

Pre-Registration of Rpr - Reproduction of Malcomb et al 2014 Vulnerability modeling for sub- Saharan Africa

Joseph Holler, Department of Geography, Middlebury College, Middlebury VT 05753 Drew An-Pham, Department of Geography, Middlebury College, Middlebury VT 05753 Kufre Udoh, Department of Geography, Middlebury College, Middlebury VT 05753 Peter Kedron, School of Geographical Sciences & Urban Planning, Arizona State University, Tempe AZ 85281

Version 1.0 | Created June 25, 2021 | Last Updated June 25, 2021

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).

Sampling plan

Existing data and data exploration

This registration is based upon knowledge of the original study based upon a thorough reading of the original research article only.

In full transparency, we have already attempted the reproduction study, including acquisition and analysis of all of the secondary data sources required. This preregistration serves two purposes for us:

1. Revisiting the original paper to document the most complete knowledge of the data and methodology possible without acquiring data or beginning analysis, represented in this preregistration.

2. Practicing and demonstrating the full workflow of a reproduction study in the human-environment and geographical sciences, including preregistration, reproducible computational research practices, and publication of a reproduction report.

Therefore, we have not taken steps to not explore or analyze the data. Rather, we are practicing the intellectual challenge of practicing a scenario in which we are preregistering the reproduction study as if we had only read the original 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).

Data collection and spatial sampling

The original study included over 70 qualitative semi-structured interviews with officials from development agencies, government, and non-governmental organizations (1.4). 47 specific interviewee titles are listed in a "full list of the interviews" (A1), leaving uncertainty around the remaining 23 (or more) interviewees. The study included 11 village focus groups (1.4). The interviews and focus groups were selected to focus on villages and organizations connected at different levels of organization to "externally designed adaptation projects" (3.2). We presume that the interviewees and focus groups were selected based upon their association with externally designed and funded climate change adaptation projects, which are neither equally nor justly distributed (Barrett 2014). No minimum sample size or stopping criteria were defined.

The original authors used exclusively secondary data for the vulnerability models and did not subset or sample from the secondary data. The published results based on DHS surveys include data 203 traditional authorities in 2010 (F4), whereas the authors state that there are more than 250 populated traditional authorities in Malawi (4.4). The 2010 household resilience data is based upon 24,850 DHS household surveys (5.2). Furthermore, "not every traditional authority had surveys conducted within its administrative boundaries" (5.2). This suggests possible spatial sampling problems in the DHS survey data in the context of its application in this study. The authors' rationales for using the third level administrative units of traditional authorities include matching the political level at which many projects are planned and assessed, and identifying hotspots of vulnerability that are lost in the aggregation to second level districts (4.4). The original paper does not contain any further detail on the spatial sampling or distribution of DHS surveys *vis a vis* traditional authorities.

Data description

Household resilience 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).

The 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.

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)

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).

Variables

All variables used in this study are derived from secondary data. There are no experimentally manipulated variables. Eighteen independent variables are represented as indicators and drivers of four component themes of vulnerability/resilience: 1) *assets* and 2) *access* combine to create *adaptive capacity*, 3) *livelihood sensitivity*, and 4) *physical exposure*. Although the eighteen indicators are often discussed as driving vulnerability or resilience outcomes, they are not formally used as **predictor** or **response** variables in any statistical tests.

The variables are presented below as described in various sections of the original paper, with the weighted values assigned to them for analysis of household resilience and climate vulnerability. Text is copied almost verbatim from the original study to facilitate finding the correct variables in data.

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)

Livelihood sensitivity

- 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)

Physical exposure

- 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)

Aggregate thematic concepts

The variables 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)
- 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)

The authors use the terms and definitions of "adaptive capacity", "vulnerability" and "resilience" inconsistently in the paper. Depending on context, both "vulnerability" and "resilience" may refer only to the assets and access portion of the model or to the full model inclusive of adaptive capacity (assets and access), livelihood sensitivity, and exposure. In some sections and formulas, "adaptive capacity" is used, while in others it appears as its components (assets and access) or even as "household resilience". The data mapped in figures 3 and 4 are described in different sections of the original paper as resilience, adaptive capacity, and/or vulnerability. The data mapped in figure 5 are described in different sections of the article as vulnerability or resilience. Confounding matters further, the authors' definition for resilience is typical for resilience theory, but the formula for

resilience is more typical of the well-known Intergovernmental Panel on Climate Change (IPCC) operationalization of "vulnerability" (Gallopín 2006, Smit and Wandel 2006).

For the purposes of the reproduction study, the data visualized in figure 4 of the original study and referred to as 40% Adaptive Capacity in table 2 will be referred to as **adaptive capacity** while the data visualized in figure 5 of the original study and referred to as 100% Household Resilience in table 2 will be referred to as **vulnerability**.

Attribute variable transformations

The original study's description of attribute variable transformations is confusing, and we attempt to present all of the evidence in the original study clearly below. All variables are normalized between zero and five (4.3 and 5.6) with zero representing the worst or poorest condition and five representing the best or richest condition (4.3). The normalization method is not described, but the poorest and richest conditions are described as "quintiles" with values of zero and five (4.3). "Quintiles" suggests classification into five classes with equal counts or frequencies. "Normalization" suggests transforming each variable into a normal distribution. A minimum of zero and maximum of five suggests rescaling the data to a range from zero to five or classifying the data into six quantiles assigned integer values zero through five. Four variables were described as (Y/N) nominal data (T1), for which it is not clear how to transform the data into ordinal data with more than two classes. A similar concern arises for the market access variable with three classes (Rural, Peri-urban, and Urban). Regardless, there is ambiguity in the method of normalizing, scaling, or classifying each variable.

Adaptive capacity is analyzed with the weighted combination of all asset theme indicators and access theme indicators. Weights are given in the Variables section. The formula for this combination is not specified. For 2004, the authors report a minimum household adaptive capacity score of -0.80 and maximum of 39.33 (5.2). A maximum score near 40 is intuitive as a theoretical maximum of 40% for the adaptive capacity category. It is not clear whether the calculation should be a weighted average, weighted sum, weighted combination specified to achieve a maximum of 40%, or some other form of weighted combination. Results of the average adaptive capacity score for traditional authorities in 2010 are visualized in figure 4.

Livelihood sensitivity is presumably calculated with the weighted combination of its four indicators. Weights are given in the Variables section. The formula for this combination is not specified. Results for this theme are not presented.

Physical exposure may be calculated with the weighted combination of its two indicators. Weights are given in the Variables section. The formula for this combination is not specified. It is not specified whether any geographic transformation is required to combine these two variables. Results for this theme are not presented.

Geographic transformations

Adaptive capacity is analyzed in the spatial units of traditional authorities (4.4, 5.2, F3 and F4). This aggregation from households to traditional authorities is accomplished with a spatial join (5.2) with an average of the individual household resilience scores (5.2, 6.3, F3 and F4), and classified into four classes with the Natural Breaks Jenks classification

method (F3 and F4). In the methodology section, the authors state that "DHS Indicators were disaggregated to the village level" (4.4). Since the DHS data is described as using the household level of aggregation, it is not clear how or why the data would be "disaggregated" to the village level.

Vulnerability is analyzed in the spatial units of gridded cells (F5). Each theme was converted to the raster grid data format (4.6). The paper does not specify the parameters for raster conversion, including the relationship between vector polygons and raster grid cells or the spatial resolution of the grid cells. Presumably, only adaptive capacity and livelihood sensitivity were converted, since physical exposure is already in a raster grid format. The paper does not specify if or how any geographic transformation is required for the biophysical risk grids from the UNEP Global Risk Data Platform. It may be possible to infer the resolution and methods from close inspection of the final climate vulnerability map (F5).

Analyses

The final climate vulnerability analysis is calculated with map algebra on a raster grid for each theme using the formula:

```
household resilience = adaptive capacity + livelihood sensitivity - physical exposure
```

(4.6).

The results are presented as a continuous raster grid with a continuous color gradient. No descriptive statistics of the results are provided.

Geographical characteristics

The **coordinate reference system(s)** used in the original study are not specified. However, the study does not appear to include any distance or area calculations: therefore the analysis should not be sensitive to the coordinate reference system used as long as each layer is stored or reprojected into one consistent system.

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 **first order spatial effects**, **second order spatial effects**, or **spatial anisotropies**.

Temporal characteristics

The **temporal extent** of the original study was stated as "2004–2010 based on the availability of the Malawi DHS datasets with GPS data" (4.5). The study also references Dartmouth Flood Observatory Data from 1999 to 2007 (F5) used to indicate flood risk and a Malawi Vulnerability Assessment Committee (MVAC) Household Economy Approach (HEA) baseline survey conducted between May and July 2003 used to indicate livelihood sensitivity (3.6). Therefore, it appears that the 2004–2010 temporal extent applies strictly to the household resilience analysis, and not to the climate vulnerability analysis.

The **temporal support** for the household resilience analysis was longitudinal DHS Survey data collected in 2004 and 2010. The **temporal support** for the climate vulnerability analysis is an aggregation of data from different sources, ranging from 1999 to 2010 (F5). **Temporal effects** are not measured or accounted for, although the authors discuss differences between household resilience in 2004 and 2010 (5.3 and 5.4)

Data exclusion

Some traditional authorities are **missing data** for DHS surveys because no DHS surveys were conducted within their administrative boundaries. These traditional authorities were excluded from the analysis of household resilience and labelled as 'Areas Missing DHS Data' (5.2). No **inferences** were made to fill the missing data (5.2). Some traditional authorities are clearly symbolized with the diagonal hashmarks representing missing data for household resilience (F4), and appear to be similarly excluded from the analysis of climate vulnerability (F5). However, several other traditional authorities appear white—not one of the four categories of household resilience (F4). Some of these areas appear to be excluded from the climate vulnerability analysis while others are not (F5).

The study does not analyze the presence of **outliers** or exclude them. The study does not **weight samples**.

Analytical specification

There are no hypothesis tests requiring explanation of analytical specifications.

Inference criteria and robustness

There are no hypothesis tests requiring explanation of inference criteria or robustness.

Exploratory analyses and contingency planning

The entire study is an **exploratory analysis** in soliciting local Malawian's opinions and perspectives on climate change vulnerability and constructing descriptive geographic vulnerability models based upon those opinions and perspectives. The authors do not

discuss any pre-described **contingency plans**, but do describe "several iterations of the model using alternative weighting schemes result[ing] in a final product that best reflects Malawi's contextual and perceptual vulnerability—a stated goal" (4.3). It is unclear whether the model iterations were planned or unplanned.

Reproduction study design

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.
2. Adaptive capacity scores will be calculated for each household, and then household scores will be spatially joined by traditional authority and averaged.
3. Figure 5 represents vulnerability, composed of adaptive capacity, livelihood sensitivity, and physical exposure.
4. 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.
5. High ranks (4) will be assigned to better and safer conditions for each indicator.
6. 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
- 7. Each thematic indicator will be rasterized or resampled to the UNEP/GRID data input most closely resembling the spatial resolution of figure 5.
- 8. 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)
- 9. Any traditional authority missing adaptive capacity data from DHS surveys will be removed / masked from the final vulnerability analysis.

Evaluating the reproduction results

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.

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.

We expect relatively stronger correlations for the adaptive capacity analysis than the vulnerability analysis because of some uncertainty in digitizing correct continuous values from a georeferenced image of the figure 5 map, particularly when considering the influence of text and other symbology on the map in potentially altering effects.

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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1016/j.gloenvcha.2006.03.008).