

Infrastructure, Borders, and Conflict in East Africa

Project Analysis Plan

GISC412

Advanced Methods in Geospatial Data Science

Thomas Waldin

17775654

University of Canterbury

Date Submitted: 21 August 2025

1. INTRODUCTION

The aim of this report is to present a topic for investigation and discuss its feasibility as a GISC412 final project. The datasets, methods for analysis, and research questions will be discussed.

The research question that will be discussed is as follows:

‘Investigate how infrastructure, urbanization, and borders influence the spatial distribution of conflict events in East Africa, and whether these relationships vary across space.’

More specifically the project will assess:

- Where are conflict events most concentrated in East Africa?
- Does proximity to borders increase likelihood of conflict?
- Are conflicts more strongly associated with roads and urban areas in some regions than others?
- How do these relationships vary spatially?

For the study, Kenya, Uganda, Ethiopia, and South Sudan are considered East Africa. The study will focus on the last decade (2014-2024).

2. DATASET DESCRIPTIONS

Multiple geospatial datasets will be integrated to explore the drivers of conflict in East Africa.

The Armed Conflict Location and Event Data (ACLED) dataset provides detailed records of political violence and includes location through lat/long, date, event type, actors, and estimated fatalities. Spatial aggregation will be used to create an outcome variable, likely rate of conflict, which may need to be log-transformed.

Population data will be obtained from WorldPop (2020) as a raster which can be used to normalize to per-capita rates. Aggregation to match the chosen spatial units will be required.

Infrastructure data will be obtained from OpenStreetMap. Density of roads could be calculated as total length per square kilometre.

NOAA data of night-time lights measured with Visible Infrared Imaging Radiometer Suite (VIIRS) will be used as a measure of urbanization. Raster values will likely be averaged by each spatial unit.

Borders and boundaries will be obtained from GADM to define country and administrative borders.

All datasets will need to be projected into a consistent coordinate reference system.

3. ANALYTIC TECHNIQUES

All spatial data will need to be reprojected to the same coordinate system and aggregated to the chosen spatial units. All data will need to be cleaned.

Exploratory data analysis will be undertaken, including choropleth maps for spatial variables and global Moran's I to examine spatial autocorrelation. Correlation and relationships between variables will be examined using scatter and correlation plots.

Geographically weighted regression (GWR) will be used to model the varying relationship between conflict and predictors.

Local coefficient maps and a local r-squared map will be produced from the GWR and can be compared with a hotspot map of conflicts.

4. FEASIBILITY

While this project is ambitious there are several opportunities to reduce the scope if it proves infeasible. It is difficult at this stage to accurately assess the workload. If internal deadlines are not met, it may be sensible to reduce the scope in the following ways:

- Reduce years of study.
- Reduce number of countries.
- Reduce potential predictors.

5. CONCEPTUAL WORKFLOW AND TIMELINE

The following internal deadlines have been set:

- The data should be obtained and cleaned by 7th September (two weeks).
- Processing and aggregation should be completed by 21st September (two weeks).
- Exploratory spatial data analysis should be completed by 28th September (one week).
- Modelling and write-up will be done by the due date (four weeks).

Figure 1 demonstrates the workflow in a visual flowchart.

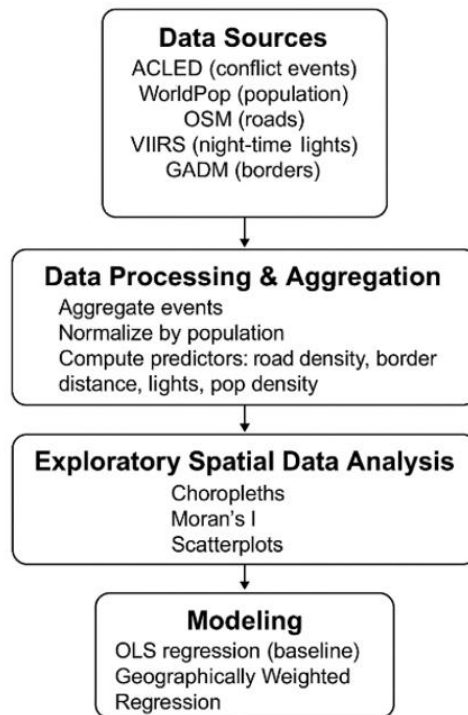


Figure 1. *A visual representation of conceptual workflow. This flowchart was made with assistance from generative AI.*