VILNIUS UNIVERSITY
FACULTY OF CHEMISTRY AND GEOSCIENCES
DEPARTMENT OF CARTOGRAPHY AND GEOINFORMATICS

Simon Philipp Herbst

A Thesis Presented for the Degree of Master of Science in Cartography

INTEGRATING VOLUNTEERED GEOGRAPHIC INFORMATION AND SAR DATA FOR DISASTER IMPACT ANALYSIS

Supervisor: Dr. Andrius Balčiūnas

Vilnius, 2022

Herbst S. Integrating Volunteered Geographic Information and SAR Data for Disaster Impact Analysis. Master Thesis. Vilnius: VU. 2022

ABSTRACT

Floods are a common natural disaster with increasingly often devastating results. Disaster management is a highly complex and localized process that is difficult to generalize or extract from context. With radar data based on modern satellite constellations, the raw extent of floods can be quickly assessed at high spatial and temporal resolution, irrespective of cloud cover or time of day. In addition, Volunteered Geographic Information in the form of OpenStreetMap now offers an extensive database of individual pieces of infrastructure, in particular buildings, on a worldwide scale and free of use. Together with a newly available global land cover dataset, these data can be integrated to allow for quick, highly reproducible, and globally adaptive spatial impact analysis in the wake of flood events.

Potvyniai yra dažna stichinė nelaimė, kurios padariniai vis dažniau tampa pražūtingais. Efektyvus ir greitas reagavimas į stichines nelaimes – lokalizuotas ir itin sudėtingas procesas, apie kurį sunku kalbėti bendrai ar išskirti iš įvykio konteksto. Naudojant radarų duomenis, pagrįstus šiuolaikinių palydovų konsteliacijomis, galima greitai ir didele erdvine ir laiko raiška įvertinti pirminį potvynių mastą, nepaisant debesuotumo ar paros laiko. Be to, OpenStreetMap, kaip Bendruomenės kuriamos geografinės informacijos (angl. Volunteered Geographic Information, VGI) dalis, nemokamai siūlo plačią atskirų infrastruktūros objektų, ypač pastatų, duomenų bazę pasauliniu mastu. Šiuos duomenis galima sujungti su naujai prieinamais pasauliniais žemės paviršiaus duomenimis, kad po potvynių būtų greitai ir paprastai atkurta ir pasauliniu mastu pritaikyta stichinių nelaimių teritorijų poveikio analizė.

CONTENTS

List of Figures

List of Tables

INTRODUCTION

In 2021, 280 billion dollars in economic losses and more than 11,000 deaths were attributed to natural disasters [1]. Among these disasters, floods stand out both for how common they are and how dangerous they are to any part of the world mostly irrespective of socioeconomic development. It is estimated that about 29% of the total world population is exposed to flood risk [2]. In 2021, several devastating floods, among others in Europe, China, and Australia, illustrated this point once more. When looking at the destructive results of these floods and where they occur, it quickly becomes clear that even countries that claim to have large amounts of planning and warning facilities can be hit hard by flood events and their emergency response paralyzed by poor oversight and data availability.

Much attention is paid to figuring out how to prevent or foresee flood events, or how to estimate damages in the long run in order to adapt insurance policies accordingly. Less attention is paid to how available data now can be used in order to minimize loss of life and property during and in the immediate aftermath of flood events. Much of this available data is public and/or open-source and can be used irrespective of other available data by private sources that may be out of reach financially for many users. In particular, two types of data combine to create an attractive scenario for aiding disaster impact analysis:

- Satellite data by public sources such as the European Space Agency's (ESA) Copernicus program is reliable and available at high spatial and temporal resolutions.
- Volunteered Geographic Information (VGI) datasets such as OpenStreetMap (OSM) contain enormous amounts of high-resolution geospatial data with a consistent framework on a global scale that can be exported and used easily.

Given these data sources, an up to date assessment of their potential needs to be made in order to evaluate how useful they can be not only for scientific purposes, but also for aiding disaster response. Currently, there are only few works of research integrating the data described above. As such, the novelty of this work lies not only in the specific application for this case study, but also in developing this scientific approach.

The purpose of the thesis therefore is to develop and implement a methodology for integrating Synthetic Aperture Radar satellite data and VGI data to evaluate their usefulness for flood disaster impact analysis. The tasks are as follows:

1. Investigate the state of research for flood disaster impact analysis and usefulness of Volunteered Geographic Information for this task.
2. Evaluate existing algorithms for assessment of satellite data for disaster impact analysis.
3. Define appropriate data sets and algorithm for a methodological model.
4. Implement methodology technically.
5. Assess results and give recommendations.

1   LITERATURE REVIEW AND RESEARCH PROBLEM

1.1 Existing Flood Mapping Approaches

Floods are among the most significant natural disasters. They are the most frequently occurring natural disaster and cause significant damage worldwide [3]. As an effect of climate change, floods are also likely to become more frequent and more hazardous as the entire global hydrological cycle intensifies [4–6]. In light of this background, vast resources have been and are being spent to predict, prepare for, and diminish the results of flood events all around the globe, much of which is beyond the scope of this work. Where attention has to be paid however is all those scientific efforts that fall within the broader (extensive) field of Cartography and Geographic Information Sciences (GIS) and coincide with this work. Because of the enormous scope and relevance of flood events, this is not as much of a limitation as it may be in other fields of research. A further limitation is the ambition of this work to use satellite imagery and volunteered geographic information in order to tackle this set of problems – as such, special attention will be paid to previous research that integrated either or both of these methodologies. Finally, the directions these research activities take broadly correspond to the sections of events laid out by general disaster management theory. As such, this theory shall first of all be explained shortly.

Disaster management theory generally defines a cycle of four phases [7]. The Mitigation phase consists of activities that reduce the effect that any disaster could have – building and zoning codes, maintenance of infrastructure such as dams and bridges and general land-use planning, especially when it comes to high-risk areas such as flood meadows. The Preparedness phase similarly describes activities to be done before, in between disasters or in an ongoing manner. This includes making sure that the public is aware of existing dangers as well as what to do in case of a disaster, but also logistical aspects like setting up strategic stores of food, water, medicine, and necessary equipment. Finally, this phase also includes making sure that warning systems exist and work. Once a disaster does occur, the Response phase begins. This is the phase during and immediately after a disaster where human lives are most at risk and emergency operations are ongoing. As such, this phase includes directing relief and emergency resources in order to minimize property as well as personal damage. Because of this, one of the most crucial aspects of this phase is receiving real-time or near-real-time information as fast as possible. Without information, rescue and emergency operations cannot be done in an efficient manner. The Response phase ends as soon as all remaining active hazards are removed or otherwise conclude. Now, the Recovery phase starts. This phase describes the process in which affected areas and communities return to a normal way of life – rebuilding homes and structures, repairing damaged infrastructure and restarting essential services.

GIS research activities with the limitations set above can be broadly grouped into two groups. The first group combines the Mitigation and Preparedness phases into what this work will call the section of *Flood Vulnerability Analysis*. Within this section fall efforts to use GIS to mitigate the effects of flood events by attempting to predict which areas or structures are especially at risk of flood damage. These attempts can also help prepare for potential flood events. The Recovery phase of disaster management is mostly restricted to physical "on-the-ground" activities, such as rebuilding, as well as reconnecting social structures of affected communities. Cartographic activities play less of a

part here, but where they do, they can be grouped into the first section. The second group of GIS research activities is where they relate to the Response phase of disaster management and will be called *Flood Disaster Response* in this work. This part of the field attempts to aid immediate emergency operations in the wake of flood events.

In the following section, the two groups as described above will be introduced in their current state of research for the broader GIS field. While the focus of this work is on the Response phase, interesting developments, especially those regarding satellite data and VGI, are also found in research about other phases of disaster management and have to be considered, not only to give wider context but also to explore where different applications may be possible.

1.1.1 Flood Vulnerability Analysis

Vulnerability in the context of disasters is generally defined in regard to disaster hazards and risks. A disaster hazard can be described by a multitude of factors (for example, a 100-year flood as an event that has a probability of 1% of occurring each year), and any element that after analysis is deemed "vulnerable" to this hazard becomes an "element at risk" [8]. Whether an element or structure is regarded as vulnerable depends on vulnerability analyses. As mentioned previously, these analyses are therefore closely linked to the Mitigation and particularly the Preparedness phases of disaster management. GIS research can play an important part in this analysis, not only for its ability to link quantitative assessment with spatial data, but also cartographic works that can give concise overviews to stakeholders involved with preparing for disasters, such as emergency services, policy makers, but also to inform the general public.

Flood risk management in general, and flood vulnerability analysis in specific, take on diverse forms but can be traced back to the conception of flood risk management in the early 21st century [9]. Beginning with this paradigm shift, the field was further developed, in the German-speaking world especially by the works of Merz, Thieken and Herrmann [10,11]. What these works have in common is a focus on extraordinary thoroughness, where models are developed for a single river at a time. In this traditional approach, flow regimes are combined with digital elevation models in order to create a flow accumulation model that can be used to produce vulnerability maps. This methodology is shown in use by Herrmann et al. for the Elbe river [11]. Isma'il et al. produced a vulnerability map for the Kaduna River in Nigeria with a similar approach [12]. More recently, attempts have been made to introduce VGI data to the field. This was done by Schelhorn et al. with the conclusion that infrastructure data such as is available from OSM can make for an alternative approach that is less costly than the in-depth methods described above, and can based on its nature also involve local citizens, with downstream effects for general disaster preparedness [13]. Cerri et al. found that the usefulness of OSM data for flood vulnerability analyses lies in the fact that as this data is consistent structurally across the globe, vulnerability models that were developed for one region can be transferred to other regions in a cost-effective and efficient manner [14]. The advent of VGI has impacted the field in a positive sense, if only (for now) as an addition to rather than a replacement of classical types of research. The issue of whether VGI data in its current form can adhere to the strict

quality requirements needed to effectively prepare for disasters is a nontrivial one and beyond the scope this work. VGI in general is further discussed in 2.1.2 as well as 2.3.

In many ways, this section of disaster research is in contrast to disaster response (see 2.1.2). Whereas speed is the deciding factor of any disaster response analysis, it plays less of a role for vulnerability analyses that focus on long term studies. On the other hand, quality of data has the highest priority for preparing for disasters, where every involved party must know precisely to what degree which element is in danger – for disaster response, quality of data is important, but not as important as knowing quickly which areas are particularly hard hit and where people and property are affected the most. Of course, there is an interchange here – a preparatory analysis of the highest quality may make ad hoc analysis in the wake of a disaster unnecessary. But recent events show that natural disasters continue to be and may increasingly be highly volatile and unpredictable events that defy traditional risk management. For this reason, disaster response continues to be an important field of research, and all available information must be used to their utmost potential to create methodologies for worst case scenarios. The next section and the rest of this work will focus on disaster response.

1.1.2 Flood Disaster Response

Disaster response in the context of this work encompasses all research within the wider field of GIS aimed at *responding* to a disaster, be it an ongoing one or in the immediate aftermath. The focus in this phase of disaster management is on providing as fast as possible the information necessary to prevent further loss of life and property. A common theme at the outbreak of any disaster that obstructs precisely that goal is a general deficit of information [15]. Emergency operations, policy/decision makers and the general public all need to know what is happening in as much detail as possible. One important factor in this situation is that most of the relevant information is of a spatial nature, i.e., *where* something is happening. It quickly becomes clear that GIS is in a crucial position to help fill that spatial information gap. This is a fact that has become accepted in recent years [15].

GIS for disaster response shot into the spotlight in the aftermath of the 2010 Haiti earthquake. As one of the poorest developing countries, not even the most basic spatial infrastructure information was available digitally for Haiti. This lack of information severely hampered disaster relief at the outset of the catastrophe, as emergency operations could neither rely on official sources of information that simply did not exist, nor on public sources such as OpenStreetMap or Google Maps, datasets taken for granted in developed countries but that were largely uncharted territory for Haiti. In response, countless volunteers began working to fix this gap by downloading satellite images and mapping the outlines of roads and buildings en masse [16]. These so-called "crisis mappers" showcased the enormous benefits of VGI for disaster response for the first time. This of course goes especially for developing countries that do not have the means to develop significant spatial databases independently, but it was a lesson for how powerful VGI for disaster response can be, especially in combination with more traditional sources of spatial information [16]. Since then, the Humanitarian Open Street Map Team (HOT OSM) [17] has formed to tackle similar problems on a global scale, springing into action especially where other sources of spatial information are not available. Poiani et al. showed the successful collaboration of HOT OSM members in the aftermath of the 2015

earthquake in Nepal [18]. Westrope et al. discussed the mapping activities of the OSM community in the wake of Typhoon Hayian in the Philippines and in particular utilized on-the-ground research to review reports made by remote mappers about the level of destruction of individual buildings [19]. They found a significant lack of accuracy – only 26 percent and 40 percent of accurately classified (as to their level of damage) buildings for urban and rural locations respectively. This result highlights an important issue – VGI for disaster response is powerful in the sense that a large amount of data can be collected quickly and at low cost, but at the downside of potentially compromising accuracy. However, the Response phase of the Disaster Management cycle requires timely information, first and foremost. While highly accurate and qualitative data is always preferred, any data that can estimate disaster impact to a reasonable degree is good enough when lives are in danger. Having an approximate report of which areas are most affected is the basis on which decisions for directing relief and emergency efforts can be made.

When it comes to disaster response for floods in particular, closing the knowledge gap means determining the flood extent first and foremost. This means quantitatively detecting water from whichever data sources is being used for the analysis. Details about these methods and which methodology was used for this work will be discussed in 2.2 and 3.2 respectively. Aside from just detecting the water however, it is also important to know what has been affected by the flood extent. Traditionally, these kinds of analyses have been performed using damage models. These models generally use water depth as the factor for determining what has been damaged and are accordingly called depth-damage functions [20]. Such functions describe a relationship between water depth and level of damage based on certain land-use classes such as infrastructure or forest. They are then combined with damage values (i.e. a certain amount of money per m^2) to produce an estimate of the value of damaged property. As early as 2012, Jongman et al. described the downsides of this approach, in particular the wide range of existing depth-damage curves, damage values and the resulting uncertainty of damage estimates [20]. However, the approach remains widely in use today (for example in the Dutch Standard Method [21] or the American Federal Emergency Management Agency [22]).

More recently, some researchers have used the depth-damage model as part of a wider approach on socioeconomic damage in the wake of (flood) disasters. Oddo et al. were among the first to incorporate publicly available population data by SEDAC as well as VGI infrastructure from OSM to produce a thorough damage assessment in near real-time for the 2011 Southeast Asia flood and in a follow-up paper innovated further by producing a model for rerouting emergency services based on their previous findings [23,24]. Additionally, their conclusions stress the lack of real-time or near-real time approaches for disaster response as well as highlighting once more the importance of data with acceptable quality coming in time as opposed to perfect quality data coming too late.

To conclude, disaster research in GIS has a long history and is intimately tied to traditional types of disaster management. The increasing availability and complexity of VGI presents a kind of paradigm shift, which must be explored further in order to fully grasp its potential, in particular for disaster response where time is of the essence and the (at times) lacking quality of VGI data is less of an obstacle.

This work will continue in the vein of these recent conclusions to help provide more methodological resources precisely for near-real time disaster impact analysis. One of the most

important limitations of any impact analysis is the raw base data from which conclusions are taken. The next chapter will explore data available from satellite sources as will be used for this work.

1.2 Existing Satellite Data Methodologies for Disaster Impact Analysis

The vital role played by satellite remote sensing in disaster management has been well established [25,26]. High accuracy, low revisit times and the number of different sensors available are all benefits of this technology in addition to the sheer scale of data that can be quantitatively investigated. Generally speaking, satellite remote sensing works can be split into active and passive sensors. Passive sensors collect and make use of radiation that has already been emitted, while active sensors themselves emit energy pulses that are then received back. Data from passive sensors, or optical remote sensing, can create a number of different kinds of images based on the spectral information collected from various wavelengths. The information from the spectral bands can be combined to produce, for example, true color images based on the red, green, and blue bands visible to the human eye. Meanwhile, the most widely used active sensor technology is RADAR (Radio Detection and Ranging) [27]. As the name suggests, this technology works by actively sending out and receiving back pulses of electromagnetic energy with wavelengths in the radio frequency. The information received is not directly spectral information, but rather information about objects relative to the sensor from which the energy pulses were reflected or scattered, such as distance (ranging) and angle. Images can be produced from calculations based on this information. Of particular importance for remote sensing is Synthetic Aperture Radar (SAR). SAR works based on the satellite-borne sensor sending out radio waves measuring the same scene multiple times at different times or locations (as the satellite is moving), the resulting information of which bears data about ground topography and deformations, which can be used to create highly accurate reconstructions such as Digital Elevation Models (DEM) [28]. The difference in absorption and reflection rates between water and land areas [29] means the technology can also be used to detect water surfaces – including flooded surfaces of land area. In addition, radio waves penetrate clouds and SAR does not rely on having to capture already emitted radiation (i.e. from sunlight). This means that SAR is not reliant on weather or time of day [30], making it the ideal instrument for flood disaster response [31]. However, despite being held in high esteem, SAR sensors have traditionally carried with them limitations, especially their high revisit times and low accuracy / large pixel size [27].

Among the most recent remote sensing missions (with data publicly available) are Sentinel-1 and Sentinel-2 by the European Space Agency Copernicus program. Sentinel-2, initially launched in 2015, is a constellation of the two satellites Sentinel-2A and Sentinel-2B and provides multi-spectral optical imagery. While optical methods are ideal for real color visualization and long-term studies of natural disasters [32,33], optical imagery is reliant on cloud-free skies, a prerequisite that is problematic for observing many parts of the planet, and in particular flood events that are often accompanied by precipitation events and cloudy skies. Sentinel-1 on the other hand is a constellation of two satellites (Sentinel-1A and 1B) carrying C-band microwave SAR instruments (see 2.2.2). It boasts a low revisit time of 12 days, cut down to just 6 days since the launch of Sentinel-1B [34]. Together with its high spatial resolution, it leaves behind many of SAR's traditional limitations that

would have made it much less usable for disaster response. However, Sentinel-1B is unavailable since December 2021 based on technical difficulties and is assumed to be out for the long term [35]. In the meantime, Sentinel-1A remains available, though the longer revisit times have an effect on its effectiveness for disaster applications. In the following section, existing algorithms for using both optical (passive sensors) and radar (active sensors) satellite data for disaster impact analysis will be introduced.

1.2.1 Optical Imagery

As established in 2.2, optical remote sensing passively collects emitted radiation of various wavelengths. Multispectral systems such as the Sentinel-2 constellation collect their data in several separate bands, each of which is sensitive to a small part of the electromagnetic spectrum. The image output of these bands can be combined at will to emphasize elements of the environment that absorb and/or reflect specific parts of the electromagnetic spectrum. An example of this approach is the Normalised Difference Vegetation Index (NDVI) [36]. The NDVI is calculated as the ratio of reflected red to near-infrared light on a scale of -1 to +1. Because the chlorophyll compound in vegetation absorbs red light, whereas the mesophyll leaf structure reflects near-infrared light, this index is an efficient indicator of absence or presence of vegetation. Optical imagery is also widely used in long term studies such as the CORINE land cover program by the European Union. Similarly, it can be used to cover the long term effects of disasters on the landscape. Balamurugan et al. studied the long term land cover changes in an earthquake-affected area of Gujarat, India [33]. These studies are valuable for better understanding of how large disasters can have lasting effects on land cover patterns and help decision-making for the Mitigation and Preparedness phases of disaster management.

When it comes to studying the immediate impact of (flood) disasters, many of these ideas can be used as well. Already in 2006, Brakenridge et al. recommended the use of detecting flood (pixels) by calculating the variation in NDVI in "typical" and "flooded" conditions [37]. This methodology works based on the assumption that NDVI values below a certain threshold in "flooded" conditions can be considered inundated. Giordan et al. used NDVI variation in addition to a modified Normalized Difference Water Index (NDWI) variation to detect flood extents from Sentinel-2 satellite imagery [38]. The NDWI was initially derived to detect water stress in plants as a consequence of drought and defined as the ratio of near infrared (NIR) and short-wave infrared (SWIR) channels, where NIR is related to the mesophyll leaf structure (see above) and SWIR is negatively related to leaf water content [39]. In its modified form (MNDWI), it relates to soil moisture rather than leaf water content and is defined as the ratio between the red and SWIR channels. Soil moisture then is an additional indicator of flooding. Giordan et al. also used SAR data in their analysis, which will be discussed in 2.2.2, as well as concluding that satellite imagery is useful for providing a quick, cost-effective overview for larger areas, while lacking in urban areas [38]. Huang et al. similarly used an NDWI index for near real-time flood detection in addition to using VGI in the form of tweets and reports to form a flood probability distribution as a tool for guiding emergency responders [40]. Other spectral-based indices commonly used include the Land Surface Water Index (LSWI) and the Modified Land Surface Water Index (MLSWI) [41–43]. The Standard Method described in 2.1.2 calculates damage estimates

according to the land cover types affected by a flood. Land cover datasets such as CORINE are typically made through a combination of field survey and (automatic) classification (supervised or unsupervised) of optical remote sensing imagery [44,45].

Further algorithms for disaster management and prevention are also made available by the cloud platform for remote sensing *Sentinel Hub* as a collection of user-made JavaScript scripts [46].

Optical imagery is easily interpretable (especially in its channels visible to the human eye), which is a factor not to be underestimated – disaster response involves presenting data to various stakeholders and decision makers, many of which are not familiar with remote sensing. The data made available to these stakeholders must be easily and rapidly understandable. With its many spectral channels, optical image data can also take on a variety of forms for different purposes and repurposed when necessary and found to be successful (such as the MNDWI). Clearly, optical imagery is highly useful data for analyzing the impact of disasters. However, its biggest drawback here is the reliance on cloud free weather, which is not a given for many latitudes and certainly not in the wake of a flood event that is inherently correlated with precipitation [47]. Many visualizations done with optical imagery can rely on images taken at different dates in order to create perfect cloud free mosaics. This is not the case for disaster events, where data must be quickly and reliably available. For this reason, radar data will be used in this work.

1.2.2 Radar Data

Radar sensors are unaffected by weather or daytime conditions. With the advent of satellite constellations that offer higher resolutions and revisit times, they have thus become popular systems for flood impact analysis. Examples for used satellite systems are TerraSAR-X [48], ALOS-2 PALSAR-2 [49] and the Sentinel-1 constellation launched in 2014 and 2016 (see 2.2).

Bioresita et al. successfully applied a rapid mapping approach to European flood events using Sentinel-1 Ground Range Detected (GRD) data with double polarization as well as developing a processing chain for automatic extraction of flood extents [50]. Uddin et al. implemented a methodology with the same type of data for flooding in Bangladesh, which stands as an example of optical imagery being unhelpful for disaster response actions as a result of the almost complete lack of cloud-free images during the flood season [51]. They validated their results using cloud-free Landsat-8 images that were classified using NDWI and NDVI (see 2.2.1). Landsat images were also used to produce land cover maps for estimating more specific flood damage. Zhang et al. introduced a fuzzy logic-based refinement for postprocessing flood extent images to remove areas that looks like water, further increasing accuracy, on top of developing flood detection methods that are more suited to challenging climates such as the semi-arid areas of Pakistan [52]. In addition, the initial flood delineation was done using thresholding based on Probability Density Functions (PDF) – comparing the PDFs of flooded and non-flooded areas. Li et al. focused specifically on the issue of flood mapping in urban areas using SAR: settlements, in particular densely built-up areas, cause complex backscattering effects to occur, which can make SAR flood mapping unreliable in these regions [53]. Their solution was to use SAR interferometric coherence in addition to the SAR intensity with promising results. The issue of flood mapping accuracy, or lack thereof, in urban areas has been

discussed in several other examples [54,55]. In addition to using more complex SAR base data, a solution can be to use extremely high resolution, locally sourced data such as from drones or even down to using handheld cameras with GPS devices [38] as well as using LiDAR-derived data [56].

Similar to 2.2.1, Sentinel Hub also offers a number of custom-made JavaScript scripts for disaster management with radar data, in particular for flood mapping [57]. The algorithm focuses on separation between permanent water bodies and surface water in flood events.

Regardless of its current limitations, SAR remains the most consistent and reliable data source for flood impact analysis, especially in the context of rapid disaster response.

1.3 Volunteered Geographic Information

Volunteered Geographic Information (VGI) is an umbrella term used to describe the various forms of crowd-sourced geographic information available on the internet. It was first coined by Goodchild et al. [58]. In general, it describes geospatial datasets put together in a community effort by volunteers in their free time. This distinguishes VGI from efforts made by private or public institutions to collect data in an organized and centralized way. VGI instead depends on a large number of volunteers to contribute the kind of quantity of data that can compete with the often large budgets of public or private institutions. While volunteers may have considerable local knowledge, they are not necessarily trained professionals in any relevant surveying field. As a rule then, VGI is different from traditional surveying or data collection in two important ways: First, the data is collected in a *decentralized* and not necessarily organized manner, often creating a patchwork of data that is more pronounced in areas where many volunteers coincide and less pronounced where there are fewer contributors. The consequence of this dynamic is a divide in available data between urban and rural areas on the one hand, such that urban areas are mapped more quickly than rural areas with less contributors [59], and on a bigger scale also often a divide between highly developed countries with many members contributing data and developing countries on the other hand, although there is a lack of conclusive studies for this scale [60]. The second way in which VGI is different from traditional (spatial) data collection is that volunteers are generally not trained professionals in this field – the surveying effort is by and large a *non-professional* one.

These factors highlight the biggest advantages of VGI on the one side: It is cheap because of the voluntary nature of contributions and can be quickly updated because of the direct action taken by volunteers that does not rely on lengthy planning, procurement and quality assurance processes that accompany data collection efforts by larger public and private institutions. However, they also highlight the biggest downside of VGI: even with community standards and crosschecking efforts in place, the final quality of the dataset cannot be wholly assured to the same level as it could with more centralized, larger budget organizations. The quality and credibility of VGI in general has been discussed at length, in a discussion that is often tied to a larger discussion about user-generated data in a Web 2.0 [61,62]. Among its drawbacks are not only the varying quality, but also the lack of an overreaching scientific approach and incompleteness of datasets [63]. Quality assurance varies from platform to platform.

Rather than relying on VGI, even relatively complete datasets such as OSM, there are generally more comprehensive, detailed, and complete datasets for infrastructure available on a local, regional, or national level that could be used for disaster impact analysis. The main advantage of using OSM as a global dataset however is the possibility to apply the same methodology used in this work to any location in the world. Just as satellite data can be used on a global scale irrespective of borders, OSM now provides not only the theoretical possibility to intersect this remote sensing information with quantitative data about manmade structures (buildings, roads, railways etc.) anywhere in the world, but also a consistent technical framework that makes such an approach comparable and reproducible. The innovative approach of using OSM for a disaster impact analysis approach is going beyond the many administrative levels – national, regional, and local, each of which with their own approach to

data collection, processing, and presentation – and producing a methodology that works on a global scale.

Another aspect that makes using VGI for disaster research and analysis attractive is the well-established benefits of including the public particularly within the environmental domain in "citizen science" efforts [64,65]. Not only is the data offered by volunteers useful and different from other more static sources of data, but the result (if well communicated) is a kind of positive feedback loop where volunteering data becomes more enticing for the public when they see it being used to great effect in humanitarian and environmental accomplishments, which makes existing datasets more complete and appealing for researchers.

2   METHODOLOGY

The methodology for this work was chosen in a process lasting several months. First was the idea of using OSM data for analyzing the impact of disasters in a quick, and cost-effective, and (in theory) globally applicable manner. Initially, it was thought to combine this data with optical imagery for the most intuitive and presentable result. While this is possible in theory, the issue of cloud cover means that in practice it is simply unfeasible – in the case of a disaster, the stakes are too high to rely on whether or not a cloud-free image is available. For this reason, SAR data was chosen as the raw data for detecting the flood extent instead. SAR carries with it other issues – especially in its lack of interpretability, which also has effects on its presentability as raw data. However, in this case, not having to rely on weather and daytime means choosing accuracy over interpretability. Next, the Rhineland floods of 2021 were chosen as a case study for their recency as well as catastrophic magnitude in a highly developed region. The level of completeness of OSM data in this region was another reason for choosing it as a kind of prototype for this methodology. When it became clear that the statistical output of the analysis does not necessarily reflect the extent of the disaster, the 2021 floods in the Henan province of China were chosen as an additional case study.

The reasons for choosing the 2021 Henan floods lay in maximizing the comparability of the results of the analysis on the one hand: Choosing the same year, and indeed the same month (July 2021) was important for checking the validity of integrating OpenStreetMap data. Because the OpenStreetMap database is continuously being updated, the data exported from this database reflects the most recent real-time situation possible (depending on how up to date the area of interest is). It therefore makes sense to pick a recent flood example to compare this data too, or at the very least pick an example from a similar time. On the other hand, the 2021 Henan floods were significant for their damage done to urban areas, in particular in the cities of Xinxiang and Zhengzhou. They therefore were a great opportunity to test the conclusions taken from analyzing the first example, specifically whether the methodology can serve a purpose in visualizing damage done to settlements, although more precise, locally-sourced data (as described in 2.2.2) was not available for these case studies and only raw SAR data was used to detect the flood extent.

2.1 Case Studies

Two case studies were selected for this work. In this section, the events leading up to the floods will be shortly introduced.

2.1.1 Rhineland, Germany

Beginning on the 13th of July 2021 and lasting up to the 15th of July, areas of Western Europe experienced torrential rain with precipitation as high as 272 mm within 48 hours in Belgium as well as 154 mm within 24 hours at the Köln Stammheim station, more than double the monthly average for July [66]. Resulting flash floods led to catastrophic results, especially in the narrow and steep river valleys of Rhine tributaries in the Rhineland-Palatinate and North Rhine-Westphalia states of Germany. The worst occurred along the Ahr river where 20 of the 35 bridges as well as many streets and villages were devastated. Within just hours from the 14th to the 15th of July, the water level of the Ahr river had risen to more than 7 meters before malfunctioning (**Error! Not a valid bookmark self-reference.**) [67]. In total, 184 people died, including 90 along the Ahr river and 27 in Euskirchen north of the Ahr [68]. The case study in this work will focus both on the Ahr valley as well as the Euskirchen district.
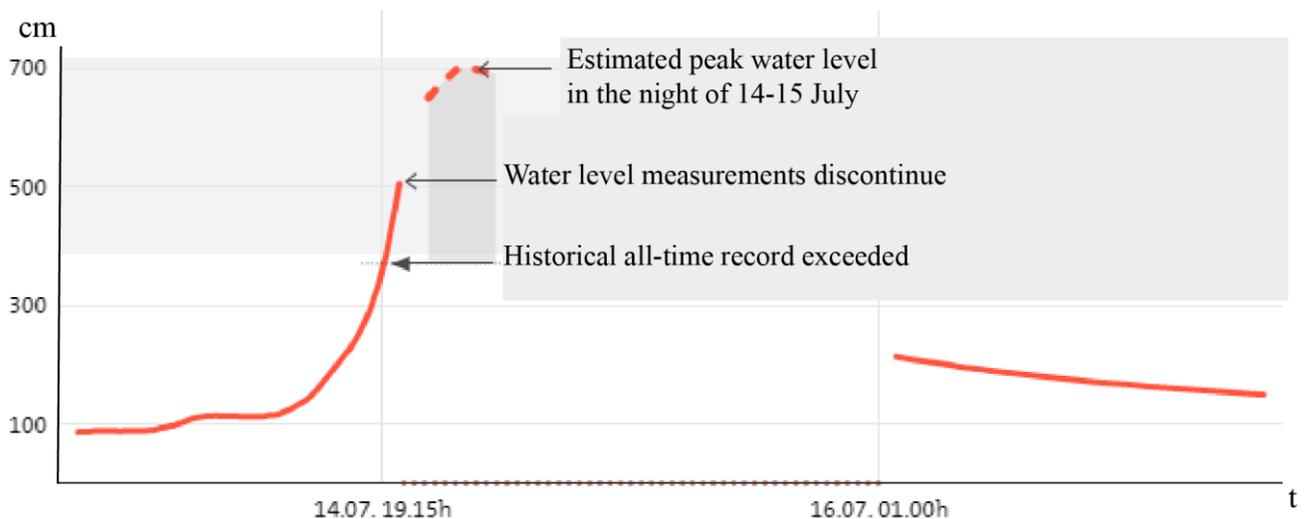
Figure 1: Sequence of water levels at Altenahr point from the 14th to the 16th of July 2021 [67] (modified).

2.1.2 Henan, China

From the 16th to the 20th of July, the Henan province of China experienced record amounts of rain (Figure 2).The provincial capital of Zhengzhou was hardest-hit with 201.9 mm of rain within just an hour and up to 622.7 mm between the 17th and 20th of July, which is more than 95% of the average annual rainfall [69]. Other areas particularly affected were the city of Xinxiang with more than 250 mm of precipitation within the period as well as the city of Xinmi with more than 400 mm [70]. The official death toll stands at 398 [71]. More than a million people were evacuated during the event and up to 500,000 households affected by damaged or collapsed structures [72]. Direct economic losses amount to €12.5 billion [72].
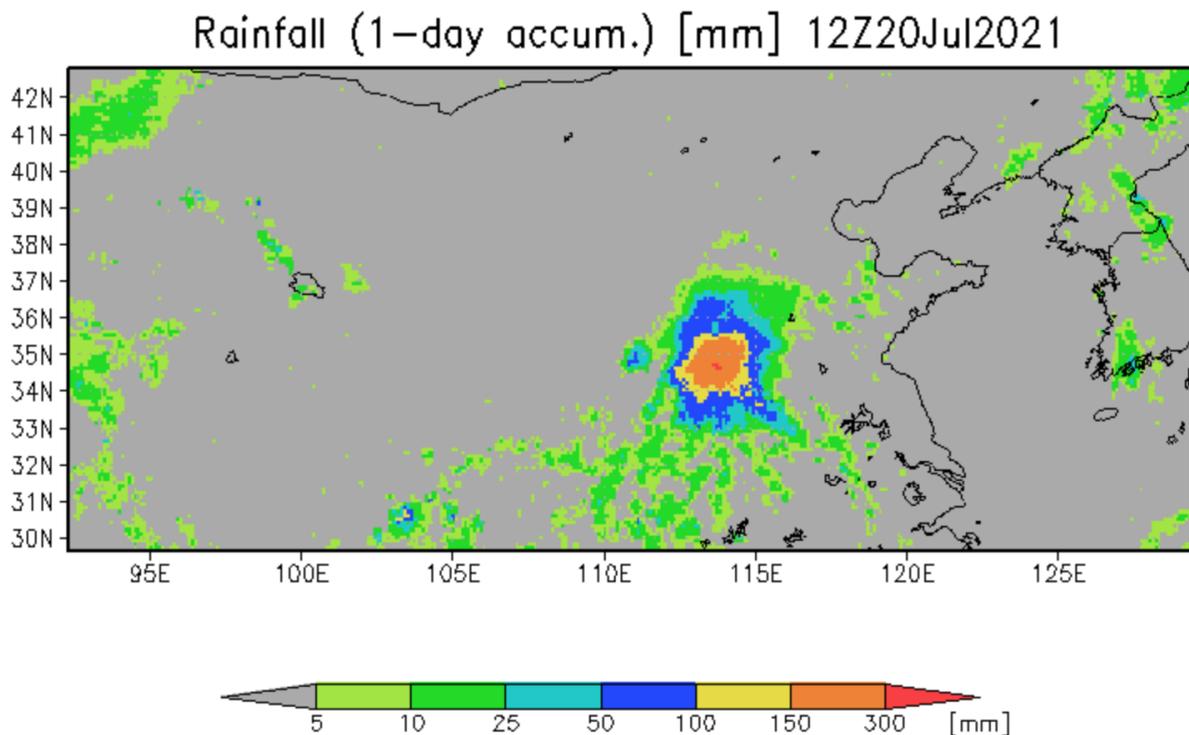
Figure 2: 1-day accumulated rainfall in Northeastern China on the 20th of July 2021 [73].

2.2 Classification of SAR Data

Sentinel-1 imagery is freely available as part of the Copernicus program by the European Space Agency on the Copernicus Open Access Hub [74]. The goal was to select a date ideally during the flood (disaster image) in order to satisfy the aims of this work to fit within the disaster response timeline – in other words, to produce an analysis that could be considered real-time or near real-time. To produce a flood extent, a date before the disaster was necessary to compare the disaster image with. All satellite products were obtained in Interferometric Wide swath (IW) mode and as Ground Range Detected (GRD) products at 20 m x 20 m resolution. A GRD product has been multilooked (adjacent pixel values averaged to reduce noise) and projected to ground range coordinates before being made available on the Copernicus hub. In a SAR GRD product, pixel values are representative of the detected amplitude, which means that they first have to be calibrated in order for the product to be usable for analysis.

For the Germany case study, two products were initially chosen for the investigation: the 15th of July, coinciding with the peak of the flood events, was ideal for the disaster image. For the before date, the 3rd of July was chosen. The reasoning behind choosing this date lay in the repeat cycle of the constellation: each satellite (Sentinel-1A as well as 1B) completes 175 orbits per 12-day cycle before repeating its orbits and differs only in a 180° orbital phasing difference. So while more dates are technically available – both from Sentinel-1B as well as from Sentinel-1A (but from a different orbital angle) - sticking to the revisit time of 12 days guarantees the same orbital angle (or *relative orbit number*) as well as an exact corresponding geometry. This not only simplifies the image processing stage of the analysis, but also makes possible an exact overlay of the before and disaster scenes. This on the other hand means that the result of the analysis can be a stack of images, or in other words a layer simply representing the change in pixel values from before to after, and therefore a reliable flood extent classification.

The two scenes were imported into the Sentinel Application Platform (SNAP) software for preprocessing. SNAP is an open-source image processing and analysis platform, specifically developed in the context of the Sentinel mission [75]. The raw amplitude products were loaded. They could be preprocessed separately, but their corresponding geometry means that the steps taken were precisely the same. Instead, the SNAP toolbox allows batch processing for multiple scenes. First, the scenes were cropped to a subset which covered only the area of interest in question. This decreased the file size and processing time by more than 70%. Next, a thermal noise reduction was applied, and a radiometric calibration algorithm transformed the raw amplitude pixel values to a calibrated backscatter coefficient. In this case, σ_0 (Sigma nought) was selected as output band. This is consistent with previous studies on SAR processing of flood images [76,77], and σ_0 has been found to separate well between water and land surfaces [50]. It carries with it the downside of not taking topography into account (as it represents a mathematical ratio projected onto a horizontal plane) [29], however, for the relatively flat areas of interest in this work, this could be ignored. After applying a speckle filter, the exact orbits for the images were computed (either downloaded from the internet or taken from the product metadata) and the images were projected onto a geographic reference system, in this case WGS 84 (EPSG:4326). Finally, the pixel values were transformed from linear to logarithmic (dB) values in order to improve readability of the pixel histogram. This histogram (of the disaster

image) was investigated in order to find an acceptable threshold for creating the binary flood extent image. This method takes advantage of the difference in radar backscatter values between flooded areas (i.e. areas of water) and land areas, such as agricultural or grass land. Urban areas show higher backscatter values still, appearing light on the raw image. Because most of the image is still non-flooded areas, investigating the histogram of the entire image usually yields little results: The non-flooded areas dominate statistically and create a histogram result not dissimilar from a Gaussian distribution. To get a more accurate view, SNAP toolbox allows creating geometry containers (Figure 3). Several smaller areas showing flood evidence were selected with the polygon drawing tool and thus included into the geometry container. This container could now be investigated statistically, separate from the image as a whole, with the result of a much clearer image of the essentially binary nature of the image (Figure 3.4) – areas of low backscatter value (water) and areas of higher backscatter value (land). The valley in between these two peaks was defined as the threshold for the next step. In this next step, both images were altered with the Band Maths function of SNAP toolbox. Using the threshold defined in the previous step. a binary image was created which separates the many backscatter values into simple ranges of "larger" or "smaller" than the threshold.
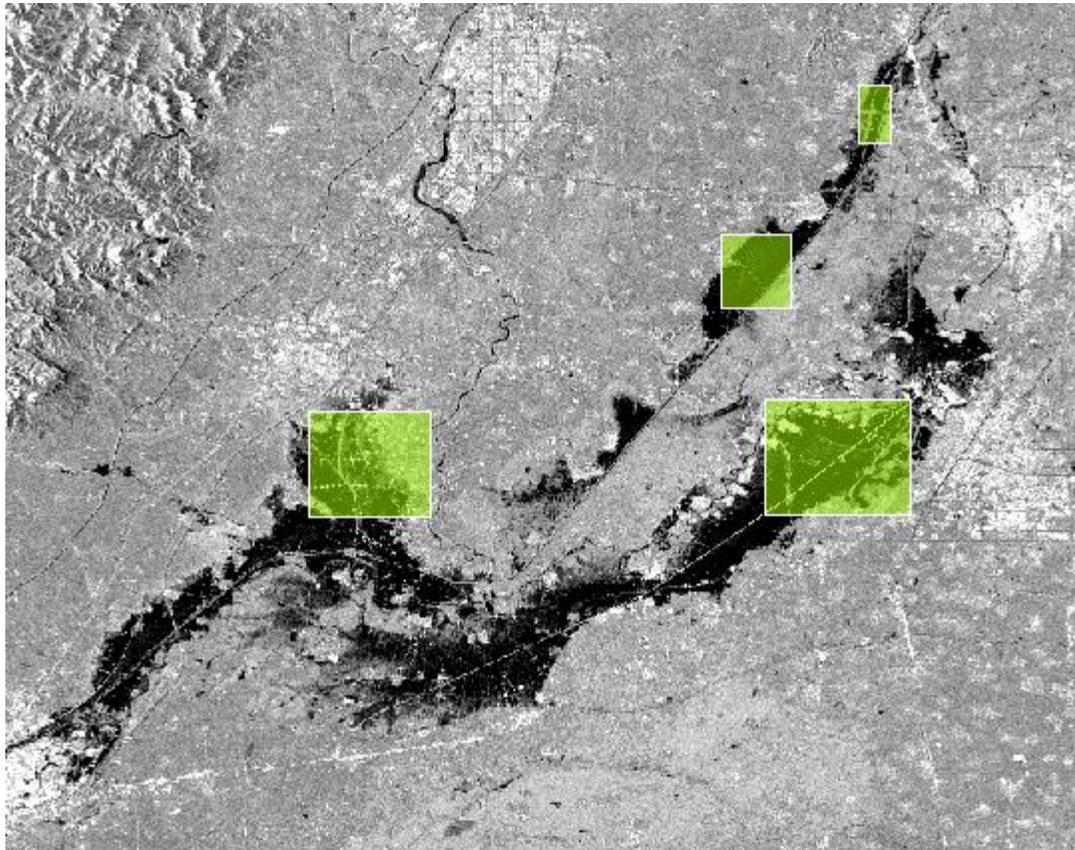
Figure 3: Henan Province preprocessed SAR flood extent and geometry containers for thresholding (SNAP Toolbox).
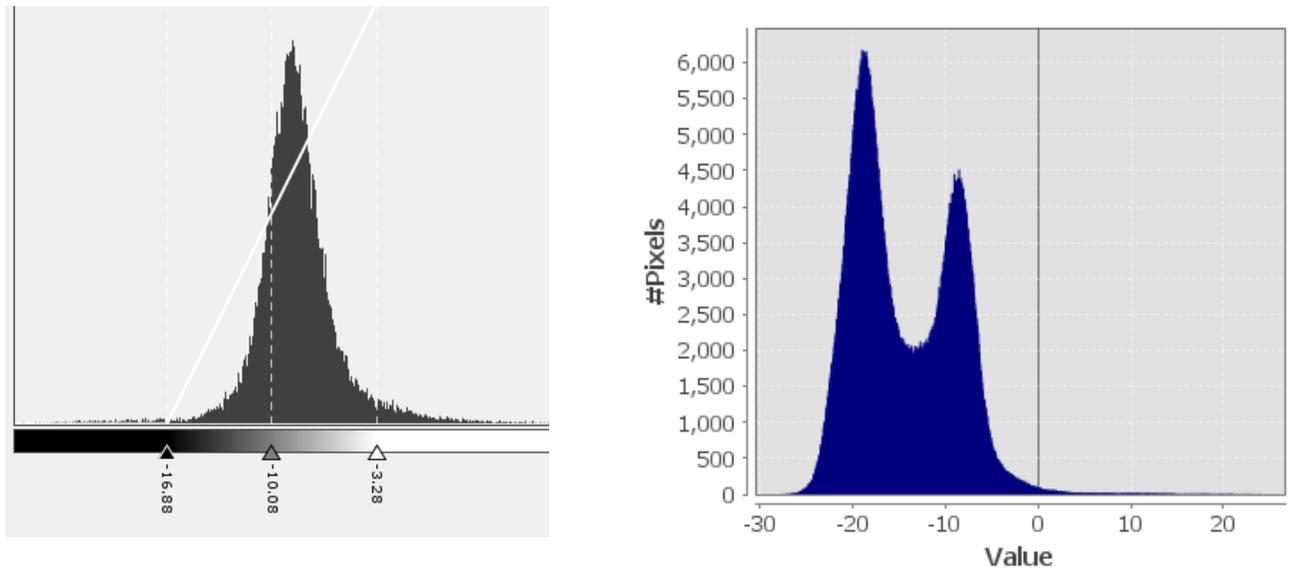
Figure 4: Pixel value histogram of the raw image (left) and geometry container (right).

The final step in processing the images was to overlay them into a single stack image and finally subtract the values of the before image from the disaster image. Since both of them were now binary images, and all light areas were representative of water (or at the very least low backscatter values), this step was equivalent to removing from the after image all low backscatter values that were already present in the before image. Put another way, all sources of water that are irrelevant for the flood analysis (standing water bodies, rivers etc.) were removed from the classified flood extent. What was left was simply the difference in low backscatter values between the two dates – since they were just 12 days apart, it could be said with some certainty that the difference represented the extent of the flooding which happened in the meantime.

For analyzing the Henan floods, two SAR products were used: Similar to the previous example, one scene before and one during the disaster was preferred. In the case of Henan however, the satellite flyover dates did not perfectly correspond to the flood event. As established in 3.1.2, the rains that caused the floods happened largely between the 16th and 20th of July. The only scenes available near these dates were from 15th of July, which was satisfactory for the before image, but in order to satisfy the geometry limitations (that make analyzing the images properly possible), the next image with the same relative orbit number had to be chosen – 27th of July. While already six to seven days after the initial flood event, an initial overview of the amplitude band showed that evidence of the flood event was still very much visible and therefore adequate for a comparative analysis at this stage. The two products were subject to the same processing stages as in the previous case study, with one difference – the Sentinel product could not be subset suitably based on the enormous scale of the Henan floods as well as the unique geometry of the product with Zhengzhou city on one end and the worst hit areas around Xinxiang along the Wei River on the other end meant that the product could not be cropped satisfactorily. However, the result of this difference was only one of computing time: Batch processing the two images took around 6 minutes as opposed to half a minute for the Germany case study.

These final flood extents were exported as a georeferenced raster (geoTIFF format) and imported into QGIS.

2.3 Incorporation of VGI Data

For exporting infrastructure data from OSM, the HOT OSM Export tool [78] was selected. The HOT OSM Export tool has several advantages over other OSM export tools: First, it is possible to export by custom bounding box rather than a predefined area (such as a state, or county). As this work looks at the impact of a natural disaster that in its extent bears no similarity to predefined regions, this is both a way to optimize data storage as well as a more precise way of working. Second, it is possible to select the precise type of data wanted for export, by specifying through the OSM tag tree as well as through a preset XML file. This is preferable instead of a bulk download as the properties of the data received are much clearer, as well as again optimizing data storage. Third, it is possible to clone an export configuration. This is especially advantageous for scenarios where data for the same area, but for different feature types is needed.

The *Extract layer extent* algorithm in QGIS was used to create the minimum bounding box covering all features within the flood extent layer as exported in 3.2. This box in form of a GeoJSON file was imported into HOT OSM Export as the spatial extent of the export file. Further, only building extent polygons mapped in OSM were selected through the tag tree, which is based on all objects identified by the *building* key. The result of this operation was a vector layer containing the (polygon) geometries of individual buildings mapped in OSM. This raw export layer included 47136 buildings for the China case study within an area (minimum bounding box of the flood extent) of 20501 km^2, and 664233 buildings for the Germany case study in an area of 3971 km^2. This procedure was repeated for the OSM data types *Road* and *Railway* as classified through the tag tree by the *highway* and *railway* keys, *"highway = yes"* being the identifying value for any road, street, or path within OSM.

The raw export layers were then imported into QGIS and reprojected to the respective local UTM projection.

In the following figures, the exported OSM data is shown over the case study extent in raw form.
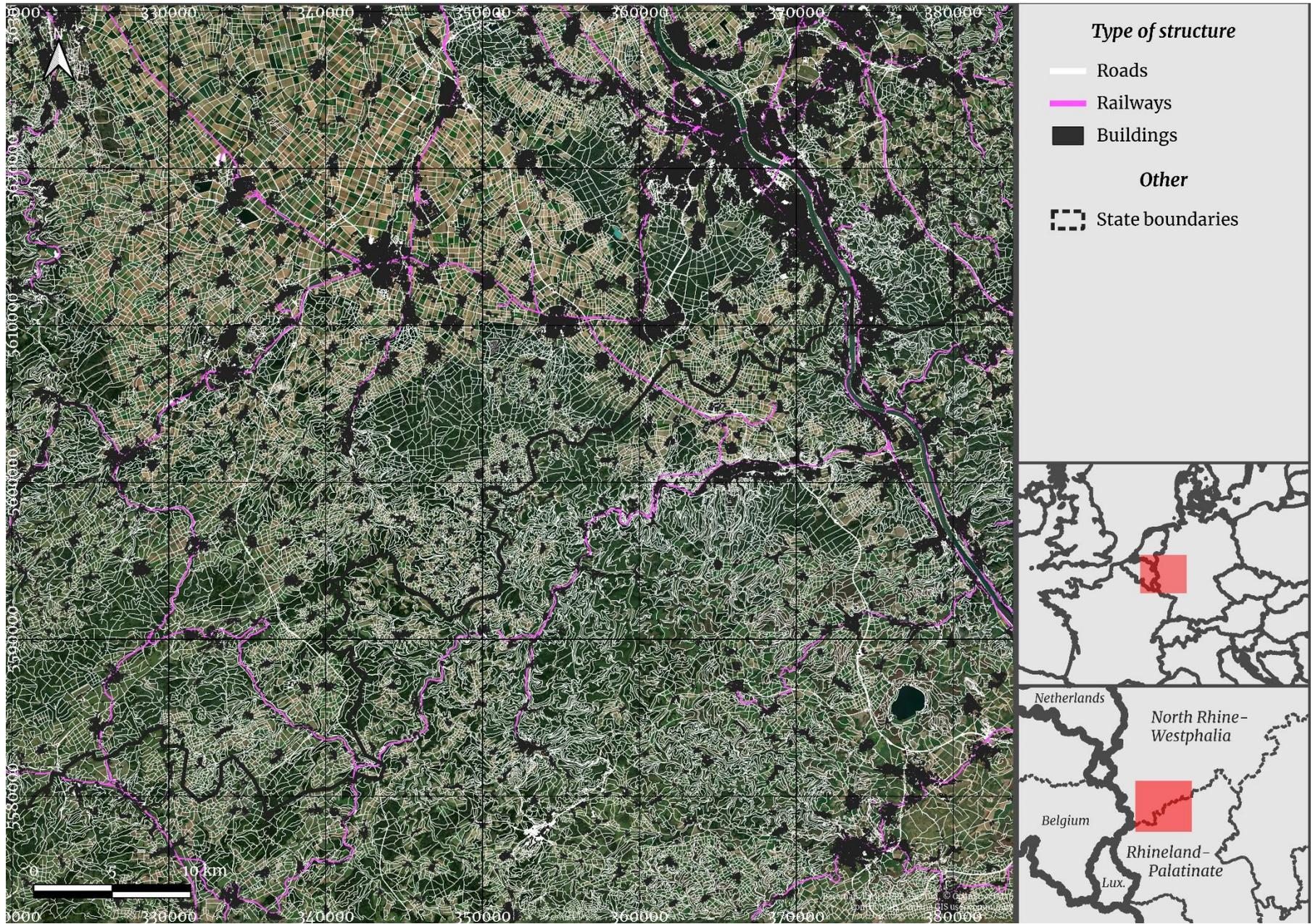
Figure 5: Density of raw infrastructure network as exported from OSM. Germany case study at 1:280 000.
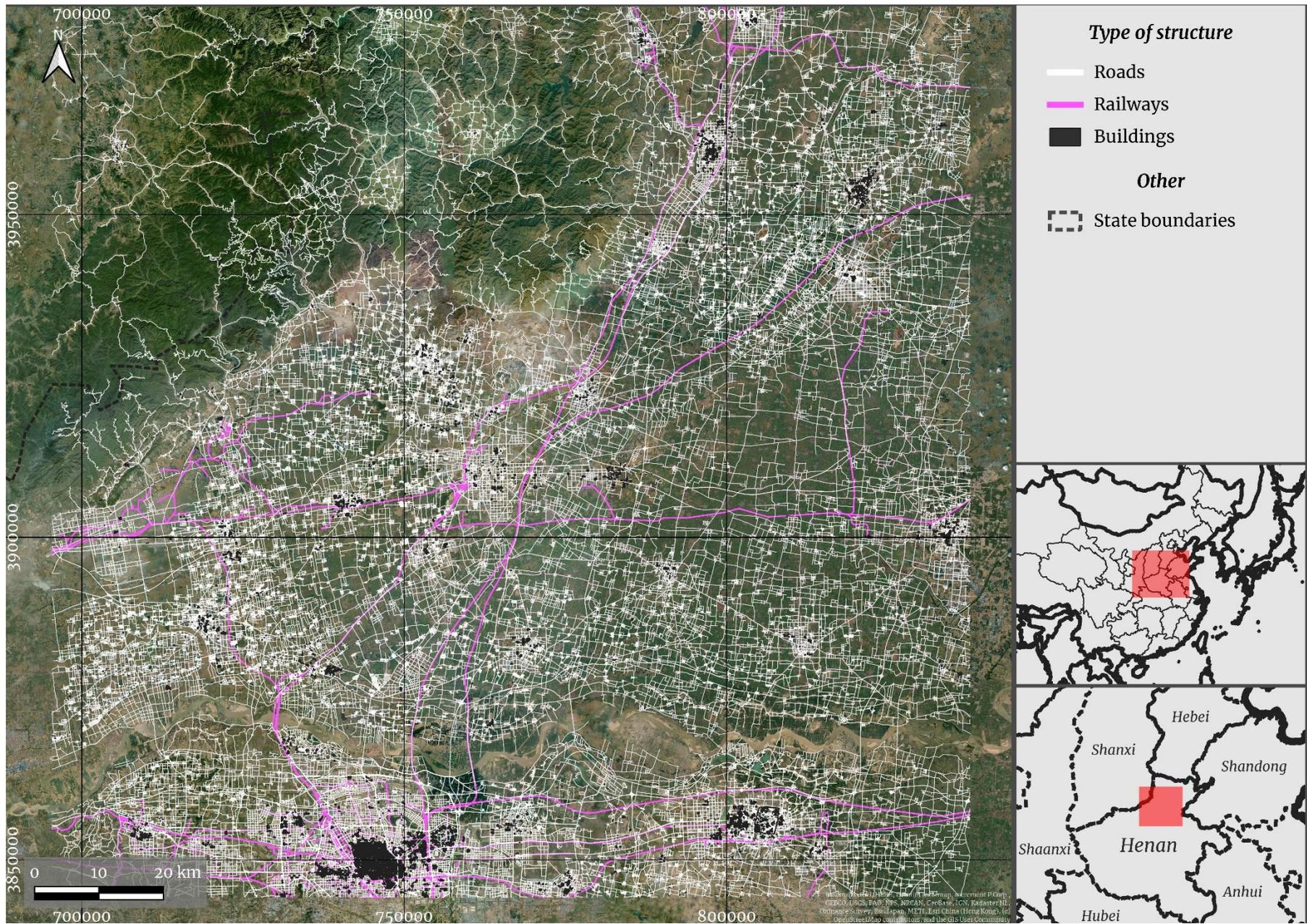
Figure 6: Density of raw infrastructure network as exported from OSM. China case study at 1:680 000.

2.4 Integration of Data Sources and Impact Analysis

The raw flood extent layers were location-checked, vectorized using the *Polygonize (Raster to Vector)* function and reprojected to the respective UTM zone (UTM Zone 32N / EPSG:32632 for Germany; UTM Zone 49N / EPSG:32649 for China). Next, the intersection of the flood extent layer and the exported *Building* dataset from OSM was calculated using the *Extract by location* function. While *Extract by location* does not technically calculate the exact overlap with the flood, this is sufficient for buildings as it is less important to know which parts of a building may be affected and higher priority lay on getting an overview of which buildings could be affected. Next, the procedure was repeated for the *Road* and *Railway* datasets, with the difference that here the exact *Intersect* of the two layers was calculated to find out which exact parts of roads and railways were affected. Thereafter, the total length of all affected road and railway pieces was calculated. As all layers had been previously reprojected to the appropriate UTM zone (i.e., the map units were now meters), this could be simply achieved with the Field Calculator. In order to classify bridges affected by the floods, the dataset was filtered for where "*bridge = yes*", taking advantage of the bridge attribute included in the initial dataset. This attribute does not distinguish between bridges across water and other types of bridges, i.e. across valleys or elevated stretches of roads. Since the main idea of analyzing the impact of floods on bridges is the potentially impaired ability to cross waterways, this is not an ideal scenario. However, since a better classifier was not readily available within the dataset, this was accepted as a limitation of the analysis. Another constraint of the raw dataset was the OSM custom to map some types of roads in parallel, especially large roads such as highways, where each direction of the road is mapped as a separate road "piece", with the consequence of showing up in the dataset with a separate ID and therefore as an independent road. In order to get around this limitation, the OSM attribute "oneway" was used to identify roads where this was the case. For the post-analysis dataset (the affected pieces of infrastructure), the number of one way roads was then divided by two in order to get a better general estimate of possibly damaged roads, so that the final estimate is calculated as follows:

$$n_{total} = n_{twoway} + \frac{n_{oneway}}{2}$$

where n_{total} = final estimate of number of affected infrastructure,
n_{twoway} = number of affected infrastructure where oneway=NULL,
n_{oneway} = number of affected infrastructure where oneway=yes,

This calculation was used for both roads and bridges. The Railway dataset did not show similar patterns.

Further, the flood extent was intersected with a land cover raster in order to get more information about what types of land were affected by the floods. This step also helps with identifying potential limitations of the approach taken in this work (see 4.3). The land cover raster used here was taken from ESA Worldcover [79], a newly available global land cover map at 10m resolution and in 3x3 degree tiles based on Sentinel-1 and Sentinel-2 input data. The land cover rasters covering the extent of the floods (N48E006 for the Germany case study; N33E111 and N33E114 for the Henan case study) were imported into QGIS and reprojected to the case study projection (UTM Zone 32N for the Germany case study, UTM Zone 49N for the Henan case study). In the case of Henan, the two rasters

were first cropped to the approximate area of interest to save computing time and then merged. For both case studies, the rasters were then clipped to the precise extent of the flooded area using the *Clip raster by mask layer* function and subsequently spatial statistics were calculated for the clipped raster using *Raster layer unique values report*. This tool generates a HTML report of the pixel values within a raster both for number of pixels as well as area covered in m^2.

3   RESULTS

3.1 Impact Analysis Results of Germany Case Study

The total flood extent as classified was 10.8 km^2. A total of 171 buildings and 1042 roads of all types covering a length of 46.6 km were shown to be affected by the disaster, including 10 bridges. Additionally, 19 railway lines or pieces with a total length of 819 m were shown to be within the extent of the flood.

In terms of land cover, the overwhelming amount of terrain was cropland at 6.27 km^2 or just over 58%. Other major land cover types were grassland at 3.05 km^2 or 28% and tree cover at 1.17 km^2 or just under 11%. Settlements (classified as built-up area) included only 2.2% of the total flood extent.

The methodology failed to classify significant amounts of flooded land in the Ahr valley, where the worst of the flooding occurred, and the cost of lives was highest (see 3.1.1). A possible reason for this is the narrow nature of the river valley, which led to flash flooding and fast receding of the water level. However, the satellite data was obtained at the peak of the disaster on 15th of July 2021, where the water level of the Ahr should have still been above normal values, especially when considering the partly reconstructed water level (see 3.1.1) on that day. Another possible reason is the hilly nature of the surroundings, in contrast to the flat landscape of the Erft and Swist valleys further north (e.g. Figure 4.3), which may result in misclassification of radar data, depending on the processing. In this work, no special requirements for mountainous regions were considered, as the majority of the area of interest was flat landscape. Other classifications of the flood such as the one by the Dartmouth Flood Observatory show similar results for the Ahr valley [80].

The classification of the landscape further north around Euskirchen generated a much more complete outcome with regards to expectations. Large areas, mostly crop- and grassland, were shown to be inundated at the time of the satellite flyover. The percentage of urban or built-up area was however low.

The following tables show the spatial statistics as calculated from intersecting infrastructure and land cover data with the raw flood extent. The figures show an overview of the classified flood extent as a whole as well as several large-scale views of different areas affected by the disaster.

Table 4.1: Classified flood extent by land cover type and total area in km^2 (Germany).

Land Cover Type	Area[1]
Tree Cover	1.166
Shrubland	0.000
Grassland	3.049
Cropland	6.270
Built-up	0.198
Bare / Sparse vegetation	0.084
Herbaceous Wetland	0.002
Total area	**10.769**
Misclassified water	*2.067*

[1] in square km

Table 4.2: Infrastructure affected by the flood (Germany).

Type	Amount	Length[1]
Roads	1042	46.6
Railways	19	0.8
Bridges	10	NA
Buildings	171	NA

[1] in km
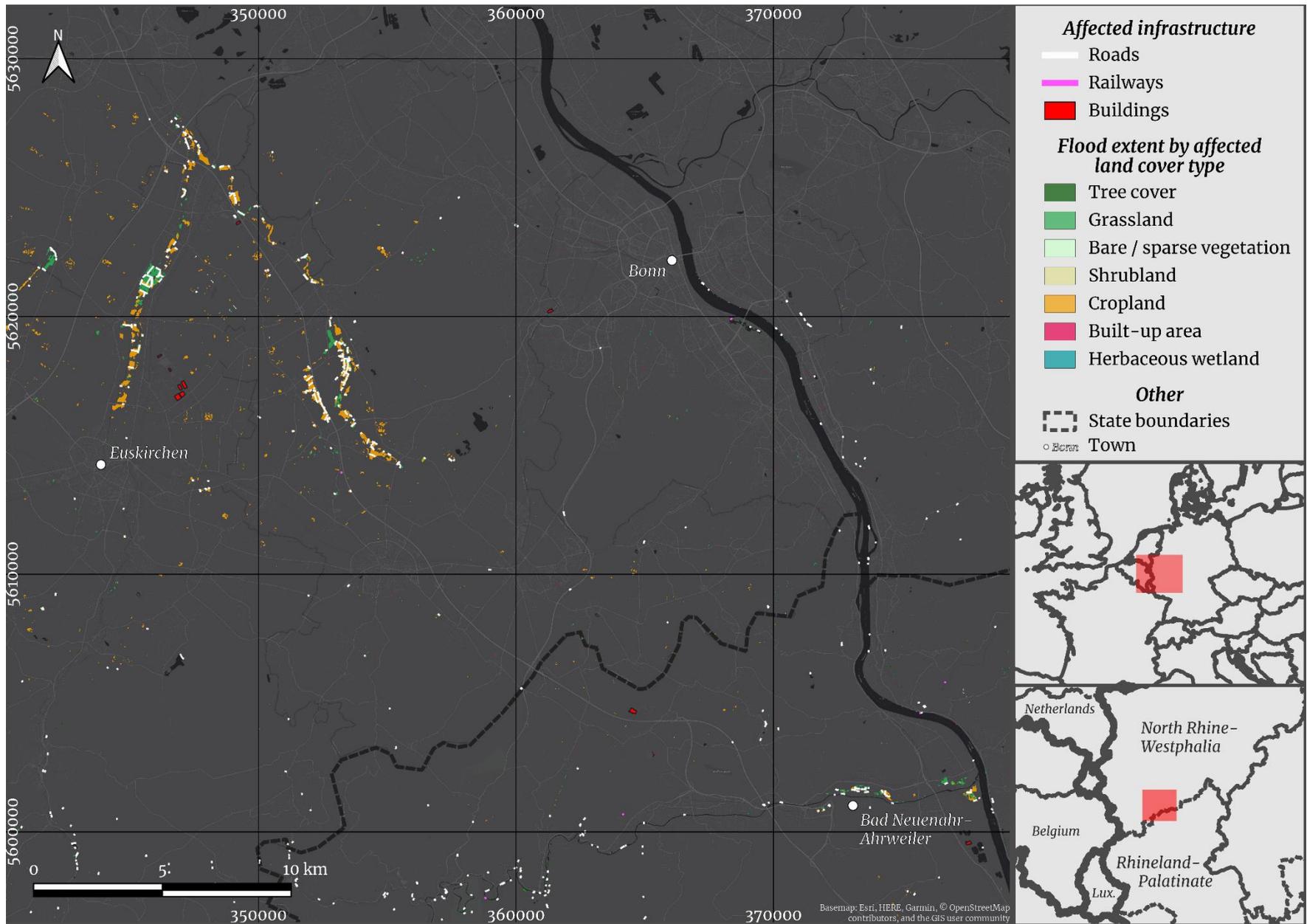
Figure 7: Small-scale overview of the Germany case study at 1:170 000.

Figure 8: Large-scale view of part of the Germany study at 1:15 000. Village of Lommersum, North Rhine-Westphalia.

Figure 9: Large-scale view of part of the Germany study at 1:15 000. Swist Valley, North Rhine-Westphalia.
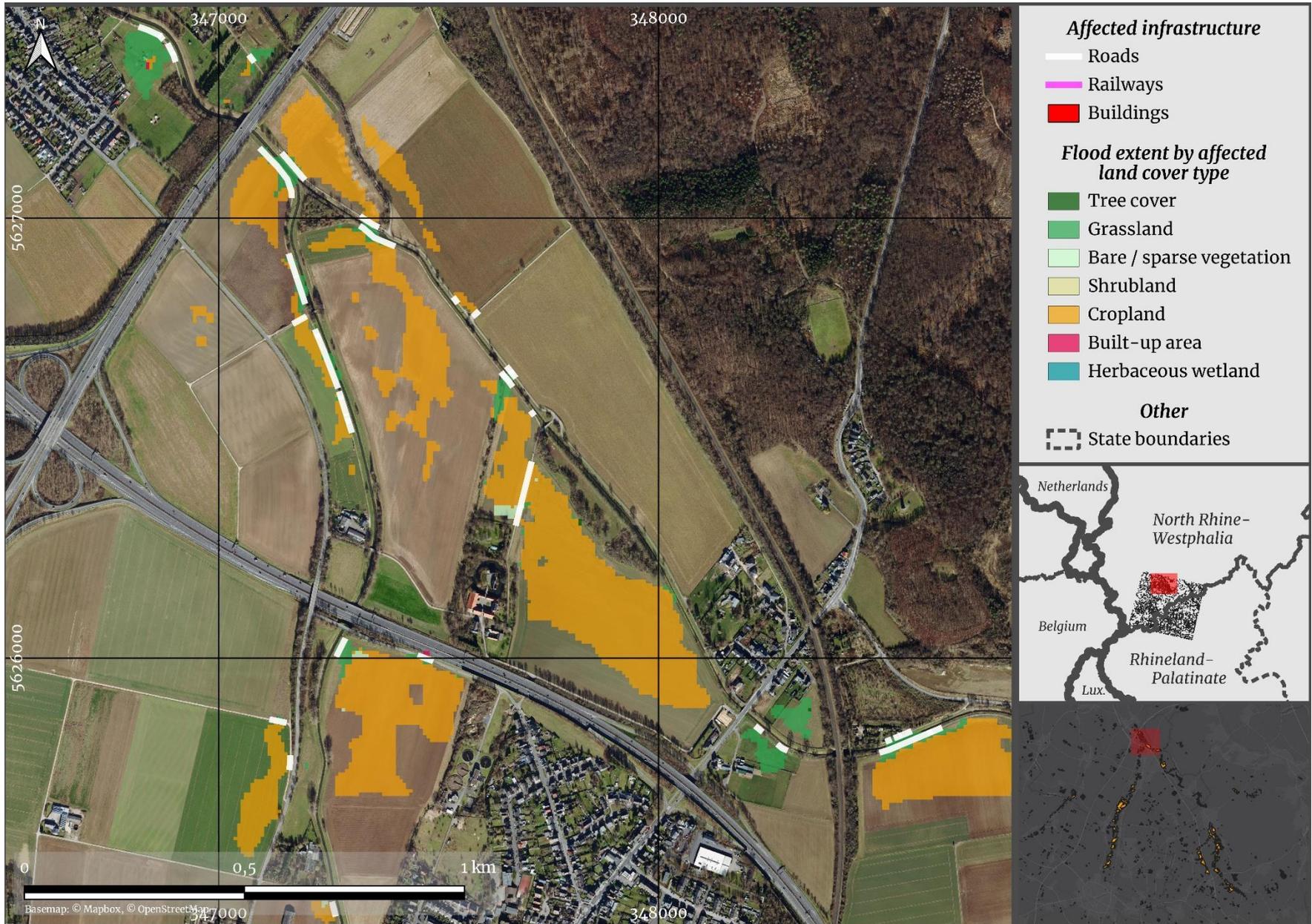
Figure 10: Large-scale view of part of the Germany study at 1:10 000. Confluence of Swist and Erft Rivers, North Rhine-Westphalia.
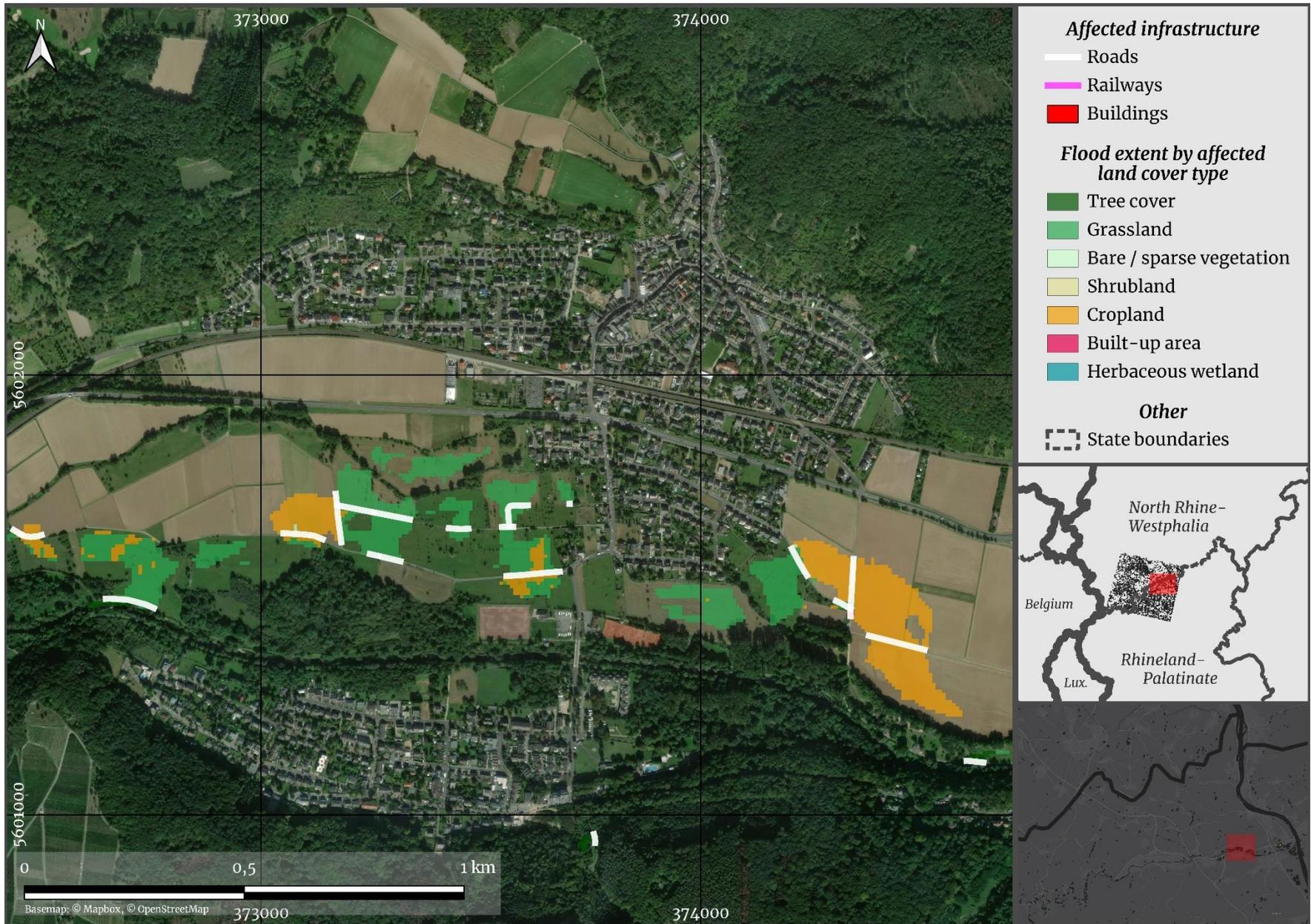
Figure 11: Large-scale view of part of the Germany study at 1:10 000. Village of Bad Bodendorf, Rhineland-Palatinate.

3.2 Impact Analysis Results of China Case Study

The total flood extent as classified was 437 km^2. 618 buildings of the OSM dataset were shown to be affected by the disaster. Additionally, 4809 roads of all types covering 941 km as well as 376 bridges were affected (see 3.4 and 4.3 for discussion on this result). 232 railway lines or pieces covering 109 km were classified as affected by the flood.

84.4% of the flood extent was classified as cropland, with only 18.9 km^2 or 4.47% classified as built-up.

The lack of identified buildings is a result of the fragmentary state of China within OSM. Relatively few individual structures are mapped and while this is less the case for the road and railway network, the dataset remains widely incomplete. The OSM Analytics tool notes 100,091 individual buildings and less than 200 contributors for Henan province [81], a region with a population of 99 million [82]. This issue is clearly shown in the figures below. The fragmentary nature of mapped buildings is especially clear in Figures 4.8 and 4.10, with large built-up areas classified as such, but very few or no individual buildings affected.

The vast majority of classified flooded land was cropland. In the first place, this is consistent with Henan province as a rural province and a major grain exporter with a relatively low urbanization rate [83]. On the other hand it is less consistent with the devastating human cost of the flood (as described in 3.1.2), most of which centered on urban areas and especially the city of Zhengzhou. This is further discussed in 4.3.

The following tables show the spatial statistics as calculated from intersecting infrastructure and land cover data with the raw flood extent. The figures show an overview of the classified flood extent as a whole as well as several large-scale views of different areas affected by the disaster.

Table 4.1: Classified flood extent by land cover type and total area in km^2 (China).

Land Cover Type	Area[1]
Tree Cover	8.990
Shrubland	0.046
Grassland	2.366
Cropland	357.094
Built–up	18.916
Bare / Sparse vegetation	35.431
Herbaceous Wetland	0.137
Total area	**422.979**
Misclassified water	*14.800*

[1] in square km

Table 4.2: Infrastructure affected by the flood (China).

Type	Amount	Length[1]
Roads	4809	941.0
Railways	232	109.0
Bridges	376	NA
Buildings	618	NA

[1] in km
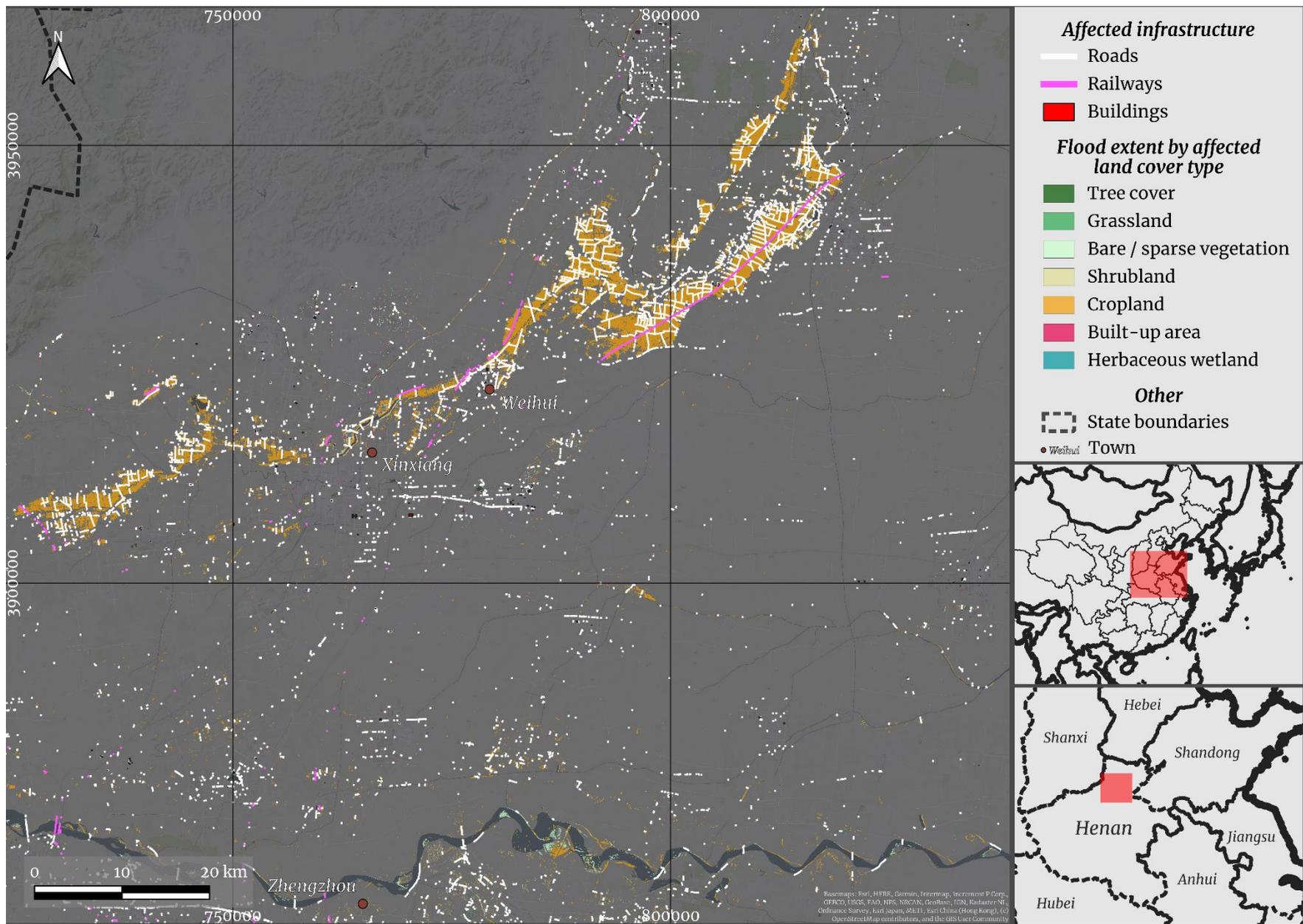
Figure 12: Small-scale view of the China case study at 1:1 000 000.
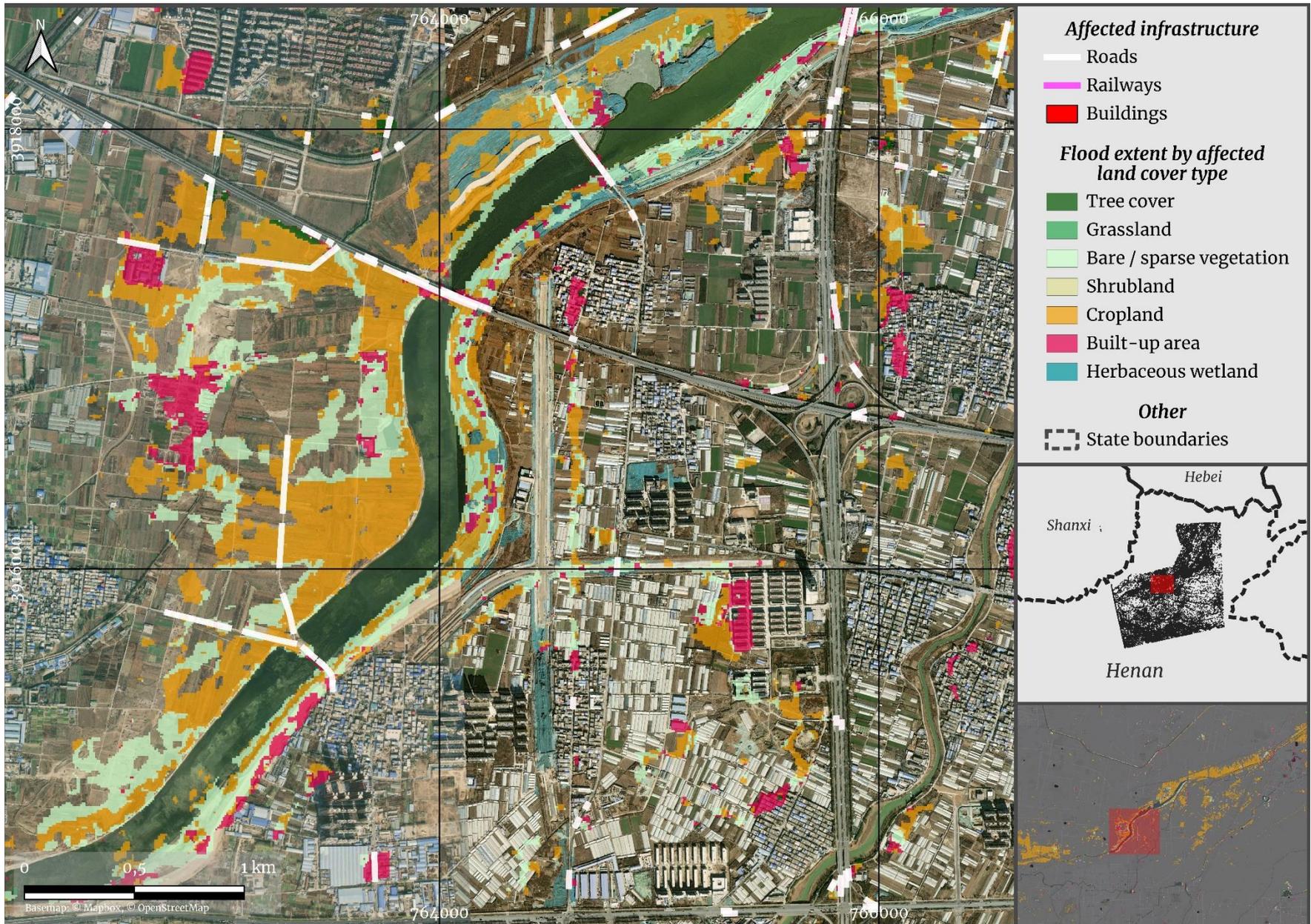
Figure 13: Large-scale view of part of the China case study at 1:20 000. Gongchanzhuyi Canal, Fengquan District, Henan Province.
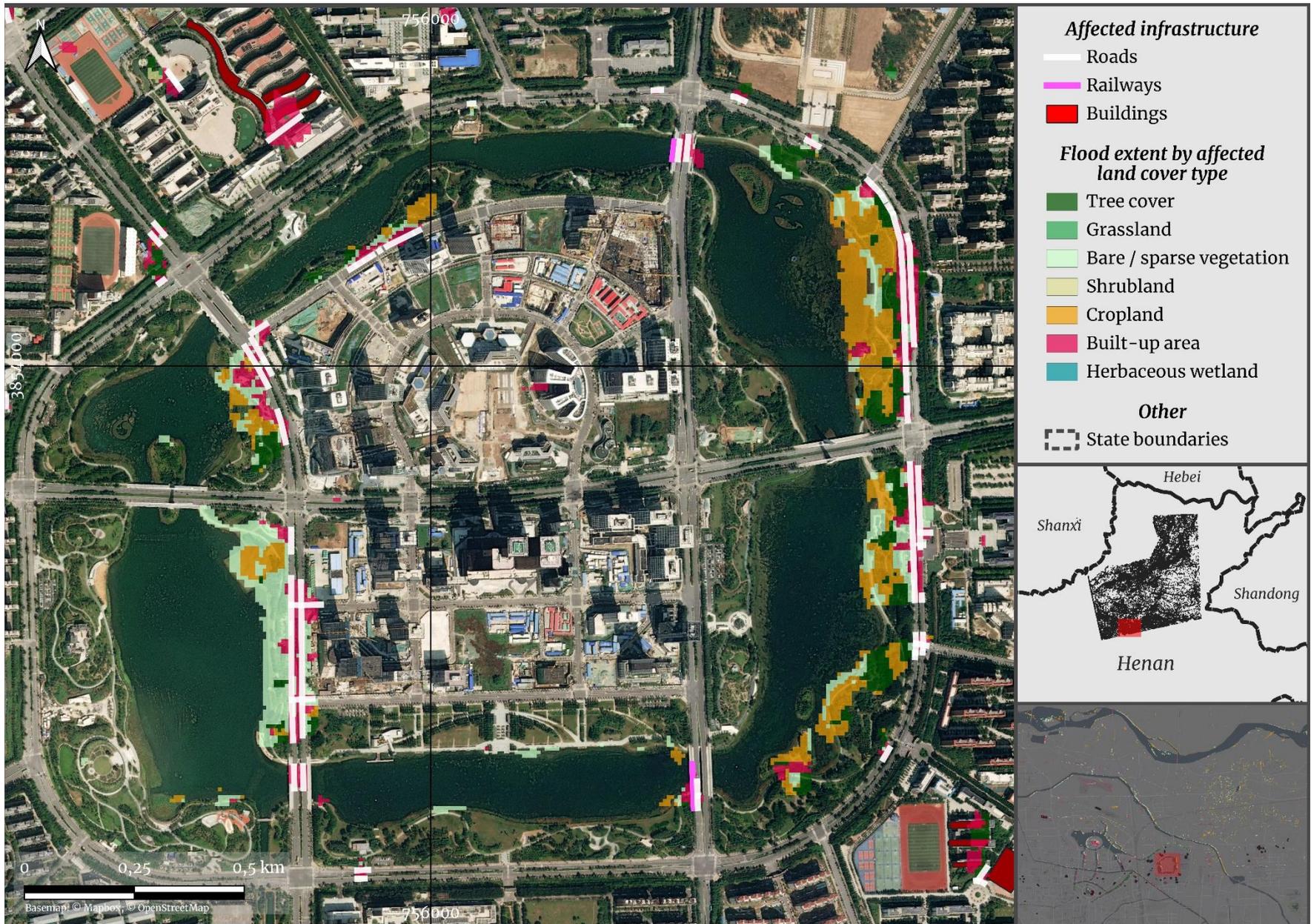
Figure 14: Large-scale view of part of the China case study at 1:10 000. Zhengzhou, Henan Province.
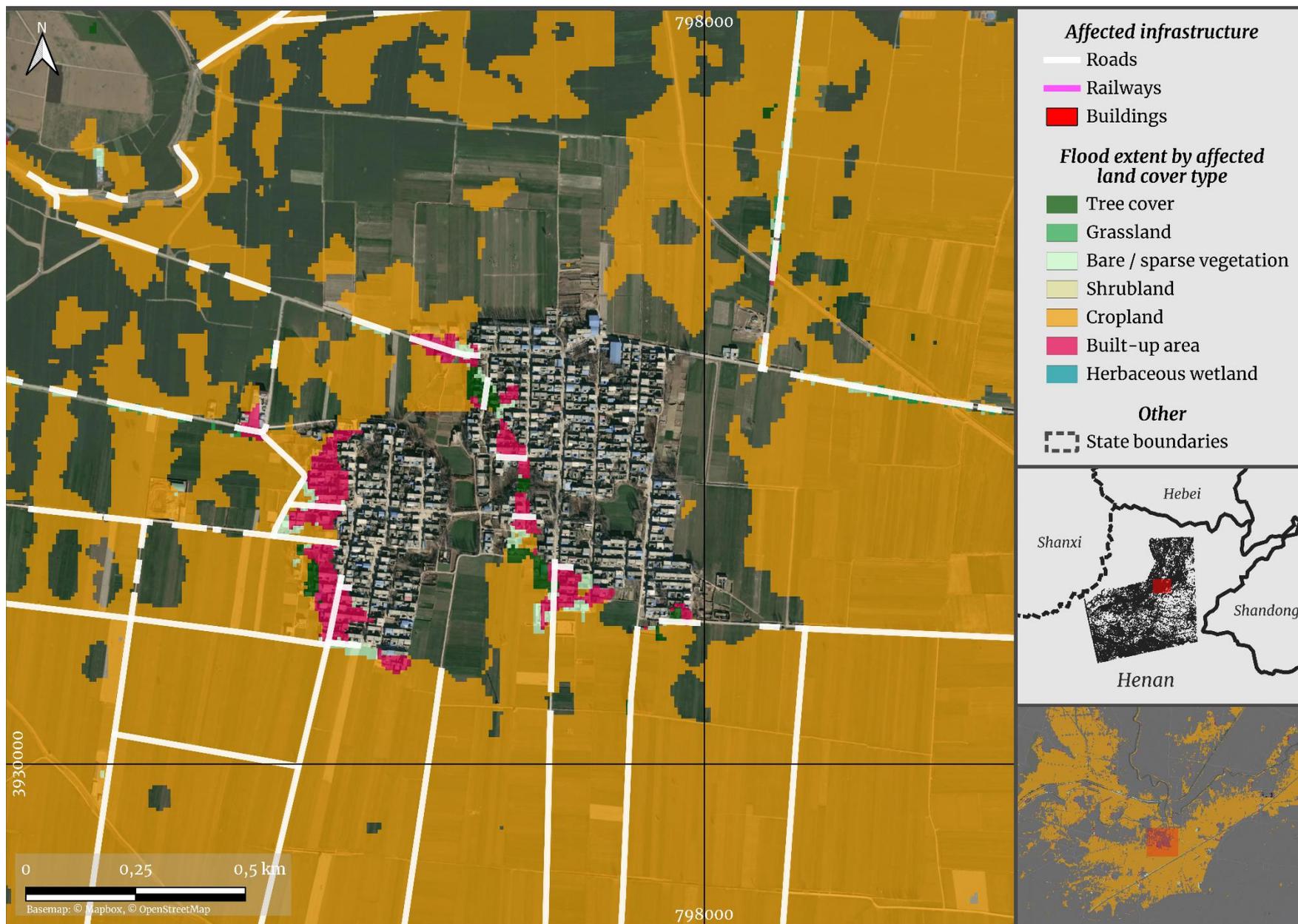
Figure 15: Large-scale view of part of the China case study at 1:10 000. West of Xinzhen, Xun County, Henan Province.
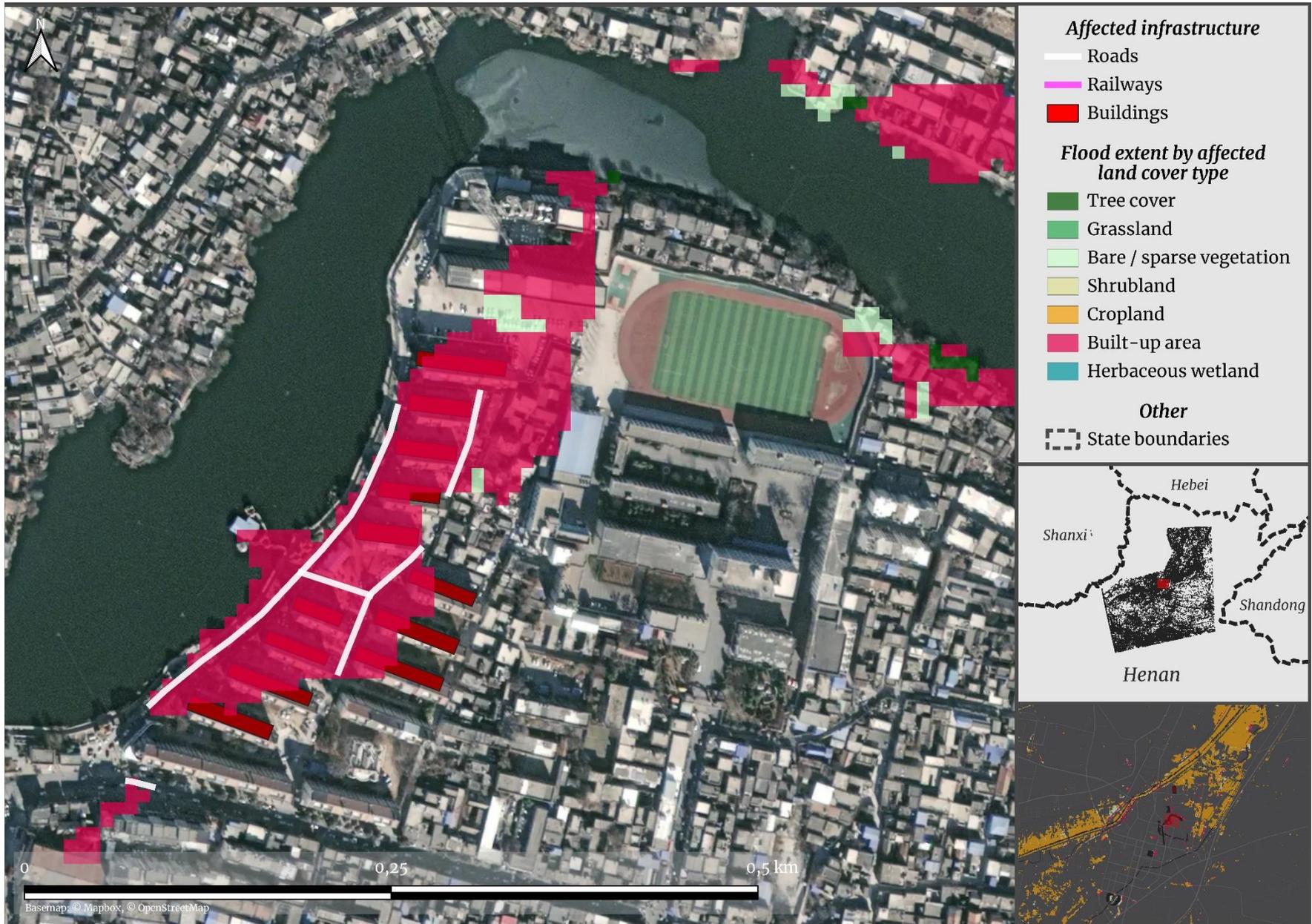
Figure 16: Large-scale view of part of the China case study at 1:3 000. Weihui City, Henan Province.

3.3 Discussion

The methodology allows a relatively quick overview of the flood extent. After downloading the data, the integration of data is a matter of hours. The largest time investments are with preprocessing the data ahead of calculating the flood impact. This is the case both for the raw SAR data, which must be preprocessed according to many factors, such as the topography of the landscape in question, as well as the OSM data. The landcover dataset can be used with minimal preparation. The longest potential delay when using this exact methodology lies in the availability of Sentinel-1 data, which at the current moment is affected by the malfunction of Sentinel-1B, prolonging revisit times to 12 days. Another consequence of this problem is the fact that a flood disaster may occur at any time and may not overlap with the revisit times at hand. As shown in this work, the Henan flood event peaked around the 17th to the 20th of July, but the next Sentinel-1 flyover did not happen until the 27th of July.

Certainly the biggest advantage of the methodology is in its global approach: SAR data from Sentinel-1, OSM data as well as the land cover dataset sourced from the Worldcover project are datasets at global scale that can be applied to any region on the planet. However, there are some caveats to be considered when using these datasets.

First, raw SAR data allows only nontrivial interpretation ahead of processing with tools such as SNAP and subsequent thresholding. When responding to an active disaster, comparative datasets are not available, and it is up to the user to decide whether thresholding was achieved in an appropriate and accurate manner. Another result was the misclassification of permanent water bodies as flooded land. For Germany, an area of 2.07 km^2 or 16.1% was misclassified, for Henan 14.8 km^2 or 3.38%. This error may be the consequence of thresholding inaccuracies in step 3.2, where some part of the permanent water bodies may have not been recognized as such in the "before" scene. Another possibility is an inaccuracy within the land cover dataset used for the spatial analysis. While the spatial resolution itself is very high at 10 m, the authors acknowledge a 74.4% global accuracy (though this ranges from 68-81% based on the continent). While this is high, it could explain some misattribution [84]. For comparison, the CORINE land cover program declares thematic accuracies of >85% [85]. This is a second caveat to consider and a potential source of misclassification in this work.

A general limitation for using OSM infrastructure is the unclear number of actual roads described within a dataset. While this problem is mitigated by intersecting the road and railway network with the precise extent of the flood, it is an issue of the dataset at large. Many separate roads marked with their own unique identifier are actually pieces of the same larger road, and on a bigger scale, the network of both roads and railways is more likely to reflect a pattern of user mapping styles than a perfect register of real-world infrastructure. Clearer and/or more standardized rules for displaying and mapping roads and railways in OSM could help mitigate this issue in the future.

A limitation for estimating the number of affected bridges was the lack of intrinsic partition within OSM between bridges across water and other types of bridges, as discussed in 3.4. The results here may therefore be viewed as overestimating the effect of the flood on actually crossing bodies of water.

The calculation as described in 3.4 may be seen as a simplification of the real situation, as there are evidently one-way roads that are not highways and do not have a parallel road running in the opposite direction. However, these roads are less numerous than (pieces of) highway or large roads,

especially in China, so that while both the initial dataset as well as the modified version are reduced versions of the real picture, the goal of the analysis was to make a post-disaster estimate, which the modified version comes closer to doing.

Finally, the integration of VGI and SAR data for Chinese territory is feasible only at the most basic level. China remains largely unmapped on OpenStreetMap not only in terms of specific individual buildings, but at every level of infrastructure mostly due to the unclear legal basis of VGI activity within Chinese territory. For further reading, [86] explored the sometimes paradoxical relationship between state, entrepreneurs and OSM contributors in China. In the meantime, an integrated analysis of flood damage with the help of exported infrastructure data from OSM is not possible at the same level as in Europe. To illustrate the stark difference in the extent of mapping; the Rhineland-Palatinate state of Germany, where the worst of the 2021 summer floods happened, has a population of about 4 million [87] and is represented by more than 1.7 million buildings mapped in OSM [88]. Meanwhile Henan province, center of the 2021 floods, is represented by just over 100,000 buildings [89], despite its population of more than 99 million [82].

It is clear that the impact analysis did not come close to estimating the real damage that transpired in the wake of these disasters. Only 171 buildings were flagged as being affected by the flood in (the extraordinarily well-mapped) Germany, while in reality whole villages were devastated and the total cost has been estimated at $40 billion [90]. Similarly, built-up area was only a small percentage of the classified flood extent for the China case study, and especially the denser city environments (Zhengzhou, Xinxiang cities) showed up in a fragmentary way at best. Integration of OSM data for China was complicated further, as the issues surrounding VGI in China mean that the majority of real world structures are missing in OSM and cannot be included in an analysis in the first place. The result is paradoxical – more urban environments that are more likely to be mapped at this incomplete stage (see 2.3) are less likely to be picked up by SAR data thresholding efforts, while more rural settlements are successfully detected as inundated, but few individual structures are mapped in these areas and subsequently show up in statistics. However, the methodology is fundamentally highly reproducible and applicable globally, so that it may be adapted quickly and developed further.

CONCLUSIONS

1. Flood disaster impact analysis has begun incorporating VGI and found it to be a useful addition to the more traditional methods of disaster management, but more case studies and research is needed.
2. Flood disaster impact analysis by satellite can be achieved using both active and passive remote sensing, although active remote sensing has the crucial advantage of not being reliant on weather or daytime.
3. A thresholding method for Sentinel-1 SAR data was chosen as the appropriate algorithm for detecting the flood extent. OSM infrastructure data and a global landcover dataset were chosen additionally in order to estimate the impact of the floods.
4. Two flood case studies, in Germany and China respectively, were chosen and the methodology was implemented technically on these examples. Flood extents were classified and intersected with the OSM and landcover data in order to produce statistical results as well as cartographic products in the form of maps.
5. The methodology worked well for free-standing areas such as grass- and cropland but did not yield satisfactory results for settled areas. Few individual building footprints were shown to intersect with the floods. Integration of OSM data for Chinese case studies is feasible only at the most basic level based on a lack of completeness of the dataset. However, the methodology is applicable quickly and globally and may be refined in the future.

FUTURE RESEARCH SUGGESTIONS

In order to further test the methodology, more case studies of floods should be considered in general. In particular, the impact of the time delay between peak flooding and satellite flyover may be tested in a more quantitative manner. The methodology should also be tested for more mountainous terrain, modifying the preprocessing algorithm used for the SAR data. To test the completeness of OSM data for this purpose, other regions of the world should be included. Different levels of completeness may be analyzed quantitatively to find a fit-for-purpose threshold, in particular as developing countries continue to lag behind in terms of OSM mapping.

To improve the performance of the methodology in built-up / urban areas, additions to the SAR thresholding classification could be considered. The most promising methods currently include LiDAR or other very high resolution terrain data sourced locally, such as by drone.

As Sentinel-1B continues to malfunction, Sentinel-1A's elapsed active mission time is at the time of writing more than a year beyond its schedule, forcing the question of whether or not the Sentinel program will continue to be a reliable source of SAR data. While future satellites with SAR technology have been announced by ESA, this is not a short-term solution and future data accessibility may well pose a problem. While other sources of SAR data are available, different providers may offer data structure and preprocessing that is different from Sentinel's Copernicus program. As such, other satellite programs should be investigated for their potential in this methodology. Data by the TerraSAR-X constellation may be considered for higher resolution SAR data.

Simon Philipp Herbst
Integrating Volunteered Geographic Information and SAR Data for Disaster Impact Analysis

SUMMARY

Floods are becoming more frequent and increasingly destructive. Disaster management is a well-developed field of research but relies on detailed data and localized procedures as well as context. Disaster response as the phase of disaster management in the immediate wake of the disaster is a complicated process that ties together decision makers, emergency response and the affected communities. In combination, disaster response often does not or cannot make use of data that is publicly and openly available.

Satellite data is becoming increasingly accurate and reliable. At the same time, much of this data is publicly available. Synthetic aperture radar data from such a satellite (Sentinel-1 of the Copernicus program by the European Space Agency) allows detecting the extent of floods irrespective of weather or daylight conditions and is accessible at good spatial and temporal resolution. This in theory allows real-time or near real-time data collection of the scale of flood disasters.

Volunteered Geographic Information (VGI) is a term coined to describe the location-based data collection and accumulation efforts by volunteers on the internet. OpenStreetMap (OSM) is an example of this process, having gathered millions of pieces of infrastructure, such as building footprints, on a single platform with consistent data structure throughout. This data has been mapped globally and separately by individuals in a crowdsourcing effort and can be exported at will for a multitude of use cases. While the quality and reliability of VGI in general and OSM in particular are part of an ongoing debate and can be flawed, they are a major source of infrastructure data that can be applied across the globe.

The main aim of this Master Thesis is to bring together these two types of data in order to evaluate their usefulness for flood disaster impact analysis.

Two case studies in Germany and China were selected. Synthetic aperture radar data from the Sentinel-1 satellites was classified for the flood extent with a thresholding method. VGI data from OSM was exported and intersected with the flood extent. Further, a global land cover dataset was also intersected with the flood extent.

The results showed that the methodology works well for free-standing areas such as grass- and cropland. It works less well with built-up urban areas, which is mainly a consequence of structures such as buildings causing complex backscattering effects for radar pulses.

The results additionally showed the fragmentary state of mapping of China in OSM. As such, individual building footprints are only rarely available.

The methodology can be applied rapidly and globally in a matter of hours once the algorithm is in place. The longest delays are caused by data preprocessing. The impact analysis is contingent on satellite data being available, which is limited by satellite flyover (revisit) times. Accurate analysis of affected structures is also dependent on how well-mapped the area of interest is within OSM.

Keywords: disaster impact analysis, flood disaster response, SAR data, volunteered geographic information, openstreetmap.

Simon Philipp Herbst
Bendruomenės kuriamos geografinės informacijos panaudojimas kartu su SAR duomenimis stichinių nelaimių teritorijų analizei

SANTRAUKA

Potvyniai tampa vis dažnesniais ir keliančiais vis daugiau niokojančių padarinių. Efektyvus ir greitas reagavimas į stichines nelaimes remiasi duomenų apie paveiktas teritorijas kokybe, kuri priklauso nuo naudojamų mokslinių tyrimų metodų, duomenų išsamumo, gamtinės ir urbanizuotos aplinkos, valstybės duomenų politikos ir pan. Įvykus stichinei nelaimei itin svarbiu tampa reakcijos laikas, per kurį gebama tikslingai organizuoti pagalbą. Tai pirmasis nelaimių valdymo etapas, kuris susieja sprendimus priimančiuosius su nukentėjusiomis bendruomenėmis ir greitosios pagalbos institucijomis. Tačiau dažnai, reaguojant į nelaimes, turimi duomenų rinkiniai dėl jų geografinės turinio aprėpties trūkumo ir aktualumo stokos (stichinės nelaimės smarkiai pakeičia aplinką) tampa mažai naudingais.

Palydoviniai duomenys tampa vis tikslesni ir patikimesni, taip pat daugelis šių duomenų yra viešai prieinami. Vieno tokių palydovų (Europos kosmoso agentūros "Copernicus" programos Sentinel-1) SAR duomenys leidžia nustatyti potvynių mastą nepriklausomai nuo oro ar dienos šviesos sąlygų ir šie duomenys yra prieinami gera erdvine ir laiko raiška, tinkama pagalbos darbų organizavimui. Teoriškai, šie duomenys leidžia beveik realiuoju laiku rinkti duomenis apie potvynių mastą.

Bendruomenės kuriama geografinė informacija (angl. Volunteered Geographic Information, VGI) – tai terminas, kuriuo apibūdinamos savanorių pastangos rinkti ir kaupti vietos informacija pagrįstus duomenis internete. Šio proceso pavyzdys yra OpenStreetMap (OSM), surinkęs daugybę infrastruktūros objektų, tokių kaip pastatų žymenys, vienoje platformoje su nuoseklia visų duomenų struktūra. Šiuos duomenis pasauliniu mastu savarankiškomis pastangomis žemėlapiuose žymi pavieniai žmonės. Surinktus duomenis galima eksportuoti pagal poreikį įvairiems panaudojimo būdams. Nors OSM, kaip VGI dalies, kokybė ir informacijos patikimumas yra nuolatinių diskusijų dalis ir gali turėti trūkumų, tai yra pagrindinis infrastruktūros duomenų šaltinis, kurį galima naudoti visame pasaulyje.

Pagrindinis šio magistro darbo tikslas – sujungti šiuos, dviem būdais gautus, duomenų tipus, kad galėtų būti įvertintas potvynių poveikis aplinkai.

Buvo pasirinkti tirti Vokietijos ir Kinijos pavyzdžiai. Sintetinės apertūros radaro duomenys iš palydovo Sentinel-1 buvo klasifikuojami potvynio mastui nustatyti taikant slenksčio nustatymo metodą. VGI duomenys iš OSM buvo eksportuoti ir susieti su potvynio mastu. Pasaulinis žemės paviršiaus duomenų rinkinys taip pat buvo integruotas su potvynio masto duomenimis.

Rezultatai parodė, kad ši metodika ypač tinkama taikyti neapstatytoms teritorijoms, tokioms kaip laukai ar dirbamos žemės. Metodika prasčiau veikia, kai taikoma užstatytoms miesto teritorijoms, kas daugiausiai susiję su tokiomis struktūromis kaip pastatai, sukeliančiomis sudėtingą radaro impulsų grįžtamosios sklaidos poveikį.

Be to, rezultatai parodė, kad Kinijos žemėlapiai OSM sistemoje yra fragmentiški, todėl pavienių pastatų žymenys prieinami tik retais atvejais.

Sukūrus algoritmą, metodiką galima taikyti greitai ir globaliai per kelias valandas, kur ilgiausiai užtrunka pirminis duomenų apdorojimas. Poveikio analizė priklauso nuo to, ar yra palydovinių duomenų, o tai riboja palydovų sugrįžimo į buvusią padėtį užtrunkamas laikas. Tiksli paveiktų struktūrų analizė taip pat priklauso nuo to, kaip išsamiai OSM žemėlapyje yra sužymėta tiriama teritorija.

Raktažodžiai: nelaimės poveikio analizė, reagavimas į potvynius, SAR duomenys, Bendruomenės kuriama geografinė informacija, Openstreetmap.

BIBLIOGRAPHY

[1]   Swiss Re Institute, Natural catastrophes in 2021: the floodgates are open, (2022).

[2]   J. Rentschler, M. Salhab, People in Harm's Way : Flood Exposure and Poverty in 189 Countries, 2020. https://openknowledge.worldbank.org/handle/10986/34655.

[3]   S.N. Jonkman, J.K. Vrijling, Loss of life due to floods, J. Flood Risk Manag. 1 (2008) 43–56. https://doi.org/10.1111/j.1753-318x.2008.00006.x.

[4]   IPCC, Summary for Policymakers, Cambridge University Press, 2012.

[5]   Q. Zhang, X. Gu, V.P. Singh, P. Shi, P. Sun, More frequent flooding? Changes in flood frequency in the Pearl River basin, China, since 1951 and over the past 1000 years, Hydrol. Earth Syst. Sci. 22 (2018) 2637–2653. https://doi.org/10.5194/hess-22-2637-2018.

[6]   D. Panagoulia, G. Dimou, Sensitivity of flood events to global climate change, J. Hydrol. 191 (1997) 208–222. https://doi.org/10.1016/S0022-1694(96)03056-9.

[7]   N. Altay, W.G. Green, OR/MS research in disaster operations management, Eur. J. Oper. Res. 175 (2006) 475–493. https://doi.org/10.1016/j.ejor.2005.05.016.

[8]   B. Barroca, P. Bernardara, J.M. Mouchel, G. Hubert, Indicators for identification of urban flooding vulnerability, Nat. Hazards Earth Syst. Sci. 6 (2006) 553–561. https://doi.org/10.5194/nhess-6-553-2006.

[9]   E.J. Plate, Flood risk and flood management, J. Hydrol. 267 (2002) 2–11. https://doi.org/10.1016/S0022-1694(02)00135-X.

[10]   B. Merz, A.H. Thieken, Flood risk analysis: Concepts and challenges, Österreichische Wasser- Und Abfallwirtschaft. (2004) 27–34.

[11]   U. Herrmann, A.H. Thieken, U. Suhr, K.-E. Lindenschmidt, Flood risk analyses on the River Elbe - Methods and data resolution , Osterr. Wasser- Und Abfallwirtschaft. 59 (2007) 151–162. https://www.scopus.com/inward/record.uri?eid=2-s2.0-37749053071&doi=10.1007%2Fs00506-007-0137-7&partnerID=40&md5=cc7dc5f12211675cd0d6d642f91891ab.

[12]   M. Isma'il, I.O. Saanyol, Application of Remote Sensing (RS) and Geographic Information Systems (GIS) in flood vulnerability mapping : Case study of River Kaduna, Int. J. Geomatics Geociences. 3 (2013) 618–627.

[13]   S.J. Schelhorn, B. Herfort, R. Leiner, A. Zipf, J.P. De Albuquerque, Identifying elements at risk from OpenStreetMap: The case of flooding, ISCRAM 2014 Conf. Proc. - 11th Int. Conf. Inf. Syst. Cris. Response Manag. (2014) 508–512.

[14]   M. Cerri, M. Steinhausen, H. Kreibich, K. Schröter, Are OpenStreetMap building data useful for flood vulnerability modelling?, Nat. Hazards Earth Syst. Sci. 21 (2021) 643–662. https://doi.org/10.5194/nhess-21-643-2021.

[15]   B. Tomaszewski, M. Judex, J. Szarzynski, C. Radestock, L. Wirkus, Geographic Information Systems for Disaster Response: A Review, J. Homel. Secur. Emerg. Manag. 12 (2015) 571–602. https://doi.org/10.1515/jhsem-2014-0082.

[16]   M. Zook, M. Graham, T. Shelton, S. Gorman, Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake, World Med. Heal. Policy. 2 (2010) 6–32. https://doi.org/10.2202/1948-4682.1069.

[17]   Humanitarian OpenStreetMap Team | Home, (n.d.). https://www.hotosm.org/ (accessed April 19, 2022).

[18]   T.H. Poiani, R.D.S. Rocha, L.C. Degrossi, J.P. De Albuquerque, Potential of collaborative mapping for disaster relief: A case study of openstreetmap in the Nepal earthquake 2015, Proc. Annu. Hawaii Int. Conf. Syst. Sci. 2016-March (2016) 188–197. https://doi.org/10.1109/HICSS.2016.31.

[19] C. Westrope, R. Banick, M. Levine, Groundtruthing OpenStreetMap building damage assessment, Procedia Eng. 78 (2014) 29–39. https://doi.org/10.1016/j.proeng.2014.07.035.

[20] B. Jongman, H. Kreibich, H. Apel, J.I. Barredo, P.D. Bates, L. Feyen, A. Gericke, J. Neal, J.C.J.H. Aerts, P.J. Ward, Comparative flood damage model assessment: Towards a European approach, Nat. Hazards Earth Syst. Sci. 12 (2012) 3733–3752. https://doi.org/10.5194/nhess-12-3733-2012.

[21] M. Kok, H.J. Huizinga, A. Barendregt, Standard Method 2004: Damage and Casualties Caused by Flooding, (2005) 56.

[22] C. Scawthorn, P. Flores, N. Blais, H. Seligson, E. Tate, S. Chang, E. Mifflin, W. Thomas, J. Murphy, C. Jones, M. Lawrence, HAZUS-MH Flood Loss Estimation Methodology. II. Damage and Loss Assessment, Nat. Hazards Rev. 7 (2006) 72–81. https://doi.org/10.1061/(ASCE)1527-6988(2006)7:2(72).

[23] P.C. Oddo, A. Ahamed, J.D. Bolten, Socioeconomic impact evaluation for near real-time flood detection in the Lower Mekong River Basin, Hydrology. 5 (2018) 1–17. https://doi.org/10.3390/hydrology5020023.

[24] P.C. Oddo, J.D. Bolten, The Value of Near Real-Time Earth Observations for Improved Flood Disaster Response, Front. Environ. Sci. 7 (2019) 1–11. https://doi.org/10.3389/fenvs.2019.00127.

[25] C. Van Westen, Remote sensing for natural disaster management, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 33 (2000) 1609–1617.

[26] K.E. Joyce, S.E. Belliss, S. V. Samsonov, S.J. McNeill, P.J. Glassey, A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters, Prog. Phys. Geogr. 33 (2009) 183–207. https://doi.org/10.1177/0309133309339563.

[27] T.W. Gillespie, J. Chu, E. Frankenberg, D. Thomas, Assessment and prediction of natural hazards from satellite imagery, Prog. Phys. Geogr. 31 (2007) 459–470. https://doi.org/10.1177/0309133307083296.

[28] G. Metternicht, L. Hurni, R. Gogu, Remote sensing of landslides: An analysis of the potential contribution to geo-spatial systems for hazard assessment in mountainous environments, Remote Sens. Environ. 98 (2005) 284–303. https://doi.org/10.1016/j.rse.2005.08.004.

[29] L. Selmi, Flood mapping using the Sentinel-1 imagery and the ESA SNAP S1 Toolbox, (2021).

[30] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, G. Ruello, Unsupervised Rapid Flood Mapping Using Sentinel-1 GRD SAR Images, IEEE Trans. Geosci. Remote Sens. 56 (2018) 3290–3299. https://doi.org/10.1109/TGRS.2018.2797536.

[31] K. Voormansik, J. Praks, O. Antropov, J. Jagomagi, K. Zalite, Flood mapping with terraSAR-X in forested regions in estonia, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (2014) 562–577. https://doi.org/10.1109/JSTARS.2013.2283340.

[32] V. Falťan, M. Bánovský, M. Blažek, Evaluation of land cover changes after extraordinary windstorm by using the land cover metrics: a case study on the high Tatras foothill, Geografie. 116 (2011) 156–171. https://doi.org/10.37040/geografie2011116020156.

[33] G. Balamurugan, S.M. Aravind, Land use land cover changes in pre- and postearthquake affected area using geoinformatics - Western Coast of Gujarat, India, Disaster Adv. 8 (2015) 1–14.

[34] European Space Agency Copernicus Program, Sentinel-1 - Mission Summary, (2014). https://sentinel.esa.int/web/sentinel/missions/sentinel-1/overview/mission-summary (accessed April 24, 2022).

[35] Copernicus Sentinel-1B anomaly (5th update) - Copernicus Sentinel-1B anomaly (5th update) - Sentinel Online, (n.d.). https://sentinels.copernicus.eu/web/sentinel/-/copernicus-sentinel-1b-anomaly-5th-update/1.2 (accessed May 1, 2022).

[36] N. Pettorelli, J.O. Vik, A. Mysterud, J.M. Gaillard, C.J. Tucker, N.C. Stenseth, Using the satellite-derived NDVI to assess ecological responses to environmental change, Trends Ecol. Evol. 20 (2005) 503–510. https://doi.org/10.1016/j.tree.2005.05.011.

[37] R. Brakenridge, E. Anderson, MODIS-based flood detection, mapping and measurement: the potential for operational hydrological applications, in: Transbound. Floods Reducing Risks Through Flood Manag., Springer, Dordrecht, 2006: pp. 1–12.

[38] D. Giordan, D. Notti, A. Villa, F. Zucca, F. Calò, A. Pepe, F. Dutto, P. Pari, M. Baldo, P. Allasia, Low cost, multiscale and multi-sensor application for flooded area mapping, Nat. Hazards Earth Syst. Sci. 18 (2018) 1493–1516. https://doi.org/10.5194/nhess-18-1493-2018.

[39] B.C. Gao, NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space, Remote Sens. Environ. 58 (1996) 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3.

[40] X. Huang, C. Wang, Z. Li, A near real-time flood-mapping approach by integrating social media and post-event satellite imagery, Ann. GIS. 24 (2018) 113–123. https://doi.org/10.1080/19475683.2018.1450787.

[41] Y. Kwak, B. Arifuzzanman, Y. Iwami, Prompt proxy mapping of flood damaged rice fields using MODIS-derived indices, Remote Sens. 7 (2015) 15969–15988. https://doi.org/10.3390/rs71215805.

[42] Y. Kwak, J. Park, K. Fukami, Near real-time flood volume estimation from MODIS time-series imagery in the indus river basin, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (2014) 578–586. https://doi.org/10.1109/JSTARS.2013.2284607.

[43] K. Chandrasekar, M.V.R. Sesha Sai, P.S. Roy, R.S. Dwevedi, Land Surface Water Index (LSWI) response to rainfall and NDVI using the MODIS vegetation index product, Int. J. Remote Sens. 31 (2010) 3987–4005. https://doi.org/10.1080/01431160802575653.

[44] J. Keuchel, S. Naumann, M. Heiler, A. Siegmund, Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data, Remote Sens. Environ. 86 (2003) 530–541. https://doi.org/10.1016/S0034-4257(03)00130-5.

[45] V. Walter, Object-based classification of remote sensing data for change detection, ISPRS J. Photogramm. Remote Sens. 58 (2004) 225–238. https://doi.org/10.1016/j.isprsjprs.2003.09.007.

[46] Collection of custom scripts | Sentinel-Hub custom scripts, (n.d.). https://custom-scripts.sentinel-hub.com/ (accessed April 30, 2022).

[47] D. Sun, S. Li, W. Zheng, A. Croitoru, A. Stefanidis, M. Goldberg, Mapping floods due to Hurricane Sandy using NPP VIIRS and ATMS data and geotagged Flickr imagery, Http://Dx.Doi.Org/10.1080/17538947.2015.1040474. 9 (2015) 427–441. https://doi.org/10.1080/17538947.2015.1040474.

[48] V. Herrera-Cruz, F. Koudogbo, TerraSAR-X rapid mapping for flood events, Proc. Int. Soc. Photogramm. Remote Sens. (Earth Imaging Geospatial Information), Hann. Ger. (2009) 170–175. http://cmsv021.rrzn.uni-hannover.de/fileadmin/institut/pdf/isprs-Hannover2009/Herrera-Cruz-170.pdf.

[49] M. Ohki, M. Watanabe, R. Natsuaki, T. Motohka, H. Nagai, T. Tadono, S. Suzuki, K. Ishii, Flood area detection using ALOS-2 PALSAR-2 data for the 2015 heavy rainfall disaster in the Kanto and Tohoku area, Japan, J. Remote Sens. Soc. Japan. 36 (2016) 348–359. https://ci.nii.ac.jp/naid/130005475775/.

[50] F. Bioresita, A. Puissant, A. Stumpf, J.P. Malet, A method for automatic and rapid mapping of water surfaces from Sentinel-1 imagery, Remote Sens. 10 (2018). https://doi.org/10.3390/rs10020217.

[51] K. Uddin, M.A. Matin, F.J. Meyer, Operational flood mapping using multi-temporal Sentinel-

1   SAR images: A case study from Bangladesh, Remote Sens. 11 (2019). https://doi.org/10.3390/rs11131581.

[52] M. Zhang, F. Chen, D. Liang, B. Tian, A. Yang, Use of sentinel-1 grd sar images to delineate flood extent in Pakistan, Sustain. 12 (2020) 1–19. https://doi.org/10.3390/su12145784.

[53] Y. Li, S. Martinis, M. Wieland, S. Schlaffer, R. Natsuaki, Urban flood mapping using SAR intensity and interferometric coherence via Bayesian network fusion, Remote Sens. 11 (2019). https://doi.org/10.3390/rs11192231.

[54] S. Martinis, Automatic near real-time flood detection in high resolution X-band synthetic aperture radar satellite data using context-based classification on irregular graphs, (2010).

[55] L. Giustarini, R. Hostache, P. Matgen, G.J.P. Schumann, P.D. Bates, D.C. Mason, A change detection approach to flood mapping in Urban areas using TerraSAR-X, IEEE Trans. Geosci. Remote Sens. 51 (2013) 2417–2430. https://doi.org/10.1109/TGRS.2012.2210901.

[56] D.C. Mason, I.J. Davenport, J.C. Neal, G.J.P. Schumann, P.D. Bates, Near real-time flood detection in urban and rural areas using high-resolution synthetic aperture radar images, IEEE Trans. Geosci. Remote Sens. 50 (2012) 3041–3052. https://doi.org/10.1109/TGRS.2011.2178030.

[57] Flood Mapping With Sentinel-1 Script | Sentinel-Hub custom scripts, (n.d.). https://custom-scripts.sentinel-hub.com/sentinel-1/flood_mapping/ (accessed April 30, 2022).

[58] M.F. Goodchild, Citizens as sensors: The world of volunteered geography, GeoJournal. 69 (2007) 211–221. https://doi.org/10.1007/s10708-007-9111-y.

[59] J. Jokar Arsanjani, M. Bakillah, Understanding the potential relationship between the socio-economic variables and contributions to OpenStreetMap, Int. J. Digit. Earth. 8 (2015) 861–876. https://doi.org/10.1080/17538947.2014.951081.

[60] M. Minaei, Evolution, density and completeness of OpenStreetMap road networks in developing countries: The case of Iran, Appl. Geogr. 119 (2020) 102246. https://doi.org/10.1016/j.apgeog.2020.102246.

[61] A.J. Flanagin, M.J. Metzger, The credibility of volunteered geographic information, GeoJournal. 72 (2008) 137–148. https://doi.org/10.1007/s10708-008-9188-y.

[62] K. Poser, D. Dransch, Volunteered geographic information for disaster management with application to rapid flood damage estimation, Geomatica. 64 (2010) 89–98.

[63] M.F. Goodchild, L. Li, Assuring the quality of volunteered geographic information, Spat. Stat. 1 (2012) 110–120. https://doi.org/10.1016/j.spasta.2012.03.002.

[64] P. Sultana, P. Thompson, C. Green, Can England learn lessons from Bangladesh in introducing participatory floodplain management?, Water Resour. Manag. 22 (2008) 357–376. https://doi.org/10.1007/s11269-007-9166-z.

[65] L.S. Fore, K. Paulsen, K. O'Laughlin, Assessing the performance of volunteers in monitoring streams, Freshw. Biol. 46 (2001) 109–123. https://doi.org/10.1046/j.1365-2427.2001.00640.x.

[66] Summer Floods in Europe 2021 | JBA Risk Management Event Response, (n.d.). https://www.jbarisk.com/flood-services/event-response/summer-floods-in-europe-2021/ (accessed April 30, 2022).

[67] Hochwasser im Ahrtal: Protokoll der Flutkatastrophe - SWR Aktuell, (n.d.). https://www.swr.de/swraktuell/rheinland-pfalz/flut-rekonstruktion-ahrtal-protokoll-100.html (accessed May 16, 2022).

[68] Germany Floods: Climate Change Moves to Center of Campaign as Toll Mounts - The New York Times, (n.d.). https://www.nytimes.com/2021/07/17/world/europe/germany-floods-climate-change.html (accessed April 30, 2022).

[69] How a deadly downpour flooded Zhengzhou, China - The Washington Post, (n.d.). https://www.washingtonpost.com/weather/2021/07/21/zhengzhou-china-record-rain-flooding/

(accessed April 30, 2022).

[70]   China – Over 350,000 Evacuated, 33 Dead, 8 Missing After Zhengzhou and Henan Floods – FloodList, (n.d.). https://floodlist.com/asia/china-henan-zhengzhou-floods-update-july-2021 (accessed April 30, 2022).

[71]   Chinese officials arrested for concealing true scale of flood death toll | China | The Guardian, (n.d.). https://www.theguardian.com/world/2022/jan/23/chinese-provincial-officials-concealed-scores-of-deaths-from-flood-disaster (accessed April 30, 2022).

[72]   Post Event Report: Henan Flood - July 17-21, (n.d.). https://www.guycarp.com/insights/2021/07/post-event-report-henan-flood-july-17-21.html (accessed May 18, 2022).

[73]   2021 Flooding China 5096, (n.d.). https://floodobservatory.colorado.edu/Events/5096/2021China5096.html (accessed May 18, 2022).

[74]   Open Access Hub, (n.d.). https://scihub.copernicus.eu/ (accessed May 18, 2022).

[75]   SNAP – STEP, (n.d.). http://step.esa.int/main/toolboxes/snap/ (accessed May 1, 2022).

[76]   D. O'Grady, M. Leblanc, A. Bass, The use of radar satellite data from multiple incidence angles improves surface water mapping, Remote Sens. Environ. 140 (2014) 652–664. https://doi.org/10.1016/j.rse.2013.10.006.

[77]   H. Lee, T. Yuan, H.C. Jung, E. Beighley, Mapping wetland water depths over the central Congo Basin using PALSAR ScanSAR, Envisat altimetry, and MODIS VCF data, Remote Sens. Environ. 159 (2015) 70–79. https://doi.org/10.1016/j.rse.2014.11.030.

[78]   HOT OSM Team, HOT Export Tool, (2022). https://export.hotosm.org/en/v3/ (accessed April 15, 2022).

[79]   D. Zanaga, R. Van De Kerchove, W. De Keersmaecker, N. Souverijns, C. Brockmann, R. Quast, J. Wevers, A. Grosu, A. Paccini, S. Vergnaud, O. Cartus, M. Santoro, S. Fritz, I. Georgieva, M. Lesiv, S. Carter, M. Herold, L. Li, N.-E. Tsendbazar, F. Ramoino, O. Arino, ESA WorldCover 10 m 2020 v100, (2021). https://doi.org/10.5281/ZENODO.5571936.

[80]   2021 Flooding Germany 5095, (n.d.). http://floodobservatory.colorado.edu/Events/5095/2021Germany5095.html (accessed May 16, 2022).

[81]   OSM Analytics Tool, (n.d.). https://osm-analytics.org/#/ (accessed May 16, 2022).

[82]   National Bureau of Statistics in China, Communiqué of the Seventh National Population Census (No. 3), (2021). http://www.stats.gov.cn/english/PressRelease/202105/t20210510_1817188.html (accessed April 15, 2022).

[83]   C. Zi, M. Qian, G. Baozhong, The consumption patterns and determining factors of rural household energy: A case study of Henan Province in China, Renew. Sustain. Energy Rev. 146 (2021) 111142. https://doi.org/10.1016/J.RSER.2021.111142.

[84]   Release of the 10 m WorldCover map, (n.d.). https://blog.vito.be/remotesensing/release-of-the-10-m-worldcover-map (accessed May 18, 2022).

[85]   CORINE Land Cover — Copernicus Land Monitoring Service, (n.d.). https://land.copernicus.eu/pan-european/corine-land-cover (accessed May 18, 2022).

[86]   W. Lin, Volunteered Geographic Information constructions in a contested terrain: A case of OpenStreetMap in China, Geoforum. 89 (2018) 73–82. https://doi.org/10.1016/j.geoforum.2018.01.005.

[87]   S. Landesamt Rheinland-Pfalz, Statistisches Jahrbuch Rheinland-Pfalz 2020, (2020).

[88]   OSM Analytics Tool - Rhineland-Palatinate, (n.d.). http://osm-analytics.org/#/show/polygon:qxhd%40chppHmni%40xtj%40ww%60%40tcD_rAv%60Pxa%

5Eji%5CejFryCu%7DnBwjOcwy%40neHx%60%40x%7CXahZllSx%60S~%7DYiqfAdq%5E_lmB%7CpNylm%40gx%7C%40m~Euk%5Ddy%5Bqpp%40%7BfVujKvjZ_wf%40fzW%7DiMfa_Ar~JteTewPc_%5EorNhfC%7D%7BJcbg%40mnL%60hZkb_%40sfCuePwp%5Cs_KztaAa%7DcAngj%40paf%40hv~Bh_b%40~rKzgSlpOe%7DDo%7DBfgVf%7DmAsrBbgDjkMdre%40dqN/buildings/recency (accessed May 18, 2022).

[89]   OSM   Analytics   Tool   -   Henan,   (n.d.).   http://osm-analytics.org/#/show/polygon:mvp%60TwxtqE_%60aFpuhJsj%7CCzvaAcobGe%7DPe_%5Dtj%7CAgjuGzpbBo_eCig%7CApoDobcBrjuB%7Dba%40bijAomaB%7DqkC%7Bj%7CAckFig~AqdjBpqaAwfmAyiy%40lioAetcB%7CkgCohL%7CqyB%7BufBadnFojyDlv%7DBn%7Co%40zrx%40ocgAxbaF%7Bz%5Blb%7DBlilDbbkFjme%40/buildings/recency (accessed May 18, 2022).

[90]   Natural disasters cost $280 billion in 2021: German insurance firm | News | DW | 10.01.2022, (n.d.).   https://www.dw.com/en/natural-disasters-cost-280-billion-in-2021-german-insurance-firm/a-60378575 (accessed May 9, 2022).

APPENDICES

Appendix A – List of Abbreviations

GIS –   Geographic Information System
VGI –   Volunteered Geographic Information
OSM –   OpenStreetMap
SAR –   Synthetic Aperture Radar
ESA –   European Space Agency
DEM –   Digital Elevation Model
NDVI –   Normalized Difference Vegetation Index
NDWI –   Normalized Difference Water Index
NIR –   Near Infrared
SWIR –   Short-Wave Infrared
LSWI –   Land Surface Water Index
MLSWI –   Modified Land Surface Water Index
PDF –   Probability Density Function
IW –   Interferometric Wide
GRD –   Ground Range Detected

Appendix B – Code Listings

This section contains the R code used to generate the tables displayed in the results sections 3.1 and 3.2.

```r
library(tidyverse) # Data Management
library(gt) # Tables
library(webshot) # for export function
webshot::install_phantomjs() # for export function

###### Land Cover Results ######
data <- read_csv2("lulc_long_ch.csv")
tablc <-
  data %>%
  gt() %>%
  tab_options(
    column_labels.border.top.color = "white",
    column_labels.border.top.width = px(3),
    column_labels.border.bottom.color = "black",
    table_body.hlines.color = "white",
    table.border.bottom.color = "white",
    table.border.bottom.width = px(3)
  ) %>%
  cols_label(
    type = "Land Cover Type",
    area = "Area",
  ) %>% # Changing the font
  opt_table_font(
    font = list(
      google_font(name = "Merriweather")
    )
  ) %>% # Marking text bold
  tab_style(
    style = list(
      cell_text(weight = "bold")
    ),
    locations = list(
      cells_body(
        columns = everything(),
        rows = 8
      ),
      cells_column_labels(
        columns = everything()
      )
    )
  ) %>% # Footnote
```

```r
    tab_footnote(
      footnote = "in square km",
      locations = cells_column_labels(
        columns = area)
      ) %>% # Misclassified cell font style
    tab_style(
      style = cell_text(style = "italic"),
      locations = cells_body(
        columns = everything(), rows = 9
      )) %>% # Misclassified cell font style
    tab_style(
      style = list(
        cell_text(color = "firebrick", style = "italic")
      ),
      locations = cells_body(
        columns = area, rows = 9
      )
    ) %>%
    cols_width(type ~ px(250)) %>%
    tab_style(
      style = cell_borders(
        sides = "top",
        color = "black",
        weight = px(2)
      ),
      locations = cells_body(
        columns = everything(),
        rows = 8
      )
    ) %>% # Increase font size
    tab_style(
      style = cell_text(
        size = px(14)
      ),
      locations = cells_body(
        columns = everything(),
        rows = 9
      ))
  tablc %>%
    gtsave("ch_lulc.png")


###### Infrastructure table ######
data_infra <- read_csv2("infra_de.csv")
tabinf <-
  data_infra %>%
  gt() %>%
```

```r
  cols_label(
    type = "Type",
    amount = "Amount",
    length = "Length"
  ) %>%
  tab_options(
    column_labels.border.top.color = "white",
    column_labels.border.top.width = px(3),
    column_labels.border.bottom.color = "black",
    table_body.hlines.color = "white",
    table.border.bottom.color = "white",
    table.border.bottom.width = px(3)
  ) %>%
  tab_footnote(
    footnote = "in km",
    locations = cells_column_labels(
      columns = length)
  ) %>%
  opt_table_font(
    font = list(
      google_font(name = "Merriweather")
    )
  ) %>%
  tab_style(
    style = list(
      cell_text(style = "italic")
    ),
    locations = cells_body(
      columns = length, rows = 3:4
    )
  ) %>%
  tab_style(
    style = list(
      cell_text(weight = "bold")
    ),
    locations = (
      cells_column_labels(
        columns = everything()
      )
    )
  ) %>%
  fmt_number(
      columns = length, rows = 1:2,
    decimals = 1
  )
tabinf %>%
  gtsave("de_infra.png")
```