dialogue.

Download the Malawi 2010 survey data and geographic points.

Load tabular data of household surveys

Load geographic data of household survey clusters. Some household survey points are erroneously placed at the WGS 1984 coordinate reference system origin (Equator and Prime Meridian).

DHS Data Transformations

In order to simultaneously maximize reproducibility while avoiding direct redistribution of DHS GPS data, we spatially join the GPS data to the Traditional Authority enumeration areas. Adaptive capacity is ultimately mapped by traditional authority, but the data comes from household-level surveys. Surveys are grouped into clusters with one geographic point. Therefore, the traditional authority to which each survey will be assigned must be spatially joined to the cluster point, and then joined by attribute to the household survey. The adaptive capacity calculation at the household level also requires urban/rural status, which is stored in the cluster.

Many household surveys contain inconclusive answers (e.g. “I don’t know”) or are missing data for survey questions used in the adaptive capacity calculation. The livestock variable will be calculated as a sum of four livestock types, so we remove any household with uncertain answers about any of the livestock types and remove households with missing data for all livestock types. Households with answers about some livestock types and missing data for others are still included in the data.

We remove incomplete household surveys.

Prior observations

Some of the authors had already examined the data and attempted a reproduction study prior to writing the preregistered analysis plan.

Bias and threats to validity

The **spatial extent** of the study was the country of Malawi (OSM relation 195290), excluding large bodies of water, national parks or similarly reserved land, and areas missing data (4.5). 203 traditional authority areas were included in the original study (F4).

The authors suggest that the scale of the phenomena of vulnerability dynamics in the context of climate change is at the **household level** (1.4, 2.2, 3.1 and 4.4). The authors use the third administrative level (traditional authorities) as the **spatial scale** and **units of analysis** of household resilience (4.4 and F4). The **spatial support** for the final analysis of climate vulnerability is a raster grid (4.6, F5) with unknown spatial resolution—appearing finer than the size of traditional authorities and the smallest unit on the scale bar, which is 12.5 kilometers (F5). We presume that the spatial resolution may be identical to at least one of the gridded physical exposure raster inputs.

Edge effects and neighboring countries will not be addressed in the analysis (4.2). The spatial analysis techniques in this study are not sensitive to edge effects.

The analysis does not include creation of any **spatial subgroups** and does not measure or account for any **spatial autocorrelation**, **spatial heterogeneity**, or **spatial anisotropies**.

Analysis

Planned differences from the original study

The replication study will focus on reproducing 2010 household resilience (F4) and climate vulnerability (F5), excluding the 2004 household resilience analysis (F3). The aim of this reproduction is to produce results identical to the original study. Therefore, we will not collect new interview or focus group data. Additionally, qualitative interview and focus group data was not provided with the original study. Therefore, we will not attempt to reinterpret any qualitative data or determine new themes, indicators or weights for the models. The reproduction study will use the indicators and weights as they are described in the original study.

The replication study will use a different software environment, using replicable open source software over proprietary software. Specifically, the study will be completed using The R Project for Statistical Computing version 3.6.1 or later using RStudio version 1.3.1 or later, and the research will be completed in full on both Windows 10 and MacOS operating systems. A complete list of required R packages is not known at the time of preregistration, but will be reported with the final publication.

The study will attempt to reproduce the original methods exactly, but some differences may be inevitable due to ambiguous or conflicting information in the original article. We will plan to make the following reasonable decisions, which may differ from the authors' intentions: 1. Figure 4 represents adaptive capacity, composed of assets and access. 1. Adaptive capacity scores will be calculated for each household, and then household scores will be spatially joined by traditional authority and averaged. 1. Figure 5 represents vulnerability, composed of adaptive capacity, livelihood sensitivity, and physical exposure. 1. Every indicator will be rescaled to a 0 to 4 scale using the formula: $\text{percent rank} * 4$. This method is a compromise from the uncertainty caused by a 0 to 5 scale, quintiles, and nominal indicators. 1. High ranks (4) will be assigned to better and safer conditions for each indicator. 1. Weighted aggregation will be formulated so that the aggregate scores have a theoretical minimum of 0 and maximum of the assigned percentage for the thematic concept. - Assets = $([\text{land}] * 0.06 + [\text{livestock units}] * 0.04 + [\text{wealth}] * 0.04 + [\text{number sick}] * 0.03 + [\text{orphans}] * 0.03) * 25$ - Access = $([\text{water}] * 0.04 + [\text{cell phone}] * 0.04 + [\text{radio}] * 0.03 + [\text{electricity}] * 0.03 + [\text{cooking fuel}] * 0.02 + [\text{urban/rural}] * 0.02 + [\text{female household}] * 0.02) * 25$ - Livelihood sensitivity = $([\text{subsistence food}] * 0.06 + [\text{wage income}] * 0.06 + [\text{cash crop income}] * 0.04 + [\text{disaster coping}] * 0.04) * 25$

- Physical exposure = (flood risk * 0.2 + drought exposure * 0.2) * 50 1. Each thematic indicator will be rasterized or resampled to the UNEP/GRID data input most closely resembling the spatial resolution of figure 5. 1. Vulnerability will be calculated so that the aggregate scores have a theoretical minimum of 0 and maximum of 100. This is achieved by inverting physical exposure. - Vulnerability = Assets + Access + Livelihood sensitivity + (40 - Physical Exposure) 1. Any traditional authority missing adaptive capacity data from DHS surveys will be removed / masked from the final vulnerability analysis.

Adaptive Capacity

The variables for adaptive capacity are aggregated into thematic concepts and referenced in the original paper as outlined below:

- 40%: Adaptive capacity (T2)
 - “adaptive capacity” defined as “household-level assets to recover from disasters and access to resources” (2.2) and referred to as:
 - * “adaptive capacity”, “capacity score”, or “adaptive capacity score” (3.3, 4.6 formula, 5.2, 5.4, 6.3)
 - * “assets” and “access” (3.3, 5.2, F3 and F4)
 - * “assets” and “access” included, but not “adaptive capacity” (1.4, T1 theory, F5)
 - * “resilience”, “household(-level) resilience” or “resilience scores” (5.2, 5.3, F3, F4 and F5, 6.4)
 - * “vulnerability” (4.1, 4.4, 4.5, 5.1, 5.3, 5.4, 5.5, 6.1, 6.2, 6.3, 6.4, 7.2)
 - measured as a positive condition (4.6)
- 20%: Assets (T2)
 - defined only as a component of adaptive capacity: “assets to recover from disasters” (2.2) and referred to as:
 - * “assets” (1.4, 3.3, 3.4, T1 theory, F5)
 - measured as a positive condition (4.6)
- 20%: Access (T2)
 - defined only as a component of adaptive capacity: “access to resources” (2.2) and referred to as:
 - * “access” (1.4, 3.3, 3.5, T1 theory, F5)
 - measured as a positive condition (4.6)

Rescale adaptive capacity indicators

Calculate percent rank for each component of household adaptive capacity. We had to make many assumptions about calculating individual components, e.g. about how to aggregate different forms of livestock, and which values to invert such that high numbers correspond to low capacity (e.g. number of orphans or sick members of the household). Rescaling to a quintile rank as described in the original study is unclear, especially considering the number of discrete or even binary inputs. We have made a judgement call to do this by calculating percent rank and multiplying by 4, producing a theoretical domain of 0 to 4 similar to that of quintiles.

Household adaptive capacity

Calculate household-level adaptive capacity scores based on original study Table 2 weights. The indicators have already been rescaled to a possible domain of 0 to 4, and the weights sum to 0.4, giving a possible domain of adaptive capacity scores from 0.0 to 1.6.

Summary statistics of adaptive capacity and its components at the household level.

Traditional authority adaptive capacity

Aggregate household adaptive capacity scores to traditional authorities. The original paper found adaptive capacity scores for 203 TAs, of which we found 6 TAs were conservation areas, leaving 197 meaningful TA scores. We created an additional 9 TAs from errors from three features on Lake Malawi, so if the original authors did not notice those errors, we could expect scores for 206 TAs.

Now that household adaptive capacity data has been aggregated, they may be saved to the `data\derived\public` directory.

Load aggregated public adaptive capacity data. `## RUN THIS ## RUN THIS ##`

```
## Reading layer `ta_v' from data source
## `C:\Users\bcordola\Documents\GitHub\RPr-Malcomb-2014\data\derived\public\ta_v.gpkg'
## using driver `GPKG'
## Simple feature collection with 239 features and 15 fields
## Geometry type: MULTIPOLYGON
## Dimension: XY
## Bounding box: xmin: 32.67152 ymin: -17.12721 xmax: 35.91505 ymax: -9.363796
## Geodetic CRS: WGS 84

## # A tibble: 6 x 18
##   ta_id capacity_avg capacity_min capacity_max capacity_sd n_hh livestock_avg
##   <dbl>      <dbl>      <dbl>      <dbl>      <dbl> <int>      <dbl>
## 1     1         0.667         0.384         0.855         0.139     20         0.120
## 2     2         0.380         0.100         0.846         0.146    315         0.982
## 3     3         0.390         0.0960         0.846         0.152    450         1.05
## 4     4         0.635         0.227         0.993         0.163    301         0.404
## 5     5         0.370         0.164         0.761         0.158     24         1.41
## 6     6         0.431         0.203         0.741         0.137     54         0.524
## # i 11 more variables: sick_avg <dbl>, land_avg <dbl>, wealth_avg <dbl>,
## #   orphans_avg <dbl>, water_avg <dbl>, electricity_avg <dbl>,
## #   cooking_avg <dbl>, femalehh_avg <dbl>, cellphone_avg <dbl>,
## #   radio_avg <dbl>, urban_avg <dbl>
```

Count TAs with adaptive capacity data.

```
## [1] 215 TAs have adaptive capacity data
```

Finding scores for 215 traditional authorities is surprising, and most likely relates to differences in discovery and treatment of geometry errors and missing data. The reason(s) for these differences cannot be determined with the content of the original manuscript.

Mapping adaptive capacity

Join adaptive capacity data to geographic TAs and rescale in attempt to match original publication. The original publication figure 4 shows ranges from 11.48 to 25.77, but after rescaling indicators to domains of 0 to 4 and multiplying by percentages in table 2 (which sum to 0.4), the theoretical domain is only 0 to 1.6. We might suppose that the authors had rescaled adaptive capacity to a possible domain of 0 to 40 in accordance with the 40% weight of adaptive capacity in the overall vulnerability model. Therefore, we may multiply our possible domain of 0 to 1.6 by 25 to achieve a possible domain of 0 to 40.

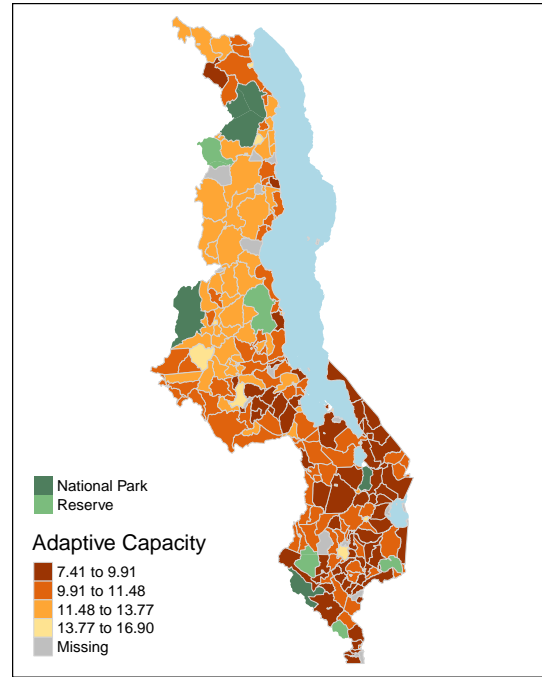
	rpac_unscaled	rpac
nbr.val	215.00	215.00
nbr.na	24.00	24.00
min	0.30	7.41
max	0.68	16.90
range	0.38	9.48
median	0.43	10.66
mean	0.44	10.99
std.dev	0.07	1.80

The original publication uses the Jenks Natural Breaks method to classify the data.

rpac_class	n
1	67
2	80
3	53
4	15
NA	24

Reproduction figure 4

Map reproduction results for comparison to figure 4.



Evaluate adaptive capacity reproduction

In order to test the adaptive capacity results, we will georeference the original figure 4 map using the QGIS3 georeferencer plugin. Using a vector dataset of traditional authorities and the georeferenced map, we will then use zonal statistics to extract the average brightness values, (which represent four classes of adaptive capacity) for each traditional authority. We will use an interior buffer of the traditional authority polygons, optimized in order to avoid summarizing border symbol in zonal statistics while capturing as much of the choropleth color symbol as possible. After inspecting a histogram of the mean brightness values, we will reclassify the values as closely to the four classes on the original figure 4 as possible and then manually adjust the attribute values for any misclassified traditional authorities. We will compare original and reproduction household resilience results by creating a confusion matrix, calculating the Spearman's Rho correlation coefficient (expecting a value of 1 for perfect positive correlation), and creating a thematic map of the difference between the original results and replication results.

Digitize original study figure 4 Ordinal data from figure 4 was digitized in QGIS with the following procedure:

1. Copy image from the original publication pdf file using Adobe Acrobat Pro
2. Paste the image and save as a .png file with pixel dimensions 1982 by 2811

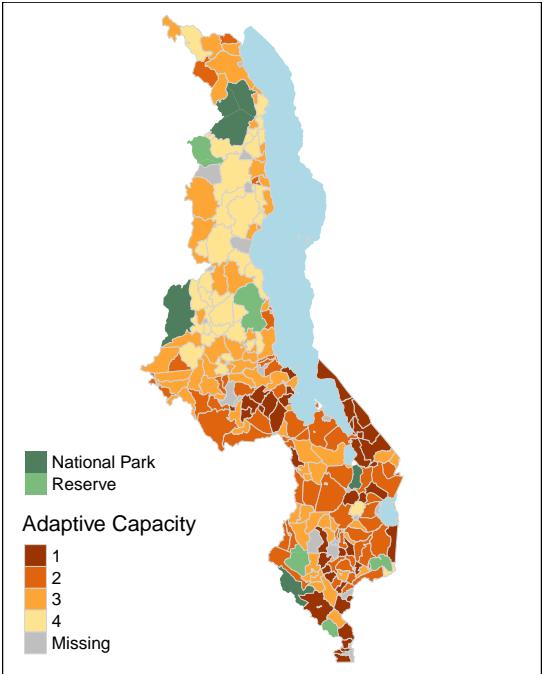
3. Use QGIS 3.26.3 Georeference the map image to match `ta_v.gpkg` using WGS 84 geographic coordinates (epsg:4326). Use linear georeferencing with points in `metadata\malcomb_fig4.png.points`
4. Make internal buffer to reduce the noise from boundary line symbology.
 1. Project `ta_v` to UTM 36S epsg:32736: `ta_v_fig4.gpkg:utm36s`.
 2. Calculate an internal buffer of -600m: `ta_v_fig4.gpkg:utm36s`.
 3. Project back to WGS 84 epsg:4326: `ta_v_fig4.gpkg:buffer_wgs84`.
5. Extract the average and standard deviation of the original map's red, green, and blue bands for each traditional authority using the zonal statistics algorithm: `ta_v_fig4.gpkg:r`, `ta_v_fig4.gpkg:rb` and `ta_v_fig4.gpkg:rbg`
6. Join the zonal statistics results to the `ta_v` layer by the ID_2 attribute: `ta_v_fig4.gpkg:ta_v_fig4`
7. Classify the results in a new field `orac` (original adaptive capacity) using the field calculator and CASE statements, choosing break points that classify most traditional authorities correctly.
8. Visually inspect results and edit the `orac` attribute for any mis-classified area.
9. The original map contains data in six conservation areas, noted with digitized point features in `ta_v_fig4.gpkg:fig4_errors`. Other areas are coded as follows:

code	description
-3	polygon too small to discern color or pattern fill
-2	white fill not matching any legend item
-1	pattern fill for "missing DHS data"
1	lowest adaptive capacity
2	...
3	...
4	highest adaptive capacity

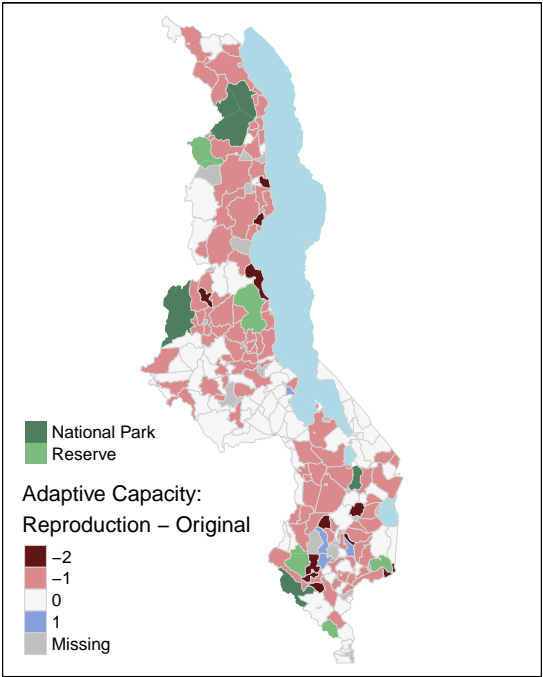
Original study figure 4 Load digitized figure 4 data and display counts of results. Convert all forms of missing data to NA to be excluded from mapping and statistics. Join original figure 4 adaptive capacity results to `ta_v`.

orac	n
-3	3
-2	30
-1	3
1	38
2	56
3	72
4	37

Map original figure 4.



Compare adaptive capacity result Calculate and map difference between the two maps.



##					
##		1	2	3	4
##	1	34	27	6	0
##	2	4	26	44	5
##	3	0	0	19	29
##	4	0	0	0	3

```
##
## Spearman's rank correlation rho
##
## data: ta_v$rpac_class and ta_v$orac
## S = 268637, p-value < 2.2e-16
## alternative hypothesis: true rho is greater than 0
## sample estimates:
##      rho
## 0.7891711
```

Vulnerability

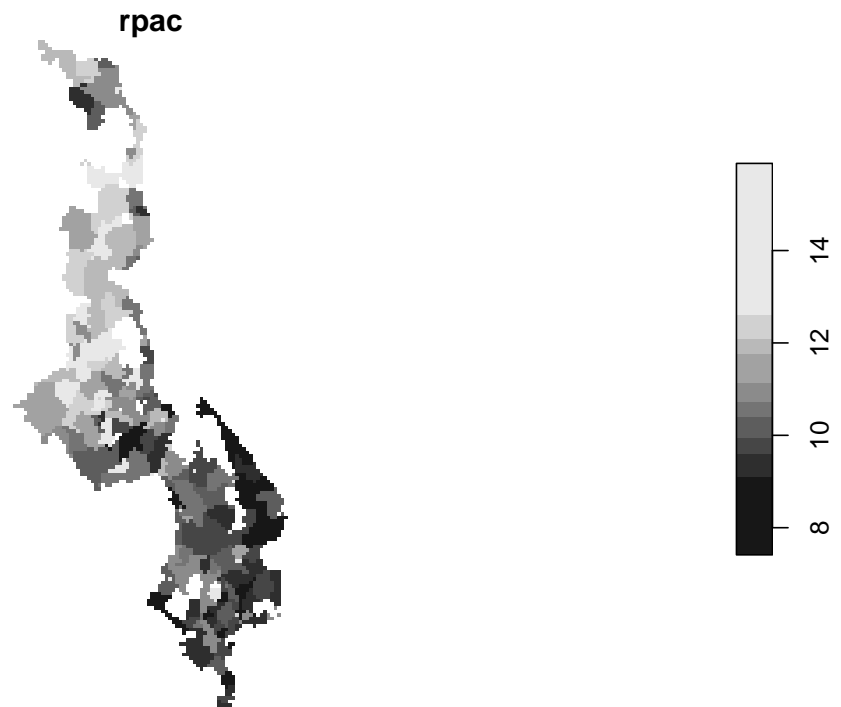
- 40%: Adaptive Capacity
- 20%: Livelihood Sensitivity (T2)
 - “sensitivity” defined as “degree to which a system will respond to an external disturbing force” (2.2) and referred to as:
 - * “livelihood sensitivity” (1.4, 3.3, 3.6, T1 theory, 4.6 formula, F5)
 - measured as a positive condition (4.6)
- 40%: Physical exposure (T2)
 - “exposure” defined as the “magnitude and frequency of forces that could stress a system” (2.2) and referred to as:
 - * “physical exposure” (1.4, 3.3, 3.7, 4.6 formula, T2)
 - * “biophysical exposure” (T1 theory)
 - * “exposure to floods and droughts” (F5)
 - measured as a negative condition (4.6)
- 100%: Household Resilience (T2)
 - “resilience” defined as “ability of a household to prepare for, respond to and recover from complex drivers of vulnerability” (2.2, 5.6) and referred to as:
 - * “household resilience” calculated as “Adaptive Capacity + Livelihood Sensitivity - Physical Exposure” (4.6 formula)
 - * “vulnerability to climate change” calculated as “assets + access + livelihood sensitivity - physical exposure” (F5)
 - * “vulnerability” (title, 3.3, 3.6, 4.3, 4.5, 6.5, 7.1, 7.2)

Extent and spatial resolution

Create bounding box representing the spatial extent of Malawi. Create a raster grid frame matching the extent of the bounding box and the spatial resolution of the drought exposure raster, which is 0.041667 decimal degrees. Although the flood risk raster has a coarser spatial resolution, visual inspection of the original figure 5 suggests that the finer spatial resolution of drought exposure was used for the original analysis.

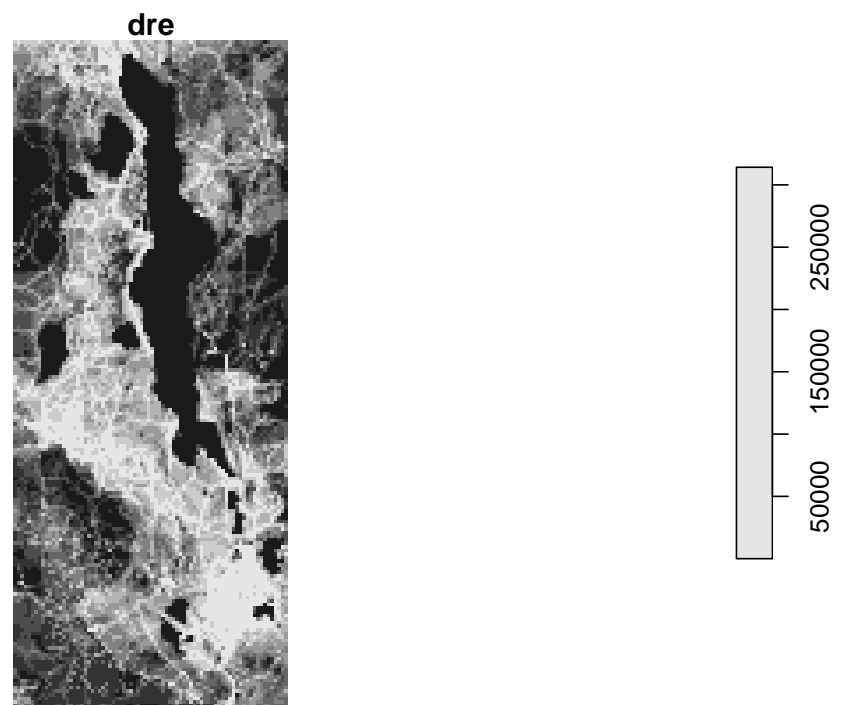
Adaptive capacity

Convert adaptive capacity to raster grid.

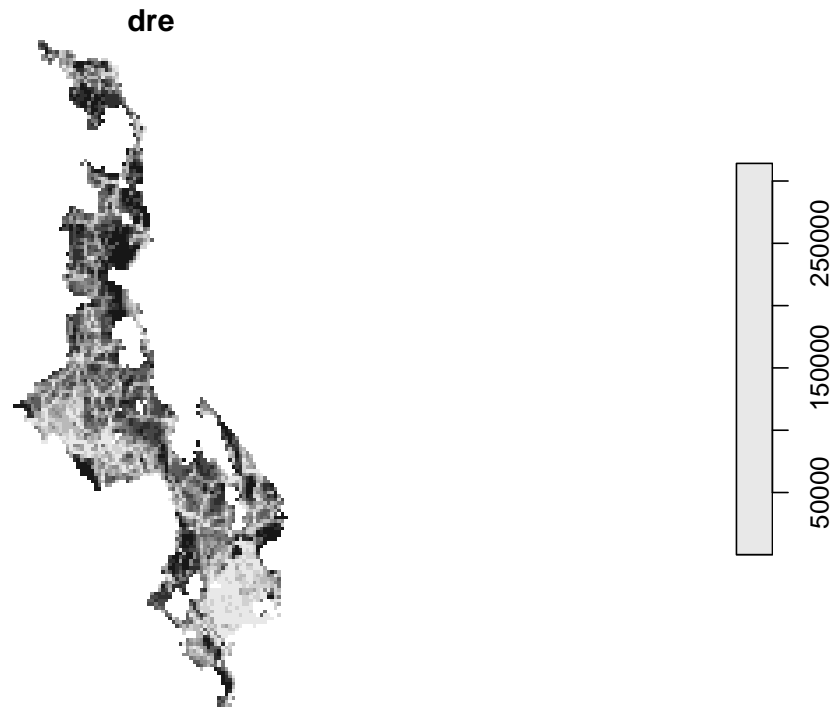


Drought exposure

Clip and warp drought exposure to match our extent and spatial resolution.



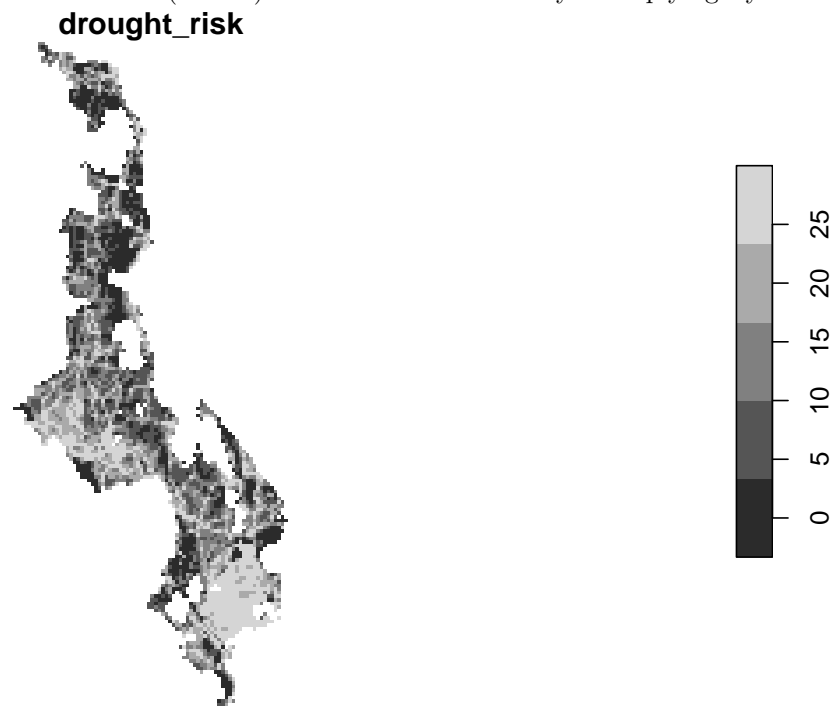
Create a mask with the adaptive capacity results so that lakes, conservation areas, and traditional authorities with no data will not skew the classification / rescaling of drought exposure. Apply this mask to drought exposure. Masking is our own decision based on intuition: it is not specified in the original publication.



I am concerned with the quality of the drought exposure data. It is essentially a population map, because although touted as a global drought risk, the units are expected average annual population (2007 as the year of reference) exposed. This raster is also double the resolution of the flood map. Therefore, I will scale the drought layer to be worth double what the flood layer is worth in the vulnerability calculation. This means drought will be worth 26.666% and flood 13.333%, as opposed to the previous 20% and 20%.

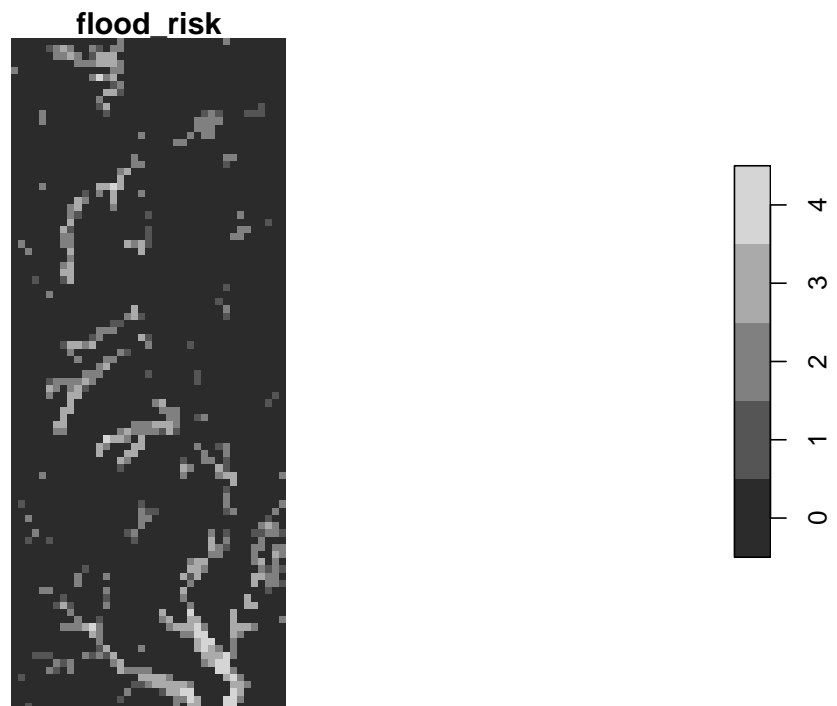
Here, you can set which percentages you want to use for the study.

Classify drought exposure into quintile classes (0 to 4) Then rescale to 20% by multiplying by 6.666.

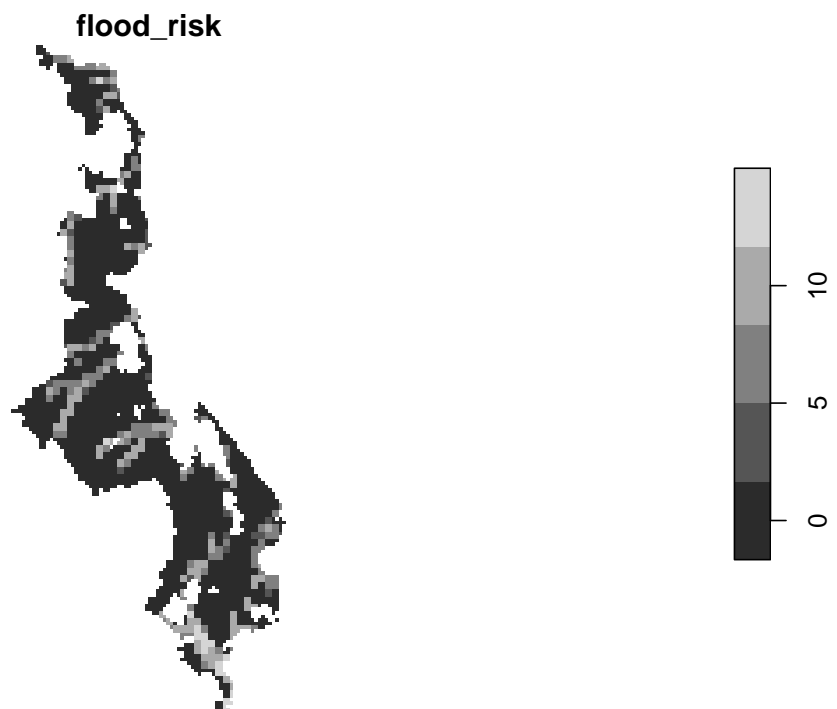


Flood risk

Clip and warp flood risk to match our extent and spatial resolution.



Mask and rescale flood. ## My CHANGE - MULTIPLY BY 3.333 INSTEAD OF 5 ### Since flood is already on scale from 0 to 4, simply multiply by 3.333 to achieve the 26.666% weight.



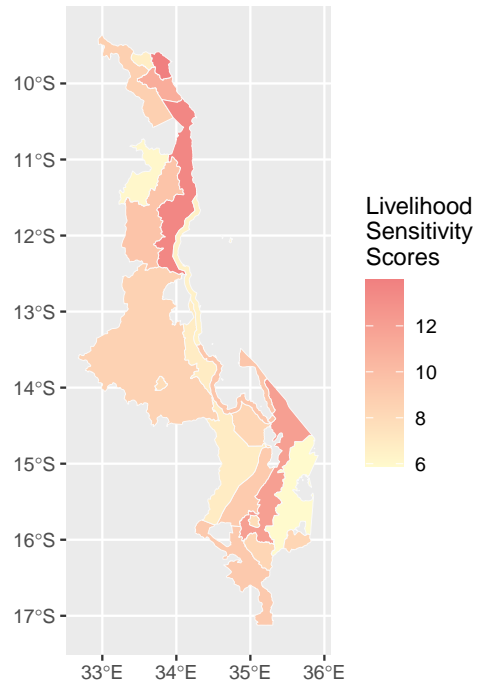
Livelihood sensitivity

Calculate livelihood sensitivity indicators from FEWSnet livelihood zone baseline profiles of poor households according to table 2.

Rescale livelihood sensitivity indicators into quantiles.

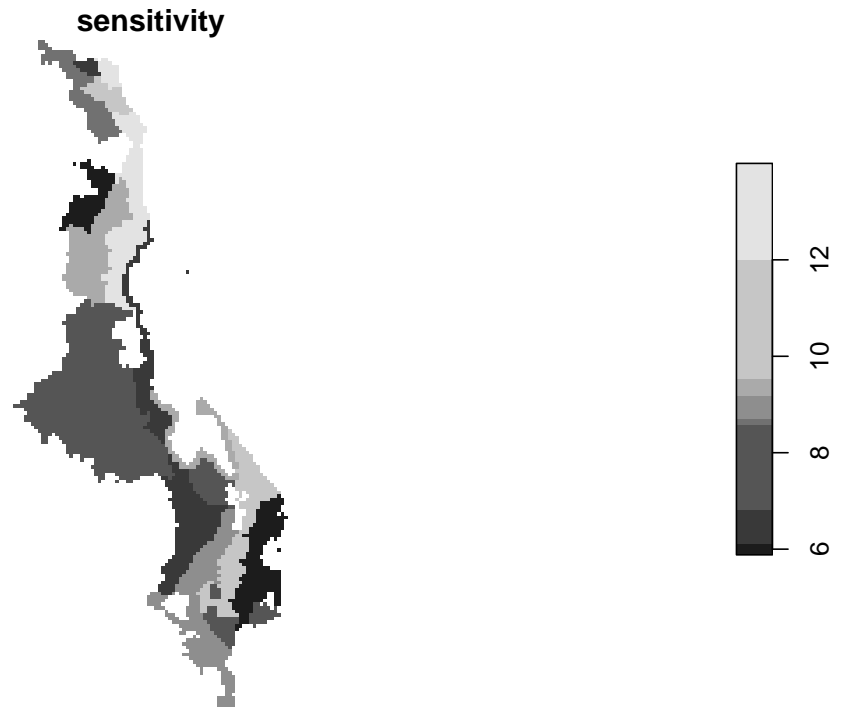
##	pctOwnCrop	pctIncWage	pctIncCashCrops	pctDisasterCope	ownCrop
## nbr.val	18.0	18.0	18.0	18.0	18.0
## nbr.null	0.0	0.0	13.0	1.0	1.0
## nbr.na	0.0	0.0	0.0	0.0	0.0
## min	29.4	9.7	0.0	0.0	0.0
## max	88.0	50.3	75.1	71.9	4.0
## range	58.6	40.6	75.1	71.9	4.0
## sum	1059.3	489.6	171.8	236.5	36.0
## median	55.0	24.7	0.0	8.8	2.0
## mean	58.9	27.2	9.5	13.1	2.0
## SE.mean	3.1	2.6	5.3	3.7	0.3
## CI.mean.0.95	6.6	5.5	11.2	7.9	0.6
## var	176.8	121.3	507.8	251.0	1.6
## std.dev	13.3	11.0	22.5	15.8	1.3
## coef.var	0.2	0.4	2.4	1.2	0.6
##	wageIncome	cashCropIncome	disasterCope		
## nbr.val	18.0	18.0	18.0		
## nbr.null	1.0	1.0	1.0		
## nbr.na	0.0	0.0	0.0		
## min	0.0	0.0	0.0		
## max	4.0	1.2	4.0		
## range	4.0	1.2	4.0		
## sum	36.0	17.6	36.0		
## median	2.0	1.2	2.0		
## mean	2.0	1.0	2.0		
## SE.mean	0.3	0.1	0.3		
## CI.mean.0.95	0.6	0.2	0.6		
## var	1.6	0.1	1.6		
## std.dev	1.3	0.4	1.3		
## coef.var	0.6	0.4	0.6		

Calculate aggregate livelihood sensitivity score



```
##          sensitivity
## nbr.val      18.00
## nbr.null      0.00
## nbr.na        0.00
## min           5.88
## max          14.00
## range         8.12
## sum          161.65
## median        8.65
## mean          8.98
## SE.mean       0.56
## CI.mean.0.95  1.18
## var           5.65
## std.dev       2.38
## coef.var      0.26
```

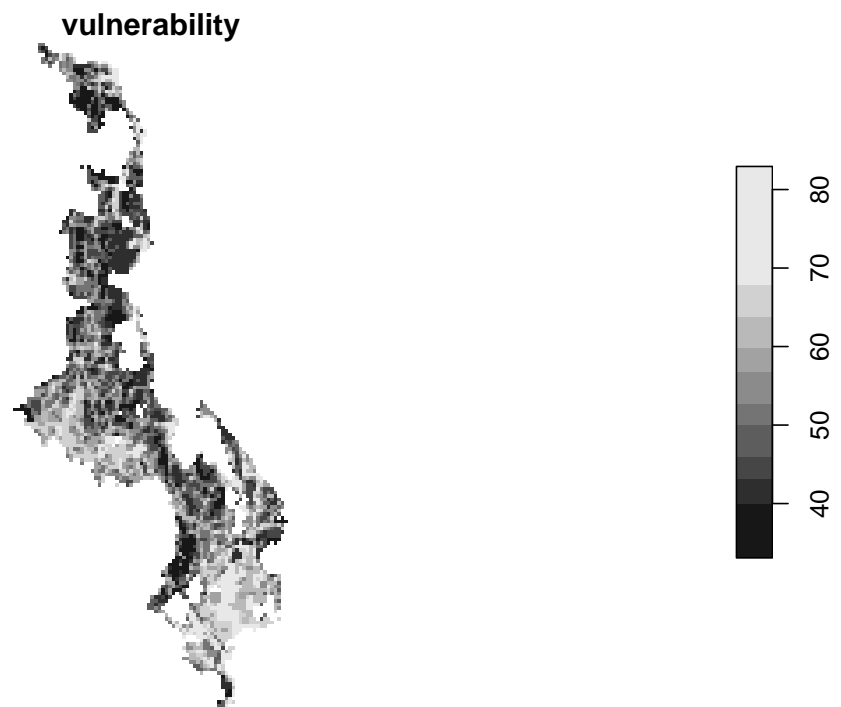
Convert livelihood sensitivity into raster grid



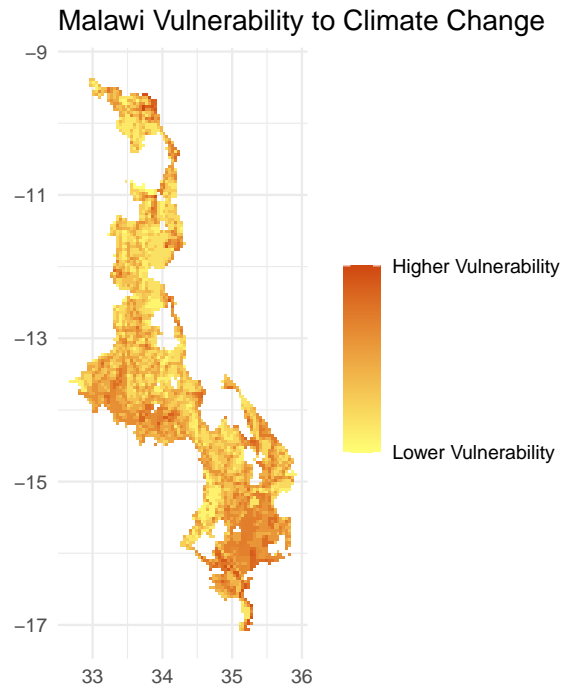
Vulnerability score

Calculate an aggregated vulnerability score by adding low adaptive capacity (invert adaptive capacity by subtracting from the maximum score of 40), livelihood sensitivity, drought exposure, and flood risk.

$$Vulnerability = (40 - AdaptiveCapacity) + LivelihoodSensitivity + DroughtExposure + FloodRisk$$



Reproduction figure 5



Evaluate vulnerability reproduction

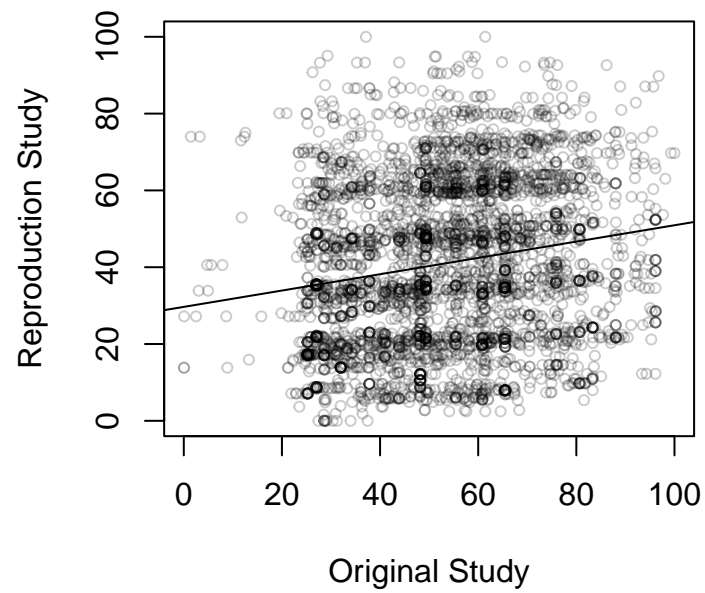
In order to compare the Malawi vulnerability results, we will georeference the original figure 5 map using QGIS georeferencer plugin. We will vectorize the UNEP-Grid raster input most closely matching the published map and summarize the red, green, and blue brightness values of the original map using zonal statistics. We will add the green and blue brightness values together to convert the original color ramp into a linear scale of continuous values. We will compare original and reproduction Malawi vulnerability results by creating a scatterplot, Spearman's Rho correlation coefficient (expecting a value near 1 for perfect positive correlation), and thematic map of the difference between the original results and replication results.

Original study figure 5 Comparing the reproduction of figure 5 with the original figure 5 requires first digitizing the original figure 5 (unclassified choropleth map with yellow to red gradient) in QGIS as follows:

1. Copy image of figure 5 from the original publication pdf file using Adobe Acrobat Pro
2. Paste the image and save as a .png file with pixel dimensions 1949 by 2811
3. Use QGIS 3.26.3 Georeference the map image to match `ta_v.gpkg` using WGS 84 geographic coordinates (epsg:4326). Use linear georeferencing with points in ...
4. Convert `ta_capacity.tif` raster to vector polygons
5. Extract the average blue and green bands from the georeferenced map image using `zonal statistics`
6. Save results as `georef_bg.gpkg`.

To approximate data values from the yellow to red gradient of the original map, the blue and green bands are then added, inverted, and rescaled to a range from 0 to 100.

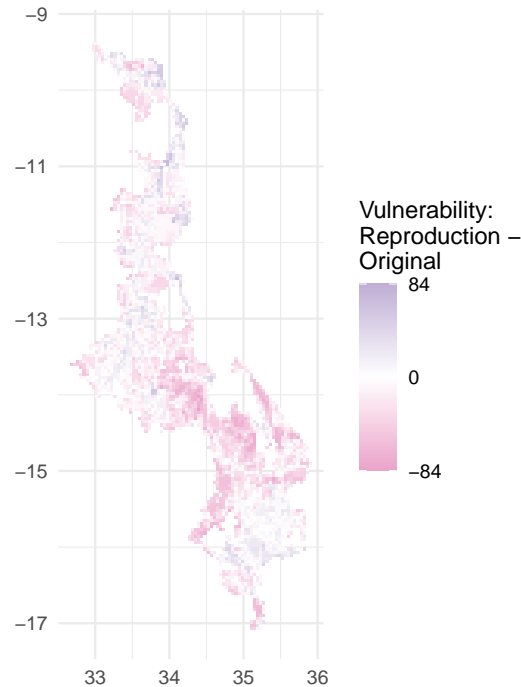
Vulnerability Scores



Compare vulnerability result

```
##  
## Spearman's rank correlation rho  
##  
## data: vulnerability_p$orv and vulnerability_p$rpv  
## S = 7058278811, p-value < 2.2e-16  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
##      rho  
## 0.2007671
```

Map differences in Figure 5



Discussion

When replicating this study, I am most concerned by the handling and resolution of the drought risk and flood risk raster layers. The cell size of these rasters is neither included in the metadata nor noted anywhere else in the replication. This seems like a large oversight, especially because these two layers make up 40% of the final vulnerability index, as much as all of the surveyed factors contributing to adaptive capacity, also at 40%. While I understand why the authors would have gone ahead with using these raster layers, I feel like the uncertainty they introduce, both in the low resolution and data source is glossed over.

The drought raster is essentially just a population map, because although touted as a global drought risk, the units are expected average annual population (2007 as the year of reference) exposed. This might be fine in the context of building an index concerned with number of people at risk or affected, but the vulnerability index is for the entire country of Malawi, and should not overrepresent areas with higher population.

The flood raster has its own problems, and is at a significantly lower resolution than all other data layers. When this layer gets baked into everything else and split into a higher resolution, a lot of uncertainty is introduced, made even worse by the fact that the layer is responsible for 20% of the total vulnerability.

While I do not know exactly how much weight these layers should have in the Household Resilience Model, I do intuitively think that, because the flood map is half the resolution as the drought map, their weights should not be valued equally (drought map cell size is 4.6 and flood map cell size is 9.5km). So, just to see how much impact a change in the weight of the drought map would have, I will weight the drought map as double the weight of the flood map. To make up the total 40%, this means I will assign 13.333% to the flood map, and 26.666% to the drought map.

I am not certain that drought should make up exactly 26.666% of the vulnerability index, and flood 13.333%. Rather, I am just pointing out the discrepancy in the precision of the rasters in two layers that are weighted the same amount in the model. On top of this, it is worth reminding the reader that the drought raster is just a population map, and not a measure of rainfall across the world.

Conclusions

By comparing my modified final map with the original, it is clear that the weighting does make a difference to the overall vulnerability. While more could be said with statistics about how much difference, it seems that weighting the drought layer doubly compared to flood layer increases overall vulnerability in the country.

Integrity Statement

This report and its preregistration were written after already attempting the reproduction study, including acquisition and analysis of all of the secondary data sources required. However, the preregistered analysis plan was written *as if* we had no prior knowledge of the data other than what is documented in the study. Holler has previously reviewed and compared other climate vulnerability models for Malawi, and conducted a scoping study in the Lilongwe and Mangochi districts of Malawi in 2015, including meeting with the Regional Centre for Mapping of Resources for Development (RCMRD) consultants who created the Malawi Hazards and Vulnerability Atlas (2015).

References

Referencing the original paper

Malcomb, D. W., E. A. Weaver, and A. R. Krakowka. 2014. Vulnerability modeling for sub-Saharan Africa: An operationalized approach in Malawi. *Applied Geography* 48:17–30. DOI:10.1016/j.apgeog.2014.01.004](<https://doi.org/10.1016/j.apgeog.2014.01.004>).

Sections

1. Introduction
2. Complex vulnerability
3. Evidence-based Indicators
4. Methodology
5. Results
6. Discussion
7. Conclusion

Tables, figures, other elements

- T1 Evidence-based complex vulnerability indicators
- T2 Weighted indicators by metatheme
- F1 Map of Malawi
- F2 Vulnerability web
- F3 Malawi Household Resilience (2004)
- F4 Malawi Household Resilience (2010)
- F5 Malawi Composite Vulnerability Index
- A1 Appendix 1
- R References

Other References

Barrett, S. 2014. Subnational Climate Justice? Adaptation Finance Distribution and Climate Vulnerability. *World Development* 58:130–142. DOI: 10.1016/j.worlddev.2014.01.014. Gallopín, G. C. 2006. Linkages Between Vulnerability, Resilience, and Adaptive Capacity. *Global Environmental Change* 16 (3):293–303. DOI: 10.1016/j.gloenvcha.2006.02.004. Rufat, S., E. Tate, C. G. Burton, and A. S. Maroof. 2015. Social vulnerability to floods: Review of case studies and implications for measurement. *International Journal of Disaster Risk Reduction* 14:470–486. DOI: 10.1016/j.ijdr.2015.09.013. Smit, B., and J. Wandel. 2006.

Adaptation, adaptive capacity and vulnerability. *Global Environmental Change* 16 (3):282–292. DOI: 10.1016/j.gloenvcha.2006.03.008.