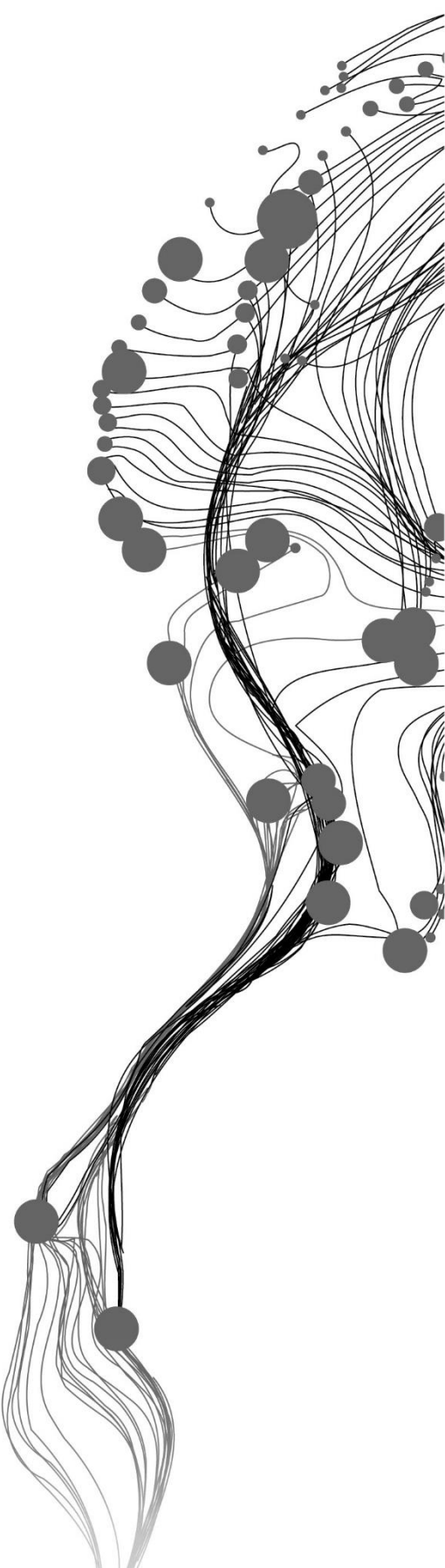# THE INFLUENCE OF URBAN MORPHOLOGY ON FLOOD SUSCEPTIBILITY IN SLUMS IN A DATA SCARCE ENVIRONMENT USING MACHINE LEARNING

JANE-MARIE MUTHONI MUNYI
July, 2024

SUPERVISORS:
Dr. M. Kuffer (First)
Dr. B. Van den Bout (Second)
Trento Lorraine Oliveira (PhD Advisor)

THE INFLUENCE OF URBAN MORPHOLOGY ON FLOOD SUSCEPTIBILITY IN SLUMS IN A DATA SCARCE ENVIRONMENT USING MACHINE LEARNING

JANE-MARIE MUTHONI MUNYI
Enschede, The Netherlands, July, 2024

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Urban Planning and Management

SUPERVISORS:

Dr. M. Kuffer (First)
Dr. B. Van den Bout (Second)
Trento Lorraine Oliveira (PhD Advisor)

THESIS ASSESSMENT BOARD:

Dr. rer. nat. D. Reckien (Chair)
Dr. S. Georganos (External Examiner, Karlstads Universitet)
Marija Bockarjova (Procedural Advisor)

The Influence of Urban Morphology on Flood Susceptibility in Slums in a Data-Scarce Environment Using Machine Learning

DISCLAIMER

The Influence of Urban Morphology on Flood Susceptibility in Slums in a Data-Scarce Environment Using Machine Learning

ABSTRACT

The causes of flooding in Nairobi are multifaceted, with climate change and rapid urbanization making flooding of great concern, particularly in slums as they are the most vulnerable settlements. Climate change risks are heterogeneously distributed in cities such as Nairobi, with the urban poor and slum settlements experiencing higher flood exposure and aggravated impacts. Despite slums becoming potential locations for increased urbanization, their association with being 'illegal' have made them receive minimal attention from governments, making the problems they face, such as flooding invisible and barely understood since the risks they face are not quantified. The influence of slum morphologies on flooding such as density, spatial distribution and arrangement has been insufficiently studied, with key focus being placed on their locations in flood-prone areas.

The overall aim of this research was to (i) investigate the influence of urban morphology by quantifying flood susceptibility and (ii) explore the distribution of flood susceptibility between slums and formal settlements. This aim was achieved following three objectives: identifying morphological flood factors alongside hydrological, environmental and geomorphological Flood Influencing Factors (FIFs), constructing a flood inventory (flood no flood locations) for Flood Susceptibility Mapping (FSM) using Machine Learning (ML) as the last objective, at a grid level of 100 meters by 100 meters. FIFs were derived from Remote Sensing (RS) while urban morphological flood factors were quantified into measurable characters (morphometrics). The flood inventory was generated by combining results from a flood simulation modelled using Fast Flood - a fast browser flood simulation tool, and Citizen Science (CS) flood information. Flood susceptibility was modelled using 2 Random Forest (RF) models by predicting the probability of susceptibility based on morphometric factors and FIFs (Model 1) and solely on FIFs (Model 2).

Susceptibility values ranged from 0-1 with values near 1 indicating very high flood susceptibilities. Results from both models exposed disproportional distribution of flood susceptibility between slum and formal settlements. Slums are observed to be highly susceptible compared to formal settlements with median susceptibility values of 0.65 (in Model 1) and 0.5 (in Model 2) for slums and 0.3 in both models for formal settlements. Additionally, the results imply that urban morphology has a significant influence on flood susceptibility as the overall accuracy of Model 1 increased to 84.71% from 71.76% in Model 2 with the inclusion of morphometric factors. The findings suggested that distance to rivers was the most influential susceptibility factor, followed by building adjacency and mean inter-building distance morphometric characters respectively. The space and room left for water to flow and infiltrate as a result of the spatial arrangement in slums and their occupation in floodplains were discovered as the fundamental reasons as to why slums face more flood risks than formal settlements.

Considering the bias towards focusing on the effects of floods on development (i.e. exposure), building development regulatory policies such as impact assessments and building arrangement guidelines ought to be formulated to evaluate the influence of development on flooding. Additionally, structural and non-structural measures were also provided as flood mitigation and adaptation measures. Some recommendations for future flood investigations lay stress on the use of CS information and high-resolution data for accurate flood mapping and spatial transferability of the model across geographically diverse regions.

ACKNOWLEDGEMENTS

This MSc journey, specifically this last year has tested my strength and resilience. The trials and tribulations that I have faced have been quite something, from crying alone and to others, joking to drop out and going back home, in Kiswahili we say "wacha Mungu aiitwe Mungu" translating to 'let God be called God', because, I had to get this degree, which in my opinion is an absolute honour. Thank you, ITC, for awarding me this scholarship.

Taking up this research topic enabled me to academically challenge myself releasing me from the shackles of staying within my comfort zone. I would therefore start by expressing my sincerest gratitude to my supervisors, Dr Monika Kuffer and Professor Bastian van den Bout for their support, belief and invaluable guidance throughout this phase especially in the formative stages of developing the topic. Many thanks to Dr. Kuffer who helped me land an internship where I gained a plethora of knowledge on flooding from my host supervisor Professor Jacques Teller and advisor Fenosoa Ramiaramanana (Nata), they encouraged me and solidified my passion in the field of flooding, and I look forward to working and publishing with them.

A special thanks to Lorraine Trento, my wonderful PhD Advisor who I have had the privilege of learning from. Lorraine has been there from the onset, providing assistance, advice and dedicating her time to reviewing my work, low-key she became my pillar of motivation. We have come so far, from me naively bombarding her with questions to us conducting fieldwork workshops together in Nairobi where she became a 'Nairobian', using slang words such as 'choche choche' (alleys), taking 'boda boda' (motorcycle) rides throughout and realizing that time culture in Nairobi is a more of a suggestion unlike in Europe. This research is without a doubt the result of her commitment to being an advisor, I wish her the very best and much success in her PhD, hope we can work together in the future.

I would also like to thank the Space4All team for being very supportive, most especially for facilitating the fieldwork through their many connections and resources. It was really lovely being in a team that fed into everyone and supported each other by giving encouragement, feedback and attending defences. The African ITC Community and my UPM batchmates have been a great source of support during my entire MSc serving as an escape from school and I am truly grateful for our de-stressing catch-ups. I would like to thank Abdul, he has been a great pillar of support and motivation, especially during my weakest moments, his optimistic nature and perspective have been incredible sources of energy. If anything, he knew my work more than I knew his, I thank him for his selflessness.

Deepest appreciation to my parents, brother and grandmother for their prayers, and emotional and financial support, I hope that I have made them proud and will keep making them proud. I know my mom wants to call me Dr.Muthoni, education is not easy! we will have to see what happens in the future.

Last but not least, I want to thank me! for doing all this hard work, for having no days off and especially for never quitting! And the almighty God for giving me favour.

The Influence of Urban Morphology on Flood Susceptibility in Slums in a Data-Scarce Environment Using Machine Learning

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ANN	Artificial Neural Network
APHRC	African Population and Health Research
BD	Building Density
BuA	Building Adjacency
CAR	Covered Area Ratio
CBO	Community-Based Organization
CLPC	Clay Soil Percentage
CS	Citizen Science
DEM	Digital Elevation Model
DEM	Digital Elevation Model
DistRivers	Distance to Rivers
FD	Fractal Dimension
FFS	Fast Flood Simulation
FIF	Flood Influencing Factors
FIFs	Flood Influencing Factors
Flow_Acc	Flow Accumulation
FN	False Negative
FP	False Positive
FSM	Flood Susceptibility Mapping
GEE	Google Earth Engine
GIS	Geographic Information System
GOB	Google Open Builidngs
IBD	Mean Inter-building Distance
IPCC	International Panel on Climate Change
IQR	Inter-quartile Range
KNBS	Kenya National Bureau of Statistics
LIDAR	Light Detection and Ranging
LMICs	Low and Middle-Income Countries
LULC	Land Use Land Cover
ML	Machine Learning
MLP	Multi-Layer Perception
NCCG	Nairobi City County Governmnet
NDBI	Normalized Differentiated Built-up Index
NDi	Neighbour Distance
NDVI	Normalized Differentiated Vegetation Index
NDWI	Normalized Differentiated Wetness Index
NGO	Non-Governmental Organization
OBIA	Object-Based Image Analysis
OSM	Open Street Map
RF	Random Forest
RS	Remote Sensing
SDGs	Sustainable Development Goals

SDI	Sum Dwellers International
SPI	Stream Power Index
SSA	Sub-Saharan Africa
SVM	Support Vector Machine
SWR	Shared Wall Ratio
TN	True Negative
TP	True Positive
TWI	Topographic Wetness Index
UAV	Unarmed Aerial Vehicle
UN-DESA	United Nations Department of Economic and Social Affairs
UN-HABITAT	United Nations Human Settlement Programme
VHR	Very High Resolution
XGBoost	Extreme Gradient Boosting

1. INTRODUCTION

It is without a doubt that 21st-century cities confront tons of urbanization challenges (UN-HABITAT, 2009) such as climate change, which is identified as a principal challenge in the context of sustainable development. The intensification of the hydrological cycle as an effect of human-induced climate change (IPCC, 2023) influences erratic and extreme precipitation, significantly increasing urban flood risk (Tabari, 2020). Despite the effects of climate change being felt by all, the poorest populations characterized by having minimal carbon footprints, constrained adaptive capacities and limited decision-making powers, are particularly struck the most by climate change impacts, especially flooding (IPCC, 2022).

1.1. Background

Floods have become the most frequent and catastrophic climate change-related risks Bradford et al. (2012); Paul (2015), where in the last 20 years, flooding has affected more than 2.3 billion people globally, and in 2019 alone, it accounted for 43.5% of deaths from natural disasters (Suhr & Steinert, 2022). Considering that floods are projected to be heightened by 2050 (Towfiqul Islam et al., 2021), with extreme daily precipitation events set to exacerbate by approximately 7% for every 1^0C of global warming (IPCC, 2021) flood events are set to become common disasters.

World Bank (2006), contend that people in low-income countries are four times more likely to die from natural disasters compared to people in high-income countries. With poor communities and populations being primary victims of such disasters, the increasing probability and severity of floods ought to be viewed as a social and physical construct (Anwana & Owojori, 2023).

Dumedah et al. (2021) emphasizes that the proliferation of slums is one of the challenges of rapid urbanization. However, the effects of floods in cities often vary between formal and slum. While flood risks have been reported to be extremely high in slums (WHO & UN-Habitat, 2016), Hamidi et al. (2022) suggest that slums are more likely to experience heightened impacts due to higher flood levels and increased exposure.

Approximately a third of the urban population in Low and Middle-Income Countries (LMICs) are hosted in slums (UN–Habitat, 2015). Nearly 1.1 billion people globally, were reported to live in slums in 2020, with a further 2 billion expected to live in slums in the next 30 years (United Nations, 2023). At a global scale, Africa and Asia are anticipated to account for roughly 90% of the future's urbanization (UN-DESA, 2019).

A study conducted by Tschakert et al. (2010) highlighted that Africa was the second hardest-hit continent by floods after Asia. In addition, according to Jha et al. (2011), by the year 2000, the number of flood incidences in Africa was higher compared to the rest of the world. Climate variability, altering patterns of flooding and prolonged rains have worsened flooding in many Sub-Saharan African (SSA) cities (Douglas et al., 2008), for example:

   i)   Residents in Accra, Ghana observed that from 2000, the heavy rains that occurred in June and July would start before June or continue beyond July.
   ii) The predictable cycles of flooding in Kampala, Uganda that occurred in the rainy seasons of April-May and October-November altered, resulting in unpredictable and erratic flooding.
   iii) In Nairobi, Kenya, long-term slum dwellers provided that areas that never used to flood began to experience flooding within two decades.

iv) In Maputo, Mozambique residents asserted that since 1980, flooding events became worse, claiming that a one-day rain event would cause floods that persisted for 3 days.

Flooding is not exclusively attributable to climate change, but also to anthropogenic factors (Anwana & Owojori, 2023) and urbanization dynamics (Ramiaramanana & Teller, 2021). The lack of data to justify climate change as a probable cause has resulted in its unfair associations with flooding, placing blame on it while overlooking human activities influencing flooding (Dumedah et al., 2021). Human activities, especially those resulting from human settlements have modified urban landscapes creating urban morphologies that increase the propensity of flooding (Santos & Reis, 2018).

According to Mensah & Ahadzie (2020) key drivers of increased flood risks in Africa are related to intensive development. Notably, slums, characterized by their densities, location in, and obstruction of flood-prone areas make slum dwellers frequent and primary victims (Kuffer et al., 2021; United Nations, 2023), emphasized by their direct relationship with the intensities of disaster vulnerabilities (Abunyewah et al., 2022). Excluding climate variability and the existence of slums, poor city planning, improper waste management strategies, poor drainage infrastructure, human activities near rivers and uncontrolled growth have been considered as driving causes of urban floods in African cities (Abass, 2022; Amoako, 2012; Douglas et al., 2008).

Efforts have been made to understand, quantify, analyse and predict the impacts of flooding worldwide (Mudashiru et al., 2021). Despite the complexity surrounding flooding Douglas (2017), Flood Susceptibility Mapping (FSM) has been used to understand flood risks by assessing the correlation between floods and their influencing factors (Seydi et al., 2022). Quantitative, hydrological-based and statistical models have been used and developed for FSM (Mudashiru et al., 2021). However, recently, Machine Learning (ML) models such as Artificial Neural Networks (ANN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Multi-Layer Perception (MLP) and Support Vector Machine (SVM), have gained traction in FSM due to their capabilities to model complex events such as flooding (Seydi et al., 2022).

1.2.   Research Problem and Gaps

Over the past three decades, the role of urbanization and climate change in urban flooding has been heavily contested, bearing no resolution (Amoako, 2012). As climate change increases flood hazards in Africa (Dottori et al., 2018) and with Africa forecasted to face the majority of the future's urbanization (UN-DESA, 2019), flooding is becoming an acute problem demanding immediate action (Zhu et al., 2019).

1.2.1.   Relationship Between Slums and Flooding

Unfortunately, the association of slums as being 'illegal' have made them receive minimal attention (Huang, 2021) and ineffective government responses (UN-Habitat, 2009). Subsequently, they are excluded within the formal urban systems, making the problems they face quite difficult to recognise (Huang, 2021), especially in relation to climate change adaptation as they are 'invisible'. For this reason, literature has overlooked flood challenges in slum contexts such as the geographical space left for water to flow during flooding (Dumedah et al., 2021).

Challenges such as flooding in slums are not well understood, specifically the differences between slums and formal settlements in terms of flood exposure and susceptibility. This is attributable to the lack of databases that quantify disaster risks faced by slums compared to better-off settlements, yet, on numerous occasions, the urban poor are witnessed to be hit the hardest (Kuffer et al., 2021).

1.2.2. Flood Quantification

Efforts have been made to quantify the risk, damage, vulnerability, and spatial extent of floods by focusing on understanding, predicting and estimating flood hazards (Seydi et al., 2022). Flood mapping studies such as those by Kia et al. (2012) and Tehrany et al. (2014), highlight that most efforts consider hydrological, geomorphological and environmental factors referred to as Flood Influencing Factors (FIFs). However, current studies have shown an association between urban floods and anthropogenic factors (Anwana & Owojori, 2023; Lee & Brody, 2018; Lin et al., 2021).

Urbanization has a direct positive relationship with an increase in flood events, even when there is no rainfall variability (Walsh et al., 2012). The depiction of urban growth in terms of urban morphology, referring to the spatial analysis of urban structures, land use, street patterns, buildings, open spaces, and the nature of human settlements, is suggested to add to the understanding of flooding (Dumedah et al., 2021).

Few flood hazard modelling studies exist for slums in Africa (Tom et al., 2022), which is concerning, especially given the (i) increasing exposure of slums and (ii) projection of variable climate in Africa (Amoako, 2012) which will increase the risk of flooding. The minimal or lack of investigations done for flooding in slums provides little evidence to support flood risk management (Tom et al., 2022) and quantify their flood risks. Despite having few flood mapping studies, a small proportion have attempted to utilize urban morphological elements in FSM, thus indicating the need for research that targets this gap.

1.2.3. Data Scarcity

Flood mapping is challenged by data paucity of relevant datasets (Juma et al., 2023), especially in the global south (Hawker et al., 2020), as well as in the least developing and developed regions (Omonge et al., 2022). However, the availability of open-access global datasets has been instrumental in providing mapping data (Ceola et al., 2022), albeit having coarse resolutions and low accuracies (Carr et al., 2024). Despite their benefits, the effectiveness of global datasets is uncertain, especially due to downsizing processes that approximate local conditions (Sun et al., 2022).

Al-Aizari et al. (2024) assert that flood inventories (flood and no-flood locations) are important for FSM, as historical and previous events can be used to predict the likelihood of future flood events (Towfiqul Islam et al., 2021). Data scarcity concerning flood inventories is a common problem in many cities, most especially in urban areas, as (pluvial) flooding tends to be dispersed (Al-Aizari et al., 2024).

Hydrological monitoring stations are few, particularly in Africa and in most cases, do not cover flood locations, therefore contributing to low spatial coverage of areas susceptible to flooding (Al-Aizari et al., 2024). Similarly, Nairobi faces such challenges, where most of the rivers are ungauged (Mulligan et al., 2019) leading to the lack of historical discharge datasets – which are most especially useful for fluvial flooding (Juma et al., 2023).

1.3. Significance and Purpose

This research intends to investigate and quantify the nexus between slums and floods as a result of inadequate knowledge and minimal efforts in substantiating flooding challenges and quantifying flood risks faced by slums. Effectively understanding this relationship is essential for it to be addressed in line with the Sustainable Development Goals (SDGs), Urban Agenda 2030 and the Sendai Framework.

Since most flood studies exclude anthropogenic flood factors and heavily focus on the location of slums and their uncontrolled growth as the main influencing flood factors, this research focuses on understanding the role of the spatial urban structures of slums using urban morphology. This research will add knowledge

in the fields of informality and flooding, by providing new insights through the creation of empirical evidence, in addition to addressing the intricacies of urban landscapes to flooding, especially in slums.

The integration of urban morphology in flood mapping is relatively new, representing a novel approach to understanding flooding. Researchers such as Dumedah et al. (2021); Lin et al. (2021) denote correlations between urban morphology elements and flooding. By investigating these correlations, the findings of this research will contribute to an enhanced understanding of the relationship between the built environment and flood dynamics.

This research aims to understand the complexity of flooding and its dynamics in slums in SSA, presenting Nairobi, Kenya, as a case study, envisioned to investigate the relationship between flood susceptibility and the physical form of urban areas using urban morphological elements.

1.4. Objectives

The overall goal of the research is to investigate how urban morphology influences flood susceptibility in slums and compare how susceptibility varies between slums and formal settlements. On this basis, the objectives are:

1. To identify and derive flood factors using openly available datasets.

   RQ1: What urban morphological elements have the likelihood of influencing flooding?
   RQ2: What Flood Influencing Factors (FIFs) are important for Flood Susceptibility Mapping (FSM) based on existing methodologies and datasets?

2. To generate a flood inventory in a data-scarce environment.

   RQ1: What tools and datasets can be used to create a flood inventory?
   RQ2: How can qualitative and quantitative datasets be combined for validation?

3. To develop a city-wide flood susceptibility map For Nairobi.

   RQ1: To what extent are slums more susceptible to flooding than formal settlements?
   RQ2: Which morphological element influences susceptibility the most?

2.   LITERATURE REVIEW

2.1.   Slums

The complexity surrounding the definition of slums is a result of (i) differential local context matter, (ii) location (iii) official and unofficial description and (iv) differences in issues covered (Bird et al., 2017). Subsequently, resulting in no standard definition, adding a level of subjectivity in defining slums (Mahabir et al., 2016). The UN defines slum households as households where inhabitants lack one or more of the following household deprivations (UN-Habitat, 2008, 2015, 2018).

1. Lack of access to improved water sources.
2. Lack of access to improved sanitation.
3. Lack of sufficient living space.
4. Lack of housing durability.
5. Lack of secure tenure.

UN-Habitat (2008) argued that household-level shelter deprivation fails to fully capture the degree of deprivation experienced by slum households, the severity of combined deprivations and the changing dynamics of deprivation over time. To address these challenges UN-Habitat, (2008) enhanced the definition by grouping slum households into 3 categories: moderately deprived (one-shelter deprivation), severely deprived (two-shelter deprivation) and extremely deprived (three-shelter deprivation)

Despite the improvement, the definition suffers from the lack of a social dimension and has difficulty in capturing information relating to tenure security within slums (Mahabir et al., 2016) since it is a non-physical expression of slum conditions that deals with legality (UN-Habitat, 2008). Khalifa (2011) argues that despite significant strides toward alleviating one or more deprivations, the changes in slum status would not be recognized. Khalifa (2011) further provides that shelter deprivations do not consider the risk posed to human life and argues that people can live without tenure security but cannot survive when their houses are located in hazardous areas such as floodplains or landslide-prone locations.

Informal settlements and slums are the most commonly interchangeably used terms (Kuffer et al., 2016), however, according to (Mahabir et al., 2016; UN-Habitat, 2015), there is a difference between the terms, with informal settlements referring to area deprivation while slums are defined as household deprivation.

This research uses the term slums because informal settlements are perceived to be real estate speculations that can be occupied by both affluent and poor urban residents, and specifically focuses on the formal status of land tenure and structures (United Nations, 2023). Moreover, given that the term 'slum' explicitly refers to the physically deprived conditions (Wang et al., 2019), it is well suited for this research since the study investigates the influence of the physical form on flooding.

2.1.1.   Slum Characteristics

Globally and locally, the definition of slums varies, resulting in poor classifications of what slums are or are not. Attempts to classify slums based on the UN-Habitat definition by Engstrom et al. (2015), resulted in nearly the whole city being classified as a slum. One alternative for the misclassification issues is the characterization of slums using morphological components, with Kuffer et al. (2016) outlining components such as small roof sizes, high densities, irregular patterns, and locations in hazardous sites. These specific morphological characteristics can be grouped into 5 dimensions: building geometry, density, arrangement, roofing materials and site characteristics.

Additionally, to enhance slum characterization, Kohli et al. (2012) proposed a morphological classification at different spatial levels of: environs, settlements and objects. Slum experts from a survey done by Kohli et al. (2012) delineated slums from imageries using characteristics such as irregular roads, small roofs, absence of roads, lack of vegetation and open spaces, compact density, irregular settlement shape, and locality.

The location of slums, according to Ezzati et al. (2018) is perceived to be a trade-off between poverty and risk that contributes to the exposure of slums. Ezeh et al. (2017) argue that the urban poor have increased exposure to multiple hazards, because of their decision to live in hazardous locations influenced by poverty and proximity to economic opportunities.

2.2. Urban Morphology

Moudon (1997) terms urban morphology as the study of the city as a human habitat, defining it as the study of the complex structures of human habitats. Gauthier & Gilliland (2006), define urban morphology as the study of the urban form of cities. Barau et al. (2015) explain urban morphology as the study of human settlements and the process of their transformation, considering urban morphology as the spatial analysis of urban physical structures, land use, street patterns, buildings, and open spaces.

Urban morphology can be studied at different urban scales to identify recurring patterns in the structure and configuration of the built environment and to understand how its elements work together (Kropf, 2014). The spatial scales of urban morphology, according to Fleischmann et al. (2022) consist of plots, buildings, streets, squares, blocks, and neighbourhoods.

2.2.1. Morphological Comparison between Slums and Formal Settlements

Differentiating slums from formal settlements using pattern, density, and size morphological features, Kuffer et al. (2014) explain slum morphologies as exhibiting organic patterns, which are usually more complex, irregular, diverse, and dense than formal areas. Useful comparisons between slums and formal settlements are provided in Table 1 (Baud et al., 2010; Kuffer et al., 2016; Kuffer & Barros, 2011; Scott et al., 2017).

Table 1: Morphological Differences between Slum and Formal Settlements

Morphological features	Slum settlements	Formal settlements
Size	✓ Small building sizes	✓ Larger building sizes
Density	✓ Lack of green spaces ✓ High density (of at least 80% of roof coverage) ✓ Insufficient distance between buildings ✓ Narrow access paths	✓ Planned green spaces ✓ Low - moderate density ✓ Larger distances between buildings
Patterns	✓ Organic and irregular structure with irregular layout patterns ✓ Irregular streets or unidentifiable streets	✓ Regular layout patterns ✓ Regular street patterns
Site characteristics	✓ In proximity to hazardous locations	✓ Buildings built on suitable land

2.2.2. Morphometrics

The physical structure of cities can be characterized by quantifiable elements known as morphological features (Abascal et al., 2022), by quantifying the features through meaningful measurements of their geometrical forms and spatial relationships (Wang et al., 2023). Quantifying these features relies on consistent and reliable data on elements of urban morphology such as buildings, open spaces, plots, and streets (Mumford, 1961). Building footprints are globally mapped and easily accessible from various sources such as Open Street Map (OSM), Bing Satellite Maps, Google Open Buildings (GOB) and Microsoft Buildings.

Datasets such as OSM are inconsistent as they are manually delineated, with datasets such as GOB and Microsoft Buildings being consistent as they are delineated using Artificial Intelligence from Very High Resolution (VHR) images. Street data from sources such as Microsoft Roads are not consistently mapped and lack connectivity especially in slums where there is minimal coverage (Wang et al., 2023).

2.2.3. Urban Morphology and Flooding

Lee & Brody (2018) claim that anthropogenic factors such as unplanned development and alteration of natural landscapes with impervious surfaces are highly correlated with flooding. According to Lin et al. (2023), artificial impervious surfaces continuously occupy green spaces that consequently increase surface run-off and alter hydrological conditions.

In this context, building metrics such as covered area ratio, building congestion degree and density of buildings heavily influence flooding (Lin et al. (2021). Shepherd (2005) claims that high-density developments tend to concentrate runoff volume subsequently getting waterlogged, especially in flood seasons because of the urban rain island effect. Walsh et al. (2012) further contribute that flooding due to increased surface run-off is directly increased by building density as it increases surface sealing and impervious areas.

According to Dumedah et al. (2021):

1. The disjointed arrangement of buildings in slums rarely leaves open spaces for flood water to flow and their limited drainage infrastructure, makes spaces between houses the only space left for surface run-off.
2. Horizontal spaces between buildings play significant roles in flow accumulation during flooding as minimal distances between buildings increase the accumulation of surface run-off.
3. Irregular spatial arrangements, modify and artificially extend the path of surface run-off that increases the volume of water as flow accumulation is promoted.
4. The orientation of buildings in relation to topography controls the level of room left for water to flow. Relative to the direction of water flow, buildings aligned perpendicularly can narrow the flow path leading to localised flooding, while buildings built in parallel allow for water to freely flow.

2.3. Flood Susceptibility

In the context of flooding, susceptibility is considered to be one of the determinants of flood vulnerability, alongside exposure and resilience (Salami et al., 2017). In this regard, susceptibility is considered to relate to the extent to which elements such as economic assets, buildings, and people in flood-prone and exposed areas are likely to be affected by a flood event (Anees et al., 2019).

Flood susceptibility is explained to be the probability of flood occurrence in a specific area with a certain intensity (Anees et al., 2019), due to an area's landscape and geographical region (Mudashiru et al., 2021).

Unlike flood risk, flood susceptibility is not related to life and property loss (Dottori et al., 2018; Wang et al., 2023).

Borrowing from Miranda et al. (2023), this research ought to focus on flood susceptibility related to the physical environment, with 'propensity to flooding' being a good synonym to conceptualize flood susceptibility. For this research, flood susceptibility is understood to be the likelihood of a flood event to occur.

2.3.1. Flood Susceptibility and Flood Influencing Factors (FIFs)

Developing flood models such as FSM demands the understanding of FIFs for specific regions (Kia et al., 2012; Masahiro et al., 2021), which are usually hydrological, geomorphological and environmental factors that influence the occurrence of flood events (Tehrany et al., 2017).

Given the vast array of factors to consider for FSM, the number of factors selected varies and is often selected from previous studies. Mahmoud & Gan (2018) used 10 FIFs and following a sensitivity analysis they found that flow accumulation, runoff and soil type were the most influential factors for flood susceptibility. They concluded that flood susceptibility maps ought to include more than 6 FIFs, however, other studies argue that reducing the number of FIFs can result in inaccurate results (Miranda et al., 2023).

The majority of flood susceptibility studies have used similar FIFs derived from literature and fieldwork. Research studies such as ones done by (Kia et al., 2012; Mudashiru.,2021; Tehrany et al., 2014, 2017, 2019) considered the following similar FIFs:

1) **Elevation:** As a prime factor in controlling flooding (Pradhan, 2009), higher elevation areas are seen to be less likely to be affected by flooding compared to low-elevation areas as water continuously flows into these areas causing them to quickly flood (Das, 2019). With flood-prone regions characterized by low-elevations (Mahmoud & Gan, 2018).

2) **Slope:** Flood-prone areas, characterized by having low surface slopes flood more easily and faster compared to steep slopes (Mahmoud & Gan, 2018). High slopes tend to increase run-off flow and decrease the time for surface infiltration increasing the likelihood of flooding (Mojaddadi et al., 2017), unlike low slopes where the run-off speed is decreased (Kabenge et al., 2017), increasing the propensity to flooding.

3) **Aspect:** As aspect shows the deviation of the slope from the geographic north which varies between 0 – 360 degrees (Farhadi & Najafzadeh, 2021). It affects factors such as moisture and vegetation cover depending on aspect direction affecting the micro-climate of a region (Al-Aizari et al., 2024).

4) **Curvature:** According to Tehrany et al. (2017) flooding mostly occurs in areas with flat curvature, as the flat terrain in such areas is suitable for flooding due to the role curvature has on the concentration of flow through infiltration and run-off processes (Cao et al., 2016).

5) **Flow accumulation:** Mahmoud & Gan (2018) explain flow accumulation as the concentration of water flowing from surrounding paths into another cell. They further contribute that elements located in high-flow accumulation are more susceptible to flooding as they act as convergent points for surface runoff.

6) **Stream Power Index (SPI):** SPI is an index that represents the erosive capability of water flow that contributes to stream channel erosion and transportation of sediments (Barker et al., 2009). The measure is based on the local slope gradient, identifying areas likely to flood (Ahmad et al., 2019). Mojaddadi et al. (2017) explain that areas with low power streams are highly susceptible to flooding, attributable to the location of regions with higher SPI values on steep areas where flooding is less likely to occur.

7) **Topographic Wetness Index (TWI):** TWI is an elevation-derived index associated with and used as a proxy for soil moisture (Cao et al., 2016), (Kopecký et al., 2021) and soil water storage capacity

(Chowdhury, 2023), which shows the amount of flow accumulation in a drainage basin and the capacity of water to travel downstream with gravity (Tehrany et al., 2019).

8) **Distance to rivers:** Regions located far from river systems are less likely to suffer from flooding than regions in closer proximity as they are situated along the flow path of surface runoff, consequently increasing their predisposition to flooding (Mahmoud & Gan, 2018). In addition to being situated along the flow path of water, flooding could be a result of overflow and over-capacity resulting in higher flood risk for areas closer to the rivers (Kabenge et al., 2017).

9) **Clay soil content:** Soils and their associated ecosystems and physical properties provide regulating contribution through physical mechanisms such as infiltration, storage and slow run-off release during flooding that consequently delay run-off and lower flood volumes (Saco et al., 2021). Soil clay content plays a role in influencing the intensity of flooding, as high clay content reduces infiltration and increases surface run-off (Shah & Shah, 2023).

10) **Land Use Land Cover (LULC):** Different LULC typologies have different hydrological responses (García-Ruiz et al., 2008) and contributions to flooding. Mojaddadi et al. (2017) suggest that urban areas are more prone to flooding due to the composition of bare land and impervious surfaces that increase storm-water runoff compared to vegetated areas.

11) **Normalized Difference Vegetation Index (NDVI):** NDVI is used to calculate the amount of vegetation in an area (Farhadi & Najafzadeh, 2021). Areas with high vegetation density decrease the runoff speed (Tehrany et al., 2017), deeming such areas as having low susceptibility to flooding in comparison to areas with low vegetation densities.

12) **Normalized Difference Built-up Index (NDBI):** NDBI analyses built-up areas as they are prone to flooding due to their impervious nature decreasing infiltration and increasing run-off (Mojaddadi et al., 2017).

13) **Normalized Difference Wetness Index (NDWI):** NDWI identifies the presence of water bodies and permanent water since areas occupied by waterbodies always pose threats to adjacent land by over-saturating soil leading to inevitable flooding and higher flood risks (Farhadi & Najafzadeh, 2021).

14) **Rainfall:** Susceptibility to flooding is known to increase with higher rainfall intensities. Extreme rainfall within a short time period can lead to flooding especially when water cannot be quickly absorbed by the soil or evaporate, resulting in increased surface flow (Douglas, 2017). Additionally flooding experienced in a given area is highly dependent on rainfall regardless of the catchment area or environmental condition(Nyarko, 2014; Segond et al., 2007).

2.3.2. Flood Inventory

FSM is treated as a binary classification task with the flood inventory being classified into flood points and no-flood points with binary labels of 1 and 0, respectively as dependent variables for susceptibility predictions (Towfiqul Islam et al., 2021). Which are used to determine the flood classification of an area and its probability of belonging to a specific class (Wang et al., 2023).

ML models heavily depend on sufficient flood inventory data as references for training and validation, however, the lack of such data limits the accuracy of FSM, particularly in data-scarce regions (Yu et al., 2023). Notably, a study conducted by Nsangou et al. (2021) in Cameroon, managed to work with limited data by using 50 flood points for a region occupying 95.6 km². Similarly, using an ANN algorithm, Falah et al. (2019) conducted FSM in an Iranian city spanning 120 km², based on 58 flood points.

Scarce data poses a challenge to the generalization ability of a ML model, specifically when using an excessive number of flood factors as sparse distances between sample points can result in overfitting (Ying, 2019). Moreover, when basic inputs are used in supervised ML models, they face difficulty in capturing complex

interactions between flood factors, resulting in model underfitting as sample points with similar attributes become undistinguishable (Leandro et al., 2016).

As a result, alternative data sources such as field observations, historical data records, RS imagery, topographic maps, perception of residents, topographic flood information, flood damage reports, and field surveys have been widely used to construct flood inventories (Al-Aizari et al., 2024; Farhadi & Najafzadeh, 2021; Towfiqul Islam et al., 2021; Youssef et al., 2022).

Recently, studies such as Assumpção et al. (2018) emphasize the role of Citizen Science (CS) in flood mapping in collecting flood data for use in flood prediction, calibration and validation. CS has been employed in collecting historical flood events given that it is cost-effective (Buytaert et al., 2014), offering more spatially distributed and relatively accurate, detailed and reliable data by providing context-specific observations. (Helmrich et al., 2021; Zeng et al., 2020). There is high potential for the use of CS in collecting flood and no-flood points on the basis that residents report more flood information (de Bruijn et al., 2019), since they are knowledgeable of their surroundings, are the witnesses and sometimes victims of flood events (Sy et al., 2020).

2.3.3. State of the Art in FSM

Over the years hydraulic models such as MIKE-FLOOD (Kadam & Sen, 2012), SWAT (Lee et al., 2017) and HEC-RAS (Juma et al., 2023; Lea et al., 2019) have been used for FSM and seen to be efficient, however, these models depend on detailed data accumulated over a long time, which causes challenges in areas that suffer from insufficient detailed data (Nguyen et al., 2022).

Researchers such as Ding et al. (2021); Hermas et al. (2021), carried out spatial FSM by integrating RS, and Geographic Information Science (GIS) data with statistical models as this integrated technique had the capabilities of handling large amounts of spatial data. However, these statistical models suffer from the problems of covariance, normal distribution predictions and linear relationships (see Table 2) which reduces the accuracy of statistical-based models (Prasad et al., 2021).

ML algorithms have been increasingly used for FSM, due to their high computation efficacy (Prasad et al., 2021; Wang et al., 2023) compared to physically based models. Additionally, their advantage in overcoming the challenges in statistical-based models, by predicting complex non-linear relationships with high accuracies and coping with limited data (Pham et al., 2021) has been preferred.

According to Bentivoglio et al. (2022), ML is considered to be supreme in susceptibility estimations, where after a systematic review of different ML models for binary classifications, most models produced accuracies surpassing 80%. The most commonly used ML models are supervised models that adopt instance-to-instance learning schemes such as neural networks, logistic regression, SVM and RF (Liu et al., 2022), which introduce mathematical relationships to learn the relationships between flood factors and their labels (Wang et al., 2023).

As ML models face the issues of overfitting and optimization problems, researchers have proposed the use of hybrid ML models as they have shown significantly better performances compared to traditional individual models (Nguyen et al., 2022). The superior performance of hybrid ML models has led to their advanced use and development, however since there is no guide in selecting the best hybrid model for FSM, researchers have to develop and test new hybrid algorithms (Nguyen et al., 2022).

Table 2: Overview of Existing Flood Models

Flood Models	Advantages	Disadvantages
Hydraulic	➢ High dependency on details collected over long periods	➢ Faces efficacy concerns in areas that suffer from limited data especially since they rely on detailed long-term information
Statistical	➢ Can handle large amounts of spatial data	➢ Has limitations in covariances, prediction of normal distribution and linear relationships ➢ Lower accuracies due to its inherent limitations
Machine Learning	➢ Overcomes limitations faced in statistical models ➢ Can predict non-linear relationships ➢ Higher accuracies	➢ Has challenges of model overfitting and optimization
Hybrid Machine Learning	➢ Have higher prediction power and performance than traditional and single ML models	➢ The combination of models depends on the characteristics of the area of interest

3. STUDY AREA

3.1. Study Area and Brief History

Nairobi City County, Kenya's capital city is located in the Southwest of Kenya and lies at an average elevation of 1,798 meters above sea level Government of Kenya, 2018; UN-Habitat, 2020). Given Nairobi's hilly topography, elevation decreases towards the Eastern boundary to the Athi River and is traversed by three main rivers, namely: Ngong River, Mathare River and Nairobi River (see Figure 1).

Nairobi is strategically located between Mombasa and Kisumu, its location was initially an uninhabited swamp (Mundia, 2017) up until it was selected to serve as a railway depot during the construction of the Mombasa - Uganda Railway (Oyugi, 2018) due to its adequate supply of water, comparatively flat terrain, cooler grounds, and availability of land (UN-Habitat, 2020). Nairobi has an area of 703.9 km², consisting of a National Park that covers 117 km².

According to the 2019 population census, Nairobi had a total population of 4,397,073 people (KNBS, 2019) with approximately more than 2.5 million slum dwellers representing 60% of the population, settling on 6% of the urban land (Rajula, 2016). With majority of slums in Kenya located in Nairobi as depicted in Figure 1.

The name Nairobi originates from "Enkare Nyirobi', a Maasai phrase, that translates to "a place of cool waters", the city was dubbed as such due to its expansive blue spaces at the margins of an arid region (UN-Habitat, 2020). Additionally, known by the presence of green infrastructure the city was coined as the "green city under the sun", however, due to rapid urbanization, Nairobi has been experiencing a reduction in environmental quality (Oyugi, 2018).

3.1.1. Rationale for Study Area Selection

Nairobi was selected for this research based on its large slum population and the proliferation of complex settlement patterns as a result of rapid urbanization (Sverdlik, 2021). According to Mulligan et al. (2017), Nairobi faces severe urban climate threats, where flooding is a major threat in Nairobi slums due to their positioning on the major river systems (Douglas et al., 2008). Additionally, the lack of global datasets such as the Global Flood Awareness System (GloFAS)[1] and MODIS Near Real-Time Global Flood Product[2] in capturing floods in Nairobi motivated this selection. Despite the recently added flood dataset (10/05/2024) in UNOSAT[3], urban flooding was minimally captured with a focus on fluvial flooding.

---

[1] Global Flood Awareness System – global ensemble streamflow forecasting and flood forecasting (copernicus.eu)

[2] MODIS NRT Global Flood Product | Earthdata (nasa.gov)
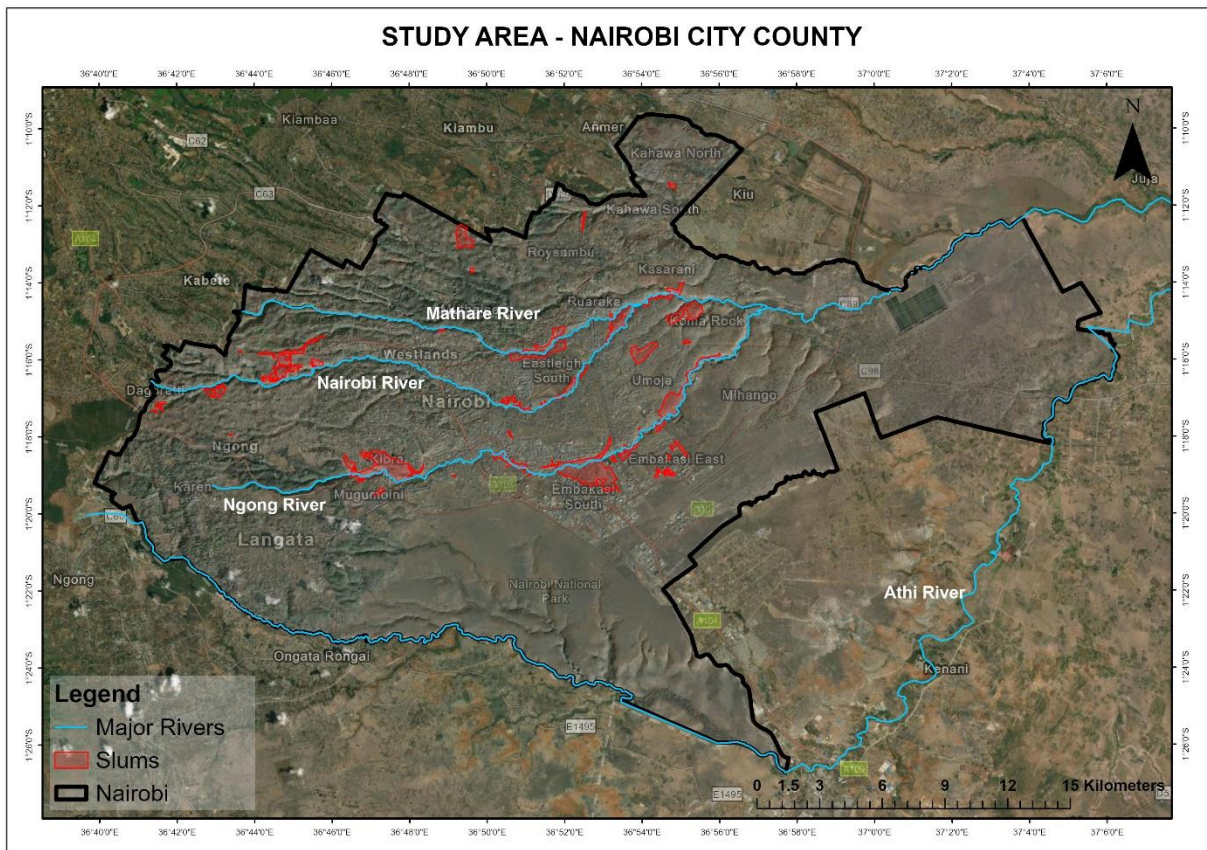
[3] UNOSAT

Figure 1: Location of Slums in Nairobi

3.2. Slums in Nairobi

The origin of slums in Nairobi can be dated back to the colonial period when racial segregation dictated city planning and settlement location. Access to designated residential areas by Africans was controlled and limited as they were reserved for Europeans and Asians (Mutisya & Yarime, 2011). As a result, natives retaliated by creating their settlements outside planned areas and the Central Business District (CBD) giving rise to informal settlements, after independence, residential segregation transformed into social status segregation (Mutisya & Yarime, 2011). The failure of the new administration (after independence) to settle the landless citizens is attributed to catalysing the expansion and growth of new slums (Slum Dwellers International, n.d.).

Slums occupied poor quality and inhabitable land, as it was their only alternative, with dwellers situating themselves in steep slopes, riparian reserves, swamps, refilled quarries, dumping sites, utility and infrastructural reserve land (Slum Dwellers International, n.d.). The oldest and major slums in Nairobi are Kibera, Mathare and Mukuru. Kibera slum emerged in 1912 around an ethnic migrant core, as it was initially designated for demobbed Sudanese Nubian soldiers, which was taken to be a military reserve (United Nations Human Settlements Programme, 2003). Mathare and Mukuru emerged in 1920, 1958 respectively (Mutisya & Yarime, 2011; Wanjiru & Matsubara, 2017). After independence, the continued proliferation of slums has been on a steady rise due to inadequate housing and the failure of poorly planned efforts to reduce them (Obudho & Aduwo, 1989).

3.2.1. Characteristics of Slums in Nairobi

Focusing on the physical attributes that can be detected using RS, Nairobi slums conform to the typical morphological characteristics of slums provided by Kohli et al. (2012) ; Kuffer et al. (2016). As seen in Figure 2, slums have low housing characterized by dense metal housing built using iron sheets, with some houses built with wood and mud (Scott et al., 2017). From aerial imagery, they appear to have small-sized roofs, irregular arrangements of buildings and road layouts, little vegetation cover and located near hazardous areas such as rivers. According to Scott et al. (2017) slums have limited access to basic and public utilities and services. In Nairobi, the scarcity has led to the commercialization of essential services such as water provision, waste collection and sanitation often provided by exploitative and informal enterprises (Huchzermeyer, 2008).

Kibera   Mathare   Kangemi

Jerusalem   Viwandani   Kwa Rueben
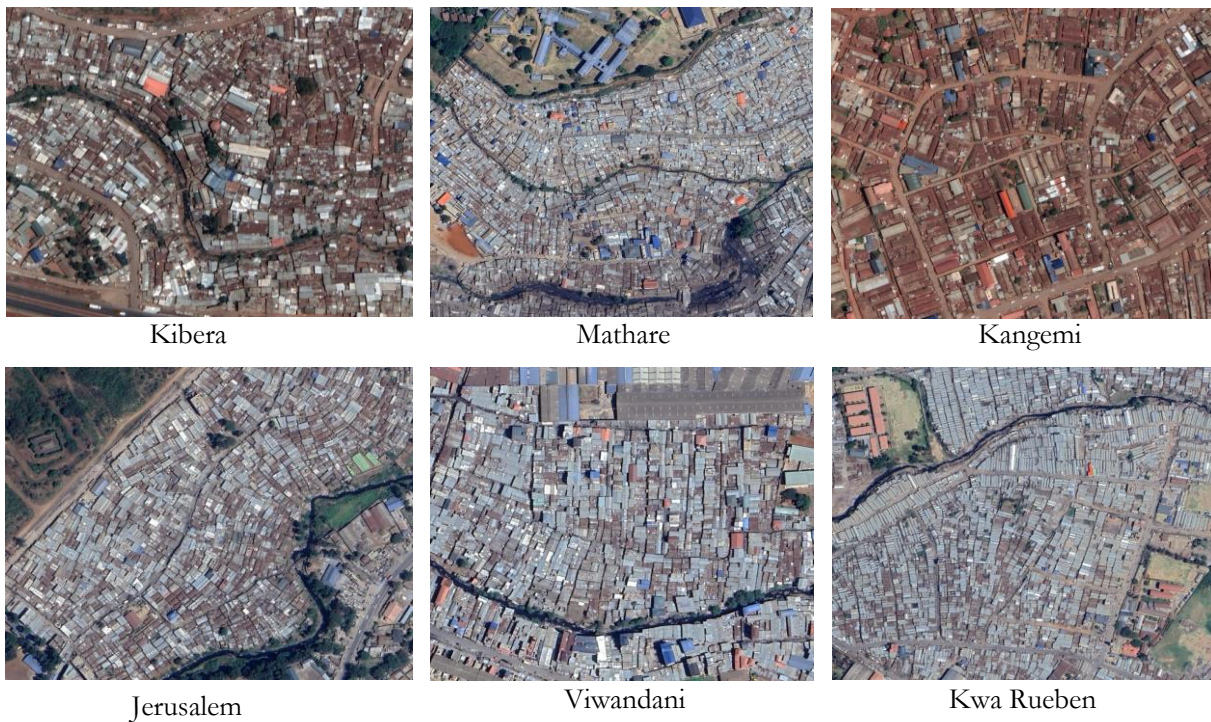
Figure 2: Aerial View Showing the Characteristics of Nairobi Slums

3.2.2. Urban Morphology of Slums and Formal Areas in Nairobi

UN-Habitat (2008) argues that the most distinguishing morphological element used to differentiate the quality of settlements in Nairobi is the presence of vegetation cover, as slums have very little vegetation, compared to formal areas. The difference in morphology between slum and formal settlements in Nairobi is typical, as explained in Section 2.2.1. The notable distinctive morphological attributes between slum and formal settlements after lack of vegetation are high building densities, small building sizes and low building heights, as they are the most explicit slum characteristics (Scott et al., 2017; Taubenböck & Kraff, 2014).

4.   METHODOLOGY

4.1.   Methodological Approach

The overall workflow provided in Figure 3, is structured in 3 stages following the objectives in Section 1.4. Stage 1 involved identifying and deriving flooding factors, both morphological and FIFs, as independent and explanatory variables for the FSM. Stage 2, focused on creating a flood inventory (flood (1) and no-flood points (0)), a critical step in FSM (Sarkar & Mondal, 2020), to be used as dependent variables. Lastly, flood susceptibility was modelled in stage 3.

Each stage is related to the next providing this research with a holistic approach, using various methodologies and approaches to accomplish the research objectives. Objective 1 employed the use of morphometrics (urban morphology metrics) to quantify urban morphological characters related to flooding. Objective 2 sought the use of Fast Flood - a browser simulation tool, and Citizen Science (CS) techniques to create a flood inventory while Objective 3 employed a RF model for FSM.

Fast Flood Simulation Rationale

Challenged by data scarcity in flood information such as historical flood extents and flood depths, this research drew upon Fast Flood[4] – an open-source browser simulation tool developed in 2022 by Van Den Bout et al., (2023), serving as a novel alternative for acquiring flood information. The purpose of this approach was to derive substitute flood and no-flood data needed for FSM given the lack of such data, especially from historical, government flood records and satellite imagery. Fast Flood simulation results were combined with flood validation data based on citizen and expert perceptions (CS data) to create the flood inventory. Despite the existence of other flood models, physically based models that represent physical processes that govern flooding such as OpenLisem are computationally expensive (Van Den Bout et al., 2023) compared to Fast Flood, which offers a short computing time as it is 1500 times faster, hence its selection for this study.

Random Forest Model Rationale

Following the FSM approaches of Lee et al., (2017); Lin et al. (2023); Pourghasemi et al. (2020) RF, a supervised ensemble ML model introduced by Breiman (2001), was used to predict the probability of flooding. The preference for RF in this research was influenced by its wide use in data-driven modelling in the water resources field, its ability to handle imbalanced data, and its low sensitivity to multicollinearity (Chen et al., 2020; Safaei-Moghadam et al., 2023).

Additionally, RF can handle large and high dimensional data, it reduces challenges of overfitting related to decision tree models due to its characteristic of conducting majority voting and is capable of interpreting complex non-linear relationships between flood explanatory factors (Farhadi & Najafzadeh, 2021; Golkarian et al., 2018; Yu et al., 2023). Moreover, given that RF was highly preferred as it efficiently handles diverse datasets from various sources, does not make statistical assumptions regarding the distribution of data and has a higher predictive performance (Breiman, 2001; Prasad et al., 2006).

---

[4] [FastFlood | FastFlood website, free super-fast flood mapping tool.](FastFlood)

Data sources

Data Preparation

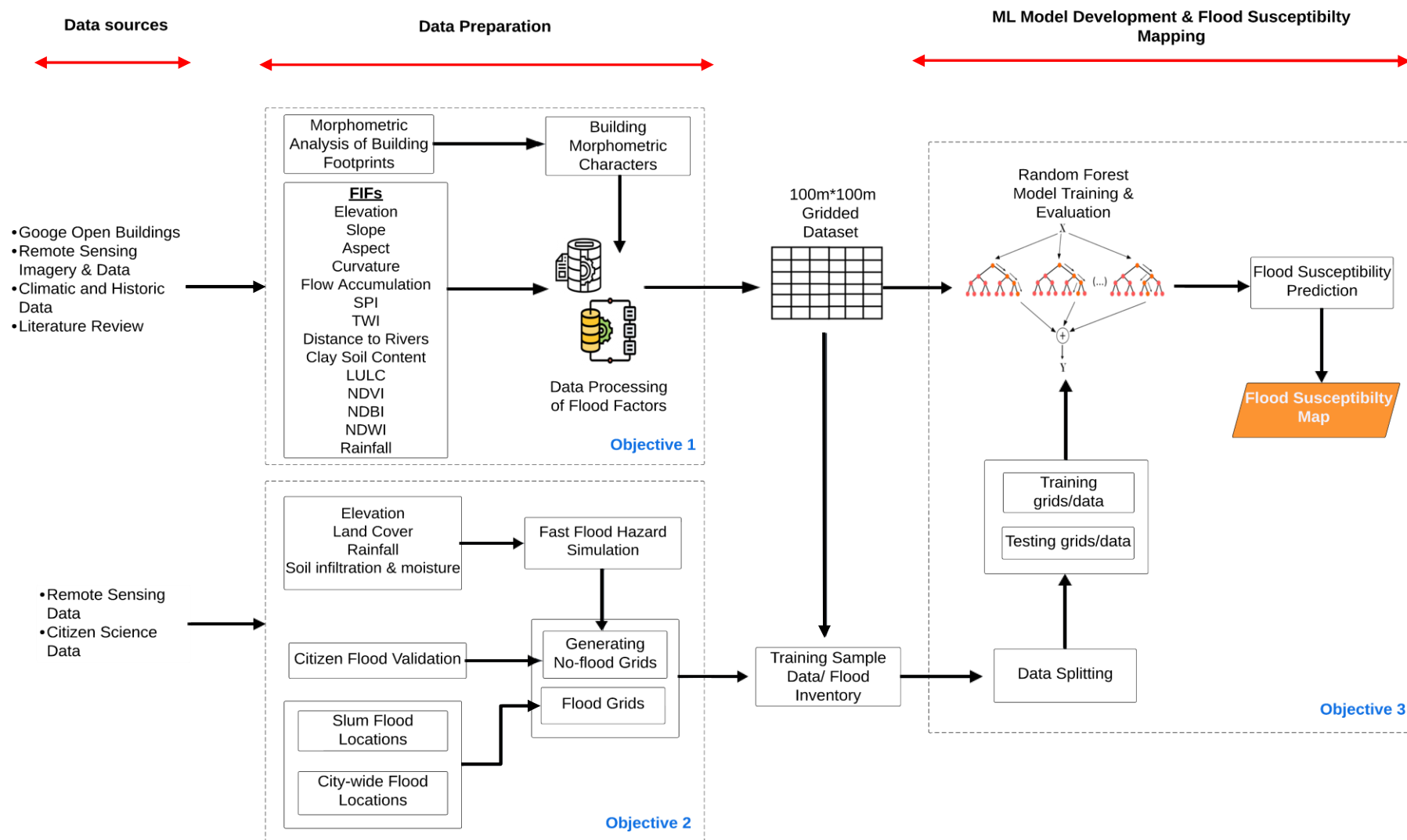**ML Model Development & Flood Susceptibilty Mapping**

Figure 3: Research Methodological Flow Chart

4.2. Data and Data Sources

This research utilized both primary and secondary data at various stages and for different objectives. Geospatial information and data (secondary data) were obtained from global remote sensing and satellite imagery datasets from sources such as Google Earth Engine (GEE) and Open Topography. Primary data such as flood hotspot points was gathered from residents and city experts during fieldwork. Table 3 below provides an overview of data and data sources with further explanations provided in Sections 4.3 and 4.4

Table 3: Overview of Data and Data Sources

Objective	Data Required			Source
1	Building Footprints		Secondary	Google Open Buildings
	FIFs	Digital Elevation Model (DEM) and derivatives (slope, curvature, aspect, rivers, TWI,SPI, flow accumulation		Open Topography
		LULC		GEE
		NDVI		
		NDWI		
		NDBI		
		Rainfall		
		Soil		ISRIC
		Waste		SLUMAP Project
2	Flood inventory	Flood points	Primary	Fieldwork
		Non-flood points		Flood hazard

4.2.1. Secondary data

As FSM necessitates the use of relevant and effective flood factors, an extensive literature review was conducted to inform the selection of pertinent flood factors. Associated flood influencing building morphology factors referred to in Section 2.2.3, were quantitatively derived based on building footprints. FIFs based on data availability included elevation, slope, aspect, curvature, flow accumulation, Stream Power Index (SPI), Topographic Wetness Index (TWI), distance to rivers, clay soil content, Land Use and Land Cover (LULC), Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), Normalized Difference Wetness Index (NDWI) and rainfall.

Factors such as drainage and waste identified to increase flooding were not used in FSM due to the unavailability of drainage data and city-wide waste data (waste data was limited to slums). However, these anthropogenic factors influence flooding hence their inclusion in the ranking of flood factors during the fieldwork workshops (explained in Section 4.2.2).

Input factors for the Fast Flood Simulation (FFS) such as elevation, land cover, soil infiltration and soil moisture were automatically downloaded from their respective global datasets within Fast Flood.

4.2.2. Primary Data - Fieldwork Data Collection

4.2.2.1. Participant Selection

Prior to fieldwork, a flood hazard map (uncalibrated) was simulated using Fast Flood (see Section 4.4.2.1) which formed the basis for selecting slum settlements to collect data from. Given the financial and time constraints, it was not feasible to conduct fieldwork in all slums, therefore, slum selection was based on:

- High severity of flooding identified from the uncalibrated simulation.
- Slums that are constantly faced with harsh effects of flooding based on news reports.
- Slums with high SPI values as high values pose a great risk due to the strong erosive capability of flowing water.

Kibera, Mukuru, Mathare and Korogocho were selected as they befitted the criteria. In collaboration with Community Mappers, a Community-Based Organization (CBO) affiliated with Slum Dwellers International (SDI), community leaders from each settlement selected 10 participants to be involved in the fieldwork activities. Emphasis was placed on ensuring participant diversity in terms of age, gender, flood-related knowledge and roles in the settlement (e.g. Red Cross disaster risk respondent). Additionally, to ensure spatial representation and coverage, especially for validation, the selected participants were from different villages within their settlements.

To address flooding in the formal settlements and at a City level, 23 city experts from Nairobi City County Government (NCCG), Non-Governmental Organizations (NGOs) such as the African Population and Health Research (APHRC) and SDI and academics, accepted an invite to participate in the fieldwork activities. The selection of these institutions and individuals was based on their existing relationships with Space4All[5] and their expertise in flooding in Nairobi.

4.2.2.2. Fieldwork Activities

Following a CS approach, flood validation and flood factor ranking were done using expert and citizen knowledge (Molinari et al., 2019) to determine the performance of Fast Flood, the validity of the uncalibrated FFS and understand the location-specific importances of the flood factors.

The participatory activities were conducted in 5 workshops, 1 for each slum settlement and 1 for the city experts. Additionally, flood locations were collected in the selected slums using My Maps tool – a Google-based tool used for spatial data collection and handling, spearheaded by the Space4All project. The purpose of using My Maps was to have readily available and easily shareable flood data, especially for the settlements with the additional benefit of increasing flood awareness for flood risk measures.

Flood Validation: Using preliminary Fast Flood results (uncalibrated simulation), flood maps were generated for Kibera, Mukuru, Mathare, Korogocho using the city-scale simulation. To validate the maps smaller groups were created to enhance active participation from all, participants were divided into 2 groups of 5 individuals each. The groups were diverse, considering participant characteristics such as gender, age, and the slum village of residence

Using marker pens, pins and sticky notes, the participants (slum and city experts) sketched False Positive (FP) and False Negative (FN) areas, pinpointed locations where structures typically get destroyed and flood hotspots, giving short descriptions of flood characteristics and their impacts (see Appendix 1). Thereafter each group presented their results explaining and discussing their validation processes and outcomes to the larger group, where they confirmed each other's results to be accurate.

Factors Ranking: The flood factors were grouped into 7 broader and general groups to allow for easy understanding as shown in Appendix 2. Pair-wise ranking, a participatory approach, was used to rank the grouped factors allowing individual factors to be compared in a pair (Vallely et al., 2007). The ranking was

---

[5] Space4all (itc.nl)

done collectively which allowed for the group to discuss and debate before voting on the most important factor in a given pair. Moreover, it allowed all involved participants to view their opinions providing more context. Voting to determine the most important factor between a given pair of factors was done by a show of hands, with the factor that received more votes, chosen as the most important.

A pairwise matrix was used to analyze a specific (constant) factor against other factors, with the final ranking based on the number of occurrences of the constant factor in its specific column of comparison (see

Appendix 3). To determine the overall importance of the factors, the most important factor was determined by majority votes and the least important factor by the least votes.

My Maps flood locations: 2 representatives from each slum were trained to collect data using My Maps to prevent inaccuracies as a result of user error. Flood data collected was based on a defined attribute table that detailed the type of information such as flood depth, duration etc., to be recorded by the residents depicted in Appendix 4. In aims of benefitting from their local knowledge and experiences, they collected flood locations that are prone to flooding based on previous and recent flood events (as of April 2024).

Additionally, during fieldwork 2 transect walks, accompanied by residents, were conducted along the rivers traversing Mukuru (Ngong River) and Mathare (Mathare River) to collect flood depth and their locations. Flood depth measurements at various accessible points were carried out using Image Meter Pro – a measuring application, using existing flood depth markings drawn by the Red Cross after extreme flood events and the highest visible watermarks on buildings.

4.3. Deriving Flood Factors

4.3.1. Urban Morphometrics

Meaningful numerical measurements of building morphological characters following the morphometric approach adopted by Fleischmann et al. (2022) was conducted using building footprints. Google Open Buildings (GOB) dataset – a large-scale open-source building footprint dataset, derived from high-resolution imagery (50cm) was used to obtain building footprints using Google Colab[6] and were inspected using GEE [7]. The research used GOB version 3 with inference carried out in May 2023 covering 58,000,000 km² of Africa, South and South-East Asia, Latin America. These building footprints were generated using deep learning algorithms, resulting in confidence score ranges of (0.65 – 0.70), (0.7 – 0.75) and (> =0.75).

Momepy - an open-source Python library with a repository of tools for morphometric assessments was used to derive building morphometric characters[8]. A selection of morphometric characters related to flooding across 3 urban form categories: (i) shape (ii) spatial distribution, and (iii) intensity recognised by Fleischmann et al. (2021) were computed as in Table 4. However, building density was implemented at the grid level (see section 4.5).

---

[6] Open Buildings - download region polygons or points. - Colaboratory (google.com)

[7] https://code.earthengine.google.co.in/467492e7eb648fbdd6e65540dcbfa0a6

[8] http://docs.momepy.org/

Table 4: Computed Morphometrics

Morphometric characters	Urban form categories	Function/Formula	Units	Alias
Orientation	Spatial Distribution	momepy.Orientation	Range of 0-1	Orientation
Alignment		momepy.SharedWallsRatio	Range of 0-1	Alignment
Share walls ratio		momepy.Alignment	Range of 0-1	SWR
Neighbour distance		momepy.NeighborDistance	Meters	NDi
Mean inter-building distance		momepy.MeanInterbuildingDistance	Meters	IBD
Building adjacency		momepy.BuildingAdjacency	Range of 0-1	BuA
Fractal dimension	Shape	momepy.FractalDimension	Range of 0-1	FD
Covered area ratio	Intensity	momepy.AreaRatio	Range of 0-1	CAR
Building density		$\frac{\text{Number of buildings}}{\text{Area}}$	Buildings/10,000m^2	BD
*Values close to 1 for covered area ratio and shared wall ratio suggest high covered area and shared walls				
*Values close to 1 for fractal dimension indicate simple building shapes.				
*Values close to 0 for building adjacency denote that buildings are close to each other				

Before generating the measurements, multi-polygons were split into individual polygons and the dataset was projected into a (local) coordinate system for calculation purposes. Using Voronoi tessellations, morphological tessellations were created from the building footprints (Fleischmann et al., 2022), as the smallest spatial division, they were used to identify topological relationships between buildings and measure spatial distribution characters (Fleischmann et al., 2021).

4.3.2. Flood Influencing Factors

4.3.2.1. Elevation Derivatives

Elevation-based factors such as slope, aspect, curvature, flow accumulation, TWI and SPI provided in Table 5, were derived from a Copernicus 30m DEM acquired from Open Topography[9]. Despite the availability of higher resolution DEMs such as ALOS Plasar with a resolution of 12.5m, the Copernicus DEM was preferred as its processing was done to include the consistent flow of rivers and its higher accuracy detailed in Section 4.4.2.1. Using ArcGIS Pro, the DEM was reprojected to WGS 1984 UTM Zone 37S (EPSG 32737), followed by DEM processing to derive slope, aspect and curvature using the *surface spatial analyst tool*. Using terrain preprocessing techniques, flow accumulation, SPI, TWI, and rivers were derived.

Following a heuristic approach for stream definition, flow accumulation values for generating rivers were conditioned to an optimal threshold of 800 (flow accumulation pixel value) as the threshold captured major rivers as well as significant streams, especially in slums. Visual inspection using Google Earth highlighted that the generated rivers had an offset from the rivers, a disparity presumed to be caused by the resolution of the DEM. The distance to rivers (meters) was calculated using the *Euclidean distance tool* in ArcGIS Pro as informed by (Choubin et al., 2023).

---

[9] OpenTopography - Copernicus GLO-309 Digital Elevation Model

The TWI was calculated using the *calculate Topographic Wetness Index (TWI) using surface parameters tool,* while SPI was calculated using the *raster calculator tool* in ArcGIS Pro, both using the previously derived slope and flow accumulation rasters based on the formula adapted from Sevgen et al. (2019) as below:

TWI = ln(A_s/tanβ)
SPI = A_s/tanβ
Where:

A_s = flow accumulation/specific catchment area
β = slope (radians) = (slope (degrees) * 0.017453)

4.3.2.2. Sentinel and Landsat Products

GEE was used to access and derive NDWI and NDVI from the Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-1C 10 meters resolution[10] (see Table 5). A 1-year (2023/01/01 – 2024/01/01) composite was generated to get the mean NDVI and NDWI values instead of using values that are only present during certain seasons of the year. The imageries used for the composite were based on scenes with less than 10% cloud cover to avoid interference from cloud reflection.

The reflectance in the near-infrared (NIR) band (B8) and the reflectance in the red band (B4) were used to calculate NDVI (Choubin et al., 2023; Kuc & Chormański, 2019). Whereas the reflectance in the green band (B3) and the reflectance in the near-infrared (NIR) band (B8) were used to calculate NDWI (Du et al., 2016), using the following equations:

NDVI = *(B8 – B4) / (B8 + B4)*
NDWI = *(B3 – B8) / (B3 + B8)*

Similarly, the mean NDBI values were derived from a 1-year Landsat-8 OLI/TIRS composite (2023/01/01 – 2024/01/01) with a resolution of 30m, using scenes with less than 10% cloud coverage from GEE[11]. The built-up index was calculated using the reflectance in the middle infrared reflectance (MIR) band (B6) and the reflectance in the NIR band (B5) (Kshetri, 2022) as below.

NDBI = *(B6 – B5) / (B6 + B5)*

LULC for the study was derived from Sentinel 2 ESA WorldCover 10m v100[12] using GEE at a resolution of 10 meters.

4.3.2.3. CHIRPS Rainfall Intensity Product

CHIRPS[13] a 30+ year quasi-global satellite-based rainfall product with in-situ station data was preferred due to its fine spatial resolution of 0.05º which improves the efficacy of CHIRPS in capturing precipitation heterogeneity, yielding better performances in hydrological modelling (Dhanesh et al., 2020). Using a 5-year

---

[10] Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-1C | Earth Engine Data Catalog | Google for Developers

[11] USGS Landsat 8 Level 2, Collection 2, Tier 1 | Earth Engine Data Catalog | Google for Developers

[12] ESA WorldCover 10m v100 | Earth Engine Data Catalog | Google for Developers

[13] CHIRPS Daily: Climate Hazards Group InfraRed Precipitation With Station Data (Version 2.0 Final) | Earth Engine Data Catalog | Google for Developers

time range (2019/01/01 – 2024/01/01), the highest precipitation (115.194 mm/day) recorded on 2021/11/28 was used to generate a gridded rainfall intensity raster at a resolution of 5566m from GEE.

4.3.2.4. Soil Data

Soil clay content in percentage was derived from the Soil and Terrain Database for Kenya (KENSOTER) Version 2 at a scale of 1:1 million, compiled by the Kenya Soil Survey via the International Soil Reference and Information Centre (ISRIC) World Soil Information[14]. With soil properties such as texture and drainage typically used for flood mapping, the lack of data on such properties in a somewhat large section of the study area was missing, with the only available soil data for the 'no-data' section being clay percentage.

4.3.2.5. Waste Data

Waste data was acquired in a gridded format of 100m*100m from a land cover classification modelled from WorldView-3 2019 VHR imagery using image segmentation an Object-Based Image Analysis (OBIA) technique and RF classification by SLUMAP[15], a RS project for mapping and characterizing slums in SSA cities.

Table 5: Derived Flood Influencing Factors (FIFs)

Data	Data Type	Dataset	Alias
1. Elevation 2. Slope 3. Aspect 4. Curvature 5. Flow accumulation 6. SPI 7. TWI 8. Distance to rivers	**Dataset:** Global **Type:** Raster **Resolution**: 30m **Units:** • Elevation – Meters • Slope – Degrees • Distance to rivers – Meters	**Source:** GEE **Dataset:** Copernicus GLO-30 Digital Elevation Model	Elevation Slope Aspect Curvature Flow_Acc SPI TWI DistRivers
9. LULC	**Dataset**: Global **Type:** Raster **Resolution:** 10m	**Source**: GEE **Dataset**: Sentinel 2 ESA WorldCover 10m v100 Copernicus GLO-30 Digital Elevation Model	LULC
10. NDVI		**Source:** GEE **Dataset:** Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-1C	NDVI
11. NDWI			NDWI
12. NDBI	**Dataset:** Global **Type:** Raster **Resolution:** 30m	**Source:** GEE **Dataset:** USGS Landsat 8 Level 2, Collection 2, Tier 1	NDBI
13. Rainfall	**Dataset:** Global **Type:** Raster **Resolution:** 5566m **Unit:** Millimetres	**Source:** GEE **Dataset:** CHIRPS daily: Climate Hazards Group InfraRed Precipitation with Station Data (Version 2.0)	Rainfall

---

[14] ISRIC Data Hub

[15] SLUMAP research project - Mapping slums with remote sensing (ulb.be)

14. Soil clay content	**Dataset:** Global **Type:** Vector **Scale:** 1,000,000 **Unit:** Percentage	**Source:** ISRIC Data Hub **Dataset:** Soil and Terrain Database for Kenya (KENSOTER), version 2.0	CLPC
15. Waste	**Dataset:** Local **Type**: Raster **Scale:** 100m*100m Grid level **Unit:** Percentage	**Source:** SLUMAP **Dataset:** WorldView-3 Imagery	

4.4.   Flood Inventory Creation

The flood inventory procedures were done after fieldwork which entailed flood validation and the collection of flood hotspot locations. Given that the unit of analysis for this research was at a built-up grid level of 100m*100m (explained in Section 4.5), the inventory was created after the flood factors were processed and integrated into the grid. This was undertaken to ensure that (i) the flood inventory points, particularly the flood hotspots collected using My Maps intersected with their respective grid cells, and (ii) the corresponding flood inventory grid cells had no missing values and belonged to built-up grids.

4.4.1.   Flood Sample Points

The flood sample points were combined from (i) flood points identified by experts during their workshop, providing city-wide level flood locations (both in formal and slum settlements) and (ii) flood points collected in the selected slums using My Maps. The identified flood points were based on historical and recent flood (April 2024) events and totalled to 134 observed flood points, with 93 points from My Maps and 41 points digitized from the expert flood validation workshop using ArcGIS Pro.

Upon intersection with the grids, 116 points intersected with the grid cells, indicating that 18 flood points were observed in non-built-up areas. These 116 flood grids were considered susceptible to flooding regardless of flood depth and were labelled as belonging to class 1 to be used for RF model training and evaluation.

4.4.2.   No-flood Sample Points

As standard practice (Al-Aizari et al., 2024; Towfiqul Islam et al., 2021; Youssef et al., 2022), an equal number of no-flood samples were generated based on a no-flood extent simulated and calibrated using Fast Flood.

4.4.2.1.   Fast Flood Simulation (FFS)

Elevation, land cover, soil infiltration and moisture data were automatically downloaded from their respective global datasets within Fast Flood, while rainfall data was sourced from a local dataset (see Table 6) were used to simulate a previous flood event. To account for upstream flooding, the catchment area (Athi River Basin), illustrated in Appendix 5 was included in the simulation. This allowed for rainfall-driven modelling, which ensured that all the rainfall that contributed to the flood event was considered in the simulation.

Table 6: Fast Flood Input Parameters

Fast Flood input data	Fast Flood Data source			Data Type
Elevation	Copernicus	OpenTopography - Copernicus GLO-30 Digital Elevation Model	Global dataset	Type: Raster Resolution: 40m Units: Meters

LULC & Surface Roughness (Mannings' N)	Copernicus World Cover 10m product	ESA WorldCover 2020	Global dataset	Type: Raster Resolution: 10m
Rainfall	TAHMO	About TAHMO - TAHMO	Local dataset	Type: Excel rainfall station data Units: Millimetres
Soil infiltration Soil moisture	SoilGrids	SoilGrids250m 2.0	Global dataset	Type: Raster Resolution: 250m

Elevation: The Copernicus DEM was auto-downloaded at a resolution of 40m to reduce the computational demand as the catchment area was relatively large. The Shuttle Radar Topography Mission (SRTM) DEM was not used due to its high vertical errors (Van Den Bout et al., 2023) which could have significantly affected the flow direction, flood extent and inundation depths of the simulation as they are sensitive to even minor errors, specifically in low-elevation areas (Horritt & Bates, 2022; Simpson et al., 2015).

As elevation errors act as artificial sinks or obstacles that either retain or divert simulated flows, resulting in misleading flood results (Meadows et al., 2024), the Copernicus DEM was preferred since newer DEMs derived from the TanDEM-X Mission such as Copernicus have better accuracies and are highly recommended for flood mapping following vertical accuracy assessment studies such as one by Meadows et al. (2024).

Surface Roughness Mannings: The frictional force applied by the terrain on flowing water determined by land cover captured using Manning's value (Van Den Bout et al., 2023) was used in the FFS to account for the different surface run-off and discharge responses of various land cover typologies. Manning's values used in the FFS were default values provided by Fast Flood after automatically downloading land cover from the European Space Agency (ESA) World Cover 10m land cover dataset.

As the default Manning's values were seen to correspond with other globally used Manning's values such as the LULC 2020-ESRI provided by Soliman et al. (2022) presented in Appendix 6, the default values were used as provided in the simulation.

Rainfall: Rainfall parameters such as duration, intensity and rain shape were generated using the rainfall analysis tool in Fast Flood based on precipitation data (mm/hr) acquired from the Trans-African Hydro-Meteorological Observatory (TAHMO)[16] over 5 years from January 2019 to January 2024. The precipitation measurements for 13th May 2021 were used for the FFS as:

1. The precipitation resulted in significant flooding explained to be a 1-in-5-year event by Juma et al. (2021).
2. The precipitation event had the highest rainfall peak of 62.017 mm/hr with a total precipitation of 100.74 mm.
3. The precipitation date was used for flood mapping in Kibera by Juma et al. (2023) verifying that it was an extreme event.

---

[16] About TAHMO - TAHMO

Using (i) a comma-separated value file, with hourly rainfall measurements and (ii) 99% of the total precipitation in the rainfall analysis (as it produced a representative rainfall peak intensity of 62.748mm/hr similar to that of the precipitation event). The analysis tool provided a rainfall intensity of 10.103 mm/hr, duration of 9 hours and rain shape of 3.105 as rainfall parameters for flood simulation (Appendix 7).

Soil data: Soil infiltration and moisture information auto-downloaded from SoilGrids for the FFS used default soil depths of 5-15cm. However, the majority of the built-up area did not have soil information as a result of soil sealing and the presence of thin soil as noted by Van Den Bout et al. (2023). The FFS relied on the default infiltration and moisture rates provided by Fast Flood.

Two FFS were modelled, an (uncalibrated) intermediate FFS used for fieldwork flood validation and a subsequent (calibrated) FFS after fieldwork.

FFS Validation

The validation of the FFS employed a fusion of quantitative and qualitative validation techniques from both primary and secondary data sources. Validation was carried out twice, on an uncalibrated simulation and on a calibrated simulation. The uncalibrated simulation was validated by (i) using expert and resident knowledge and expertise as explained in Section 4.2.2.2 and (ii) comparing the simulation flood extent and depth with the location and depths of the observed flood points obtained from My Maps and the transect walks.

After calibration, the calibrated simulation was validated using a satellite-observed flood extent acquired from UNOSAT using PLEIADES imagery (50cm resolution) for a flood event that occurred in May 2024. Validation was performed by identifying the True Positive (TP) (common areas between the satellite flood extent and the calibrated FFS). Additionally, provided with a hydrodynamic flood model for Kibera, FFS validation was also done at a local level using Kibera slum.

Kounkuey Design Initiative (KDI) - an international NGO working on projects in Kibera, through one of their PhD students Juma et al., (2023), provided the Space4All team with a flood model for Kibera. The flood model was modelled using a 1D/2 HEC-RAS model using a 5m LiDAR DEM, land use Manning's N, and precipitation measurements for 13th May 2021, making it suitable for benchmarking as both the KDI model and calibrated FFS modelled the same event. Similar to the satellite-observed flood extent, the TP areas between the hydrodynamic model and FFS were derived.

FFS Calibration

The flood hotspot points collected using My Maps (see Section 4.2.2.2), were used as input for calibration as they contained flood depth information (see Appendix 8). Flood depth was recorded using different parts of the human body, such as the ankle, knee, waist and above-the-head level as it provided residents with visual estimations and references for measuring and describing depth levels easily (Sy et al., 2020). Flood depths were converted into quantitative flood heights of 0.10, 0.50, 1.2 and above 2 meters respectively before calibration, following a similar methodology by Luke et al. (2018).

Sensitivity analysis, the most important part of calibration was implemented to examine the interaction between model parameters and the obtained simulation results (Wałęga, 2016). Calibration was carried out using 5 variations for infiltration and Mannings' parameters each, chosen on regular intervals between 50% and 150% of their original values (Van Den Bout et al., 2023). 25 simulations for each unique combination of the two parameters were done based on the intervals of original values and the number of calibration parameters.

The optimal parameter settings (multipliers) provided by the calibration process were 1.5 for Mannings' and 0.5 for infiltration (see Appendix 8), which were used to calibrate the intermediate simulation to produce a calibrated flood simulation.

4.4.2.2. Classification of flood and no flood extent

The calibrated FFS was classified into five depth ranges of 0-0.11m, 0.11-0.45m, 0.45-1.0m, 1.0-1.7m and above 1.7m borrowing from Sanders et al. (2023). Depths ranges of 0-0.03 and 0.03-0.11 were merged since depths below 0.03m are considered nuisance flooding due to their very-low level of inundation (Moftakhari et al., 2018). Various studies employed different depth thresholds such as 0.30m by Lin et al. (2023) and 0.10m by Ward et al. (2013) arguing that the threshold depths have minimal impact on livelihoods.

As this research used Sanders et al. (2023) flood classification depths, and the 0.10m threshold being widely used, 0.11m was used as the flood threshold in this research being the lowest hazard level and close to the 0.10m threshold. Thus, depths below 0.11m were considered as no-flood depths and labelled as 0 creating a binary map extent of 0 and 1, with 1 being the flood extent of depths above 0.11m.

4.4.2.3. Extracting no-flood grids

The no-flood extent from the calibrated simulation was intersected with the 100m*100m grids to create no-flood grids based on the 0.11m threshold. 'Pure' no-flood grids for model training, were created by excluding the known flood grids from the overall grids, which comprised the 116 flood sample grids and grids intersecting with slum polygons known to flood, namely Kibera, Mukuru and Mathare.

Using a k-means grid-based clustering approach, the resultant grids (after excluding flood grids) were clustered using elevation following an approach by Wang et al. (2023) to get no-flood grids. However, given that Wang et al. (2023), used high-elevation areas, this study used all the elevation clusters as flooding was reported (by the city experts) to occur even in high-elevation areas. Additionally, it was one of the most important flood factors in increasing flooding provided by residents and city experts.

The elevation data was scaled and normalized before clustering (Tardioli et al., 2018) to prevent the influence of outliers on the optimal number of clusters created by the algorithm (Joshi et al., 2022). 5 elevation clusters were created informed by the elbow method (see Figure 4) and exploration of 3, 4 and 5 k-cluster results. Stratified random sampling was then used to select an equal number of no-flood grids per cluster (23) to complement the 116 flood grids.
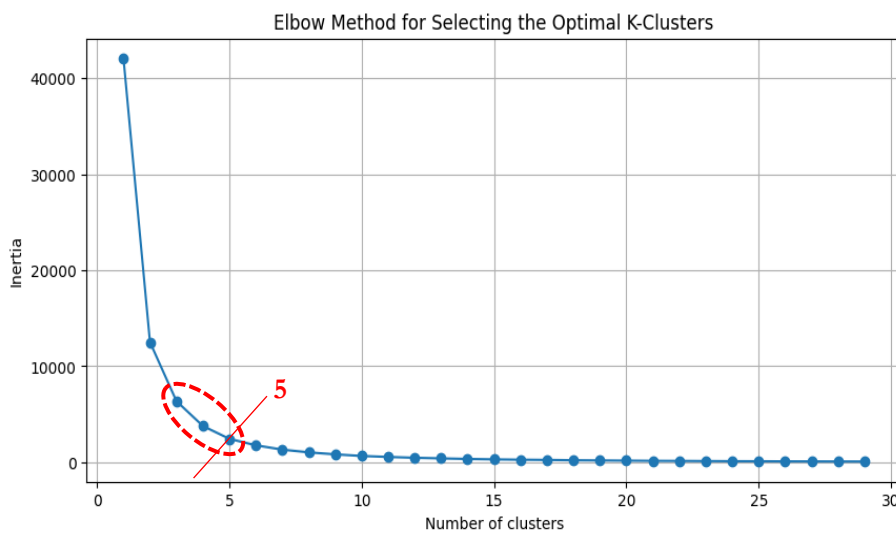

Figure 4: Elbow Method Determining Optimal K-Clusters

4.4.3.   Training Sample Grids/ Flood Inventory

The flood grids and no-flood grids were combined to create training sample grids totalling to 231 grids. The flood points obtained from the experts and residents were not based on any spatial selection, but rather on areas that flood regardless of flood depth. Consequently, the distribution of the flood points varied across the study area resulting in cluster imbalance, with some clusters having more training grids, than others due to the flood grid imbalance as depicted in Appendix 16 .

4.5.   Data Preparation and Processing

The unit of analysis used for this research was at a grid level of 100m*100m, as grid cell delineation is more dynamic and adaptive to distinguish different built-up form typologies compared to administrative boundaries since it considers the physical layout of cities (Marwal & Silva, 2023). Additionally, in spatial analysis, grid cells provide a high degree of spatial granularity that allows for the recognition of subtle spatial patterns and relations, with the 100m*100m grid size conforming to most global datasets and protecting population privacy (Kuffer et al., 2021).

The grid was created in ArcGIS Pro using the *Grid Index Feature* referencing a 100m resolution snap raster, to ensure alignment between the grid and the FIFs raster datasets resampled into a 100m resolution using the same snap raster. The building morphometrics (vector), and the FIFs (vector and raster) were combined at a grid-cell level, assigning flood factor attributes to each grid cell. The morphometric dataset was spatially joined to the grid, with each grid cell receiving the mean value of the building footprints' morphometrics present within specific grid cells.

The FIFs derived in their respective resolutions were first resampled into resolutions of 100m using a snap raster before being integrated into the grid. This ensured alignment between the rasters and the grid, which would have otherwise resulted in misalignment due to overlapping raster extents. Using the *Extract Values To Points* tool in ArcGIS Pro, the raster values were extracted to the grid centroids and concatenated with the other flood factors at each grid cell. Soil vector data was intersected with the grid cells assigning each cell with soil data.

4.5.1.   Data Cleaning

With a focus on urban morphology specifically building morphology, non-built-up grid cells were discarded from the grid. Additionally, grid cells with null values and -9999 values (no data) were also removed from the grids resulting in a total number of 42,020 grids.

4.5.2.   Data Splitting

Prior to model training, the 231 training sample grids were split into training and testing sample sets using the spatial block technique adopted from Mahoney et al. (2023), which resulted in spatially uncorrelated training and testing set. Informed by the distribution of the training sample grids, a grid of six rows and six columns was created to form the spatial blocks, blocks that did not intersect with any training sample grids were dropped to prevent training and testing blocks from having no data as in Figure 5.

From the upper-left corner, moving horizontally across each row from left to right starting from the upper left block, every 3rd block was selected to be a testing block as seen in yellow in Figure 5. After the iterations, 146 sample grids (14 spatial blocks) were assigned to the training set while 85 sample points (6 spatial blocks) were assigned to the testing set.
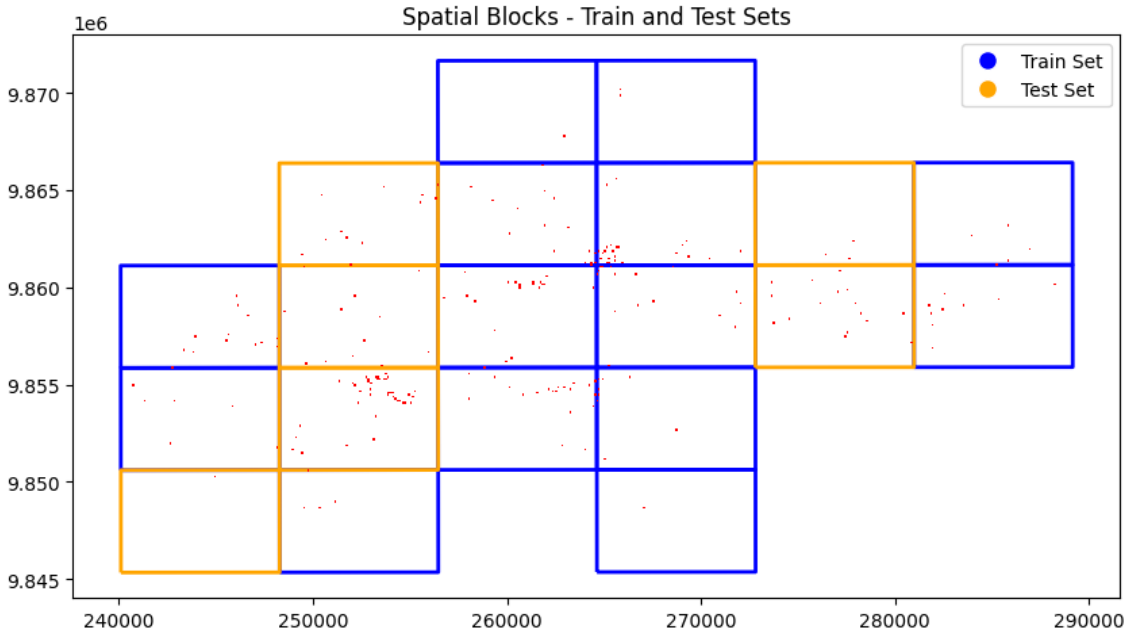
Figure 5: Training and Testing Data Splitting

4.6.   Random Forest Model Training and Evaluation

RF classification was conducted using the Scikit-Learn library in the Python environment. Flood susceptibility prediction was conducted using 2 models, where Model 1 utilized all the flood factors while Model 2 utilized the FIFs. To boost the performance of the models hyperparameters such as the number of trees (n_estimators), the maximum number of features for node splitting (max_features), the maximum level of each decision tree (max_depth), the minimum sample required to split a node (min_samples_split) and the minimum samples required in a leaf node (min_samples_leaf) were optimized (Al-Aizari et al., 2024). A grid search cross-validation technique of five folds was applied for hyperparameter optimization on a defined grid of candidate values (shown below) to derive the best-performing parameter sets.

N_estimators : [100, 200, 300, 400, 500]
Max_features : [sqrt]
Max_depth : [10, 20, 20, 40, 50]
Min_samples_split: [2, 5, 10]
Min_samples_leaf: [1, 2, 4]
Random_state: [42]

The best parameter sets were then used to train the models by fitting the algorithms to the 146 binary classified training grid samples (0 and 1), while the 85 binary testing samples were used to evaluate the performance of the models. The robustness and classification accuracies of the models were evaluated using statistical measures such as overall accuracy, recall and precision, which were calculated based on the True Negative (TN), False Positive (FP), False Negative (FN) and True Positive (TP) parameters (Chen et al., 2020; Pourghasemi et al., 2020) using the formulas given below:

Overall Accuracy = TP + TN/(TP+TN+FP+FN)
Precision = TP/(TP +FP)
Recall = TP/(TP + FN)

Where:

- ➢ TP and TN = correctly classified instances
- ➢ FP and FN = incorrectly classified instances

Feature importance was calculated by the RF models, determining the hierarchical importance of the contribution of each flood factor based on the mean decrease in Gini impurity. The flood factors ranking results from the workshops which represented local flood dynamics (explained in Section 4.2.2.2) were used to compare and evaluate the RF feature importance lists.

4.7.   Susceptibility Prediction

The 42,020 grids generated as explained in Section 4.5.1, were used as a predictor set and were fitted to the trained models, which predicted the class labels of the unlabelled grid cells as no-flood or flood. Using the predict probability function (predict_proba), the probability scores of each grid belonging to its predicted class were calculated based on a threshold of 51% as in Appendix 9, with scores above 51% belonging to class 1 and vice versa.

The susceptibility values were classified using natural breaks and interpreted categorically, ranging from very low susceptibility to very high susceptibility which were then used to assess the degree of flood susceptibility between slums and formal settlements. The influence of urban morphology on flood susceptibility was assessed by comparing the model performances and flood susceptibility maps of both Models 1 and 2.

5. RESULTS

5.1. Flood Factors (Independent Variables)

5.1.1. Building Morphometrics

High values of Shared Wall Ratio (SWR), Covered Area Ratio (CAR) and Building Density (BD) and low values in Building Adjacency (BuA) are depicted in red as in Appendix 11, are seen to explicitly delineate slum settlements. Lower values of BD are dominantly observed to be in East, Southwest and Northwest of the city.

Buildings in the East and Southwest have large distances between each other as illustrated by the Building Neighbour Distance (NDi) and Mean Inter-building Distance (IBD) metrics in Appendix 11, with the mean distance between buildings being 23.36m and 22.56m respectively (see Appendix 10). It is observed that high values of SWR, CAR, BD, and low values of BuA in slums e.g. Kibera are associated with small distances between buildings. In comparison, in some formal settlements such as Kilimani and Kileleshwa, lower values of SWR, CAR, BD, and higher values of BuA are also associated with smaller distances between buildings as seen in Figure 6.
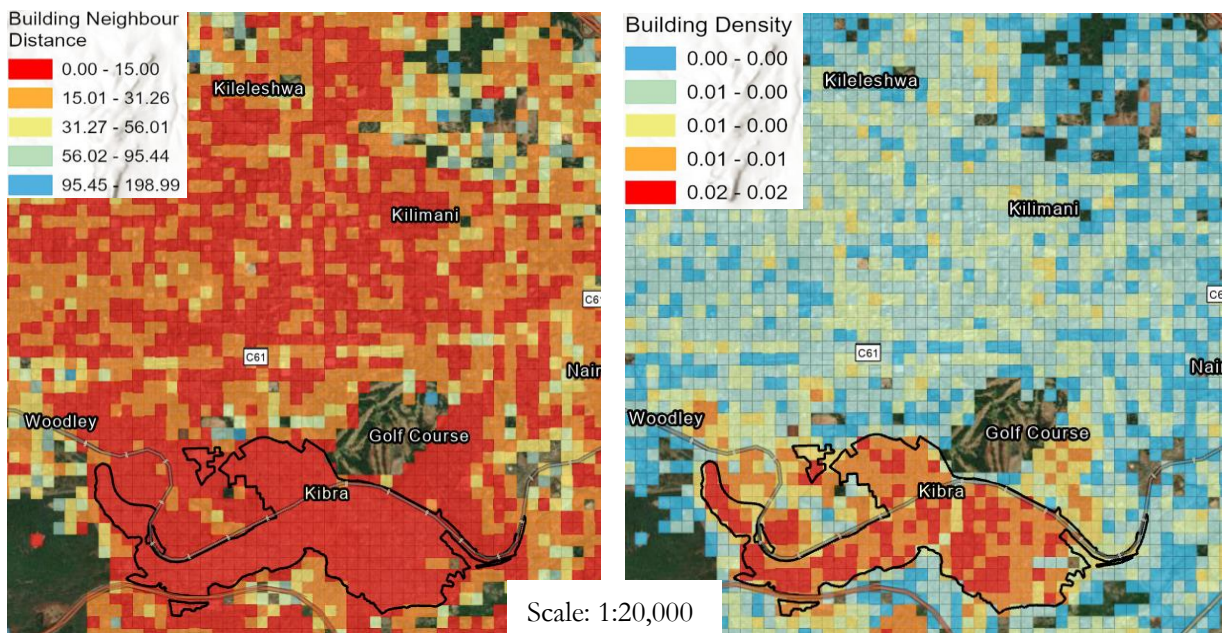

Figure 6: Relationship between Building Neighbour Distance (NDi) and Building Density (BD) in Kibera Slum and Formal Settlements (Kilimani and Kileleshwa).

With Fractal Dimension (FD) having a mean value of 1.034 and a low standard deviation of 0.042 (see Appendix 10). The majority of buildings in Nairobi are assumed to have regular shapes indicated by the mean values being near 1, implying that buildings have minimal departure from Euclidean geometry (Taubenböck et al., 2019).

5.1.2. Flood Influencing Factors

The Elevation values for the study area ranged from 1,462.05 to 1,933.98 meters, with most slums found in the Elevation ranges of 1,573.69 and 1,610.89 meters (depicted in green and dark green) as in Appendix 11. The mean Slope was 3.031° with the highest slope value of 25.618° found in the North-west. The mean

distance to rivers is 303.394 meters (see Appendix 10), with most slums observed to be located along and near rivers.

High values of Flow Accumulation and SPI are observed to be associated with proximity to rivers and vice versa, where the shorter the distance to rivers, the higher the values of Flow Accumulation and SPI (see Appendix 11). The LULC shows that the majority of the city is built-up, with vegetation/tree cover mostly in the West. The mean NDVI value for the study area is 0.0296, with the least NDVI (negative values) located in slum settlements.

Interestingly, a direct relationship is seen between elevation and NDVI and a negative relationship between Elevation and TWI. As elevation decreases from West to East, NDVI decreases in the same manner (from West to East), whilst TWI values increase, implying that the lower the elevation the higher the TWI values as in Appendix 11.

5.1.3. Correlation Amongst Flood Factors

The relationships between independent variables were analyzed using the Pearsons's correlation as shown in Appendix 12 to understand the relationship between variables. The morphometric variables showed strong positive correlations among NDi, IBD and BuA, with CAR and BD having a strong positive relationship of 0.87.

The strongest negative relationship of -0.96 is observed between NDVI and NDWI, suggesting that higher NDVI values are associated with lower NDWI values. Conversely, NDVI and NDBI had the strongest positive relationship of 0.75. Curvature, FD and distance to rivers were seen to have weak correlations with other variables, as shown by their low correlation values.

5.1.4. Covariance and Correlation between Variables and Flood Classes

To understand how the variables relate to the flood labels, the covariance and correlation were done, providing a comprehensive overview of which variables are strongly related to flooding or no flooding. SPI and Flow accumulation had the highest positive covariances of 1337.48 and 1238.55 respectively depicted in Appendix 13, with distance to rivers having the highest negative covariance of -49.17.

According to the correlation results, CAR, BD, and SWR are portrayed to have significant magnitude towards flooding given their high correlations of 0.51, 0.40 and 0.36, respectively, seen in Figure 7. BuA, IBD and distance to rivers showed strong negative relationships with flooding indicated by their correlations of -0.55, -0.47 and -0.47 respectively. Building orientation, elevation, curvature and soil clay content had values closer to 0, implying either minimal or no linear relationship to flooding based on the data.
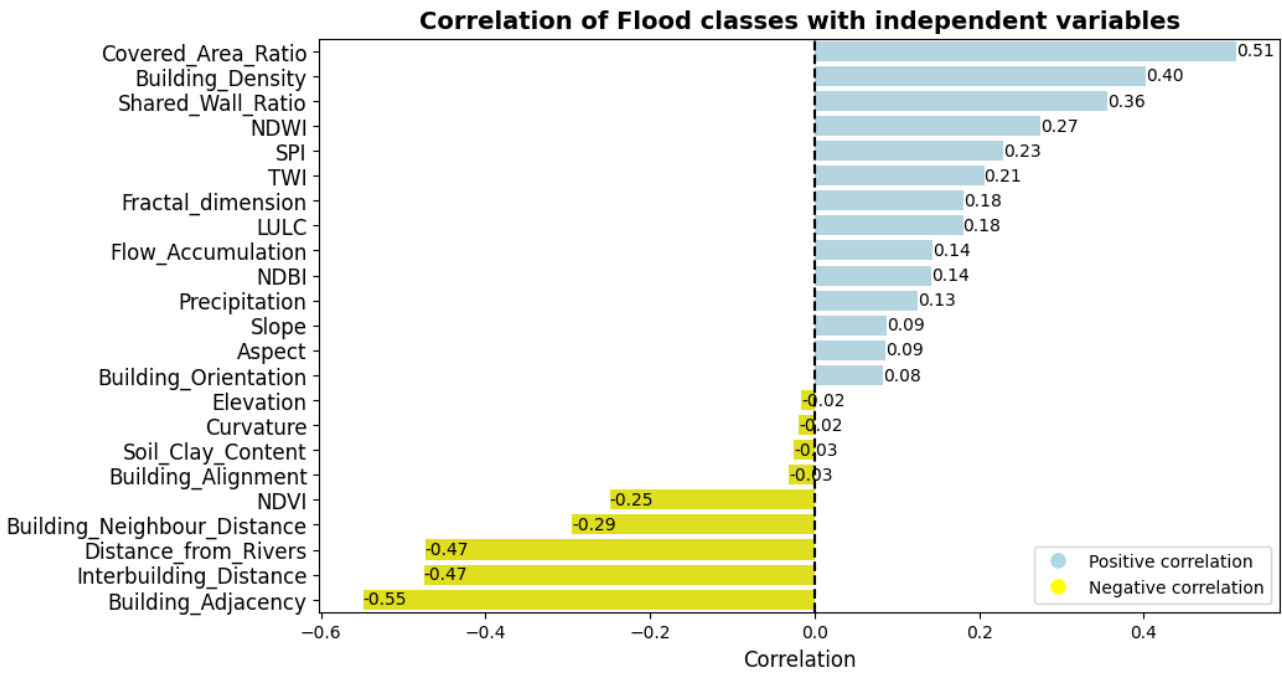
Figure 7: Correlation of Flood Factors with Flood Labels

5.2. Flood Inventory

The Fast Flood uncalibrated simulation shown in Appendix 14 shows extensive flooding in various parts of the study areas with a maximum flood depth of 17.51m. The flood depths on the Western side are seen to be low compared to the Eastern side which has depths above 1.7m. High flood depth values are seen along the major rivers, where most of the slums are located. Hence observing that the slums are more exposed to fluvial flooding.

5.2.1. Fast Flood Calibration

A new simulation was run using My Maps flood observation points (91 points) resulting in a maximum depth of 16.47m, indicating a maximum depth decrease from the uncalibrated simulation. The simulation with observation points was then calibrated using multipliers of 0.5 and 1.5 for infiltration and Mannings', respectively, resulting in a calibrated FFS with a maximum depth of 13.82m illustrated Figure 8.

The calibrated simulation had lower flood depths compared to the uncalibrated simulation with a difference of 3.69m between their maximum flood depths. Even after using observed flood points located far from rivers to calibrate the simulation, both simulations were seen to dominantly capture fluvial flooding. Despite the decrease in depth, the extent of the calibrated simulation significantly reduced in the Eastern region compared to the uncalibrated simulation, as seen in Appendix 15.
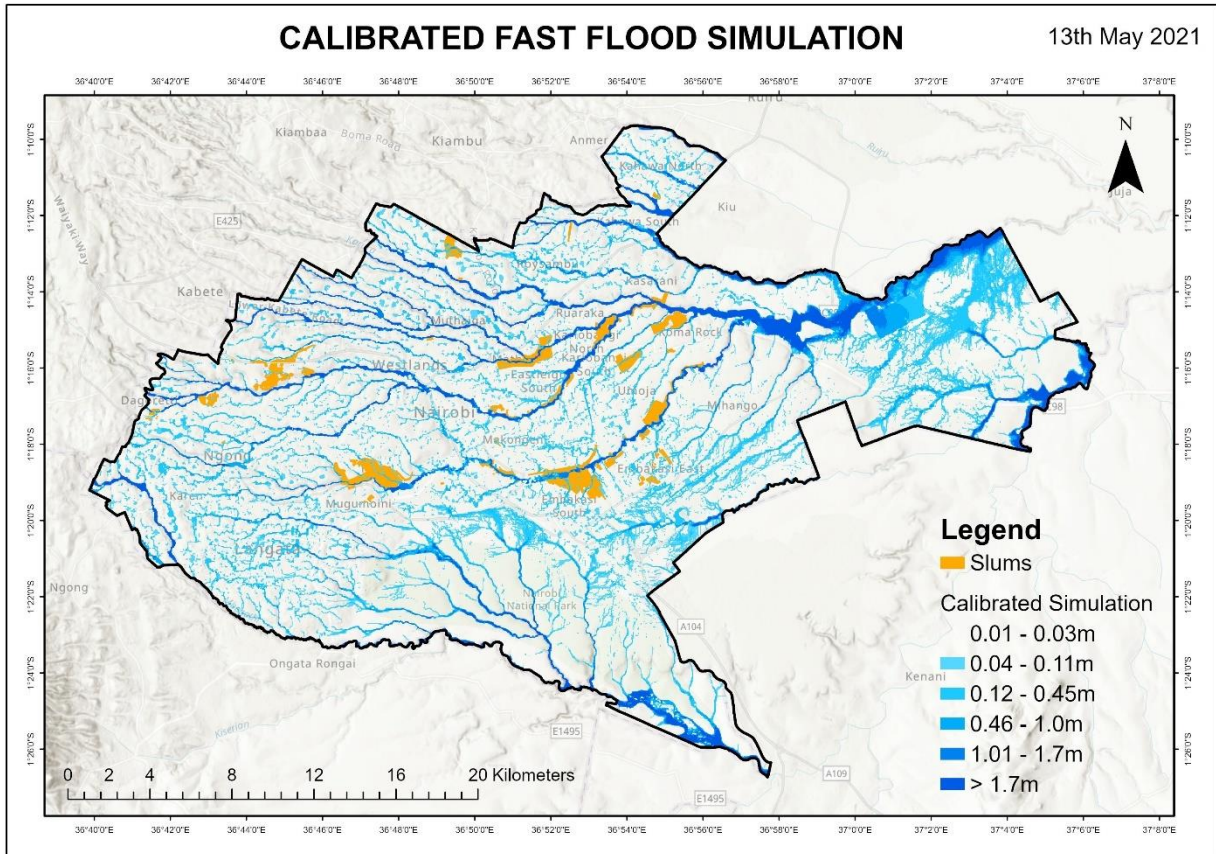
Figure 8: Calibrated Fast Flood Simulation

5.2.2. Fast Flood Validation

5.2.2.1. Validation of the Uncalibrated Simulation

The uncalibrated FFS was qualitatively validated in the field by both experts and slum residents from the selected slums. The stakeholders claimed that the simulation did not capture localized flooding with some inaccuracies in fluvial flooding extent. In some areas of the slum settlements, the residents claimed that the simulated flood extents were over or under-simulated. Additionally, dozens of collected My Maps flood points did not overlay with the simulated flood extent, especially in Kibera. However, they confirmed the fluvial FFS as a true depiction of river flooding.

5.2.2.2. Validation of the Calibrated Simulation

The calibrated simulation was compared with an observed satellite flood extent captured on May 1st, 2024, by getting the TP extent areas denoted in green in Figure 9. The Eastern region is seen to experience much flooding as observed in both extents. Additionally, apart from capturing some of the urban floods in the Eastern region, the satellite imagery was only able to capture minimal fluvial flooding along major rivers.

The TP extent covered an area of 20 km², with the UNOSAT extent within Nairobi covering 25 km² and the calibrated FFS occupying an area of 82 km². Additionally, upon further inspection, the UNOSAT fluvial flood extent was minimal compared to the Fast Flood fluvial flood extent. Additionally, when comparing using the My Maps flood points in Korogocho and Mukuru (as they were the only selected slums observed by the UNOSAT flood imagery), the flood areas missed by the FFS were also missed by the satellite extent.
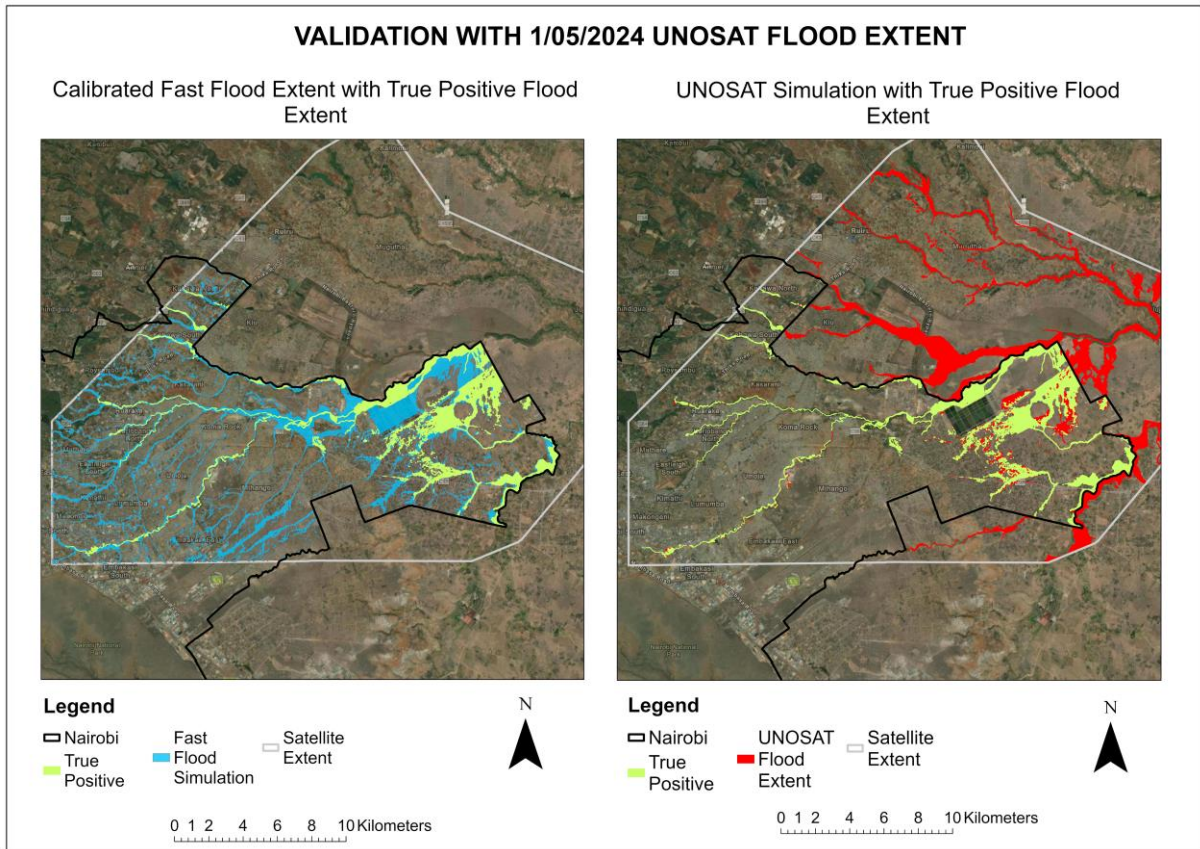
Figure 9: Validation of Calibrated Model with Satellite Observed Flood Extent

The KDI Kibera urban flood map modelled using a HEC-RAS hydrodynamic model captured finer details of urban flooding, such as flood occurrences along roads as illustrated in Figure 10. Additionally, the flood extent of the hydrodynamic model was larger compared to the FFS with higher depths, especially along rivers. The highest depths of the FFS are observed to be along the lower river entering the Nairobi Dam, with the Nairobi Dam also having high depths.

The golf course bordering Kibera is seen to have high flood depths of above 1.7m according to the HEC-RAS model. However, in comparison to the FFS, the maximum depth is observed to be 0.45m, resulting in a difference of 1.25m. Both simulations did not capture the flood extent of some flood points purported to frequently experience flooding as in Figure 10. Additionally, the FFS in the Eastern region of Kibera captured more flood areas and had larger extents compared to the HEC-RAS model.
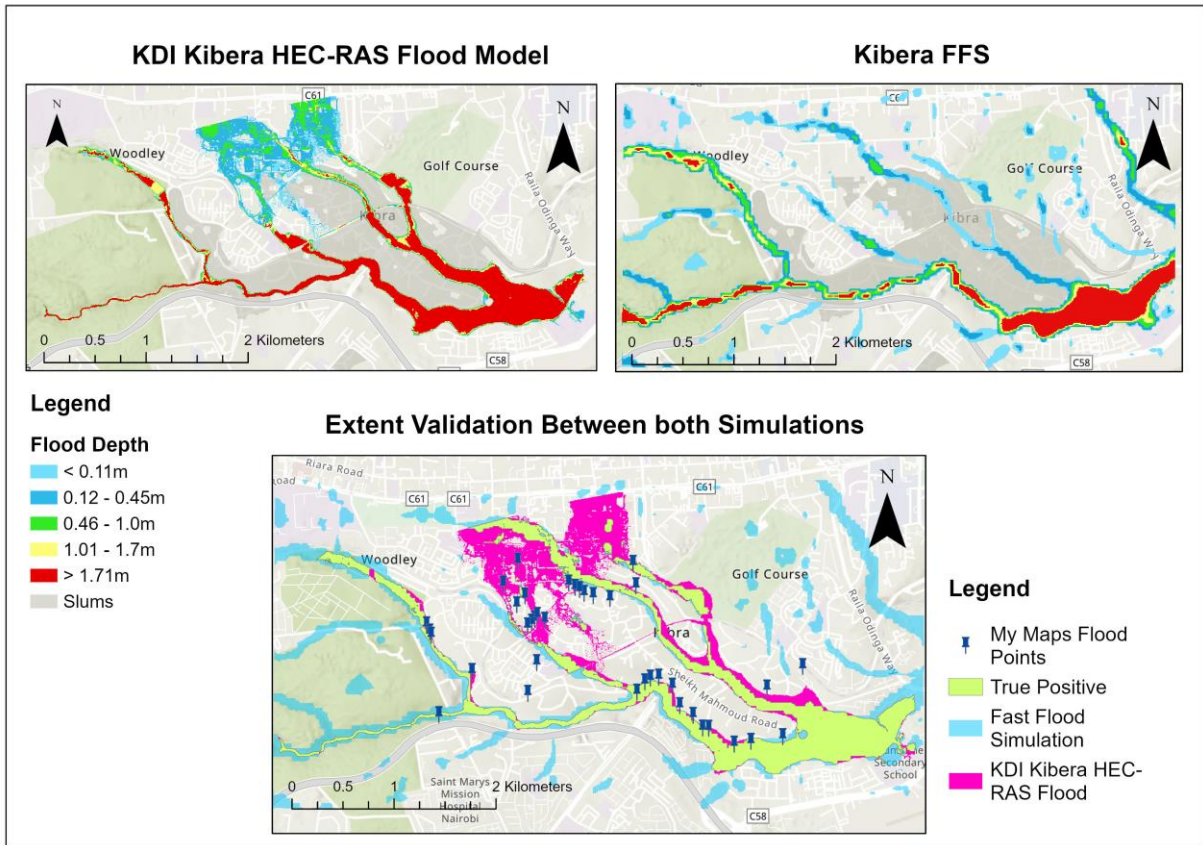
Figure 10: Kibera Slum Validation Against Kibera HEC-RAS Flood Model

5.2.3. Sample Training Grids

Appendix 16 shows that the no-flood grids were spatially distributed across the study area, with an equal number of grids in each elevation cluster. Majority of the flood grids were in the slums where the My Maps exercises were conducted, as seen in clusters 0, 2 and 3. The number of flood grids in each cluster varied, with cluster 2 having the highest number of grids (42 grids) and cluster 4 having the least number of grids (4 grids), as in Appendix 17. The variations in the flood grids resulted in an imbalance in the number of sample training grids in each cluster.

5.3.   Flood Susceptibility

5.3.1.   Data splitting

The splitting technique divided the study area into different spatial non-overlapping regions, resulting in spatially uncorrelated training and testing grid sets. The splitting resulted in 146 training grids and 85 testing grids, as in Table 7, with each training and testing block capturing the different diversity of features in the study area, ensuring that the model did not learn to over-perform in specific spatial regions. As a result of the split, the training grids had more flood grids than no-flood grids while testing grids had more no-flood grids than flood grids as seen in Table 7.

Table 7: Distribution Split of Sample Training Grids

	Flood Grids (class 1)	No-flood grids (class 0)	Total Number
Sample Training Grids	116	115	231
Training Grids	76	70	146
Testing Grids	40	45	85

5.3.2. Model 1 – All independent variables

As Model 1 used all the variables, the grid search cross-validation fitted a total of 1125 fits for 225 candidates, yielding a best cross-validation score of 0.753. The best RF hyperparameter combinations for Model 1 provided in Table 8 with a random state of 42 for reproducibility, resulted in an overall accuracy of 84.71%, a precision score of 0.935 and a recall score of 0.725. The feature importance for Model 1 presented distance to rivers, SPI, BuA, IBD and elevation as the top 5 important features. Flow Accumulation, clay soil content (CLPC) and LULC were implied to have minimal influence on flood susceptibility, as in Figure 16.

Figure 12 depicts that low susceptibility values are mostly seen in the East, South-west and North-west regions of Nairobi. However, in these same areas of low susceptibility, high susceptibility values are witnessed along the river systems which relates to the high importance of distance to rivers. Observing that areas near rivers are very susceptible to flooding, slum settlements are also seen to be highly susceptible to flooding, though, just not in slum areas that are close to rivers.

A comparative analysis between Kangemi and Kawangware slums and Loresho and Masiwa formal settlements, exhibit that low BuA and high CAR that characterize slums and high-density settlements have moderate to high susceptibilities. Whereas formal areas with high BuA, are depicted in Figure 11 are exposed to very low-moderate flood susceptibilities.

Building Adjacency ranges from 0-1, with values close to 0 indicating buildings are close to each other

Susceptibility ranges from 0-1, with values close to 1 indicating very high susceptibility
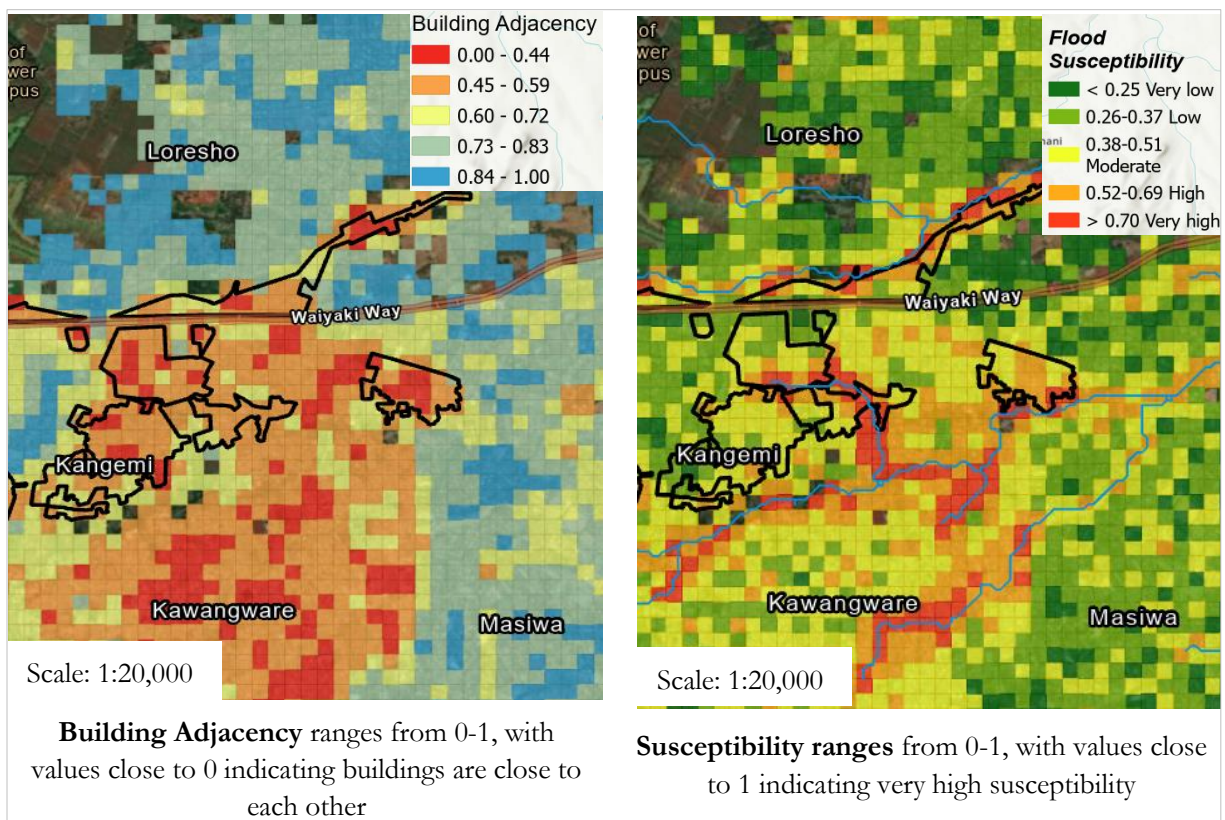
Figure 11: Relationship between Building Adjacency and Flood Susceptibilities between Kangemi and Kawangware Slums and Loresho and Masiwa Formal Settlements
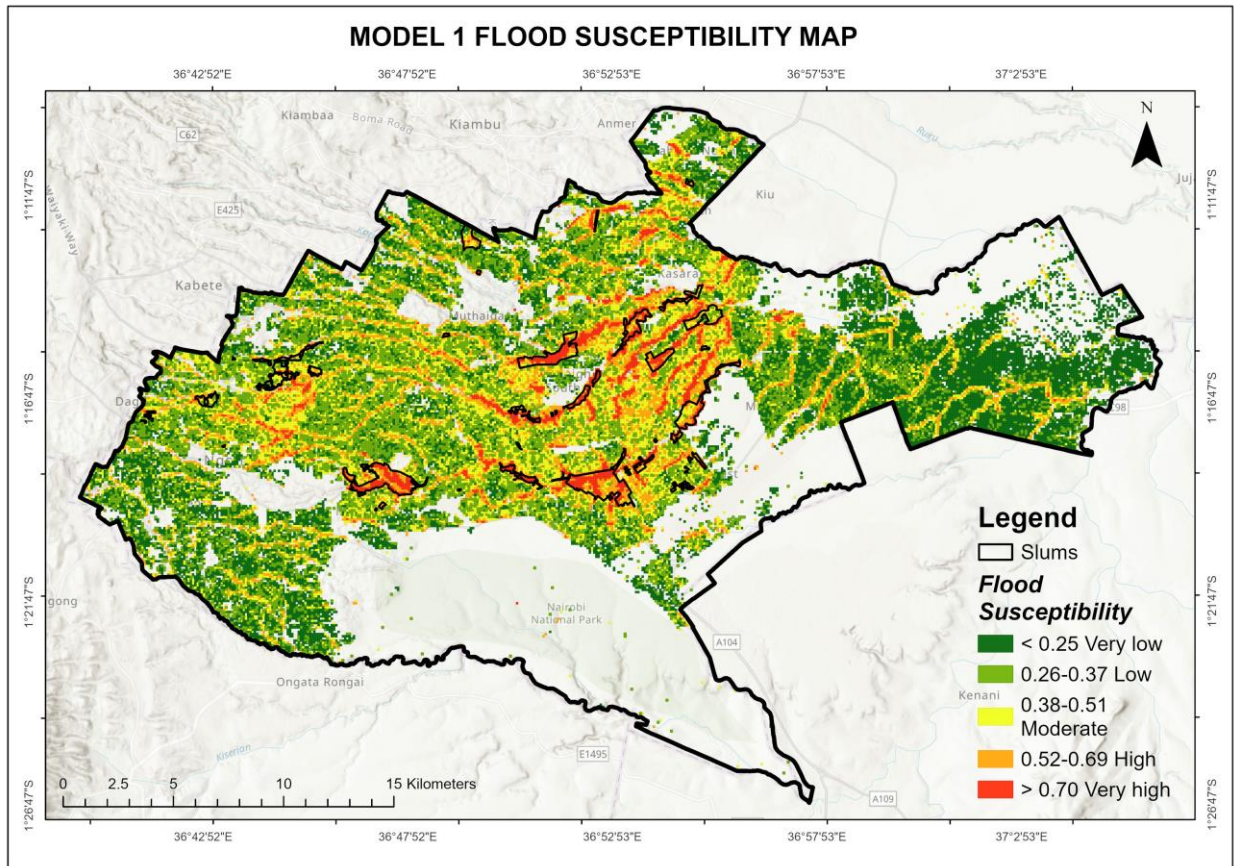
Figure 12: Model 1 Flood Susceptibility Map

5.3.3. Model 2 – Using the FIFs variables

While training the model using the 14 FIFs variables to predict susceptibility, the grid search technique similar to the one employed in Model 1 provided a best validation score of 0.760. The best hyperparameter combinations for Model 2 provided in Table 8, resulted in an overall accuracy of 71.76%, with a precision score of 0.833 and a recall score of 0.5. The top 5 important features were distance to rivers, SPI, elevation, NDWI and curvature depicted in Figure 16, with the least important FIFs being the same as in Model 1.

The predicted susceptibility illustrated in Figure 13, is observed to be strongly influenced by the distance to rivers, with areas in proximity to rivers being more susceptible to flooding. Areas near rivers in the Eastern region (formal settlement areas) had high to very high susceptibilities.

In both Model 1 and 2, flood susceptibility values are seen to increase as elevation decreases (from West to East) up to a certain location, where, as elevation decreases flood susceptibility decreases as in Appendix 11, Figure 12 and Figure 13. Moreover, despite higher TWI being associated with increased flooding, the Eastern region does not conform to this relationship implying that TWI does not heavily influence flooding in the East.
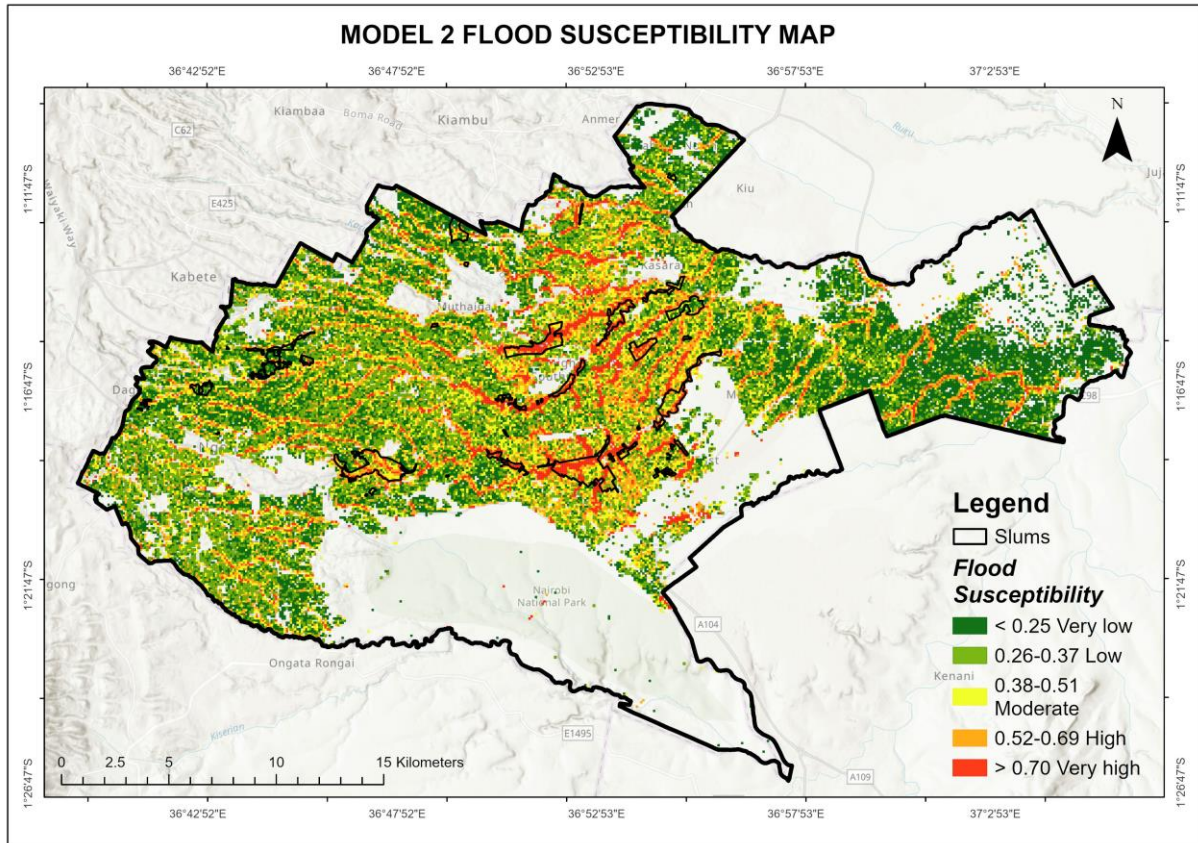
Figure 13: Model 2 Flood Susceptibility Map

Table 8: Random Forest Hyperparameters and Accuracy Metric

Model	Variables used	Hyperparameter combinations	Overall Accuracy	Precision	Recall
1	All	n_estimators: 200 Max_features: 4 Max_depth: 10 Min_samples_split: 4 Min_samples_leaf: 10 Random_state: 42	0.8471	0.935	0.725
2	FIFs	n_estimators: 200 Max_features: 4 Max_depth: 10 Min_samples_split: 4 Min_samples_leaf: 10 Random_state: 42	0.7176	0.833	0.5

5.3.4.   Differences in Model Predictions

Susceptibility values in slums in Model 2 are observed to be lower compared to the susceptibility values in Model 1, with similarities in slum areas near rivers being very highly susceptible to flooding (See Figure 14). Formal areas such as Langata in the Southwest region in Model 1 predominantly have very low–moderate susceptibility with exceptions in areas near rivers, while in Model 2 the susceptibility values range quite higher.
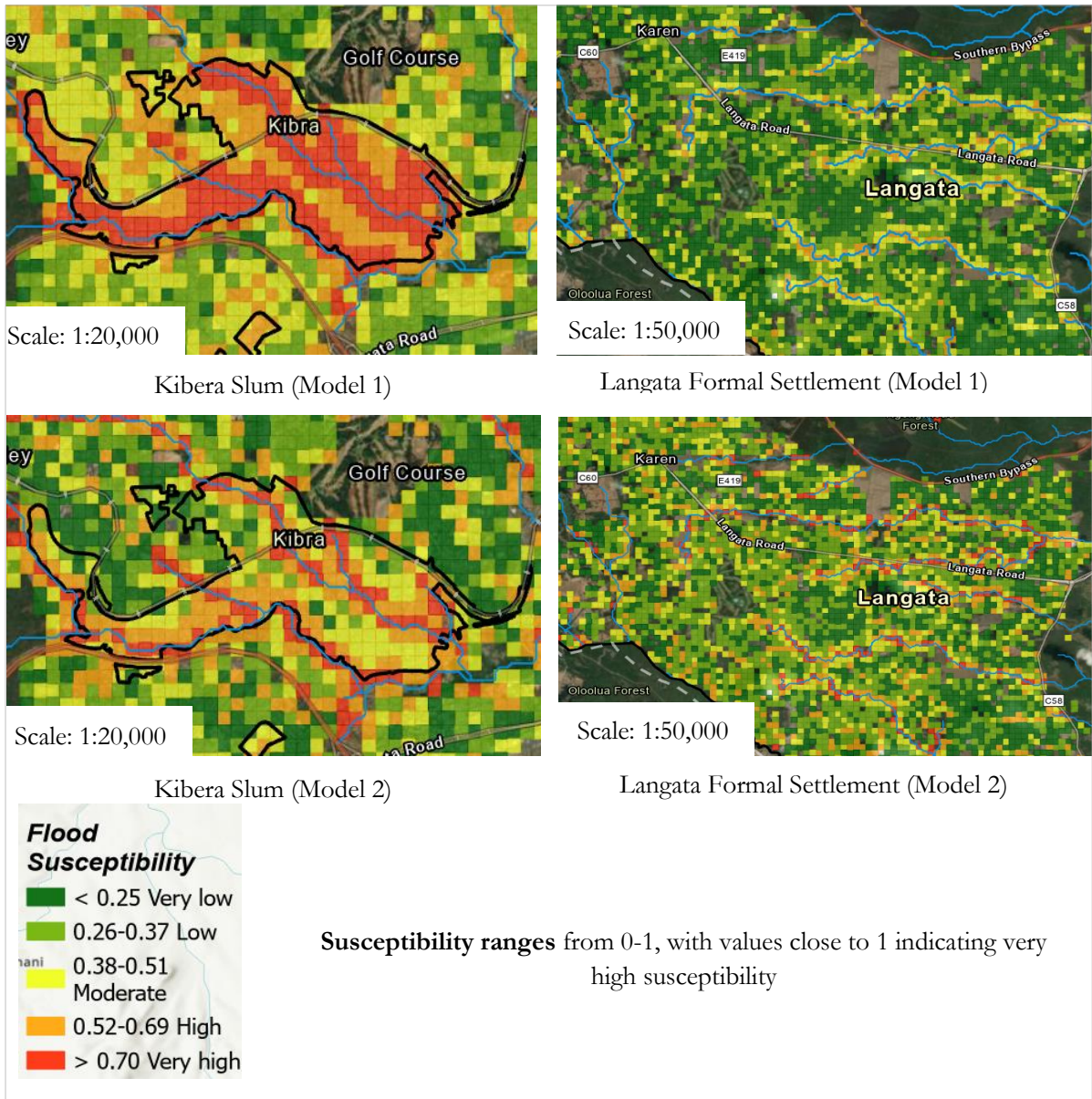
Figure 14: Comparison Between Model 1 and Model 2 Predictions in both Formal and Slum Settlements

Based on the 51% threshold probability of an unlabeled grid belonging to a particular flood class label, Model 1 had 8,425 flood grids with Model 2 having the highest number of predicted flood grids of 10,852 (see Table 9). The disparity in predicted flood grids critically infers that the exclusion and or inclusion of the morphometric variables affects the prediction outcomes. However, as much as Model 2 predicted more flood grids by relying on FIFs, its overall accuracy (71.76%) is significantly lower than that of Model 1.

Table 9: Grids per Flood Classification Prediction

Models	Flood Class Prediction (based on a 51% probability threshold of belonging to a specific class)	
	Number of flood grids (1)	Number of no-flood grids (0)
1	8,425	33,595
2	10,852	31,168

5.4. Flood Susceptibility in Formal and Slum Settlements

With formal areas being grids that do not intersect with the slum extent, the box plots in Figure 15 depict high slum flood susceptibilities with median susceptibility values of roughly 0.65 in Model 1 and 0.55 in Model 2. In formal settlements, the median susceptibility appears to be consistent in both models with mean susceptibility values of approximately 0.3, which is significantly lower than in slum areas.
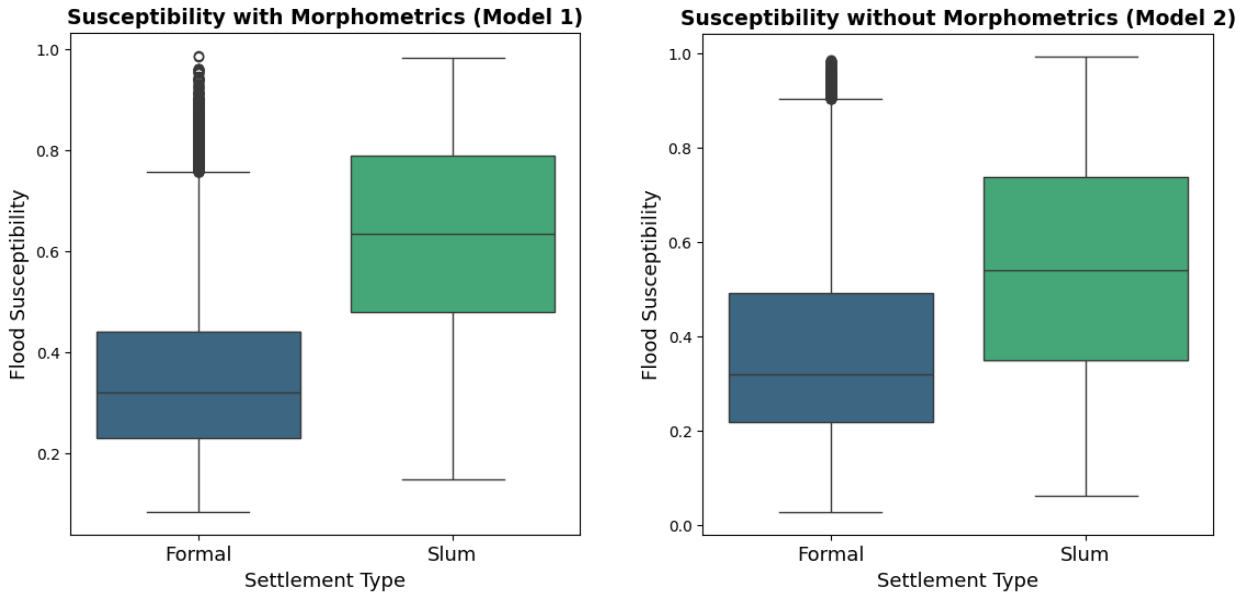

Figure 15: Flood Susceptibility Comparison between Formal and Slum Settlements (Model 1 and 2)

In both models, the Inter-Quartile Range (IQR) is wider for slums, showing more variability indicating broader ranges of susceptibility values, suggesting that some slums or some slum areas have lower or higher susceptibilities than the median. Formal areas in comparison, have narrower IQR, revealing that susceptibility values are close to the median value and closer to each other.

However, as formal areas are seen to be least susceptible, the outliers suggest that there are some formal areas with extremely high susceptibilities, with the majority of outliers seen with the inclusion of morphometric variables. The outliers are seen in high-density compact areas, with some areas located near rivers. Slum susceptibility in Model 1 is portrayed to be higher given the inclusion of morphometric variables, compared to Model 2, informed by:

- The higher minimum susceptibility value depicted by the lower whisker
- The IQR (representing 75% of the slum grids) having susceptibility values of approximately 0.45 to 0.8, compared to values of approximately 0.3 to 0.75 in Model 2.

5.5. Comparison of RF Feature Importance and CS Feature Ranking

Using Model 1 as it is the main model, building in low areas and near rivers representing elevation, elevation derivatives and distance to rivers were ranked as the most important factors that cause flooding at a city scale level (experts voting) and in Kibera with a score of 6 votes as in Figure 17. Coincidentally, distance to rivers, SPI and elevation were ranked the 1st, 2nd and 5th most important variables by the RF algorithm.

Morphometric factors such as the arrangement of buildings, the density of buildings and the area covered by buildings were ranked as the 4th, 5th and 6th influential factors with 3, 2 and 1 votes respectively, by the experts for both slum and formal areas. Arrangement of buildings encompassing spatial distribution of

buildings, building orientation, alignment and fractal dimension had the highest votes in Mukuru and Korogocho (4 votes) and 2 votes in Kibera.
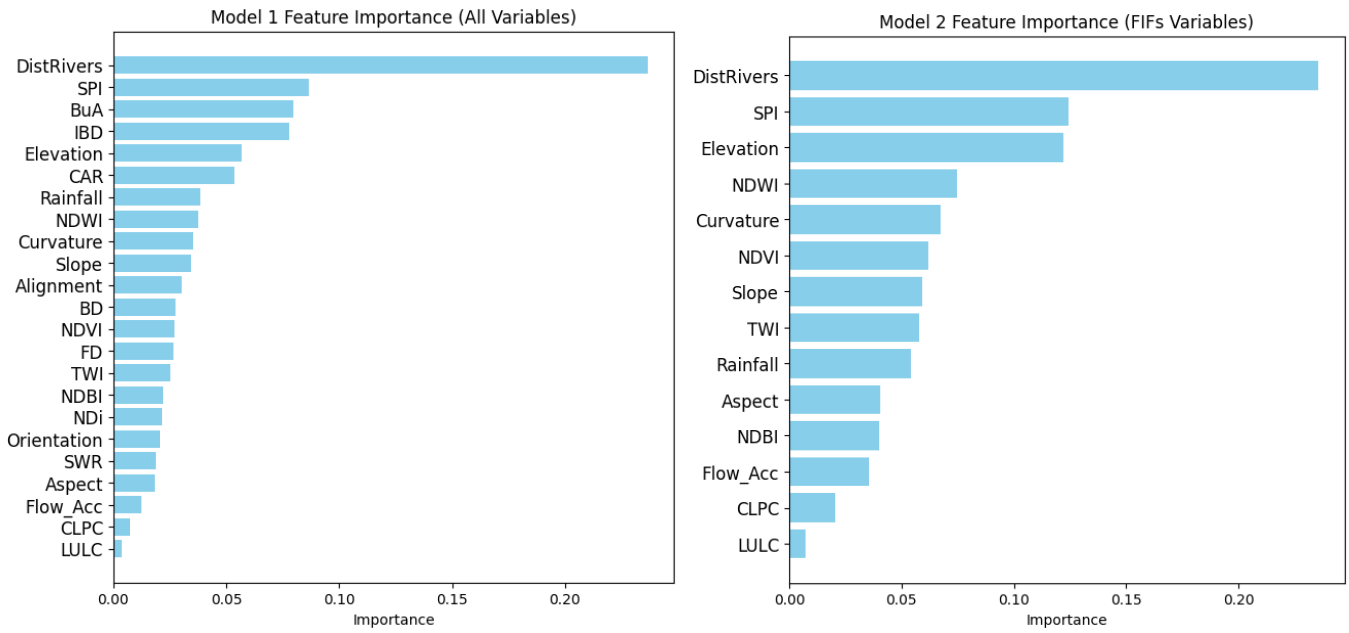

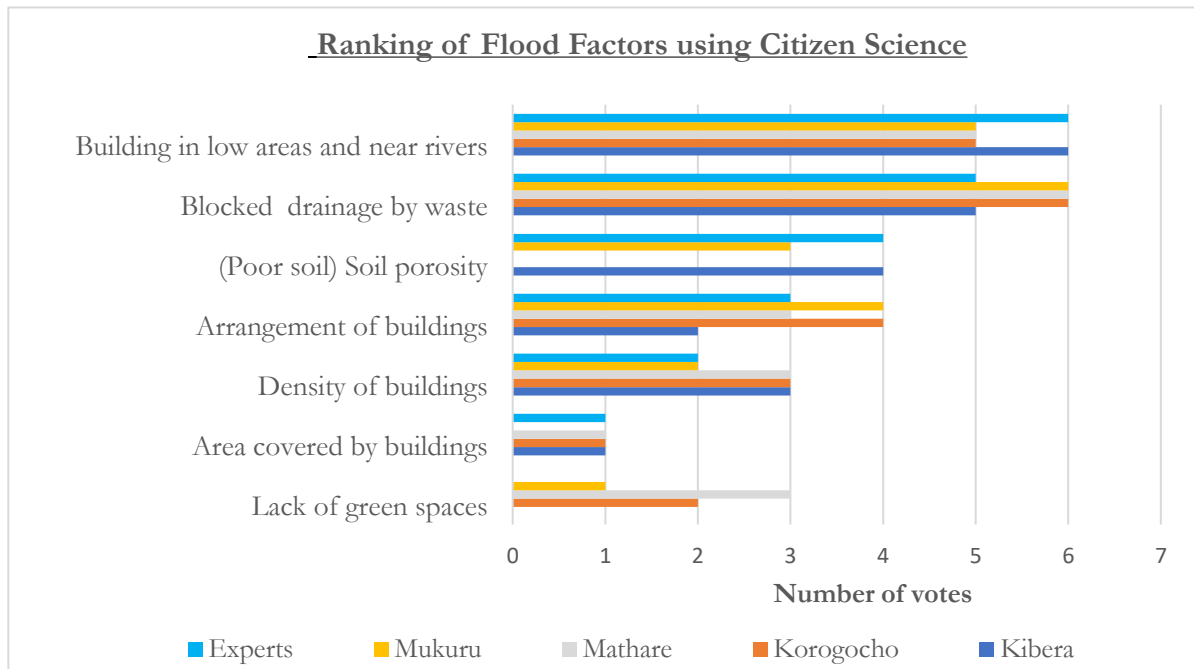
Figure 16: RF Model 1 and 2 Feature Importance Lists

Figure 17: Citizen Science Pairwise Feature Ranking Importances

During fieldwork, the most common flood incidences were reported to be along rivers proving that distance to rivers significantly contributes to heightened flood risks. Buildings were seen to be built either on the same or at elevated levels in floodplains, with some buildings constricting the river paths as seen in Figure 18.

Figure 18: Distance to Rivers in Slums

Building adjacency and mean inter-building distance both relating to the spatial distribution of buildings were ranked 2nd and 3rd by the RF algorithm shown in Figure 16 while the arrangement of buildings ranked 3rd in Mukuru and Korogocho (see Figure 17).

The area covered by buildings was ranked as the 6th important influential flooding factor in Mathare, Korogocho, and Kibera and at a city scale level similar to the ranking of CAR by the RF algorithm in Figure 16. Interestingly, Mathare was the only settlement that highly ranked the lack of green spaces, assigning it with 4 votes, tying at 3rd place with arrangement and density of buildings, while lack of green spaces was ranked as the least important factor by experts.

To quantitatively analyse the factor blocked drainage by waste, slum waste data was acquired from SLUMAP research project[17] in 2019 - a project that employed RS for slum mapping and characterization in SSA cities. The analysis exposed Mukuru as having high amounts of waste across the settlement, especially along the main river system. Furthermore, Mathare and Soweto slums had areas with very high percentages of waste ranging from 12.6% - 32.64%, compared to other slums, also found along rivers as in Figure 19.

Kibera is seen to have lower percentages of waste, with the highest percentage of waste levels being in the 2.16% - 5.39% category. Comparing the amount of waste in the selected slums, blocked drainage by waste ranked 2nd in Kibera and 1st in both Mathare and Mukuru, reflecting the proportion of waste in the slums as seen in Figure 19. Moreover, areas with high levels of waste were also seen to have high values of flood susceptibility, especially in Mukuru.

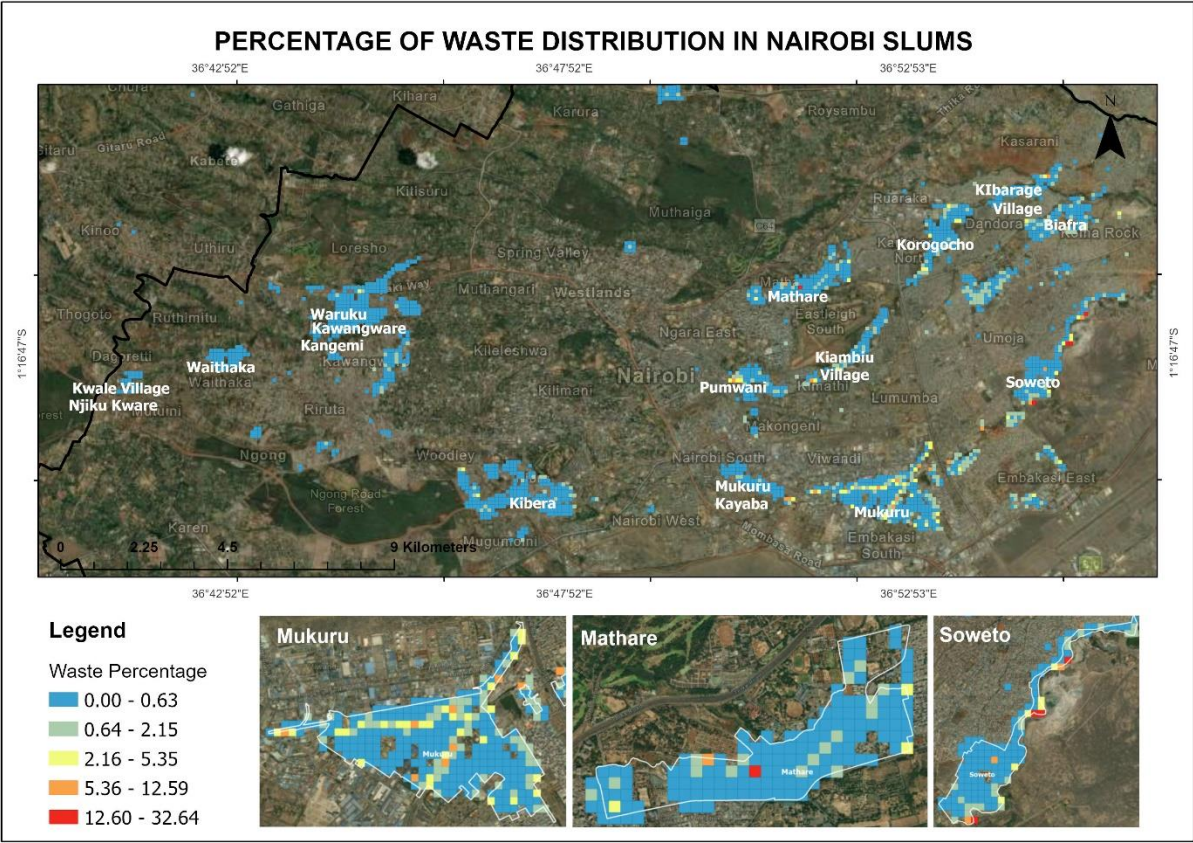[17] SLUMAP research project - Mapping slums with remote sensing (ulb.be)

Figure 19: Percentage of Waste Distribution in Nairobi Slums

6. DISCUSSION

6.1. Role of Urban Morphology on Flooding

The differences in the overall accuracies between Model 1 (85%) and Model 2 (71.76%) suggest that despite FIFs being crucial in determining flood susceptibility, flood prediction solely using FIFs has limited ability to capture the complexities of urban flooding. The lower predictive performance of Model 2 provides evidence that FIFs do not account for micro-topographic features that govern flood flow dynamics (Safaei-Moghadam et al., 2023). These micro-topographic variations often result from anthropogenic modifications to landscapes and landforms (Chirico et al., 2021), which aggravates hydrological dynamics and processes (Douglas, 2011).

The novelty of including building morphometrics in FSM in Model 1 and the improved model accuracy in flood susceptibility prediction, suggests that its incorporation accounted for micro-topographic features relating to building developments that were overlooked by the FIFs. Successively, urban morphology is seen to be vital in capturing the complex intricacies of flooding.

Postulated that dense urban areas tend to have the most pronounced landscape effects of anthropogenic activities (Chirico et al., 2021), the influence of morphometrics was seen in slums where their susceptibilities greatly increased in Model 1, indicating a correlation between density and flooding. Inversely, formal areas characterized by having lower densities and larger distances between buildings exhibit reduced flood susceptibilities in Model 1. Huang (2021) notes that slums are often left out of official and formal plans resulting in unregulated development – a situation also observed in Nairobi. Formal settlements, unlike slums, adhere to building and zoning ordinances such as minimal plot coverage, plot ratio, maximum densities, and plot setbacks (Kuffer et al., 2016) resulting in lower values of SWR, CAR, BD which reduce flood susceptibilities.

This occurrence is illustrated by the narrower IQR of the formal settlements and its highest susceptibility value of 0.75 (excluding outliers) in Model 1 compared to Model 2 which had a larger IQR and highest susceptibility value of 0.9 (excluding outliers), seen in Figure 15. The urban morphological features seen to affect flood susceptibility were building adjacency, mean inter-building distance and covered area ratio, which ranked 3rd, 4th and 6th in the top 10 important features.

6.2. Use of Global Datasets

The use of global datasets at a city scale resulted in low spatial resolution and accuracy, resulting in datasets that inadequately capture fine-scale details required for efficient flood mapping (Carr et al., 2024). However, leveraging global datasets provided this research with comprehensive data for FSM, with their use resulting in RF model accuracies of above 70% in susceptibility prediction.

Despite their limitations, the performance of global datasets in both Fast Flood and RF models proved their effectiveness in capturing fundamental susceptibility factors, especially as Model 1 yielded an overall accuracy of 85%. Given that they are readily and openly available they are useful in data-scarce contexts, particularly in SSA cities, as they can facilitate easy scalability and replicability of this approach allowing uniform and consistent FSM comparisons.

6.3. Inclusion of Citizen Science in FSM

The contribution of CS has been significantly useful in this research with residents and experts providing insightful and complementary information, considering that they are knowledgeable about their

environment, flood-related issues and witnesses of local flood events as asserted by Sy et al. (2020). CS information helped in understanding the settlement-specific underlying flood factors as they provided hierarchical importance of both morphometric and FIF factors, making CS a valuable source of primary information.

Additionally, given that the global datasets were not able to capture fine details, by virtue of the residents understanding their current issues, they provided flood information related to the interconnectedness of various issues to flooding such as poverty, poor governance, failure of urban planning, ignorance and neglect by authorities. CS was particularly useful in the acquisition of accurate flood data as slum residents collected flood inventory points by identifying flood hotspots also highlighted by Tran et al. (2024) which for this study reflected the influence of anthropogenic, hydrological and environmental factors on urban flooding. This participatory approach remarkably tackled the limitations of data paucity as the residents (slum participants and city experts) were able to provide historical flood data needed for susceptibility prediction.

6.4. Limitations

Despite the results of this research, the approach and study had various limitations such as the inherent limitations of global datasets such as the downsizing of data to approximate local conditions, which reduced details that would have accounted for the intrinsic nature of flooding. The use of local VHR DEMs produced by Light Detection and Ranging (LIDAR) and Unarmed Aerial Vehicles (UAV) would improve accurate elevation information resulting in accurate flood mapping (McClean et al., 2020).

6.4.1. Limitations and Uncertainties of Fast Flood

The simulation result provided by Fast Flood illustrates that the tool predominantly accounts for fluvial flooding and hydrological-based flooding. Based on fieldwork, city experts asserted that Nairobi greatly suffers from urban flooding referring to it as artificial flooding. These floods are mainly caused by excessive water run-off in built-up areas where water has no place to go, exerting an immense amount of pressure on drainage systems, intensified by impervious surfaces (Sandink et al., 2016).

From the simulation result, Fast Flood does not seem to fully capture urban flooding, however, some areas are seen to flood despite being positioned far from the rivers, as in Figure 8. Areas notorious for flooding were simulated to have minimal or no flooding, which was also highlighted in fieldwork by both slum residents for their respective slums and by city experts at a city-scale level.

Despite Fast Flood offering the option to include urban data from OSM, such as buildings and roads, simulations done with and without the urban elements resulted in similar simulation outcomes. This is a result of the primary role of OSM data in Fast Flood for exposure analysis. Buildings as obstacles were not included because the resolution of the DEM (Copernicus 40m) was coarse, especially for small buildings, as they would have been amalgamated to form an elevation obstacle.

6.4.2. Limitations of Satellite Flood Imagery

Given that in Nairobi, the April 2024 - May 2024 flood events reported many cases of urban flood events, they were not fully captured by the UNOSAT satellite imagery in its specific region of interest (see Figure 9). The minimal reflection of fluvial flood extent and urban floods critically reveals the difficulty of capturing the intricacies of flooding while using satellite imagery. Moreover, this highlights that most flood-detection and mapping focus more on fluvial flooding compared to urban flooding, which is easier to detect as urban floods are usually shallow and short-lived with their ponding nature creating discontinuous flood extents increasing detection difficulty (Tanim et al., 2022).

6.4.3.   Limitations of CS in Flood Data Collection and Flood Inventory

Despite the results and usefulness of CS data, challenges such as coverage, reliability and accuracy exist. Low participation is often seen to result in low volumes of data affecting the coverage and reliability of the collected data (Le Coz et al., 2016). In this research due to the easy manipulation of My Maps data points, 2 participants were selected per slum to collect flood points, in specific slums like Mathare the points did not cover the full extent of the settlement, resulting in poor coverage data.

As the flood points provided were based on areas that flood with no selection criteria like the no-flood point/grids, they resulted in cluster class imbalance (see Section 5.2.3). The class imbalance in ML models can result in models that are not generalizable, affecting the robustness. Given that clusters 1 and 4 had 9 and 4 flood grids respectively, with 23 no-flood grids each, the model could have possibly learned how to predict non-flood grids in clusters 1 and 4 compared to flood grids. This could have led to the misprediction of (no)flood grids due to the lack of an equal number of flood grids to train the model with.

6.4.4.   Limitations in Data and Data Processing

Aggregating data to the 100m*100m downscaled the resolutions of the datasets such as LULC, and NDVI from 10m resolution to 100m resolution might have reduced the accuracy of the model as fine details are omitted, producing flood susceptibilities that might over or under-estimate the likelihood of flooding. Additionally, given that data is lost in aggregation, variations in the grid cells (100m*100m) were not captured, ergo creating an ecological fallacy by assuming that for example, all the areas in a grid cell have high susceptibilities, however, it could just be one specific area and not the entire grid.

Also, as the reference slum layer did not capture all the slums present in Nairobi, some slums were considered as formal settlements, which affected the susceptibility values of formal settlements represented by the outliers in Model 1, which when inspected, were in some areas resembling slums.

7.   CONCLUSION AND RECOMMENDATIONS

7.1.   Conclusion

The findings of this research revealed that slum settlements in Nairobi are highly susceptible to flooding, with a susceptibility median value of 0.65. The majority of the formal areas have very low to low susceptibility, especially in the Eastern and Southwest parts of Nairobi, with a median of 0.3, except along rivers that have higher susceptibility values in Model 1. Despite formal areas near rivers having higher susceptibilities, susceptibility values in slum areas near rivers are extremely higher, especially in Kibera, Mukuru and Mathare.

The results of the quantification of flood risk using susceptibility conform with the assertions that slums are hit hardest during disaster risks as sharp susceptibility differences are observed between slums and formal areas. The findings of this research expose distance to rivers, SPI, building adjacency, mean interbuilding distance and elevation as the top 5 influential factors.

As much as hydrological and environmental factors influence the probability of flooding, the top urban morphology factors such as building adjacency, mean interbuilding distance and covered area ratio significantly influence flooding specifically in slums. The improved accuracy with the inclusion of morphometrics in the RF predictive Model 1, particularly highlights that understanding and examining flooding solely based on FIFs is not effective and does not lead to a comprehensive understanding of flooding in the city. Thus suggesting the importance of incorporating urban morphology in understanding the urban-induced complexities of flooding.

From the findings, it is apparent that the critical morphological element influencing flood susceptibility is the availability of space for water. The spatial configuration of settlements influenced by morphological elements such as building adjacency and mean interbuilding distance, significantly alter how water moves and accumulates during flood events. This natural hydrological response is dictated by the amount of space left for water to flow.

The covered area ratio, identified in the top 10 important features in Model 1, has a vital role in influencing surface run-off accumulation by dictating the amount of impervious space, controlling the available space left for water infiltration. Increased soil sealing, resulting from highly covered areas reduces soil permeability leading to increased inundation, specifically in slums given their dense nature.

The participatory techniques used to validate and confirm flooding and rank the importance of flood factors have been seen to be essential in flood mapping, especially at the local scale, where much flood information is unknown while using large-scale data. Citizen-based data in flood modelling ought to be explored as it has great potential to improve flooding studies, especially in the global south where flood data is often scarce

7.2.   Recommendations

With the increase in climate variability intensifying flood events, and the 'static' nature of hydrological and environmental conditions, urban planners and disaster risk management authorities ought to pay attention to 'changeable' flood factors such as the nature and typologies of urban devlopment. Which have been realised to play a great role in controlling the dynamics of flood waters.

The duality of urban development and flooding has been biased with a keen focus on the effect of flooding on urban elements, discounting the effects that urban elements have on flooding. Scrutinizing the potential

impacts urban development, particularly, urban morphologies has on flooding needs to be conducted, most especially for slums since they are the most vulnerable and highly susceptible settlements. This necessitates the inclusion and or recognition of slums as part of the formal system, to purposefully plan their organization and arrangement as their morphologies and locations in flood-prone areas increase their propensity to flooding.

Based on the strong relationship between flooding and urban morphology, morphological responses to reduce flood susceptibilities, planning authorities ought to manage urban development to ensure harmony between the built and natural environment. Approaches such as controlling development, especially densification through zoning and building ordinances can be done to control urban flooding that results from development pressure.

From a bottom-up perspective, community empowerment and participation through workshops, meetings and other public engagement forums can be employed to create awareness, encourage community-led mitigation measures and ensure public involvement in the planning process, especially in the formulation of disaster and settlement plans.

Hydrological and environmental responses to reducing susceptibilities could be through the provision of:

1. Non-structural measures, especially nature-based solutions that aim to reduce surface run-off, create room for water, and increase infiltration.
2. Structural measures include canalising major rivers, constructing flood walls, and improving drainage infrastructure.

7.3. Research Recommendations and Areas of Further Research

This research could be built upon, by collecting both flood and non-flood locations using a CS approach, to capture 'on-ground' flood depths as it would account for multiple sources of flooding, especially artificial flooding. Local and high-resolution datasets ought to be employed in future research for accurate flood mapping.

As this research focuses on one specific geographical region, a similar methodology can be conducted across different geographical settings to compare how varying slum morphologies influence flooding. Such a study could be done with a key focus on identifying common or distinctive morphological elements that influence flood susceptibility given the diversity and heterogeneity of slums.

As this research focused on building morphology, explicitly 2D building morphologies, research on flood susceptibility can be steered towards using a broader set of building morphologies such as 3D building morphologies, street and plot morphologies to provide deeper insights into flood dynamics.

8. ETHICAL CONSIDERATIONS AND DATA MANAGEMENT PLAN

8.1. Ethical Considerations

The research fieldwork was conducted ethically as it took informed consent of participants and respected their privacy and confidentiality (see Appendix 18). At no point did this research use any individual information regarding the participant's identity or detailed specific locations in the selected slums that would expose them or put them at risk. Data findings from the field or research findings were not fabricated and or altered, thus upholding the integrity of this research. The research properly acknowledged other works used by referencing the respective authors and citing the data sources for data used.

Ethical concerns for the research including fieldwork through requesting an ethical review (request number 240110), were examined and approved by an ethical committee, receiving positive advice from the reviewer.

8.2. Data Management Plan (DMP)

To allow for the reproducibility and replicability of this research, datasets and codes used in this study were stored and made available on Google Drive and GitHub Platforms[18] respectively, with Appendix 19 providing the DMP summary.

---

[18] https://github.com/MarieMuthoni/Jane-marie_Msc.git

LIST OF REFERENCES

Abascal, A., Rodríguez-Carreño, I., Vanhuysse, S., Georganos, S., Sliuzas, R., Wolff, E., & Kuffer, M. (2022). Identifying degrees of deprivation from space using deep learning and morphological spatial analysis of deprived urban areas. *Computers, Environment and Urban Systems*, *95*. https://doi.org/10.1016/j.compenvurbsys.2022.101820

Abass, K. (2022). Rising incidence of urban floods: understanding the causes for flood risk reduction in Kumasi, Ghana. *GeoJournal*, *87*(2), 1367–1384. https://doi.org/10.1007/s10708-020-10319-9

Abunyewah, M., Okyere, S. A., Diko, S. K., Kita, M., Erdiaw-Kwasie, M. O., & Gajendran, T. (2022). Flooding in Informal Communities: Residents' Response Strategies to Flooding and Their Sustainability Implications in Old Fadama, Accra. In *Disaster Risk Reduction for Resilience* (pp. 435–461). Springer International Publishing. https://doi.org/10.1007/978-3-030-72196-1_18

Ahmad, I., Dar, M. A., Teka, A. H., Gebre, T., Gadissa, E., & Tolosa, A. T. (2019). Application of hydrological indices for erosion hazard mapping using Spatial Analyst tool. *Environmental Monitoring and Assessment*, *191*(8). https://doi.org/10.1007/s10661-019-7614-x

Al-Aizari, A. R., Alzahrani, H., AlThuwaynee, O. F., Al-Masnay, Y. A., Ullah, K., Park, H. J., Al-Areeq, N. M., Rahman, M., Hazaea, B. Y., & Liu, X. (2024). Uncertainty Reduction in Flood Susceptibility Mapping Using Random Forest and eXtreme Gradient Boosting Algorithms in Two Tropical Desert Cities, Shibam and Marib, Yemen. *Remote Sensing*, *16*(2). https://doi.org/10.3390/rs16020336

Amoako, C. (2012). *Emerging issues in urban flooding in African cities-The Case of Accra, Ghana*. www.afsaap.org.au

Anees, M. T., Abu Bakar, A. F. Bin, San, L. H., Abdullah, K., Nordin, M. N. M., Rahman, N. N. N. A., Ishak, M. I. S., & Kadir, M. O. A. (2019). Flood vulnerability, risk, and susceptibility assessment: Flood risk management. In *Decision Support Methods for Assessing Flood Risk and Vulnerability* (pp. 1–27). IGI Global. https://doi.org/10.4018/978-1-5225-9771-1.ch001

Anwana, E. O., & Owojori, O. M. (2023). Analysis of Flooding Vulnerability in Informal Settlements Literature: Mapping and Research Agenda. In *Social Sciences* (Vol. 12, Issue 1). MDPI. https://doi.org/10.3390/socsci12010040

Assumpção, T. H., Popescu, I., Jonoski, A., & Solomatine, D. P. (2018). Citizen observations contributing to flood modelling: Opportunities and challenges. *Hydrology and Earth System Sciences*, *22*(2), 1473–1489. https://doi.org/10.5194/HESS-22-1473-2018

Barau, A. S., Maconachie, R., Ludin, A. N. M., & Abdulhamid, A. (2015). Urban morphology dynamics and environmental change in Kano, Nigeria. *Land Use Policy*, *42*, 307–317. https://doi.org/10.1016/j.landusepol.2014.08.007

Barker, D. M., Lawler, D. M., Knight, D. W., Morris, D. G., Davies, H. N., & Stewart, E. J. (2009). Longitudinal distributions of river flood power: The combined automated flood, elevation and stream power (CAFES) methodology. *Earth Surface Processes and Landforms*, *34*(2), 280–290. https://doi.org/10.1002/esp.1723

Baud, I., Kuffer, M., Pfeffer, K., Sliuzas, R., & Karuppannan, S. (2010). Understanding heterogeneity in metropolitan India: The added value of remote sensing data for analyzing sub-standard residential areas. *International Journal of Applied Earth Observation and Geoinformation*, *12*(5), 359–374. https://doi.org/10.1016/J.JAG.2010.04.008

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2022). Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrology and Earth System Sciences*, *26*(16), 4345–4378. https://doi.org/10.5194/HESS-26-4345-2022

Bird, J., Montebruno, P., & Regan, T. (2017). Life in a slum: Understanding living conditions in Nairobi's slums across time and space. *Oxford Review of Economic Policy*, *33*(3), 496–520. https://doi.org/10.1093/oxrep/grx036

Bradford, R. A., O'Sullivan, J. J., Van Der Craats, I. M., Krywkow, J., Rotko, P., Aaltonen, J., Bonaiuto, M., De Dominicis, S., Waylen, K., & Schelfaut, K. (2012). Risk perception - Issues for flood management in Europe. *Natural Hazards and Earth System Science*, *12*(7), 2299–2309. https://doi.org/10.5194/nhess-12-2299-2012

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T. C., Bastiaensen, J., De Bièvre, B., Bhusal, J., Clark, J., Dewulf, A., Foggin, M., Hannah, D. M., Hergarten, C., Isaeva, A., Karpouzoglou, T., Pandeya, B., Paudel, D., Sharma, K., Steenhuis, T., Zhumanova, M. (2014). Citizen science in hydrology and water resources: Opportunities for knowledge generation, ecosystem service management, and sustainable development. *Frontiers in Earth Science*, *2*, 104024. https://doi.org/10.3389/FEART.2014.00026/BIBTEX

Cao, C., Xu, P., Wang, Y., Chen, J., Zheng, L., & Niu, C. (2016). Flash flood hazard susceptibility mapping using frequency ratio and statistical index methods in coalmine subsidence areas. *Sustainability (Switzerland)*, *8*(9). https://doi.org/10.3390/su8090948

Carr, A. B., Trigg, M. A., Haile, A. T., Bernhofen, M. V., Alemu, A. N., Bekele, T. W., & Walsh, C. L. (2024). Using global datasets to estimate flood exposure at the city scale: an evaluation in Addis Ababa. *Frontiers in Environmental Science*, *12*. https://doi.org/10.3389/fenvs.2024.1330295

Ceola, S., Domeneghetti, A., & Schumann, G. J. P. (2022). Unraveling Long-Term Flood Risk Dynamics Across the Murray-Darling Basin Using a Large-Scale Hydraulic Model and Satellite Data. *Frontiers in Water*, *3*, 797259. https://doi.org/10.3389/FRWA.2021.797259/BIBTEX

Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., Wang, X., Bian, H., Zhang, S., Pradhan, B., & Ahmad, B. Bin. (2020). Modelling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. *Science of The Total Environment*, *701*, 134979. https://doi.org/10.1016/J.SCITOTENV.2019.134979

Chirico, P. G., Bergstresser, S. E., DeWitt, J. D., & Alessi, M. A. (2021). Geomorphological mapping and anthropogenic landform change in an urbanizing watershed using structure-from-motion photogrammetry and geospatial modelling techniques. *Journal of Maps*, *17*(4), 241–252. https://doi.org/10.1080/17445647.2020.1746419

Choubin, B., Hosseini, F. S., Rahmati, O., & Youshanloei, M. M. (2023). A step toward considering the return period in flood spatial modelling. *Natural Hazards*, *115*(1), 431–460. https://doi.org/10.1007/s11069-022-05561-y

Chowdhury, M. S. (2023). Modelling hydrological factors from DEM using GIS. *MethodsX*, *10*. https://doi.org/10.1016/j.mex.2023.102062

Das, S. (2019). Geospatial mapping of flood susceptibility and hydro-geomorphic response to the floods in Ulhas basin, India. *Remote Sensing Applications: Society and Environment*, *14*, 60–74. https://doi.org/10.1016/j.rsase.2019.02.006

de Bruijn, J. A., de Moel, H., Jongman, B., de Ruiter, M. C., Wagemaker, J., & Aerts, J. C. J. H. (2019). A global database of historic and real-time flood events based on social media. *Scientific Data 2019 6:1*, *6*(1), 1–12. https://doi.org/10.1038/s41597-019-0326-9

Dhanesh, Y., Bindhu, V. M., Senent-Aparicio, J., Brighenti, T. M., Ayana, E., Smitha, P. S., Fei, C., & Srinivasan, R. (2020). A Comparative Evaluation of the Performance of CHIRPS and CFSR Data for Different Climate Zones Using the SWAT Model. *Remote Sensing 2020, Vol. 12, Page 3088*, *12*(18), 3088. https://doi.org/10.3390/RS12183088

Ding, L., Ma, L., Li, L., Liu, C., Li, N., Yang, Z., Yao, Y., & Lu, H. (2021). A Survey of Remote Sensing and Geographic Information System Applications for Flash Floods. *Remote Sensing 2021, Vol. 13, Page 1818*, *13*(9), 1818. https://doi.org/10.3390/RS13091818

Dottori, F., Szewczyk, W., Ciscar, J. C., Zhao, F., Alfieri, L., Hirabayashi, Y., Bianchi, A., Mongelli, I., Frieler, K., Betts, R. A., & Feyen, L. (2018). Increased human and economic losses from river flooding with anthropogenic warming. In *Nature Climate Change* (Vol. 8, Issue 9, pp. 781–786). Nature Publishing Group. https://doi.org/10.1038/s41558-018-0257-z

Douglas, I. (2011). Urban Hydrology. In *The Routledge Handbook of Urban Ecology* (pp. 148–158).

Douglas, I. (2017). Flooding in African cities, scales of causes, teleconnections, risks, vulnerability and impacts. *International Journal of Disaster Risk Reduction*, *26*, 34–42. https://doi.org/10.1016/j.ijdrr.2017.09.024

Douglas, I., Alam, K., Maghenda, M., Mcdonnell, Y., Mclean, L., & Campbell, J. (2008). Unjust waters: Climate change, flooding and the urban poor in Africa. *Environment and Urbanization*, *20*(1), 187–205. https://doi.org/10.1177/0956247808089156

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., & Li, X. (2016). Water bodies' mapping from Sentinel-2 imagery with Modified Normalized Difference Water Index at 10-m spatial resolution produced by sharpening the swir band. *Remote Sensing*, *8*(4). https://doi.org/10.3390/rs8040354

Dumedah, G., Andam-Akorful, S. A., Ampofo, S. T., & Abugri, I. (2021). Characterizing urban morphology types for surface runoff estimation in the Oforikrom Municipality of Ghana. *Journal of Hydrology: Regional Studies*, *34*. https://doi.org/10.1016/j.ejrh.2021.100796

Engstrom, R., Sandborn, A., Yu, Q., Burgdorfer, J., Stow, D., Weeks, J., & Graesser, J. (2015, June 9). Mapping slums using spatial features in Accra, Ghana. *2015 Joint Urban Remote Sensing Event, JURSE 2015*. https://doi.org/10.1109/JURSE.2015.7120494

Ezeh, A., Oyebode, O., Satterthwaite, D., Chen, Y. F., Ndugwa, R., Sartori, J., Mberu, B., Melendez-Torres, G. J., Haregu, T., Watson, S. I., Caiaffa, W., Capon, A., & Lilford, R. J. (2017). The history, geography, and sociology of slums and the health problems of people who live in slums. In *The Lancet* (Vol. 389, Issue 10068, pp. 547–558). Lancet Publishing Group. https://doi.org/10.1016/S0140-6736(16)31650-6

Ezzati, M., Webster, C. J., Doyle, Y. G., Rashid, S., Owusu, G., & Leung, G. M. (2018). Cities for global health. *The BMJ*, *363*. https://doi.org/10.1136/bmj.k3794

Falah, F., Rahmati, O., Rostami, M., Ahmadisharaf, E., Daliakopoulos, I. N., & Pourghasemi, H. R. (2019). Artificial Neural Networks for Flood Susceptibility Mapping in Data-Scarce Urban Areas. *Spatial Modeling in GIS and R for Earth and Environmental Sciences*, 323–336. https://doi.org/10.1016/B978-0-12-815226-3.00014-4

Farhadi, H., & Najafzadeh, M. (2021). Flood risk mapping by remote sensing data and random forest technique. *Water (Switzerland)*, *13*(21). https://doi.org/10.3390/w13213115

Fleischmann, M., Feliciotti, A., & Kerr, W. (2022). Evolution of Urban Patterns: Urban Morphology as an Open Reproducible Data Science. *Geographical Analysis*, *54*(3), 536–558. https://doi.org/10.1111/gean.12302

Fleischmann, M., Romice, O., & Porta, S. (2021). Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities. *Environment and Planning B: Urban Analytics and City Science*, *48*(8), 2133–2150. https://doi.org/10.1177/2399808320910444

García-Ruiz, J. M., Regüés, D., Alvera, B., Lana-Renault, N., Serrano-Muela, P., Nadal-Romero, E., Navas, A., Latron, J., Martí-Bono, C., & Arnáez, J. (2008). Flood generation and sediment transport in experimental catchments affected by land use changes in the central Pyrenees. *Journal of Hydrology*, *356*(1–2), 245–260. https://doi.org/10.1016/j.jhydrol.2008.04.013

Gauthier, P., & Gilliland, J. (2006). Mapping urban morphology: a classification scheme for interpreting contributions to the study of urban form. *Urban Morphology*, *10*(1), 41–50.

Golkarian, A., Naghibi, S. A., Kalantar, B., & Pradhan, B. (2018). Groundwater potential mapping using C5.0, random forest, and multivariate adaptive regression spline models in GIS. *Environmental Monitoring and Assessment*, *190*(3), 1–16. https://doi.org/10.1007/S10661-018-6507-8/TABLES/5

Government of Kenya. (2018). *Nairobi County Integrated Development Plan (CIDP) 2018-2022*.

Hamidi, A. R., Jing, L., Shahab, M., Azam, K., Tariq, M. A. U. R., & Ng, A. W. M. (2022). Flood Exposure and Social Vulnerability Analysis in Rural Areas of Developing Countries: An Empirical Study of Charsadda District, Pakistan. *Water (Switzerland)*, *14*(7). https://doi.org/10.3390/w14071176

Hawker, L., Neal, J., Tellman, B., Liang, J., Schumann, G., Doyle, C., Sullivan, J. A., Savage, J., & Tshimanga, R. (2020). Comparing earth observation and inundation models to map flood hazards. *Environmental Research Letters*, *15*(12), 124032. https://doi.org/10.1088/1748-9326/ABC216

Helmrich, A. M., Ruddell, B. L., Bessem, K., Chester, M. V., Chohan, N., Doerry, E., Eppinger, J., Garcia, M., Goodall, J. L., Lowry, C., & Zahura, F. T. (2021). Opportunities for crowdsourcing in urban

flood monitoring. *Environmental Modelling & Software*, *143*, 105124. https://doi.org/10.1016/J.ENVSOFT.2021.105124

Hermas, E. S., Gaber, A., & El Bastawesy, M. (2021). Application of remote sensing and GIS for assessing and proposing mitigation measures in flood-affected urban areas, Egypt. *The Egyptian Journal of Remote Sensing and Space Science*, *24*(1), 119–130. https://doi.org/10.1016/J.EJRS.2020.03.002

Horritt, M. S., & Bates, P. D. (2022). Evaluation of 1D and 2D numerical models for predicting river flood inundation. *Journal of Hydrology* , *268*, 87–99. www.elsevier.com/locate/jhydrol

Huang, J.-C. (2021). Redevelopment or retreat for informal settlers? A case study in Shezidao, Taipei, Taiwan. *Journal of Environmental Studies and Sciences* , *11*, 404–411. https://doi.org/10.1007/s13412-021-00687-0/Published

Huchzermeyer, M. (2008). Slum Upgrading in Nairobi within the Housing and Basic Services Market. *Http://Dx.Doi.Org/10.1177/0021909607085586*, *43*(1), 19–39. https://doi.org/10.1177/0021909607085586

Intergovernmental Panel for Climate Change. (2021). *Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* https://www.ipcc.ch/report/ar6/wg1/chapter/summary-for-policymakers/

Intergovernmental Panel on Climate Change. (2022). *Mitigation of Climate Change.* www.ipcc.ch

Intergovernmental Panel on Climate Change (IPCC). (2023). Weather and Climate Extreme Events in a Changing Climate. *Climate Change 2021 – The Physical Science Basis*, 1513–1766. https://doi.org/10.1017/9781009157896.013

Jha, A., Lamond, J., Bloch, R., Bhattacharya, N., Lopez, A., Papachristodoulou, N., Bird, A., Proverbs, D., Davies, J., & Barker, R. (2011). *Five Feet High and Rising Cities and Flooding in the 21 st Century.* http://econ.worldbank.

Joshi, M. Y., Rodler, A., Musy, M., Guernouti, S., Cools, M., & Teller, J. (2022). Identifying urban morphological archetypes for microclimate studies using a clustering approach. *Building and Environment*, *224*. https://doi.org/10.1016/j.buildenv.2022.109574

Juma, B., Olang, L. O., Hassan, M. A., Chasia, S., Mulligan, J., & Shiundu, P. M. (2023). Flooding in the urban fringes: Analysis of flood inundation and hazard levels within the informal settlement of Kibera in Nairobi, Kenya. *Physics and Chemistry of the Earth*, *132*. https://doi.org/10.1016/j.pce.2023.103499

Juma, B., Olang, L. O., Hassan, M., Chasia, S., Bukachi, V., Shiundu, P., & Mulligan, J. (2021). Analysis of rainfall extremes in the Ngong River Basin of Kenya: Towards integrated urban flood risk management. *Physics and Chemistry of the Earth*, *124*. https://doi.org/10.1016/j.pce.2020.102929

Kabenge, M., Elaru, J., Wang, H., & Li, F. (2017). Characterizing flood hazard risk in data-scarce areas, using a remote sensing and GIS-based flood hazard index. *Natural Hazards*, *89*(3), 1369–1387. https://doi.org/10.1007/s11069-017-3024-y

Kadam, P., & Sen, D. (2012). Flood inundation simulation in ajoy river using mike-flood. *ISH Journal of Hydraulic Engineering*, *18*(2), 129–141. https://doi.org/10.1080/09715010.2012.695449

Khalifa, M. A. (2011). Redefining slums in Egypt: Unplanned versus unsafe areas. *Habitat International*, *35*(1), 40–49. https://doi.org/10.1016/j.habitatint.2010.03.004

Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A., & Moradi, A. (2012). An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environmental Earth Sciences*, *67*(1), 251–264. https://doi.org/10.1007/s12665-011-1504-z

KNBS. (2019). *2019 Kenya Population and Housing Census Volume I: Population by County and Subcounty*.

Kohli, D., Sliuzas, R., Kerle, N., & Stein, A. (2012). An ontology of slums for image-based classification. *Computers, Environment and Urban Systems*, *36*(2), 154–163. https://doi.org/10.1016/j.compenvurbsys.2011.11.001

Kopecký, M., Macek, M., & Wild, J. (2021). Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. *Science of the Total Environment*, *757*. https://doi.org/10.1016/j.scitotenv.2020.143785

Kropf, K. (2014). Ambiguity in the definition of built form. *Urban Morphology*, *18*(1), 41–57. https://doi.org/10.51347/jum.v18i1.3995

Kshetri, T. B. (2022). NDVI, NDBI and NDWI calculation using Landsat 7 and 8. *Geomatics for Sustainable Development* , *2*, 32–34. https://www.researchgate.net/publication/327971920

Kuc, G., & Chormański, J. (2019). Sentinel-2 imagery for mapping and monitoring imperviousness in urban areas. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(1/W2), 43–47. https://doi.org/10.5194/isprs-archives-XLII-1-W2-43-2019

Kuffer, M., & Barros, J. (2011). Urban morphology of unplanned settlements: The use of spatial metrics in VHR remotely sensed images. *Procedia Environmental Sciences*, *7*, 152–157. https://doi.org/10.1016/j.proenv.2011.07.027

Kuffer, M., Barros, J., & Sliuzas, R. V. (2014). The development of a morphological unplanned settlement index using very-high-resolution (VHR) imagery. *Computers, Environment and Urban Systems*, *48*, 138–152. https://doi.org/10.1016/j.compenvurbsys.2014.07.012

Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space-15 years of slum mapping using remote sensing. In *Remote Sensing* (Vol. 8, Issue 6). MDPI AG. https://doi.org/10.3390/rs8060455

Kuffer, M., Thomson, D. R., Maki, A., Vanhuysse, S., Georganos, S., Sliuzas, R., & Persello, C. (2021). Eo-based low-cost frameworks to address global urban data gaps on deprivation and multiple hazards. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2106–2109. https://doi.org/10.1109/IGARSS47720.2021.9554094

Kuffer, M., Wang, J., Thomson, D. R., Georganos, S., Abascal, A., Owusu, M., & Vanhuysse, S. (2021). Spatial Information Gaps on Deprived Urban Areas (Slums) in Low-and-Middle-Income-Countries: A User-Centered Approach. *Urban Science*, *5*(4). https://doi.org/10.3390/urbansci5040072

Le Coz, J., Patalano, A., Collins, D., Guillén, N. F., García, C. M., Smart, G. M., Bind, J., Chiaverini, A., Le Boursicaud, R., Dramais, G., & Braud, I. (2016). Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand. *Journal of Hydrology*, *541*, 766–777. https://doi.org/10.1016/J.JHYDROL.2016.07.036

Lea, D., Yeonsu, K., & Hyunuk, A. (2019). Case Study of HEC-RAS 1D–2D Coupling Simulation: 2002 Baeksan Flood Event in Korea. *Water 2019, Vol. 11, Page 2048*, *11*(10), 2048. https://doi.org/10.3390/W11102048

Leandro, J., Schumann, A., & Pfister, A. (2016). A step towards considering the spatial heterogeneity of urban key features in urban hydrology flood modelling. *Journal of Hydrology*, *535*, 356–365. https://doi.org/10.1016/J.JHYDROL.2016.01.060

Lee, J. E., Heo, J. H., Lee, J., & Kim, N. W. (2017). Assessment of Flood Frequency Alteration by Dam Construction via SWAT Simulation. *Water 2017, Vol. 9, Page 264*, *9*(4), 264. https://doi.org/10.3390/W9040264

Lee, S., Kim, J. C., Jung, H. S., Lee, M. J., & Lee, S. (2017). Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Natural Hazards and Risk*, *8*(2), 1185–1203. https://doi.org/10.1080/19475705.2017.1308971

Lee, Y., & Brody, S. D. (2018). Examining the impact of land use on flood losses in Seoul, Korea. *Land Use Policy*, *70*, 500–509. https://doi.org/10.1016/J.LANDUSEPOL.2017.11.019

Lin, J., He, X., Lu, S., Liu, D., & He, P. (2021). Investigating the influence of three-dimensional building configuration on urban pluvial flooding using random forest algorithm. *Environmental Research*, *196*. https://doi.org/10.1016/j.envres.2020.110438

Lin, J., Zhang, W., Wen, Y., & Qiu, S. (2023). Evaluating the association between morphological characteristics of urban land and pluvial floods using machine learning methods. *Sustainable Cities and Society*, 104891. https://doi.org/10.1016/j.scs.2023.104891

Lin, L., Tang, C., Liang, Q., Wu, Z., Wang, X., & Zhao, S. (2023). Rapid urban flood risk mapping for data-scarce environments using social sensing and region-stable deep neural network. *Journal of Hydrology*, *617*. https://doi.org/10.1016/j.jhydrol.2022.128758

Liu, J., Wang, J., Xiong, J., Cheng, W., Li, Y., Cao, Y., He, Y., Duan, Y., He, W., & Yang, G. (2022). Assessment of flood susceptibility mapping using support vector machine, logistic regression and their ensemble techniques in the Belt and Road region. *Geocarto International*, *37*(25), 9817–9846. https://doi.org/10.1080/10106049.2022.2025918

Luke, A., Sanders, B. F., Goodrich, K. A., Feldman, D. L., Boudreau, D., Eguiarte, A., Serrano, K., Reyes, A., Schubert, J. E., Aghakouchak, A., Basolo, V., & Matthew, R. A. (2018). Going beyond the flood insurance rate map: Insights from flood hazard map co-production. *Natural Hazards and Earth System Sciences*, *18*(4), 1097–1120. https://doi.org/10.5194/NHESS-18-1097-2018

Mahabir, R., Crooks, A., Croitoru, A., & Agouris, P. (2016). The study of slums as social and physical constructs: Challenges and emerging research opportunities. *Regional Studies, Regional Science*, *3*(1), 399–419. https://doi.org/10.1080/21681376.2016.1229130

Mahmoud, S. H., & Gan, T. Y. (2018). Multi-criteria approach to develop flood susceptibility maps in arid regions of Middle East. *Journal of Cleaner Production*, *196*, 216–229. https://doi.org/10.1016/j.jclepro.2018.06.047

Mahoney, M. J., Johnson, L. K., Silge, J., Frick, H., Kuhn, M., & Beier, C. M. (2023). *Assessing the performance of spatial cross-validation approaches for models of spatially structured data.* http://arxiv.org/abs/2303.07334

Marwal, A., & Silva, E. A. (2023). Exploring residential built-up form typologies in Delhi: a grid-based clustering approach towards sustainable urbanisation. *Npj Urban Sustainability*, *3*(1). https://doi.org/10.1038/s42949-023-00112-1

McClean, F., Dawson, R., & Kilsby, C. (2020). Implications of Using Global Digital Elevation Models for Flood Risk Analysis in Cities. *Water Resources Research*, *56*(10), e2020WR028241. https://doi.org/10.1029/2020WR028241

Meadows, M., Jones, S., & Reinke, K. (2024). Vertical accuracy assessment of freely available global DEMs (FABDEM, Copernicus DEM, NASADEM, AW3D30 and SRTM) in flood-prone environments. *International Journal of Digital Earth*, *17*(1). https://doi.org/10.1080/17538947.2024.2308734

Mensah, H., & Ahadzie, D. K. (2020). Causes, impacts and coping strategies of floods in Ghana: a systematic review. In *SN Applied Sciences* (Vol. 2, Issue 5). Springer Nature. https://doi.org/10.1007/s42452-020-2548-z

Miranda, F., Franco, A. B., Rezende, O., da Costa, B. B. F., Najjar, M., Haddad, A. N., & Miguez, M. (2023). A GIS-Based Index of Physical Susceptibility to Flooding as a Tool for Flood Risk Management. *Land*, *12*(7). https://doi.org/10.3390/land12071408

Moftakhari, H. R., AghaKouchak, A., Sanders, B. F., Allaire, M., & Matthew, R. A. (2018). What Is Nuisance Flooding? Defining and Monitoring an Emerging Challenge. *Water Resources Research*, *54*(7), 4218–4227. https://doi.org/10.1029/2018WR022828

Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N., & Ghazali, A. H. bin. (2017a). Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Geomatics, Natural Hazards and Risk*, *8*(2), 1080–1102. https://doi.org/10.1080/19475705.2017.1294113

Molinari, D., De Bruijn, K. M., Castillo-Rodríguez, J. T., Aronica, G. T., & Bouwer, L. M. (2019). Validation of flood risk models: Current practice and possible improvements. *International Journal of Disaster Risk Reduction*, *33*, 441–448. https://doi.org/10.1016/j.ijdrr.2018.10.022

Moudon, A. V. (1997). Urban morphology as an emerging interdisciplinary field. *Urban Morphology*, 3–10.

Mudashiru, R. B., Sabtu, N., & Abustan, I. (2021). Quantitative and Semi-quantitative Methods in Flood Hazard Susceptibility Mapping. *Arabian Journal of Geosciences*, *14*.

Mudashiru, R. B., Sabtu, N., Abustan, I., & Balogun, W. (2021). Flood hazard mapping methods: A review. In *Journal of Hydrology* (Vol. 603). Elsevier B.V. https://doi.org/10.1016/j.jhydrol.2021.126846

Mulligan, J., Bukachi, V., Gregoriou, R., Venn, N., Ker-Reid, D., Travers, A., Benard, J., & Olang, L. O. (2019). Participatory flood modelling for negotiation and planning in urban informal settlements. *Proceedings of the Institution of Civil Engineers : Engineering Sustainability*, *172*(7), 354–371. https://doi.org/10.1680/JENSU.17.00020

Mulligan, J., Harper, J., Kipkemboi, P., Ngobi, B., & Collins, A. (2017). Community-responsive adaptation to flooding in Kibera, Kenya. *Proceedings of the Institution of Civil Engineers: Engineering Sustainability*, *170*(5), 268–280. https://doi.org/10.1680/jensu.15.00060

Mumford, L. (1961). *The city in history: its origins, its transformations, and its prospects*.

Mundia, C. N. (2017). *Nairobi Metropolitan Area* (pp. 293–317). https://doi.org/10.1007/978-981-10-3241-7_15

Mutisya, E., & Yarime, M. (2011). Understanding the Grassroots Dynamics of Slums in Nairobi: The Dilemma of Kibera Informal Settlements. In *International Transaction Journal of Engineering* (Vol. 2, Issue 2). http://TuEngr.com/V02/197-213.pdf

Nguyen, H. D., Vu, P. L., Ha, M. C., Dinh, T. B. H., Nguyen, T. H., Hoang, T. P., Doan, Q. C., Pham, V. M., & Dang, D. K. (2022). Flood susceptibility mapping using advanced hybrid machine learning and CyGNSS: a case study of Nghe An province, Vietnam. *Acta Geophysica*, *70*(6), 2785–2803. https://doi.org/10.1007/s11600-022-00940-2

Nsangou, D., Kpoumié, A., Mfonka, Z., Bateni, S. M., Ngouh, A. N., & Ndam Ngoupayou, J. R. (2022). The Mfoundi Watershed at Yaoundé in the Humid Tropical Zone of Cameroon: A Case Study of Urban Flood Susceptibility Mapping. *Earth Systems and Environment*, *6*(1), 99–120. https://doi.org/10.1007/S41748-021-00276-9/FIGURES/11

Nyarko, B. K. (2014). Application of a Rational Model in GIS For Flood Risk Assessment in Accra, Ghana. In *J. of Spatial Hydrology* (Vol. 2, Issue 1). https://www.researchgate.net/publication/205947704

Obudho, R. A., & Aduwo, G. O. (1989). *Slum and squatter settlements in urban centres of kenya: towards a planning strategy*.

Omonge, P., Feigl, M., Olang, L., Schulz, K., & Herrnegger, M. (2022). Evaluation of satellite precipitation products for water allocation studies in the Sio-Malaba-Malakisi river basin of East Africa. *Journal of Hydrology: Regional Studies*, *39*, 100983. https://doi.org/10.1016/J.EJRH.2021.100983

Oyugi Maurice Onyango, B. (2018). *University of Eldoret school of environmental studies department of environmental monitoring, planning and management modelling the effects of urban morphology on environmental quality of Nairobi city, Kenya*.

Paul, M. (2015). The Economic Impacts of Natural Disaster. *Journal of Economic Issues*, *49*(1), 313–315. https://doi.org/10.1080/00213624.2015.1013902

Pham, B. T., Luu, C., Phong, T. Van, Trinh, P. T., Shirzadi, A., Renoud, S., Asadi, S., Le, H. Van, von Meding, J., & Clague, J. J. (2021). Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modelling? *Journal of Hydrology*, *592*, 125615. https://doi.org/10.1016/J.JHYDROL.2020.125615

Pourghasemi, H. R., Kariminejad, N., Amiri, M., Edalat, M., Zarafshar, M., Blaschke, T., & Cerda, A. (2020). Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-60191-3

Pradhan, B. (2009). Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. In *Journal of Spatial Hydrology* (Vol. 9, Issue 2). https://www.researchgate.net/publication/230875021

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, *9*(2), 181–199. https://doi.org/10.1007/S10021-005-0054-1/TABLES/5

Prasad, P., Loveson, V. J., Das, B., & Kotha, M. (2021). Novel ensemble machine learning models in flood susceptibility mapping. *Geocarto International*, *37*(16), 4571–4593. https://doi.org/10.1080/10106049.2021.1892209

Rajula, E. R. (2016). East african Medical Journal challenges in managing and sustaining urban slum health programmes in Kenya. In *African Medical Journal* (Vol. 93, Issue 5).

Ramiaramanana, F. N., & Teller, J. (2021). Urbanization and floods in Sub-Saharan Africa: Spatiotemporal study and analysis of vulnerability factors—case of Antananarivo agglomeration (Madagascar). *Water (Switzerland)*, *13*(2). https://doi.org/10.3390/w13020149

Saco, P. M., McDonough, K. R., Rodriguez, J. F., Rivera-Zayas, J., & Sandi, S. G. (2021). The role of soils in the regulation of hazards and extreme events. *Philosophical Transactions of the Royal Society B*, *376*(1834), 2021. https://doi.org/10.1098/RSTB.2020.0178

Safaei-Moghadam, A., Tarboton, D., & Minsker, B. (2023). Estimating the likelihood of roadway pluvial flood based on crowdsourced traffic data and depression-based DEM analysis. *Natural Hazards and Earth System Sciences*, *23*(1), 1–19. https://doi.org/10.5194/NHESS-23-1-2023

Salami, R. O., von Meding, J. K., & Giggins, H. (2017). Urban settlements' vulnerability to flood risks in african cities: A conceptual framework. *Jamba: Journal of Disaster Risk Studies*, *9*(1). https://doi.org/10.4102/jamba.v9i1.370

Sanders, B. F., Schubert, J. E., Kahl, D. T., Mach, K. J., Brady, D., Aghakouchak, A., Forman, F., Matthew, R. A., Ulibarri, N., & Davis, S. J. (2023). nature sustainability Large and inequitable flood risks in Los Angeles, California. *Nature Sustainability* |, *6*, 47–57. https://doi.org/10.1038/s41893-022-00977-7

Sandink, D., Kovacs, P., Oulahen, G., & Shrubsole, D. (2016). Public relief and insurance for residential flood losses in Canada: Current status and commentary. In *Canadian Water Resources Journal* (Vol. 41, Issues 1–2, pp. 220–237). Taylor and Francis Ltd. https://doi.org/10.1080/07011784.2015.1040458

Santos, P. P., & Reis, E. (2018). Assessment of stream flood susceptibility: a cross-analysis between model results and flood losses. *Journal of Flood Risk Management*, *11*, S1038–S1050. https://doi.org/10.1111/jfr3.12290

Sarkar, D., & Mondal, P. (2020). Flood vulnerability mapping using frequency ratio (FR) model: a case study on Kulik river basin, Indo-Bangladesh Barind region. *Applied Water Science*, *10*(1), 1–13. https://doi.org/10.1007/S13201-019-1102-X/FIGURES/6

Scott, A. A., Misiani, H., Okoth, J., Jordan, A., Gohlke, J., Ouma, G., Arrighi, J., Zaitchik, B. F., Jjemba, E., Verjee, S., & Waugh, D. W. (2017). Temperature and heat in informal settlements in Nairobi. *PLoS ONE*, *12*(11). https://doi.org/10.1371/journal.pone.0187300

Segond, M. L., Wheater, H. S., & Onof, C. (2007). The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK. *Journal of Hydrology*, *347*(1–2), 116–131. https://doi.org/10.1016/j.jhydrol.2007.09.040

Sevgen, E., Kocaman, S., Nefeslioglu, H. A., & Gokceoglu, C. (2019). A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ann and random forest. *Sensors (Switzerland)*, *19*(18). https://doi.org/10.3390/s19183940

Seydi, S. T., Kanani-Sadat, Y., Hasanlou, M., Sahraei, R., Chanussot, J., & Amani, M. (2022). Comparison of Machine Learning Algorithms for Flood Susceptibility Mapping. *Remote Sensing*, *15*(1). https://doi.org/10.3390/rs15010192

Shah, R. K., & Shah, R. K. (2023). GIS-based flood susceptibility analysis using multi-parametric approach of analytical hierarchy process in Majuli Island, Assam, India. *Sustainable Water Resources Management*, *9*(5). https://doi.org/10.1007/s40899-023-00924-0

Shepherd, J. M. (2005). A review of current investigations of urban-induced rainfall and recommendations for the future. *Earth Interactions*, *9*(12). https://doi.org/10.1175/EI156.1

Simpson, A. L., Balog, S., Moller, D. K., Strauss, B. H., & Saito, K. (2015). An urgent case for higher resolution digital elevation models in the world's poorest and most vulnerable countries. In *Frontiers in Earth Science* (Vol. 3). Frontiers Research Foundation. https://doi.org/10.3389/feart.2015.00050

Slum Dwellers International. (n.d.). *An Inventory of the Slums in Nairobi*.

Soliman, M., Morsy, M. M., & Radwan, H. G. (2022). Assessment of Implementing Land Use/Land Cover LULC 2020-ESRI Global Maps in 2D Flood Modelling Application. *Water (Switzerland)*, *14*(23). https://doi.org/10.3390/w14233963

Suhr, F., & Steinert, J. I. (2022). Epidemiology of floods in sub-Saharan Africa: a systematic review of health outcomes. *BMC Public Health*, *22*(1). https://doi.org/10.1186/s12889-022-12584-4

Sun, G., Wei, Y., Wang, G., Shi, R., Chen, H., & Mo, C. (2022). Downscaling Correction and Hydrological Applicability of the Three Latest High-Resolution Satellite Precipitation Products (GPM, GSMAP, and MSWEP) in the Pingtang Catchment, China. *Advances in Meteorology*, *2022*(1), 6507109. https://doi.org/10.1155/2022/6507109

Sverdlik, A. (2021). *Nairobi: City Scoping Study*.

Sy, B., Frischknecht, C., Dao, H., Consuegra, D., & Giuliani, G. (2020). Reconstituting past flood events: The contribution of citizen science. *Hydrology and Earth System Sciences*, *24*(1), 61–74. https://doi.org/10.5194/HESS-24-61-2020

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports 2020 10:1*, *10*(1), 1–10. https://doi.org/10.1038/s41598-020-70816-2

Tanim, A. H., McRae, C. B., Tavakol-davani, H., & Goharian, E. (2022). Flood Detection in Urban Areas Using Satellite Imagery and Machine Learning. *Water (Switzerland)*, *14*(7). https://doi.org/10.3390/w14071140

Tardioli, G., Kerrigan, R., Oates, M., O'Donnell, J., & Finn, D. P. (2018). Identification of representative buildings and building groups in urban datasets using a novel pre-processing, classification, clustering and predictive modelling approach. *Building and Environment*, *140*, 90–106. https://doi.org/10.1016/J.BUILDENV.2018.05.035

Taubenböck, H., & Kraff, N. J. (2014). The physical face of slums: a structural comparison of slums in Mumbai, India, based on remotely sensed data. *Journal of Housing and the Built Environment*, *29*, 15–38. https://doi.org/10.1007/s10901-013-9333-x

Taubenböck, H., Wurm, M., Geiß, C., Dech, S., & Siedentop, S. (2019). Urbanization between compactness and dispersion: designing a spatial model for measuring 2D binary settlement landscape configurations. *International Journal of Digital Earth*, *12*(6), 679–698. https://doi.org/10.1080/17538947.2018.1474957

Tehrany, M. S., Jones, S., & Shabani, F. (2019). Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *Catena*, *175*, 174–192. https://doi.org/10.1016/j.catena.2018.12.011

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology*, *512*, 332–343. https://doi.org/10.1016/j.jhydrol.2014.03.008

Tehrany, M., Shabani, F., Neamah Jebur, M., Hong, H., Chen, W., & Xie, X. (2017). GIS-based spatial prediction of flood prone areas using standalone frequency ratio, logistic regression, weight of evidence and their ensemble techniques. *Geomatics, Natural Hazards and Risk*, *8*(2), 1538–1561. https://doi.org/10.1080/19475705.2017.1362038

Tom, R. O., George, K. O., Joanes, A. O., & Haron, A. (2022). Review of flood modelling and models in developing cities and informal settlements: A case of Nairobi city. In *Journal of Hydrology: Regional Studies* (Vol. 43). Elsevier B.V. https://doi.org/10.1016/j.ejrh.2022.101188

Towfiqul Islam, A. R. M., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., Kuriqi, A., & Linh, N. T. T. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, *12*(3). https://doi.org/10.1016/j.gsf.2020.09.006

Tran, H. N., Rutten, M., Prajapati, R., Tran, H. T., Duwal, S., Nguyen, D. T., Davids, J. C., & Miegel, K. (2024). Citizen scientists' engagement in flood risk-related data collection: a case study in Bui River Basin, Vietnam. *Environmental Monitoring and Assessment*, *196*(3). https://doi.org/10.1007/S10661-024-12419-2

Tschakert, P., Sagoe, R., Ofori-Darko, G., & Codjoe, S. N. (2010). Floods in the Sahel: An analysis of anomalies, memory, and anticipatory learning. *Climatic Change*, *103*(3), 471–502. https://doi.org/10.1007/s10584-009-9776-y

United Nations. (2023). *The Sustainable Development Goals Report.*

United Nations Department of Economic and Social Affairs. (2019). *World Urbanization Prospects: The 2018 Revision.*

United Nations Human Settlements Programme. (2003). *The Challenge of Slums: Global Report on Human Settlements.*

United Nations Human Settlements Programme. (2008). *State of the world's cities 2008/2009 : harmonious cities.* Earthscan.

United Nations Human Settlements Programme. (2009). *Planning Sustainable Cities: Global Report on Human Settlements.*

United Nations Human Settlements Programme. (2015). *Slum Almanac 2015/2016: Tracking Improvement in the Lives of Slum Dwellers.*

United Nations Human Settlements Programme. (2018). Adequate Housing and Slum Upgrading. In *SDG Indicator 11.1.1 Training Module.*

United Nations Human Settlements Programme. (2020). *Public Space Inventory and Assessment: Reclaiming The Green City In The Sun.* www.unhabitat.org

United Nations Settlements Programme. (2015). Issue Paper 22 - Informal Settlements. In *Habitat III Issue Papers* (pp. 150–155).

Vallely, A., Shagi, C., Kasindi, S., Desmond, N., Lees, S., Chiduo, B., Hayes, R., Allen, C., & Ross, D. (2007). The benefits of participatory methodologies to develop effective community dialogue in the context of a microbicide trial feasibility study in Mwanza, Tanzania. *BMC Public Health, 7.* https://doi.org/10.1186/1471-2458-7-133

Van Den Bout, B., Jetten, V. G., Van Westen, C. J., & Lombardo, L. (2023). *A breakthrough in Fast Flood Simulation.*

Van Den Bout, B., Jetten, V., Van Westen, C., & Lombardo, L. (2023). *Fastflood.org, an open-source super-fast flood model in the browser.* https://doi.org/10.5194/egusphere-egu23-3611

Walęga, A. (2016). The importance of calibration parameters on the accuracy of the floods description in the Snyder's model. *Journal of Water and Land Development, 28*(1), 19–25. https://doi.org/10.1515/jwld-2016-0002

Walsh, C. J., Fletcher, T. D., & Burns, M. J. (2012). Urban Stormwater Runoff: A New Class of Environmental Flow Problem. *PLoS ONE, 7*(9). https://doi.org/10.1371/journal.pone.0045814

Wang, J., Fleischmann, M., Venerandi, A., Romice, O., Kuffer, M., & Porta, S. (2023). EO + Morphometrics: Understanding cities through urban morphology at large scale. *Landscape and Urban Planning, 233.* https://doi.org/10.1016/j.landurbplan.2023.104691

Wang, J., Kuffer, M., & Pfeffer, K. (2019). The role of spatial heterogeneity in detecting urban slums. *Computers, Environment and Urban Systems, 73*, 95–107. https://doi.org/10.1016/j.compenvurbsys.2018.08.007

Wang, Z., Lyu, H., & Zhang, C. (2023). Pluvial flood susceptibility mapping for data-scarce urban areas using graph attention network and basic flood conditioning factors. *Geocarto International*, *38*(1). https://doi.org/10.1080/10106049.2023.2275692

Wanjiru, M. W., & Matsubara, K. (2017). Slum Toponymy in Nairobi, Kenya. *Urban and Regional Planning Review*, *4*(0), 21–44. https://doi.org/10.14398/urpr.4.21

Ward, P. J., Jongman, B., Weiland, F. S., Bouwman, A., Van Beek, R., Bierkens, M. F. P., Ligtvoet, W., & Winsemius, H. C. (2013). Assessing flood risk at the global scale: Model setup, results, and sensitivity. *Environmental Research Letters*, *8*(4). https://doi.org/10.1088/1748-9326/8/4/044019

World Bank. (2006). Social Dimensions of Disaster. In *Hazards of Nature, Risks to Development Natural Disasters: An IEG Evaluation of World Bank Assistance for Natural Disasters* (pp. 43–52). The World Bank. https://doi.org/10.1596/978-0-8213-6650-9

World Health Organization, & United Nations Human Settlements Programme. (2016). *Global report on urban health : equitable, healthier cities for sustainable development.*

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*(2), 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

Youssef, A. M., Pourghasemi, H. R., & El-Haddad, B. A. (2022). Advanced machine learning algorithms for flood susceptibility modelling — performance comparison: Red Sea, Egypt. *Environmental Science and Pollution Research*, *29*(44), 66768–66792. https://doi.org/10.1007/s11356-022-20213-1

Yu, H., Luo, Z., Wang, L., Ding, X., & Wang, S. (2023). Improving the Accuracy of Flood Susceptibility Prediction by Combining Machine Learning Models and the Expanded Flood Inventory Data. *Remote Sensing*, *15*(14). https://doi.org/10.3390/RS15143601

Zeng, Z., Lan, J., Hamidi, A. R., & Zou, S. (2020). Integrating Internet media into urban flooding susceptibility assessment: A case study in China. *Cities*, *101*, 102697. https://doi.org/10.1016/J.CITIES.2020.102697

Zhu, Y., Geiß, C., & So, E. (2019). Using deep neural networks for predictive modelling of informal settlements in the context of flood risk. *Journal of Physics: Conference Series*, *1343*(1). https://doi.org/10.1088/1742-6596/1343/1/012032

APPENDICES

Appendix 1

Figure 20: Flood Validation Workshop Outputs

Appendix 2

Table 10: Grouping of Flood Factors into 7 Broader Categories

No	Factor Groupings	Flood Factors
1	Area covered by buildings	Area covered by buildings
2	Density of buildings	Density of buildings
3	Arrangement of buildings	Fractal dimension Neighbour distance Shared walls Mean interbuilding distance Building adjacency Orientation
4	Building in low areas and near rivers	Distance to rivers Elevation & its derivatives
5	Lack of green spaces	NDVI, NDBI & LULC
6	Poor soils	NDWI Soil infiltration
7	Blocked drainage by waste	Waste

Appendix 3

Table 11: Flood Factors Pairwise Ranking Matrix

	Covered area ratio	Building density	Arrangement of buildings	Building in low areas and rivers	Lack of green spaces	Soil porosity	Blocked drainage by waste	Votes
Covered area ratio								
Building density								
Arrangement of buildings								
Building in low areas and near rivers								
Lack of green spaces								
Soil porosity								
Blocked drainage by waste								

Explanatory visuals of the urban morphology and hydrological factors

Participants discussing the ranking of factors

Pairwise ranking of flood factors

Figure 21: Factor Ranking Methodology and Output for Mathare

Appendix 4
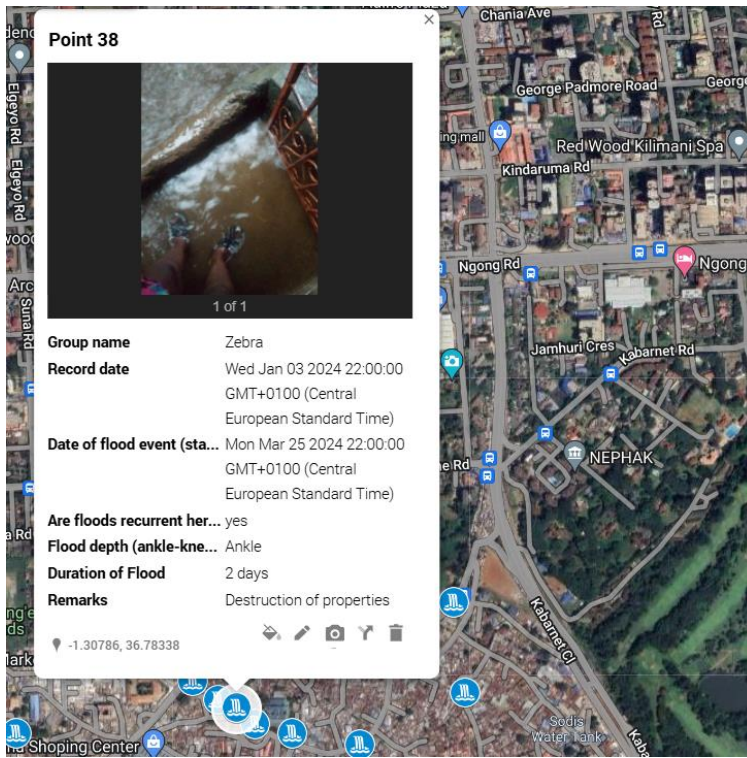
Figure 22: Screenshot of Flood Information Collected using My Maps

Appendix 5
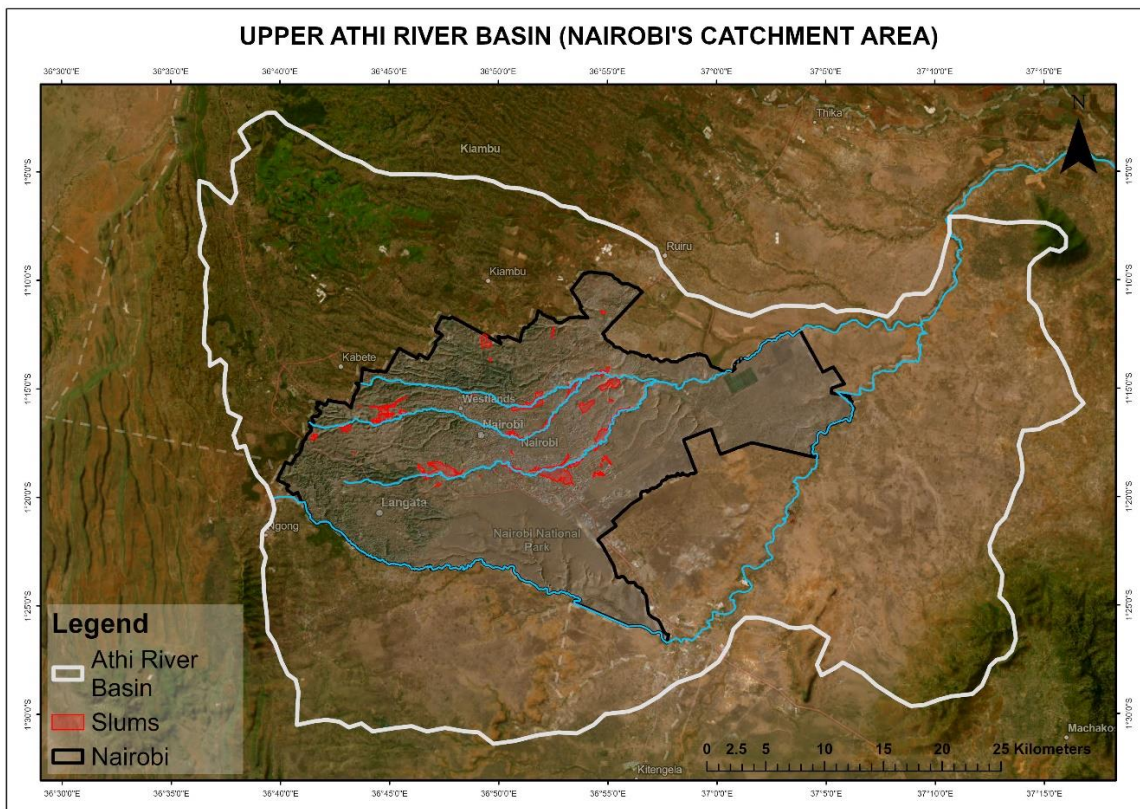
Figure 23: Upper Athi River Catchment Basin

Appendix 6

Table 12: Comparison between Fast Flood and ESRI Mannings' Value

Default FastFlood Manning's Values		LULC 2020 - ESRI			
ESA land cover	Value	Esri LULC	Minimum value	Maximum value	Weighted average
1. Forest	0.12	1. Water	0.025	0.05	0.038
2. Shrubs	0.1	2. Trees	0.079	0.174	0.126
3. Grass	0.03	3. Grass	0.025	0.05	0.038
4. Crops	0.05	4. Flooded vegetation	0.05	0.085	0.061
5. Building	0.02	5. Crops	0.02	0.05	0.035
6. Bare	0.01	6. Shrubs	0.07	0.16	0.115
7. Snow	0.01	7. Built Area	0.064	0.119	0.092
8. Water	0.01	8. Bare ground	0.23	0.03	0.027
9. Wetlands	0.09				
10. Mangroves	0.14				
11. Moss	0.07				

Appendix 7
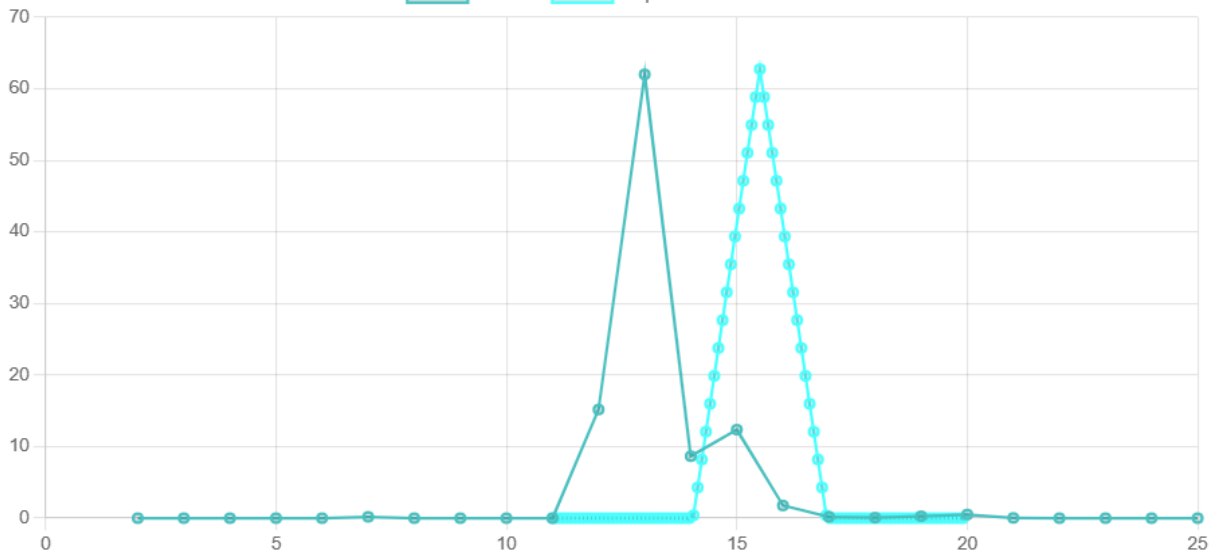
Figure 24: Rainfall Parameters Derived from the Rainfall Analysis

Appendix 8
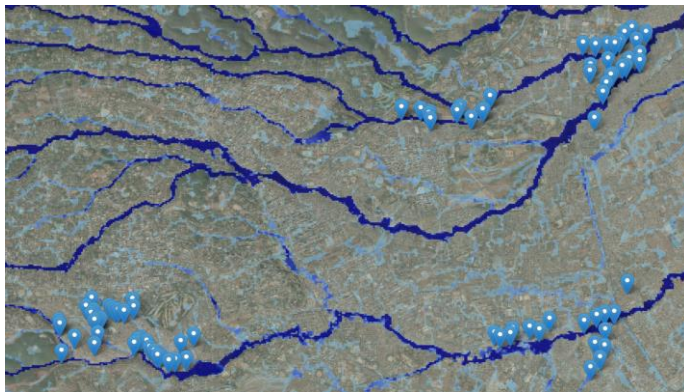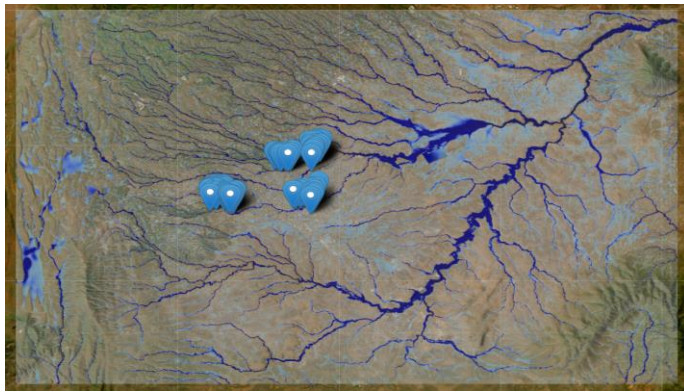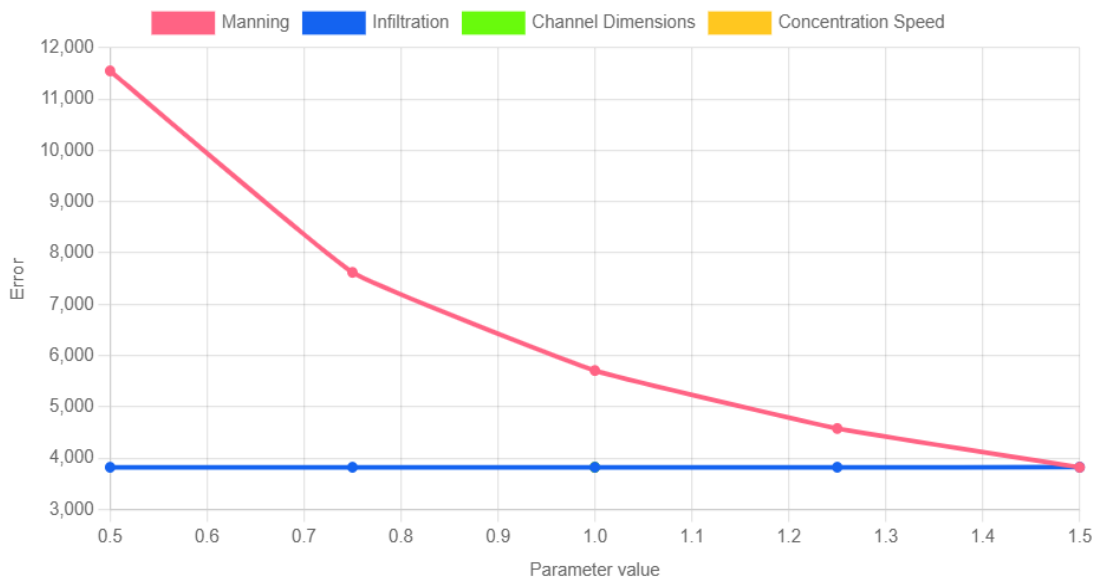
Figure 25: Calibration Settings and Output

Appendix 9

```
        Class_Label  Probability_Class_0  Probability_Class_1
0                 0             0.682330             0.317670
1                 0             0.759547             0.240453
2                 0             0.870027             0.129973
3                 0             0.875427             0.124573
4                 0             0.682127             0.317873
...             ...                  ...                  ...
42015             0             0.579587             0.420413
42016             0             0.625861             0.374139
42017             0             0.726669             0.273331
42018             1             0.492610             0.507390
42019             1             0.330080             0.669920

[42020 rows x 3 columns]
```

Figure 26: Flood Class Prediction and Probabilities

Appendix 10

Table 13: Descriptive Statistics of Flood Factors

Flood Factors	Mean	Standard Deviation	Minimum	Maximum
Building Orientation	17.910	10.302	0	45
Shared Wall Ratio	0.162	0.140	0	1.768
Building Alignment	6.452	4.070	0	39.097
Building Neighbour Distance	23.366	22.411	0	198.989
Mean Inter-building Distance	22.558	14.190	0	198.989
Building Adjacency	0.716	0.158	0	1
Fractal Dimension	1.034	0.042	0.210	8.454
Covered Area Ratio	0.213	0.170	0.0002	1.014
Building Density	0.002	0.002	0.0001	0.015
Elevation	1667.138	117.407	1462.05	1933.98
Slope	3.031	2.876	0	25.618
Curvature	0.006	0.428	-3.445	3.635
Aspect	152.031	105.031	-1	359.995
Distance to Rivers	303.394	221.618	0	1560.29
Flow Accumulation	9.492e+02	1.492e+02	0	1.65
LULC	35.856	17.085	10	80
TWI	8.542	2.273	4.515	23.3
SPI	923.065	11041.919	0	596846
NDVI	0.296	0.194	-0.15	0.843
NDWI	-0.368	0.162	-0.79	0.381
NDBI	0.054	0.128	-0.6	0.533
Precipitation	113.443	36.238	59.342	182.302
Clay Content	48.211	26.336	0	71

Appendix 11
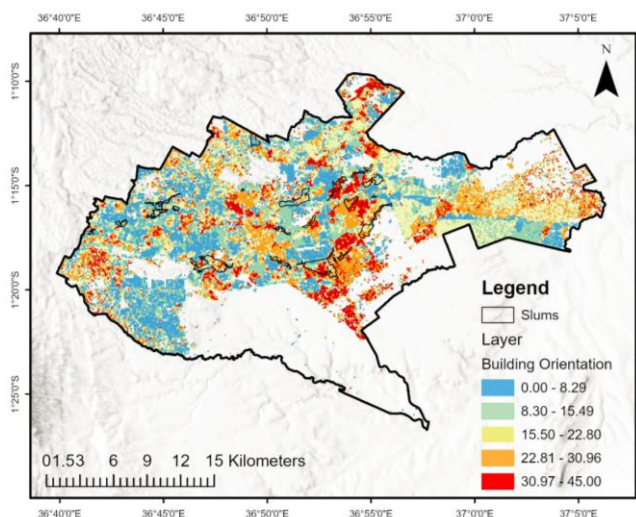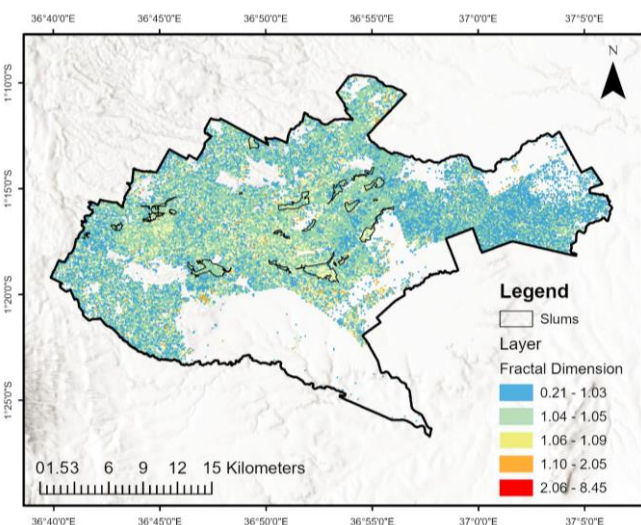

Building Orientation
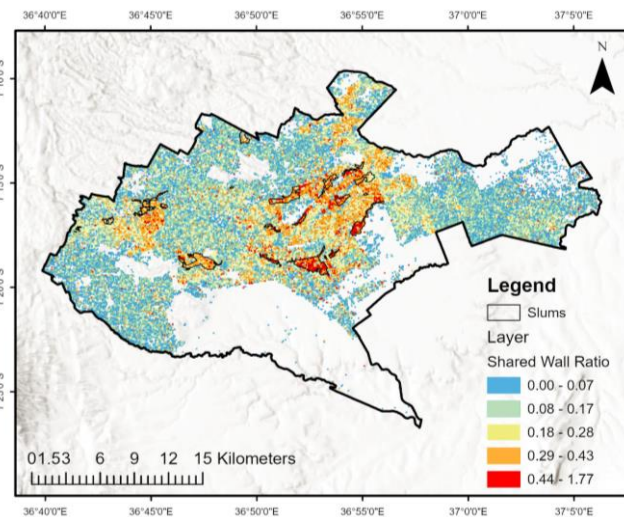
Shared Wall Ratio
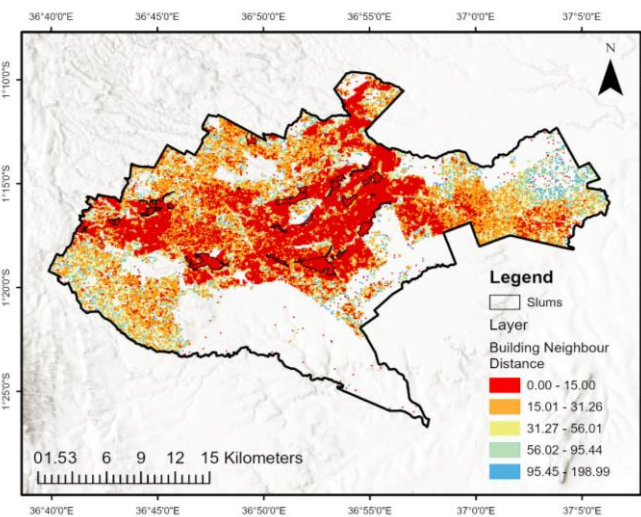
Building Neighbour Distance
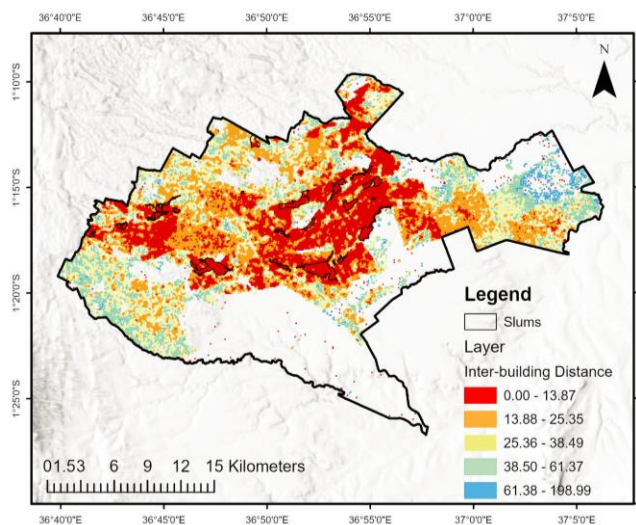
Mean Inter-building Distance
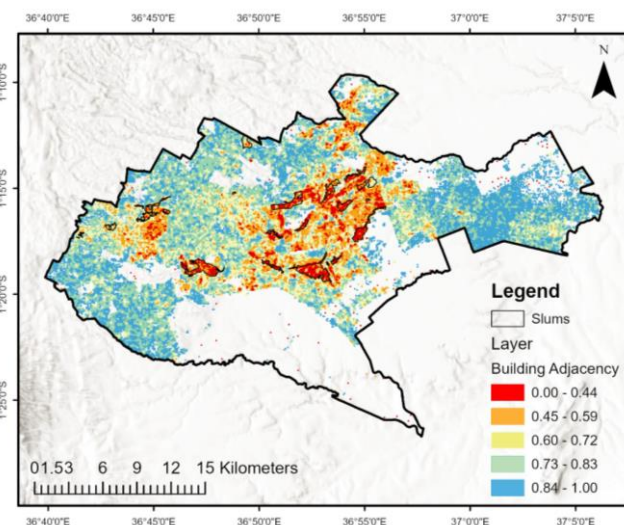
Building Adjacency

Fractal Dimension
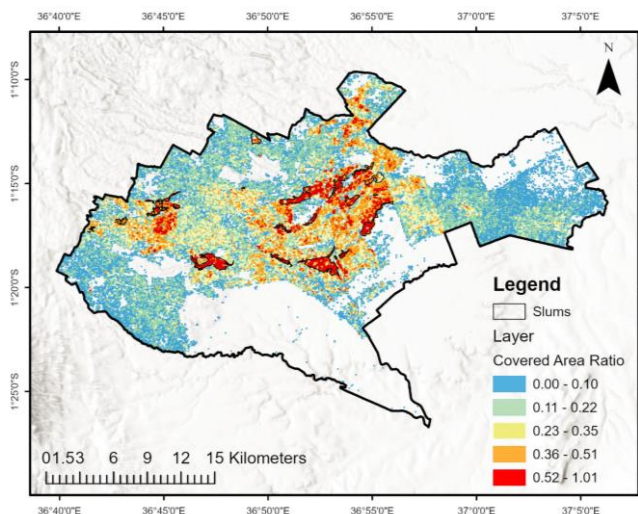
Covered Area Ratio
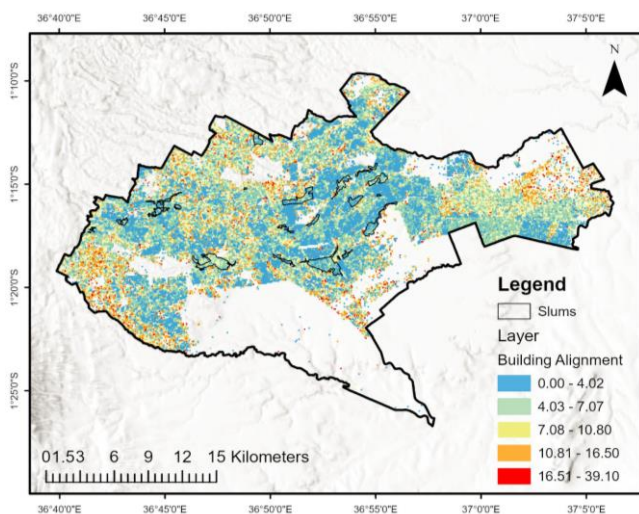
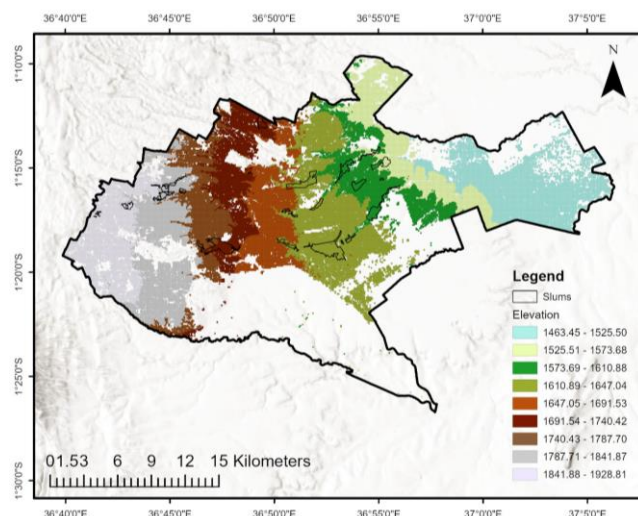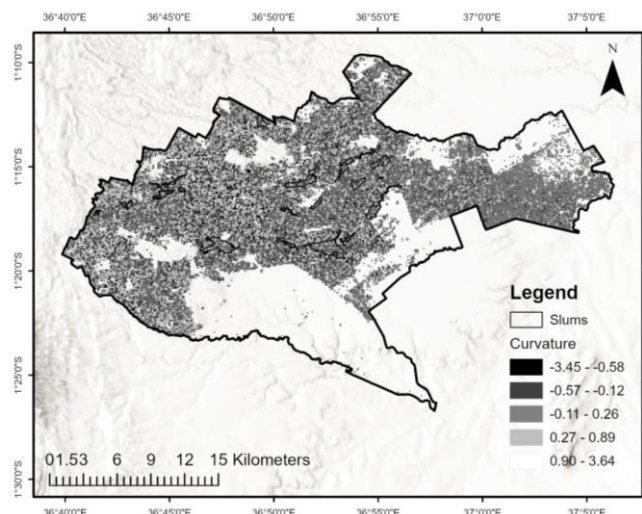**Building Alignment**
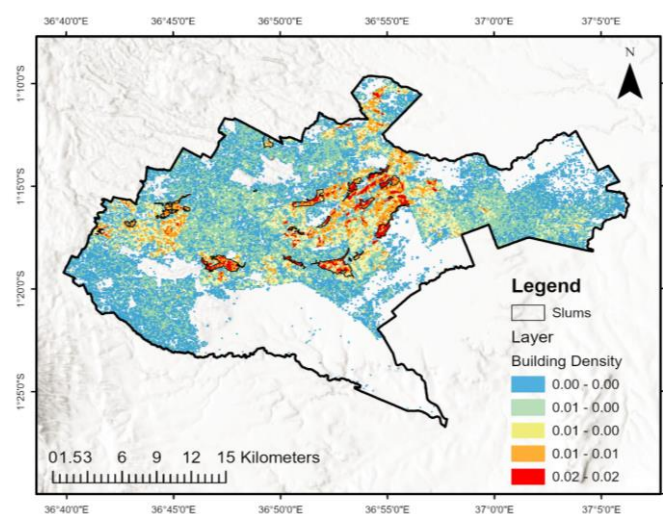
Building Density

Elevation
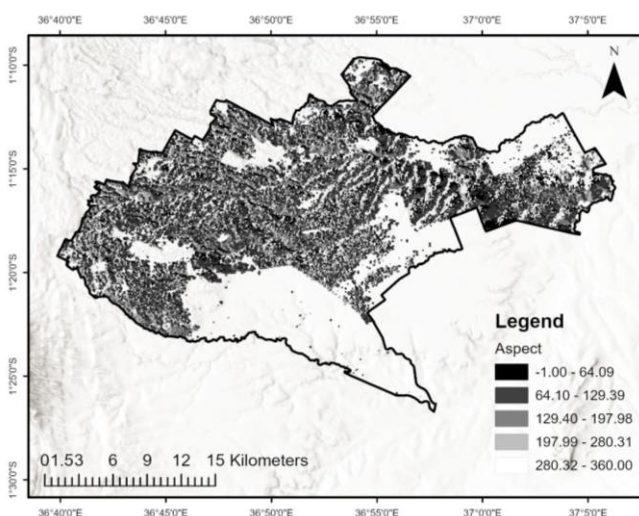
Aspect
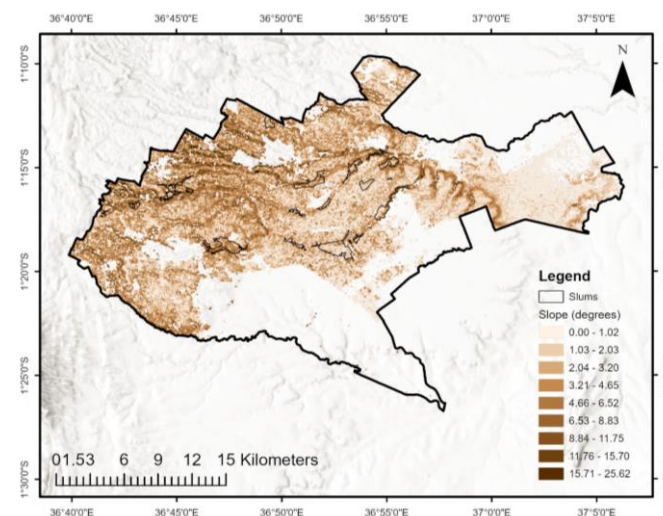
Slope

Curvature
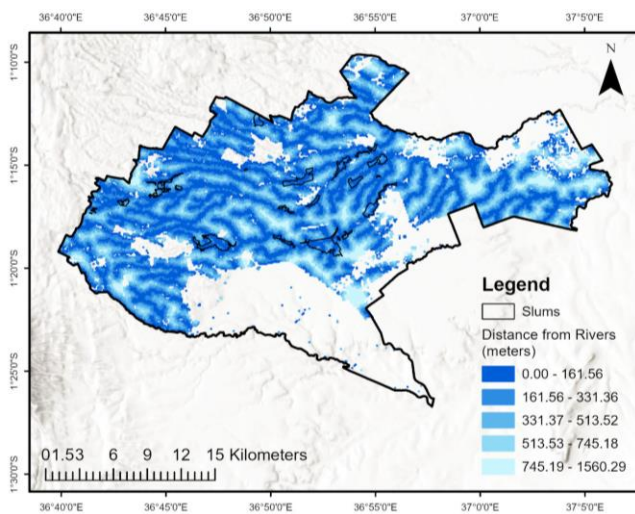
Distance to Rivers
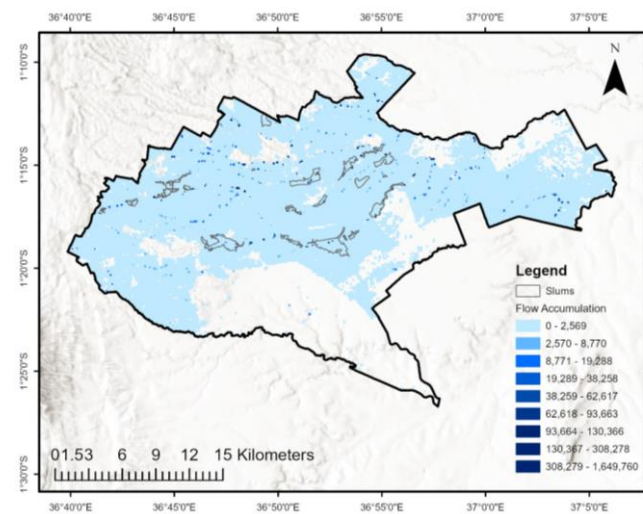
Flow Accumulation
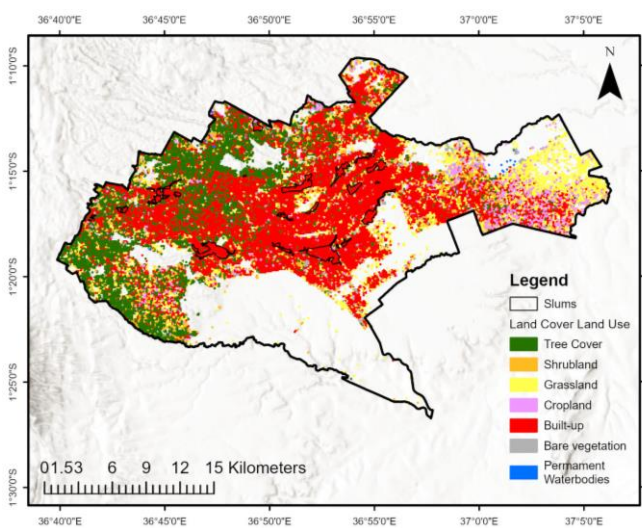
Land Cover Land Use
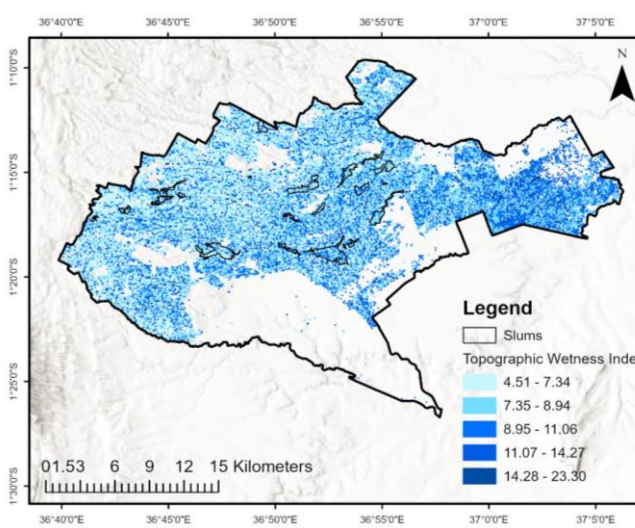
Topographic Wetness Index
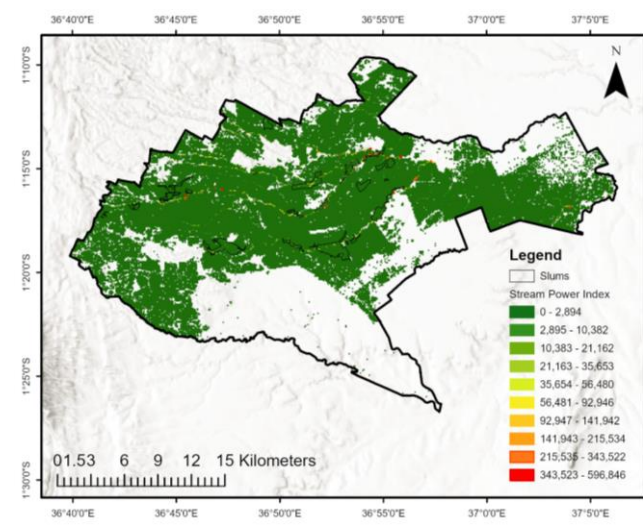
Stream Power Index

NDVI

NDWI

NDBI

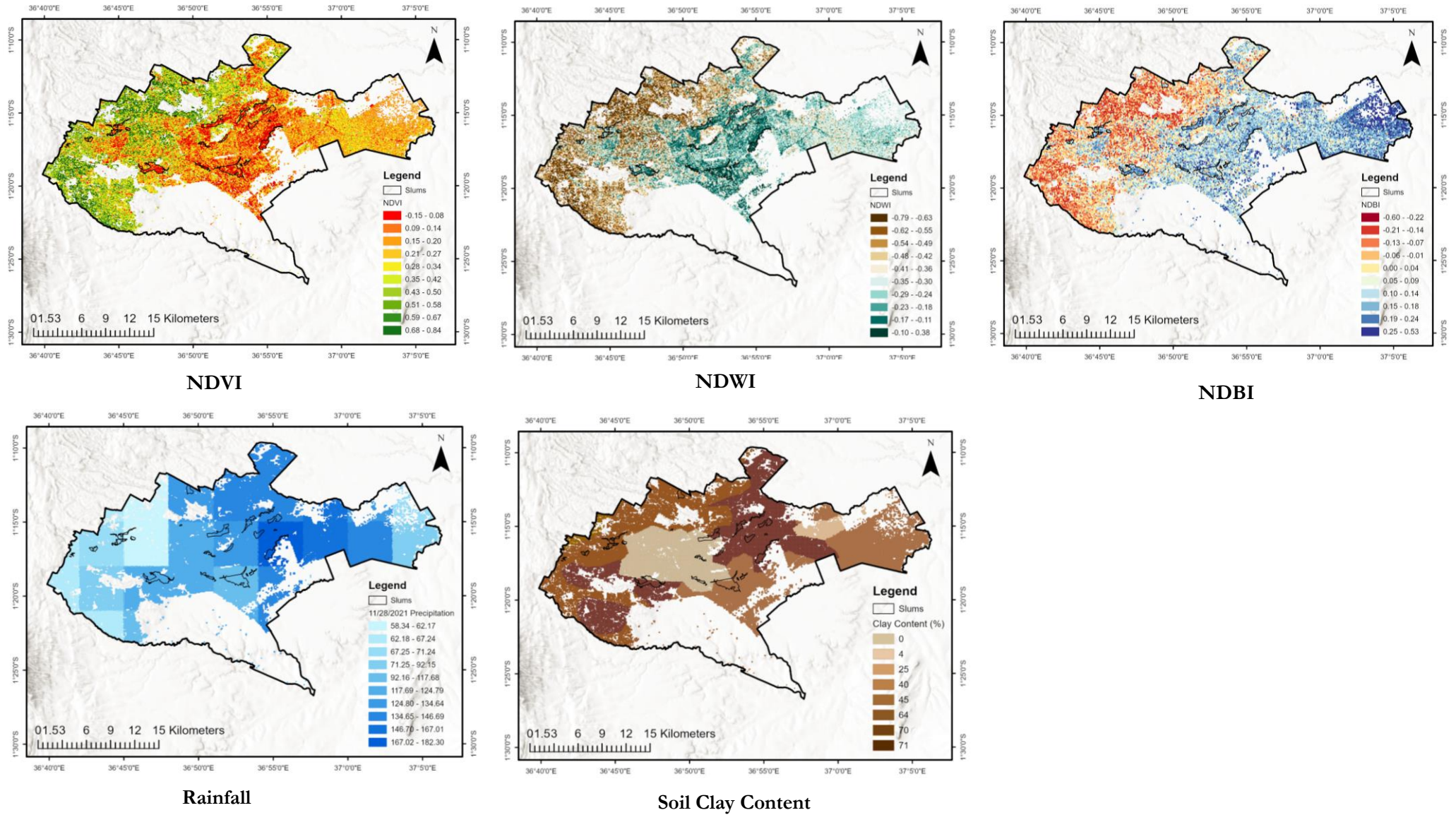**Rainfall**

Soil Clay Content

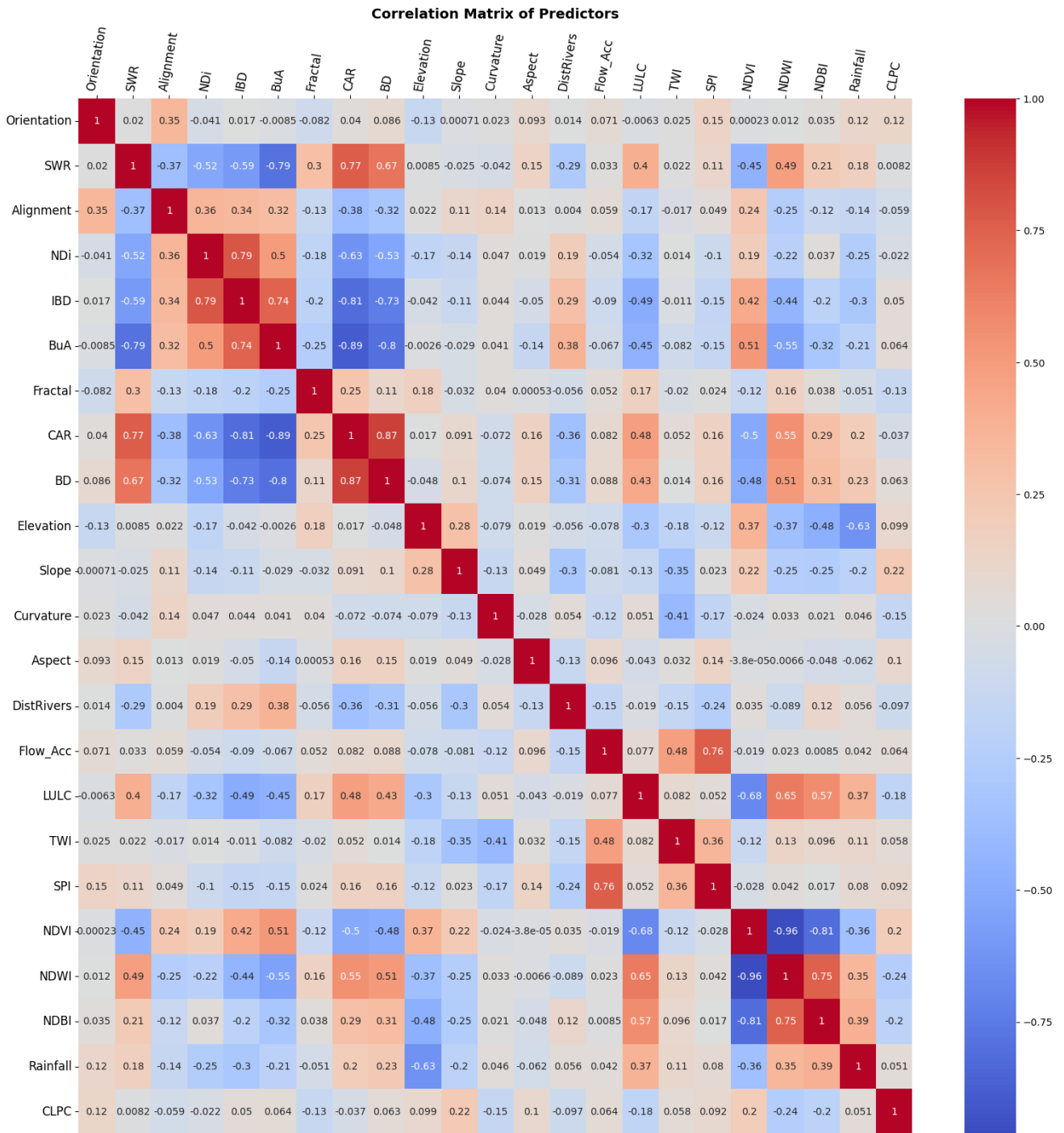Figure 27: Flood Factors Represented in Grid Level (100m*100m)

Appendix 12

Figure 28: Flood Factors Correlation Matrix

Appendix 13

Table 14: Covariances of Flood Factors

| Variable	Variables Alias	Covariance with classes
Building Orientation	Orientation	0.41
Shared Wall Ratio	SWR	0.03
Building Alignment	Alignment	-0.06
Building Neighbour Distance	Ndi	-3.45
Interbuilding Distance	IBD	-3.15
Building Adjacency	BuA	-0.05
Fractal dimension	FD	0.00
Covered Area Ratio	CAR	0.05
Building Density	BD	0.00
Elevation	Elevation	-0.83
Slope	Slope	0.12
Curvature	Curvature	0.00
Aspect	Aspect	4.28
Distance from Rivers	DistRivers	-49.17
Flow Accumulation	Flow_Acc	1238.55
Land Use Land Cover	LULC	1.43
Topographic Wetness Index	TWI	0.30
Stream Power Index	SPI	1337.48
Normalized Difference Vegetation Index	NDVI	-0.02
Normalized Difference Built-up Index	NDBI	0.02
Normalized Difference Wetness Index	NDWI	0.01
Precipitation	Precip	1.89
Soil Clay Content	CLPC	-0.34

Appendix 14
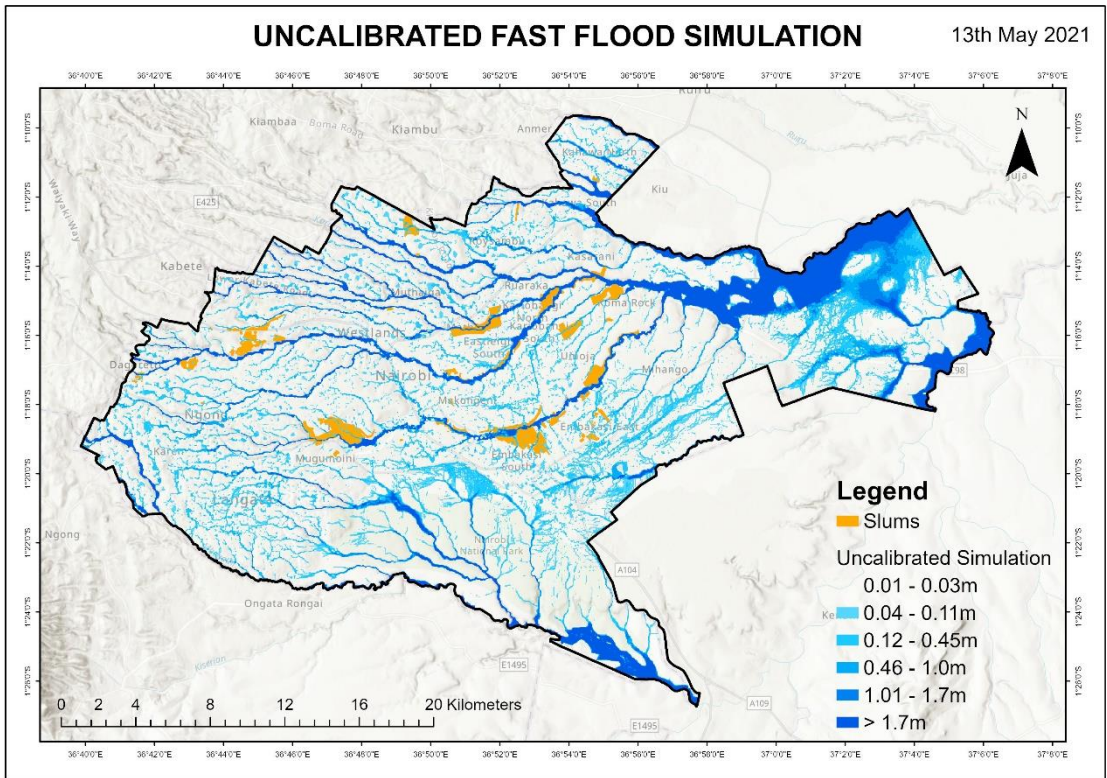
Figure 29: Uncalibrated FFS
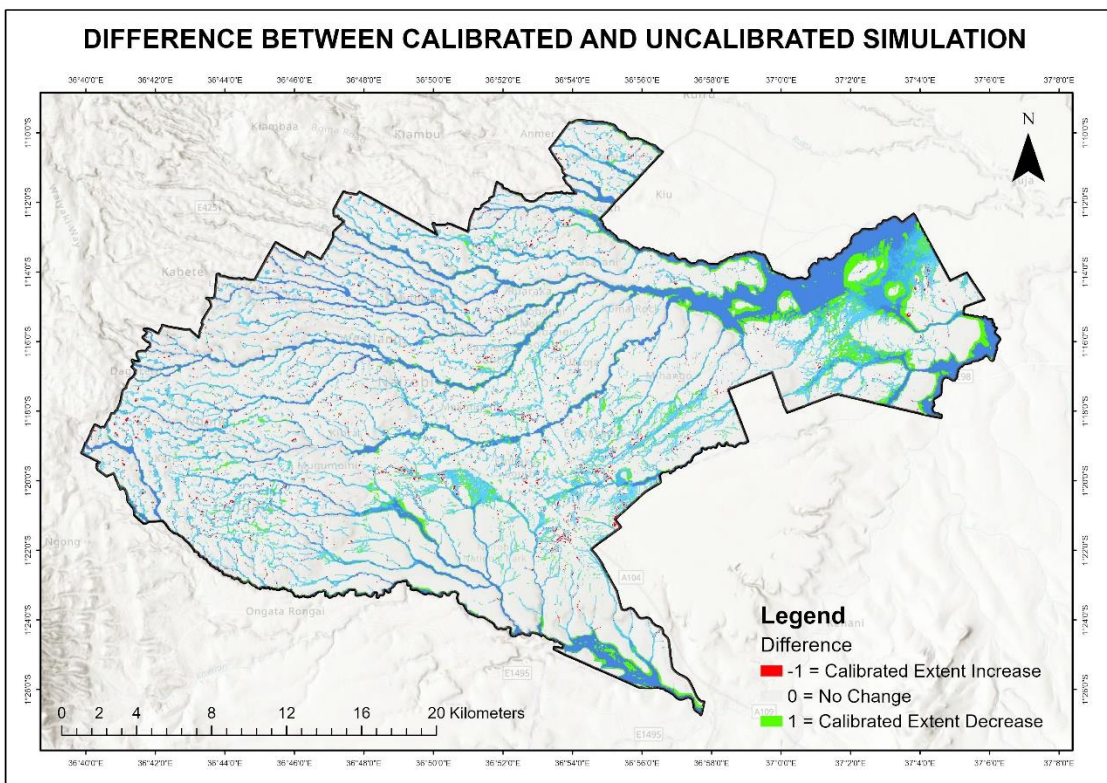
Appendix 15

Figure 30: Difference between the Calibrated and Uncalibrated Simulation
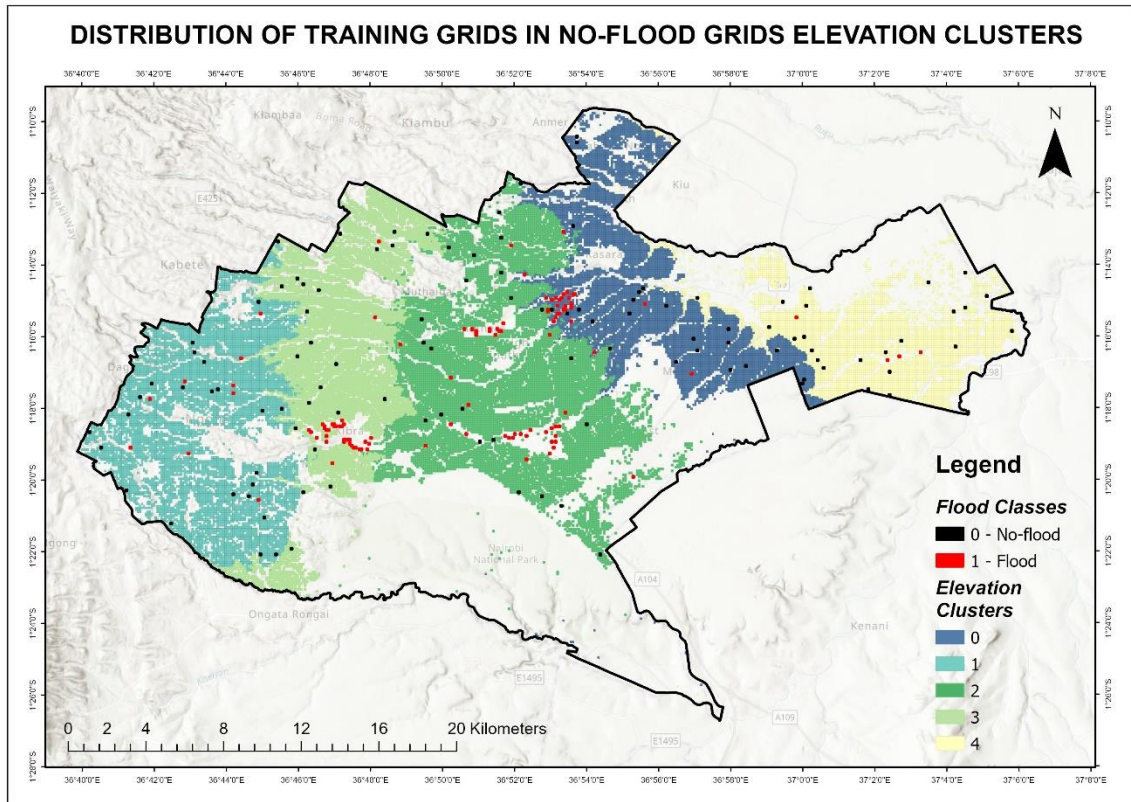
Appendix 16

Figure 31: Distribution of Training Sample Points in No-flood Elevation Clusters
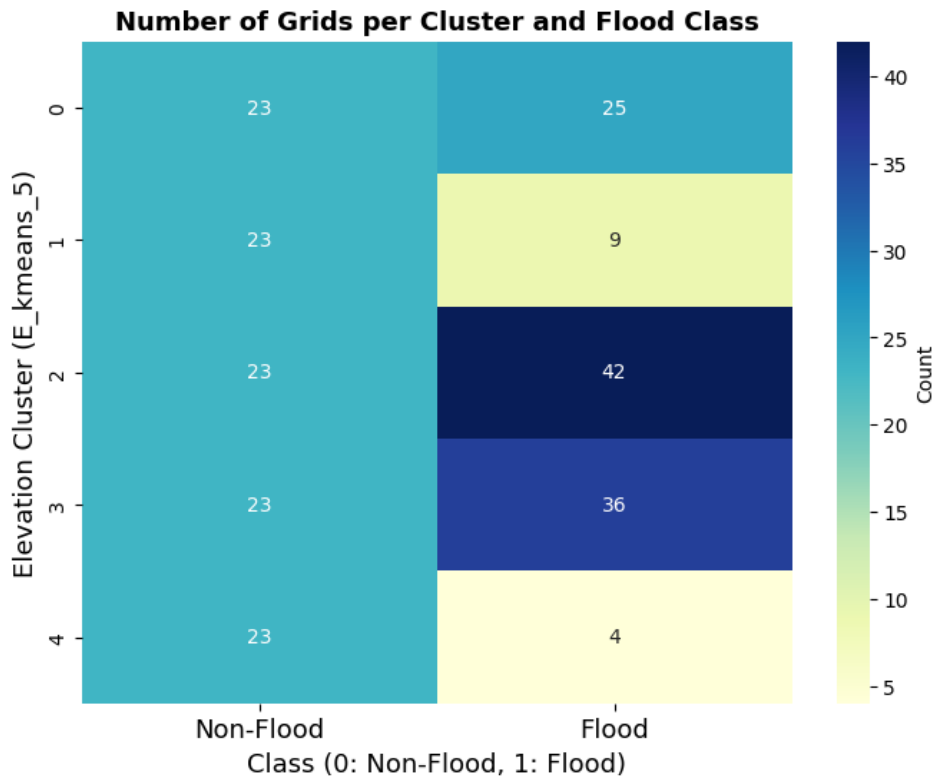
Appendix 17

Figure 32: Cluster Imbalance within Elevation Grids

Appendix 18

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
UNIVERSITY OF TWENTE, ENSCHEDE, THE NETHERLANDS

CONSENT LETTER (Residents)

Dear respondents,

I am Jane-marie Munyi, an MSc student in Urban Planning and Management at the Faculty of Geo-information and Earth Observation (ITC), at the University of Twente set to complete my thesis to graduate.

I am researching on the influence of urban morphology on flooding in deprived areas (slums), and I am reaching out for your assistance to help me understand flooding in your settlements and Nairobi. I am asking for help to learn about why floods affect slums more than formal settlements.

Your participation is voluntary and the information you will share is for academic purposes and be treated as very confidential. By sharing your knowledge and participating, you will help your community know which areas flood more than others and why some areas flood more than others which will help raise awareness about flooding.

Your support will be very helpful in getting this information. If you have any questions about participating, please feel free to ask.

Thank you for considering my request. Your support will be greatly appreciated

1) Do you agree to participate in this survey/interview/ focus group discussion?

    a. No   b. Yes

    *(If **yes**, please put your details in the table below)*

2) Community of engagement _____

3) Name of community leader _____   Contact _____

4) Date _____

5) Duration of engagement, from _____ to _____

Participant *(name optional)*	Gender	Location of residence
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
UNIVERSITY OF TWENTE, ENSCHEDE, THE NETHERLANDS

CONSENT LETTER (City Experts)

Dear respondent,

I am Jane-marie Munyi, an MSc student in Urban Planning and Management at the Faculty of Geo-information and Earth Observation (ITC), at the University of Twente set to complete my thesis to graduate.

I am researching on the influence of urban morphology on flooding in deprived areas (slums) and reaching out for your assistance to gain knowledge about flooding in Nairobi, in the context of flood characteristics and influencing factors. As a professional, I would like to understand from your expertise how urban morphology has shaped flooding in the City and gain insights as to why slums or other areas are seen to be hit harder by flood events as compared to formal settlements.

Your participation would greatly enrich my research and contribute to a more thorough understanding of the complex flood dynamics in Nairobi. Your participation is voluntary and the information you will share is solely for academic purposes and will be treated as very confidential.

Your professional expertise and insights would be valuable in contributing to the depth and relevance of my study. If you have any questions about participating, please feel free to ask.

Thank you for considering my request. Your support will be greatly appreciated.

1) Do you agree to participate in this survey/interview/ focus group discussion?

    a. No   b. Yes

2) Name of official (optional) _____

    Contact _____

3) Function/Department/Role _____

4) Date _____

5) Duration of engagement, from _____ to _____

Figure 33: Residents and Experts Fieldwork Consent Forms

Appendix 19

Table 15: DMP Summary

NAME OF DATA FILE	SOURCE (SECONDARY DATA)	IF SECONDARY, WHO IS THE OWNER?	RESTRICTIONS AND LICENSE	DATA FORM	DATA FORMAT	CONTAINS PERSONAL DATA (Yes/No)	DATASET AVAILABILITY
Copernicus GLO-30 Digital Elevation Model (DEM)	Secondary	European Space Agency (ESA)	Free but not for commercial use.	Raster	GeoTiff	No	2019-2026
Slope	Secondary	Derived from Copernicus GLO-30 DEM	Free	Raster	GeoTiff	No	
Aspect	Secondary	Derived from Copernicus GLO-30 DEM	Free	Raster	GeoTiff	No	
Curvature	Secondary	Derived from Copernicus GLO-30 DEM	Free	Raster	GeoTiff	No	
Stream Power Index	Secondary	Derived from Copernicus GLO-30 DEM	Free	Raster	GeoTiff	No	
Topographic Wetness Index	Secondary	Derived from Copernicus GLO-30 DEM	Free	Raster	GeoTiff	No	
Flow Accumulation	Secondary	Derived from Copernicus GLO-30 DEM	Free	Raster	GeoTiff	No	
Distance to rivers	Secondary	Derived from Copernicus GLO-30 DEM	Free	Raster	GeoTiff	No	
NDVI Sentinel-2MSI	Secondary	European Space Agency (ESA)	Free but not for commercial use.	Satellite Raster	GeoTiff	No	2017 - present
NDWI Sentinel-2MSI	Secondary	European Space Agency (ESA)	Free but not for commercial use.	Satellite Raster	GeoTiff	No	2017 - present
NDBI Landsat 8 OLI	Secondary	U.S Geological survey (USGS)	Open data policy	Satellite Raster	GeoTiff	No	2013 - present
LULC ESA WorldCover	Secondary	ESA/VITO/Brockmann Consult/CS/GAMMA	CC-BY-4.0.	Satellite Raster	GeoTiff	No	2020 - 2021

		Remote Sensing/IIASA/WUR	The policy allows free and public access to data.				
CHIRPS Rainfall	Secondary	UCSB/CHG	Public access	Raster	GeoTiff	No	1981 - present
Soil Clay Content	Secondary	Soul and Terrain database for Kenya (KENSOTER)	CC-BY-3.0. The policy allows free and public access to data.	Vector	Shapefile	No	Made available in 2004
Waste	Secondary	SLUMAP	Proprietary public access to data	Vector	Shapefile	No	Made available in 2021
Slums	Secondary	Spatial Collective	Proprietary public access to data	Vector	Shapefile	No	Made available in 2020
Building Footprints	Secondary	Google Open Buildings	CC-BY-4.0 and ODbl v1.0. The policy allows free and public access to data	Excel sheet	csv	No	2021 - present
Flood Hotspot Points/Locations	Primary			Survey data	kml	No	Collected in 2024
Kibera Flood Map	Secondary	Kounkuey Design Initiative (KDI)	Requested access to data	Flood map	Raster	No	2023
Nairobi flood extent	Secondary	Airbus Defence and Space		Satellite imagery	Vector	No	Made available in 2024
Soil infiltration and moisture	Secondary	SOILGRIDS	CC-BY-4.0. The policy allows free and public access to data.	Raster	GeoTiff	No	Latest release in 2020
Rainfall	Secondary	TAHMO	Requested access to data	Excel sheet	csv	No	2019-2024