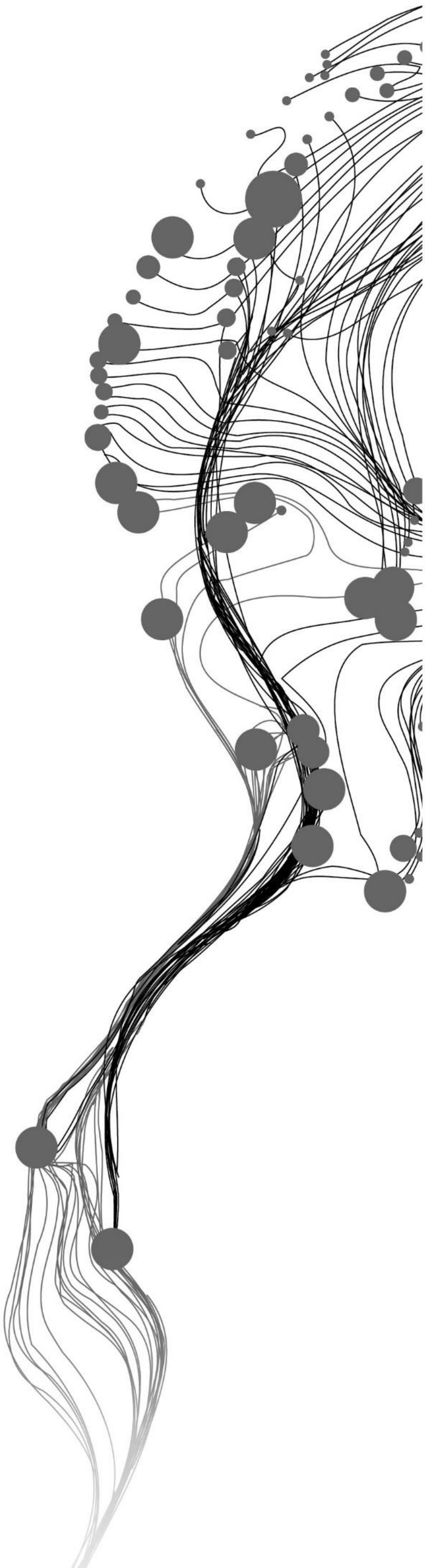
Assessment of Population's Exposure to Flood in A Leptospirosis Endemic Slum Area in Brazil Through the Integration of 2D and 3D Data

SURAKSHYA DHAKAL
August 2021

SUPERVISORS:
dr. M.N. Koeva
dr. M. Kuffer

ADVISORS:
dr. P.L Brito, Politechnic School of Federal University of Bahia (UFBA)
dr. J.C. Pedrassoli, Politechnic School of Federal University of Bahia (UFBA)



Assessment of Population's Exposure to Flood in A Leptospirosis Endemic Slum Area in Brazil Through the Integration of 2D and 3D Data

SURAKSHYA DHAKAL
Enschede, The Netherlands, August 2021

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 (UPM)

SUPERVISORS:
dr. M.N. Koeva
dr. M. Kuffer

ADVISORS:
dr. P.L Brito, Politechnic School of Federal University of Bahia (UFBA)
dr. J.C. Pedrassoli, Politechnic School of Federal University of Bahia (UFBA)

THESIS ASSESSMENT BOARD:
dr. J. A Martinez (Chair)
dr. V. Andreo (External Examiner)

ABSTRACT

Flood is a risk factor for the spread of leptospirosis disease, which is caused by *Leptospira* bacteria that spread through water. Slums that receive intense rainfalls, lie in floodplains, and lack storm water drainage systems are prone to flooding. Since leptospirosis is endemic in Salvador, Brazil, the slum communities like Alto do Cabrito are thought to be at a high risk of leptospirosis as the flood brings people in contact with the *Leptospira* bacteria. It is difficult to eradicate leptospirosis but exposure to flood water that spread the *Leptospira* bacteria in the slums can be reduced. To reduce the exposure, it is important to identify flood-prone areas and those in contact with the flood water. This study aimed to estimate building-level population in Alto do Cabrito, Salvador, identify flood-prone areas by simulating floods, and assess exposure of buildings and population to flood waters. Since leptospirosis is endemic in Salvador, an assumption that the exposure to flood is proportional to the exposure to *Leptospira* motivated the aim. To disaggregate the census-sector population of Alto do Cabrito to building-level population, a 3D dasymetric method that utilized building footprints derived from aerial images and building heights derived from digital terrain models was used. To simulate the flood water depths for 1-in-2-year, 1-in-5-year, and 1-in-10-year design storms, a 2D hydraulic modelling was conducted. To visualize flooding, 3D models of the water depths during and at the end of a 24-hr rainfall event were produced. To inform the wet periods that influence *Leptospira* bacterial growth and leptospirosis incidence, PERSIANN CDR, a satellite rainfall product, was analysed for seasonality of rain. To assess the exposure of buildings and people to flood, the number and percentage of buildings and population in close proximity to pooled water was estimated. To visualize the exposure, water depths juxtaposed with buildings and building-level population normalized by building heights were produced. Additionally, in a small subset of the population for which leptospirosis sero-survey data was available for Alto do Cabrito, an association between the sero-status and the exposure to pooled water was assessed. The results of the population disaggregation showed that the numerical measure used to summarize the building height, which is used to estimate the building volume and the population, as well as the rounding method to convert the estimated population to an integer approximation affected the precision of the disaggregated population. For Alto do Cabrito, the maximum of pixel values as the numerical measure of the building height with the population rounded to the nearest integer resulted in the 2010 building-level population with a low error. Consequently, this building-level population was used for exposure analysis in this study. According to the PERSIANN rainfall data, the rainfall varied annually. The months of April to June comprised the wet season but PERSIANN underestimated the rainfall. The flood simulations showed that high intensity rainfalls caused high rainfall-runoffs. At the end of the 24-hr rainfall events, most of the rainwater flowed away from the study area to the lowlands. The remaining water pooled along roads and around buildings. During the rainfall events, the water rose to the levels that inundated buildings of 6-7m and less in height. Since the water pooled across the study area, the use of proximity thresholds to assess the exposure showed that almost all of the study area was exposed to flood. The exposure to flood waters was likely to be more in small buildings housing a larger number of people. However, the small Odds Ratios (ORs) and the 95% confidence intervals (CIs) that contained the null value showed uncertainty in the association between leptospirosis and floods. In conclusion, the flood simulations aided to identify areas where water pooled after heavy 24-hr rainfall events. With the building-level population, a better estimate of people exposed to flood could be made. The 3D visualization of the flood was helpful to understand the exposure to pooled water during and after the heavy rainfalls. The association between leptospirosis and flood was uncertain. The uncertainty in the association is attributed to the study design, exposure misclassification, and various risk factors that were not considered in this study.

ACKNOWLEDGEMENTS

I would like to acknowledge my supervisors, Mila Koeva and Monika Kuffer, and my advisors, Patricia Brito and Julio Pedrassoli, for their helpful guidance and constructive feedbacks on my research, and especially, for believing in me and giving me the freedom to construct my own interdisciplinary research question.

I thank Anderson Freitas and Adriana Tavares from Politechnic School of Federal University of Bahia (UFBA) for helping me acquire data from Brazil and providing explanations of data acquisition methods.

I would express my sincere thanks to Tom Rientjes and Gabriel Parodi for guiding me through hydrology concepts and flood modelling. I would also like to recognize Zhoobin Rahimi, a hydrologist and a graduate of ITC, for introducing features of Hec-RAS that I could use for my research.

I thank Anurag Kulshrestha, a PhD candidate at ITC, for guiding me through image processing and use of radar in flood detection and mapping.

I would also like to express my thanks to my friends and colleagues at ITC for their help and friendship throughout my MSc journey. And, special thanks to those who accompanied me on long evening walks during the COVID lockdown period.

I reserve my special thanks for Andre Da Silva Mano for being a wonderful teacher and a mentor, without whom I wouldn't be at ITC. I am very grateful for his help in guiding me through the MSc admission process and in transitioning from a part-time distance course learner to a full-time student at ITC.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Motivation

Sustainable Development Goal 3 (SDG 3) aims to "ensure healthy lives and promote well-being for all at all ages" (United Nations, 2015). To achieve the goal by 2030, countries around the world have increased spending in public health. Yet, for infectious diseases, such as HIV/AIDS, tuberculosis and malaria, there have been mixed results in terms of incidence, prevalence and per-capita spending (Micah et al., 2020). For a group of neglected tropical diseases (NTDs) identified by the World Health Organization (WHO), there is inequity in global health financing (Addisu et al., 2019). Yet still for emerging infectious diseases (EIDs), such as leptospirosis, in addition to poor allocation of resources to tackle the diseases (Jones et al., 2008), there is lack of surveillance, intervention, and research into health determinants and risk factors (Costa et al., 2015). Hence, there is a need to fill in the gaps in knowledge associated with EIDs.

The gaps in knowledge of the health determinants and risk factors associated with infectious diseases is important to consider. So far, the global health research and financing for both infectious and non-infectious diseases have focused on medicines, and structural and behavioural changes to prevent and control diseases. While addressing health system constraints (Micah et al., 2020), which are distal or indirect factors, is important, it is also necessary to understand both the proximal (those which cause immediate effects) and the interconnected factors (Cohen, Wilson, & Aiello, 2007), and intervene where and when necessary. This is important for diseases like leptospirosis for which incidence is dependent on structural, behavioural, socio-economic, occupational as well as environmental factors (WHO, 2011), but which have not been thoroughly investigated.

The gap in knowledge and research on infectious diseases can be effectively filled by applying tools and methods that have been developed outside of the public health domain, with the benefit of introducing novel frameworks and methods to investigate health issues. To date, few studies have investigated the environmental and social risk factors of infectious diseases, including NTDs, using geographic information system and earth observation (Hamm, Soares Magalhães, & Clements, 2015; Linard & Tatem, 2012; Souza, Uberti, & Tassinari, 2020) though uses of satellite images, UAV images, LiDAR data and 3D modelling tools show big potentials to uncover knowledge about the risk factors and exposure. This holds true for EIDs as well, such as leptospirosis (Ledien et al., 2017; Malloy, Horack, Lee, & Newton, 2019; Parselia et al., 2019; Skouloudis & Rickerby, 2015). Therefore, this research aims to integrate 2D and 3D information to better understand the risk factors and exposure related to leptospirosis.

1.2. Background

1.2.1. Leptospirosis: An Emerging Infectious Disease

Leptospirosis is an EID. Emerging pathogens are those which have recently evolved in strain, entered the human population, or were historically present in the human population but have recently increased in incidence (Jones et al., 2008). According to Jones et al. (2008), the EIDs are dominated by zoonoses, accounting for 60.3% of EIDs. Zoonoses are diseases that are transmitted from animals to humans.

Leptospirosis is an example of a zoonotic disease, with transmission of *Leptospira* bacteria to humans occurring from pigs, cattle, horses, dogs, sheep, racoons, rats, mice, marsupials and bats (Bharti et al., 2003). In urban areas, rats are the major animal reservoirs for *Leptospira*. Transmission usually occurs when humans are exposed directly to the urine of infected carriers or indirectly through contaminated soil and water. Routes of exposure are ingestion, inhalation, dermal contact, and penetration (Hartskeerl, Collares-Pereira, & Ellis, 2011). Once infected, humans show signs and symptoms ranging from headache, chills, cough, muscle pain, fever, nausea/vomiting, diarrhoea, jaundice, enlarged liver, renal failure to pulmonary haemorrhage (Bharti et al., 2003). On the other hand, the non-human carriers of *Leptospira* are usually asymptomatic. Some infected people also remain asymptomatic, but transmission through them is considered unlikely. It is difficult to eradicate Leptospirosis, as it has many natural reservoirs (Hartskeerl et al., 2011). Hence, the focus should be on identification of vulnerable groups, risk factors, and prevention of exposure.

The incidence and prevalence of leptospirosis in human population are influenced by socio-economic, environmental, occupational and ecological factors as well as recreational activities that involve immersion in water (Bharti et al., 2003; Jones et al., 2008). The disease is prevalent in tropical and sub-tropical regions. However, the incidence rates of leptospirosis is underreported due to misdiagnosis – undifferentiated fever is misdiagnosed as malaria or dengue (De Francesco Daher et al., 2017), inadequate rapid diagnostics, lack of access to laboratory testing, and insufficient public awareness (Bharti et al., 2003; Hartskeerl et al., 2011). While the disease is emerging in both industrialized and developing countries, the slum populations are thought to be more vulnerable to the leptospirosis infection due to the poor living conditions (Khalil et al., 2021). Moreover, in informal settlements, estimation of the risk of leptospirosis is hindered by lack of up-to-date demographic data and information on risk factors at local level. This has, thus, hindered the identification of vulnerabilities and prevention of exposure as well as early diagnosis in patients who come from high-risk zones.

1.2.2. Slums and Leptospirosis

Traditionally, leptospirosis was considered an occupational and rural disease, as the infection was found in people who came in contact with domestic or wild animals (Bharti et al., 2003). However, the disease is now found in urban and suburban areas, with slums being disproportionately affected (Hartskeerl et al., 2011; Ko, Reis, Dourado, Johnson, & Riley, 1999) due to poor socioeconomic and environmental conditions (National Academies of Sciences & Medicine, 2018). The slums also have disproportionate burden of other infectious diseases, including but not limited to meningitis, hepatitis, multidrug-resistant TB, sepsis, influenza, HIV, and sexually transmitted diseases. Thus, to understand the burden of leptospirosis, there is a need to thoroughly investigate the different conditions and co-morbidities in the slums.

1.2.3. Determinants and Risk Factors of Leptospirosis: What is Known

According to Baquero and Machado (2018) study that examined spatio-temporal dynamics and risk factors of human leptospirosis in Brazil, increased soil moisture, precipitation, poverty and decreased proportion of urban households were associated with leptospirosis. However, contrary to some studies (López et al., 2019), the authors found that temperature had a preventive effect, hypothesizing that high temperature reduced the soil moisture and thereby impeded the growth of *Leptospira* bacteria. It is possible that rainfall and temperature have varying effects depending on the region and scale of study area (Chadsuthi, Modchang, Lenbury, Iamsirithaworn, & Triampo, 2012) as well as inclination (Barcellos & Sabroza, 2001). These factors are necessary to consider, as studies examining the effect of climate change on leptospirosis (Cucchi et al., 2019) seem to ascertain that increased precipitation leads to increased soil moisture which is associated with increased leptospirosis cases that are observed with time lag (Bierque, Thibeaux, Girault, Soupé-Gilbert, & Goarant, 2020; Cucchi et al., 2019). But this remains to be examined in conjunction with temperature, which is affected by climate change too as well as inclinations and seasons. Furthermore, as a critique to Baquero and Machado (2019) study, the decreased proportion of urban households could imply either sub-urban/rural areas or open spaces that form habitats for rodents and other hosts of *Leptospira*, and hence, require further nuanced investigation. In a separate but related study conducted in the slum community of Pau da Lima, which is on the outskirts of Salvador, Brazil, Reis et al. (2008) found a strong association between acquiring *Leptospira* antibodies and household environmental factors like residence in flood-prone region with open sewers, proximity to accumulated refuge and rat sightings. In a more recent study in Pau da Lima, N. J. Santos, Sousa, Reis, Ko, and Costa (2017) showed that the increased rodent infestation is associated with households that were less than 25 meters from open sewers or located at the lowest point of the valley, with distance to open sewers highly correlated with household elevation. In another study in Salvador, Brazil, Felzemburgh et al. (2014) found that distance to open sewers is significantly related to secondary infection, that is, reinfection through repeated exposure to pathogenic *Leptospira* bacteria. The association with rodent infestation, contact with mud, and lower household elevation is also reported in Hagan et al. (2016) study of leptospirosis in slums of Salvador, Brazil. The N. J. Santos et al. (2017) study in Pau da Lima, also found that dilapidated fences and walls, used as indicators of environmental condition, were highly correlated with rodent infestation. But in a separate study of the spatial distribution of leptospirosis in Sao Paulo, Brazil, the effect of sloped terrain on the locations of leptospirosis cases was not significant (Ferreira, 2016). The authors explain that higher case rates were found in census sectors of higher river density and in households closer to rivers, thus further motivating research into the role of space and elevation to explain their effect on leptospirosis prevalence. In another study, Gracie, Barcellos, Magalhães, Souza-Santos, and Barrocas (2014) found that the incidence of leptospirosis is dependent on the geographical scale of analysis. At the regional scale, the incidence correlated with slum population. At the municipal level, there was no significant correlation, but at the local level, the percent of areas prone to flooding was significantly correlated with leptospirosis incidence (Gracie et al., 2014). The variation in leptospirosis prevalence across small spatial scale is also mentioned by Hagan et al. (2016). Flooding as being an important risk factor for increased incidence of leptospirosis is stated in other studies as well (Lau, Smythe, Craig, & Weinstein, 2010; Mohd Radi et al., 2018; Vanasco et al., 2008).

In summary, studies in Brazil, together with those in other countries (Briskin et al., 2019; Calderón-Sierra, Jaimes-Bernal, & Pedraza-Bernal, 2019; Pandji Wibawa Dhewantara et al., 2020; Garba et al., 2018; Halliday et al., 2013; Kawaguchi et al., 2008; Smith, Young, Wilson, & Craig, 2013), clearly establish that poor households and slum neighbourhoods are vulnerable to leptospirosis. Important risk factors are rodent infestation, distance to canals/sewers and flooding, and several other risk factors, as mentioned above, that need further investigation.

1.2.4. 3D Modelling

3D modelling, which utilizes data on the third dimension, adds value in various applications (Filip Biljecki, Stoter, Ledoux, Zlatanova, & Çöltekin, 2015) compared to 2D modelling. By using a 3D city model of the Netherlands, F. Biljecki, Ohori, Ledoux, Peters, and Stoter (2016) show that using building height information in a population model provides a good approximation of the population in a region. The authors argue that 3D city models add value in population estimates by forgoing expensive and time-consuming field surveys and are useful for countries where census surveys are not carried out for long periods of time (F. Biljecki et al., 2016). Similarly, in a case study assessing the application of 3D modelling in public health and environmental planning, Maroko, Maantay, Pérez Machado, and Barrozo (2019) examine the population distribution and exposure to air pollution in 3D in New York and Sao Paulo, Brazil, two cities with varying degrees of data availability. Based on the results, the authors argue that, compared to traditional methods, 3D modelling provides better estimates of the population at risk and exposure, both disaggregated in three dimensions.

Maroko et al. (2019) also claim that 3D disaggregation is useful to assess population exposure to noise pollution, urban heat island effect, and extreme heat events – there are numerous publications on the use of 3D data and models for analysing pollution and heat effect (Kumar, Ledoux, Commandeur, & Stoter, 2017; Kurakula et al., 2007; Lánský, Ceccarelli, Mastorakis, Jongh, & Li, 2019; Ujang, Azri, Zahir, Rahman, & Choon, 2018; Wästberg, Tornberg, Billger, Haeger-Eugensson, & Sjr berg, 2013; Zahran, Smith, & Bennett, 2013) – as well as to other adverse events in areas with multistorey buildings, and where population at risk is expected to vary vertically (Maroko et al., 2019).

The other adverse events could be related to safety, sanitation and health in urban areas (Shih, Chang, & Popper, 2018). For example, in Japan and Taiwan, urban regulation for home construction, and broadly property development, was largely driven by a need to prevent exposure to infectious diseases like malaria and cholera that are related to unsanitary and congested environments. As such, the urban regulation took into consideration ground floor elevation, building height, building material, minimum ceiling-to-floor height, and hours of exposure to sunlight (Shih et al., 2018). Thus, with such urban planning, exposure to and risk of infectious diseases associated with sanitation and building conditions is expected to vary vertically. This is observed in the case of tuberculosis (TB), another infectious disease spread person-to-person through inhalation of air containing TB bacteria. The number of TB cases varied vertically according to a study that examined the risk of TB in high-rise and high density dwellings (Lai et al., 2013). The study concluded that people living in lower floors had higher risk of TB than those living in higher floors. The result was ascribed to more sun exposure in higher floors than in lower floors. However, the paper cautions that the observed risk could also be due to differences in socio-economic status of residents, where those living in higher floors are wealthier and have better nutrition. In their study, the sun exposure was measured using sky view factor (SKV), which was measured using 3D urban raster and vector databases.

Another adverse event that shows vertical variation is flooding. Mioc et al. (2011) state that, compared to 2D maps with descriptive statistics, 3D visualization of flood provides a more informative and realistic outcome during floods and is therefore more helpful in making decisions about when buildings and infrastructures are unsafe for use. The added value of 3D model output in flood risk assessment and communication is also discussed in other papers (Kolbe, Gröger, & Plümer, 2005; Kumar, Ledoux, & Stoter, 2018; Van Ackere, Glas, Beullens, et al., 2016; Van Ackere, Glas, Vandenbulcke, et al., 2016). While all these papers emphasize the need for 3D modelling and visualization, the approaches adopted for the studies differ. The Mioc et al. (2011) study is based on LiDAR data that was collected for flood monitoring and was combined with hydrological modelling. Using the data, the authors extracted Digital Elevation Model (DEM) and building footprints, modelled buildings, and utility infrastructures in 3D, added attribute data, and intersected 3D models with the terrain model. With different flood scenarios to model different flood levels, the authors showed where and when the buildings and infrastructure would be inundated. The authors used CityGML and ArcGIS to accomplish this. In contrast, (Van Ackere, Glas, Beullens, et al., 2016; Van Ackere, Glas, Vandenbulcke, et al., 2016) produced a web-based 3D and 4D (with time as the 4th dimension) dynamic flood visualization tool using Ol3-CesiumOpenlayers, whereas Kumar et al. (2018) integrated a 3D city model represented by CityJSON with Cesium 3D webglobe.

While the existing studies show a potential for application of 3D modelling in public health scenarios, the use of 3D modelling is currently in a nascent stage in public health research. As such, the advantages it offers in measuring and visualizing hazards, exposure and risk are yet to be explored and documented. Specifically, in the context of leptospirosis, factors that vary vertically but which are difficult and expensive to capture through survey methods or use of 2D data alone, such as population number and density, height of the building, flooding, mud mobilization and contamination of water, and elevation and inclination of the ground can be investigated through 3D modelling.

1.3. Research Gap

The understanding of the dynamics of leptospirosis is hindered by lack of data on demographics and its health determinants and risk factors, especially in informal settlements that are stigmatized and marginalized from governmental policies and interventions (Addisu et al., 2019). To quantify the risk of leptospirosis as well as its changing trend, demographic data is vital. However, in many countries, census data are updated once every decade and do not provide a precise estimate of the underlying population at risk (Patricia Brito et al., 2020; Kuffer, Persello, Pfeffer, Sliuzas, & Rao, 2019). While changing climate conditions, such as rainfall, flooding and warm temperature are attributed to epidemics of leptospirosis (Abela-Ridder, Sikkema, & Hartskeerl, 2010; Alderman, Turner, & Tong, 2012; Ansdell, 2017), how these would affect slum neighbourhoods are lacking from the literature. Furthermore, within the informal settlement, information is missing on the population at high risk of leptospirosis when the area is flooded. To understand the population at risk, delineating areas at risk of flooding and where direct contact with rodents and contaminated soil and water is likely is needed.

With regards to the use of data and tools, there are studies that use GIS and earth observation data and tools (PL Brito, 2010; Cucchi et al., 2019; P. W. Dhewantara et al., 2019; Pandji Wibawa Dhewantara et al., 2020; Ledien et al., 2017) to examine a restricted number of structural, environmental and social risk factors associated with Leptospirosis. However, these are limited to 2D and do not inform risk at the local level. Over these methods, 3D modelling has clear advantages in providing estimates of population at risk (F. Biljecki et al., 2016), visualizing simulated floods, assessing exposures, and quantifying and visualizing risks (Filip Biljecki et al., 2015; Kolbe et al., 2005) in areas and neighbourhoods of interest. Although application of 3D modelling to answer questions related to leptospirosis is not found in the literature, different factors, especially population can be modelled using 3D data and exposure of the population to flood can be visualized in 3D. Furthermore, Maroko et al. (2019) found that 3D modelling was useful to provide good approximation of population in Sao Paulo, Brazil even though data availability was not on par with cities such as New York. This facilitated better exposure assessment of air pollution. The usefulness of 3D modelling in population estimation and informing exposure assessment remains to be substantiated with more investigations conducted in different areas with different data availability.

1.4. Research Objectives and Questions

This research aims to investigate exposure of Alto do Cabrito and its population to floods that are attributed to epidemics of leptospirosis by integrating 2D and 3D data.

1.4.1. Sub-objectives

S1. Review methods that use 2D and 3D data to disaggregate population

S2. Review methods to model flood and conduct 3D visualization

S3. Disaggregate population at building-level using 2D and 3D data

S4. Examine seasonality of rainfall and rainfall volume in the study area

S5. Determine flood extents and depths resulting from different design storms

S6. Estimate the number of buildings and population in contact with pooled water

S7. Evaluate contact of leptospirosis positive and negative cases with pooled water

1.4.2. Research questions

S1. Review methods that use 2D and 3D data to disaggregate population

1. What methods currently exist to model population?
2. What type of input data do the models need to disaggregate population?

S2. Review methods to model flood and conduct 3D visualization

1. What methods currently exist to model and visualize flood?
2. What type of input data does flood modelling and visualization need?

S3. Disaggregate population at building-level

1. Based on literature review and available data for the study area, what is the most suitable method that integrates 3D data to model population?
2. What would be the limitations to disaggregate projected population at building-level?

S4. Examine seasonality of rainfall and rainfall volume in the study area

1. How does the rainfall differ by season in the study area?
2. How does the rainfall differ by volume in the study area?

S5. Determine flood depths resulting from different design storms

1. What is the water depth at the end of 24-hr rainfall?
2. What is the maximum depth during the 24-hr rainfall?

S6. Estimate the number of buildings and population in contact with pooled water

1. What number of buildings are in contact with pooled water?
2. What number of people are in contact with pooled water?

S7. Evaluate contact of leptospirosis positive and negative cases with pooled water

1. How many people are diseased and exposed to pooled water?
2. How many people are not diseased and exposed to pooled water?

1.5. Scientific Significance

While traditional environmental epidemiology has been useful in studying the risk of leptospirosis, there has always been an uncertainty in estimating the population at risk that is used in quantifying the exposure and risk of leptospirosis at the local level. And, while 2D spatial information have been used to estimate the spatial distribution of hazards and populations, use of 3D data is scarce. Over these methods, 3D modelling has advantages in disaggregating population and visualizing hazards, such as floods.

1.6. Thesis Outline

This thesis comprises of seven chapters and an Appendix. Chapter one introduces the motivation and the background to the research problem, research gaps, and research objectives and questions. Chapter two provides literature review of methods for population disaggregation and flood modelling using 2D and 3D data as well as a review of exposure assessment methods. Chapter three provides information about the study area and the methodology used to achieve the research sub-objectives S3-S7. Chapter four describes the results as per the research sub-objectives. Chapter five discusses the results and their implications. Chapter 6 provides conclusions based on the results and the discussions. Chapter seven discusses the limitations of the research and the recommendations for future work. Lastly, the Appendix provides auxiliary analyses related to the research.

2. LITERATURE REVIEW

2.1. Population Modelling

Up-to-date and accurate population data at the local level is necessary for estimating the population at risk of disease, exposure, and health related policy and decision making. The main sources of population data are typically the national census that is usually conducted every 10 years, and registries for births and deaths which are used to update the census. However, the size of the enumeration and temporal resolution of the national census is not always useful for public health inquiry. Also, in resource poor settings, birth and death registers may not exist or are poorly kept. This sets the need to model and forecast population at different spatial and temporal scales (Eichhorn, 2020; Silva, Guerrero, & Peña, 2011; Wardrop et al., 2018). The research into population disaggregation at spatial and temporal scale is vast and an exhaustive review is not intended for this research. Rather, an overview of common methods is provided below.

An accepted conceptualization to disaggregate population treats population as continuously varying surface whose value can be measured at any location (Mennis, 2003). To create the population surface, areal interpolation and dasymetric methods are used. In the process, the pycnophylactic property, according to which people are neither created nor destroyed during transformation, is preserved.

Areal interpolation transfers data from one set of spatial units to another (Eicher & Brewer, 2001). The common techniques in areal interpolation include areal weighting, inverse distance weighting (IDW) and regression methods (Eicher & Brewer, 2001; Langford, 2006; Mennis, 2003). In areal weighting, the area of the target areal unit is divided by the total area of the host (also termed source) areal unit that is then applied to the population of the host areal unit. The resultant estimate is the population estimate of the target area (Langford, 2006; Mennis, 2003). For raster data, the target areal units are grid cells, and for vector data, the target units are polygonal units in the host area (Mennis, 2003). IDW assumes that the population densities of close by areas are similar than those of areas that are farther away and that the influence of the population density of an area on the other decreases as per some distance-decay function. As such, in IDW, a moving window is applied to search for areas neighbouring the target area. Weights are calculated based on the distance of the neighbouring area to the target area. These weights are then applied to the population of the neighbouring areas and an average is calculated. The weighted average is then assigned as the population estimate of the target area (Mennis, 2003). Regression methods use information about the study area to estimate the population (Eicher & Brewer, 2001). The procedure involves determining a statistical relationship between population (independent variable) and dependent variables (or covariates) (Langford, 2006). However, ordinary regression methods do not capture spatial autocorrelation and geographical variation in the relationship between the population and covariates. As such, researchers use spatial statistical methods, such as cokriging and geographically weighted regression to interpolate population (Sridharan & Qiu, 2013).

Dasymetric mapping is a kind of areal interpolation that uses ancillary data, such as inhabited and uninhabited areas, land cover data, and settlement patterns (Eicher & Brewer, 2001; Eichhorn, 2020; Langford, 2006; Mennis, 2003). Similarly, areal interpolation can also use ancillary data, such as land cover, so the distinction between the two methods may appear subtle (Sridharan & Qiu, 2013). Unlike areal interpolation, however, the dasymetric approach does not reaggregate the data to a preferred enumeration unit (Eicher & Brewer, 2001). The dasymetric approach involves using the additional data to sub-divide an area into smaller, relatively homogenous spatial units and distribute the population into smaller units using weights. The boundaries of the smaller units are abrupt unlike that for enumeration area. To create a continuous population surface, such as a population raster, the population of the smallest area in consideration is divided evenly among the grid cells that make up the area (Eichhorn, 2020; Mennis, 2003). Dasymetric mapping that use 2D data generally works fine in homogeneous regions, such as the suburbs, but does not work well in heterogeneous regions that are comprised of various types of buildings (Zhenyu Lu, Im, & Quackenbush, 2011). The problem concerns underestimation of the population in areas with high-rise buildings and overestimation in areas with low-rise buildings (Sridharan & Qiu, 2013). To overcome this, researchers use 3D data.

With the availability of high-resolution images and LiDAR point clouds, several researchers have used 3D data for population mapping (F. Biljecki et al., 2016; Zhenyu Lu et al., 2011; Wang, Tian, Zhou, Liu, & Lin, 2016). The procedure involves extraction of building footprints and heights from the images or the point clouds, followed by 3D building reconstruction of residential buildings and fine scale population estimation. Depending on the data, the extraction of building footprints and heights differ. For example, Wang et al. (2016) used an object-based method (morphological operation) to extract the buildings from images and shadow post processing to extract building height. Zhenyu Lu et al. (2011) used a modified morphological algorithm on LiDAR derived surfaces to delineate building polygons, land parcel polygons to identify residential building polygons, and LiDAR derived height, which was estimated by subtracting bare earth surface (DEM) from first and last return surfaces. The authors considered last return height to take into account occlusions from trees. F. Biljecki et al. (2016) used building data from the national register of addresses and buildings (BAG – Basisregistraties Adressen en Gebouwen) in the Netherlands and elevation data from Actueel Hoogtebestand Nederland (AHN).

Following building reconstruction, the population is estimated using either a mathematical or statistical relationship between floor space, building height, living area, and height per floor, depending on data availability (F. Biljecki et al., 2016; Wang et al., 2016). But, despite providing more details on 3D reconstruction than other studies, F. Biljecki et al. (2016) fail to describe and discuss the statistical model they used. Thus, their work, while useful to understand the application of 3D data, cannot be considered reproducible for population estimation. More complex population estimations regress target population against other covariates, such as total area and volume of buildings, category of single or multiple family housing, and land cover (Alahmadi, Atkinson, & Martin, 2013; Zhenyu Lu et al., 2011). In their study, Zhenyu Lu et al. (2011) compared population estimates from volume-based and area-based regression models and found that volume-based models out-performed area based models. In their previous publication, Z. Lu, Im, Quackenbush, and Halligan (2010) had found that area-based models outperformed volume-based models in homogeneous study areas that mostly consisted of single family houses and where the volume-based approach was sensitive to classification errors, such as due to tall trees. Whichever method is used, the procedure should conserve both volume and population and the accuracy analyses are usually conducted on building detection, height retrieval, and population estimation (F. Biljecki et al., 2016; Wang et al., 2016).

Building upon the use of 3D data, Sridharan and Qiu (2013) and Maroko et al. (2019) used the dasymetric method to disaggregate population. Sridharan and Qiu (2013) estimated the population at building-level by interpolation, where the volume of each building was multiplied by a factor, which was the total population in the source area divided by the sum of the volume of the buildings in the source area. Following this, the researchers fine-tuned the building-level population by using regression parameters obtained by regressing the building-level population against building volumes. To maintain the pycnophylactic property, the researchers applied iterative adjustment to redistribute errors of estimation. Using the adjusted building-level population, source-level population was estimated and compared against the true source-level population. Discrepancies between the estimated and true population was calculated and applied to each building to get the final adjusted estimate of the building-level population. Maroko et al. (2019) used tax lot level total residential area and total building volume to estimate residential area for each building in the tax lot. Then, the researchers estimated building-level population by multiplying total census block level population by a ratio of a building's residential area to total residential area at the block level. Contrary to the claims made by Maroko et al. (2019), neither Maroko et al. (2019) nor Sridharan and Qiu (2013) disaggregated population vertically. Both studies used the dasymetric method to estimate the population at building-level, but did not further disaggregate that population vertically, such as by the number of floors in a building.

The methods reviewed above are classified as top-down approaches, where the population of small areas are estimated by disaggregating census-level population. However, due to reasons of conflict and political instability, national census figures may not exist. In such cases, a bottom-up approach is used to estimate the population for small areas which can be aggregated to administrative and national levels. Conceptually, in a bottom-up approach, a micro-census – a complete count of the population – is taken of a small area and statistical methods are used to estimate the population in other areas. Similar to the top-down approach, the bottom-up methods also utilize ancillary data, such as socio-economic data, distance to road, elevation, slope, as well as remotely sensed data, such as vegetation index, surface temperature, and night time light (Wardrop et al., 2018).

2.1.1. Tools Integrating Population Models and 3D visualization

With the advent of mapping technology, population mapping is often achieved through GIS (Eicher & Brewer, 2001; Wardrop et al., 2018) Among the reviewed articles, F. Biljecki et al. (2016) mention using 3dfier to create a 3D city model using which population is estimated. Though Maroko et al. (2019) provide 3D visualization of both the buildings and population, they do not mention the type of software and tools used to create the visualizations in their article.

2.1.2. Population Modelling Data Requirement

From the review provided above, population modelling utilizes national census, birth and death registries, survey data, and various remotely sensed products of which land cover data is a common one. The ancillary data, such as the land cover, can be in vector or raster data types. The type of data acquired and used differ by top-down or bottom-up approaches for population estimation.

2.2. Flood Modelling

Here, a distinction is made between fluvial and pluvial flooding. Fluvial floods, also called riverine floods, are localized in the alveolar part of water bodies. They occur when excessive rainfall causes a channel to overflow. The excess water affects operations and eventually breaches dams and dikes (Rubinato et al., 2019). Pluvial floods also occur due to excessive rainfall but are independent of overflowing channels of water. Pluvial floods can be separated into surface water floods (SWFs) that occur when natural and/or man-made drainage systems are overwhelmed leading to ponded water and overland flow, and flash floods that occur when there is a sudden release of water, which reach downstream at erosive speeds in a relatively short time (Rubinato et al., 2019). In SWFs, small scale features, such as heights of kerbs, dimensions of road cambers, as well as road and pathway networks control the water flow and ultimately determine areas that are flooded (de Almeida, Bates, & Ozdemir, 2018). The focus of this research is on surface water flooding due to intense rainfall over Alto do Cabrito, Salvador, Brazil, an area that is mostly built-up.

Different hydraulic models are used to model flow and transport processes along water channels and floodplains. Specifically, one dimensional (1-D), two dimensional (2-D) and 1D and 2D coupled hydraulic modelling that partially or totally solve St. Venant equations are used to model floods, depending on whether the flow is in the channel, floodplain, or channel-floodplain interface (Bates & De Roo, 2000; Criado, Martínez-Graña, San Román, & Santos-Francés, 2019; de Almeida et al., 2018). In this regard, 1-D hydraulic modelling entails selection of several discrete cross-sections along a water channel. A 1D solution of the St. Venant equations results in an average water surface elevation (WSE) and velocity at the discrete cross-sections, where speed is considered a variable only in the direction of the streamflow. 2-D modelling entails selection of a water-affected area that is modelled as a grid or mesh. 2D equations are then solved to compute WSE and velocity on each cell of the grid or mesh. The coupled model links 1D for channels with 2D for floodplains.

Simulating the flood system in the study area with a 1D model is possible. However, it means that all the streets need to be connected at their intersections with a node. The water balance of the node requires the solution of the momentum equation. This equation solves the node with a description of the external forces acting over the node, the gravity, the roughness, and the water mass inside the node. This is very common in river junctions, but on a rectangular street pattern this solution is extremely cumbersome and affected by overlapping of sections that are not allowed in numerical solutions. A 2D model simplifies the problem, as the solution does not differ from the solution of the St-Venant general computations at cell level. Hence, in this research, the 2D hydraulic modelling is of interest.

Both 2D and the coupled 1D and 2D models are computationally heavy (Maksimović et al., 2009). Previously, 1D hydraulic models were favoured for flood simulation and forecasting along a water channel, but now, with the availability of software utilizing different algorithms to solve the simplifications of the full St. Venant equations, both 2D and coupled hydraulic models are becoming more common in use. For pluvial floods, where above ground surface features need to be considered, a 2D hydraulic model can be used with modified terrain data. Even so, the current modelling techniques and model efficiencies are not on par with those for fluvial flooding (Chen, Evans, Djordjević, & Savić, 2012). Some of the methods used to incorporate features like buildings are: unstructured mesh, grid coarsening, using roof elevations in fine grids, and using ground elevations along with increased local roughness (Chen et al., 2012).

2.2.1. Methods and Tools Integrating Flood Models and 3D visualization

Costabile et al. (2021) had used Aquaveo's SMS (Surface-water Modelling System) to conduct 2D hydraulic modelling to simulate flooding in the old town of Cosenza, Italy, which lies along the Crati River. The authors then used SketchUp Make 2014 to produce realistic 3D visualization of flood by integrating results of 2D flood simulation with 3D views of the town that were obtained using 3D point cloud data from terrestrial laser scanner. Previously in 2018 and 2019, some of the same authors had attempted to justify the use of 3D representation of urban flood for risk communication purposes (De Santis, Macchione, Costabile, & Costanzo, 2018, 2019). In the 2018 paper, the authors compared building footprint extrusion in a GIS environment and texture-mapping in Blender™, suitable for neighbourhood level analysis, to use of Terrestrial Laser Scanner (TLS) data to produce 3D models of geometrically valid and realistic virtual environments with more information at the building-level (De Santis et al., 2018). In the 2019 paper, the authors promoted the use of Potree, which is an open source WebGL based point cloud renderer, for visualization of water surfaces in an urban flooding scenario (De Santis et al., 2019). Around the same time, Khoury et al. (2017) built an animated 3D flood visualization system using WebGL. As input, the author used image formats of outputs of a WCA 2D flood model. Both (De Santis et al., 2019; Khoury et al., 2017) approaches aimed towards producing an interactive visualization of flooding with the ability to visualize effects on individual features, such as buildings. Yet, there are other publications that have used Building Information Modelling (BIM) to integrate building-level information with flooding. BIM is a modelling process widely used in construction and architecture sectors to design buildings and infrastructure. Since BIM takes a building-centred approach, a key objective is to integrate flood model results with BIM to aid in disaster response. For instance, Sam Amirebrahimi, Rajabifard, Mendis, and Ngo (2015) used MIKE 21 to model a 1-in-100-year flood and exported the spatial distribution of depth and velocity as ESRI shapefiles. The authors then used an in-house tool to extract flood parameters into an XML file for integration with a BIM model. In another paper, Sam Amirebrahimi, Rajabifard, Mendis, and Ngo (2016) mention the suitability of MIKE, TUFLOW, and 3D Smoothed Particle Hydrodynamics for tasks that aim to integrate flood parameters with BIM for flood damage assessment. While BIM helps to provide needed estimates for emergency evacuation and damage assessment of a building, it does not provide information at the city level. To overcome this, researchers use a 3D city model developed using CityGML or CityJson, which allows them to represent, store and use attributes of different urban features (Elfouly & Labetski, 2020; Kumar et al., 2018; Lee, Park, Park, & Jang, 2016).

In the above cited literature, flood was modelled using different modelling software. If, however, modelling is not the core concern, but risk assessment and communication is, instead of water depth information from flood modelling software, researchers use flood polygons obtained from government and institutes working on hydrology and hydrological disasters. For example, Adda, Mioc, McGillivray, Morton, and Fraser (2010), utilized flood polygons to examine the flood risk to government infrastructure. Though the paper does not explicitly mention how the flood polygons were created, it is inferred from the references listed by the authors that the flood polygons are created in ArcGIS by comparing the water surface TIN with bare earth terrain elevations obtained from LiDAR data. Though scenario-based flooding is mentioned in this paper, there is no mention of flood modelling and how the water surface TIN was originally derived. To integrate flood polygons with 3D models of buildings and utilities, the authors used Google SketchUp. Herman, Russnak, and Řezník (2017) also used flood polygons from Prague Institute of Planning and Development (IPR) which were provided as shapefiles for various water levels (5, 20, and 100-year floods). The authors used the QGIS plug-in QGIS2threejs for visualization in a web browser.

2.2.2. Flood Modelling Data Requirement

Hydrological modelling entails solving the water balance over time so that the volume of water in store for any time interval can be estimated. Hydraulic modelling, on the other hand, uses additional information on the shapes and geometries through which the water flows. This enables evaluation of the dynamic variables – water depth, velocity, and energy - in time, at each calculation cell. Thus, in general, hydrological and hydraulic modelling require information on precipitation, temperature, evaporation, soil infiltration, soil moisture, water levels and discharges at gauging stations, as well as land use and topographic data, such as DEM and Digital Surface Model (DSM), and delineation of basins and sub-basins. However, the type of data acquired and used differ by the conceptualization and objective of flood modelling. For the validation of flood extent and in the absence of ground data, remotely sensed data, such as flood extent maps derived from SAR, have been used (Barneveld, Silander, Sane, & Malnes, 2008).

2.3. Exposure Assessment

Flood risk assessment characterize flood hazard in terms of extent, depth, velocity, and their spatial and temporal dynamics. Exposure analysis is used to identify elements at risk of flooding and vulnerability analysis is conducted to determine the potential of the elements to be adversely affected by flooding (EXCIMAP, 2007; Foudi, 2013). In this research, exposure assessment is conceptualized differently. Exposure is defined as contact with the flood and not merely as identification of elements at risk of flooding (Adgate, 2011). Hence, exposure assessment involves estimating the number of buildings and people in the study area who come in contact with the flood.

Using GIS, exposure can be measured by spatial coincidence and distance-based methods (Chakraborty & Maantay, 2011; Maroko et al., 2019). Spatial coincidence assumes that exposure occurs within a predefined geographic unit, such as an administrative unit, postal code, or census tract that contains the hazard. As such, the location of the hazard becomes the surrogate for exposure. Elements, such as building and people, are then considered exposed if they fall within the geographic unit. Distance-based methods measure proximity of elements to hazards. They assume exposure declines with distance from the hazard. As such, elements are considered unexposed beyond a threshold distance. Fixed distance buffers, variable distance buffers, network buffers, and continuous distance are examples of distance-based methods. Once spatial relationships between different sets of data are established, overlay operations are conducted to combine the information. A review of literature that assesses exposure to flooding is presented below.

With an aim to support disaster-related decision making for the city of Fredericton, New Brunswick, Canada, Adda et al. (2010) intersected a digital terrain model (DTM), flood polygons representing water depths, with 3D models of infrastructures, such as buildings and utilities, to produce realistic visualizations of affected infrastructures during different flooding scenarios and forecasted flood levels. For this purpose, the authors utilized LiDAR data for the area to extract DEM and building footprints. The latter was edited to correct planimetry errors by comparing to cadastral data. Thereon, the buildings and utilities were modelled in Google SketchUp. In a study of spatiotemporal multi-hazard exposure assessment in Austria, Fuchs, Keiler, and Zischg (2015) intersected a 1-in-100-year flood data from the Austrian Insurance Association and 1-in-30-, 100-, and 300-year water levels and flood polygons derived from hydrological models, details of which are not provided in the paper, with building and population data acquired from building and population registries kept by municipalities in Austria. The authors also used digital building footprints to test the accuracy of the building locations and to assign flood hazard information to the buildings. The exposure assessment assumed that flood damages a building even if only a few parts of it are in contact with the water. Hence, any building that intersected any flood polygons representing the four hazard scenarios were considered exposed. Following this, the number of exposed people for each building was calculated using the population data. Similarly, Röthlisberger, Zischg, and Keiler (2017), in their study to identify spatial clusters of flood exposure in Switzerland, intersected population data with building footprints and flood polygons. The building footprints were extracted from Topographic Landscape Model (TLM) and point-referenced population data from the federal Buildings and Dwelling Statistics (GWS). The communal flood hazard maps, with classes defined by a combination of intensity and probability of events, were acquired from the 26 Swiss cantons as well as complimentary flood data from an Aquaprotect dataset from the Federal Office for the Environment. The research considered a building exposed if it either overlapped with the Aquaprotect layer or the communal flood hazard maps. In a further assessment, a hotspot analysis was conducted on the normalized aggregation of exposed buildings and inhabitants to detect statistically significant clusters of flood exposure. In another study, Criado et al. (2019) overlayed a flood hazard map with land use and ground feature layers to estimate length of infrastructure and number of buildings affected during different flooding scenarios.

The above-mentioned methods are suitable for areal-scale analysis. At micro or individual-scale, where impact of flooding on infrastructures, such as buildings, need to be assessed on a case-by-case basis, integration of BIM that provides building-level semantic and geometric information with flood inundation models in GIS has been useful (Sam Amirebrahimi et al., 2015; Sam Amirebrahimi, Abbas Rajabifard, Priyan Mendis, & Tuan Ngo, 2016; Sam Amirebrahimi, Rajabifard, Mendis, Ngo, & Sabri, 2016; S. Amirebrahimi, Rajabifard, Sabri, & Mendis, 2016). As the current research focuses on the area-level and not the building-level, an in-depth review of methods integrating BIM and GIS is not provided here.

3. METHODOLOGY

3.1. Study Area

The study area is the neighbourhood of Alto do Cabrito in Salvador, Brazil which is shown in **Figure 1**.

The city of Salvador is in the State of Bahia. It is the fourth largest city in Brazil and the main economic powerhouse of the North-eastern region. The city has grown drastically over the last five decades, resulting in increased built-up area and decreased vegetation cover (Machado, Oliveira, & Lois-González, 2019). Geomorphologically, Salvador is prone to flooding and landslides, with the highest recorded number of disasters due to intense rainfall. On average, Salvador receives precipitation of 2123 mm per year, with intense rainfall and higher frequency of floods occurring between April and July (Brandão, Santos, & Carelli, 2016a, 2016b; A. P. P. d. Santos et al., 2016). The risk of flooding is made worse with the loss of vegetation cover, obstructions of drainage channels, floodplain occupation and soil sealing (Machado et al., 2019).

The study area Alto do Cabrito consists of a large slum community. In the slum neighbourhood, the houses are densely built with narrow footpaths. Houses lie in flood zones, making them prone to flooding. Further, the literature report poor physical and environmental conditions put the area at risk of various infectious diseases, such as leptospirosis, cholera, typhoid, and hepatitis (Patricia Brito et al., 2020; Machado et al., 2019). Unfortunately, there is lack of updated socio-economic and environmental data for Alto do Cabrito (Patricia Brito et al., 2020) that hinders examination of associations between different risk factors and the diseases.

While several studies, as cited in the Background section **1.2**, identify flooding as an important risk factor for leptospirosis in slum areas, there is, to-date, no study conducted in Alto do Cabrito, Salvador that has examined exposure to floods within the context of endemic leptospirosis. For these reasons, the selection of the study area for this research is appropriate and important.

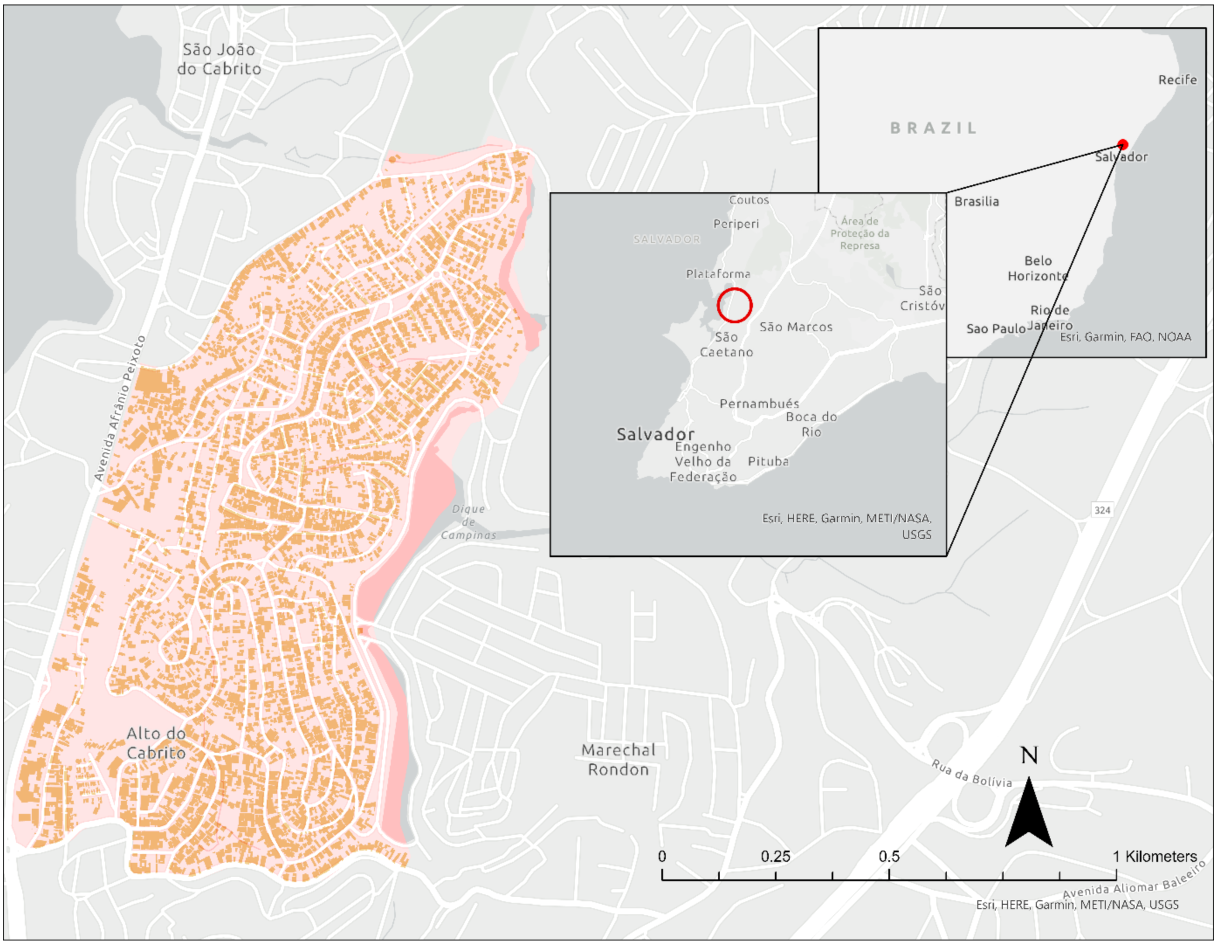


Figure 1. Map of the study area, Alto do Cabrito, Salvador, Brazil

3.2. Methodology Workflow

To achieve sub-objectives S1 and S2, literature review of the methods used to disaggregate population and model and visualize flood was conducted. The aim was to use the review to select suitable methods for population disaggregation (S3) and flood modelling (S5). **Table 1** presents the selection criteria used to select the methods to achieve sub-objectives S3 and S5. The selection was constrained by the available data.

The flow diagram below (**Figure 2**) presents a broad overview of the methodology used to achieve sub-objectives S3-S7. The details on the methods can be found in the following methodology sub-sections. The building footprints, DEM, DSM, lowland polygons have x, y locations and make up the 2D data. The height data derived from the DSM and DEM add the third dimension, z-value, and together with the x, y locations form the 3D data used in this research. Hence, essentially, 3D data is used to disaggregate population in the study area at the building-level. The overland flood is modelled using 2D numerical solution of St. Venant's equation utilizing the height data from the DSM. The resulting water height information from the flood model is visualized in 3D. The buildings and estimated building-level population are used in exposure assessment. Lastly, in a subset of population, the proximity of leptospirosis positive and negative cases to flood is examined.

Table 1. Selection criteria for population and flood modelling methods

Objective	Available Data	Selection Criteria	Selected Method
Population disaggregation	- Census-sector population - Building footprints - DSM and DEM	- Method should be able to disaggregate census-sector level population - Method should utilize 3D data in the form of terrain model or LiDAR point clouds - Method should not require auxiliary data from field survey or mapping for data processing - Method can be quickly reproduced for other areas	A 3D dasymetric method that uses census-level population and building volume to disaggregate population.
Flood modelling	- DSM and DEM	- Method should be amenable for use in urban flood simulation - Method should allow flow in various directions and paths - Method should utilize terrain data - Method should not require immediate field surveys to characterize study area conditions	2D hydraulic modelling

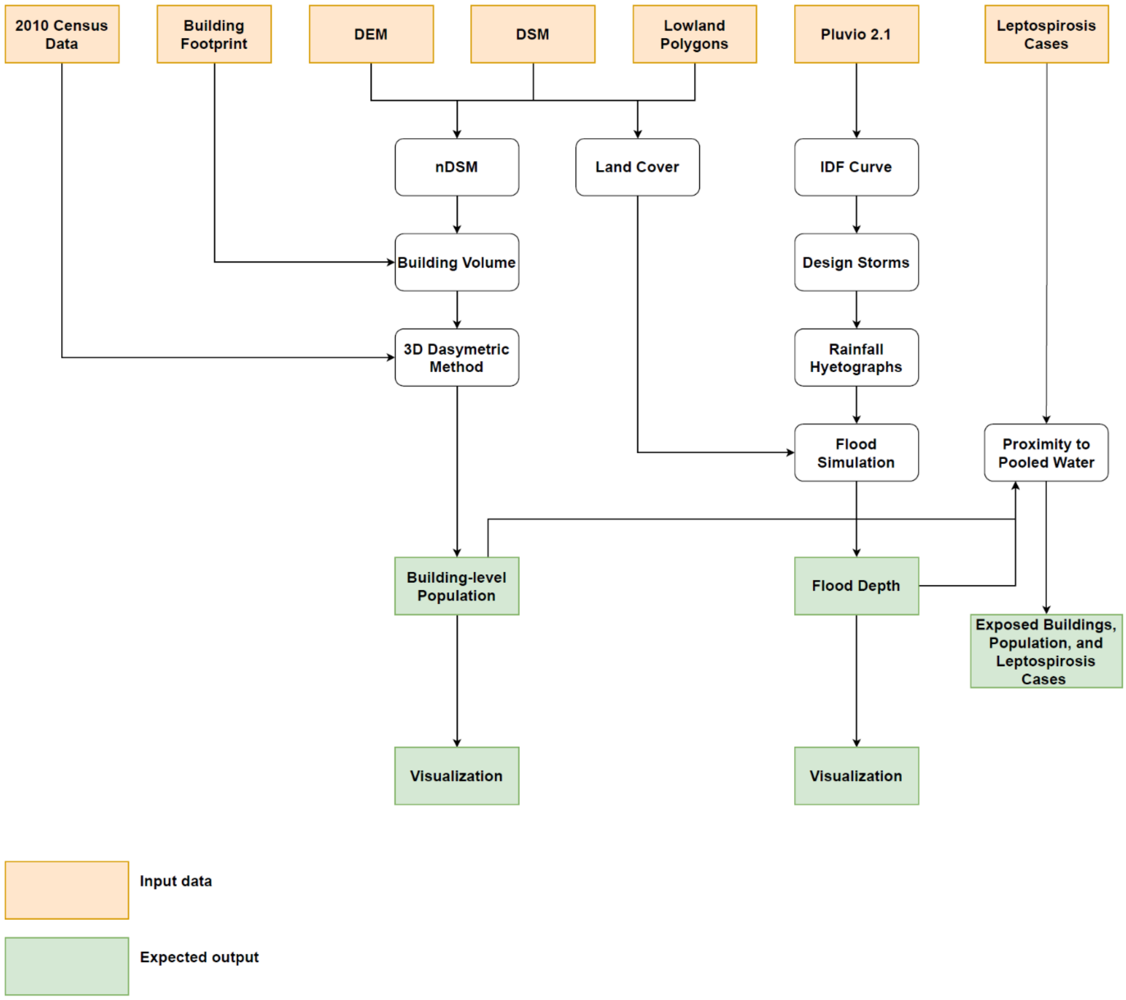


Figure 2. A general workflow for sub-objectives S3-S7

3.3. Data and Software Applications

Table 2 presents the data used in this study, the location for which the data is available, and their respective sources. The citations for the data are provided next to their descriptions in the following sections.

Table 2. Data used in the study and their sources

Data	Location	Sources
2010 census-sector population	Alto do Cabrito	Brazilian Institute of Geography and Statistics (IBGE), Brazil
Building footprint	Alto do Cabrito	City Hall, Salvador, Brazil
DEM and DSM tiles (0.5m × 0.5m)	Alto do Cabrito	City Hall, Salvador, Brazil
PERSIANN rainfall	Alto do Cabrito and Ondina	Google Earth Engine (GEE)
Gauged rainfall	Ondina	Ondina Weather Station
Gauged and interpolated rainfall	Copper Basin	MSc Thesis, Federal University of Bahia
Intensity-Duration-Frequency (IDF) parameters	Salvador	PLUVIO 2.1
Leptospirosis sero-survey	Alto do Cabrito	Institute of Collective Health, Federal University of Bahia, Salvador, Brazil

PLUVIO 2.1, a software application that compiles IDF parameters for different parts of Brazil (Fiorio, Duarte, Rodrigues, Miranda, & Cooke, 2012; Xavier, Cecílio, Pruski, & Lima, 2014), was used to extract IDF parameters for Salvador, Brazil. It is produced by the Water Resources Research Group at the Federal University of Viçosa (GPRH, n.d).

The Hydrologic Engineering Centre-River Analysis System (HEC-RAS), a software application developed by the US Army Corps of Engineers, was used for flood modelling. The version used for this research is 6.0 Beta 2 for Windows. HEC-RAS was chosen for the following reasons: 1) it uses a volume finite difference calculation scheme allowing the representation of the results at terrain level resolution and not at calculation grid level; 2) it is an opensource software; and 3) it is used both by industries and academia for both practice-oriented and research purposes. HEC-RAS provides an added advantage to visualize model results in 3D.

Additionally, QGIS, ArcGIS Pro, R, Python, and GEE were used to visualize and process satellite and other geospatial data.

3.4. Population Estimation

To estimate the population of Alto do Cabrito, two sets of data were used: the 2010 population census survey data of Brazil, and remotely sensed data on building footprints, DSM and DEM of the study area. The 2010 census data was collected by the government of Brazil between August and December 2010 (Brazilian Institute of Geography and Statistics (IBGE), 2012; IHME, 2021). The building footprints, and the DSM and DEM tiles were acquired from the City Hall, Salvador, Brazil (Municipality of Salvador). The cartographic management team of the City Hall produced the building footprints from aerial images acquired between August 19, 2016 and February 13, 2017 using stereophotogrammetric restitution technique. The team produced the 0.5 × 0.5 m² resolution DEM and DSM tiles from point clouds acquired between August 19th, 2016 and February 13th, 2017 using Riegel LiDAR/LASER sensor. The metadata for the building footprints and DSM and DEM do not provide additional information on data processing.

Since the 2010 population figures are aggregated by census sector, the population was disaggregated at building-level. To disaggregate the population, a 3D dasymetric method that uses building footprint (length measurements along X and Y axes) and building height (length measurement along Z axis) data was used. The measurement along X, Y, and Z axes were used to calculate the volume of each building. The volume was multiplied by a factor, which is the total population in the census sector divided by the sum of the volume of the buildings in the census sector (Equations 1 and 2).

$$\widehat{\mathrm{P}}_{\mathrm{ij}} = V_i \times \frac{P_j}{\sum_{i=1}^{N_B^j} V_i} \qquad Equation\ (1)$$

$$\mathrm{V_i} = \mathrm{A_i} \times \mathrm{h_i} \qquad Equation\ (2)$$

where,
P_{ij} = Population of building i in census sector j
P_j = Population of census sector j
A_i = Area of the footprint for building i
h_i = Height of the building i
V_i = Volume of building i
N_B^j = Number of buildings in census sector j

As this method requires estimation of building height, the DSM (0.5 m × 0.5 m res) and DEM (0.5 m × 0.5 m res) tiles from City Hall were mosaiced into new rasters in ArcGIS Pro with assignment of maximum value for overlapping pixels from different raster tiles. Next, normalized Digital Surface Model (nDSM), representing height of above ground objects, was created by subtracting DEM from DSM and recoding negative values to zero (**Figure 4**). The building footprint shapefile was cleaned to contain only valid geometries. **Figure 3** shows examples of intersecting and self-intersecting polygons that were cleaned using topological corrections and buffers in QGIS. The nDSM and building footprints data, then, were read into R, where all nDSM pixels lying over each building footprint were extracted. The pixel values were averaged and assigned to each building as the building height. This procedure was repeated to assign the maximum of pixel values and centroid value as the building height. At this point, census-level population data was merged with the building height dataset. The population factor was estimated, which is simply the total population of each sector divided by the sum of the volumes of the buildings for that sector, where the building volume is the building height times the area of the building footprint. In the equations above, the population estimate is related to building volume, which in turn is a function of the building height. Changes in building height, then, can affect the estimated population. Since there is no literature guiding the choice of the building height and in the absence of ground validation data, the building volume calculation and, hence, the population calculation was done using all three estimates – the centroid height, the averaged height, and the maximum height of the building. The population living in each building was then estimated by multiplying the population factor with the volume of each building.

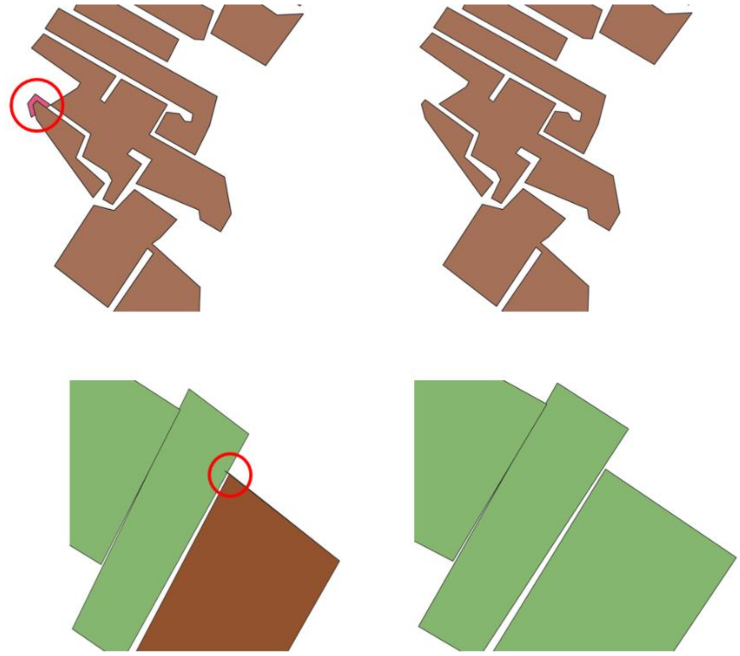


Figure 3. Examples of invalid geometries that were cleaned. The red circles indicate self-intersecting (top left) and intersecting polygons (bottom left).

It should be noted that the 2010 population should be disaggregated using 2010 building footprints, but the building footprint data used in this study was digitized between 2016 and 2017. However, the footprints were deemed appropriate for use in 2010 population disaggregation, because of large number of missing buildings in the data. This is apparent when the building footprint is overlayed with the 2017 DSM for the study area.

To assess the accuracy of the population estimation, Root Mean Square Error (RMSE) was used to compare the observed total census sector population with the total of the estimated building-level population. The estimated building-level population can be summed without rounding, by rounding up, rounding down, or rounding to the nearest integer. As the literature does not provide guidance on the choice of rounding, RMSEs were computed and compared for these four types of rounding. The total census sector population was also compared with the reported population for Aldo do Cabrito in Patricia Brito et al. (2020) to test for accuracy of acquired data.

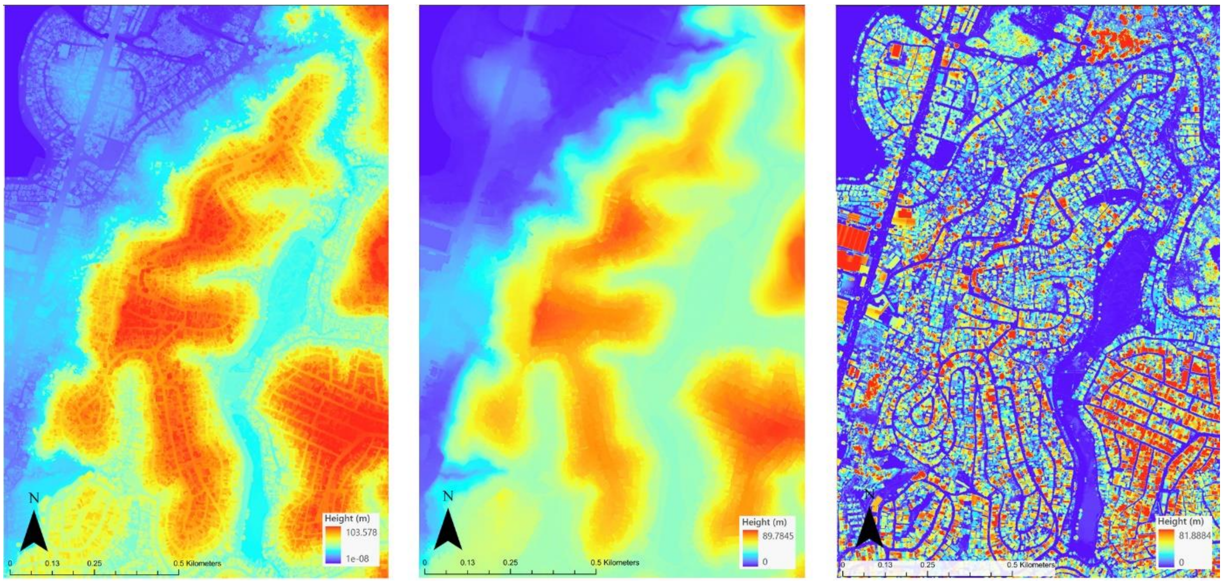


Figure 4. DSM (left), DEM (center), and nDSM (right) for the study area

With the absence of the growth rate for Alto do Cabrito, the population for the year 2017 was estimated using the averaged population growth rate for Brazil as well as the growth rate for metropolitan Salvador as per the exponential growth model (Equation 3) (Schacht, 1980). The growth rates for Brazil were downloaded from The World Bank databank for the years 2010-2017 (The World Bank Group, n.d.) and are shown in **Table 3**. The population growth rate for Brazil has been declining since 2010, but the annual rate of change is small. Therefore, it is reasonable to average the growth rates between 2010 and 2017. The growth rates for Salvador were downloaded from a web platform that collects the world's population statistics (World Population Review, n.d). The growth rates for the metropolitan Salvador are higher than for Brazil but show a decreasing trend for recent years (**Table 4**). It should be noted that the growth rate of 1.46% for Salvador is used in this research only for comparison against the national growth rate. Since the reliability of the data source and the growth rate for Salvador is currently unverified, it should not be used to draw conclusions from the comparisons presented here.

$$N_t = Pe^{rt} \qquad Equation\ (3)$$

where,
N_t is the future population at time t
P is the current population
r is the rate of increase in percent, and
t is the number of years over which growth is measured

Table 3. Annual population growth rate for Brazil

Year	2010	2011	2012	2013	2014	2015	2016	2017
Growth Rate %	0.938	0.916	0.894	0.874	0.856	0.839	0.824	0.807

Table 4. Annual population growth rate for metropolitan Salvador, Brazil

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Growth Rate %	1.46	1.46	1.46	1.46	1.46	1.46	1.46	1.46	1.46	1.14	1.12

3.5. Rainfall Analysis

Precipitation is the driving force for flooding as well as an important variable associated with epidemics of leptospirosis because the seasonality and volume of rainfall affects the timing of disease outbreaks as well as floods (refer to section **1.2** for background on leptospirosis and its risk factors). It also affects soil moisture, another environmental parameter associated with *Leptospira* bacteria but not studied in this research.

Since gauged rainfall data is not available for the study area, three different datasets representing historic rainfall for the study area were acquired to compare and assess the rainfall. The three datasets are PERSIANN-CDR data, Ondina weather station data, and combined rainfall data used for an MSc research project on floods in Copper Basin, Salvador, Brazil. The last dataset combines data from different stations that were in operation during 1945 to 2019 and interpolates for missing data.

PERSIANN is satellite-based rainfall data that is developed by the Center for Hydrometeorology and Remote Sensing at the University of California, Irvine (UC-IRVINE/CHRS) with the aim to provide a consistent, long-term, high-resolution (daily at 0.25° spatial resolution), global (60°S-60°N) precipitation dataset. The dataset is created by running the PERSIANN algorithm on Gridded Satellite (GridSat-B1) infrared (IR) data that are derived from merging ISCCP B1 IR data, along with GPCP version 2.2. For this study, precipitation data on a monthly scale spanning 1990 to 2020 was downloaded from the PERSIANN-CDR dataset with Google Earth Engine for Alto do Cabrito and Ondina, Salvador, Brazil. At least 30 years of data are required to assess rainfall variability over space and time.

The Ondina Weather station dataset contains monthly maximum of 24-hr precipitation and monthly aggregated precipitation for the years 1963 to 2019 for Ondina, Brazil. It is the only active weather station in Salvador and is approximately 11.6 km from Alto to Cabrito (**Figure 5**).

Figure 5. Google Earth view of Alto do Cabrito and Ondina, Salvador, Brazil

While some studies state that PERSIANN has good agreements with ground-based rain gauge and radar data (Ashouri et al., 2015; Sorooshian, Hsu, Braithwaite, Ashouri, & NOAA CDR Program, 2014), there are other studies that show that PERSIANN data tends to underestimate heavy rain (T. Zhang, Yang, Dong, & Gui, 2021) and at high elevation (Romilly & Gebremichael, 2011). Therefore, a comparison between PERSIANN data for Ondina and Ondina weather station data is warranted. Discrepancy between the two datasets will be calculated and applied to PERSIANN data for Alto do Cabrito to correct bias.

The rainfall datasets were examined for trend and seasonality. Next, the four datasets were merged and examined for correlation. In the absence of gauged rainfall at Alto to do Cabrito, the bias correction of PERSIANN data for Alto do Cabrito (P_{ADC}) relied on error estimation from the PERSIANN derived rainfall for Ondina (P_{Ondina}) and rainfall from Ondina weather station (OWS). Following this, three different bias correction methods were applied to P_{ADC}. With the first bias correction method (BC1), the difference in means of the P_{Ondina} and OWS were added to P_{ADC}. With the second (BC2), the difference between P_{Ondina} and OWS was added to P_{ADC}. With the third (BC3), P_{ADC} was multiplied by the ratio of sum of P_{Ondina} and sum of OWS. Lastly, the bias-corrected and uncorrected time series data were compared.

3.6. Frequency Analysis

The frequency analysis uses historical time series of rainfall to relate magnitudes of extreme events in a certain prefixed interval or duration to their probability of occurrence. As the number of observations increases, the expected error in expected rainfall decreases. Thus, rainfall data spanning at least 30 years is needed and for extreme events, longer time series will be required. Additionally, there should be no trend in the time series, as only homogenous data can be used for frequency analysis (Dirk, 2013; Lecture, 2009).

The frequency analysis consists of sorting single maximum independent precipitation events of a certain fixed duration that occur in the study period. Since the events must be independent, only one event per year for the study period is selected. The output of the frequency analysis is the exceedance probability that a certain event with a certain intensity and duration occurs. It is then related to the return period, which is normally expressed in years. Next, a distribution is fit to the data to estimate rainfall depths for different return periods (Dirk, 2013; Fetter, 2001; Lecture, 2009; Oosterbaan, 1994).

The frequency analysis in this study was conducted on the bias uncorrected P_{ADC}, bias-corrected (BC2) P_{ADC}, and OWS data that were processed to provide the monthly maximum of 1-day rainfall. The data were ranked in descending order. The probability associated with the rank was calculated using the Weibull formula (Equation 4) and the return period was calculated by inversing the probability (Equation 5). These two equations define a discrete treatment of rainfall frequency. Next, the rainfall intensity was plotted against the return period. Prediction for other return periods were made using the lognormal distribution (Equation 6) and the Gumbel Extreme Value distribution (Equations 7 and 8). Next, the return periods and associated rainfall intensities from the two distributions were compared and goodness of fit was assessed using the Kolmogorov-Smirnov test (Massey, 1951).

The probability that a certain event x is greater than rainfall in the rank "m" is given by:

$$F(x > x_m) = \frac{m}{n+1} \qquad Equation\ (4)$$

The return period is given by:

$$T = \frac{(n+1)}{m} \qquad Equation\ (5)$$

where,
T = return period
n = number of years of observation
m = rank

For the lognormal distribution,

$$Y = \alpha + \beta \times \ln(X) \qquad Equation\ (6)$$

where,
Y is precipitation corresponding to the return period (T);
X is return period (T);
α and β are model estimators.

For the Gumbel distribution,

$$x_T = \bar{x} + K_T * s \qquad Equation\ (7)$$

where,
x_T is precipitation corresponding to the return period (T);
$\bar{x}$ is mean of the annual maximum precipitation values for a given duration;
s is the standard deviation of the annual maximum precipitation values; and
K_T is the frequency factor.

$$K_T = -\frac{\sqrt{6}}{\pi}(0.5772 + ln[ln(\frac{T}{T-1})]) \qquad Equation\ (8)$$

The Kolmogorov-Smirnov test assesses the following hypotheses:
H_0: The precipitation data follows a specified distribution.
H_A: The precipitation data does not follow a specified distribution.

Here, the specified distributions are the lognormal and the Gumbel extreme value distributions. The significance level chosen for the Kolmogorov-Smirnov test is α = 5%.

The next step before rainfall-runoff modelling is the construction of an IDF curve. Since three different bias-correction methods for PERSIANN resulted in different results and data to validate the results of the bias correction is unavailable, it was reasoned that PERSIANN rainfall data should not be used for further analysis. Moreover, the temporal resolution at which the data was available was not adequate to create an IDF curve. Hence, a literature search was conducted to identify an established IDF equation for Salvador, Brazil. PLUVIO 2.1 software application that compiles IDF equations for different parts of Brazil (Fiorio et al., 2012; Xavier et al., 2014) was identified and downloaded. Next, the IDF parameters for Salvador Brazil (Equation 9) was extracted and an IDF curve constructed.

$$I = \frac{KT^a}{(t+b)^c} \qquad Equation\ (9)$$

where,
I is average maximum intensity of rain (mm/hr);
T is return period (yr);
t is duration of rain (min); and
K, a, b, c are local adjustment constants.

The parameters values for Salvador, Brazil are as follows: K = 1288.500, a = 0.200, b = 22.00, and c = 0.810

Information from the IDF curve was used to create a design rainfall hyetograph, which forms the forcing for rainfall-runoff modelling in HEC-RAS. The rainfall hyetograph describes the distribution of rainfall over a given duration (Lecture, 2009). For example, for a 24-hr rainfall, the rainfall distribution curve gradually increases till it reaches a maximum and then gradually decreases. There are many methods to create a rainfall hyetograph. Here, the alternative block method was used since it assures that the maximum intensity happens once the initially available soil moisture deficit is fulfilled. This method distributes the total rainfall depth over *n* successive time steps of duration Δt over a total duration $T = n \times \Delta t$. In practice, for a given design storm, intensities are taken from an IDF curve and multiplied by the duration to get precipitation depths. Differences between precipitation depths are calculated successively and sorted in descending order. These differences represent rainfall increments per time step. Next, the maximum increment (or block) is assigned to the centre of total duration T. Then, the next largest is assigned to the next time interval and the next largest to the previous time interval from the maximum, and so on in an alternating manner (Lecture, 2009).

For this study, rainfall hyetographs were created for design storms of return periods 2, 5, and 10 years.

3.7. Rainfall-Runoff Modelling

The following assumptions were made for rainfall-runoff modelling:

i. Floods occur after high intensity rainfall during wet periods when soil moisture is likely to be nearly saturated. This assumption is based on the terrain of the study area as well as reports of heavy runoff after intense rainfall.
ii. There is increased runoff in the study area due to the expansive built-up area that reduces soil infiltration.
iii. For the 24-hr rainfall simulation event, evapotranspiration is ignored.

Due to the unavailability of drainage data, the absence of storm-water drainage system was imposed as a model constraint.

For 2D modelling, HEC-RAS was used. The DSM of the study area was imported as the terrain data into the software's RAS Mapper (**Figure 6**). DSM helps to represent buildings as obstacles around which water is forced to flow instead of freely flowing through the terrain. Next, the 2D computational area was created over the terrain by generating a 5m × 5m mesh. A finer grid is desirable but is computationally prohibitive. The terrain extent was chosen as the boundary for the calculation mesh to model rainfall over the entire study area. Next, a land cover layer was added to RAS Mapper to set the Manning's roughness coefficient (refer to **Table 5**) for the flow area. The land cover layer was created in ARC GIS Pro with binary thresholding of DSM to create two distinct classes – roads and buildings. Later, in QGIS, polygons for lowlands were created, rasterized, and added to the binary thresholding raster (**Figure 6**). Null pixels around the edges of the classified DSM were replaced by zero and the final raster was converted into an integer format that HEC-RAS can read.

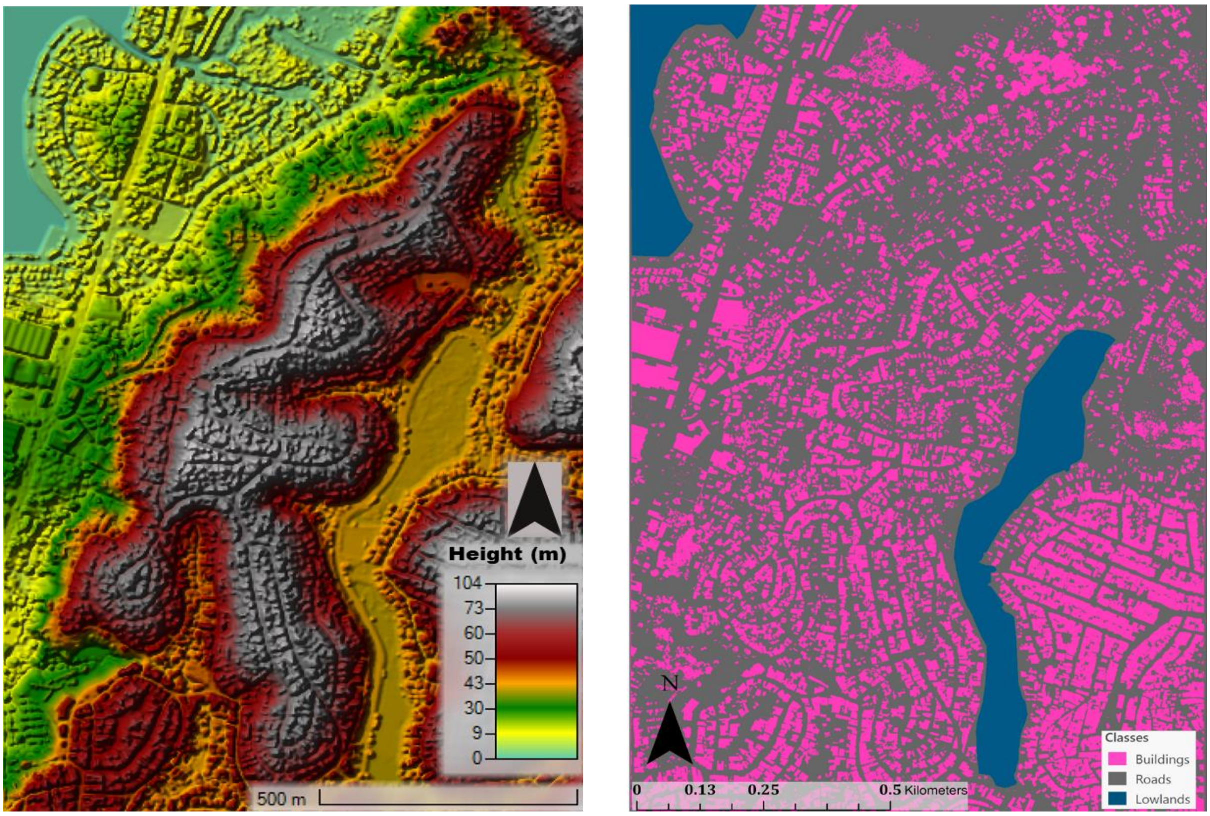


Figure 6. Map of the terrain for the study area (left) and land cover (right)

Three different sets of Manning's roughness coefficients (**Table 5**) (Jung, Park, Park, Lee, & Kim, 2011; Marcus, Roberts, Harvey, & Tackman, 1992; Papaioannou et al., 2018; USACE, 2010) were tested before deciding on Set 3 for modelling. This was an attempt to evaluate the response of the model to the change in roughness and is not indicative of the variety that can be found in the study area. The use of DSM and Manning's coefficient helps to keep the buildings impervious so that water will flow around the buildings.

Table 5. Manning's Coefficients for different land cover types

	Manning's N		
Land Cover Type	**Set 1**	**Set 2**	**Set 3**
Lowlands	0.031	0.031	0.031
Buildings	0.200	1	0.200
Roads	0.013	0.016	0.016

Next, the rainfall hyetographs for different return periods were used as boundary conditions for 2D unsteady flow analysis. Initially, the simulation duration was set to 24-hrs with a computation interval of 10 minutes. Later, the simulation duration was set to 3 days for model outputs to stabilize. At the end of the simulations, outputs for water depths were exported. Since pooled water is important for *Leptospira* growth and survival as well as exposure to humans, the final events from simulations are more important than the intermediate time steps.

Lastly, visualizations of the water depths were produced in 2D using Arc GIS and in 3D using HEC-RAS and ArcScene.

3.8. Exposure Assessment

The water depth rasters of the final time step of the simulations from HEC-RAS were converted into polygons and aggregated. HEC-RAS exports simulation results as a float-point raster which must be converted into an integer raster before performing the raster to polygon conversion. Three different distances were tested for aggregation – 1m, 3m, and 5m – before deciding on 3m. At 3m, adjacent polygons are aggregated without crossing the buildings footprints. This procedure was repeated for all three return periods.

Next, to estimate the number of buildings exposed, the location selection tool was used to select buildings that were within a distance (Euclidean distance) from flood polygons. The distance thresholds considered were 0m, 1m, and 2m. As water collects along most roads in the study area, a distance greater than 2m does not need to be considered. Separate layers were created for the exposed buildings, which were joined with population data. Next, the exposed population was calculated by taking the sum of the estimated number of people in each exposed building. This was performed for all three return periods.

3.9. Leptospirosis Risk Assessment

To assess the risk of leptospirosis, disease data is required. A sub-set of the leptospirosis data collected as part of the sero-survey conducted in Salvador, Brazil is used here. The method for data collection is described in (Khalil et al., 2021). An examination of the data shows that each surveyed household has varying numbers of negative and positive cases. The maps depicting the survey locations and disease statuses are not provided here to protect the identities of the surveyed households.

To assess whether leptospirosis sero-status and the flood are associated, following hypotheses are assessed.
H_0: There is no association between the leptospirosis sero-status and the flood.
H_A: There is an association between the leptospirosis sero-status and the flood.

To assess the association between leptospirosis and flood, ORs are calculated along with the 95% confidence intervals associated with the ORs. An odds ratio measures the odds of an outcome in the exposed to the odds of an outcome in the unexposed (**Table 6**, Equation 10). In this study, the OR measures the odds of sero-positive case in the exposed group to the odds of sero-positive case in the unexposed group. The exposure statuses are either being in contact with the pooled water or not.

Table 6. Contingency table

Exposure	Outcome	
	Yes	No
Yes	a	b
No	c	d

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a \times d}{b \times c} \qquad Equation\ (10)$$

4. RESULTS

4.1. Population Estimation

The results of population disaggregation show that the numerical measure used for the building height and the rounding method will affect the estimated populations. The disaggregation that used the maximum building height with no rounding gave the smallest error (RMSE = 0 people) followed by rounding to the nearest integer (RMSE = 17.93 people). **Figure 7** shows the 2010 population of Alto do Cabrito that was disaggregated at the building level using the maximum building height and the rounding set to the nearest integer. The building-level population ranges between 0 and 77 persons. The overall mean building-level population is 3 persons and standard deviation is 4 persons. It should be noted that there was no height information for the buildings on the south-west corner of the study area. So, the population for these buildings were estimated as zero.

For 2010, the acquired census sector data that is used in this research has 17, 655 people in Alto do Cabrito. For the same year, the population of Alto do Cabrito is reported as 17,051 in Patricia Brito et al. (2020). So, there are 604 more people in the acquired census sector data than reported for Alto do Cabrito. This difference also leads to different number of populations forecasted for 2017. Using the acquired data and with an average growth rate of 0.8685%, the 2017 estimated population in Alto do Cabrito is 18,762, whereas using the reported 2010 population and the same average growth rate, the estimated population is 18,120. The difference between the two forecasted population is 642 people. Using the growth rate of metropolitan Salvador and based on the acquired census-sector data, the 2017 estimated population for Alto do cabrito is 19,555, whereas based on the reported population, it is 18,886. The difference between the two is 669 people.

The population forecast based on the acquired census sector data shows that 1,900 more people are living in Alto do Cabrito than in 2010. This is a large increase. In the absence of an updated and accurate building footprint for 2017, the forecasted population should not be disaggregated at the building-level, because the disaggregation will produce inaccurate building-level population.

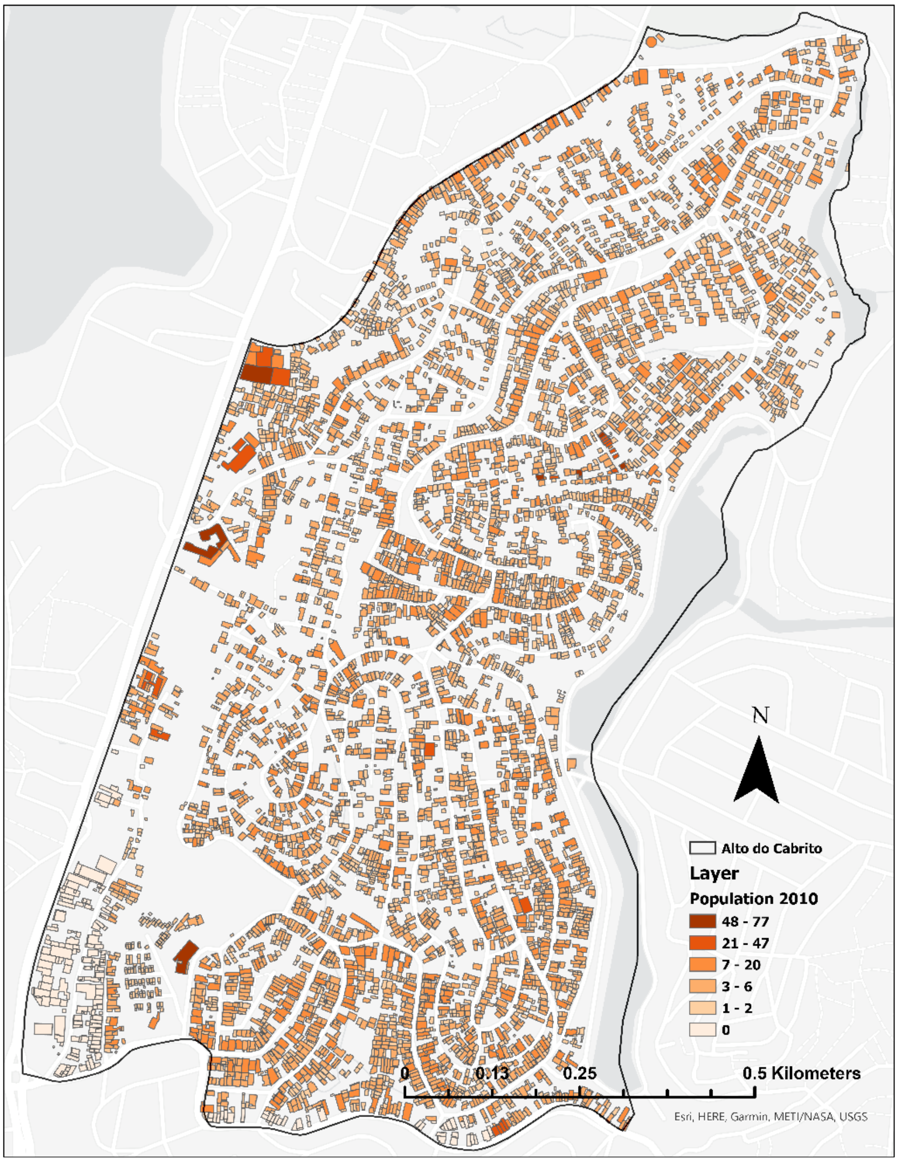


Figure 7. 2010 population of Alto do Cabrito at building level

4.2. Rainfall Analysis

Figure 8 presents PERSIANN derived daily precipitation at monthly scale for the year 1990 to 2020 for Alto do Cabrito and Ondina, Salvador. The rainfall varies annually, with some years receiving significantly more rainfall than others (minimum of 326 mm in 1993 and maximum of 1175 mm in 1999). The average and standard deviation of the total annual rainfall for this period are 714 mm and 214 mm, respectively. But no general trend of increase or decrease in the rainfall is observed during this period (**Figure 9**).

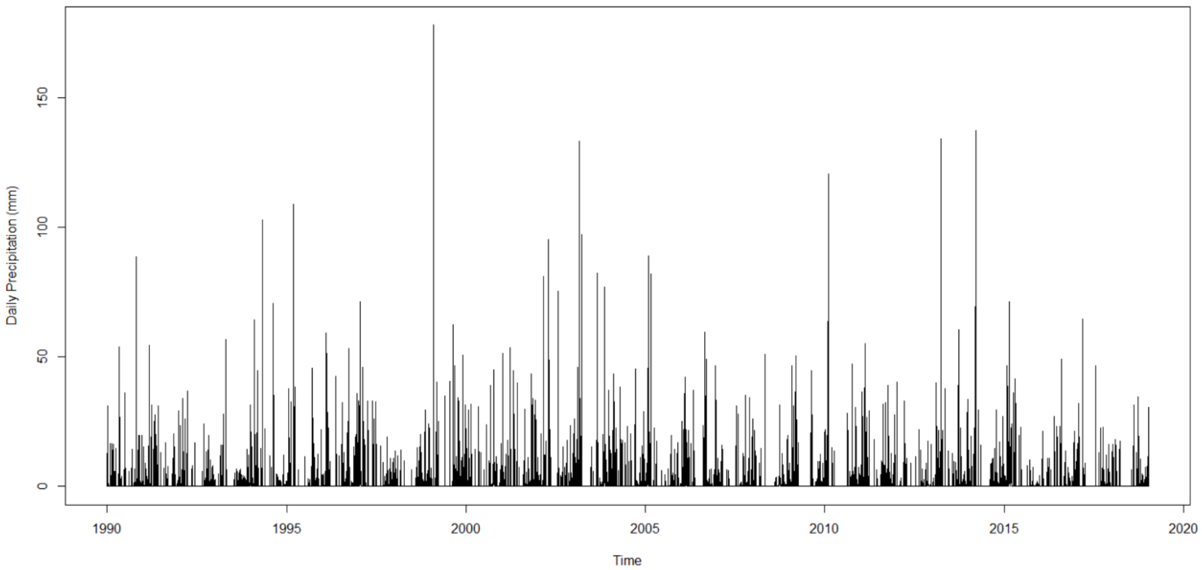


Figure 8. Time series of PERSIANN rainfall for Alto do Cabrito, Salvador, Brazil

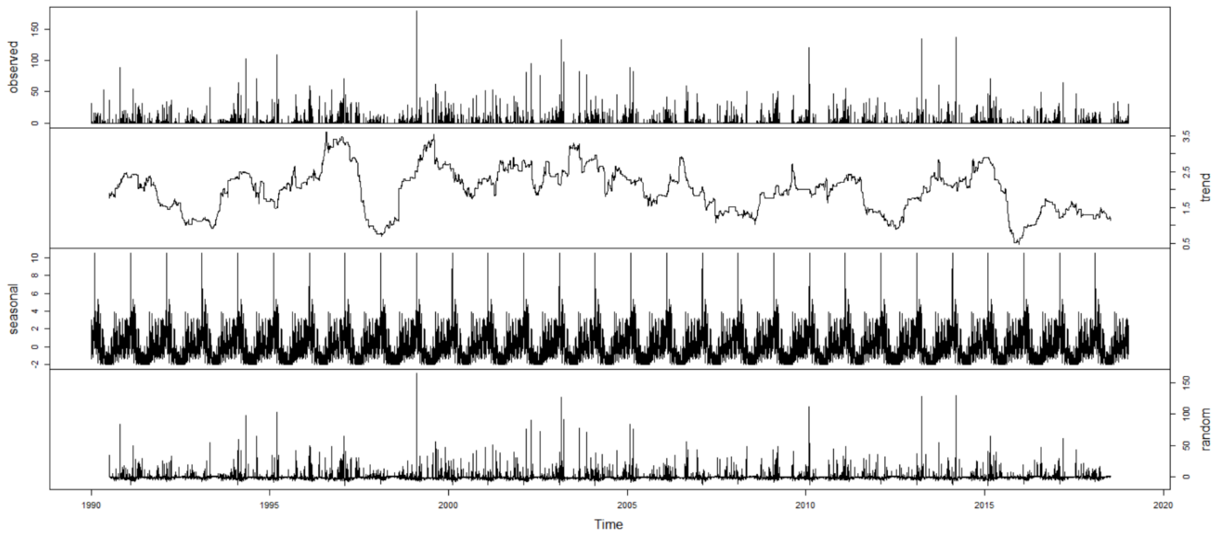


Figure 9. Decomposition of time series of PERSIANN rainfall for Alto do Cabrito, Brazil

Similarly, **Figure 10** presents PERSIANN derived daily precipitation at monthly scale for the year 1990 to 2020 for Ondina, Salvador. The average and standard deviation of the total annual rainfall are 745 mm and 203 mm, respectively. The minimum rainfall of 379 mm occurs in 1993 and the maximum of 1150 mm occurs in 1997. Like Alto do Cabrito, there is no general trend of increase or decrease in the rainfall over the years considered (**Figure 11**).

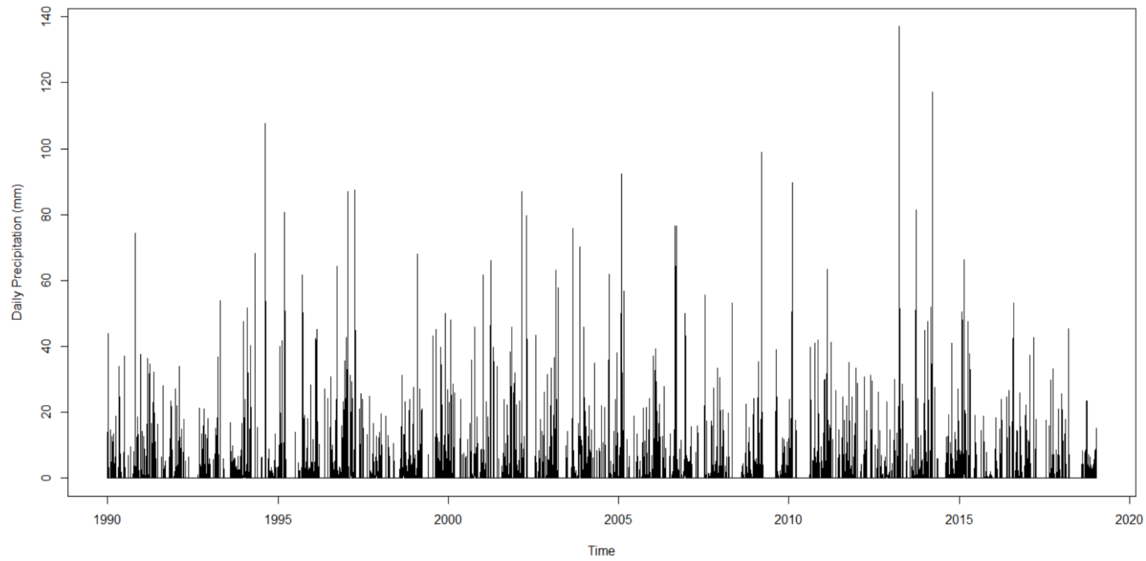


Figure 10. Time series of PERSIANN rainfall for Ondina, Salvador, Brazil

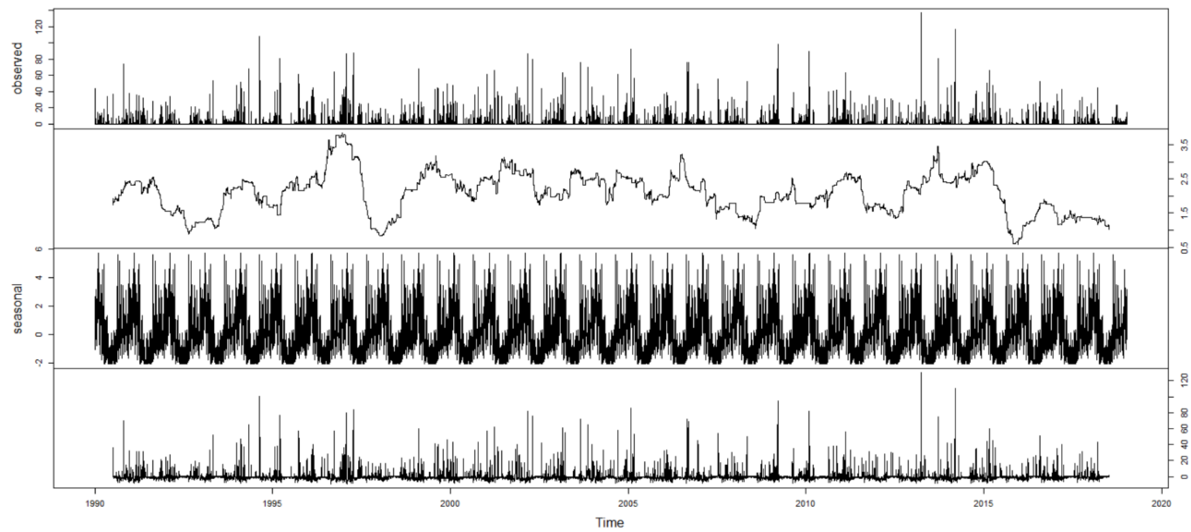


Figure 11. Decomposition of time series of PERSIANN rainfall for Ondina, Brazil

Unlike the PERSIANN rainfall data, the OWS data and combined data are not daily time series of 24-hr rainfall. Instead, these data are monthly maxima of 24-hr rainfall for years 1963 – 2019 and 1945 – 2019 respectively. Therefore, for comparison, **Table 7** presents the summary statistics of 24-hr rainfall for months with greater frequency of monthly maxima for the four data sources.

According to the statistics, the range and average rainfall volume vary largely across the data sources. However, all show that the wet months are April, May, and June. According to the combined data, there is high rainfall in the months of February, March, and November as well.

Table 7. Statistics of monthly maximum of 24-hr rainfall for four different rainfall data sources

Data Source	Months	Min (mm)	Mean (mm)	Max (mm)	SD (mm)
P_{ADC}	April	55.1	100.4	177.9	50.7
	May	36.7	79.8	137.3	37.2
	June	56.6	97.2	134.1	31.9
P_{Ondina}	April	63.2	78.3	92.4	14.8
	May	42.7	76.7	117.2	25.4
	June	53.9	86.1	137.1	36.7
OWS	April	73.8	121.5	232.5	37.9
	May	68.7	115.2	186.4	34.1
	June	75.8	108.3	141.0	27.2
Combined	February	67.8	111.3	159.9	39.4
	March	64.8	110.9	180.2	45.9
	April	54.3	113.8	232.5	39.2
	May	68.7	102.1	186.4	29.1
	June	47.1	97.0	141.0	29.5
	November	61.9	101.9	135.2	26.6

The Pearson's correlation shows that for the common years of 1990 – 2019, the yearly maximum 24-hr precipitation data from PERSIANN is very poorly correlated with OWS and combined data (**Table 8**). The perfect correlation between the Ondina station and combined data is because the latter takes the maximum precipitation value from the Ondina Station for 1990-2019. If the OWS is taken as the ground truth, PERSIANN data underestimates rainfall in Salvador, Brazil and should be corrected for errors before further use. The results and discussion of the bias correction methods can be found in the **Appendix**.

Table 8. Correlation of rainfall from different data sources for common years (1990 – 2019)

	P_{ADC}	P_{ondina}	OWS	Combined
P_{ADC}	1			
P_{ondina}	0.67	1		
OWS	0.17	0.03	1	
Combined	0.17	0.03	1	1

4.3. Frequency Analysis

The **Figure 12** shows the return period and the associated rainfall intensity computed using the Weibull formula. As expected, the rainfall intensity increases with increasing return period. A noticeable difference is in the rainfall intensities derived from the bias corrected and uncorrected P_{ADC}. The computed rainfall intensities for bias-uncorrected values are smaller than for bias-corrected values.

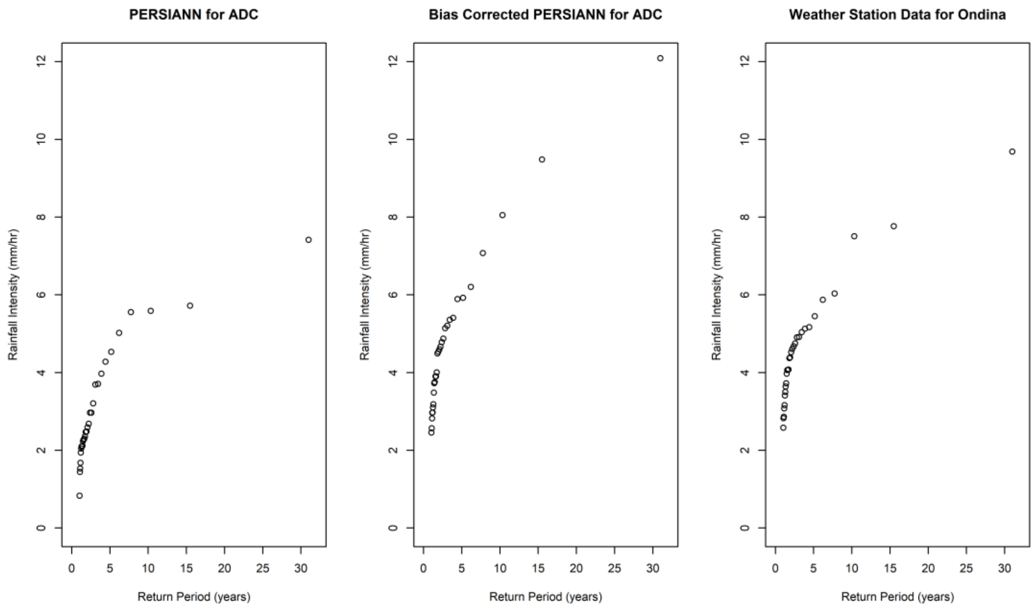


Figure 12. Plot of rainfall intensities against return periods

Table 9 shows the results of the prediction of return periods and associated rainfalls from lognormal and Gumbel extreme value distributions. According to the Kolmogorov-Smirnov test, the null hypothesis is not rejected for neither lognormal nor Gumbel extreme value distribution. However, visual examination of the goodness of fit through a Q-Q plot shows that the lognormal distribution fits the data better than the Gumbel distribution (**Figure 13**). However, it should be noted that the sample size is too small (n = 8) to have an absolute confidence on the test statistics. Here, the sample or the observed data consists of the original 24-hr precipitation that correspond to the following return periods, T = (1, 2, 3, 4, 5, 10, 15, 30).

Table 9. Return periods and 24-hr design rainfalls estimated from different distributions

T (years)	Log-normal Distribution			Gumbel Extreme Value Distribution		
	Uncorrected P_{ADC} (mm)	Bias-corrected P_{ADC} (mm)	OWS (mm)	Uncorrected P_{ADC} (mm)	Bias-corrected P_{ADC} (mm)	OWS (mm)
1	34.67	61.88	71.08	-	-	-
2	64.27	102.5	100.9	69	109	105.7
3	81.58	126.3	118.4	84.37	130.1	121.3
4	93.87	143.1	130.8	94.2	143.7	131.2
5	103.4	156.2	140.4	101.5	153.7	138.6
10	133	196.9	170.2	123	183.3	160.4
15	150.3	220.6	187.7	135.1	200.1	172.7
30	179.9	261.3	217.5	155.5	228.1	193.4
50	201.7	291.2	239.5	170.3	248.5	208.5
100	231.3	331.8	269.4	190.3	276.1	228.8
200	260.9	372.5	299.2	210.2	303.6	249
300	278.2	396.2	316.7	221.9	319.6	260.8
500	300.1	426.2	338.7	236.5	339.8	275.7
1000	329.7	466.8	368.5	256.4	367.2	295.8

At this point, since the results of the bias correction and the estimates from the distribution differed, and since validation data and data at finer temporal resolution are unavailable, the IDF curve (**Figure 14**) is created using the IDF equation for Salvador. Each curve in the figure shows the relationship between duration, plotted on the abscissa, and intensity of rainfall, plotted on the ordinate, for different frequencies, which is expressed as return periods. This IDF relationship was used to develop the rainfall hyetographs for return periods of 2, 5, and 10 years (**Figure A. 4**).

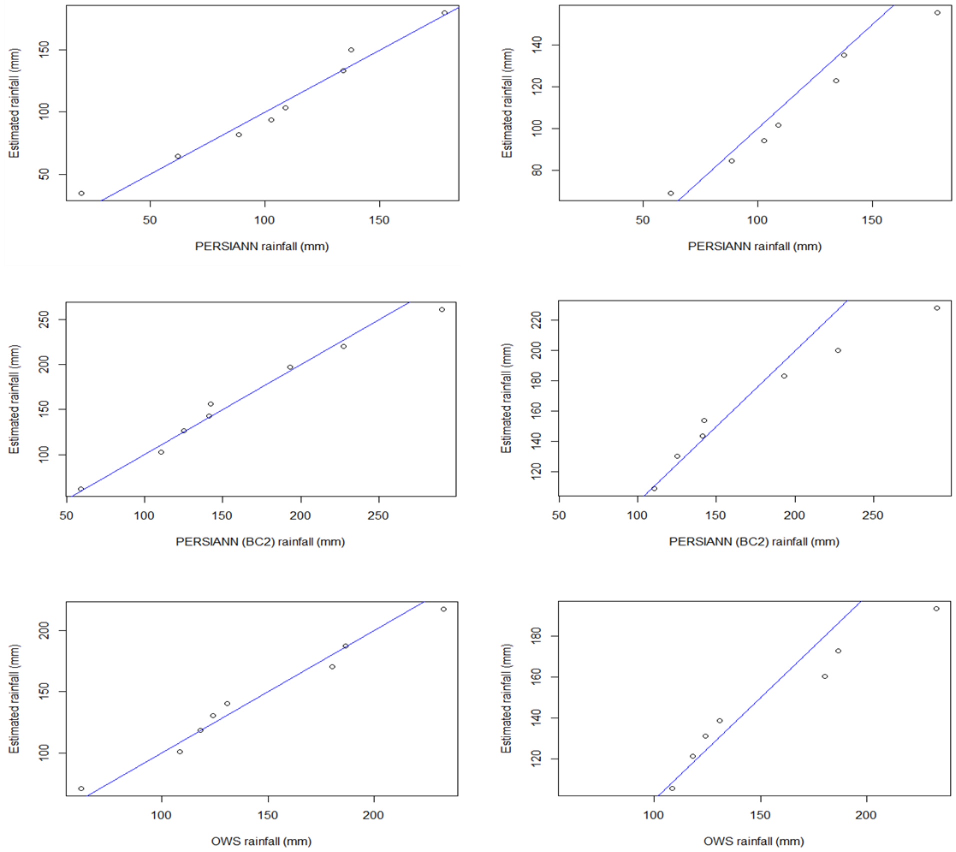


Figure 13. Q-Q plots for estimates from the lognormal (left) and the Gumbel distribution (right)

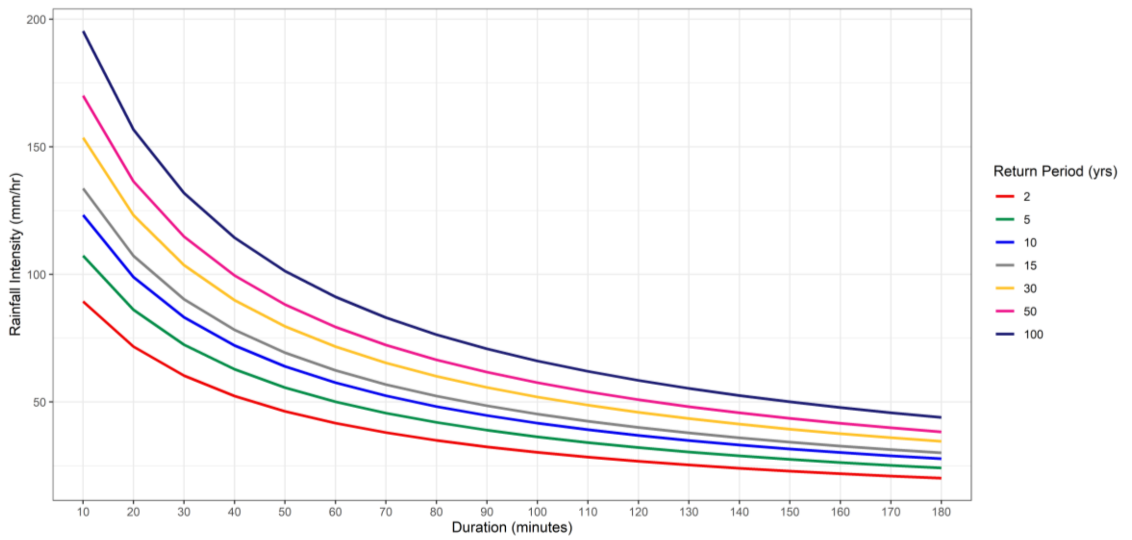


Figure 14. Intensity-Duration-Frequency (IDF) curves for Salvador, Brazil

4.4. Rainfall-Runoff Modelling

The rainfall-runoff simulation results (**Figure 15**) show that roads in Alto do Cabrito act as conduits for water. The water flows through roads and down the elevation gradient to the depression on the east and the ocean to the west. The left-over water accumulates as puddles across the study area. When the rainfall intensity increases, there is no increase in the extent of the flooded areas but there are changes in depths of collected water. The increased water depth observed along the south and south-west of Alto do Cabrito are due to the boundary condition imposed during simulation. If the storm water drainage and other water outlets were considered, the rainfall would eventually flow out of Alto do Cabrito and water would not accumulate along the edges. The increased depth observed over the ocean is a modelling artifact and should not be interpreted as increased water depth due to rainfall. There are many small areas, made up of one to two pixels (0.5m) – that show water collection, which are likely caused by the lack of soil infiltration and evapotranspiration in the model. Evapotranspiration leads to loss of water and an area only floods after the soil infiltration capacity has been saturated. To consider these parameters in the model, data needs to be collected from the field. Also, it is likely that, due to the very fine resolution of the terrain model, very small ground objects obstructed flow of water that caused water to accumulate at various locations.

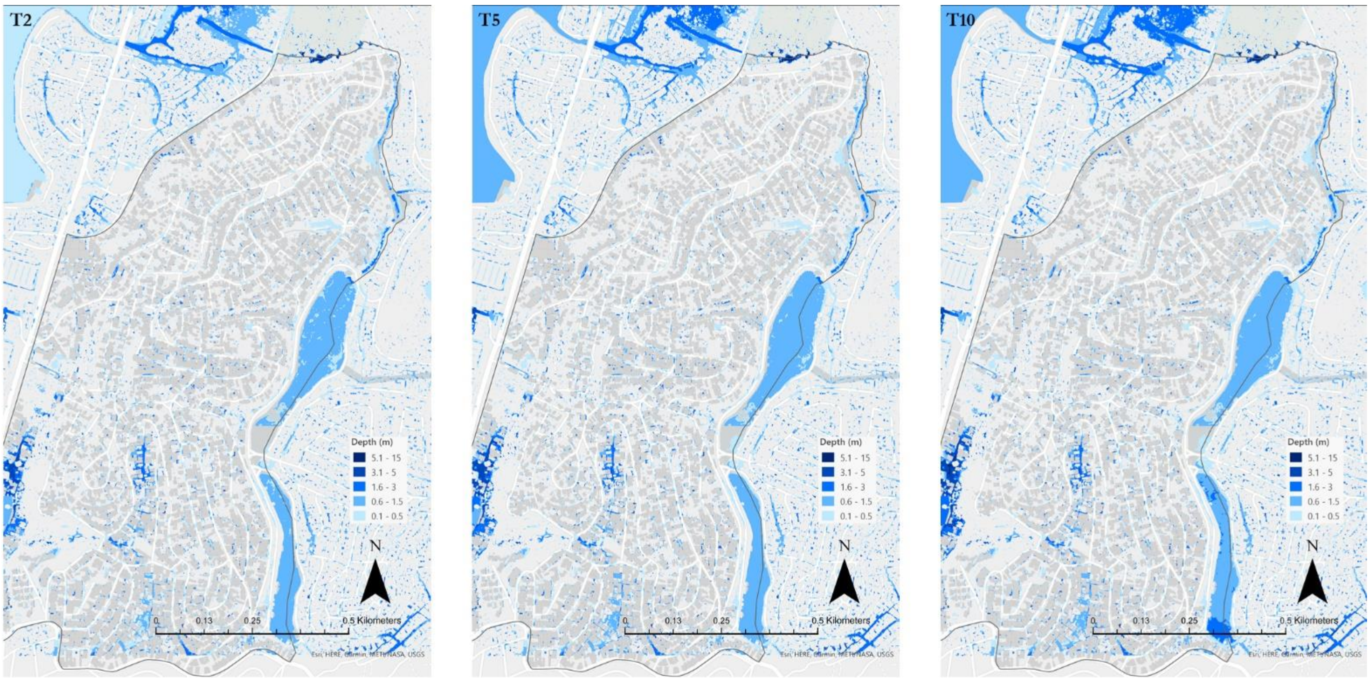


Figure 15. Water depths after rainfall of return periods of 2 (T2), 5 (T5), and 10 (T10) years

Figure 16 shows the maximum water depth for each pixel. The time when the maximum occurs can be different for different pixels and does not correspond to the peak rainfall. **Figure 17** shows a 3D view of the flooded area near Dique do Alto do Cabrito from HEC-RAS 3D visualization. **Figure 18** also shows the 3D view of the flooded area near Dique do Alto do Cabrito but with buildings coloured by their height and overlayed on maximum water depth. Here the potential of a high intensity 24-hr rainfall in Alto do Cabrito to inundate buildings of 6-7m in height and less is clearly observed.

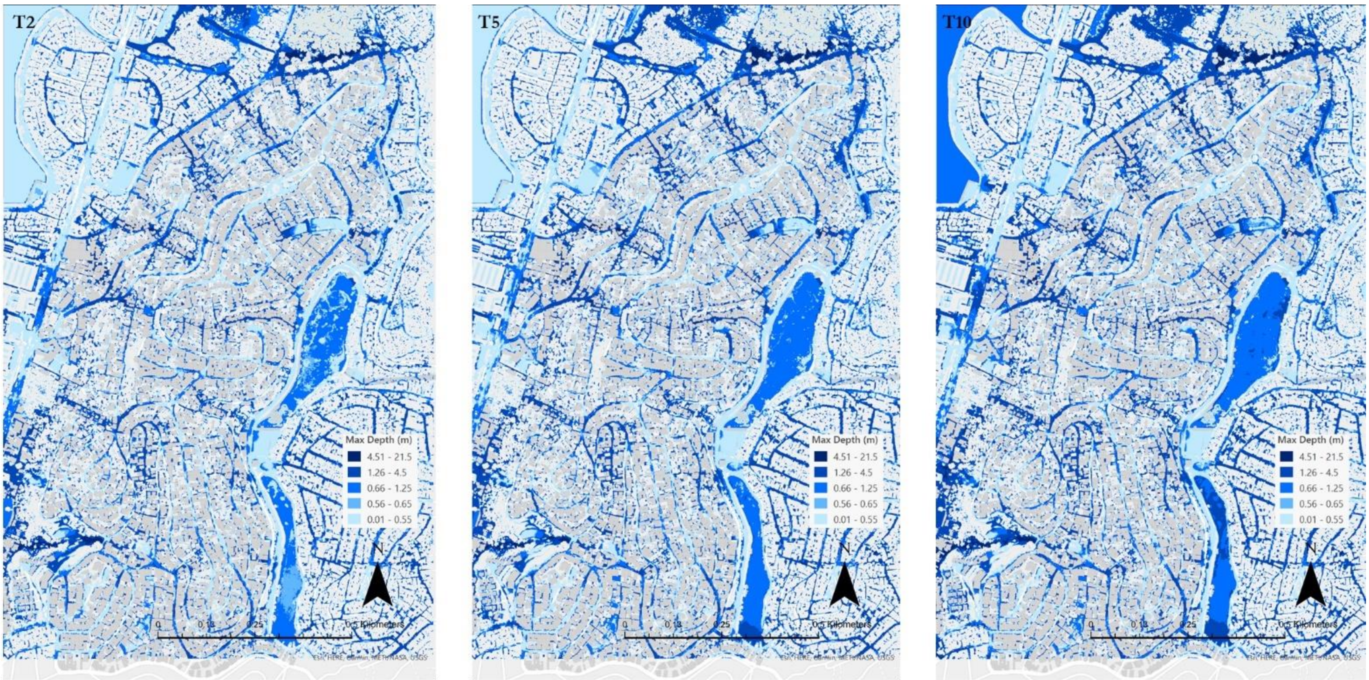


Figure 16. Maximum water depths after rainfall of return periods of 2 (T2), 5 (T5), and 10 (T10) years

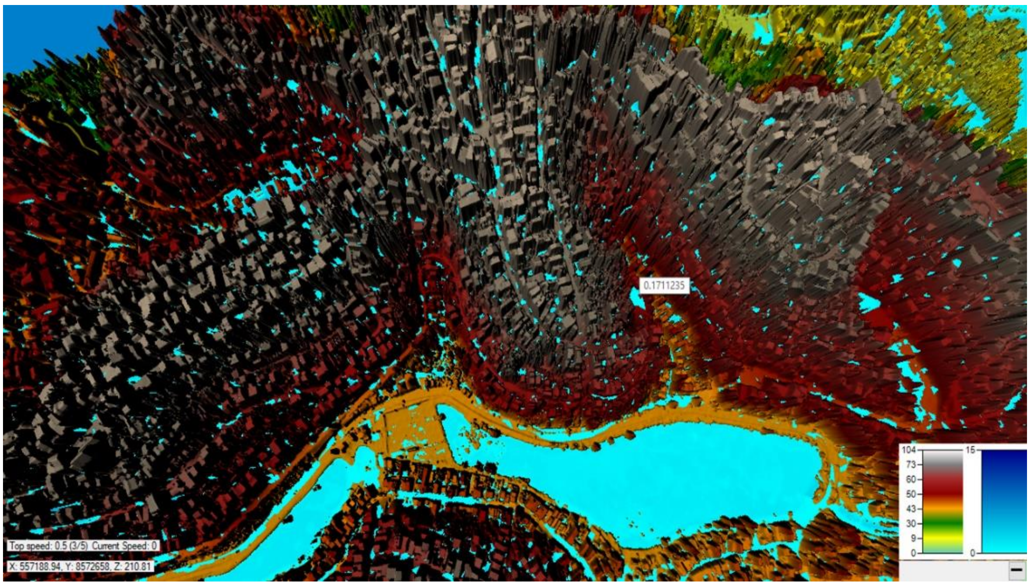


Figure 17. A 3D view of flooded area near Dique do Alto do Cabrito (T = 10 years)

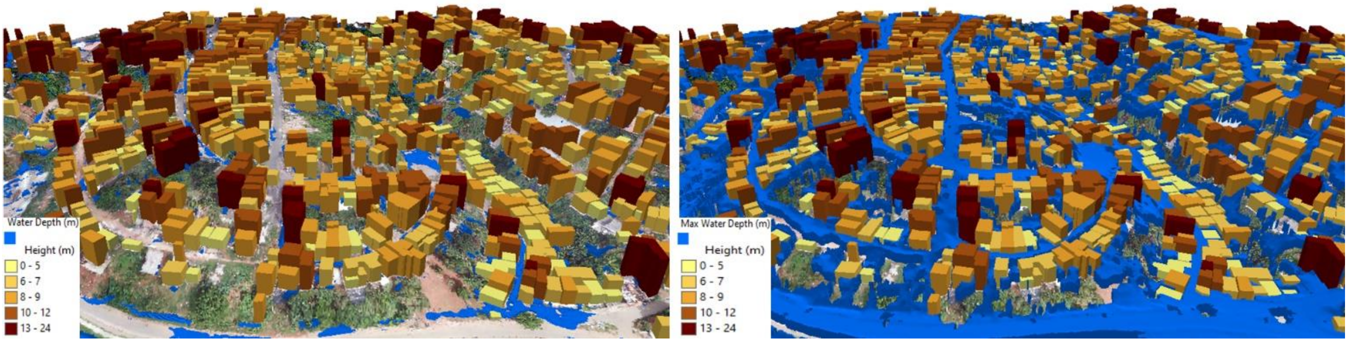


Figure 18. 3D views of flooded buildings near Dique do Alto do Cabrito with water depths at the end of 24-hr rainfall (left) and maximum water depths during the 24-hr rainfall (right).

4.5. Exposure Assessment

Table 10 shows the number and the percent of exposed buildings and people in the study area, which has a total of 6,790 buildings and 17,569 people. After most of the rainfall flows down the elevation gradient, at the end of the 24-hr rainfall, the remaining water pools along the streets and around the buildings. This occurs throughout the study area. So, most of the buildings and people living in the study area are exposed to the flood water. Within each design storm scenario, the number of exposed buildings and people increase with the distance thresholds. Between the design storms, there is a very small difference in the number and the percentage of exposed buildings and people. This is because varying the rainfall intensities lead to changes in water depths and not the spatial extents over which the water collects.

Since the difference between the last time step of 24-hr simulation and the last time step of 3-day simulation is very small, the results for the latter are presented in **Table 10**.

Table 10. Number and percent of exposed buildings and populations

Event	Distance (m)	No. of Buildings	% Building	No. of People	% Population
T = 2 yrs	0	3989	59	12,268	70
	1	5953	88	16,756	95
	2	6246	92	17,137	98
T = 5 yrs	0	3979	59	12,240	70
	1	5927	87	16,703	95
	2	6220	92	17,092	97
T = 10 yrs	0	4278	63	13,048	74
	1	5907	87	16,657	95
	2	6225	92	17,101	97

As shown in **Figure 19**, the exposure will be greater where the population density is higher. Figure 19 (a) shows the buildings in the study area extruded and coloured by their heights. Figure 19 (b) shows the population in the study area. The buildings are extruded by the building heights but coloured by the population number. Figure 19 (c) shows the population density that is calculated by normalizing the population of Alto do Cabrito by building heights. Here the buildings are extruded by the building heights and coloured by the normalized population. The red ellipses indicate relatively short buildings housing relatively larger population. During a high intensity rainfall, flood is more likely to cause the most exposure here. Figure 19 (d) overlays the normalized population with the maximum water depth.

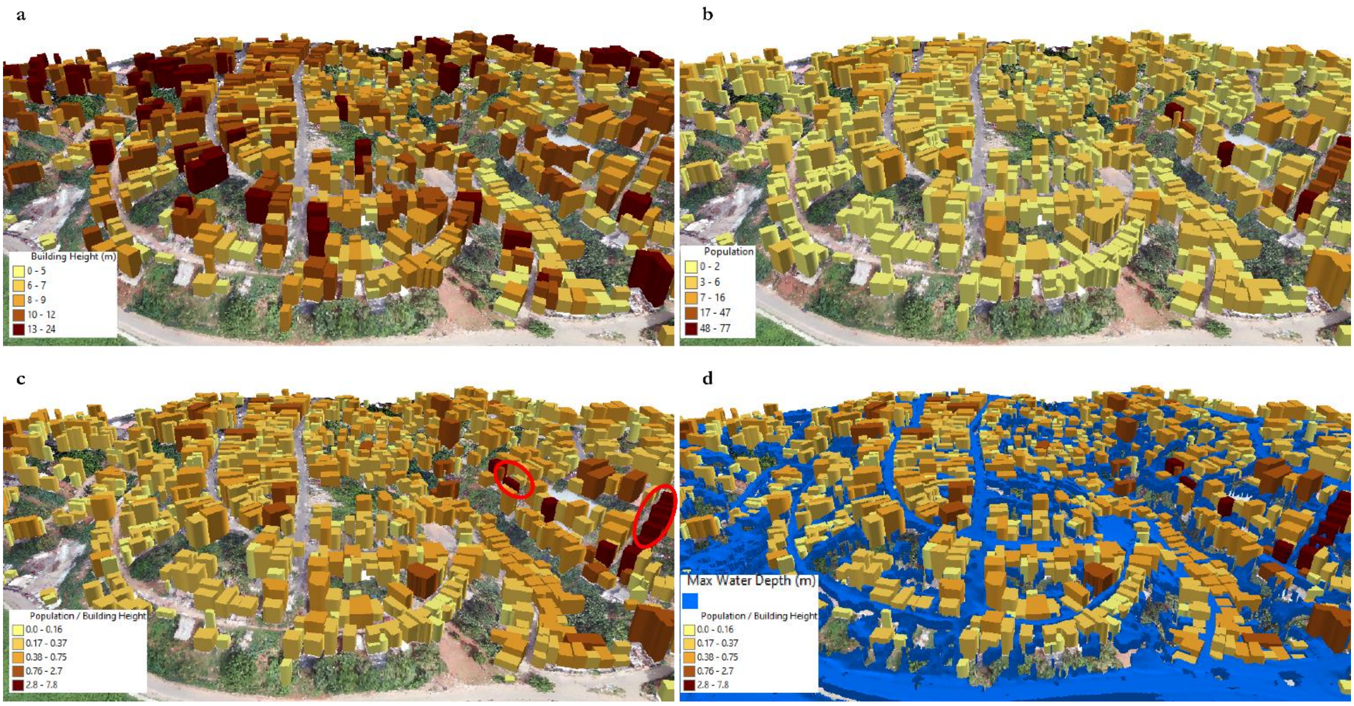


Figure 19. The buildings (a) and population (b) affected by flood. The red ellipses indicate densely populated buildings. The juxtaposition of the population density (c) with maximum water depth (d) indicates the degree of exposure that can occur.

4.6. Leptospirosis Risk Assessment

Of a sample of 375 people in the study area, 38 were leptospirosis positive and 337 were leptospirosis negative. **Table 11** shows the exposure of the positive and negative cases to pooled water by proximity thresholds under different design storm scenarios. It also provides the ORs and the associated 95% CIs. The ORs estimated for different design storm scenarios and proximity thresholds are very small. There is no consistent increase or decrease in the ORs with an increase in the distance for all scenarios considered. In addition, since the 95% CIs contain the null (value 1.0), there is an insufficient evidence that the exposed and unexposed groups are different. Hence, there is no certainty that an association between leptospirosis and exposure to flood exists.

To protect the privacy of the leptospirosis sero-survey participants, maps depicting survey locations of the leptospirosis cases and the population exposed to pooled water who may or may not be sero-positive to leptospirosis are not shown.

Table 11. Exposure and leptospirosis under different design storm scenarios

		Exposed		Unexposed		95% CI		
Scenario (T yrs)	Distance (m)	Positive	Negative	Positive	Negative	OR	L.B.	U.B.
2	0	4	28	34	309	1.30	0.43	3.92
	1	9	92	29	245	0.83	0.38	1.81
	2	20	156	18	181	1.29	0.66	2.52
5	0	4	30	34	307	1.20	0.40	3.62
	1	7	80	31	257	0.73	0.31	1.71
	2	19	160	19	177	1.11	0.57	2.16
10	0	4	50	34	287	0.68	0.23	1.99
	1	7	98	31	239	0.55	0.23	1.29
	2	18	169	20	168	0.89	0.46	1.75

L.B.: lower bound of 95% CI; U.B.: upper bound of 95% CI

5. DISCUSSION

Based on the literature review and the available data for the study area, the most suitable method that integrated 3D data to disaggregate population at the building-level was the 3D dasymetric method. This method utilized the building footprints (x, y-values) and the building heights (z-value) to calculate the building volumes, which was used to distribute the census sector-level aggregated population to building-level. No further disaggregation by floor was performed due to the absence of building-level data. Different studies have used a constant floor height to disaggregate population vertically (Xu, Cao, & Jia, 2020). This study avoided doing the same, because Alto do Cabrito is composed of heterogenous buildings with varying structures that do not conform to building standards. In this scenario, the assumption of a constant floor height would be inappropriate.

In the 3D dasymetric method, population is a function of the building volume, which is a function of the building height and the building footprint area. This means that a priori definitions of the building height and the footprint, such as the definitions found in building standards or urban plans, are necessary to consider. But remotely sensed data cannot provide that level of detail, especially for the building height. So, the published literature have either used the average (S. Zhang, Han, & Bogus, 2020), or the median (F. Biljecki et al., 2016) of pixel or point cloud values over the building footprints, or have failed to mention the measure assigned as the building height (Zhenyu Lu et al., 2011; Xu et al., 2020). Further, in areas like Alto do Cabrito, where building standards are not followed and building structures take disorganized shape, using definitions found in the building standards is impractical. However, arbitrary decisions on a summary measure should not be made either. As such, in the absence of literature guidance, this study estimated the building heights using the centroid, average, maximum pixel values. It was found that compared to the centroid and average heights, the maximum building height resulted in more accurate building-level population as assessed by the smaller RMSE associated with the estimate. This highlights the importance of the selection of a proper summary measure to assign as the building height.

This research focused on utilizing the available building footprints, and DSM and DEM to disaggregate the 2010 population for Alto do Cabrito at the building-level. The accuracy of the 3D dasymetric method depends on the accuracy of the building footprints and the terrain models. Similarly, updated building footprints and terrain models are required to accurately disaggregate forecasted population. Without the accurate updated data, population disaggregation will produce erroneous results. To produce the updated data, there is a need to regularly acquire fine resolution imageries or point cloud data which can be cost-prohibitive. Furthermore, manual digitization of building footprints is time and labour-intensive. To overcome the latter, automated building classification and delineation approaches (Wei & Ji, 2021; Wei, Ji, & Lu, 2019) should be used.

The rainfall analysis that used PERSIANN data showed an annual variation in rainfall. No trends were observed over the 30 year period which is contrary to the negative trend reported for Salvador (A. P. P. d. Santos et al., 2016). Apart from any spatial differences, the contradictory finding is most likely related to the different time periods under study. In this research, the PERSIAN data for the years 1990-2020 were analysed for Alto do Cabrito and Ondina, whereas A. P. P. d. Santos et al. (2016) analysed rainfall for 1960-2010 for Salvador. The authors reported that almost all large rainfall volumes occurred in the 1980s.

The rainfall analysis concluded that PERSIANN underestimates precipitation. Both PERSIANN and A. P. P. d. Santos et al. (2016) report 1993 as one of the driest years. But while authors reported 1235 mm of total rainfall in 1993, the PERSIANN reported 326 mm and 379 mm of total rainfall in 1993 for Alto do Cabrito and Ondina, respectively. According to the PERSIANN and OWS data, the wet seasons consisted of April, May, and June, whereas A. P. P. d. Santos et al. (2016) reported April, May, June, and July as making up the wet season for Salvador. Each of these data sources provided different intensities of rainfall during the wet seasons, with PERSIANN underestimating the rainfall compared to the other.

The significant finding of the rainfall analysis for leptospirosis was that the months of April to June, when heavy rainfalls occur, provide a suitable wet environment for the growth and survival of the *Leptospira* bacteria. Based on the background research (refer to section **1.2.1**), precipitation is a significant factor associated with the *Leptospira* bacterial growth and survival and the risk of leptospirosis. However, the differences in rainfall intensities reported by different rainfall data sources suggest caution in choosing data for research as different input data may lead to different results.

While this research assumed that the rainfall volume and rainfall seasonality are key factors for the bacterial presence in the environment and hence leptospirosis, there are other factors that need to be examined in parallel. These include but are not limited to consecutive days of rainfall, temperature, and humidity (Cunha et al., 2019; Hacker et al., 2020), as well as the presence of *Leptospira* in the soil and water (Bierque et al., 2020), and host species, such as rats (Socolovschi et al., 2011).

The 2D hydraulic model was important to simulate rainfall-runoff in Alto do Cabrito. The fine resolution DSM provided detailed terrain information that helped to represent buildings as obstructions to free flow of water. This combined with the roughness of buildings helped to correctly model water flow around the buildings. In this study, the Manning's roughness coefficient was applied to three dominant land cover classes in the study area. On the ground, land cover may have more variety. Moreover, within class variation was not considered. To make such considerations, field survey is required.

According to the flood simulations, the high intensity 24-hr rainfall combined with the steepness of the terrain in Alto do Cabrito results in high rainfall-runoff. This is in accordance to the expected rainfall-runoff in Salvador reported by A. P. P. d. Santos et al. (2016). Though the flood simulations were ran for three different design storms, the result should hold true for the different rainfall intensities observed during the wet season.

During the 24-hr rainfall events, almost all the study area is prone to inundation. Since Alto do Cabrito lies on an elevated ground, most of the rainwater drains away quickly. The remaining pooled water appears all over the study area. During the wet season, this is expected to reoccur frequently. This means that the soil across the study area will remain moist for a long duration. With regards to leptospirosis, this will form a fertile ground for *Leptospira* bacteria to breed.

In this study, the flood simulations were run with no stormwater drainage as the model constraint. Other outlets for water, such as culverts, were not considered due to the absence of the data. And the computational mesh extended to the boundary of the study area. Because of these constraints, the flood simulation overestimates water depths, especially towards the edges of the study area. The amount of overestimation can be estimated from historical flood data but this is currently unavailable for the study area. If storm drainage systems and outlets are considered and the model computation area is expanded, the high water-depths observed along the boundary of the study area is expected to decrease. So, as new data on the drainage system and outlets are made available, they should be used to refine the flood model and the exposure assessment that follows.

On the other hand, in this study, 0.5 m × 0.5 m area was represented by a single DSM and DEM raster cell. This fine scale resolution preserved small land features that otherwise would have been smoothed by coarser resolution terrain rasters. So, in this regard, the fine resolution terrain models provided more accurate estimates of water depths across the study area than a coarser resolution models would have. This is despite the fact that very small ground artifacts were also captured, which obstructed free flow of rainwater during the simulation. Ideally, a field survey at selected control points would be conducted to determine the existence and the type of obstructions and the model rerun to correct for any discrepancies. Since a field survey could not be conducted, this is recommended as the next step in the research progression.

In terms of contact of buildings and population with flood water, during heavy rainfalls, most of the buildings and people living in Alto do Cabrito are exposed. But there are accounts of local people using barriers to prevent water from rushing inside houses, pumps to take out water, and presence of some drainage networks to drain the storm water (CONDER, 2021; TV Bahia, 2017, 2021). If these are surveyed and mapped, the percentage of buildings and people in direct contact with flood water will decrease. Even so, after the 24-hr rainfall events, most of the buildings and population of Alto do Cabrito remain in close proximity to pooled water, which can provide breeding grounds for *Leptospira* bacteria. So, with respect to the exposure to *Leptospira*, the finding that pools of water form across the study area and a large percentage of buildings and people are in close proximity to the pooled water is an important finding.

According to the sero-survey data used in this study, it is uncertain that there is an association between leptospirosis and the flood. An association is more clearly ascertained when the exposed and the control groups are clearly differentiated. That differentiation was missing in this study because of the nature of the data available. If the proximity thresholds are ignored, everyone in Alto do Cabrito is exposed, because the entire study area is affected during the rainfall. Even when the proximity thresholds are considered, within a 2 m distance from the buildings, everyone is exposed to the flood water. So, the prevalence of leptospirosis sero-positive for the unexposed truly comes from the exposed population and not from a true unexposed population. Also, this study used the proximity thresholds to create groups of exposed and unexposed in Alto do Cabrito. But people who are identified as exposed by this method may truly be unexposed and those determined as unexposed may be exposed. So, the use of different proximity thresholds can lead to exposure misclassification. Since the misclassification is expected to occur regardless of the sero-survey status of the study participants, there is a non-differential misclassification of exposure. This type of exposure misclassification biases the measure of association (the OR) towards the null. This could explain the small

ORs observed in this study and the 95% CIs that contain the null values. Additionally, the limitation of this study is that a causal relationship cannot be established between the exposure and the outcome because these are determined simultaneously in this study. To establish causality, a prospective study design that records exposure statuses of study participants and follows the participants over time to observe if they develop the outcome or not should be used.

Lastly, the potential of GIS and EO to fill gaps in knowledge related to leptospirosis was one of the key motivations behind this research. The use of GIS in estimating the population of Alto do Cabrito as well as visualizing the results in both 2D and 3D improved the understanding of distribution of the population across the study area and how water pools across the study area and around buildings. The visualization of the height the water can rise to helped to identify buildings that could be inundated during a 24-hr intense rainfall. While PERSIANN rainfall data underestimated rainfall volume in the study area, it was useful to establish wet seasons and months to focus for leptospirosis outbreaks and surveys. When gauged rainfall data for the study area are available, the PERSIANN can be bias corrected and used further in the types of analyses presented here.

6. CONCLUSION

In conclusion, this study assessed the exposure to flood in a leptospirosis endemic slum area in Brazil by integrating 2D and 3D data. With an aim to select methods, the study synthesized common methods used to disaggregate population and model floods. Next, both 2D and 3D data were used to disaggregate the census-sector population to building-level population at Alto do Cabrito. A 2D hydraulic model that used terrain height information was used to simulate floods in the study area. The contact of the buildings and the population in Alto do Cabrito to flood water was assessed by using a distance-based method. Finally, the association between leptospirosis and flood was assessed.

A significant finding of this study is that even though water starts accumulating across Alto do Cabrito during high intensity 24-hr rainfall events, the roads quickly convey most of the rainwater from high to low elevations. At the end of the rainfall, the remaining water will form pools of varying depths across Alto do Cabrtio. Given the terrain of Alto do Cabrito and the characteristics of the rainfall, the seasonality of rainfall is important as it affects pooling of water for a longer duration than a single big storm. Given the terrain and rainfall-runoff characteristics, most of the buildings and the population in the study area come in contact with the pooled water. However, for the small area in Alto do Cabrtio for which leptospirosis sero-survey data was available, this study did not find an association between leptospirosis and the flood, indicating in principle that the finding holds for the entire population in Alto do Cabrito that is similarly impacted by flood.

7. LIMITATIONS AND RECOMMENDATIONS

7.1. Recommendations

With respect to the building footprints, the next logical step would be to correct the footprints using auxiliary data, such as community mapping, land parcels, and lots from 2010. With respect to the building heights, field surveys should be conducted to validate the building heights.

The 3D dasymetric method used in this study takes census sector population and distributes it to buildings in Alto do Cabrito. The method makes an intrinsic assumption that all buildings are residential buildings, every person is accounted for, and there is neither mobility nor migration of people. These assumptions may not be true. If the objective is to better understand and assess the exposure to *Leptospira* and the risk of leptospirosis, it is important to distinguish between residential and non-residential buildings as well as understand and integrate mobility and migration into 3D dasymetric mapping (Zhao et al., 2017).

In terms of water depth, field surveys to validate terrain obstructions and pooling of water should be conducted. Future research could add losses to the rainfall to account for absence of data on storm water drainage. The soil-related parameters, such as soil types, infiltration rate, and soil moisture are important for accurate flood modelling but are currently missing for the study area. Hence, documentation of hydrogeological parameters should be conducted.

With respect to the soil moisture, since the soil eventually dries after a rainfall event, the rate and the time that the soil takes to dry should be investigated. This aspect is important to account for *Leptospira* survival. Since the soil moisture is also affected by buildings that obstruct sunlight, shadow impact analysis may help to assess the soil moisture.

While the current research simplifies the exposure as contact with flood water, understanding if, when, and how the contact occurs is important. As such, in terms of human behaviour, people's conduct during and after heavy rainfall and flood should be examined. Equally important is the fact that people's perception of risk differs by demographic and socioeconomic factors (Becker et al., 2015). Therefore, it is important to understand who comes in contact with the flood and why.

As overland flow is transferred from higher elevation to lower elevation areas, sources of *Leptospira* at higher elevation, such as open garbage and sewage, can contribute to contamination of water and soil in the lower elevation. Thus, it is recommended to map possible sources of open garbage and sewage.

During this research, a need was observed to collect and organize data in a manner that is accessible to researchers. Hence, attention should be given to the development of a database system to store and manage primary and secondary data for the study area. Also, a need was seen to make available remotely sensed flood indicators for the study area as well as Salvador, Brazil. As remotely sensed data need to be calibrated and validated, ground truths are required. At present, each information is missing for the study area.

7.2. Limitations

In general, the accuracy of 3D dasymetric method is influenced by the accuracy of input data, such as the building footprints. If there are too few buildings, the population disaggregation will estimate large number of residents per building. If there are too many buildings than what is true, the population disaggregation will estimate lesser number of residents per building. Auxiliary data is, thus, required to correct the building footprints as well as validate the building-level population. So, the reliance on precise building footprint data as well as auxiliary data is a limitation of the 3D dasymetric method as these data may not be always available. In this context, the precise estimation of the building-level population in Alto do Cabrito was impeded, and hence, the estimated building-level population should be used with caution. Furthermore, the inaccuracies in building footprints affect flood simulation. The presence or absence of buildings will affect the flow of rainwater as well as the depths of collected water. All these together influence the estimated number and percentage of exposed buildings and population in Alto do Cabrito. Therefore, the results of this study are only applicable for the scenario presented in this study. Also, the true exposure and risk of Leptospirosis cannot be understood simply as contact of buildings and population with flood water. More environmental monitoring and socio-behavioural data is required to grasp the complexity of exposure in Alto do Cabrito. Therefore, at present the use of this research is limited to understanding possible points of contact that should be further explored.

LIST OF REFERENCES

Abela-Ridder, B., Sikkema, R., & Hartskeerl, R. A. (2010). Estimating the burden of human leptospirosis. *International Journal of Antimicrobial Agents, 36*, S5-S7. doi:https://doi.org/10.1016/j.ijantimicag.2010.06.012

Adda, P., Mioc, D., McGillivray, E., Morton, A., & Fraser, D. (2010). 3D Flood-Risk Models of Government Infrastructure. *ISPRS Conference Proceedings, Volume XXXVIII-4/W13*(Commission VI, WG VI/4). Retrieved from https://www.isprs.org/proceedings/XXXVIII/4-W13/ID_39.pdf

Addisu, A., Adriaensen, W., Balew, A., Asfaw, M., Diro, E., Garba Djirmay, A., . . . Ethiopia, S. I. T. N. T. D. G. (2019). Neglected tropical diseases and the sustainable development goals: an urgent call for action from the front line. *BMJ global health, 4*(1), e001334-e001334. doi:10.1136/bmjgh-2018-001334

Adgate, J. L. (2011). Exposure Modeling and Measurement: Exposure Factors. In J. O. Nriagu (Ed.), *Encyclopedia of Environmental Health* (pp. 630-635). Burlington: Elsevier.

Alahmadi, M., Atkinson, P., & Martin, D. (2013). Estimating the spatial distribution of the population of Riyadh, Saudi Arabia using remotely sensed built land cover and height data. *Computers, Environment and Urban Systems, 41*, 167-176. doi:https://doi.org/10.1016/j.compenvurbsys.2013.06.002

Alderman, K., Turner, L. R., & Tong, S. (2012). Floods and human health: A systematic review. *Environment International, 47*, 37-47. doi:10.1016/j.envint.2012.06.003

Amirebrahimi, S., Rajabifard, A., Mendis, P., & Ngo, T. (2015). A Data Model for Integrating GIS and BIM for Assessment and 3D Visualisation of Flood Damage to Building. *CEUR Workshop Proceedings, 1323*, 78-89.

Amirebrahimi, S., Rajabifard, A., Mendis, P., & Ngo, T. (2016). A framework for a microscale flood damage assessment and visualization for a building using BIM–GIS integration. *International Journal of Digital Earth, 9*(4), 363-386. doi:10.1080/17538947.2015.1034201

Amirebrahimi, S., Rajabifard, A., Mendis, P., Ngo, T., & Sabri, S. (2016). A Planning Decision Support Tool for Evaluation and 3d Visualisation of Building Risks in Flood Prone Areas. *FIG Working Week 2016: Recover from disaster.*

Amirebrahimi, S., Rajabifard, A., Sabri, S., & Mendis, P. (2016). Spatial Information in Support of 3D Flood Damage Assessment of Buildings at Micro Level: A Review. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-2/W1*, 73-81. doi:10.5194/isprs-annals-IV-2-W1-73-2016

Ansdell, V. E. (2017). Chapter 23 - Leptospirosis. In C. A. Sanford, P. S. Pottinger, & E. C. Jong (Eds.), *The Travel and Tropical Medicine Manual (Fifth Edition)* (pp. 336-344): Elsevier.

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., . . . Prat, O. P. (2015). PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bulletin of the American Meteorological Society, 96*(1), 69-83. doi:10.1175/BAMS-D-13-00068.1

Baquero, O. S., & Machado, G. (2018). Spatiotemporal dynamics and risk factors for human Leptospirosis in Brazil. *Scientific Reports, 8*(1), 15170. doi:10.1038/s41598-018-33381-3

Baquero, O. S., & Machado, G. (2019). Author Correction: Spatiotemporal dynamics and risk factors for human Leptospirosis in Brazil. *Scientific Reports, 9*(1), 7000. doi:10.1038/s41598-019-42747-0

Barcellos, C., & Sabroza, P. C. (2001). The place behind the case: leptospirosis risks and associated environmental conditions in a flood-related outbreak in Rio de Janeiro. *Cad Saude Publica, 17 Suppl*, 59-67. doi:10.1590/s0102-311x2001000700014

Barneveld, H., Silander, J., Sane, M., & Malnes, E. (2008). *Application of Satellite Data For Improved Flood Forecasting and Mapping.* Paper presented at the 4th International Symposium on Flood Defence: Managing Flood Risk, Reliability and Vulnerability, Toronto, Ontario, Canada.

Bates, P. D., & De Roo, A. P. J. (2000). A simple raster-based model for flood inundation simulation. *Journal of Hydrology, 236*(1), 54-77. doi:https://doi.org/10.1016/S0022-1694(00)00278-X

Becker, J. S., Taylor, H. L., Doody, B. J., Wright, K. C., Gruntfest, E., & Webber, D. (2015). A Review of People's Behavior in and around Floodwater. *Weather, Climate, and Society, 7*(4), 321-332. doi:10.1175/wcas-d-14-00030.1

Bharti, A. R., Nally, J. E., Ricaldi, J. N., Matthias, M. A., Diaz, M. M., Lovett, M. A., . . . Vinetz, J. M. (2003). Leptospirosis: a zoonotic disease of global importance. *The Lancet Infectious Diseases, 3*(12), 757-771. doi:https://doi.org/10.1016/S1473-3099(03)00830-2

Bierque, E., Thibeaux, R., Girault, D., Soupé-Gilbert, M. E., & Goarant, C. (2020). A systematic review of Leptospira in water and soil environments. *PloS one, 15*(1), e0227055. doi:10.1371/journal.pone.0227055

Biljecki, F., Ohori, K. A., Ledoux, H., Peters, R., & Stoter, J. (2016). Population estimation using a 3D City Model: A multi-scale country-wide study in the Netherlands. *PLOS ONE, 11*(6). doi:10.1371/journal.pone.0156808

Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., & Çöltekin, A. (2015). Applications of 3D City Models: State of the Art Review. *ISPRS International Journal of Geo-Information, 4*(4), 2842-2889. doi:10.3390/ijgi4042842

Brandão, T., Santos, R., & Carelli, L. (2016a). Eventos Hidrológicos Extremos na Cidade de Salvador-BA: Análise Espacial de Ocorrências de Alagamentos. *Investigaciones Geográficas*. doi:10.5354/0719-5370.2016.41433

Brandão, T., Santos, R., & Carelli, L. (2016b). Eventos Hidrológicos Extremos na Cidade de Salvador-BA: Análise Espacial de Ocorrências de Alagamentos [Extreme Hydrological Events in the City of Salvador-BA: Spatial Analysis of Flood Occurrences]. *Investigaciones Geográficas*. doi:10.5354/0719-5370.2016.41433

Brazilian Institute of Geography and Statistics (IBGE). (2012). Brazil Demographic Census 2010. *Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).*

Briskin, E. A., Casanovas-Massana, A., Ryff, K. R., Morales-Estrada, S., Hamond, C., Perez-Rodriguez, N. M., . . . Ko, A. I. (2019). Seroprevalence, Risk Factors, and Rodent Reservoirs of Leptospirosis in an Urban Community of Puerto Rico, 2015. *J Infect Dis, 220*(9), 1489-1497. doi:10.1093/infdis/jiz339

Brito, P. (2010). *Remote sensing in the identification of urban elements and typologies related to the occurrence of leptospirosis in the railway suburb of Salvador, Bahia.* (PhD). University of São Paulo, São Paulo. Retrieved from https://teses.usp.br/teses/disponiveis/3/3138/tde-18082010-170309/pt-br.php

Brito, P., Kuffer, M., Koeva, M., Pedrassoli, J. C., Wang, J., Costa, F., & Freitas, A. D. d. (2020). The Spatial Dimension of COVID-19: The Potential of Earth Observation Data in Support of Slum Communities with Evidence from Brazil. *ISPRS International Journal of Geo-Information, 9*(9), 557. Retrieved from https://www.mdpi.com/2220-9964/9/9/557

Calderón-Sierra, D. M., Jaimes-Bernal, C. P., & Pedraza-Bernal, A. M. (2019). Epidemiology of human leptospirosis in Colombia, 2012-2016. *Revista Cubana de Medicina Tropical, 71*(2). Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85075429951&partnerID=40&md5=0d6feb7019b9e00d7957c87e21a9f137

Chadsuthi, S., Modchang, C., Lenbury, Y., Iamsirithaworn, S., & Triampo, W. (2012). Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time–series and ARIMAX analyses. *Asian Pacific Journal of Tropical Medicine, 5*(7), 539-546. doi:https://doi.org/10.1016/S1995-7645(12)60095-9

Chakraborty, J., & Maantay, J. A. (2011). Proximity analysis for exposure assessment in environmental health justice research. In *Geospatial analysis of environmental health* (pp. 111-138): Springer.

Chen, A. S., Evans, B., Djordjević, S., & Savić, D. A. (2012). Multi-layered coarse grid modelling in 2D urban flood simulations. *Journal of Hydrology, 470-471*, 1-11. doi:https://doi.org/10.1016/j.jhydrol.2012.06.022

Cohen, J. M., Wilson, M. L., & Aiello, A. E. (2007). Analysis of social epidemiology research on infectious diseases: historical patterns and future opportunities. *Journal of epidemiology and community health, 61*(12), 1021-1027. doi:10.1136/jech.2006.057216

CONDER. (2021). CONDER's macro-drainage project benefits another community on the Dique do Cabrito. Retrieved from http://www.conder.ba.gov.br/noticias/2021-05-21/obra-de-macrodrenagem-da-conder-beneficia-mais-uma-comunidade-do-dique-do

Costa, F., Hagan, J. E., Calcagno, J., Kane, M., Torgerson, P., Martinez-Silveira, M. S., . . . Ko, A. I. (2015). Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLOS Neglected Tropical Diseases, 9*(9), e0003898. doi:10.1371/journal.pntd.0003898

Costabile, P., Costanzo, C., De Lorenzo, G., De Santis, R., Penna, N., & Macchione, F. (2021). Terrestrial and airborne laser scanning and 2-D modelling for 3-D flood hazard maps in urban areas: new opportunities and perspectives. *Environmental Modelling & Software, 135*, 104889. doi:https://doi.org/10.1016/j.envsoft.2020.104889

Criado, M., Martínez-Graña, A., San Román, J. S., & Santos-Francés, F. (2019). Flood Risk Evaluation in Urban Spaces: The Study Case of Tormes River (Salamanca, Spain). *International Journal of Environmental Research and Public Health, 16*(1), 5. Retrieved from https://www.mdpi.com/1660-4601/16/1/5

Cucchi, K., Liu, R., Collender, P. A., Cheng, Q., Li, C., Hoover, C. M., . . . Remais, J. V. (2019). Hydroclimatic drivers of highly seasonal leptospirosis incidence suggest prominent soil reservoir of pathogenic Leptospira spp. in rural western China. *PLOS Neglected Tropical Diseases, 13*(12), e0007968. doi:10.1371/journal.pntd.0007968

Cunha, M., Costa, F., Ribeiro, G. S., Carvalho, M. S., Reis, R. B., Júnior, N. N., . . . Ko, A. I. (2019). Rainfall and other meteorological factors as drivers of urban transmission of leptospirosis. *bioRxiv*, 658872. doi:10.1101/658872

de Almeida, G. A. M., Bates, P., & Ozdemir, H. (2018). Modelling urban floods at submetre resolution: challenges or opportunities for flood risk management? *Journal of Flood Risk Management, 11*(S2), S855-S865. doi:https://doi.org/10.1111/jfr3.12276

De Francesco Daher, E., de Carvalho, G. S. G., de Sousa Soares, D., Mendes, M. H., Parente Filho, S. L. A., Rocha, H. A. L., & da Silva Junior, G. B. (2017). Changing patterns in leptospirosis: a three-decade study in Brazil. *International Journal of Infectious Diseases, 60*, 4-10. doi:https://doi.org/10.1016/j.ijid.2017.04.023

De Santis, R., Macchione, F., Costabile, P., & Costanzo, C. (2018). A comparative analysis of 3-D representations of urban flood map in virtual environments for hazard communication purposes. *E3S Web Conf., 40*, 06037. Retrieved from https://doi.org/10.1051/e3sconf/20184006037

De Santis, R., Macchione, F., Costabile, P., & Costanzo, C. (2019). *Visual Risk Communication of Urban Flooding in 3D Environments Based on Terrestrial Laser Scanning*, Cham.

Dhewantara, P. W., Hu, W., Zhang, W., Yin, W. W., Ding, F., Mamun, A. A., & Soares Magalhães, R. J. (2019). Climate variability, satellite-derived physical environmental data and human leptospirosis: A retrospective ecological study in China. *Environmental Research, 176*. doi:10.1016/j.envres.2019.06.004

Dhewantara, P. W., Zhang, W., Al Mamun, A., Yin, W.-W., Ding, F., Guo, D., . . . Soares Magalhães, R. J. (2020). Spatial distribution of leptospirosis incidence in the Upper Yangtze and Pearl River Basin, China: Tools to support intervention and elimination. *Science of The Total Environment, 725*, 138251. doi:https://doi.org/10.1016/j.scitotenv.2020.138251

Dirk, R. (2013). Frequency Analysis of Rainfall Data. In: College on Soil Physics— 30th Anniversary (1983-2013), The Abdus Salam International Centre for Theoretical Physica, 244-10. In.

Eicher, C. L., & Brewer, C. A. (2001). Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science, 28*(2), 125-138. doi:10.1559/152304001782173727

Eichhorn, S. (2020). Disaggregating Population Data and Evaluating the Accuracy of Modeled High-Resolution Population Distribution—The Case Study of Germany. *Sustainability, 12*(10), 3976. Retrieved from https://www.mdpi.com/2071-1050/12/10/3976

Elfouly, M., & Labetski, A. (2020). Flood damage cost estimation in 3D based on an indicator modelling framework. *Geomatics, Natural Hazards and Risk, 11*(1), 1129-1153. doi:10.1080/19475705.2020.1777213

EXCIMAP. (2007). Handbook on good practices for flood mapping in Europe. Retrieved from https://ec.europa.eu/environment/water/flood_risk/flood_atlas/pdf/handbook_goodpractice.pdf

Felzemburgh, R. D. M., Ribeiro, G. S., Costa, F., Reis, R. B., Hagan, J. E., Melendez, A. X. T. O., . . . Ko, A. I. (2014). Prospective Study of Leptospirosis Transmission in an Urban Slum Community: Role of Poor Environment in Repeated Exposures to the Leptospira Agent. *PLOS Neglected Tropical Diseases, 8*(5). doi:10.1371/journal.pntd.0002927

Ferreira, M. C. (2016). Influence Of Topographic And Hydrographic Factors On The Spatial Distribution Of Leptospirosis Disease In Sao Paulo County, Brazil: An Approach Using Geospatial

Techniques And Gis Analysis. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Retrieved from https://www.scilit.net/article3a036912ee459d4f3166682c7a46315a

Fetter, C. W. (2001). Applied Hydrogeology (4th ed.). In (pp. 50). Upper Saddle River, NJ: Prentice-Hall.

Fiorio, P. R., Duarte, S. N., Rodrigues, G. d. O., Miranda, J. H. d., & Cooke, R. A. (2012). Comparação de equações de chuvas intensas para localidades do estado de São Paulo. *Engenharia Agrícola, 32*, 1080-1088.

Foudi, S. (2013). *Flood risk management: what are the main drivers of prevention?*

Fuchs, S., Keiler, M., & Zischg, A. (2015). A spatiotemporal multi-hazard exposure assessment based on property data. *Nat. Hazards Earth Syst. Sci., 15*(9), 2127-2142. doi:10.5194/nhess-15-2127-2015

Garba, B., Bahaman, A. R., Bejo, S. K., Zakaria, Z., Mutalib, A. R., & Bande, F. (2018). Major epidemiological factors associated with leptospirosis in Malaysia. *Acta Tropica, 178*, 242-247. doi:10.1016/j.actatropica.2017.12.010

GPRH. (n.d). Softwares. Retrieved from http://www.gprh.ufv.br/?area=softwares

Gracie, R., Barcellos, C., Magalhães, M., Souza-Santos, R., & Barrocas, P. R. G. (2014). Geographical scale effects on the analysis of leptospirosis determinants. *International journal of environmental research and public health, 11*(10), 10366-10383.

Hacker, K. P., Sacramento, G. A., Cruz, J. S., de Oliveira, D., Nery, N., Jr., Lindow, J. C., . . . Costa, F. (2020). Influence of Rainfall on Leptospira Infection and Disease in a Tropical Urban Setting, Brazil. *Emerg Infect Dis, 26*(2), 311-314. doi:10.3201/eid2602.190102

Hagan, J. E., Moraga, P., Costa, F., Capian, N., Ribeiro, G. S., Wunder, E. A., Jr., . . . Ko, A. I. (2016). Spatiotemporal Determinants of Urban Leptospirosis Transmission: Four-Year Prospective Cohort Study of Slum Residents in Brazil. *PLOS Neglected Tropical Diseases, 10*(1), e0004275. doi:10.1371/journal.pntd.0004275

Halliday, J. E. B., Knobel, D. L., Allan, K. J., de, C. B. B. M., Handel, I., Agwanda, B., . . . Breiman, R. F. (2013). Urban leptospirosis in Africa: a cross-sectional survey of Leptospira infection in rodents in the Kibera urban settlement, Nairobi, Kenya. *Am J Trop Med Hyg, 89*(6), 1095-1102. doi:10.4269/ajtmh.13-0415

Hamm, N. A. S., Soares Magalhães, R. J., & Clements, A. C. A. (2015). Earth Observation, Spatial Data Quality, and Neglected Tropical Diseases. *PLOS Neglected Tropical Diseases, 9*(12), e0004164. doi:10.1371/journal.pntd.0004164

Hartskeerl, R. A., Collares-Pereira, M., & Ellis, W. A. (2011). Emergence, control and re-emerging leptospirosis: dynamics of infection in the changing world. *Clinical Microbiology and Infection, 17*(4), 494-501. doi:https://doi.org/10.1111/j.1469-0691.2011.03474.x

Herman, L., Russnak, J., & Řezník, T. (2017). Flood Modelling and Visualizations of Floods Through 3D Open Data. In (pp. 139-149).

IHME. (2021, 22 Jun, 2021). Brazil Demographic Census 2010. Retrieved from http://ghdx.healthdata.org/record/brazil-demographic-census-2010

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature, 451*(7181), 990-993. doi:10.1038/nature06536

Jung, I.-K., Park, J.-Y., Park, G.-A., Lee, M.-S., & Kim, S.-J. (2011). A grid-based rainfall-runoff model for flood simulation including paddy fields. *Paddy and Water Environment, 9*(3), 275-290. doi:10.1007/s10333-010-0232-4

Kawaguchi, L., Sengkeopraseuth, B., Tsuyuoka, R., Koizumi, N., Akashi, H., Vongphrachanh, P., . . . Aoyama, A. (2008). Seroprevalence of Leptospirosis and Risk Factor Analysis in Flood-prone Rural Areas in Lao PDR. *The American Journal of Tropical Medicine and Hygiene, 78*(6), 957-961. doi:https://doi.org/10.4269/ajtmh.2008.78.957

Khalil, H., Santana, R., de Oliveira, D., Palma, F., Lustosa, R., Eyre, M. T., . . . Costa, F. (2021). Poverty, sanitation, and Leptospira transmission pathways in residents from four Brazilian slums. *PLOS Neglected Tropical Diseases, 15*(3), e0009256. doi:10.1371/journal.pntd.0009256

Khoury, M., Gibson, M., Chen, A., Savic, D., Vamvakeridou-Lyroudia, L., & Djordjevic, S. (2017). *A 3D Web GIS Interactive Visualisation System for Animated Floods*. Paper presented at the CCWI2017 - Computing and Control for the Water Industry: F73, Sheffield.

Ko, A. I., Reis, M. G., Dourado, C. M. R., Johnson, W. D., & Riley, L. W. (1999). Urban epidemic of severe leptospirosis in Brazil. *The Lancet, 354*(9181), 820-825. doi:https://doi.org/10.1016/S0140-6736(99)80012-9

Kolbe, T., Gröger, G., & Plümer, L. (2005). CityGML: Interoperable Access to 3D City Models. In v. O. P., Z. S., & F. E.M. (Eds.), *Geo-information for Disaster Management*: Springer, Berlin, Heidelberg.

Kuffer, M., Persello, C., Pfeffer, K., Sliuzas, R., & Rao, V. (2019, 22-24 May 2019). *Do we underestimate the global slum population?* Paper presented at the 2019 Joint Urban Remote Sensing Event (JURSE).

Kumar, K., Ledoux, H., Commandeur, T. J. F., & Stoter, J. E. (2017). Modelling urban noise in CityGML ADE: case of the Netherlands. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-4/W5*, 73-81. doi:10.5194/isprs-annals-IV-4-W5-73-2017

Kumar, K., Ledoux, H., & Stoter, J. (2018). Dynamic 3d Visualization of Floods : Case of the Netherlands. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W10*, 83-87. Retrieved from https://d-nb.info/1166853438/34

Kurakula, V., Skidmore, A., Kluijver, H. D., Stoter, J., Dabrowska-Zielinska, K., & Kuffer, M. (2007). *A GIS-Based approach for 3D noise modelling using 3D city models.*

Lai, P.-C., Low, C.-T., Tse, W.-S. C., Tsui, C.-K., Lee, H., & Hui, P.-K. (2013). Risk of tuberculosis in high-rise and high density dwellings: An exploratory spatial analysis. *Environmental Pollution, 183*, 40-45. doi:https://doi.org/10.1016/j.envpol.2012.11.025

Langford, M. (2006). Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems, 30*(2), 161-180. doi:https://doi.org/10.1016/j.compenvurbsys.2004.07.001

Lánský, I., Ceccarelli, G., Mastorakis, K., Jongh, W. D., & Li, J. (2019). *Influence of 3D City Layout on Air Quality.*

Lau, C. L., Smythe, L. D., Craig, S. B., & Weinstein, P. (2010). Climate change, flooding, urbanisation and leptospirosis: fuelling the fire? *Transactions of the Royal Society of Tropical Medicine and Hygiene, 104*(10), 631-638. doi:https://doi.org/10.1016/j.trstmh.2010.07.002

Lecture. (2009). Frequency Analysis for Hydrologic Modelling. WRS – ITC. In.

Ledien, J., Sorn, S., Hem, S., Huy, R., Buchy, P., Tarantola, A., & Cappelle, J. (2017). Assessing the performance of remotely-sensed flooding indicators and their potential contribution to early warning for leptospirosis in Cambodia. *PloS one, 12*(7), e0181044-e0181044. doi:10.1371/journal.pone.0181044

Lee, S.-H., Park, J., Park, S. I., & Jang, Y.-H. (2016). *An Urban Flooding Simulation Technique by Using 3D City Information Model.* Paper presented at the Proceedings of The 7th Civil Engineering Conference in the Asian Region (CECAR 7), Waikiki, Hawaii, USA.

Linard, C., & Tatem, A. J. (2012). Large-scale spatial population databases in infectious disease research. *International Journal of Health Geographics, 11*(1), 7. doi:10.1186/1476-072X-11-7

López, M. S., Müller, G. V., Lovino, M. A., Gómez, A. A., Sione, W. F., & Aragonés Pomares, L. (2019). Spatio-temporal analysis of leptospirosis incidence and its relationship with hydroclimatic indicators in northeastern Argentina. *Science of The Total Environment, 694*, 133651. doi:https://doi.org/10.1016/j.scitotenv.2019.133651

Lu, Z., Im, J., & Quackenbush, L. (2011). A Volumetric Approach to Population Estimation Using Lidar Remote Sensing. *Photogrammetric Engineering & Remote Sensing, 77*(11), 1145-1156. doi:10.14358/PERS.77.11.1145

Lu, Z., Im, J., Quackenbush, L., & Halligan, K. (2010). Population estimation based on multi-sensor data fusion. *International Journal of Remote Sensing, 31*(21), 5587-5604. doi:10.1080/01431161.2010.496801

Machado, R. A. S., Oliveira, A. G., & Lois-González, R. C. (2019). Urban ecological infrastructure: The importance of vegetation cover in the control of floods and landslides in Salvador / Bahia, Brazil. *Land Use Policy, 89*, 104180. doi:https://doi.org/10.1016/j.landusepol.2019.104180

Maksimović, Č., Prodanović, D., Boonya-Aroonnet, S., Leitão, J. P., Djordjević, S., & Allitt, R. (2009). Overland flow and pathway analysis for modelling of urban pluvial flooding. *Journal of Hydraulic Research, 47*(4), 512-523. doi:10.1080/00221686.2009.9522027

Malloy, S. S., Horack, J. M., Lee, J., & Newton, E. K. (2019). Earth observation for public health: Biodiversity change and emerging disease surveillance. *Acta Astronautica, 160*, 433-441. doi:https://doi.org/10.1016/j.actaastro.2018.10.042

Marcus, W. A., Roberts, K., Harvey, L., & Tackman, G. (1992). An Evaluation of Methods for Estimating Manning's n in Small Mountain Streams. *Mountain Research and Development, 12*(3), 227-239. doi:10.2307/3673667

Maroko, A., Maantay, J., Pérez Machado, R. P., & Barrozo, L. V. (2019). Improving Population Mapping and Exposure Assessment: Three-Dimensional Dasymetric Disaggregation in New York City and São Paulo, Brazil. *Papers in Applied Geography, 5*(1-2), 45-57. doi:10.1080/23754931.2019.1619092

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association, 46*(253), 68-78. doi:10.1080/01621459.1951.10500769

Mennis, J. (2003). Generating Surface Models of Population Using Dasymetric Mapping. *The Professional Geographer, 55*(1), 31-42. doi:10.1111/0033-0124.10042

Micah, A. E., Su, Y., Bachmeier, S. D., Chapin, A., Cogswell, I. E., Crosby, S. W., . . . Dieleman, J. L. (2020). Health sector spending and spending on HIV/AIDS, tuberculosis, and malaria, and development assistance for health: progress towards Sustainable Development Goal 3. *The Lancet*. doi:https://doi.org/10.1016/S0140-6736(20)30608-5

Mioc, D., Anton, F., Nickerson, B., Santos, M., Adda, P., Tienaah, T., . . . Morton, A. (2011). Flood Progression Modelling and Impact Analysis. In *Efficient Decision Support Systems-Practice and Challenges in Multidisciplinary Domains*: IntechOpen.

Mohd Radi, M. F., Hashim, J. H., Jaafar, M. H., Hod, R., Ahmad, N., Mohammed Nawi, A., . . . Farakhin Ayub, N. I. (2018). Leptospirosis Outbreak After the 2014 Major Flooding Event in Kelantan, Malaysia: A Spatial-Temporal Analysis. *The American Journal of Tropical Medicine and Hygiene, 98*(5), 1281-1295. doi:https://doi.org/10.4269/ajtmh.16-0922

Municipality of Salvador. CDGV Salvador Preliminar – Edificacao. Retrieved from http://cartografia.salvador.ba.gov.br

National Academies of Sciences, E., & Medicine. (2018). *Urbanization and Slums: Infectious Diseases in the Built Environment: Proceedings of a Workshop*. Washington, DC: The National Academies Press.

Oosterbaan, R. J. (1994). Frequency and Regression Analysis. In H.P.Ritzema (Ed.). Drainage Principles and Applications, Publication 16, Second Revised Edition. International Institute for Land Reclamation andImprovement (ILRI), Wageningen, The Netherlands. ISBN 90 70754 3 39. In.

Papaioannou, G., Efstratiadis, A., Vasiliades, L., Loukas, A., Papalexiou, S. M., Koukouvinos, A., . . . Kossieris, P. (2018). An Operational Method for Flood Directive Implementation in Ungauged Urban Areas. *Hydrology, 5*(2), 24. Retrieved from https://www.mdpi.com/2306-5338/5/2/24

Parselia, E., Kontoes, C., Tsouni, A., Hadjichristodoulou, C., Kioutsioukis, I., Magiorkinis, G., & Stilianakis, N. I. (2019). Satellite Earth Observation Data in Epidemiological Modeling of Malaria, Dengue and West Nile Virus: A Scoping Review. *Remote Sensing, 11*(16), 1862.

Reis, R. B., Ribeiro, G. S., Felzemburgh, R. D. M., Santana, F. S., Mohr, S., Melendez, A. X. T. O., . . . Ko, A. I. (2008). Impact of Environment and Social Gradient on Leptospira Infection in Urban Slums. *PLOS Neglected Tropical Diseases, 2*(4), e228. doi:10.1371/journal.pntd.0000228

Romilly, T. G., & Gebremichael, M. (2011). Evaluation of satellite rainfall estimates over Ethiopian river basins. *Hydrol. Earth Syst. Sci., 15*(5), 1505-1514. doi:10.5194/hess-15-1505-2011

Röthlisberger, V., Zischg, A. P., & Keiler, M. (2017). Identifying spatial clusters of flood exposure to support decision making in risk management. *Science of The Total Environment, 598*, 593-603. doi:https://doi.org/10.1016/j.scitotenv.2017.03.216

Rubinato, M., Nichols, A., Peng, Y., Zhang, J.-m., Lashford, C., Cai, Y.-p., . . . Tait, S. (2019). Urban and river flooding: Comparison of flood risk management approaches in the UK and China and an assessment of future knowledge needs. *Water Science and Engineering, 12*(4), 274-283. doi:https://doi.org/10.1016/j.wse.2019.12.004

Santos, A. P. P. d., Aragão, M. R. d. S., Correia, M. d. F., Santos, S. R. Q. d., Silva, F. D. d. S., & Araújo, H. A. d. (2016). Precipitação na Cidade de Salvador: Variabilidade Temporal e Classificação em Quantis. *Revista Brasileira de Meteorologia, 31*, 454-467. Retrieved from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-77862016000800454&nrm=iso

Santos, N. J., Sousa, E., Reis, M. G., Ko, A. I., & Costa, F. (2017). Rat infestation associated with environmental deficiencies in an urban slum community with high risk of leptospirosis transmission. *Cad Saude Publica, 33*(2), e00132115. doi:10.1590/01021-311x00132115

Schacht, R. M. (1980). Two models of population growth. *American Anthropologist, 82*(4), 782-798.

Shih, M., Chang, H.-t. B., & Popper, F. J. (2018). Development Rights: Regulating Vertical Urbanism in Taiwan. *Planning Theory & Practice, 19*(5), 717-733. doi:10.1080/14649357.2018.1535085

Silva, E., Guerrero, V. M., & Peña, D. (2011). Temporal disaggregation and restricted forecasting of multiple population time series. *Journal of Applied Statistics, 38*(4), 799-815. doi:10.1080/02664761003692316

Skouloudis, A. N., & Rickerby, D. G. (2015). In-situ and Remote Sensing Networks for Environmental Monitoring and Global Assessment of Leptospirosis Outbreaks. *Procedia Engineering, 107*, 194-204. doi:https://doi.org/10.1016/j.proeng.2015.06.074

Smith, J. K. G., Young, M. M., Wilson, K. L., & Craig, S. B. (2013). Leptospirosis following a major flood in Central Queensland, Australia. *Epidemiology and Infection, 141*(3), 585-590. doi:10.1017/S0950268812001021

Socolovschi, C., Angelakis, E., Renvoisé, A., Fournier, P.-E., Marié, J. L., Davoust, B., . . . Raoult, D. (2011). Strikes, flooding, rats, and leptospirosis in Marseille, France. *International Journal of Infectious Diseases, 15*(10), e710-e715. doi:https://doi.org/10.1016/j.ijid.2011.05.017

Sorooshian, S., Hsu, K., Braithwaite, D., Ashouri, H., & NOAA CDR Program. (2014). *NOAA Climate Data Record (CDR) of Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN-CDR), Version 1 Revision 1. NOAA National Centers for Environmental Information.*

Souza, I. P. O., Uberti, M. S., & Tassinari, W. S. (2020). Geoprocessing and spatial analysis for identifying leptospirosis risk areas: a systematic review. *Rev Inst Med Trop Sao Paulo, 62*, e35. doi:10.1590/s1678-9946202062035

Sridharan, H., & Qiu, F. (2013). A Spatially Disaggregated Areal Interpolation Model Using Light Detection and Ranging-Derived Building Volumes. *Geographical Analysis, 45*(3), 238-258. doi:https://doi.org/10.1111/gean.12010

The World Bank Group. (n.d., 2021). Population growth indicator. Retrieved from https://data.worldbank.org/indicator/SP.POP.GROW?locations=BR

TV Bahia. (2017). Homem usa placa de isopor para atravessar rua alagada em Salvador. Retrieved from http://g1.globo.com/bahia/noticia/2017/03/homem-usa-placa-de-isopor-para-atravessar-rua-alagada-em-salvador.html

TV Bahia. (2021). Chuva invade casas, deixa pessoas ilhadas e provoca danos em Salvador e RMS: 'Mais de R$ 10 mil de prejuízo'. *Globo*. Retrieved from https://g1.globo.com/ba/bahia/noticia/2021/04/08/chuva-invade-casas-deixa-pessoas-ilhadas-e-provoca-danos-em-salvador-e-rms-mais-de-r-10-mil-de-prejuizo.ghtml

Ujang, U., Azri, S., Zahir, M., Rahman, A. A., & Choon, T. L. (2018). Urban Heat Island Micro-Mapping via 3d City Model. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 6210*, 201-207.

United Nations. (2015). *Transforming our World: The 2030 Agenda for Sustainable Development. A/RES/70/1.* Retrieved from https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf

USACE. (2010). HEC-RAS River Analysis System: Hydraulic Reference Manual. Version 4.1. Retrieved from https://www.hec.usace.army.mil/software/hec-ras/documentation/HEC-RAS_4.1_Reference_Manual.pdf

Van Ackere, S., Glas, H., Beullens, J., Deruyter, G., Wulf, A. D., & Maeyer, P. D. (2016). *Development Of A 3D Dynamic Flood Web Gis Visualisation Tool.* Paper presented at the 5th International conference on Flood Risk Management and Response (FRIAR 2016), Venice, Italy. https://www.witpress.com/elibrary/sse-volumes/6/3/1350

Van Ackere, S., Glas, H., Vandenbulcke, A., Stal, C., Decock, M., Deruyter, G., . . . Maeyer, P. D. (2016). *4D Coastal Flood Analysis and Visualisation for Urban Areas.* Paper presented at the 6th International Conference on Cartography and GIS, 1, Chr. Smirnenski Blvd. Sofia, Bulgaria.

Vanasco, N. B., Schmeling, M. F., Lottersberger, J., Costa, F., Ko, A. I., & Tarabla, H. D. (2008). Clinical characteristics and risk factors of human leptospirosis in Argentina (1999–2005). *Acta Tropica, 107*(3), 255-258. doi:https://doi.org/10.1016/j.actatropica.2008.06.007

Wang, S., Tian, Y., Zhou, Y., Liu, W., & Lin, C. (2016). Fine-Scale Population Estimation by 3D Reconstruction of Urban Residential Buildings. *Sensors, 16*(10), 1755. Retrieved from https://www.mdpi.com/1424-8220/16/10/1755

Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., . . . Tatem, A. J. (2018). Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences, 115*(14), 3529-3537. doi:10.1073/pnas.1715305115

Wästberg, B. S., Tornberg, J., Billger, M., Haeger-Eugensson, M., & Sjr berg, K. (2013). *How to Visualize the Invisible Simulating Air Pollution Dispersions in a 3D City Model.*

Wei, S., & Ji, S. (2021). Graph Convolutional Networks for the Automated Production of Building Vector Maps From Aerial Images. *IEEE Transactions on Geoscience and Remote Sensing, PP*, 1-11. doi:10.1109/TGRS.2021.3060770

Wei, S., Ji, S., & Lu, M. (2019). Toward automatic building footprint delineation from aerial images using cnn and regularization. *IEEE Transactions on Geoscience and Remote Sensing, 58*(3), 2178-2189.

WHO. (2011). World Health Organization Geneva: Leptospirosis: an emerging public health problem. *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire, 86*(06), 45-50.

World Population Review. (n.d). Salvador Population 2021. Retrieved from https://worldpopulationreview.com/world-cities/salvador-population

Xavier, A. C., Cecílio, R. A., Pruski, F. F., & Lima, J. S. d. S. (2014). Methodology for spatialization of intense rainfall equation parameters. *Engenharia Agrícola, 34*, 485-495.

Xu, M., Cao, C., & Jia, P. (2020). Mapping Fine-Scale Urban Spatial Population Distribution Based on High-Resolution Stereo Pair Images, Points of Interest, and Land Cover Data. *Remote Sensing, 12*(4), 608. Retrieved from https://www.mdpi.com/2072-4292/12/4/608

Zahran, E.-S. M. M., Smith, M. J., & Bennett, L. D. (2013). 3D Visualization of Traffic-Induced Air Pollution Impacts of Urban Transport Schemes. *Journal of Computing in Civil Engineering, 27*, 452-465.

Zhang, S., Han, F., & Bogus, S. M. (2020). Building Footprint and Height Information Extraction from Airborne LiDAR and Aerial Imagery. In *Construction Research Congress 2020* (pp. 326-335).

Zhang, T., Yang, Y., Dong, Z., & Gui, S. (2021). A Multiscale Assessment of Three Satellite Precipitation Products (TRMM, CMORPH, and PERSIANN) in the Three Gorges Reservoir Area in China. *Advances in Meteorology, 2021*, 9979216. doi:10.1155/2021/9979216

Zhao, Y., Ovando-Montejo, G. A., Frazier, A. E., Mathews, A. J., Flynn, K. C., & Ellis, E. A. (2017). Estimating work and home population using lidar-derived building volumes. *International Journal of Remote Sensing, 38*(4), 1180-1196. doi:10.1080/01431161.2017.1280634

8. APPENDIX

8.1. Bias correction of PERSIANN rainfall data

Figure A. 1 shows the time series of P_{Ondina}, P_{Ondina}, gauged and interpolated (combined) data, and OWS before bias correction. Since there is a perfect overlap between OWS and combined data, only the time series for combined data is visible in both figures. The comparison of the P_{Ondina} with the OWS rainfall shows high variation, both in the magnitude and the shape of the curves. The associated root mean square error (RMSE) is 231.5 mm, indicating large error in P_{Ondina}. In fact, at many points, P_{ADC} is closer in magnitude and shape to OWS than P_{Ondina}. No matter what, if the OWS is correct, PERSIANN consistently underestimates the rainfall in Salvador.

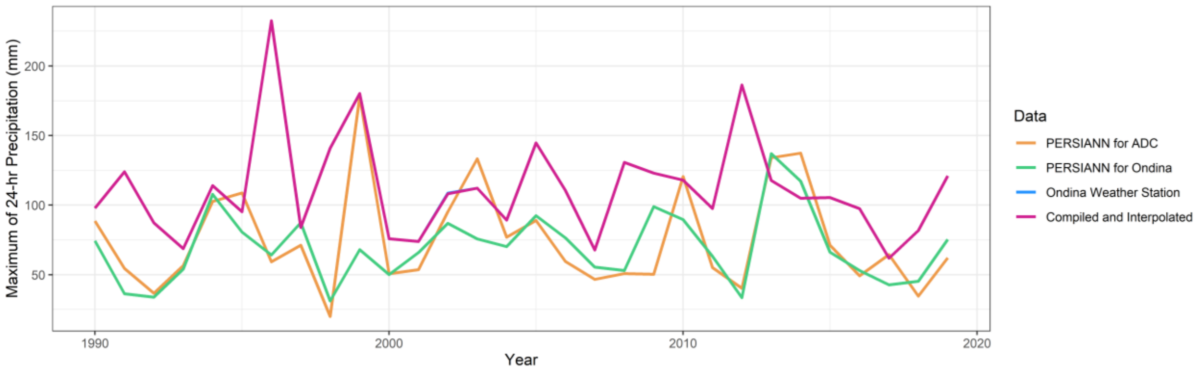


Figure A. 1. Time series of bias uncorrected monthly maximum of 24-hr precipitation

As discussed in the section **3.5**, three different bias correction methods were examined. **Figure A. 2** shows how the application of these methods on P_{Ondina} affects its time series. The bias correction methods BC1 and BC3 scale P_{Ondina}'s time series – the shape of the time series remains the same, but the magnitude is altered. The application of the difference in mean (BC1) increases the magnitude of P_{Ondina} throughout, whereas the application of the ratio of observed to satellite rainfall (BC3) decreases the magnitude. On the other hand, the application of the difference between P_{Ondina} and OWS reproduces the time series of OWS. This is the reason in **Figure A. 2**, P_{Ondina} corrected by BC2 method perfectly overlaps with the OWS time series. The associated RMSE are shown in (**Table A 1**). As would be expected, the RMSE for BC2 is 0. It is infinitely small for BC1 and rather large for BC3.

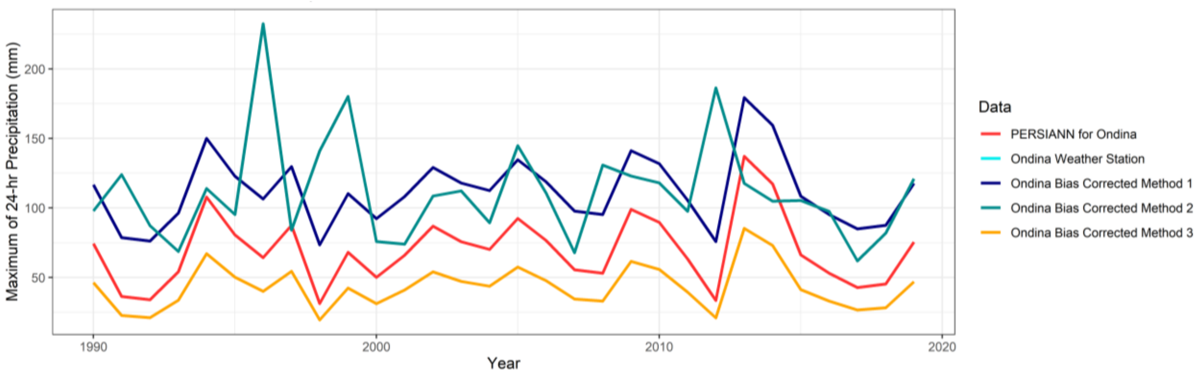


Figure A. 2. Time series plot of bias corrected P_{Ondina}

Table A 1. RMSE for bias corrected P_{Ondina}

BC1 (mm)	BC2 (mm)	BC3 (mm)
1.69×10^{-14}	0	375.4

Figure A. 3 shows the comparison of the bias corrected time series for P_{ADC} with the other rainfall data. As seen with P_{Ondina}, the BC1 and BC3 scale the magnitude of P_{ADC} keeping the shape intact. On the other hand, BC2 changes the magnitude and the shape of P_{ADC}. However, without the gauged rainfall data for Alto do Cabrito, it cannot be said with confidence that BC2 produces a reliable bias corrected P_{ADC}.

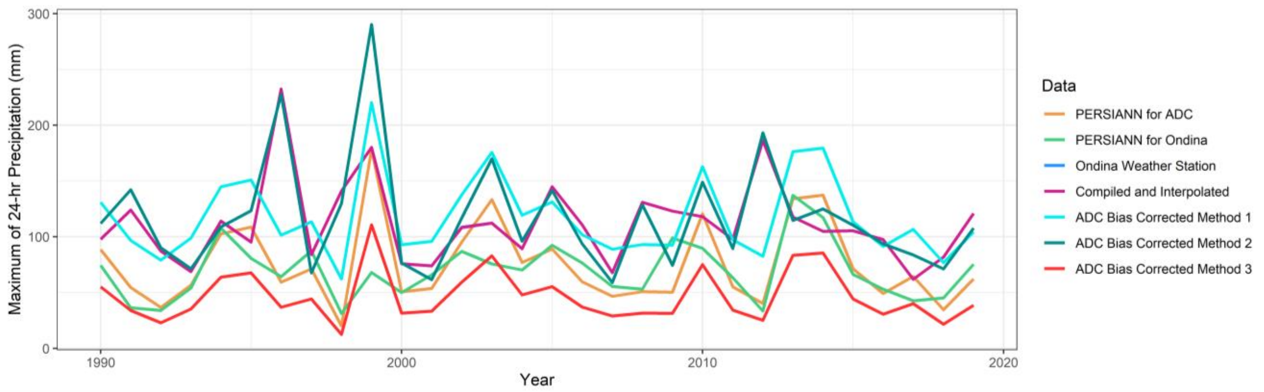


Figure A. 3. Time series of bias corrected monthly maximum of 24-hr precipitation

8.2. Rainfall hyetographs

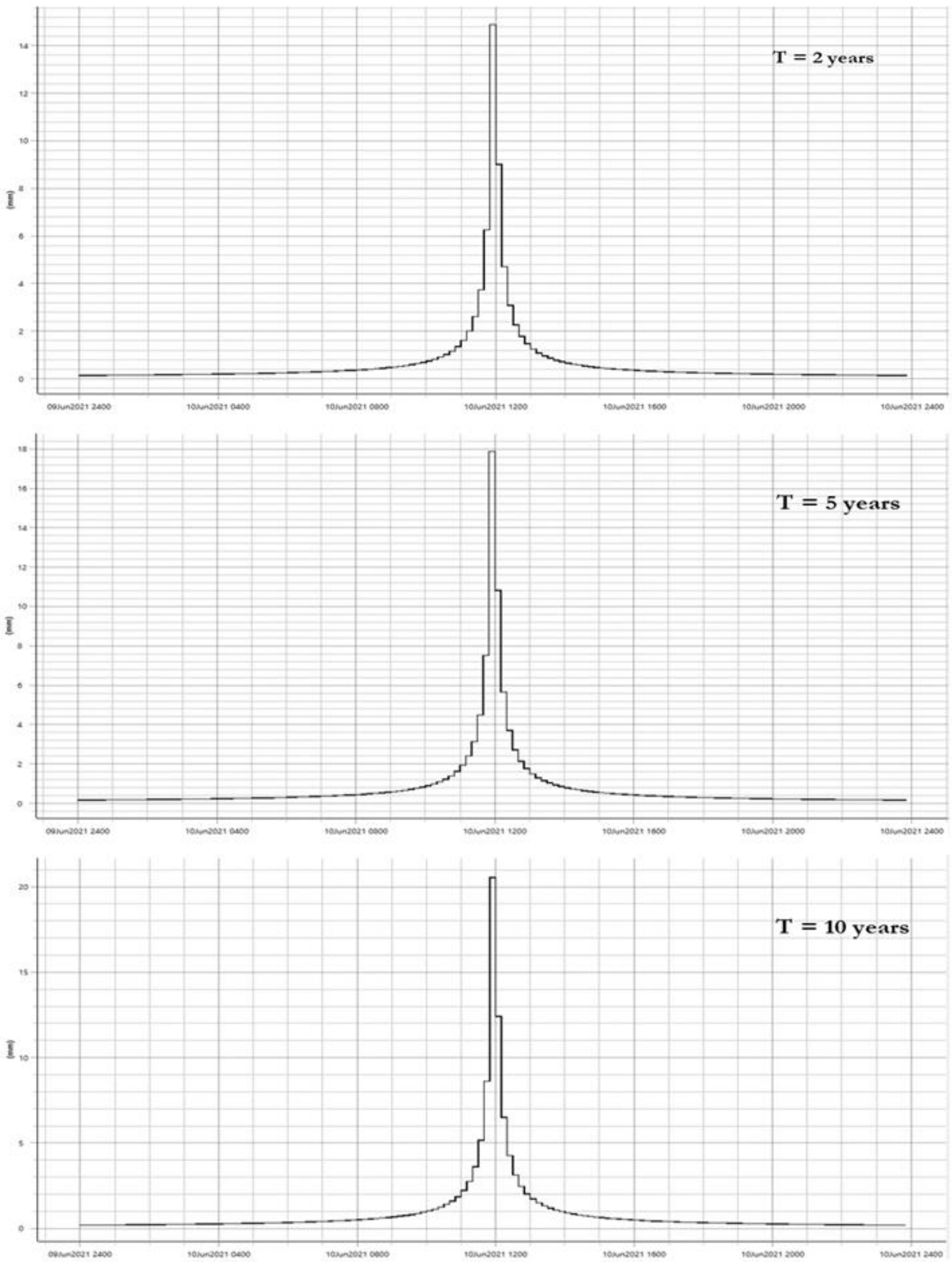


Figure A. 4. Rainfall hyetographs for return periods of 2, 5, and 10 years