Dynamic hazard risk mapping – Flooding risk in Ghana

Date: April 2023

Flowminder Foundation

Dynamic hazard risk mapping - Flooding risk in Ghana

Date: April 2023

Flowminder Foundation

All estimates and indicators presented here have been derived from pseudonymised Call Detail Records (CDRs), which have been aggregated and anonymised. Explanations of the figures and the methodology used to produce them are included to give context.

Data protection & privacy

No personal data, such as an individual's identity, demographics, location, contacts or movements, is made available to the government or any other third party at any time. All results produced by the Flowminder Foundation are aggregated results, which means that they do not contain any information about individual subscribers. This data is fully anonymised. This approach complies with the European Union's General Data Protection Regulation (EU GDPR 2016/679).

In addition, no commercially-sensitive information (of Vodafone Ghana) is made available by those results.

About the project & its partners

Authors

This report was authored by the **Flowminder Foundation**, by Thomas Smallwood and Véronique Lefebvre.

Acknowledgements

This study was made possible thanks to the anonymised data provided by Vodafone Ghana, and the Data for Good partnership.

The **Data for Good** project is a unique public-private partnership and not-for-profit initiative between **Ghana Statistical Service** (GSS), **Vodafone Ghana** and **Flowminder Foundation**, to support evidence-based decision-making for the wellbeing of all in Ghana. The project received funding support from Vodafone Foundation and the William and Flora Hewlett Foundation.

Introduction

Hazard risk analysis can include a range of different factors, including:

- The probability of an area being impacted by a given hazard
- The number of people in an area who would be exposed to a given hazard
- The vulnerability of an area to a given hazard, which may include dimensions such as:
    - Socio-economic vulnerability
    - Presence of vulnerable groups
- The capacity for authorities and institutions to respond to a given hazard.

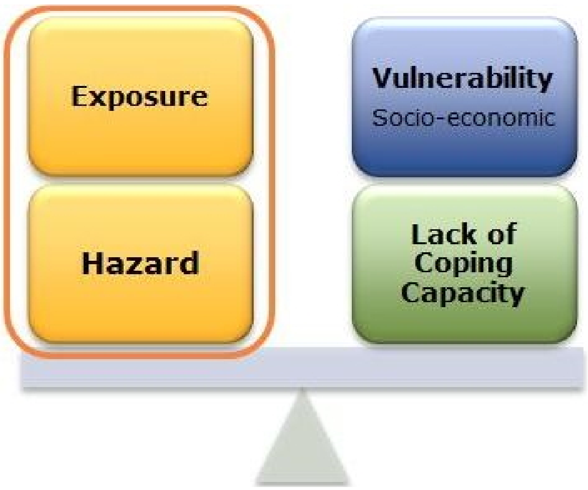

Figure 1: The INFORM Risk model balances two major forces: the hazard & exposure dimension on one side, and the vulnerability and the lack of coping capacity dimensions on the other side.
Credit: DRMKC - INFORM

The number of people who may be exposed to a hazard is an essential component of risk analysis. However, people regularly move between different areas both in terms of short trips (e.g. daily travel to work) and longer-term relocations or 'changes in residence' (e.g. migrations). This results in variations in the number of people in an area over time, both in the short-term (hourly, daily) and in the long-term (monthly, seasonal), and therefore how many people may be exposed to a hazard.

High-frequency data on mobility is difficult to obtain using traditional methods, such as surveys and censuses, due to the cost and resources associated with continuously collecting new data. In comparison, call detail records (CDRs) are routinely collected by mobile network operators for billing purposes automatically and in near-real time. As CDR data contain the necessary spatial and temporal features to study mobility, these data provide an exciting opportunity to understand population distributions and mobility at higher temporal resolution than traditional methods.

The purpose of this work is to derive dynamic indicators of hazard risk by combining static geospatial data on hazards and vulnerability, the two other components of the INFORM Risk model, with dynamic exposure indicators derived from aggregated and anonymised CDR data.

By incorporating the variation over time in the number of people present or residing in different areas, we can provide novel insights into how hazards risks vary over time: hourly, daily or seasonally. These insights can support improved decision-making for more effective disaster preparedness by supporting activities such as contingency planning, pre-positioning resources, and simulating disaster scenarios.

In this report, we demonstrate the application of our framework for combining anonymised CDR aggregates with static population data and geospatial data on hazards and vulnerability to the case study of **hourly variation in flooding risk in Ghana**. We have an ongoing partnership with Ghana Statistical Services (GSS) and Vodafone Ghana to support the use of CDR analytics by the Ghanaian government. Within this structure, we are working with GSS and Ghana's National Disaster Management Organization (NADMO), who requested that Flowminder produces dynamic indicators of hazard risk, specifically on flooding and drought.

Dynamic flooding risk indicators

Using the methodology described below, we produced a range of hazard indicators for flooding in Ghana. These include a hazard risk indicator and a percentage change in hazard risk indicator. We calculated hourly hazard indicators at district-level for the whole country and at urban division-level for Accra, for weekdays, Saturdays, and Sundays. Further intermediary indicators for dynamic population density and hazard exposure are included in the Supplementary Information section below.

Focussing first on Accra, an area with high flood hazard and vulnerability which also experiences large hourly variation in exposure, we can observe substantial temporal variation in flooding risk. In Figure 2, we can see increases in flooding risk in the central areas of Accra (highlighted in red) in the middle of the weekday, and to a lesser extent in the middle of the day on Saturday. In comparison, there is less change in risk on Sundays.

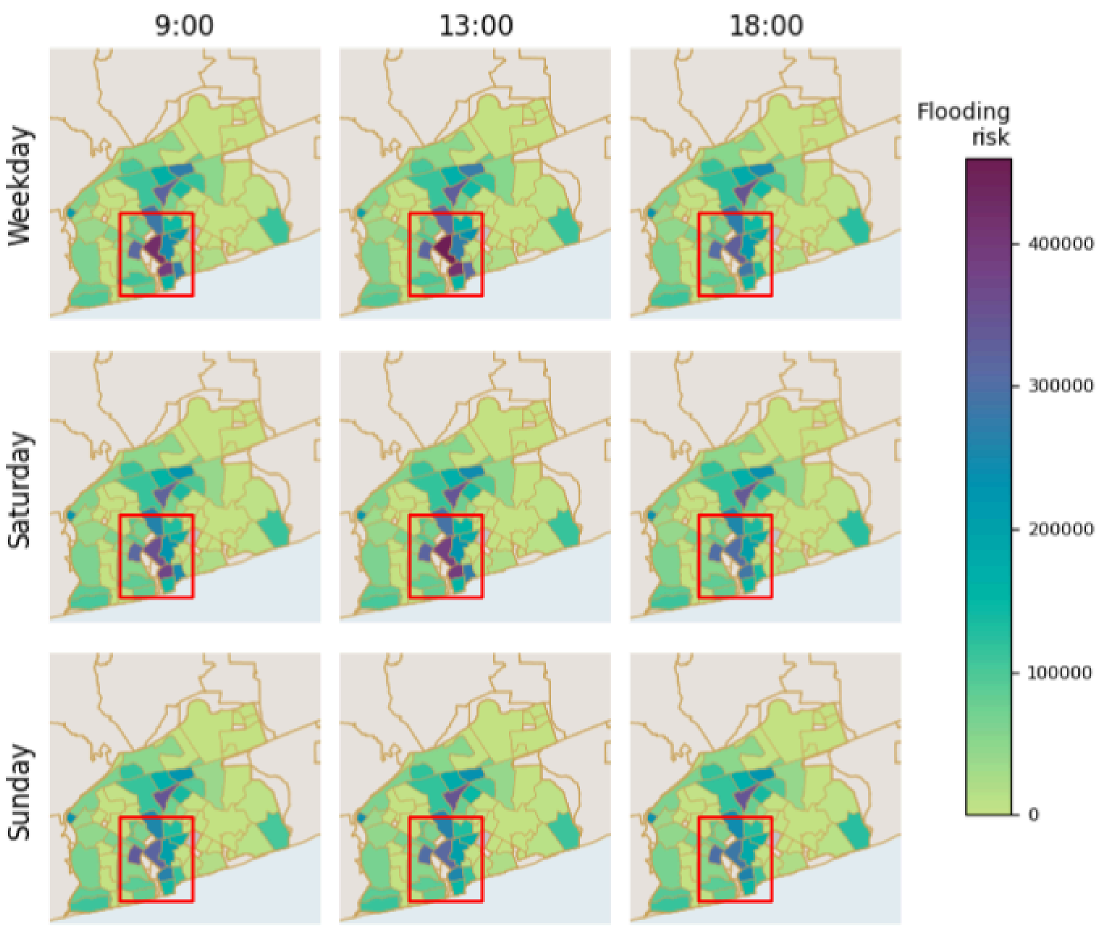

Figure 2: Urban division-level flooding risk for Accra at 09:00, 13:00, 18:00 on weekdays, Saturday, and Sundays. Central Accra experiences large increases in flooding risk (highlighted in red) during weekdays, peaking in the middle of the day. Substantial increases are also experienced on Saturdays. Flooding risk in Accra remains more stable on Sundays.

These changes are further highlighted in Figure 3 which shows the percentage change in flooding risk (relative to flooding risk at 19:00). Here, we can again see the substantial increase in flooding risk in central Accra on weekdays, and to a lesser extent on Saturdays. We can also see that only some areas with large increases in flood risk on weekdays experience similar increases on Saturdays.

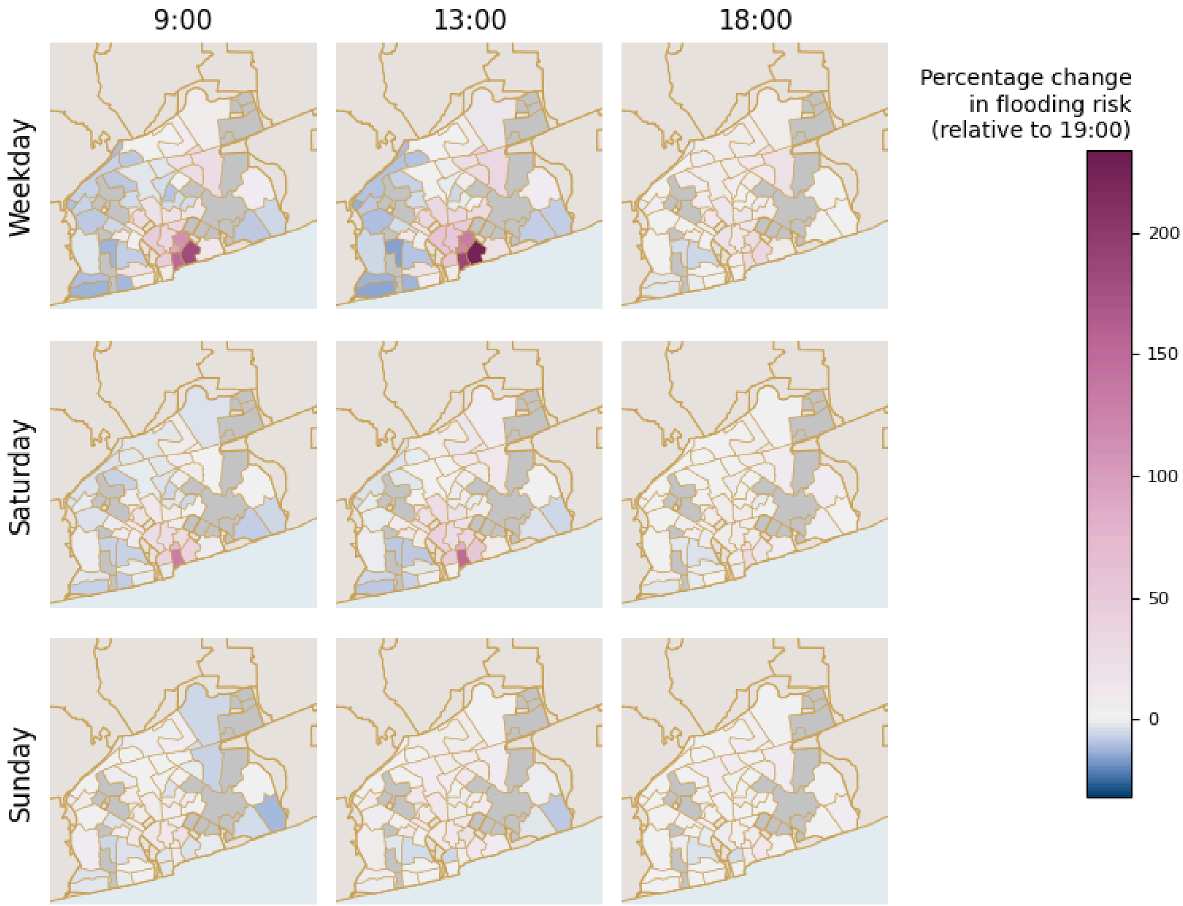

Figure 3: Urban division-level flooding risk for Accra at 09:00, 13:00, 18:00 on weekdays, Saturday, and Sundays, expressed as a percentage change relative to 19:00. Central Accra experiences large increases in flooding risk during weekdays, peaking in the middle of the day. Substantial increases are also experienced on Saturdays. Flooding risk in Accra remains more stable on Sundays.

Looking at the country as a whole, we can see the district-level flooding risk map for Ghana at different times of day and at different parts of the week. As the spatial variation in flooding risk is much larger than the variation in risk over these times scales and the largest changes occur in smaller urban districts, the change in risk between different hours is difficult to observe here.

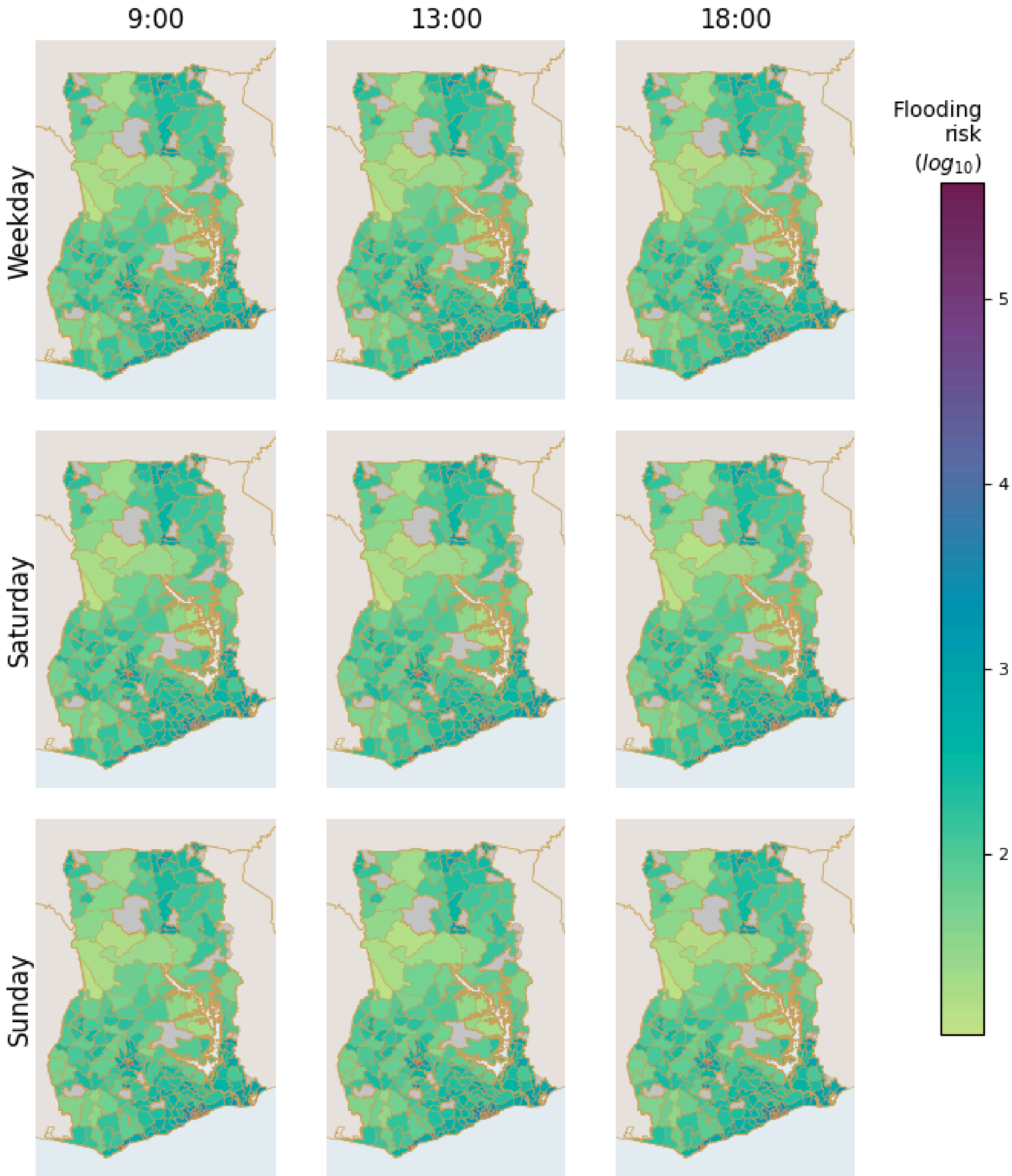

Figure 4: District-level flooding risk for the entirety of Ghana at 09:00, 13:00, 18:00 on weekdays, Saturday, and Sundays. The high spatial variation in risk and the small size of urban districts which experience the greatest changes in risk mean that the differences in risk between different hours and different changes are difficult to discern when presented as a map. However, such maps highlight high risk areas for closer investigation, such as Accra and Kumasi.

Figures 5 and 6 better demonstrate how our risk indicator varies over the course of a day, and how this differs between weekdays, Saturdays and Sundays. Figure 5 shows that, among the districts with the highest flooding risk, variation in the number of people present in an area over time can cause substantial increases in flooding risk, including in the district with the single greatest flooding risk. Figure 5 also shows that these changes differ depending on the day of the week, with very little change in risk between hours on Sundays.

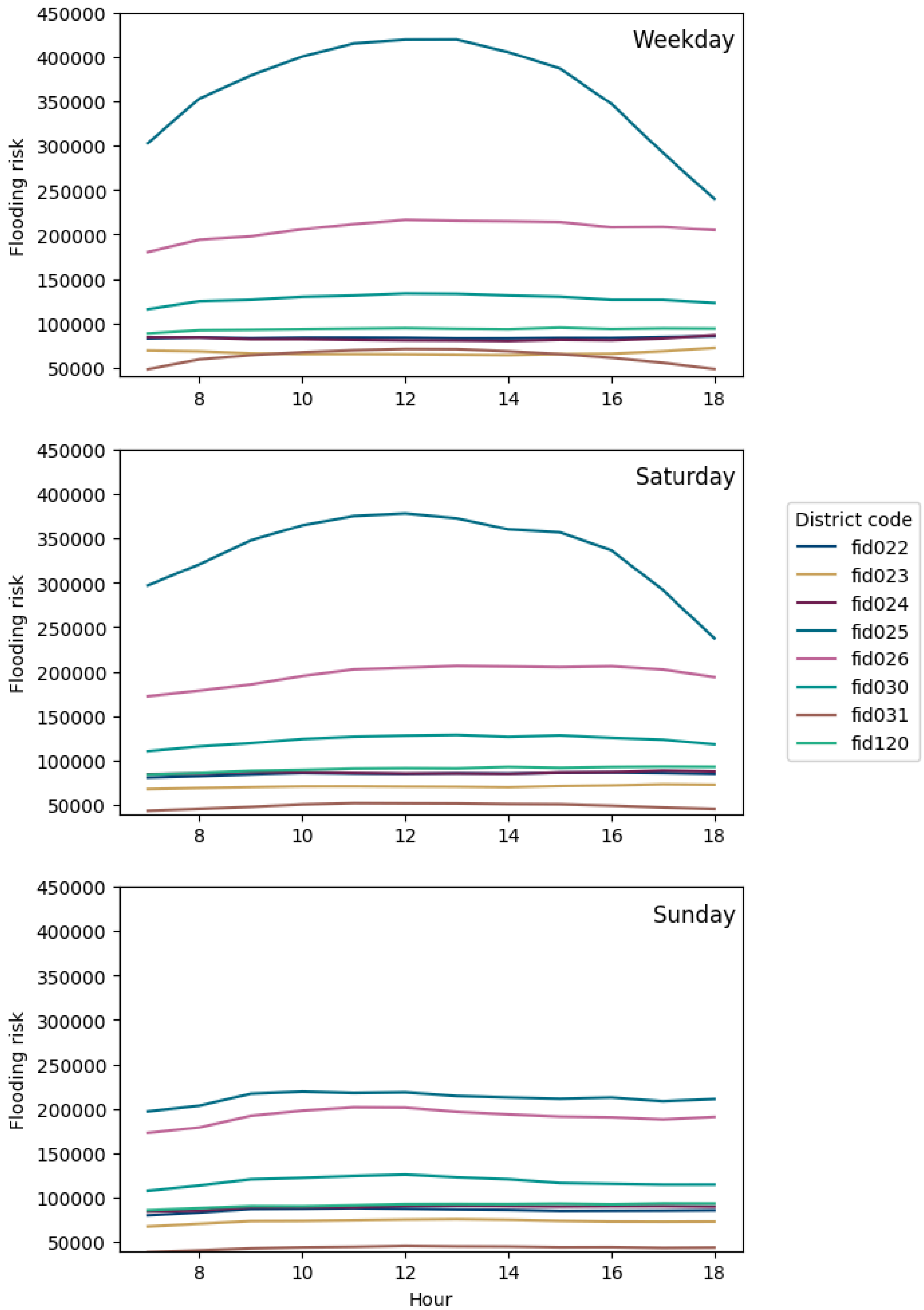

Figure 5: Variation in flooding risk over the course of a day in the eight districts with the highest flooding risk, on weekdays, Saturdays and Sundays. Some districts experience large changes in flood risk throughout the day, while others are more stable. The largest changes in flooding risk occur on weekdays, while there is relatively little change in flooding risk on Sundays.

Figure 6 further shows that some districts experience very large changes in hazard risk over the course of a day. Relative to the risk at 19:00, the district-level flooding risk can increase by as much as 100% on weekdays. Figure 6 also demonstrates that the size of these changes varies depending on the day of the week, with districts experiencing the greatest changes on weekdays and relatively small changes on Sundays.

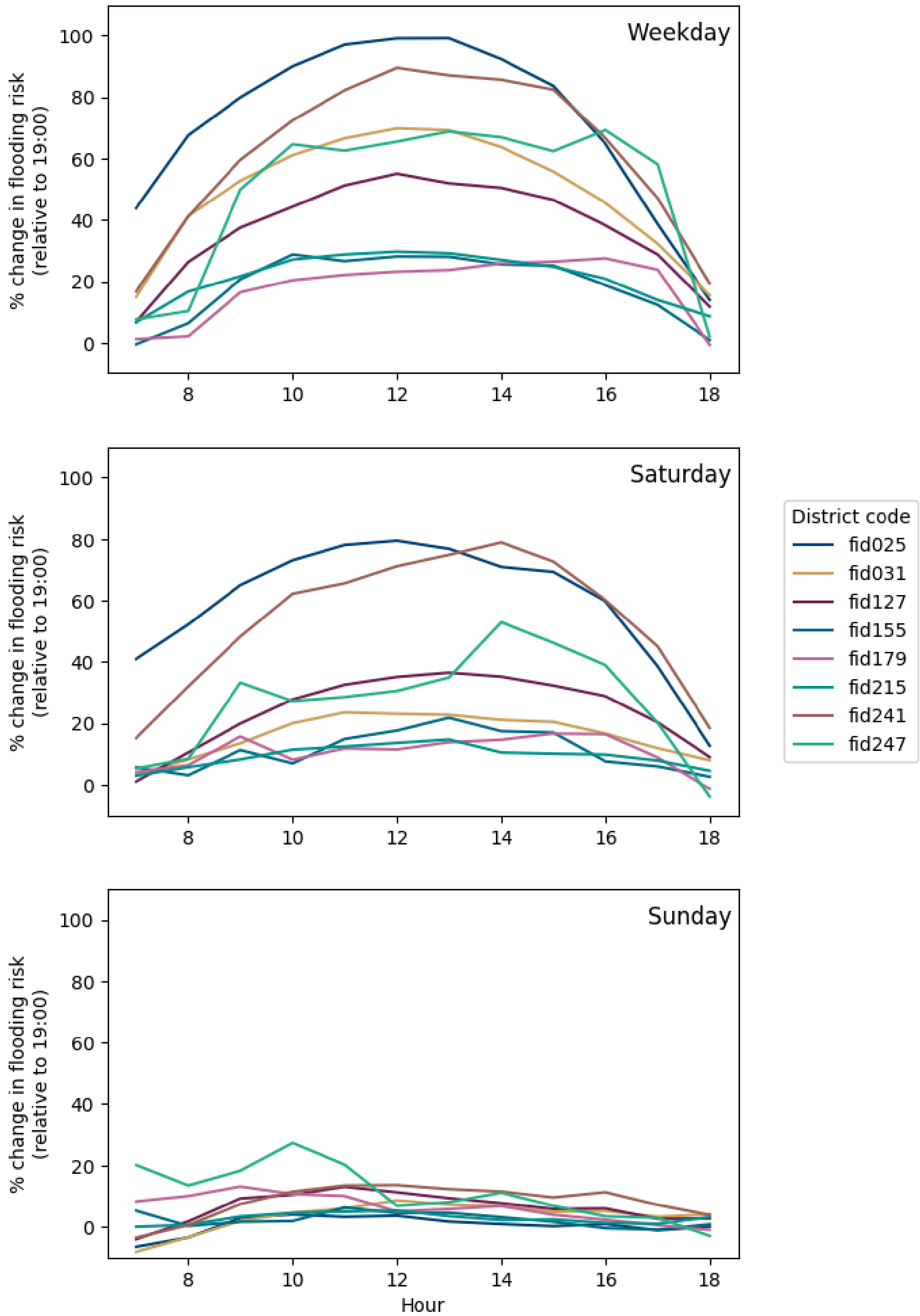

Figure 6: Variation in flooding risk, expressed as a percentage change, over the course of the day in the eight districts which experience the greatest changes in flooding risk. The flooding risk in a district can increase by

close to 100% in the middle of a weekday. Some districts also experience substantial changes in flooding risk on Saturdays. Flooding risk is relatively stable on Sundays in comparison.

Data sources

The dynamic hazard exposure mapping workflow has four data inputs:

- Anonymised CDR aggregates
- Static population data
- Geospatial hazard data
- Geospatial vulnerability data

Our methodology for combining these data is described in the next section, and is outlined in Figure 7 below.

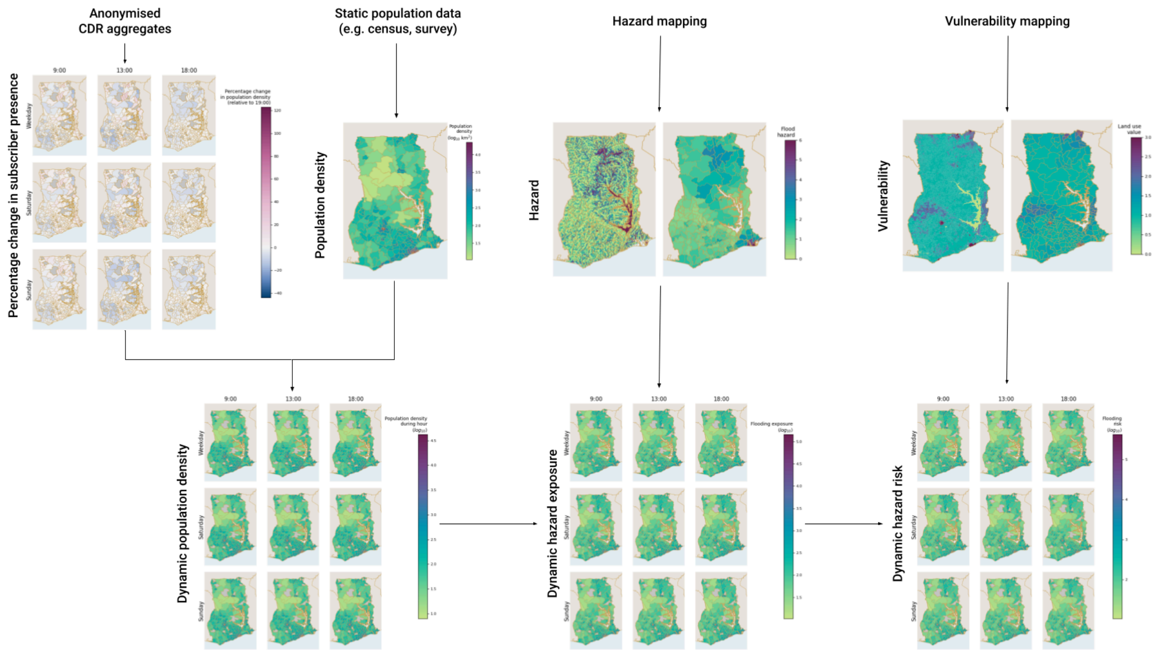

Figure 7: Flow diagram outlining how the different types of data are processed and combined to generate dynamic hazard exposure data.

Anonymised CDR aggregates

Call detail records (CDRs) are a form of data generated by activity on a mobile network, such as calls, SMS messages, and mobile data sessions, which mobile network operators routinely collect for operational purposes, primarily billing. In combination with location data for cells in the network, CDR data provide the three features required to infer mobility: a subscriber identifier (which has been pseudonymised), a timestamp (the date and time of a network event), and the location of the cell which routed the network event (which is assumed to approximate the location of the subscriber).

By aggregating the mobility of large numbers of subscribers and applying additional anonymisation protocols (such as k-anonymity), patterns of population mobility and changes in population distribution can be captured while protecting the individual privacy of mobile phone subscribers by preventing the movements of any single subscriber from being discernible.

For this report, CDR aggregates are produced using pseudonymised CDR data provided by Vodafone Ghana to calculate specific, approved aggregates such as the number of people present in each area each hour. The pseudonymised CDR data is aggregated and anonymised within the premises and firewall of the operator, and only the resulting anonymised aggregates are accessed for analysis.

Static population data

Static population data, such as population data from censuses or surveys, provide a baseline number of people residing in an area. These data are used to scale the CDR-derived indicators to produce population-scaled estimates of the number of people visiting or residing in an area over time.

For this report, we used the 2020 WorldPop population density estimates for Ghana.

Geospatial hazard data

Geospatial hazard data provide information on which areas may be affected by a hazard, such as flooding or drought.

For our use cases, this data is provided by NADMO but could be obtained from other open data sources.

Methodology

In this section, we describe the methodology used to calculate the hourly relative flooding risk in Ghana at a district level and in Accra at an urban division level.

Production of subscriber presence aggregates

The subscriber presence aggregate is calculated as the number of unique subscribers recorded in a given area in a given period of time. Subscribers are recorded in an area if a network event (e.g. phone call, SMS message, mobile data) is routed by a cell tower within that area.

We calculated three subscriber aggregates: hourly, district-level subscriber presence, hourly urban division-scale subscriber presence for Accra, and hourly, national-level subscriber presence. For this report, we used aggregates derived from CDRs produced by mobile phone calls only between May and November 2021.

Calculation of hourly subscriber presence by day type

We subset the presence aggregates grouped into weekdays, Saturdays and Sundays, and calculated the median number of active subscribers present in each district during each hour. We also calculate the median national-level count of unique active subscribers each hour for each day type.

For each day type, we then normalised the median district-level counts of unique active subscribers each hour by dividing by the median national-level hourly subscriber presence. This is expressed by the equation:

$$normalised\ subscriber\ presence_{i,t} = subscriber\ presence_{i,t} \cdot total\ \ active\ subscribers_{t}$$

where i is a given district and t is a given hour.

We used the same methodology to calculate the normalised subscriber presence by day type for each urban division in Accra.

Scaling to population size

We scaled our subscriber presence indicator to population-level using the equation:

$$population\text{–}scaled\ presence_{i,t} = static\ population_{i} \cdot (1 + X \cdot relative\ change\ in\ subscriber\ presence_{i,t})$$

where X is an adjustment factor to account for the greater mobility of mobile phone users, compared to non-mobile phone users. Based on previous mobility analyses conducted for Ghana, this was set as 0.5.

The static population in each district was calculated using the 2020 WorldPop population density estimates for Ghana which gives the estimated population count per square kilometre. To estimate the static population per district, each grid cell within a district was summed (Figure 8).

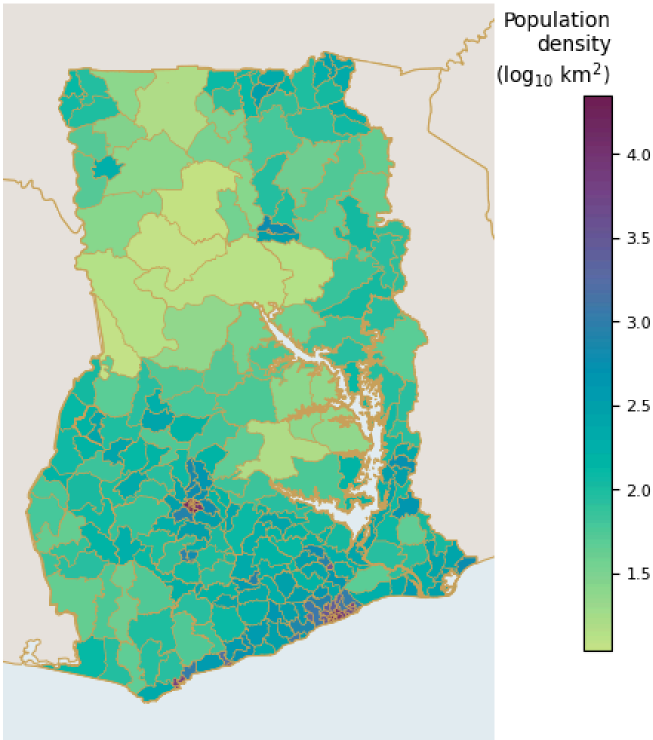

Figure 8: Static population district-level density map for Ghana.

For each day type, we calculated the relative change in the normalised hourly, district-level subscriber presence, using the normalised subscriber presence at 19:00 as a baseline. This was calculated as expressed by the equation:

$$relative\ change\ in\ subscriber\ presence_{i,t} = \frac{subscriber\ presence_{i,t} - subscriber\ presence_{i,19}}{subscriber\ presence_{i,19}}$$

We used 19:00 as a baseline for normalised hourly subscriber presence as it is the last hour with a high number of national-level active subscribers.

The same process was used to scale the subscriber presence per urban division in Accra to population size.

Calculating hourly hazard risk

We used the INFORM initiative framework to calculate hazard risk to combine hourly presence with hazard and vulnerability indicators. The hazard and vulnerability inputs were those previously used in the National Level Flood and Drought Risk Assessment and Mapping technical report produced by NADMO in February 2015.

The NADMO flood hazard data was provided as a grid of $100m^2$ cells. We estimated the district-level flood hazard by calculating the mean hazard score for all the cells in a district (Figure 9).

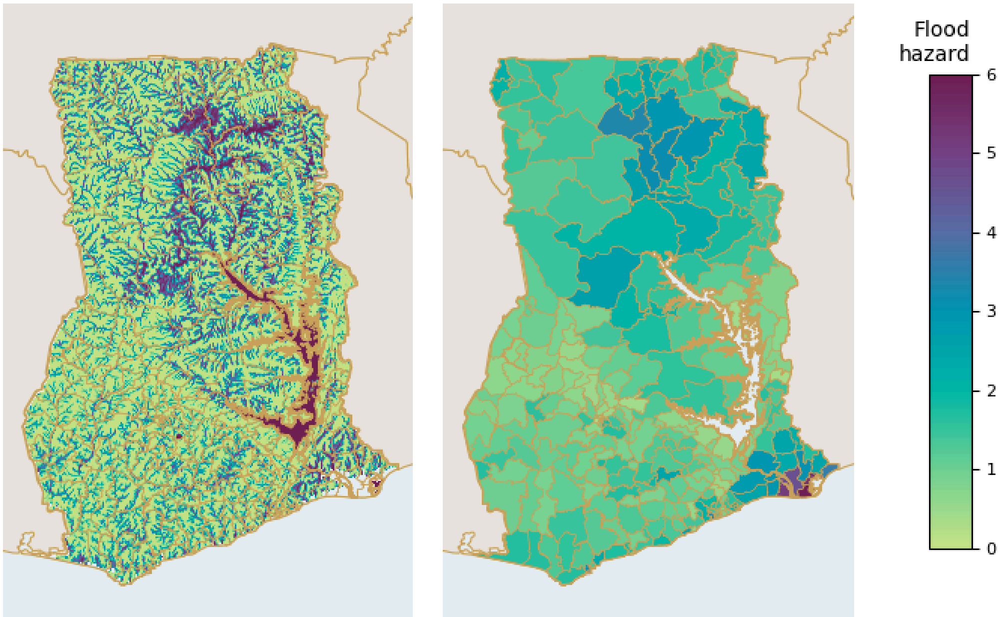

Figure 9: Flooding hazard maps for Ghana. Flow diagram outlining how the different types of data are processed and combined to generate dynamic hazard exposure data.
The map on the left shows the hazard data as a grid of $100m^2$ cells, as it was provided by NADMO. The map on the right shows the mean hazard score by district.

The previous flood risk analysis by NADMO calculated vulnerability using land use categories, which were assigned scores. We estimated the district-level vulnerability by calculating the mean land use score for each district (Figure 10).

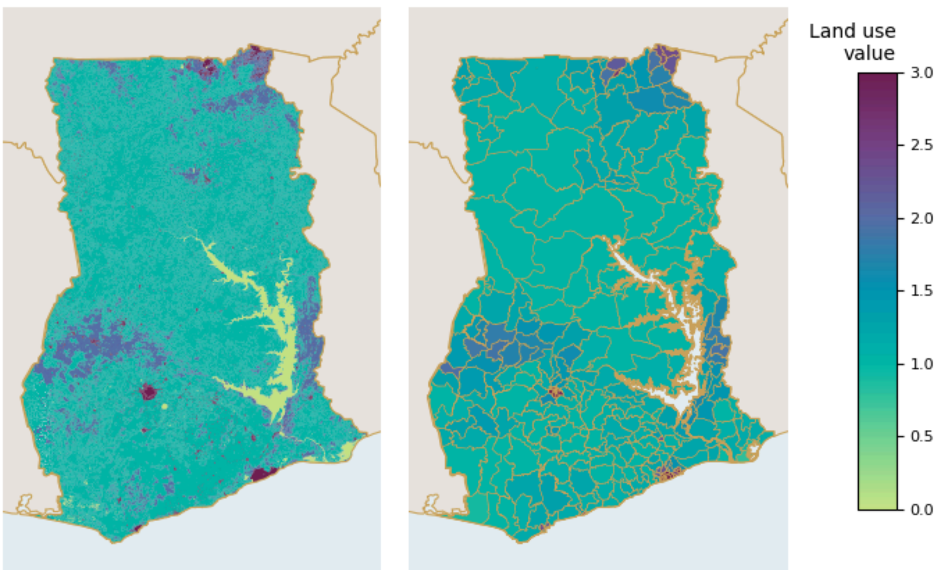

Figure 10: Land use maps for Ghana, scored by vulnerability to flooding. The map on the left shows the data as it was provided by NADMO. The map on the right shows the mean score by district.

We converted hourly, district level population size for each day type to population density by dividing by the district area (in square kilometres).

We then calculated hourly, district-level flood risk from the mean hazard score of each district, the mean vulnerability score, and hourly, district-level population density. This was calculated as expressed in the equation:

$$flooding\ risk_{i,t} = flooding\ hazard_{i} \cdot population\ density_{i,t} \cdot vulnerability_{i}$$

The capacity for authorities and institutions within Ghana to respond to flooding was assumed to be consistent across the country, and was therefore not included in the risk calculation.

We also calculated the hourly percentage change in risk, using the risk at 19:00 as a baseline. This was calculated as expressed by the equation:

$$percentage\ change\ in\ flooding\ risk_{i,t} = \frac{flooding\ risk_{i,t} - flooding\ risk_{i,19}}{flooding\ risk_{i,19}} \cdot 100$$

Some areas of Ghana have no flooding hazard, and therefore a flood risk of 0, resulting in null value for percentage change in risk. No districts of Ghana or urban divisions of Accra has a vulnerability of 0, but if areas with no vulnerability were included in the analysis the same would apply.

The same process was used to calculate hourly, urban division-level flood risk for Accra.

Next steps

In this section, we outline our ongoing work to further develop our methodology and apply it to other types of hazard.

We are consulting with NADMO about the other types of hazard within their portfolio. In addition to the flooding use case described above, we have conducted preliminary analyses on exposure to drought. Drought, which NADMO has also requested dynamic risk indicators for, is a slower-onset hazard and tends to affect larger areas than flooding, and therefore presents an interesting alternative case study. Increased temporal resolution on population density and short-term mobility may be of interest perhaps less to quantify exposure than to quantify the capacity of the area to resist the hazard or in terms of vulnerability. We are exploring a number of approaches to calculating exposure from CDR aggregates, including using the number of people residing in an area and the maximum hourly presence. We will be exploring these different approaches with NADMO to better understand their requirements for a drought risk assessment.

We are also consulting with NADMO to better understand their requirements for risk mapping. This includes understanding the spatial resolution NADMO requires for hazard risk mapping. This will vary depending on the type of hazard and CDR-derived dynamic hazard exposure maps may not be suitable for some types of hazard within NADMO's portfolio which require very high spatial resolution, such as building collapse. However, we can calculate hazard exposure at higher resolution for some areas, such as for urban subdivisions in Accra as demonstrated above.

We are currently testing an alternative approach to calculating subscriber presence, derived from the number of trips into and out of each district each hour. This alternative approach is similar to our recently developed method for calculating the number of residents in an area from the number of relocations observed. For estimating the number of resident subscribers, we found that this new method is less influenced by variation in mobile phone usage, which also varies substantially over the course of a day.

We have also recently developed a novel approach to the population-scaling and bias-adjustment of CDR aggregates by combining our long-term residents aggregates with primary and secondary survey and census data. We are currently in the process of developing similar methodologies for population-scaled and bias-adjusted short-term presence aggregates. We are now implementing our population-scaled and bias-adjusted methodology in Ghana.

Lastly, we are also developing methods to incorporate long-term (e.g. seasonal) variation in the number of people residing in each district in the framework above. This may have important implications for flooding risk in areas which experience large seasonal variation in population due to labour migration.

Summary

In this report, we demonstrate the application of anonymised CDR aggregates to the production of dynamic hazard risk indicators. This can support disaster preparedness by supporting decision-makers to understand how mobility affects risk and to take mobility into account during the planning process.

From the results presented in this report, we can see that mobility can result in large increases in hazard risk depending on the time of day and the day of the week, and that these changes differ between areas. Aggregated CDR data is well-suited to capturing changes in mobility and therefore the resulting changes in hazard risk.

We are also continuing to refine our methodologies and to work with our partners at GSS and NADMO to develop impactful indicators which are delivered in a way that facilitates their use by staff at these agencies.

Supplementary information

Change in population density

In order to calculate the hourly district-level flooding risk for Ghana, and hourly urban-division risk for Accra, we calculated the change in population density relative to the density at 19:00. These indicators are displayed on the maps below.

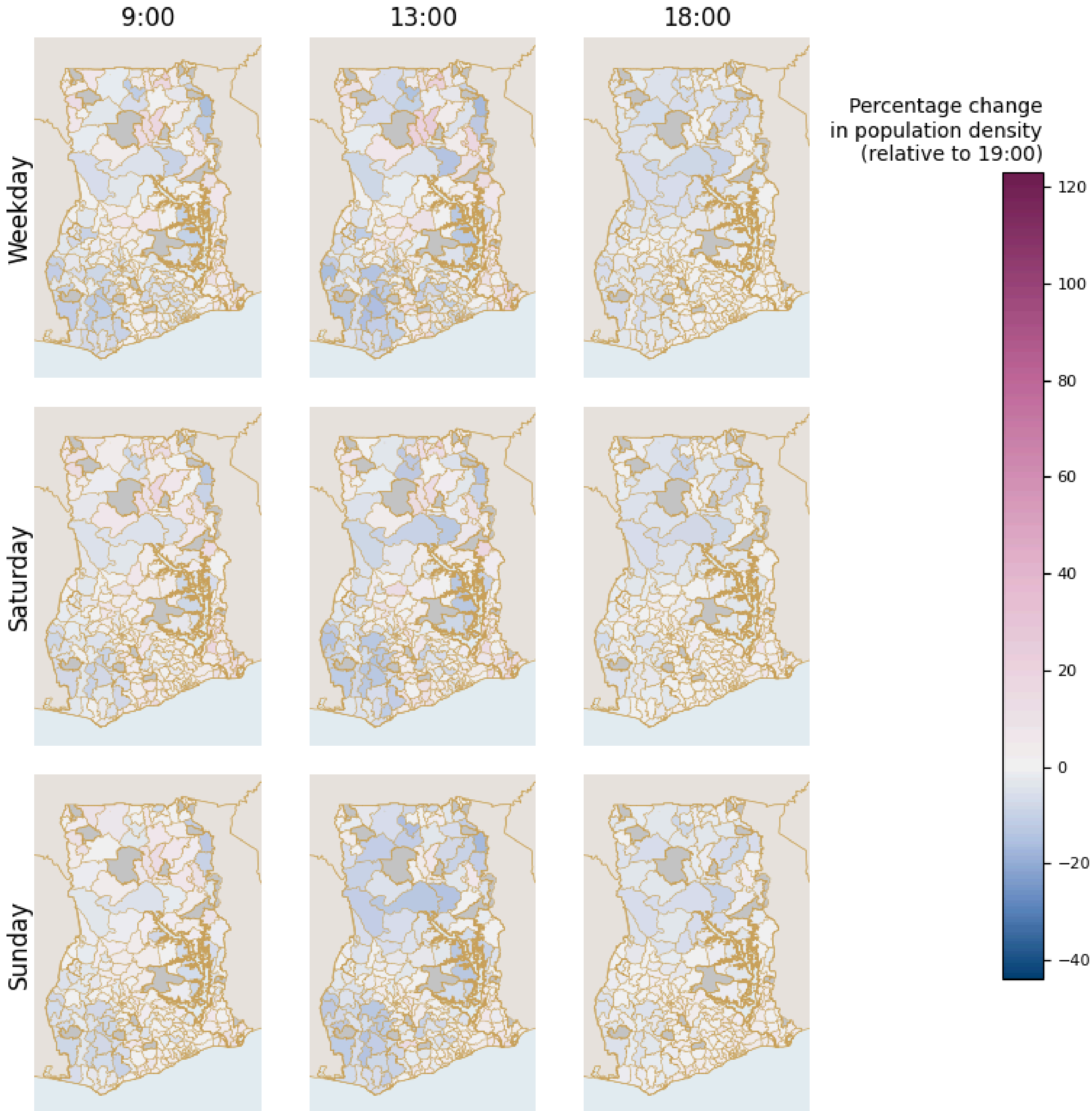

Figure S1: Change in population density of districts in Ghana, expressed as a percentage change relative to a 19:00 baseline. The largest changes occur in smaller urban districts which are not clearly visible at this scale.

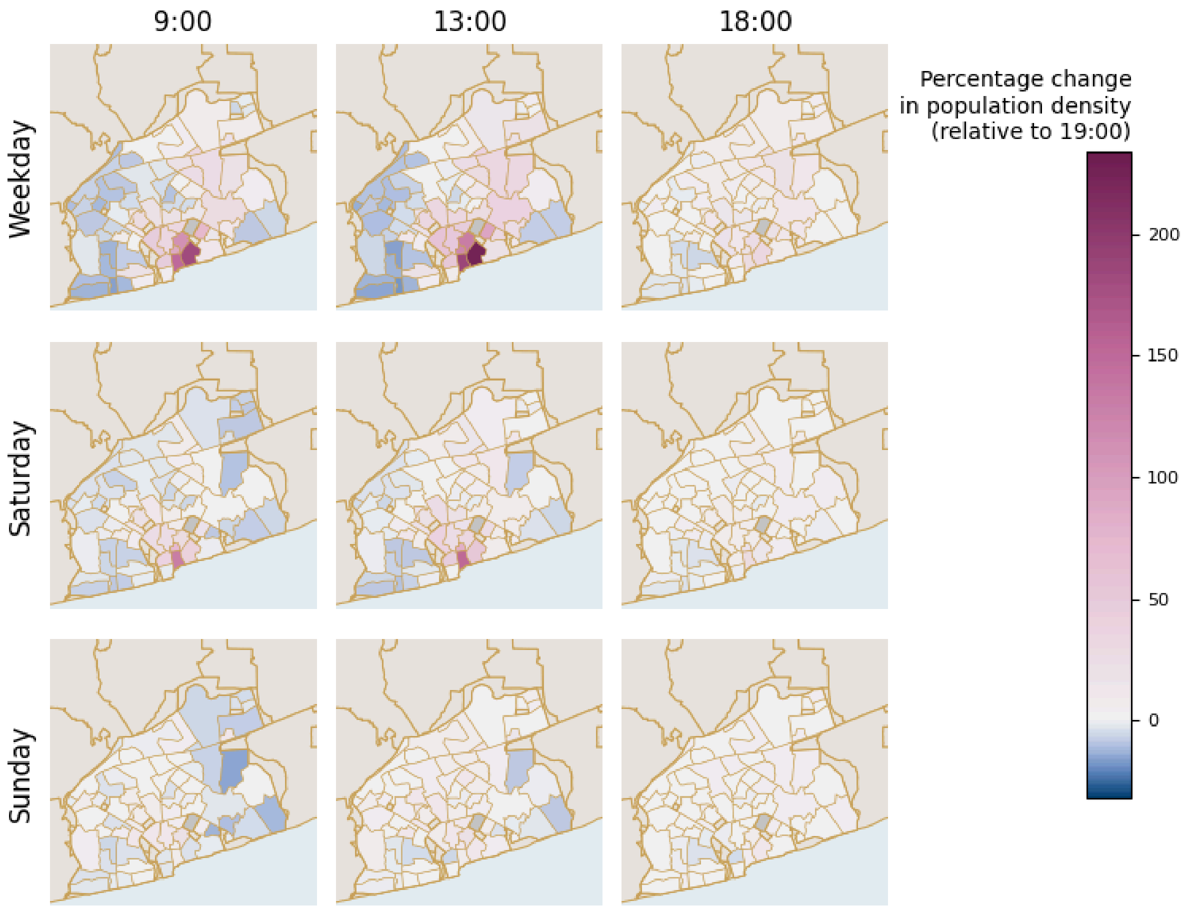

Figure S2: Change in population density of urban divisions in Accra, expressed as a percentage change relative to a 19:00 baseline. The largest increases in population density occur in central Accra during weekdays, and to a lesser extent on Saturdays. A smaller number of urban divisions experience substantial increases on Saturdays than weekdays, suggesting activity on Saturdays is focussed in specific areas. Population density in Accra is more stable on Sundays.

Change in hazard exposure

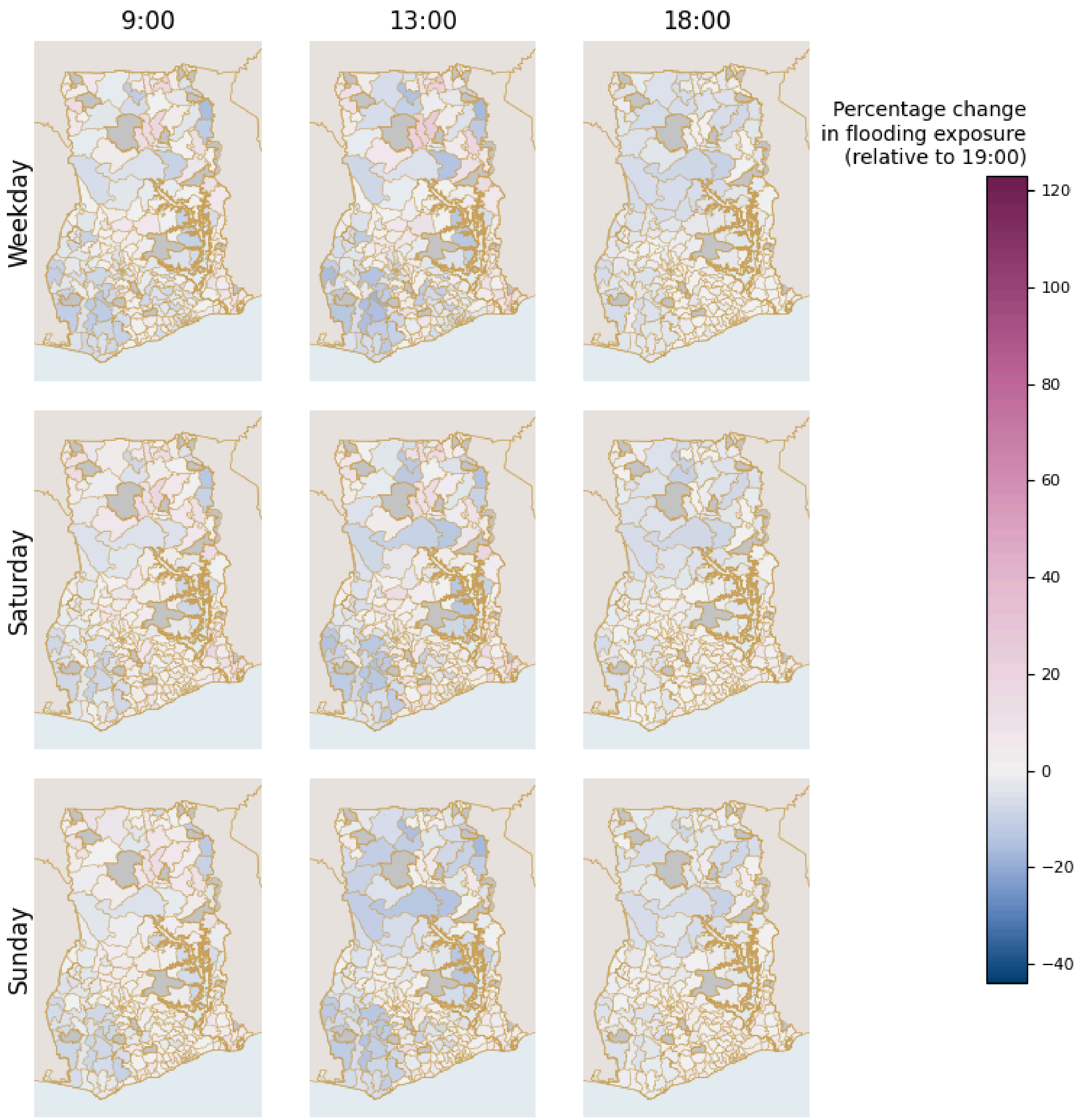

Figure S3: Change in flooding exposure of districts in Ghana, expressed as a percentage change relative to a 19:00 baseline. The largest changes occur in smaller urban districts which are not clearly visible at this scale.

Figure S4: Change in flooding exposure of urban divisions in Accra, expressed as a percentage change relative to a 19:00 baseline. The largest increases in exposure occur in central Accra during weekdays, and to a lesser extent on Saturdays. A smaller number of urban divisions experience substantial increases on Saturdays than weekdays, suggesting activity on Saturdays is focussed in specific areas. Flooding exposure in Accra is more stable on Sundays.