Modelling Social Unrest with Latent Variables

Jasmine Burgess[1], Gabriel Wallin[1], Rachel McCrea[1] and Matt Thomas[2]

[1]Lancaster University [2]British Red Cross

Social Unrest

Social unrest in Great Britain includes:

- Rioting
- Violence between police and protesters/rioters
- Attacks on asylum hotels

In Great Britain, the majority of recent incidents of social unrest have been motivated by anti-migrant sentiment. British Red Cross is interested in understanding and predicting long-term risks of social unrest.

Causes of Social Unrest

Social unrest typically has multiple interconnected causes, which are noisy and hard to measure directly, with many relevant indicators not systematically recorded. Information will be spread across diverse sources, so we will need to collect high-dimensional spatial data.

Latent Variable Models (LVMs) are a method for understanding unobservable processes. We construct underlying latent variables which generate the observed data, explain its variation and allow us to investigate social unrest risk at different locations. We will develop new LVMs incorporating different types of data to understand the causes of social unrest and provide early warning detection of impending crises.

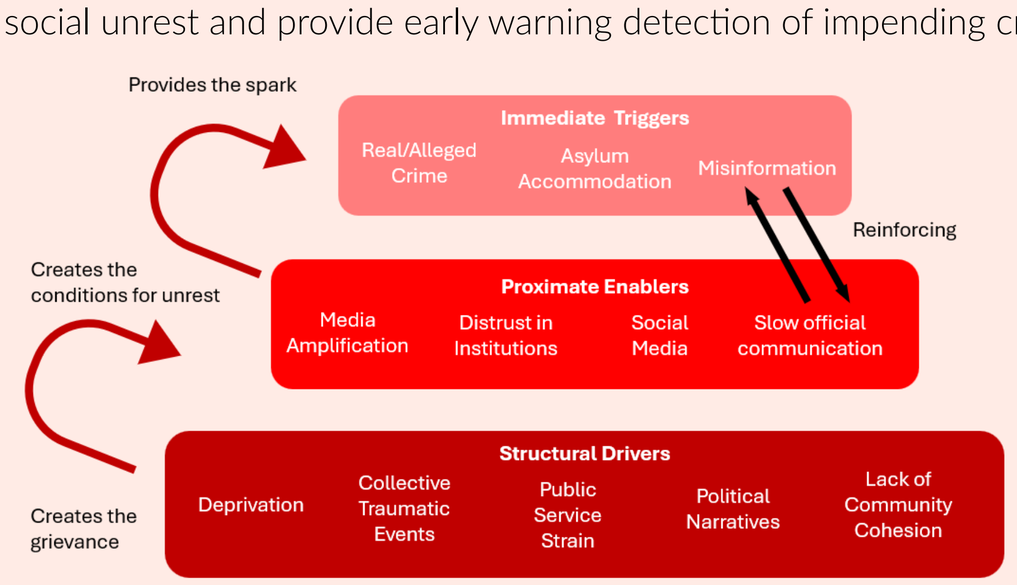


Figure 1. Representation of the Different Types of Causes of Social Unrest

Although the triggers of social unrest or rioting may be unpredictable, there are long-term structural drivers which mean that certain areas are at higher risk. For example, economic deprivation is a latent factor linked to social unrest.

A Spatial Latent Variable Model

Chacón-Montalván et al. (2025) develops a spatial latent variable model for mapping social concepts:

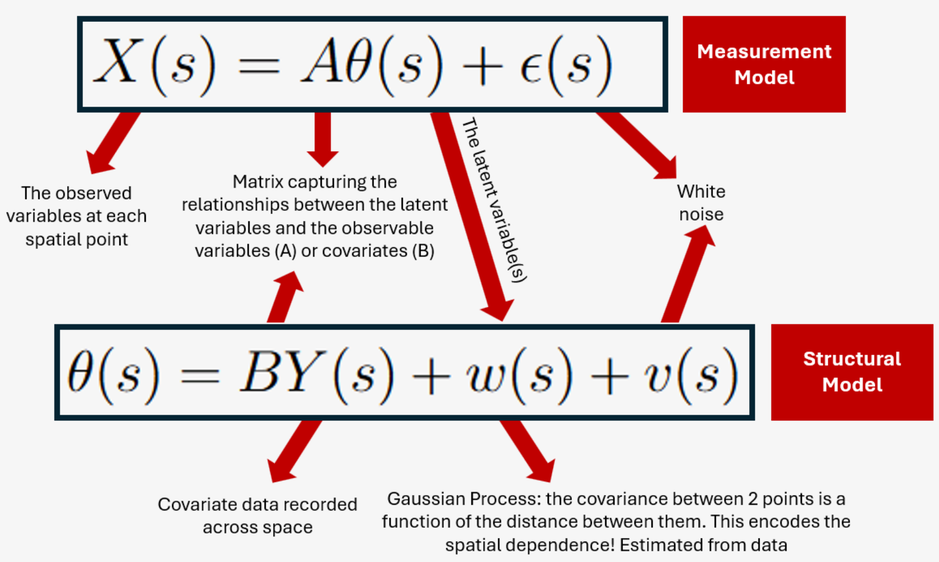
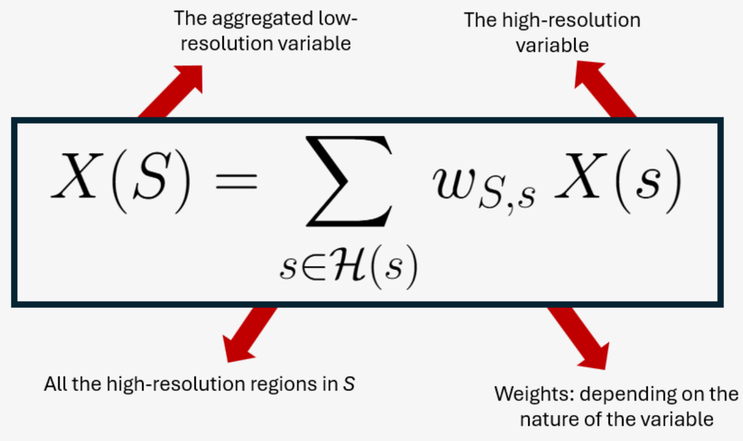


Figure 2. A model for mapping food insecurity from Chacón-Montalván et al. (2025).

The Mixed Resolution Model

One challenge is the relevant data is recorded at different resolutions. For example, crime data is available at exact longitudes and latitudes, unemployment data is available at local authority level, and most survey data is only available at region level. We propose:

1. The structural equation stays the same, with a single latent factor defined at the high-resolution.
2. We introduce a hierarchical measurement model.
3. Data at any resolution is linked to the latent factor through this model.
4. For the lower resolution measurement models, we use aggregated latent variables and aggregate the high resolution data.
5. The likelihoods of the different resolution models all inform the posterior distribution of θ, allowing all data to be used in estimation.



- For count data (e.g. crime numbers), weights may be proportional to population.
- For continuous data (e.g. deprivation indices), weights may be uniform or proportional to area.

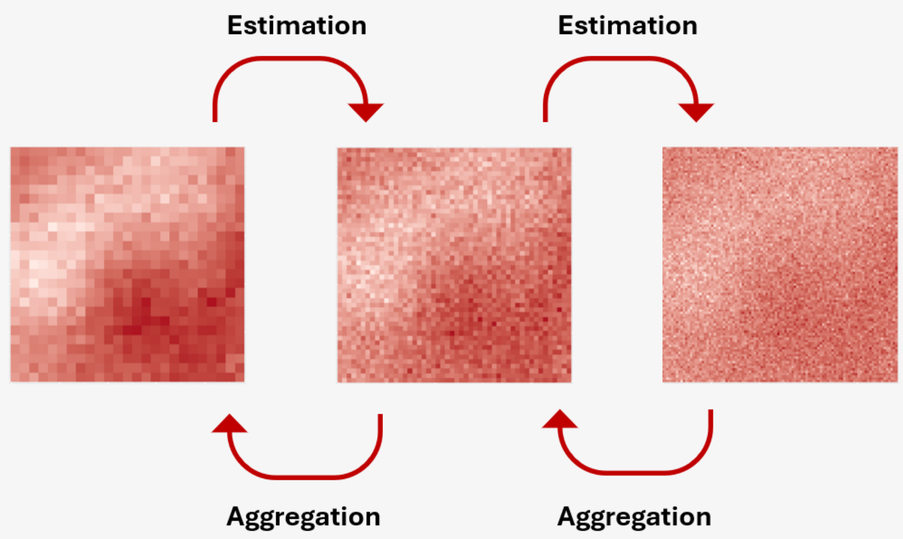


Figure 3. The process of transitioning between different spatial resolutions. Here the data is aggregated through unweighted means.

Research Direction

- We will fit our mixed resolution model with historical structured data (including counts, continuous variables and binary indicators) to test its predictive performance and identify which variables are most informative.
- We will later use **unstructured and text data** such as news articles, government reports and social media posts, alongside structured data.
- This will allow us to access high-frequency data for detecting early warning signs for social unrest.
- Text data is likely to be lower-resolution. For example, Google search data is available only at country level (England, Wales, Scotland).
- Natural Language Processing will be needed to convert text into a suitable form for LVMs.

Chacón-Montalván, E., Parry, L., Giorgi, E., Torres, P., Orellana, J., Moraga, P., and Taylor, B. (2025). Mapping food insecurity in the Brazillian Amazon using a spatial item factor analysis model. *Annals of Statistics*

jasmine.burgess@lancaster.ac.uk https://www.lancaster.ac.uk/stor-i-student-sites/jasmine-burgess/