

INAUGURAL – DISSERTATION

zur
Erlangung der Doktorwürde
der
Gesamtfakultät für
Mathematik, Ingenieur- und Naturwissenschaften
der
Universität Heidelberg

Vorgelegt von
Benjamin Herfort, M.Sc.
aus: Berlin, Deutschland

Tag der mündlichen Prüfung: 13. Juni 2023

Topic

Mapping what is not mapped

*Investigating questions of representation and data quality in
OpenStreetMap from the perspectives of humanitarian and
machine learning-assisted mapping*

Supervisors:

Prof. Alexander Zipf, Heidelberg University, Germany

Prof. João Porto de Albuquerque, University of Glasgow, United Kingdom

Abstract

Despite its omnipresence map data is still scarce and unavailable in many places and for many applications. New institutions, actors, ideas and partnerships are needed to tackle data scarcity, e.g. for monitoring progress towards the Sustainable Development Goals (SDGs). The potential of OpenStreetMap (OSM) for SDG monitoring and disaster risk reduction has been broadly recognized. Volunteer mapping communities, corporations, and humanitarian organisations are continuously addressing issues of data scarcity through OSM already.

In support of these efforts, this dissertation studies several approaches towards mapping what is not yet mapped in OSM. This is approached by analysing the spatio-temporal evolution of OSM in respect to data quality and issues of representation on a global scale since 2008. The first research question investigates the impact and scale of humanitarian mapping in OSM. The second research question addresses building data completeness and inequalities in OSM for 13,189 urban centres around the world. Finally, the third research question explores how machine learning-assisted and crowdsourcing mapping workflows could be combined to overcome the scarce availability of accurate and up-to-date human settlement data sets.

The results highlight that OSM data shows a much more spatially diverse pattern than previously considered, which is shaped by regional, socio-economic and demographic factors across several scales. For the first time this dissertation has made visible OSM's uneven spatial coverage and describes how it evolved over time. The findings show that there is a need to go beyond monitoring mapping activity and to improve the methods to monitor data quality.

This dissertation supports the general notion about the potential of machine learning systems to improve crowdsourced mapping workflows with respect to efficiency. However, similar to the biases introduced by the emergence of humanitarian mapping, machine learning-assisted mapping will put their own particular footprint on OSM. The research about the global footprint of humanitarian mapping and its effects on map data quality provides a blueprint for further investigation and monitoring of biases caused by machine learning-assisted mapping.

Researchers will be able to draw more robust conclusions once the biases in OSM's coverage are better known and can be accounted for. The results of this thesis in combination with local information make it possible to map out where mapping is needed to overcome data scarcity and issues of representation. However, real inclusion requires an active local OSM community with the capacity to generate, maintain and improve geographic data that reflects local perspectives in the long term. Researchers, humanitarian organisations and local OSM communities need to self-critically identify the structural barriers that exclude some social groups from participating in OSM mapping and work with them to develop capacity to overcome those barriers.

Kurzzusammenfassung

Trotz ihrer Allgegenwärtigkeit sind Kartendaten vielerorts noch immer knapp und für viele Anwendungen nicht verfügbar. Das Füllen dieser Datenlücken erfordert neue Ideen und Partnerschaften, ermöglicht jedoch unter anderem ein besseres Monitoring der Nachhaltigkeitsziele der Vereinten Nationen (SDGs). Das Potenzial von OpenStreetMap (OSM) im Bereich der nachhaltigen Entwicklung und im Katastrophenmanagement ist gemeinhin anerkannt. In OSM aktive Freiwillige, Unternehmen und humanitäre Organisationen tragen bereits tagtäglich zu einer besseren Verfügbarkeit von Geodaten bei.

Im Sinne dieser Bemühungen werden in der vorliegenden Dissertation verschiedene Methoden entwickelt, um die Datenlücken in OSM aufzuzeigen. So wird die räumlich-zeitliche Entwicklung von OSM seit 2008 auf globaler Ebene betrachtet. Der Fokus liegt dabei auf Fragen der Datenqualität und Repräsentation. Im Zuge der ersten Forschungsfrage wurde die Ausdehnung und der Einfluss von humanitären Kartierungsprojekten in OSM untersucht. Die zweite Forschungsfrage adressiert die Qualität von OSM-Daten für 13.189 Städte weltweit und befasst sich mit der Vollständigkeit von Gebäudedaten und ihrer räumlich ungleichen Verteilung. Die Dissertation schließt mit der dritten Forschungsfrage und einer Untersuchung, wie maschinelles Lernen und Crowdsourcing kombiniert werden können, um die Verfügbarkeit von Siedlungsdaten zu verbessern.

OSM-Daten weisen ein weitaus komplexeres räumliches Muster als bisher angenommen auf, das auf unterschiedlichen Skalenebenen durch regionale, sozioökonomische und demografische Faktoren geprägt ist. Die Ergebnisse der Dissertation machen die ungleiche räumliche Abdeckung und Vollständigkeit von OSM nicht nur sichtbar, sondern beschreiben auch, wie diese sich im Laufe der Zeit herausgebildet haben. Die Arbeit unterstreicht, dass Methoden zur Untersuchung der Datenqualität besser dazu geeignet sind räumliche Unterschiede aufzuzeigen als Ansätze, die nur Mapping-Aktivität messen.

Die Forschungsergebnisse bestätigen das Potenzial von Methoden des maschinellen Lernens zur Verbesserung der Effizienz von Crowdsourcing Mapping Workflows. Dennoch wird Datenerfassung mit der Unterstützung von Machine-Learning Systemen in OSM räumliche Ungleichheiten in der Datenverfügbarkeit nicht komplett auflösen. Der Forschungsansatz der Dissertation stellt eine gute Ausgangslage dar, die zukünftig zu erwartenden Veränderungen in OSM besser zu verstehen.

Ein besseres Verständnis über die ungleiche Verteilung von OSM-Daten erlaubt es Wissenschaftlern, diese räumlichen Unterschiede in ihren Analysen zu berücksichtigen. In Kombination mit lokalen Informationen stellen die Ergebnisse dieser Dissertation heraus, wo OSM-Daten fehlen und können so dazu beitragen, Probleme in Bezug auf Datenlücken und Repräsentation zu überwinden. Echte Teilhabe und die Einbindung lokaler Perspektiven ist jedoch nur dann möglich, wenn die lokalen OSM-Communities die Möglichkeiten haben, OSM Daten einer Region zu erstellen und langfristig zu pflegen. Wissenschaftler, humanitäre Organisationen und lokale OSM-Communities müssen dafür selbstkritisch hinterfragen, welche strukturellen Hürden die Teilhabe an OSM erschweren und Ideen entwickeln, wie diese Hindernisse überwunden werden können.

Contents

Abstract	i
Kurzzusammenfassung	iii
List of Figures	vii
List of Abbreviations	viii
I Synopsis	1
1 Introduction	3
1.1 A Short History of Mapping in OpenStreetMap	6
1.2 Technological Context: The Enabler for a Data Deluge	8
1.3 Political Context: The Call for a Data Revolution	9
2 Objective and Research Questions (RQs)	13
2.1 Research Question 1: Representation Biases and OpenStreetMap	15
2.2 Research Question 2: Data Gaps and Data Quality Assessment	17
2.3 Research Question 3: Machine Learning and AI-assisted Mapping	20
3 Scientific Contributions	23
3.1 Research Question 1	23
3.2 Research Question 2	28
3.3 Research Question 3	32
3.4 Limitations	36
4 Discussion	39
4.1 Representation Biases and Data Gaps	39
4.2 Machine Learning Systems and Data Quality	41
4.3 AI-assisted Mapping and Representation Biases	42
5 Conclusion	45
Acknowledgements	49
Bibliography	51

II	Publications	63
1	The Evolution of Humanitarian Mapping within the OpenStreetMap Community	65
2	Investigating completeness and inequalities in OpenStreetMap: spatio-temporal analysis of global urban building data	87
3	Mapping Human Settlements with Higher Accuracy and Less Volunteer Efforts by Combining Crowdsourcing and Deep Learning	119
III	Declarations	141

List of Figures

1.1	A short history of OpenStreetMap and the political and technological context. .	5
2.1	Research concept and major research questions and related research domains. .	14
3.1	Sketch of the evolution of humanitarian mapping in OSM.	24
3.2	Mismatch between the spatial distribution of mapping activity in OSM vis-à-vis the distribution of the global population.	26
3.3	Timeline of humanitarian mapping for all countries with medium and low Human Development Index.	27
3.4	Spatial distribution of OSM building completeness in 13,189 urban centers. . .	29
3.5	Spatial distribution of intra-urban OSM building completeness for selected urban centers.	31
3.6	Performance of crowdsourcing, deep learning and EO-based human settlement maps.	34
3.7	Combining crowdsourcing and deep learning for an AI-assisted mapping workflow.	35

List of Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
CNN	Convolutional Neural Networks
EO	Earth Observation
ESA	European Space Agency
GHS-POP	Global Human Settlement Population
GIS	Geographic Information System
GIScience	Geographic Information Science
GPS	Global Positioning System
GUF	Global Urban Footprint
HOT	Humanitarian OpenStreetMap Team
HRSL	High Resolution Settlement Layer
ML	Machine Learning
LULC	Land Use Land Cover
NASA	National Aeronautics and Space Administration
NGO	Non-Governmental Organisation
ODbL	Open Database License
OSHDB	OpenStreetMap History Database
OSM	OpenStreetMap
OS	Ordnance Survey
RQ	Research Question
SDG	Sustainable Development Goal
SFDRR	Sendai Framework for Disaster Risk Reduction
SHDI	Subnational Human Development Index
UAV	Unmanned Aerial Vehicle
VGI	Volunteered Geographic Information
VNL	VIIRS Nighttime Lights

Part I

Synopsis

Chapter 1

Introduction

Map makers can make a difference.

This simple rationale has motivated thousands of volunteers around the world, including myself and some of my friends and colleagues, to join the OpenStreetMap (OSM) community and to collaboratively create digital maps and geo data, which is accessible to everyone. Often, these efforts have been driven by the observation that maps are literally "missing" or existing map data was either too expensive or not disposable at all. In my case, the Typhoon Haiyan/Yolanda in the Philippines in 2013 triggered the first comprehensive engagement with the topic of digital humanitarian mapping in OSM, simply by delineating hundreds of buildings and roads from very high resolution satellite imagery. Typhoon Haiyan/Yolanda was one of the strongest ever recorded in the region and caused heavy damage across the islands of Leyte and Tacloban city. OSM data has been used in the field to provide first responders with printed maps of the most affected regions and for offline navigation with GPS systems. The mapping in OSM was coordinated by the the local OSM communities in Guagua and Tacloban, the Humanitarian OpenStreetMap Team (HOT), and first responders including the American Red Cross and the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA). Beyond that, creating digital data in OSM made it possible for people from around the world, like me and the group of students in Heidelberg, to feel part of the international humanitarian response community.

Almost ten years have passed since that first encounter and since then not only mapping or humanitarian aid, but our daily lives are becoming increasingly digitalized and are also relying more and more on automated systems. The abundance of data has changed the social, economic and political environments and impacts our actions and decisions almost on a daily basis. "Big data" are huge in volume and high in velocity, can be structured or unstructured, can be exhaustive in scope and fine-grained in resolution (Kitchin, 2013). Simultaneously, "big data" forms social processes, but has also been employed to create more accurate and detailed understandings of social processes itself (Graham and Shelton, 2013). For the domain of geography, this wealth of (georeferenced) data has already promoted data-driven approaches and methodologies, e.g. for geographic knowledge discovery or spatial modelling (Miller and Goodchild, 2015).

Despite this omnipresence of data, also described as "data deluge" (Anderson, 2008), data is still scarce and unavailable in many places and for many applications. Many examples of data scarcity are provided by UN OCHA in their report on the state of open humanitarian data, which offers a more targeted analysis outlining the need for data in regions with humanitarian response plans such as in Afghanistan or Yemen (UN OCHA Centre for Humanitarian Data, 2023). Climate impact data has been identified as a priority for the upcoming years, which will enable the prediction of vulnerability and

future hazard impacts (UN OCHA Centre for Humanitarian Data, 2023). This report also showcases the work of HOT on community mapping in South Sudan as one approach to fill existing data gaps. Also the UN Data Strategy suggests the necessity of a transformation towards a data-driven organisation, so that the UN can unlock their full potential to make "better decisions and deliver stronger support to people and planet - in the moments that matter most" (United Nations, 2020a). The strategy follows an use-case-centric approach and identifies which datasets are not available or under-exploited, e.g. in scope of the decade of action to deliver the Sustainable Development Goals (SDGs).

In retrospective, this dissertation could be described as the attempt to better understand how societal mega trends such as digitalisation and automatisisation have shaped collaborative map making in the era of big data and joint action for sustainable development, coordinated efforts to reduce disaster risks and basic approaches for coping with the climate crisis. The dissertation addresses the dichotomy of data deluge and data scarcity through analysing questions of biased representations in the OSM community and inequalities in the completeness of OSM data on a global scale. Through the lens of humanitarian mapping the dissertation will illustrate how the political context has framed the narrative of data scarcity and defined the very purpose of and need for the "missing" maps. At the same time, the potential effects of the data deluge produced by technological innovations, including machine learning and Artificial Intelligence (AI) assisted mapping, will be evaluated and discussed in the context of OSM and humanitarian mapping.

This dissertation investigates questions of representation and data quality in OpenStreetMap from the perspectives of humanitarian and machine learning-assisted mapping in order to map what is not mapped.

The broader scope of the dissertation is structured as follows. Chapter 1 of this synopsis functions as an embedding to better understand the evolution of OSM and the technological and political context. Chapter 2 provides a detailed description of state-of-art literature and derives the research structure and three major research questions of this thesis. The results and scientific contributions towards the identified research gaps are presented in Chapter 3. This includes also a characterization and an assessment of the limitations. Finally, Chapter 4 discusses the findings of this dissertation and Chapter 5 concludes by providing ideas for a future research agenda.

At first, it is worth to take a look at the momentum around collaborative mapping in OSM (see Chapter 1.1) and the essential technological developments (see Chapter 1.2) which enabled the rapid increase in data production in general and for OSM in particular in the first place. Without these innovations, contributing data to OSM, e.g. through digitizing buildings from satellite imagery or adding the opening hours for your local corner shop on your smartphone, wouldn't be as simple as it is nowadays. Furthermore, as indicated above already, the political context has shaped the very purpose why these datasets and maps are needed in the first place (see Chapter 1.3). Major political frameworks have defined data needs on a global scale, e.g. for the purpose of monitoring SDG progress. As a visual guide and summary, Figure 1.1 provides a sketch of the broader political and technological context relevant for the evolution of mapping in OSM.

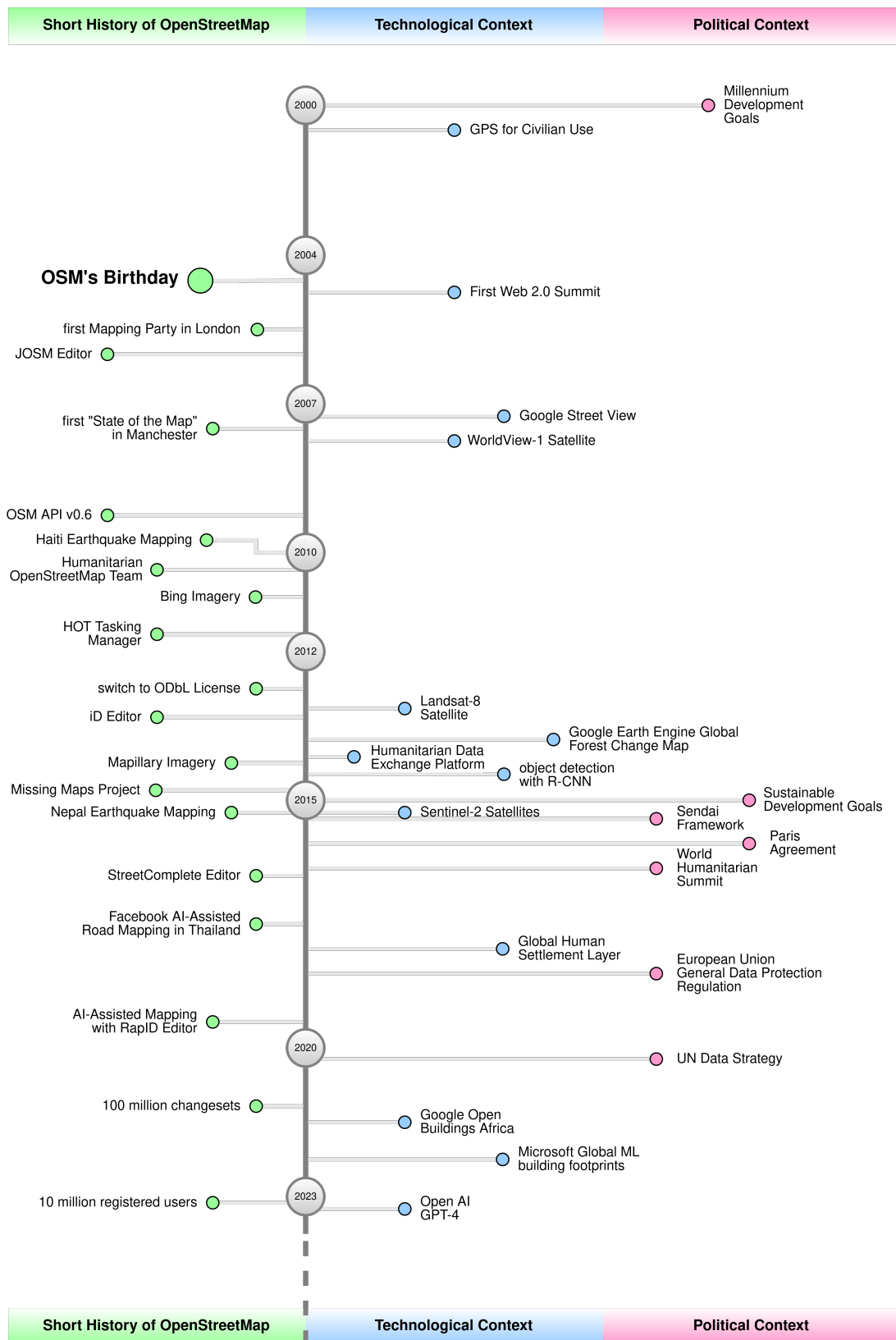


Figure 1.1: A short history of OpenStreetMap and the political and technological context.

1.1 A Short History of Mapping in OpenStreetMap

The OpenStreetMap project has evolved as a social product and forms a large community of people loosely connected through the joint work on a global geographic database. OSM is now used widely for applications such as web maps and navigation services and data from OSM has been used in domains such as urban planning (Milojevic-Dupont et al., 2020), SDG monitoring (Hoek et al., 2021), disaster management (Feldmeyer et al., 2021; Scholz et al., 2018), public health (Bhatia et al., 2018; Sturrock et al., 2018; Yeboah et al., 2021), as well as during the COVID-19 pandemic (Marco Minghini et al., 2020).

In 2004, OSM started as a project in the United Kingdom founded by Steve Coast. The community started to quickly grow from its origin to other places, but the project initially maintained a high concentration of mapping activity in European countries (Neis and Zipf, 2012). A detailed description of OSM's early history until 2008 is provided by Haklay and Weber (2008). Beyond OSM's technical stack (editing tools, geodatabase, web services, wiki) Haklay and Weber (2008) outline the importance of social collaboration as a key aspect in OSM's success. Community workshops, such as the first "mapping party" in London in 2005, and community conferences, such as the first "State of the Map" held in Manchester in the year 2007, have evolved into major events which bring together the global (and regional) OSM communities. Neis and Zipf (2012) have analyzed the rapid growth of OSM's user base between 2005 and 2011, which reached around 500,000 registered members in 2011. Due to the limited availability of satellite imagery, OSM data of the first days was usually edited based upon GPS tracks (Neis and Zielstra, 2014). Beyond the evolution of OSM's user base and the geographic regions covered, Mocnik et al. (2017) have described how the OSM folksonomy has evolved over time and how its thematic dimension has grown through the extension and modification of existing tags and the introduction of new tags.

In the aftermath of the 2010 Haiti Earthquake OSM was used for the first time on a larger scale to assist disaster response efforts first by tracing missing roads and buildings from satellite imagery and later on by adding local information captured with GPS devices on the ground (Zook et al., 2010). The Haiti Earthquake mapping opened up the way for future volunteered disaster mapping responses and led to the establishment of the Humanitarian OpenStreetMap Team (HOT). Initially since 2010 the terms "disaster mapping" and "crisis mapping" have been introduced to describe the "ability for volunteers to assist in disaster response situations via mapping and other spatial analysis" (Zook et al., 2010). In the upcoming years the broader term "humanitarian mapping" has evolved. It refers to collaborative mapping in OSM (and other platforms such as MapSwipe or Ushahidi) for both humanitarian relief responses and humanitarian purposes in general (Palen et al., 2015; Dittus et al., 2016; Porto de Albuquerque et al., 2016a). Whereas organisations such as UNOSAT had been established already, which provide satellite image analysis during humanitarian emergencies since 2003, only with the shift towards Volunteered Geographic Information (VGI) (Goodchild, 2007) and the engagement of the wider public "humanitarian mapping" has evolved as a concept describing a community within the OSM community and a field of research. This has been discussed in the wider domain of "digital humanitarianism", which more broadly encapsulates the collaboration of remote and on-the-ground individuals on humanitarian management through digital technologies (Burns, 2014).

Beyond organised humanitarian mapping, there are numerous examples of thriving OSM communities in low-income countries and the Global South. Already in 2009, the Map Kibera project started to create the first digital map of the informal settlement in the Southwest of Nairobi (Panek and Sobotova,

2015). In fact, the impact of humanitarian mapping responses such as the ones in the Philippines in 2013 (Palen et al., 2015) or in Nepal in 2015 (Poiani et al., 2016) would not have been possible without the established and well connected local OSM communities (OpenStreetMap Philippines community, 2020). The iconic photograph of OSM contributors at Kathmandu Living Lab outside their office building operational one day after the Nepal earthquake (Kathmandu Living Lab, 2015) symbolizes the relevance of local coordination and knowledge during disaster response activities. The YouthMappers network, which has been established in 2014, has engaged university students in more than 50 countries in Africa, Latin America and Asia. YouthMappers chapters are often linked with local OSM communities and address a wide range of challenges through mapping and geo data analysis, e.g. related to flooding, informal settlements, food security, power grid mapping, environmental resilience and sustainable transport planning (Solís and Zeballos, 2023).

In addition to contributions by individual volunteers (mappers), there is an intensifying trend that organized corporate mapping communities contribute to OSM in general (Anderson et al., 2019). Whereas companies have been involved in OSM right from its beginning, e.g. through sponsoring conferences and developing tools for OSM, in the meantime large tech enterprises such as Apple, Amazon, Facebook and Uber, but also smaller companies with a regional focus such as Digital Egypt or Bolt are hiring employees to directly contribute data to OSM (Veselovsky et al., 2022). These efforts have also led to the development of the RapID editor, which facilitates the editing of machine learning-based building and road detections into the OSM database (Meta, 2022). So far, machine learning-assisted mapping in OSM has mainly been pushed forward by corporates to fill data gaps in important markets, such as Facebook's attempt to map roads in Thailand starting in 2017 (Schröder-Bergen et al., 2021).

As of 2023 OSM is no longer an exclusive community of amateurs, but instead a community built of multiple smaller hobbyist, professional, amateur, and experienced mapping communities, in which professional stakeholders are sharply gaining influence on map data creation. OSM's core technology has been relatively stable over the past years, since OSM application programming interface (API) version 0.6 has been deployed in 2009 and the data license has been switched over to the Open Database License (ODbL) in 2012. Whereas OSM is often only seen from a technical standpoint as a map or geographic database, this perspective cannot fully capture and explain the specific characteristics of OSM, which need to be considered when reading the map or when utilizing the database for analysis.

For instance, as a consequence of socio-economic inequalities and the digital divide resulting in a much larger contributor basis in countries in the Global North, OSM data in general is strongly biased (Sui et al., 2013). Describing these biases from a data-oriented perspective can facilitate and promote the right usage of OSM data, however this perspective falls short to further explain the reason why these biases exist. Furthermore, neglecting the social processes and community aspects, which lead to the biases coverage of objects in OSM's database, could tempt analysts to underestimate the dynamics and potential for change. In the past almost two decades, the dynamic nature of OSM as a project has not only allowed to rapidly update its geographic data, but also had an impact on the type and granularity of the data collected. This flexibility of OSM facilitated the adaptation of different needs, e.g. for specialized routing, disaster response or SDG monitoring (Bertolotto et al., 2020). Biases in OSM are a fact, however it is important to acknowledge that with changing community practices, a shifting political context and technological innovations, they can be subject to rapid change in the future. These drivers for change will be the subject of the subsequent sections of this chapter.

1.2 Technological Context: The Enabler for a Data Deluge

Digitalisation and automatisisation are two trends which affect how maps are produced, but also have an impact on how they are consumed. Goodchild (2007) initially described the evolving world of VGI as the "widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information". At that time, it was unimaginable that in the year 2023 this revolutionary concept of map making would provide baseline data for monitoring global frameworks such as the SDGs. Goodchild (2007) also outlines the technological innovations (Web 2.0, GPS, georeferencing, broadband communication) which have enabled thousands of people to collaboratively work on projects such as OSM.

In contrast to traditional mapping, which is carried out by dedicated experts usually based in national mapping agencies following rigid protocols, OSM builds upon distributed and rather spontaneous crowdsourced data collection workflows. These approaches have been inspired by the success of projects such as Wikipedia (Heipke, 2010). Besides being able to track location through the Global Positioning System (GPS), the availability of georeferenced imagery such as aerial or satellite imagery, street-level imagery, photos from mobile phones or from unmanned aerial vehicles (UAVs) constitutes another technological foundation of many crowdsourcing applications (Heipke, 2010). In particular, the launch of Digital Globe's WorldView satellites (meanwhile acquired by Maxar) starting from 2007 and Airbus' Pléiades satellites from 2011 onwards drastically increased the timely availability of very high resolution (< 1 meter pixel resolution) imagery for almost any geographic region. From 2010 on, Microsoft Bing Maps allowed the OSM community to access its imagery for the purpose of mapping (Porto de Albuquerque et al., 2016a). Since 2014, the Mapillary project provides a mobile app and platform to collect street-level imagery under a creative commons license. This platform has developed into a source for hybrid remote-local mapping in OSM (Ma et al., 2019).

Albeit research on machine learning for (geographic) knowledge extraction has started decades before the emergence of crowdsourcing and VGI, the interaction of both has spawned yet another round of technological innovations in the domains of remote sensing and computer vision. Object based image analysis started to challenge the "pixel paradigm" in remote sensing research since the year 2000 (Blaschke, 2010). Computer vision researchers achieved breakthroughs in object detection using techniques based on deep convolutional networks and higher level feature representations of an image since 2012 (Zou et al., 2023). One example for this is the "regions with CNN features" approach (R-CNN) introduced by Girshick et al. (2014). Janowicz et al. (2020) described how the large-scale availability of high-quality data, enabled by the new culture of creating and sharing data, in conjunction with high-performance computing techniques has also advanced spatially explicit AI techniques for geographic knowledge discovery. These have been used for fast and accurate object detection, semantic classification and spatial interpolation. As such, machine learning techniques are already applied in the domain of social sensing, which utilizes (user-generated) data to model human dynamics, such as identifying human mobility patterns or human emotions (Janowicz et al., 2020).

Fuelled by freely available satellite data, e.g. Landsat-8 data by National Aeronautics and Space Administration (NASA) or Sentinel-2 data by European Space Agency (ESA), there has been an intensifying trend towards global scale derivative data products. Google Earth Engine and other cloud-based platforms have further reduced the barriers for planetary-scale geospatial analysis (Gorelick et al., 2017), e.g. to derive high-resolution global maps of forest cover change (Hansen et al., 2013). Similarly, the OpenStreetMap History Database (OSHDB) has combined easy access to the entire historical OSM data

with high querying performance (Raifer et al., 2019). As of 2023, there exist several global raster-based built-up area and population data products with varying resolutions, such as WorldPop, Global Human Settlement Layer (GHSL), World Settlement Footprint (WSF), High Resolution Settlement Layer (HRSL). In the domain of humanitarian aid and disaster management, the Humanitarian Data Exchange (HDX) platform facilitates the sharing of data across organisations and provides access to public datasets from governments, companies and open data communities including country extracts of OSM data (UN OCHA Centre for Humanitarian Data, 2023).

Only recently, several global scale vector datasets, which have been created from remote sensing data utilizing deep learning methods, have been made available under open data licences. For instance, a team at Google Research utilized a deep learning model to detect around 500 million building footprints across the entire continent of Africa using 50 centimeter satellite imagery and 1.75 million manually labelled building instances as training data (Sirko et al., 2021). A similar effort at Microsoft has produced a machine learning based building dataset on the global scale which consists of more than 1 billion building footprints including height estimations for some regions (Microsoft, 2023a). Also road network data has been made available by Microsoft (Microsoft, 2023b) and Meta (Facebook, 2023) on a global scale. In 2016, Porto de Albuquerque et al. (2016b) envisaged more comprehensive crowdsourcing projects that implement a cyclical process with different task types. This has already been partly built into human-in-the-loop mapping approaches utilizing the newly available global vector datasets, for instance using the Rapid editor (Meta, 2022). Sometimes these approaches are referred to as "AI-assisted mapping" or "machine learning-assisted mapping". Beyond providing data for mapping in OSM, these datasets have mainly been used to enrich and fill data gaps in the map services of the companies who created them in the first place, such as Microsoft Bing Maps or Google Maps.

Another dynamic research direction explores to what extent GIS analysis could be opened up to interaction with laymen, similar to how VGI has opened up geodata production (Janowicz et al., 2020). Such systems would allow users to simply ask generic questions describing the intended results of the analysis and would suggest the tools, data and workflows to answer them (Scheider et al., 2019). With the recent advances in neural network based large language models, such as OpenAI's ChatGPT, many new applications can be expected, which are not yet fully understood or foreseeable.

Nevertheless, advancements in information and communication technologies have also increased spatial inequalities and led to unevenness of representation and participation in a range of online data sets (Graham and Shelton, 2013). The growth of the data economy and policies of data-driven development might also lead to new risks of exclusion, e.g. for populations with limited access to the internet (World Bank, 2018). Technological innovations of the past two decades (and the upcoming changes in the future) cannot be reviewed without the political context condensed in global policy frameworks and legislation. Hence, this setting will be the primary focus of the next section.

1.3 Political Context: The Call for a Data Revolution

As highlighted in critical data studies and critical GIS research, data itself and the absence of data indicate new forms of power (Iliadis and Russo, 2016). This power of data has further led to the prioritization of positivistic and data-intensive approaches in the social and natural science (Iliadis and Russo, 2016). Likewise, also outside academia, global policy frameworks such as the SDGs or the UN Data Strategy attribute data with a specific power, i.e. data is framed as a strategic asset for a data-driven transformation (United Nations, 2020a).

The SDGs put a singular emphasis on the fact that "no one will be left behind" in the collaborative efforts of all countries to end poverty and hunger, protect the planet from degradation, promote economic and social prosperity and foster peace. Drawing lessons from what has been called the "immeasurability" of the Millennium Development Goals (MDGs) (Attaran, 2005) much attention has been given towards indicators and frameworks to monitor the SDGs. Reliable, timely and accessible geospatial data from various sources constitute essential contributions to this global indicator framework (Kraak et al., 2020). Furthermore, Kraak et al. (2020) illustrate how well-designed maps and diagrams do not only assist people and governments to understand progress towards the SDGs, but also how maps can reveal the spatial pattern of social, economic and environmental processes that "might otherwise go unnoticed".

The Sendai Framework for Disaster Risk Reduction (SFDRR) declares, that geospatial data is the fundament of the location-based disaster risk information utilized to create disaster risk maps for the general public, decision makers and communities at risk of exposure to disasters (United Nations Office for Disaster Risk Reduction, 2015). There is broad consensus that disaster risk management actively contributes towards sustainable development and disaster risk reduction is a necessary and cost-effective approach towards loss prevention. However, the goal of the SFDRR goes beyond reducing disaster losses, but also includes an "unprecedented emphasis on the role of science in understanding and delivering risk reduction" (Aitsi-Selmi et al., 2016).

The Paris Agreement sets out the goal to keep global temperature rise well below 2°, ideally below 1.5°, to pre-industrial levels. Besides monitoring how the climate changes, mitigation and adaptation efforts are central to the adherence of the global goal and each country's nationally determined contributions to reduce green house gas emissions. In respect to the impact of climate change, Reckien et al. (2017) show how physical exposure, high population densities and the lack of risk-reducing infrastructure and services are interlinked and lead to equity concerns, i.e. climate change usually impacts are usually more severe for residents with low social status and low income.

The temporal coincidence of these three global agreements provides a rare opportunity for coherent policies across sectors and coordinated efforts between a multitude of stakeholders (Aitsi-Selmi et al., 2016). In addition, there is considerable and intentional overlap in the goals of the Sendai Framework, Agenda 2030 and Paris Agreement, e.g. as recognizable in SDG 1.5: "By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters".

The SDGs and the SFDRR both highlight the relevance and increasing need of up-to date spatially disaggregated information. Additionally, it is evident that also effective humanitarian action and decision making rely on accurate, up-to-date spatial information (Braunschweig et al., 2020). The Sustainable Development Solutions Network have emphasized the role of the often-neglected geo-spatial and geo-referenced data to disaggregate information for almost every proposed SDG indicator (Sustainable Development Solutions Network, 2015). Hence, tackling data scarcity requires moving beyond insufficient traditional data sources to utilizing new, non-traditional sources for measuring the SDGs (Fritz et al., 2019; Sustainable Development Solutions Network, 2015). These calls for a "Data Revolution for Sustainable Development" identify the need for new institutions, actors, ideas and partnerships to successfully monitor progress towards the SDGs (United Nations, 2014).

Besides trying to tackle data scarcity for monitoring, the current global political frameworks put a particular emphasis on collaboration and participatory approach to achieve their goals. Acknowledging that data are important in general, Porto de Albuquerque et al. (2021) outline that availability of data alone is not sufficient to induce the change towards sustainable development. Sachs et al. (2019) describe six

potential pathways of transformation to achieve the SDGs and make clear that any transformation needs a broad societal buy-in, which cannot be achieved using top-down policies, but require transparency and participatory approaches. Indeed, there is supporting evidence that participation stands out as a governance strategy significantly positively linked to goal achievement related to SDG 1, 3, 10 and 17 (Glass and Newig, 2019). In the domain of disaster risk reduction, citizen science and other participatory approaches contribute to the forecasting and warning of hazards and can promote shared understanding and communication (Hicks et al., 2019).

In the light of data scarcity and lack of participation, the potential of OSM to harness the power of new information and communication technologies in support of disaster risk reduction has been recognized by several authors (Aitsi-Selmi et al., 2016; Scholz et al., 2018; Neis et al., 2010). The usefulness of base map data for sustainable development, disaster risk reduction and climate action in combination with OSM's undetermined and flexible thematic range has created a momentum, which promoted the use of OSM way beyond its initial core community of hobbyists and companies in the Global North. Through contributing to and using the open geographic database of OSM volunteer mapping communities, corporations, governmental and humanitarian organisations did already address issues of data scarcity and lack of participation at the same time in the regions that mattered most to their use cases.

Beyond this transformative power of (humanitarian) mapping in OSM, it has been shown that people directly affected by disasters, have still been marginalized in disaster response decision making. They are often excluded from information processing and interpretation, which usually happens after (map) data is collected (Mulder et al., 2016). The postulated "data revolution" intends to overcome existing inequalities by involving stakeholders beyond the traditional national mapping agencies and statistical offices (United Nations, 2014). The global policy frameworks of the past years have embraced data and this has contributed to a political context in which OSM could thrive. Nevertheless, data alone should not determine policy frameworks, as the process of data collection, presentation and interpretation can never be value-neutral and inherently leads to unevenness of representation (Graham and Shelton, 2013).

Building upon the evolution of the OSM community and by considering the technological achievements of the past decades, the subsequent chapter of this dissertation outlines the research questions addressing questions of representation and inequalities in the completeness of OSM data.

Chapter 2

Objective and Research Questions (RQs)

The objective of this research is to get a grasp of the past dynamics of mapping in OSM and the outcomes it produced. We approach this by studying the spatio-temporal evolution of OSM in respect to data quality and issues of representation. The main focus is on the use of OSM data for humanitarian aid and SDG monitoring. The investigation of the potential of machine learning approaches explores how crowdsourced mapping might evolve in the future. Consequently, the research in scope of this dissertation is tackled from three different perspectives (see Figure 2.1):

- Representation and OpenStreetMap:
RQ1: What is the impact and scale of humanitarian mapping in OSM?
- Data Gaps and Geospatial Data Quality:
RQ2: How unequal is the completeness of urban OSM building data?
- Machine Learning and AI-assisted Mapping:
RQ3: How can machine learning assist in crowdsourcing mapping workflows?

The three topics of this dissertation constitute major research domains, which are often investigated independently each within their own community and utilizing distinct methodological approaches. In fact, the main contribution of this dissertation is the analysis of the interaction between these research domains and to make visible the driving forces that make an impact on humanitarian aid and SDG monitoring. There is no single starting point for this research and indeed the thematic sequence chosen for this synopsis does not resemble the chronology of the individual publications.

First, we raise the question of who is currently represented through mapping in OSM and how this has changed over the past decade in the light of humanitarian mapping efforts (see Chapter 2.1). In Chapter 1.1, we have already described how the emerging of the OSM community has changed the way how geo data is created and used across domains. Nevertheless, representation biases in the OSM community ultimately lead to particular gaps in OSM's database. By analysing which regions and socio-economic groups are under-represented in regard to OSM mapping activity, we provide a very first overview on the potential gaps prevalent in maps based on OSM data. We will approach this question for the mapping of roads and buildings on a global scale with a particular focus on humanitarian mapping efforts. The first research question carries on VGI research and furthermore relates to discussions in the domains of Critical GIS (Goodchild, 2015) and Critical Data Studies (Iliadis and Russo, 2016).

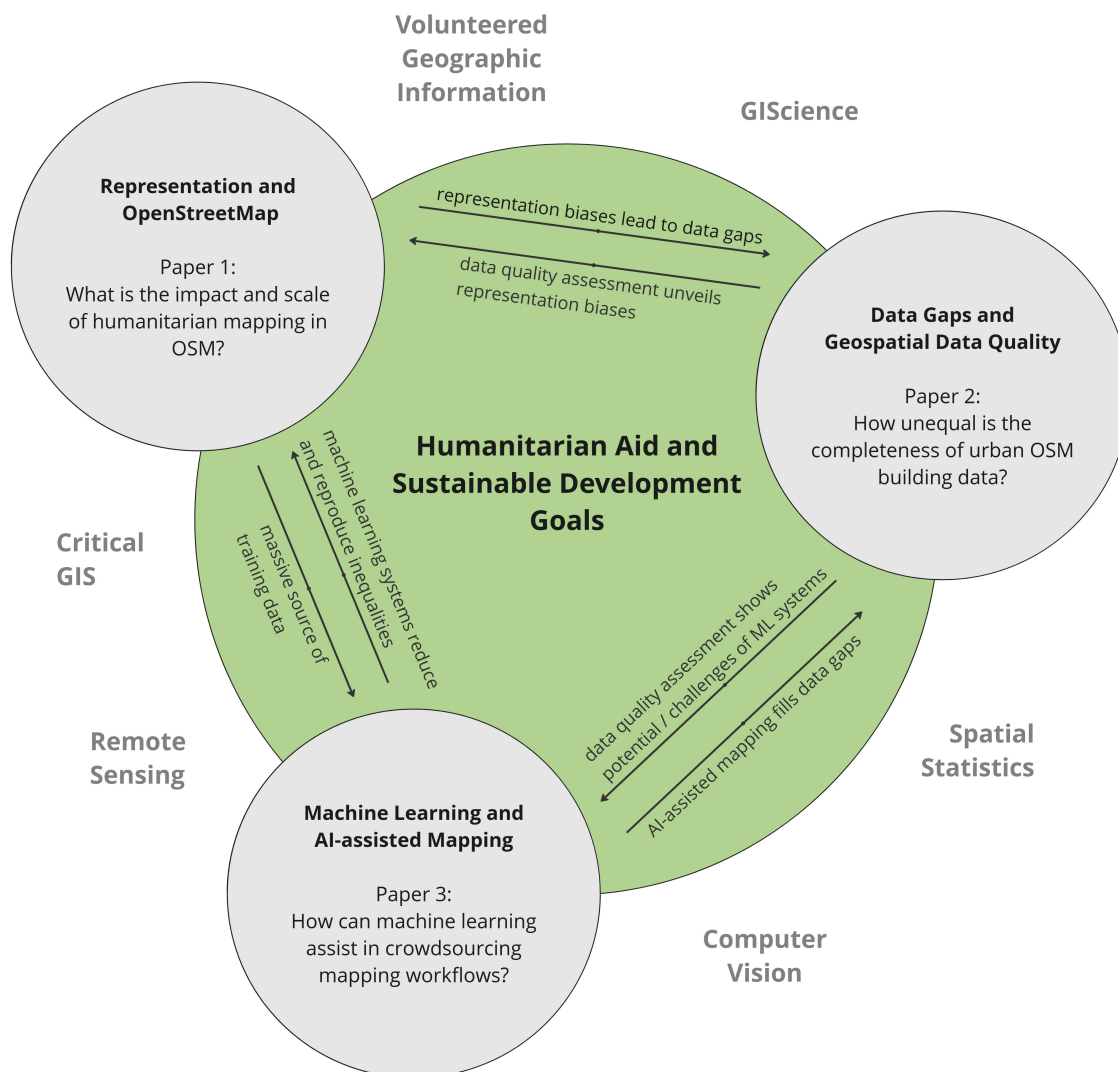


Figure 2.1: Research concept and major research questions and related research domains.

Second, we will investigate data quality in OSM and treat this as a dedicated form of representation bias (see Chapter 2.2). Here we go beyond analysing mapping activity, but investigate the completeness of OSM data. This assessment helps to unveil and better understand the outcomes of representation biases and their consequences for analyses relying on OSM data. At the same time, data quality assessment can show the potentials and challenges of approaches which aim to overcome existing gaps and biases in OSM, such as mapping workflows utilizing machine learning systems. This analysis is conducted on a global scale and studies the example of spatial-temporal dynamics in inequality and completeness of building data in urban areas. GIScience research has a long track record in providing the methods to measure spatial data quality. In addition, we will borrow methods from the domains of Spatial Statistics and Spatial Data Science to investigate inequalities in the geographical distribution.

Finally, the enabling and inhibiting aspects of machine learning and artificial intelligence assisted mapping for crowdsourcing community projects are investigated (see Chapter 2.3). The focus for this analysis lies on the potential to combine automated and crowdsourced means of creating human settlement maps. The research is performed to verify if machine learning-assisted mapping could efficiently fill existing data gaps. This third research question touches the fields of Remote Sensing and Computer Vision.

So, let's get started and dig deeper into each research question and the methodologies employed one after the other.

2.1 Research Question 1: Representation Biases and OpenStreetMap

The first research question tackles the overarching topic of who is actually represented through digital maps. Representation is an inherent issue of any map, which is deeply connected to the history of cartography as a discipline itself. With this in mind, traditional or digital maps should be read rather as a cultural text than as a mirror of nature or reality (Harley, 1989). With the spread of GIS in the 1990s the visual power of maps has been reinforced through their perception as being authoritative and scientific because they are legitimated through digital quantitative analysis and data (Pavlovskaya and Martin, 2007). Question of representation have gained renewed relevance in the era of "big data", which as a side effect could lead to the marginalisation of "small data" studies (Kitchin, 2013) or knowledge that not as easily encapsulated through sensors or data frameworks (Graham and Shelton, 2013). Policy frameworks have articulated their goals touching questions of representations, such as the general pillar of the SDGs to "leave no one behind" or the more direct addressing in SDG 10 "Reduce Inequalities" (United Nations, 2020b).

Socio-economic inequalities and the digital divide are main drivers of biases in OSM and have resulted in a much larger contributor basis in countries in the global North (Sui et al., 2013). The existing inequalities in participation within OSM question the claim that crowdsourcing and VGI can actually act as a tool to democratise geospatial data production (Gardner et al., 2020). Collaborative mapping in OSM has produced and reproduced existing pattern of exclusion. However, the expansion of geo data production to regions outside Europe has also uncovered that there are colonizing and decolonizing processes happening at the same time (Schröder-Bergen et al., 2021). Many authors have discussed issues of representation from a perspective of who creates content in OSM. For the case of Jerusalem, Bittner (2017) shows that OSM reproduced "offline" hegemonies and power imbalances as data tended to be richer in Jewish/Israeli areas in comparison to Arab/Palestinian areas. The study of Gardner et al.

(2020) analyses OSM contributions by gender and identifies differences regarding editing, tagging and overall activity in OSM for men and women. As of 2020 OSM remained significantly dominated by men (Gardner et al., 2020).

Most of these studies have focused on specific biases and usually investigated these for a restricted number of study sites. Whereas it seems to be possible to generalize some of their findings into general trends, there remains a research gap to examine the specific relationship characteristics, such as shape or intensity, on a global scale. For instance, Barrington-Leigh and Millard-Ball (2017) observed that there is no linear correlation between population density and OSM road network completeness but found an U-shaped relationship instead. This stands in contrast to what has been previously reported for urban versus rural OSM buildings completeness in Germany (Hecht et al., 2013). Hence, general trends can only provide initial guidelines when working with OSM data. There is a need for more OSM research which looks beyond dichotomous comparisons and questions simplistic linear relationships.

The Global North/South bias in OSM constitutes a primary concern for humanitarian organisations, which usually work in regions outside OSM's traditional core regions in high-income countries. However, in these data scarce environments there is often no single other source of map data available. In these settings, Zook et al. (2010) pointed out that OSM and other forms of crowdsourcing information just need to be "good enough to assist recovery workers using the maps". Regardless the existing biases in OSM, humanitarian organisations started to employ mapping through OSM in their information management workflows, because recording the presence of the millions of people "invisible" on maps is often the first step towards improved public services and humanitarian response (Bryant, 2021). At the same time it also increases accountability for those who provide aid (Bryant, 2021). Beyond their good intentions to improve OSM data, it is important that humanitarian organisations and other stakeholders involved in humanitarian mapping develop a more nuanced perspective on biases in OSM and questions of representation in general, e.g. from a post-colonial perspective (Schröder-Bergen et al., 2021). The call by the OSM Philippines community to correct and rethink narratives manifested in a 2020 documentary series produced by Amazon Web Services about the work of HOT in response to Typhoon Haiyan/Yolanda in 2013 strongly emphasizes the need to overcome persisting colonial framing of maps and mapping in OSM (OpenStreetMap Philippines community, 2020).

Previous work that analysed humanitarian mapping in OSM has mostly focused on single events or relatively short time periods. The 2010 Haiti Earthquake was one of the first attempts for which mapping in OSM was "good enough" to assist disaster response efforts and helped to make visible who was affected by the disaster event (Zook et al., 2010). There are detailed analyses of the mapping activity in OSM after the 2013 Typhoon Yolanda (Palen et al., 2015; Dittus et al., 2016), the 2014 West Africa Ebola response (Dittus et al., 2016), and the 2015 Nepal earthquake (Poiani et al., 2016). It has been shown that humanitarian remote mapping has undergone a series of changes in a socio-technological evolution between 2010 and 2013 (Palen et al., 2015). The Missing Maps project founded in 2014 has shifted the focus of large-scale humanitarian mapping from disaster response to preparedness and preventative efforts (Dittus et al., 2016; Scholz et al., 2018). The 2015 Nepal earthquake has been identified as a turning point in the adoption of OSM by a wide range of organizations, leading to open-data practices becoming more common in humanitarian work and to the constitution of a community of practice (Soden and Palen, 2016). Acknowledging the work by Palen et al. (2015), there is ongoing need for research which connects these individual events and investigates the longer-term dynamics in OSM mapping facilitated through evolving technology and policy frameworks.

A key underlying motivation behind the rich and multifaceted efforts of humanitarian mapping in

OSM has been to widen the availability of its geographic database by creating the "missing maps" needed so "everyone is counted, map data is accessible and used, everyone can engage and contribute to the map" (Humanitarian OpenStreetMap Team, 2019). However, the question arises to what extent humanitarian mapping efforts have been able to widen the footprint of OpenStreetMap to overcome the data gaps caused by socio-economic inequalities ten years after what could be considered the first large-scale humanitarian mapping event in the aftermath of the 2010 Haiti Earthquake. So far research did not address how much humanitarian mapping actually contributed to making the world-wide collaborative map more inclusive.

As this scattered picture drawn by existing literature masks the overall scale and footprint of humanitarian mapping, a methodological approach for a comprehensive assessment covering the global evolution is needed. In our work we tackle the gaps outlined above by an analysis which encompasses all humanitarian mapping projects organized through the HOT Tasking Manager since 2012 (start of the available data), enabling us — for the first time — to cast a longitudinal perspective on the intersecting effects of mapping efforts, socio-economic, and demographic characteristics. In particular the following two research questions are investigated:

- RQ 1.1: What is the effect of humanitarian mapping in the past decade on the global footprint of OSM in terms of socio-economic and demographic profile of the areas mapped?
- RQ 1.2: How did OSM humanitarian mapping activity evolve over time on a global scale and in individual countries?

Publication:

- Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., & Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific reports*, 11(1), 1-15. <https://doi.org/10.1038/s41598-021-82404-z>

2.2 Research Question 2: Data Gaps and Data Quality Assessment

The issue of who is represented through maps has often been linked to discussions about the digital divide, i.e. investigations in the geographical distribution of digital data or the uneven practices of VGI across the world (Sui et al., 2013). In many cases uneven participation in projects such as OSM has been modelled through regional differences in mapping activity or unequal data "density". For instance, Neis and Zipf (2012) examined the distribution of absolute and population normalized numbers of OSM members per continent and country. Also Research Question 1 of this thesis (Chapter 2.1) addresses this topic by "simply" comparing the overall number of buildings and roads in OSM with the population counts on a subnational level. These approaches are valid and useful as they unveil the general biases in OSM data production. However, at the same time they are limited in their scope as they fall short to provide a meaningful baseline for comparison, e.g. what is the expected baseline value of OSM users per 1 million inhabitants or the expected overall number of buildings in various regions of the world.

To overcome this limitation, the second research question attempts to transfer the question of representation into a more technical, GIScience-oriented question of data quality. This leads towards the identification of data gaps, i.e. an analysis of the completeness of OSM data. In contrast to describing the

data density, completeness analysis can reveal the desired baseline. Initial efforts have been conducted by Barrington-Leigh and Millard-Ball (2017), who linked internet access and governance indicator scores with the OSM road network completeness on a country level. It should be clear that this step involves a drastic simplification of what representation describes. However, it also offers the possibility to study it with the established geo data quality assessment methods.

Even though humanitarian organisations started to rely on OSM in their mapping workflows, they did so not because OSM already provided a perfect and complete dataset, but mostly because it put them into position to improve map data in a collaborative and distributed manner. From that perspective, data gaps in OSM were always present and are still pervasive today. Also in respect to SDG monitoring there still exist vast data gaps, for example due to the fact that as of 2019 most countries did not have the workflows in place to regularly collect data for more than 50% of the indicators (Fraisl et al., 2020). In addition, national statistics offices, who usually maintain the data, suffer from an estimated gap of \$1 billion USD globally in funding, and consequently, baseline geospatial data that should be provided by these agencies are often not accessible, not up-to-date or not available in standard formats (Braunschweig et al., 2020; Altay and Labonte, 2014). However, it has been shown that citizen science and open data communities such as OpenStreetMap (OSM) are already contributing to filling existing data gaps (Fritz et al., 2019; Scholz et al., 2018). In general, there is a need for research which can help to better quantify these contributions and further outlines for which regions data gaps could not be filled already.

Particular attention needs to be paid to gaps in coverage and quality when OSM data is utilized in global studies or to derive global data products, e.g. on critical infrastructure (Nirandjan et al., 2022), or when using "big data" in comparing urban morphology across the globe (Boeing, 2021). These global studies and datasets bear the highest risk of reproducing the inequalities of OSM and might lead to the propagation of representation issues simply because OSM systematically misses information for some regions. Consequently, describing spatial data quality is imperative to provide humanitarian organisations and other stakeholders with the necessary information to evaluate fitness for use of a dataset for their particular application (Oort, 2006).

A large community of researchers has analysed the quality of OSM data often with a focus on the road network (Barron et al., 2014; Senaratne et al., 2016; Barrington-Leigh and Millard-Ball, 2017). Neis et al. (2011) compared the OSM street network with TomTom's data and found that for Germany already in 2011 OSM exceeded the information provided by the proprietary dataset by 27%. To assess completeness and positional accuracy of OSM's road network, Haklay (2010) used data from Ordnance Survey (OS) as a reference dataset. Their results show that for England and London in 2010 OSM data was fairly accurate and captured already 29% of the area of England. Hecht et al. (2013) compare object-based and unit-based methods to assess completeness of OSM buildings as of 2012 for study regions in Germany. They compare OSM data to an authoritative dataset and found a relatively low degree of completeness for rural areas in comparison to urban areas. Also Brovelli and Zamboni (2018) utilize a reference dataset to analyse OSM building completeness and spatial accuracy using a map matching method. Studies based on authoritative or commercial reference data are able to provide detailed insights on OSM data completeness for selected cities (Hecht et al., 2013; Brovelli and Zamboni, 2018) or countries (Neis et al., 2011). However, they cannot be transferred to other regions for which reference datasets are either missing or unavailable in a format compatible with analysis. Also many of these studies investigated data quality at a single point in time. Given the changing technological and political context, there is further need to consider the temporal evolution of the OSM community and related temporal aspects of spatial data quality, such as described by Mocnik et al. (2017) for the evolution of the OSM folksonomy.

Intrinsic data quality assessment utilizing the edit history of OSM has been proposed as an alternative way to estimate completeness or currentness of OSM data (Barron et al., 2014). The completeness of OSM's road network, as of 2016, has been estimated at about 80% globally using an intrinsic mapping saturation based approach which doesn't require a reference dataset (Barrington-Leigh and Millard-Ball, 2017).

In data-sparse humanitarian mapping scenarios however, Li et al. (2020) have outlined that often "intrinsic quality analysis is not good enough, while extrinsic quality analysis is not possible". To overcome dependencies on sparsely available reference datasets and limits of intrinsic data quality approaches the use of globally available proxies such as remote sensing data, e.g. Nighttime Light, built-up-area, Sentinel 2 derived spectral indices (Goldblatt et al., 2020), or population data (Zhang et al., 2022) have been suggested to assess and predict OSM building completeness. Whereas approaches relying on proxy data show promising results for smaller scale case studies already, there is a gap in research on a global scale.

In fact, in recent years there has been a trend towards more global scale OSM data quality analysis using extrinsic and intrinsic data quality assessment (Barrington-Leigh and Millard-Ball, 2017; Zhou et al., 2022a,b). For instance, Zhou et al. (2022a) assessed the quality of land use land cover information from OSM for 168 countries by comparing to a remote sensing based global land cover dataset. Besides showing an estimate of completeness or accuracy for selected OSM features, these studies address an existing research gap as they are able to reveal the spatial biases in OSM on a global level. So far, these global studies have rarely utilized explicitly spatial methods to describe the geographic distribution of biases in data quality.

In general, there is a need for better spatial data quality assessment methods, as these would promote the adoption and (right) usage of new sources of data such as OSM and derived (global) data products. When unaccounted for, spatial bias can lead analysts and researchers to draw general conclusions which are only valid for well-represented (well-mapped) areas (Meyer and Pebesma, 2022).

In our research we tackle the topic of data quality in OSM for the specific case of completeness of building data in cities because this constitutes a common data quality requirement for many research and policy applications. For instance, SDG Indicator 11.3.1 ("Ratio of land consumption rate to population growth rate") would directly benefit from building footprint data. Our study addresses the derived research gaps by performing a global analysis considering the temporal evolution of mapping in OSM since 2008 utilizing proxy data through a machine learning approach. In particular, our work investigates building data completeness and inequality in OSM for 13,189 urban centres around the world, which are home to an estimated population of 3.5 billion people (about 50% of the global population). The focus lies on the following questions:

- RQ 2.1: What is the completeness of OpenStreetMap building data in the context of global urban analysis applications?
- RQ 2.2: How unequal is urban OpenStreetMap building data distributed within the space of a city, across continents and on the global level?

Publication:

- Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., and Zipf, A. (2023). Investigating completeness and inequalities in OpenStreetMap: spatio-temporal analysis of global urban building data. Under Review. Preprint available at: <https://doi.org/10.21203/rs.3.rs-1913150/v1>

2.3 Research Question 3: Machine Learning and AI-assisted Mapping

The technological evolution of the past two decades has dramatically shaped the way how geo data is collected and how maps are produced. First, VGI and OSM have challenged the hegemony of authoritative data production by national mapping agencies and opened it up to the crowd. In the past ten years the improved performance of machine learning systems have enabled automatisations to become much more central part in map making workflows.

Many authors have investigated the potential of deep learning technology in various satellite image processing tasks including human settlement mapping (Zhu et al., 2017). In respect to SDG 1, Jean et al. (2016) utilize a convolutional neural network to distinguish urban areas, non-urban areas, roads and water in optical satellite imagery to predict local level economic outcomes in five African countries (Nigeria, Tanzania, Uganda, Malawi, and Rwanda). Regarding land cover mapping, several authors propose workflows based on deep neural networks with a focus on urban areas (Kampffmeyer et al., 2016; Tracewski et al., 2017). Furthermore, deep learning based building footprint extraction has been a central research topic in the last years (Vakalopoulou et al., 2015; Yuan et al., 2018).

Often these machine learning systems rely on OSM data for deriving training samples. For Mozambique, Li et al. (2020) used the bounding boxes of OSM building footprints to fine tune a Single Shot Multibox Detection network to identify missing built up areas in OSM. This approach has been extended using a few shot transfer learning method to apply a pre-trained building detection network in a region distant from the initial training area by providing few additional training samples from OSM (Li et al., 2022). Transferring machine learning methods systems for object detection to various geographical regions remains a challenge. A research gap exists to comprehensively compare the results of automated mapping approaches on a broader scale. Data from OSM might be used as training samples for automated information extraction from satellite imagery, but citizens are rarely engaged in this process after the initial training phase. In general crowdsourcing datasets are extensively used to train machine learning systems, but crowdsourcing methods are still under-exploited towards assessing the quality of deep learning detections and similar automated mapping approaches (See et al., 2022).

Machine learning-assisted workflows are already tested and sometimes utilized in the map making domain and often they are motivated by the goal to overcome the existing data gaps. Yi et al. (2021) developed an machine learning-assisted workflows to derive accurate human-wildlife proximity maps. They used deep learning image classification and object detection approaches and could significantly reduce costs to about 25% by increasing the speed of spotting wildlife, human activities and livestock in aerial imagery. At Ordnance Survey (OS), UK's national mapping agency, automation has been embraced for change detection in aerial imagery and extracting attribute information such as building heights (Murray et al., 2019). Jiao (2018) presented a machine learning technique used by the industry to minimize the amount of manual work for creating high definition maps needed for autonomous driving technologies. This consists of several automated and "human-in-the-loop" mapping tasks such as traffic lights and traffic signs extraction, lane marking extraction, road curb extraction or light pole extraction. The United Nations Global Pulse Institute uses a satellite imagery analysis tool which employs neural networks and a human-in-the-loop approach for tasks such as counting structures in refugee settlements or rapid mapping of flooded areas (Logar et al., 2020). In these settings, relying on machine learning systems is considered to be more efficient in comparison to the prevalent fully manual operational pipelines, while preserving the required level of quality control (Quinn et al., 2018).

Despite the advantages of machine learning based object detection and other remote sensing techniques projects such as OSM do not systematically integrate them into their workflows, e.g. for quality assurance or data production. Nevertheless, initial research shows that bringing together these two worlds might help to overcome shortcomings of both approaches and ease data quality issues of other existing data sets (Gueguen et al., 2017). Vargas-Muñoz et al. (2019) investigated the quality of OSM data in study sites in Zimbabwe and Tanzania using a deep learning approach. This approach allows to detect missing building footprints and misalignment in the OSM data.

Only very recently, new actors have joined and existing mapping tools (such as the HOT Tasking Manager) are improving with advanced technological approaches such as deep learning/artificial intelligence, which are now finding their way into humanitarian mapping workflows as well (Li et al., 2020). Huck et al. (2020) have proposed a hybrid human-machine mapping approach which aims to overcome data gaps in "remote, rural and sparsely populated parts of the world that are unlikely to attract sufficient attention from volunteer mappers". Meta's RapiD editor utilizes buildings and roads detected through deep learning approaches and provides an interface to add them to the OSM database (Meta, 2022). The Humanitarian OpenStreetMap Team is experimenting with an open AI-assisted mapping service (fAIr) to overcome model bias, lack of feedback and lack of openness of existing building detection models (Humanitarian OpenStreetMap Team, 2023). There is a consensus that machine learning-based approaches should complement and not replace existing crowdsourcing approaches for mapping (Watkinson et al., 2022). However, more research is needed to identify the actual sweet spots of both approaches and to design workflows on how to best integrate automated ways of geo data production and existing crowdsourcing approaches.

In our work we address this research gap for the example of built-up area mapping in MapSwipe. We present a workflow to overcome the scarce availability of accurate and up-to-date human settlement data sets combining two techniques: (1) object detection and classification using deep learning algorithms (DeepVGI) and (2) crowdsourced mapping of human settlements by volunteers (MapSwipe). To combine both methods we propose a task allocation strategy that chooses classification labels either from DeepVGI or MapSwipe and conducted case studies in three different sites located in Guatemala, Laos, and Malawi to evaluate the proposed workflow. Moreover, we investigate whether our proposed methodology helps to produce better maps faster with respect to the following research questions:

- RQ 3.1: How good are crowdsourcing (MapSwipe) and deep learning (DeepVGI) with respect to generating human settlement maps in comparison to existing EO-based approaches?
- RQ 3.2: Which spatial and non-spatial characteristics of misclassifications are accompanied by applying the DeepVGI approach?
- RQ 3.3: What is the added value of the proposed task allocation strategy with respect to performance and effort?

Publication:

- Herfort, B., Li, H., Fendrich, S., Lautenbach, S., & Zipf, A. (2019). Mapping human settlements with higher accuracy and less volunteer efforts by combining crowdsourcing and deep learning. *Remote Sensing*, 11(15), 1799. <https://doi.org/10.3390/rs11151799>

Chapter 3

Scientific Contributions

This chapter discusses the scientific contributions of this thesis in respect to the research questions outlined in Chapter 2. In the following the results for each research question and their relevance are addressed as subsections. Limitations of the research are integrated in a synoptic way and reported in a separate subsection (see Chapter 3.4). The individual full papers are attached as separate chapters in the second part of the thesis (Part II)

3.1 Research Question 1

In our paper (see Chapter II.1) we investigated the spatial extent of humanitarian mapping activity in OSM predominantly organized through the web tool "HOT Tasking Manager". Previous work that analysed humanitarian mapping in OSM has mostly focused on single events or relatively short time periods. To overcome this limitation, in our paper we have put attention to the overall temporal evolution of humanitarian mapping activity in OSM which started in 2010.

The study analysed the effect of humanitarian mapping on the global footprint of OSM in terms of the socio-economic and demographic profile of the areas mapped (**RQ1.1**). Starting from this broad picture our paper investigated the evolution of OSM humanitarian mapping over time on a global scale and in individual countries (**RQ1.2**).

Figure 3.1 provides context and motivation for this temporal perspective by showing a chronology of key events in humanitarian mapping since 2010, the year HOT was founded in the aftermath of the devastating Earthquake in Haiti. It shows the immediate impact that post-disaster community "activations" have had on the contributions to OSM. Besides these punctual events, there are several additional factors with more long-term effects such as the introduction of new tools, the availability of open satellite imagery for mapping in OSM and the number of organizations involved. Furthermore, political frameworks set the policy agenda and form the institutional space in which local communities, national and international non-governmental organisations (NGOs) operate.

3.1.1 Spatial Footprint of Humanitarian Mapping

Our results revealed a general and expected significant positive relationship between population and observed mapping in OSM for buildings and highways. Furthermore, more buildings and highways have been added to OSM in regions that were impacted stronger by disasters. But the results also showed that a clear mismatch still exists between the spatial distribution of mapping activity in OSM vis-à-vis the dis-

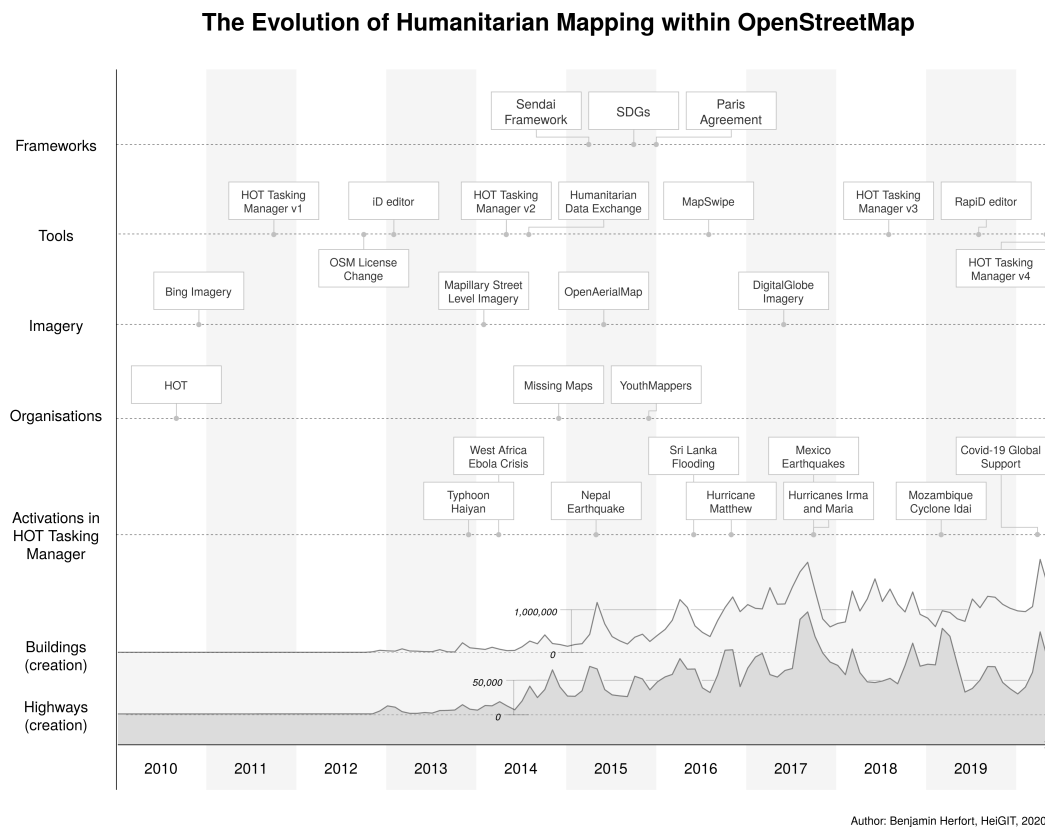


Figure 3.1: Sketch of the evolution of humanitarian mapping in OSM. Timeline of major disaster activations, the socio-technical development of the community and global political frameworks. The plots at the bottom show the number of buildings and highways added to OSM. Created using drawio 13.11.0 (<https://github.com/jgraph/drawio>).

tribution of the global population on the Earth's surface. There was a sharp contrast between population distribution per SHDI and mapping in OSM per SHDI for buildings and highways (see Figure 3.2).

Whereas only 20% of all humans lived in regions with very high human development, these regions accounted for 60% of buildings and 65% of highways created in OSM. Vice versa, in regions with low human development we counted 10% of the global population, 11% of buildings and 5% of highways in OSM. However, there was no uniform relationship between mapping in OSM and human development scores. As such, the biggest mismatch was observed for regions with medium human development, which made up for 15% of buildings and 11% of highways in OSM in overall, but represented ~36% of the global population. As these results highlight, population data could act as a first proxy variable to identify regions where OSM data is absent.

Our analysis provided further details outlining the more complex relationship between mapping activity and socio-demographic variables. The effect of SHDI on mapping activity differed clearly between overall OSM and humanitarian mapping activities. Considering overall mapping in OSM, we found a strong positive effect (more mapping in higher SHDI regions). For humanitarian mapping however, the results showed a negative effect (more mapping in lower SHDI regions). Hence, humanitarian mapping activities have helped to change the spatial footprint of OSM globally in regards to the socio-economic profile of areas mapped. To put it differently: Without humanitarian mapping efforts the observed spatial bias in OSM towards regions with very high Human Development Index would have been even more pronounced. Beyond, the results explain the relatively low coverage of OSM in regions with medium to high Human Development Index, which have been neither in the focus of "traditional" OSM communities in high-income countries nor have these regions been prioritized by humanitarian organisations.

Our results on the spatial distribution of mapping and regional variations in humanitarian mapping activity could be utilized by other researchers to obtain first insights about OSM's data coverage. For instance, Moturi et al. (2022) use information about the OSM road network to estimate access to public and private health facilities in Kenya. For Uganda, Hoek et al. (2021) examined the potential of OSM data for monitoring SDG progress in refugee settlements by measuring the spatial and temporal coverages of SDG-relevant OSM data across refugee settlements and comparing these to non-refugee settlements. Our analysis outlines that Kenya showed less-than-average humanitarian mapping contributions for buildings and highways, whereas Uganda experienced stark, above-average contributions through humanitarian mapping campaigns. This might already suggest serious differences in how well both countries are represented in OSM and should be considered when evaluating or comparing both studies.

The results do not only bear implications for areas in the focus of humanitarian organisations. The methodology proposed in the paper helps to quantify contribution bias in OSM and provides valuable meta information about which regions are represented and under-represented in OSM data on a global scale. Our research makes it obvious that analysing the OSM data density can only be the first step in understanding its spatial heterogeneity and issues of representation. The need to go from presence-only information (Where did mapping happen?) to presence-absence information (Where is mapping needed?) constitutes one of the most important lessons learned from our work in this paper.

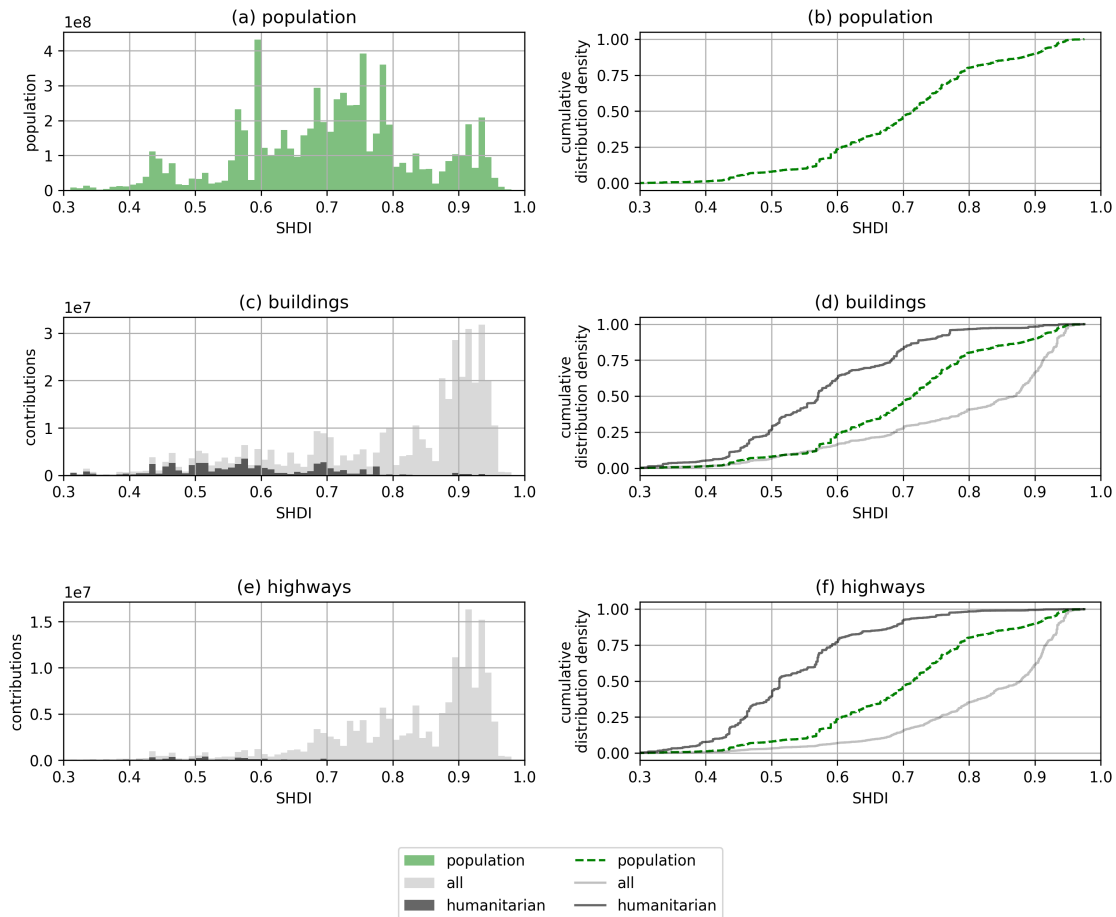


Figure 3.2: Mismatch between the spatial distribution of mapping activity in OSM vis-à-vis the distribution of the global population.

Frequency distribution and cumulative distribution density of (a)-(b) population and cumulative number of (c)-(d) buildings and (e)-(f) highways added to OSM between 2008/01/01 - 2020/05/20 in regard to overall and humanitarian mapping activities and Subnational Human Development Index. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

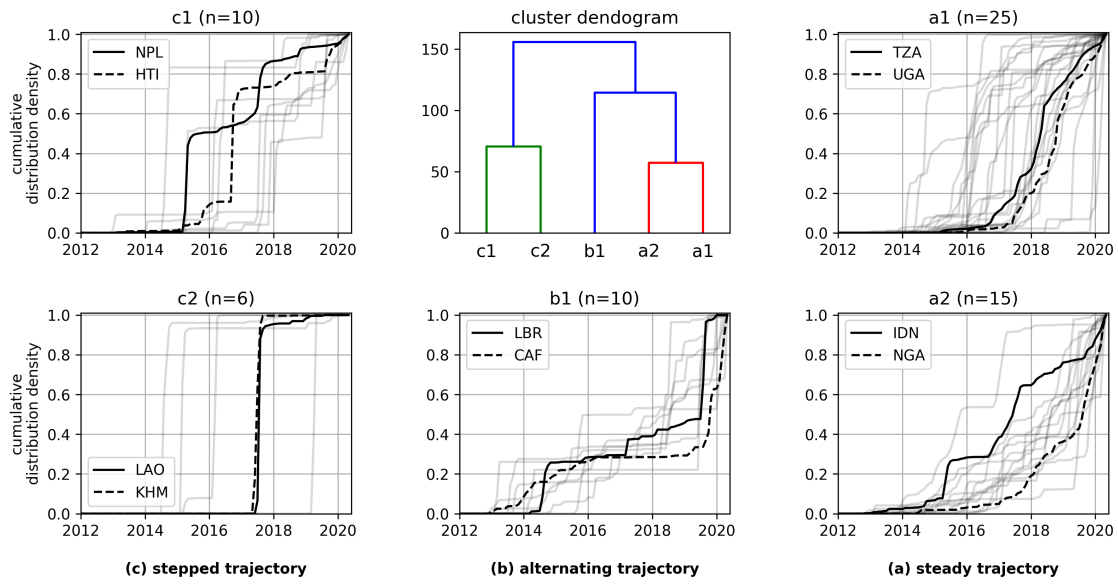


Figure 3.3: Timeline of humanitarian mapping for all countries with medium and low Human Development Index.

Cluster dendrogram and cumulative distribution density plots for monthly humanitarian OSM building mapping between 2012/01/01 - 2020/05/20. This plot depicts the values for all countries with medium and low Human Development Index (SHDI <0.7). The groups are defined by a hierarchical clustering approach based on a distance matrix derived from dynamic time warping time series similarity analysis. For each group the trajectory for the two countries with highest overall mapping activity are highlighted. We used alpha-3 ISO 3166 codes to refer to the countries: TZA - Tanzania, UGA - Uganda, NPL - Nepal, HTI - Haiti, LAO - Laos, KHM - Cambodia, LBR - Liberia, CAF - Central African Republic, IDN - Indonesia, NGA - Nigeria. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

3.1.2 Temporal Evolution of Humanitarian Mapping

In regard to the temporal evolution our results highlight that humanitarian mapping has experienced sustained growth from 2012 to mid-2017, which was followed by a phase of stagnation or a slight downward trend from mid-2017 to mid-2019. However, recent stagnation might have changed again as the number of added buildings, mappers and validators increased since mid-2019. The analysis considered mapping until May 2020.

Furthermore, the results show how this global evolution is a product of the overlaying effects of multiple mapping communities and other factors driving mapping activity at the local level (e.g. sketched in Figure 3.1). In our paper we developed a typology of the temporal evolution of OSM contributions at the country level and suggest three main types: (a) steady trajectory, (b) alternating trajectory and (c) stepped trajectory. Figure 3.3 shows the temporal evolution of humanitarian mapping for all countries with medium and low Human Development Index.

Most countries showed a steady trajectory of humanitarian building mapping in OSM over time and usually mapping in these countries followed a three stage process demonstrating the successful establishment of an active OSM community. The second mapping type consists of countries with alternating

mapping trajectories, which imply a fluctuating community with times of accelerated mapping and also times when fewer mapping happened. Finally, our results outlined countries for which times with relatively high mapping activity (e.g. during a disaster activation) have been followed by times with almost no mapping activity (e.g. after an activation). This extreme form of single event humanitarian mapping depicted in stepped trajectories puts into question the long-term effects of organized mapping campaigns in some countries towards making OSM more inclusive.

Decomposing the spatial footprint of OSM by considering the temporal aspects of mapping activity for individual countries helps us to identify distinct profiles of OSM humanitarian mapping trajectories. The methodology applied in our study might also be valuable to investigate the impact of other forms of organised mapping, such as corporate mapping, or to understand the impact of AI-assisted mapping. As such, it would be interesting to check to what extent these new forms of data creation reproduce the very same one-time-mapping patterns that have been observed in the past for some countries as revealed by our analysis of humanitarian mapping. The temporal perspective of our work emphasizes the difficulty to establish long-lasting mapping communities in some regions. In respect to questions of representation this implies that balancing the unequal spatial contribution patterns in OSM clearly goes beyond the population of a more comprehensive geographic database.

3.2 Research Question 2

The findings of the previous section highlight that particular attention needs to be paid to data inequalities and biases when OSM data is utilized in global studies or to derive global data products to inform policy and decision making. In respect to the second research question, we analyse these biases through the analysis of OSM data quality with a focus on urban areas. Despite the widespread usage of OSM building footprint data for urban analyses and policy making, an investigation of their completeness on the global scale has not been conducted so far. Consequently, our paper (see Chapter II.2) conducts research about building data completeness in OSM in the context of global urban analysis applications (**RQ2.1**). Our work further describes how unequal urban OSM building data is distributed within the space of a city, across continents and on the global level (**RQ2.2**).

In this work a novel machine-learning regression method based on a Random Forest is proposed to assess OSM building completeness. We extract an extensive collection of open building data from commercial and authoritative sources as training data and use OSM full-history data for spatio-temporal data analysis on the global scale (Raifer et al., 2019). The model further relies on information obtained from remote sensing data (land cover, population distribution, night time lights), Subnational Human Development Index (SHDI), and urban road network density as predictors.

3.2.1 Global Urban OSM Building Completeness

Our results show that for 1,848 cities (14% of the analysed amount) OSM building footprint data exceeded 80% completeness. In total, these cities were home to a population of 492 Million people (16% of the global urban population). Nonetheless, our results reveal that for 9,163 cities (69% of the analysed amount and home to 48% of the global urban population) OSM building footprint data did not reach 20% completeness. In accordance with the findings reported in Chapter 3.1, urban centers in regions with very high SHDI had the highest levels of building completeness. Beyond that and in addition to our previous work, the results of our paper were able to precisely map out where building data is missing for each city.

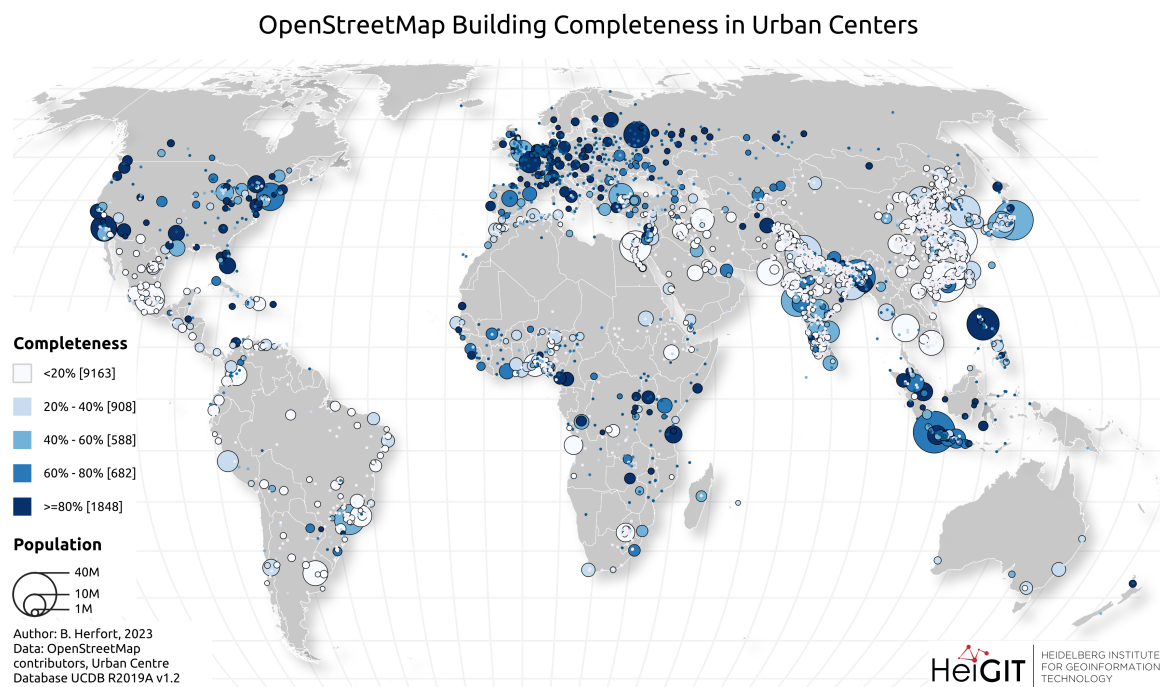


Figure 3.4: Spatial distribution of OSM building completeness in 13,189 urban centers. For each class the overall number of urban centers is reported in the squared bracket. For an interactive web map visualization visit https://hex.ohsome.org/#/urban_building_completeness. OSM data as of 2023-01-01. Created using QGIS 3.28.3 (<https://www.qgis.org/en/site/>).

Relatively high completeness was estimated for Europe & Central Asia (71%) as well as for North America (64%). Completeness values lower than the global average were observed for the regions Latin America & Caribbean (20%), East Asia & Pacific (20%), Middle East & North Africa (12%), and South Asia (9%). Sub-Saharan Africa completeness (30%) was slightly higher than the global mean.

We found that organized humanitarian mapping activities in urban centers contributed an average of about 10% of the building footprints globally. However, humanitarian contributions were focused on specific regions, especially in Sub-Saharan Africa where more than 50% of all building edits were related to organized humanitarian mapping activities. As a consequence, the completeness in low SHDI urban centers was higher than the completeness in urban centers with high SHDI. In addition, the spatial distribution of building completeness across urban centers showed a strong regional variability across that global trend previously unreported: numerous cities in any region were mapped with a very high completeness regardless of the overall completeness or mapping activity in that region (see Figure 3.4).

Besides the well-examined Global North - Global South bias in OSM, our results demonstrate that the OSM building stock as of 2023 showed a much more spatially diverse pattern than previously considered. It was shaped by regional, socio-economic and demographic factors across several scales. Most importantly, the proposed approach assesses completeness (and not only coverage) of OSM buildings at a fine scale and is genuinely suited to make visible where data is missing. For instance, our results might help to estimate in which cities the building-by-building exposure model presented by Nieves et al.

(2022) which combines information from OSM with census statistics to estimate building damage after earthquakes, can be trusted due to sufficient completeness of building footprints in OSM.

Our results confirm the similar, albeit less dichotomous, global completeness pattern of land use land cover (LULC) information from OSM (Zhou et al., 2022a) and stand in slight contrast to the relatively higher and more evenly distributed completeness for OSM's road network (Barrington-Leigh and Millard-Ball, 2017). This highlights the complexity of OSM mapping activities and the challenge for analyses based on OSM data. The unequal patterns of completeness evidenced make it essential for data analysts to assess the potential negative impact of missing data.

3.2.2 Spatial Inequalities in Urban OSM Building Completeness

Besides analysing differences in average OSM building completeness on the continental scale, the Gini coefficient and Moran's I have been applied to derive the degree of evenness and global spatial autocorrelation of urban OSM building completeness. This analysis has been conducted considering the temporal evolution of mapping in OSM.

The temporal evolution of the Gini coefficient since 2008 indicated that both globally and regionally, OSM building data distribution has become slightly more even over time. In contrast to OSM's evenness, we find that global spatial inequality in OSM building completeness sharply increased between 2008 and 2014. During that time mappers favoured cities near already well-mapped cities and the expansion of OSM mapping to distant and un-mapped regions (likely to be located in the Global South) did not happen at a significant scale. Since 2014, Moran's I declined. This means that the spatially clustered completeness pattern became less intense, albeit still clearly visible. Combined with the decrease of the Gini coefficient in the same period, our results suggests OSM building data in 2023 was much less segregated in both evenness and clustering compared to the state-of-the-map in 2014.

The intra-urban spatial heterogeneity of OSM building completeness was estimated at a resolution of 1x1 kilometers. Based on Gini coefficient, Moran's I and OSM building completeness, urban centers were classified into three different types in an agglomerative clustering approach (see Figure 3.5). Urban centers of type (1) showed very low completeness and could appear as white spots on the map, more aptly described as the "unmapped" cities. They are characterized by a high Gini coefficient and a very low Moran's I value. The second group of urban centers is characterized by a relatively high Moran's I and moderate to high evenness covering cities with wide range of completeness. For these urban centers, the overall completeness value hardly reflected the local completeness values. Finally, urban centers with the highest overall completeness and very low spatial clustering were most likely to get classified as type (3). As opposed to type (1), the urban centers of type (3) could be described as the "well-mapped" cities and represent some finality of mapping building footprints in OSM.

The methodology proposed and the results obtained from our paper will be of high value for comparative studies, which usually seek to assess a large set of indicators involving multiple cities across the globe. Our findings allow researchers to quantify biases in OSM building completeness and let them factor in inequalities in OSM in their study design. For instance, the Global Building Morphology Indicators project proposes several multi-scale measures to characterize urban areas based on building level data from OSM and other sources of open data (Biljecki and Chow, 2022). Debray et al. (2023) also predominantly rely on geo data from OSM to investigate the relationship between diversity of planning processes and the existing morphology of the urban landscape and introduce the concept of "Intensity of Plannedness". For the case of evaluating quality of OSM data for slums, Yeboah et al. (2021) highlight

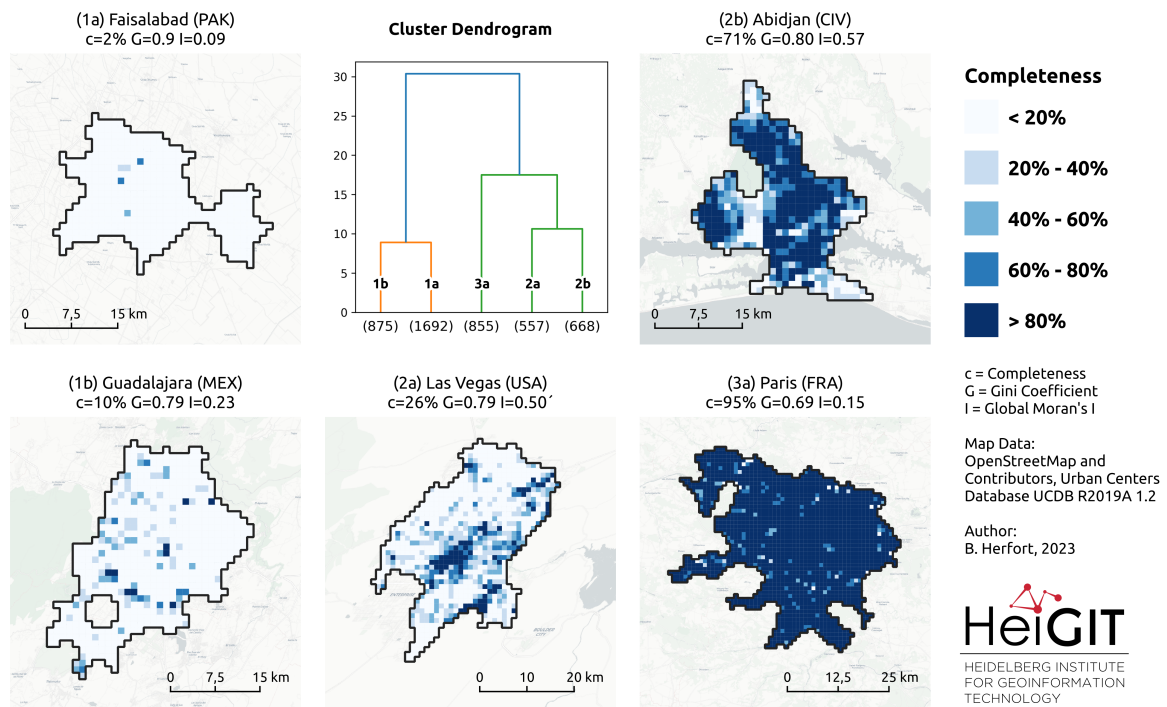


Figure 3.5: Spatial distribution of intra-urban OSM building completeness for selected urban centers.

For each urban center we report on overall OSM completeness c , Gini coefficient G and Moran's I . Cell size is always 1x1 kilometers for any urban center. The number of urban centers in each cluster is indicated in the dendrogram. For an interactive web map visualization visit https://hex.ohsome.org/#/urban_building_completeness. OSM data as of 2023-01-01. Created using QGIS 3.28.3 (<https://www.qgis.org/en/site/>) and Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

the major difficulties that arise because of the absence of external reference datasets and limitations of existing intrinsic quality analyses.

The approach proposed in our work could open up an additional avenue to tackle these challenges to estimate biases in OSM. To support this assessment, we provide an open dataset (Herfort, 2023) resulting from this study with estimated completeness maps for 13,189 urban agglomerations worldwide using a 1x1 kilometer grid.

3.3 Research Question 3

Recently, remote sensing and computer vision technologies have been successfully employed to derive information on human settlements at regional to global scale. These approaches are usually motivated by the fact that existing datasets are not available globally or — as in the case of OSM — spatially biased. Automated approaches offer the opportunity to analyse vast regions with uniform methodologies and might be able to speed up mapping workflows to overcome the existing data gaps in OSM outlined in Chapter 3.1 and Chapter 3.2. Hence, in our paper (see Chapter III.3) we investigated the potential of deep learning based building detection approaches in conjunction with crowdsourced mapping.

We addressed this by comparing crowdsourcing (MapSwipe) and deep learning (DeepVGI) approaches for generating human settlement maps to existing EO-based products (**RQ3.1**). MapSwipe is a mobile application which allows volunteers to swipe through satellite images of a region. The app collects user feedback on classification tasks. We evaluated four different approaches (MapSwipe, DeepVGI, Global Urban Footprint (GUF) and High Resolution Settlement Layer(HRSL)) by investigating the quality of the derived results against a reference data at three study sites: (a) Guatemala, (b) Laos and (c) Malawi.

We outlined the spatial and non-spatial characteristics of misclassification observed in the DeepVGI dataset in further detail (**RQ3.2**). Finally, we investigated the added value of a novel task allocation strategy, which combined deep learning and crowdsourcing with respect to performance and effort (**RQ3.3**).

3.3.1 Comparison of crowdsourcing and deep learning approaches to existing EO-based products

Our results supported the common hypothesis that while mapping on the basis of satellite imagery humans rarely identify something as a building which is not a building, but tend to miss buildings. In contrast, our results indicate that deep learning approaches and EO-based approaches tend to miss fewer buildings at the cost of also falsely detecting a range of objects which are not buildings. For the study site in Guatemala the results are visualized in Figure 3.6. The maps for the other sites can be found in Chapter III.3. The findings from our study suggest that beyond filling data gaps the potential of crowdsourcing for improving EO-based datasets in fact lies in its ability to reduce the number of false positives and to increase the specificity of EO-based or deep learning based datasets.

For all three case studies crowdsourcing outperformed the other approaches with respect to accuracy and Matthews correlation coefficient. In regard to the crowdsourcing approach based on MapSwipe, false negatives (users miss a building) were more likely than false positives (users map something as a building, which actually is not a building). As a consequence, crowdsourcing of human settlements was characterized by high specificity (99% for all study sites) and intermediate sensitivity.

In our study EO-based approaches were ranked lowest in regard to overall accuracy and Matthews correlation coefficient. Both EO-based datasets (HRSL and GUF) achieved moderate specificity. This

indicates that these datasets suffered from a relatively high number of false positives. The results revealed very low sensitivity for the GUF dataset especially in more rural study areas such as the one in Malawi or Laos. However, the more traditional EO-based approach relying on the HRSL achieved best results from the perspective of sensitivity.

The DeepVGI approach, which utilizes OSM data as training samples for deep learning based building detection, achieved lower specificity compared to the crowdsourcing results, but also reached a higher sensitivity. Hence, similar to the other EO-based approaches false positives were expectedly the major concern for the DeepVGI approach as well.

There was a clear spatial trend for all study sites, which revealed better performance for major (large) settlements. The most uncertain results were located in the areas with a mixed land-use, e.g., agricultural land and settlements. Furthermore, uncertain results seemed to be spatially clustered. The spatial and non-spatial characteristics were explored in a subsequent step to inform the task allocation procedure in a combined crowdsourcing-deep-learning approach.

3.3.2 Towards combining of crowdsourcing and deep learning approaches

Our results suggest that the task allocation strategy helped to exploit the complementary value of a sensitive method (DeepVGI) and a specific method (MapSwipe) and would improve the existing crowdsourcing approach MapSwipe uses (see Figure 3.7). Allocating 80% of the mapping tasks towards the automated DeepVGI approach and only 20% towards the human crowdsourcing approach did not have a negative effect on the overall accuracy of the data produced. As a consequence, from an efficiency perspective our results endorsed that for the creation of human settlement maps we should rely on automated approaches (e.g., machine learning) when possible, but rely on human skills (e.g., citizens science and crowdsourcing) when needed.

Combining the MapSwipe and DeepVGI methods using the confidence score-based task allocation strategy increased performance in terms of data quality measured by Matthews correlation coefficient by around 3 - 5 percentage points (compared to the crowdsourcing-only approach). For all study sites, this was caused mainly due to the ability of the combined approach to correctly identify potential false positives in the DeepVGI dataset and allocate these towards the crowdsourcing. When investigating the results of the random task allocation no such effects were observed. More dramatically, the proposed task allocation strategy reduced the volunteer efforts to one fifth (labour reduction of 80 percentage points). The performance of the combined approach did not change, when allocating 20% to 70% of the sorted mapping tasks towards crowdsourcing.

Considering the needs of humanitarian organizations (e.g. getting information on human settlements for which no other data sets exist) our approach could be used in real crowdsourcing or expert mapping applications. Indeed, recent mapping campaigns have already utilized Meta's RapID OSM editing tool, which allows OSM users to add and edit buildings or roads generated through machine learning to the map.

The findings from our study have already been used by other authors in the domain to design systems that rely on the interaction of human and machine learning approaches. For instance, the Centaur VGI describes such a hybrid human-machine mapping approach (Huck et al., 2020). In contrast to the earth observation perspective which sees crowdsourced data usually only as training samples, the human-ML mapping interaction research usually starts from a community mapping or crowdsourcing perspective and tries to explore how machine learning can increase the efficiency of human mapping approaches. Watkinson et al. (2023) take this one step further. They add the element of gamification to the equation

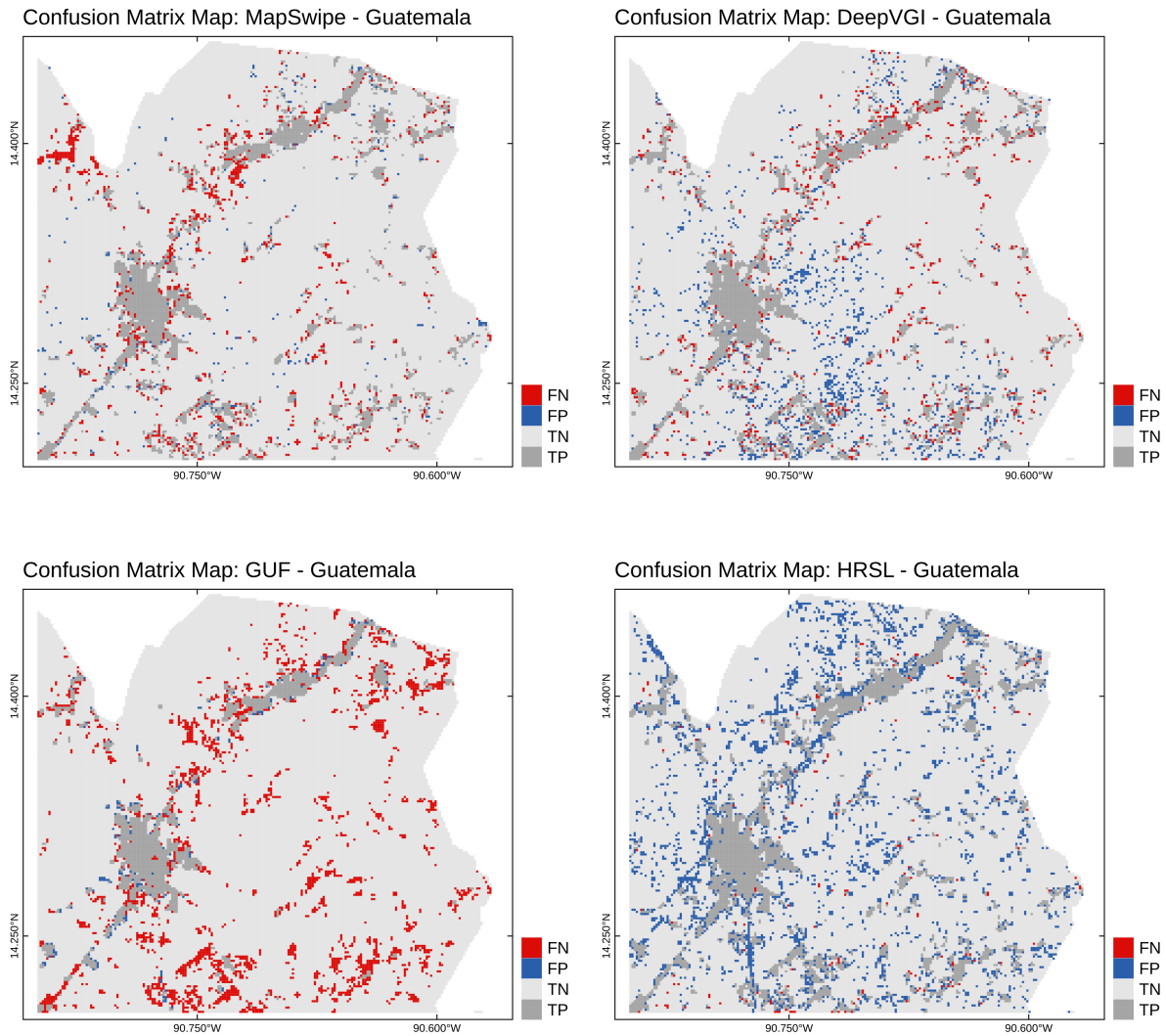


Figure 3.6: Performance of crowdsourcing, deep learning and EO-based human settlement maps.

Map representation of the confusion matrix for Guatemala: Each map shows the spatial distribution of correct "building" (TP) and "no building" (TN) classifications for a specific method in the testing area. Incorrect classifications are split into false positives (FP) and false negatives (FN). Each pixel corresponds to a single task / satellite imagery tile.

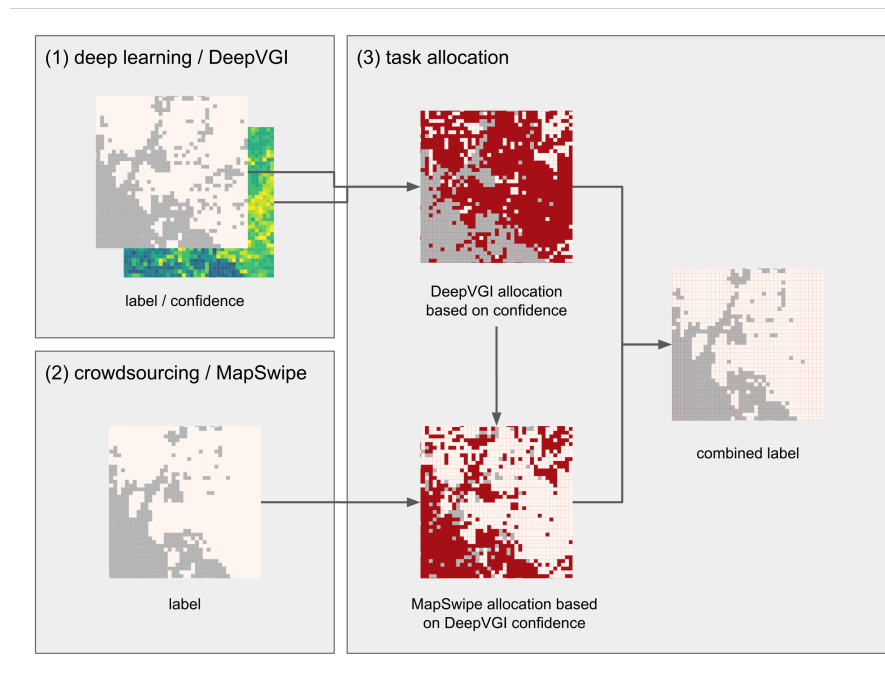


Figure 3.7: Combining crowdsourcing and deep learning for an AI-assisted mapping workflow. Proposed workflow to combine deep learning and crowdsourcing methods: Combined labels are obtained through a task allocation strategy, by choosing labels either from DeepVGI or MapSwipe based on the confidence of the DeepVGI labels.

and investigate how these elements might encourage sustained volunteer engagement without reducing contribution quality in human ML-VGI mapping platforms.

Due to the standpoint of the OSM community, until now no machine learning derived data is being added to its database completely automatically and it seems to be unrealistic that this will be the case in the foreseeable future. However, even if all data is to be reviewed by humans, the presented task allocation strategy might help to prioritize and identify data quality issues and could guide real-time or near real-time validation and feedback for mappers.

Our results show that machine learning-assisted mapping approaches could be conceptualized as "review" systems, which help mappers to identify errors in automated data creation pipelines. However, it is not yet clear to what extent these systems might be suited to overcome the prevalent representation biases and data gaps in OSM. The success of novel mapping approaches will be determined upon their ability to reduce data gaps in regions which until now have not been in the focus of traditional mapping communities in high-income countries nor humanitarian mapping campaigns.

3.4 Limitations

Although the specific restrictions in the scope of each research question are described in detail within each paper, there are some overarching aspects, which facilitate the interpretation of our findings. These limitations are grouped across three topics: a) methodological limitations, b) uncertain data and c) scope restrictions.

3.4.1 Methodological Limitations

This thesis described the inherent spatial biases in OSM, but the methods applied are also not impartial and are spatially distorted to a certain extent. This is caused by spatial autocorrelation and the associated problems in regression coefficient estimation (Dormann et al., 2007). In our work we relied on various regression modelling techniques (e.g. logistic regression, generalized linear models and random forests), which have been selected in regard to the particular goal of each research question. The results of these models, such as the random forest model used to derive building area predictions (Herfort, 2023), are conditioned by potential spatial and non-spatial biases present in the training data or by biases that arise from the algorithms themselves (Mehrabani et al., 2021). As highlighted in Herfort et al. (2021) one has to keep in mind the presence of strong spatial patterns in covariates as well as in the residuals of predictions. These effects are considered in the modelling approaches for our papers as much as possible, but might lead to optimistic model performance estimates. Addressing algorithmic uncertainty in geographic knowledge generation by testing different sets of procedures remains a key mandate for big data research in general (Kwan, 2016).

To be transparent about model uncertainty, we reported on the model performance utilizing a spatial cross validation procedures. For instance, in Herfort (2023) we utilized a cross validation procedure based on k-means clustering and we found that completeness estimation performed well with a global r^2 score of 0.9, but a slightly higher uncertainty in the predictions was observed for Sub-Saharan Africa. In regard to the work presented in Herfort et al. (2021) we relied on Spatial Eigenvector Mapping (Griffith et al., 2019; Chun and Griffith, 2013; Griffith and Peres-Neto, 2006) and also observed a very high explained deviance value in regard to overall highway mapping of 0.9, but at the same time a considerably lower value for buildings (0.79). These inherently geographic variations may constitute a starting point for local communities and researchers to design local models that can overcome the limits of the global modelling approaches used in our studies. Future research should investigate the "sweet spot" between local models, which might provide more certain predictions in smaller regions, and global models which might be easier transferred to other, more distant regions.

A second major methodological limitation applies to the depth of the selected data (quality) assessment methods. Whereas our research provides a way forward to better quantify data production and data completeness biases and how they emerge over time, Schott et al. (2021) outline that more insights are needed towards understanding what type of data of which quality is created. Also Atwal et al. (2022) have highlighted that not only the physical completeness, but also the incomplete nature of the attribute data is a shortcoming that limits the usefulness of OSM data. It is obvious that for some sectors, such as public health programs, assessment of completeness is mandatory, but still only the first step. Information on building usage is also required, but often only available for a small subset of buildings (Sturrock et al., 2018). In this regard the methods used in this thesis have very much focused on the presence/absence of map features. Biljecki et al. (2023) have outlined first approaches towards assessing

the completeness and consistency of building attribute information in OSM. Further research is needed to better understand OSM's biases in this and other dimensions of data quality on a global scale.

3.4.2 Uncertain Data

A particular overarching limitation, which affects all three studies, lies in the uncertain quality of external datasets employed, e.g. used as explanatory variables in the statistical models. Despite the abundance of data and trend towards releasing global scale datasets, information about their uncertainty is rarely provided. Apparently, information obtained from Global Human Settlement Population (GHS-POP) and SHDI, which are used in Herfort et al. (2021) and in Herfort (2023), provide only imperfect proxies for the actual population distribution and socio-economic profiles of regions. There are a few other data sets, such as WorldPop, that provide information on human population on the global scale as well (Tatem, 2017). Those could be used for comparison to estimate uncertainty. However, in our work we could not further propagate the uncertainty (and potential error) of any of the datasets utilized for the uncertainty estimation of our results. Type II spatial regression models, which are able to capture uncertainty in predictors and not only in the response, could be explored for this in future research.

The ever increasing amount of data on the global scale might ease this limitation. But it has also been shown that thematically similar datasets have related limitations, i.e. the estimation of inhabited areas is biased towards urban areas (Klotz et al., 2016). As a consequence, settlements in rural areas are more likely to be left out in global population data sets as well. This restriction was stressed in our own study results as well when comparing different EO-based human settlement datasets (Herfort et al., 2019). The consequences were considered in our research design by focusing on urban areas for the work presented in Herfort (2023), albeit this leads to additional scope restrictions.

Another concurrent uncertainty factor stems from the varying temporality of the datasets employed. For example, for the work presented in Herfort (2023), the data for ESA WorldCover, GHS-POP and VIIRS Nighttime Lights (VNL) are subject to temporal uncertainty: They reflect the situation for the year 2020, albeit they are utilized in a model to assess OSM building completeness for a time range from 2008 to 2023. Hence, the model will be most suited to predict the building stock of 2020 and not the past building stock, nor the correct building stock for 2023. The same uncertainty applies to the results of the regression analysis quantifying the effects of socio-economic and demographic profiles in regard to the global footprint of OSM (Herfort et al., 2021). Similarly, satellite imagery quality might strongly influence the performance of the human-ML task allocation strategy presented in Herfort et al. (2019) not only in respect to spatial resolution but also regarding temporal coverage. In situations where settlement pattern change rapidly, e.g., due to forced displacement, on the ground data is irreplaceable unless up-to-date satellite data and derived products become available.

While there is an abundance in data, there is still scarcity in metadata. The validation of global datasets and the estimation of their uncertainty remains a notably difficult challenge for future researchers. Benchmark datasets, such as the one provided by See et al. (2022) for validating built-up surface layers, are highly valuable. They are suitable for performing an independent audit taking into account study locations and research questions. For many domains these datasets do not exist yet. Future research could explore to what extent OSM data itself might be suited as a benchmark for some topics.

3.4.3 Scope Restrictions

A profound limitation of our work is its restriction to digital mapping which is mainly organized remotely, i.e. usually on the basis of very high resolution satellite imagery. As a consequence, our insights about humanitarian mapping in OSM only provide an incomplete picture which lacks an on-the-ground perspective. The limitation is accompanied by the fact that our analysis only focused on human settlements in general and on two particular types of mapped objects (buildings, highways). Mapping in OSM comes with a much greater variety of potential map objects (e.g. health facilities, schools, water points), which can add particular value in comparison to other geographic data sets. Both restrictions are mutually dependant to some extent, as local data generation methods usually require information which cannot be easily extracted through visual interpretation of satellite imagery.

Stemming from the limitation of uncertain data described above, the scope of our work is not exclusively restricted to urban areas, but there is a considerable preference. Our urban building completeness study (Herfort, 2023) encompasses about 50% of the global population which lives in urban agglomerations. Still one should be careful to transfer our findings to rural areas. While our other papers do not explicitly exclude rural areas in their analyses, our findings will probably be biased towards urban areas.

Finally, one should carefully read our results when trying to tackle biases in data quality and representation in OSM. Our research shows the scale of digital humanitarian mapping and provides evidence that more people than ever are currently being represented on OSM. But the methods we employed are foremost suited to describe the spatial variations in map data and the past evolution of it. They do not allow a direct prediction of the scale and location of future mapping efforts and their effects on issues of representation.

Map data is more detailed and covers more more places than ever before. However, Bryant (2021) outlines that "the means by which those mapped can access their own data have not improved at the same pace". Our work narrows down the question of representation in OSM towards an analysis about the availability of data. Unfortunately this approach is not suited to actually measure change towards sustainable development as postulated by Porto de Albuquerque et al. (2021). Our methods show where map data is missing and should be considered just as an "overview map" of the unmapped places. When zooming in, our work misses the details and without further context it is not suited to prioritize which regions should be mapped in the future. This limitation should be addressed in future work through the combination of our results with in-depth analyses of specific communities better accounting for the local perspective, e.g. using mixed qualitative-quantitative methods.

Chapter 4

Discussion

Questions of representation and data quality in OSM from the perspectives of humanitarian and machine learning-assisted mapping were the focus of this dissertation. Accordingly, this thesis has investigated the impact and scale of humanitarian mapping in OSM (RQ1). Furthermore, the analysis of urban building data has provided an example for how inequalities in OSM completeness are spatially distributed (RQ2). Finally, the potential for a combined crowdsourcing machine learning mapping workflow has been explored (RQ3). As introduced in Chapter 2, these three research questions should not be discussed separately, but need to be set into perspective with each other.

First, Section 4.1 of the discussion outlines the interconnection between representation biases and data gaps. On the one hand representation biases in OSM, e.g. caused by socio-economic inequalities, lead to data gaps (Chapter 3.1). But on the other hand the assessment of data gaps, e.g. through OSM building completeness analysis, can also unveil issues of representation (Chapter 3.2).

We have shown in detail in Chapter 3.3 that there is potential for machine learning-assisted mapping to fill data gaps. At the same time, data quality assessment can help to better specify the limitations of mapping workflows, including novel approaches utilizing machine learning systems or established workflows in humanitarian mapping (Chapter 3.2). The relationship between these two topics will be discussed in detail in Section 4.2.

OSM, as a massive source of training data for machine learning systems, has fuelled the advance of automated mapping workflows. Machine learning systems show the potential to reduce and reproduce inequalities in OSM (Chapter 3.3). In Section 4.3 we will discuss that once more this raises questions about representation and how the associated biases could be overcome.

Finally, focussing on the linkages between the three research questions will provide a guide for the conclusion of this dissertation presented in Chapter 5.

4.1 Representation Biases and Data Gaps

Mapping efforts of communities in OSM over the previous decade have made OSM a unique global database, which is accessible to all with no licensing costs. Partly due to our research, at the current state (humanitarian) organizations do already monitor their impact on OSM, e.g. through the Humanitarian OSM stats website (<https://humstats.heigit.org/>). Also the YouthMappers network is tracking mapping activity in OSM and provides an interactive dashboard to monitor the growth of the network (Solís et al., 2021). However, these attempts merely consist of counting the overall number of buildings and highways added to OSM and usually neglect the spatial and temporal components of map data quality.

The increasing contribution of corporate mappers (Anderson et al., 2019) is another pressing field where a long-term monitoring and fine-grained understanding of temporal and spatial components of mapping in OSM is needed. Our results highlight that there is a need to go beyond monitoring mapping activity and to improve the methods to monitor data quality.

By combining completeness maps with socio-demographic characteristics of the areas of interest it will be possible to ensure a fair and balanced selection of target geographic areas. This could reduce existing data inequalities and issues of representation within OSM on the global, regional or intra-urban level. In light of the critical challenges to finance high-quality data systems for addressing inequalities in SDG monitoring in both low and middle income countries (Ulbrich et al., 2019) — despite heightened demand (Sachs et al., 2022) — the creation and usage of OSM data could be promoted further as a cost-effective alternative to authoritative data sources. Data generation with OSM not only allows filling the current data gaps essential to monitor SDG progress, but could also be a pathway for equitable urban transformation by empowering local communities to have a voice and benefit from the data production process (Porto de Albuquerque et al., 2021).

This dissertation provides a good point of departure for a more holistic approach towards understanding OSM's data gaps considering questions of representation. Recently, there have been more and more studies which investigate data quality of OSM on the global scale (Barrington-Leigh and Millard-Ball, 2017; Zhou et al., 2022b,a). Beyond reporting data quality measures from a purely technical perspective it will be needed to connect these measures to mapping practices and characteristics of specific mapping communities. The work of Grinberger et al. (2021) about large-scale data production events provides a methodology to distinguish different mapping practices, but only loosely links these to representation biases.

For the case of Wikipedia researchers have already investigated question of representation in more detail on a global scale, e.g. in respect to gender (Johnson et al., 2021; Graells-Garrido et al., 2015) and locality of participation (Graham et al., 2015). By applying mixed qualitative-quantitative methods, researchers should further help to better understand how existing data gaps could be overcome, instead of just reporting about them. This could be informed by the taxonomy of knowledge gaps identified for Wikipedia and its sister projects, which have been defined as disparities in content coverage or participation of a specific group of readers or contributors (Redi et al., 2020). As defining these gaps (in Wikipedia and OSM) is an inherently value laded process, this holistic perspective needs to be driven by communities and build upon neutral criteria, flexibility and measurability (Redi et al., 2020). Similar research has not yet been conducted for aspects of representation in OSM on a global scale, but only for smaller regions and case studies. Future research needs to investigate to what extent the methodologies, which are proposed to outline global disparities in Wikipedia and other projects, could be utilized to advance studies about OSM as well.

Temporal data quality trajectories might reveal to what extent data quality dimensions, temporal evolution of mapping activity and inequality measures are mutually dependent. As such, it is about time to reproduce some of the "old" OSM data quality studies to properly account for OSM mapping and its evolution in other geographies. Often these studies solely focused on regions in western Europe at a single point in time. As one example for this, the work of Neis et al. (2013) has compared mapping activity and the composition of the OSM community for twelve urban areas around the globe, but almost all were located in high income countries. In a remarkable manner, the spatial bias in OSM-related research, approximated by the location of authors, is even bigger than the spatial bias observed in OSM data (Grinberger et al., 2022b). Perspectives from researchers and OSM community members located

in low income countries and the Global South are very much needed, but still under-represented in academia.

Providing a toolset to run OSM data quality analyses based upon common methods would be a big step forward towards a more reproducible OSM data quality science and might also allow for a wider participation from other research groups. The OSHDB and ohsome API have already reduced barriers towards the global analysis of OSM's historic evolution (Raifer et al., 2019), but have not yet included common data quality measures.

The global community of OSM data producers should be able to plan where future mapping should take place. The assessment could be based on the combination of OSM data quality layers, such as the completeness maps provided in this thesis, and additional layers on representation biases in OSM. This can help to improve coverage in OSM and balance representation biases at the same time.

4.2 Machine Learning Systems and Data Quality

The most prominent argument for the integration of automation into the mapping workflow is the need to overcome data scarcity. Similar to the framing of the "Missing Maps" project (Scholz et al., 2018), this perspective addresses urgent humanitarian needs and aims at making progress towards the SDGs. As such, OSM data quality research and technological advances towards automated geographic data production can complement and guide each other.

To overcome the problem that data about educational facilities is often inaccurate, incomplete or non-existent Yi et al. (2019) describe a machine learning approach to determine school detection at national scale in Colombia from high-resolution imagery. From a similar stand point of data scarcity, Bhuyan et al. (2022) propose a semi-automated workflow for generating a building elements-at-risk database in Kerala (India). The CentaurVGI approach developed by Huck et al. (2020) works with a semi-automated workflow for building mapping, because "inequalities in the coverage, quality, and currency of mapping persist, with significant gaps in remote and rural parts of the Global South".

There is no shortcoming in research about the possibilities that machine learning and AI-assisted mapping can provide in general and for improving OSM data quality in particular. Vargas-Muñoz et al. (2021) compiled a comprehensive list of tasks, i.e. using ML approaches to correct street network topology, extract roads from aerial images, detect geometric mismatches of building footprints or improve tagging recommendation systems. Also the human-computer interaction method proposed by Vargas-Muñoz et al. (2021) stresses on efficiency gains, as the methods allows to correct 98% of wrong OSM building annotations by analysing less than a quarter of the data. Our research findings support this general notion about the potential of machine learning systems to improve crowdsourced mapping workflows.

In regard to the creation of human settlement maps, we concluded in Herfort et al. (2019), that one should rely on automated approaches (e.g., machine learning) when possible, but rely on human skills (e.g. citizens science and crowdsourcing) when needed. In the meantime, technical developments towards machine learning based methods to extract buildings and roads from very high resolution satellite imagery are showing their impact on OSM already today (Li et al., 2020; Chen and Zipf, 2019). For some places, the building footprints in OSM might already come from a machine learning derived dataset, but they need to be accepted by a human-in-the-loop process, for instance using the mapwith.ai editor (Meta, 2022) or similar tools. A mapper utilizing Microsoft ML Building data can map a higher number of buildings in OSM compared to a participant who maps all buildings manually in the same time (Fila

and Štampach, 2022). Still, a large direct import of Microsoft building data into OSM seems unlikely at the moment, even though it has been shown for the road network that an interplay of data imports and updates by contributors could improve OSM data significantly (Witt et al., 2021; Zielstra et al., 2013).

From a data quality perspective, it is very likely that AI-assisted mapping will be most strongly employed for the very regions we identified as "incomplete" in our research to fill the gaps that existing mapping approaches have not filled so far. Data quality assessment methods can help to outline where data is scarce and further support the assessment of the effectiveness and quality of ML-based mapping. Especially in situations when data is needed urgently, quickly available ML-based approaches and data will presumably be used if no other data sources already exist. Similar to changes in other domains which rely on expert knowledge (Harig et al., 2021; van der Vaart et al., 2020), these machine learning-assisted workflows will shift the mapping more towards reviewing data, instead of creating it. Data quality validation in OSM is a time consuming task and requires more mapping experience and thus already constitutes a challenge for organisers of mapathons (Coetzee et al., 2018). Hence, it should be further monitored how shifting towards more automated mapping approaches affects mapper retention and how successful existing capacity-building strategies could to be aligned (Dittus et al., 2016). As a perceived shortcoming of experience mappers that do validation tasks in the HOT Tasking Manager exists already, it remains uncertain if organised AI-assisted mapping can engage enough experienced validators to maintain good data quality in OSM.

Similar to the biases introduced by the emergence of humanitarian mapping, machine learning and AI-assisted mapping will put their own particular footprint on OSM. It is beneficial to understand these new biases already while they are evolving. In that way local communities and other stakeholders involved in mapping can respond in time if needed. Tools such as "Show me the way" (OSM Lab, 2023) and "OSMCha" (MapBox, 2023) do already provide some real-time insights into how and where OSM changes. Future research needs to provide the methodological backbone to make sense of the OSM data edit flow in near-real time, going beyond visualizing what has changed. From our research we have learned that near-real time monitoring should not only consider counting contributions, but needs to actually monitor map data quality.

As our research highlighted the temporal dynamics in OSM mapping activity and data coverage, it is important that scientist will become able to access OSM data quality indicators for their areas of interest at the given point in time. This should support researchers in their efforts to outline the spatial disparities on several scales and how they emerge over time.

4.3 AI-assisted Mapping and Representation Biases

For more than 30 years critical cartographers have challenged the premise that mapping constitutes an unquestionable 'scientific' or 'objective' form of knowledge creation (Harley, 1989). As such, cartographic facts are always embedded in a specific cultural context. Critical GIS research and Critical Data Studies have renewed this critique to new forms of (automated) geographic knowledge creation fuelled by "big data" (Kitchin, 2013; Graham and Shelton, 2013). For the case of remote sensing, it has been argued that technology itself "is not the deus ex machina to resolve global challenges nor emancipatory struggles" (Bennett et al., 2022) and the same might be true for the application of machine learning and AI in geography.

The emergence of machine learning-assisted mapping workflows bears the "potential" for even faster one-time mapping efforts at relatively low costs combined with even fewer engagement from local com-

munities. It has been shown for OSM that the geographic shifts and socio-institutional changes already produce an intertwined dynamic between colonizing and decolonizing processes (Schröder-Bergen et al., 2021). On the one hand, these shifts reduced inequalities in OSM, but at the same time there is also the danger that local geographic knowledge is overwritten and marginalized once more (Specht, 2020). More efficient map making, e.g. through AI-assisted workflows, does not necessarily lead to more effective humanitarian response, as this would require inclusion as a key determinant (Bryant, 2021).

Consequently, probably the most controversial question in regard to the potential of AI-assisted mapping is about defining when and for what particular purposes it is actually "needed". This is not primarily a technical question of what is "possible" (as emphasized in our conclusion in (Herfort et al., 2019)), but requires putting into perspective the potential non-technical side effects that come with a more automated mapping workflow. Given the revolutionary force that many researchers and companies attribute towards new machine learning based approaches and business models, it is important to raise again our opening question of who is actually represented in OSM.

As for the general use of machine learning systems this comes not only with hopes, but also various challenges. For instance, it has been shown that technological improvements through AI-based systems may enable 124 targets across all SDGs, while they could also have a negative impact and inhibit 59 targets (Vinuesa et al., 2020). Vinuesa et al. (2020) also find that the negative impacts of AI could be prevalent for SDGs 1 (No Poverty), SDG 4 (Quality Education) and SDG 10 (Reduced Inequalities). This suggests that in general machine learning may act as an accelerator and amplifier, which will allow us to do tasks faster and more efficient, but at the same time could still provide the wrong incentives to overcome existing inequalities in OSM.

Initially, crowdsourcing and mapping in OSM have contributed a lot to shift-away from top-down approaches and introduced an alternative way of map data creation (Goodchild, 2007). For the case of deprived area mapping, Kuffer et al. (2020) underline that because automated EO-based mapping approaches are mostly top-down, they need to be complemented with bottom-up approaches to incorporate ground information and foster the interaction with urban communities and stakeholders. For humanitarian mapping, this notion could provide a general guideline how to approach and potentially incorporate human-ML based mapping workflows without falling back into "archaic" pattern of top-down map making with their related issues of unbalanced representation.

This dissertation has outlined the biases present in OSM data for the case of building completeness in urban centers and on a global scale for mapping activity in respect to roads and buildings. It has been argued that the inherent bias in user-generated data, such as buildings or roads obtained from OSM, is likely reflected in the results of machine learning models using this data for training (Mehrabi et al., 2021). Mehrabi et al. (2021) list several important types of such biases including historical bias, population bias and content production bias which our research shows are prevalent in OSM as well. This might lead to the awkward situation that machine learning-assisted mapping is applied first to the very regions not mapped in OSM, although the machine learning derived data is least fit for these places due to their under-representation in the training samples.

Future research needs to address these questions of representation and inclusion by improving the methods to make visible who actually participated in (and who was excluded from) the map making process. In any way, for OSM communities, humanitarian organisations and OSM data users this will create an increased need for technical know-how about machine learning. This will allow them to balance mapping speed, data quality concerns and costs in an informed way.

Chapter 5

Conclusion

This dissertation studied several approaches towards mapping what is not yet mapped in OSM. This has been addressed from three different perspectives: Where are the data gaps? What is the role of machine learning-assisted mapping? Who is represented? For the first time we have made visible OSM's uneven spatial coverage on multiple scales and described how it evolved over time. The results of this dissertation could be read as an "overview map" which reveals how humanitarian mapping efforts of the past decade have already made OSM more inclusive, contributing to diversify and expand the spatial footprint of the areas mapped.

The work discovers that OSM data shows a much more spatially diverse pattern than previously considered, which is shaped by regional, socio-economic and demographic factors across several scales. The unequal patterns of mapping activity and completeness we have evidenced make it essential for analysts to assess the potential impact of missing data. For example, OSM data users should investigate if their intended analysis is subject to spatial bias caused by OSM's uneven building coverage. Researchers will be able to draw more robust conclusions and avoid misleading recommendations for decision makers once the biases in OSM's coverage are better known and can be accounted for.

The anticipated volume and velocity of corporate or machine learning-assisted mapping will pose challenges for individual volunteer mappers in respect to monitoring or validation (Anderson et al., 2019). To date, neither the global footprint nor the temporal evolution of AI-assisted mapping in OSM are sufficiently known. Our research about the global footprint of humanitarian mapping and its effects on map data quality provides a blueprint for further investigation and monitoring. Extending our methods will help to quantify to what extent machine learning-powered applications might be able to contribute towards filling the data gaps. Also we have shown that some humanitarian mapping campaigns were designed in a way that led to non-sustainable outcomes. These one-time mapping efforts might have helped to improve data completeness of buildings and roads, but failed to grow local mapping communities in the long term. Similar outcomes should be avoided with respect to machine-learning assisted mapping.

The results of this thesis in combination with use-case specific local information make it possible to map out where mapping (and what type of mapping Schott et al. (2021)) is needed. The need to overcome data scarcity in order to address urgent humanitarian needs and make progress towards the SDGs has been central to the framing of the "Missing Maps" project (Scholz et al., 2018). Nowadays, the integration of automation into the mapping workflow, e.g. through machine learning systems, is motivated from a similar perspective. In accordance with most studies, we have highlighted that these novel mapping approaches help mappers to create data more effectively. However, automated mapping methods also renew questions of unbalanced representation and might lead to the comeback of top-down

approaches in mapping. Spatially explicit machine learning models and GeoAI applications are seen as potential solutions for these challenges (Janowicz et al., 2020). Nevertheless, currently they still require the high quality datasets which are usually not available in all places. Furthermore, a better understanding of methodological limits in regard to reproducibility and explainability is needed to uncover the true potential of GeoAI to fill data gaps in OSM and other geographic datasets.

The research for this dissertation has been framed by essential technological developments and the evolution of major political frameworks, which have contributed towards the momentum around collaborative mapping in OSM in the past decade. As described at the outset for the case of Typhoon Haiyan/Yolanda in the Philippines in 2013, these developments have complemented each other and revolutionised the way map data is created for humanitarian aid and for many other domains. So let us think further and focus on a few grand challenges for OSM and related research in the upcoming decade. This dissertation outlines at least two potential starting points.

First, there is a need to integrate global and local perspectives about representation in OSM through interdisciplinary approaches. It has been outlined that it is not enough to describe the biased characteristics of map data or to simply assume that more data leads to better decisions and just transformations (Porto de Albuquerque et al., 2021). To overcome issues of representation, we argue that empowerment goes beyond involving local communities in the technical aspects of collecting map data. There is a need to investigate the concrete pathways to strengthen mapping communities, which enable them to change the observed unbalanced representation in OSM. In order to achieve this, researchers, humanitarian organisations and local OSM communities need to self-critically identify the structural barriers that exclude some social groups from participating in OSM mapping. This research can build upon the geospatial methods developed within this dissertation. However, shifting the viewpoint also means that we combine these with methodologies and theoretical frameworks from political science, anthropology and other social sciences. This will allow us to reframe the opening question of this thesis to account for the fact that true representation means more than being visible on the map (Bryant, 2021). Real inclusion requires an active local OSM community with the capacity to generate, maintain and improve geographic data that reflects local perspectives in the long term. Especially in the fast developing domain of applying artificial intelligence and machine learning methods to humanitarian mapping purposes, we see a critical lack of local representation and engagement by design.

Second, there is a need to bring closer together the research *about* OSM and the research conducted *with* OSM. GIScientists are in a good position to contribute the methodological approaches for this so that anybody, no matter whether they are located in Heidelberg, Jakarta or Kinshasa, can answer this important question: Is the current state of OSM data good enough for my analysis? The free availability of Landsat and Sentinel imagery and innovations in machine learning have spilled over to a multitude of research domains. In a similar way, OSM data could become a commodity for the global research community. There are already a few harmonized datasets combining information from OSM and other sources on a global or regional scale, e.g. for critical infrastructure (Nirandjan et al., 2022) or building stock characteristics (Milojevic-Dupont et al., 2023). Furthermore, OSM data is already an essential source of information for sustainable urban transitions, e.g. for urban energy system modelling (Schiefelbein et al., 2019). Hence, it will be necessary to conduct research at the conjunction between OSM data quality assessment and OSM-based applications. For example, this would allow us to distinguish uncertainties in the models from uncertainties in OSM data. This will further increase the reliability and usefulness of OSM data and derived datasets for global analyses such as monitoring progress towards the

SDGs (Hoek et al., 2021), as well as for humanitarian activities, including disaster risk reduction (Scholz et al., 2018).

Over the past 15 years OSM has been embraced by researchers in the domain of GIScience (Jokar Arsanjani et al., 2015). Building upon the achievements of this community we can now paint a bigger picture and need to discuss a shared vision of future OSM use cases and research. The academic track at OSM’s global “State of the Map” conference has evolved into a well-established platform which can facilitate these discussions (Grinberger et al., 2022a). However, currently “OSMScience” is still influenced to a large extent by perspectives of the Global North (Grinberger et al., 2022b). It is encouraging to see that also regional OSM conferences, such as the upcoming “State of the Map Africa” in Cameroon in December 2023, will host an academic track for the very first time. The YouthMappers network is another very successful example which shows the feasibility to shift the prevailing perspectives by nurturing self-organized groups of university students around the globe (Solís and Zeballos, 2023). Learning from this approach enables us to build an international network of undergraduates, PhD students and experienced researchers which overcomes barriers and borders. Such an international collaboration could not only help to grow OSM research hubs in the places currently under-represented, but might also fuel yet another revolution in the field of map making.

Let’s build the community for this together — so that map makers can make a difference.

Acknowledgements

Working on the dissertation has taken almost five years. Still time passed very quickly due to the broad support which I have received from my supervisors, colleagues and friends.

I would like to thank Prof. Dr. Alexander Zipf for the vast range of opportunities and freedom I had in conceiving and realising this work. Without your openness and support it would not have been possible for me to build such deep connections to other researchers, practitioners and OSM communities. I'm grateful that Prof. Dr. João Porto de Albuquerque has supported my work for almost a decade now, no matter whether you were located in Germany, Brazil or the UK. Your initial trust and ongoing mentorship has helped me not only to finish the dissertation, but also encouraged me to think even more critical about the results and what they mean for a wider audience.

My colleagues at HeiGIT and GIScience Research Group have supported this dissertation in various ways. In particular, I could improve my paper writing skills a lot based upon the precise comments and insightful discussions with Prof. Dr. Sven Lautenbach. Without my colleagues and our great student assistants I could not have performed any of the analyses nor make sense of the findings. Most importantly, I'm happy that we managed to create such a friendly and open-minded work environment. I really enjoy working together with you.

Moreover, I'm thankful for all the discussions and many hours I could spend with practitioners from humanitarian organisations and OSM community members from all around the world. This continuous exchange at the State of the Map, HOT Summit or GeONG conferences and online was the true fuel of this dissertation as you are the ones that actually make a change on the ground. Learning about all the different approaches towards mapping in OSM and motivations behind it has helped me to finish this dissertation and each individual paper. Beyond the scientific community, there were you as an important audience and critics to push my research further.

Finally, a big hug and thank you goes to my family and my friends in Heidelberg, Berlin and all the other places. While there have been a few stressful moments in the past five years caused by writing this PhD thesis, overall I had a really great time with so many memorable moments and incredible people. Spending time together with you just makes me happy.

Benjamin Herfort

Heidelberg, 6 May 2023

Bibliography

- Aitsi-Selmi, A., Murray, V., Wannous, C., Dickinson, C., Johnston, D., Kawasaki, A., Stevance, A. S., and Yeung, T. (2016). Reflections on a Science and Technology Agenda for 21st Century Disaster Risk Reduction: Based on the Scientific Content of the 2016 UNISDR Science and Technology Conference on the Implementation of the Sendai Framework for Disaster Risk Reduction 2015–20. *International Journal of Disaster Risk Science*, 7(1):1–29.
- Altay, N. and Labonte, M. (2014). Challenges in humanitarian information management and exchange: Evidence from Haiti. *Disasters*, 38(S1):50–72.
- Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16(7):16–07.
- Anderson, J., Sarkar, D., and Palen, L. (2019). Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 8(5):232.
- Attaran, A. (2005). An immeasurable crisis? A criticism of the millennium development goals and why they cannot be measured. *PLoS Medicine*, 2(10):0955–0961.
- Atwal, K. S., Anderson, T., Pfoser, D., and Züfle, A. (2022). Predicting building types using openstreetmap. *Scientific Reports*, 12.
- Barrington-Leigh, C. and Millard-Ball, A. (2017). The world’s user-generated road map is more than 80% complete. *PLoS ONE*, 12(8):1–20.
- Barron, C., Neis, P., and Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, 18(6):877–895.
- Bennett, M. M., Chen, J. K., Alvarez León, L. F., and Gleason, C. J. (2022). The politics of pixels: A review and agenda for critical remote sensing. *Progress in Human Geography*, 46(3):729–752.
- Bertolotto, M., McArdle, G., and Schoen-Phelan, B. (2020). Volunteered and crowdsourced geographic information: the OpenStreetMap project. *Journal of Spatial Information Science*, (20):65–70.
- Bhatia, A., Mahmud, A., Fuller, A., Shin, R., Rahman, A., Shatil, T., Sultana, M., Morshed, K. A., Leaning, J., and Balsari, S. (2018). The Rohingya in cox’s bazar: When the stateless seek refuge. *Health and Human Rights*, 20(2):105–122.
- Bhuyan, K., Westen, C. V., Wang, J., and Meena, S. R. (2022). Mapping and characterising buildings for flood exposure analysis using open-source data and artificial intelligence. *Natural Hazards*.

- Biljecki, F. and Chow, Y. S. (2022). Global building morphology indicators. *Computers, Environment and Urban Systems*, 95:101809.
- Biljecki, F., Chow, Y. S., and Lee, K. (2023). Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes. *Building and Environment*, page 110295.
- Bittner, C. (2017). Diversity in volunteered geographic information: comparing openstreetmap and wikimapia in jerusalem. *GeoJournal*, 82:887–906.
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1):2–16. Publisher: Elsevier B.V. ISBN: 0924-2716.
- Boeing, G. (2021). Spatial information and the legibility of urban form: Big data in urban morphology. *International Journal of Information Management*, 56(October 2019):102013.
- Braunschweig, K., Eberius, J., Thiele, M., and Lehner, W. (2020). The State of Open Humanitarian Data.
- Brovelli, M. and Zamboni, G. (2018). A New Method for the Assessment of Spatial Accuracy and Completeness of OpenStreetMap Building Footprints. *ISPRS International Journal of Geo-Information*, 7(8):289.
- Bryant, J. (2021). Digital mapping and inclusion in humanitarian response.
- Burns, R. (2014). Rethinking big data in digital humanitarianism: practices, epistemologies, and social relations. *GeoJournal*.
- Chen, J. and Zipf, A. (2019). DeepVGI: Deep learning with volunteered geographic information. *26th International World Wide Web Conference 2017, WWW 2017 Companion*, (1):771–772.
- Chun, Y. and Griffith, D. A. (2013). *Spatial statistics and geostatistics: theory and applications for geographic information science and technology*. Sage.
- Coetzee, S., Minghini, M., Solis, P., Rautenbach, V., and Green, C. (2018). Towards understanding the impact of Mapathons - Reflecting on youthmappers experiences. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4W8):35–42.
- Debray, H., Kraff, N. J., Zhu, X. X., and Taubenböck, H. (2023). Planned, unplanned, or in-between? a concept of the intensity of plannedness and its empirical relation to the built urban landscape across the globe. *Landscape and Urban Planning*, 233.
- Dittus, M., Quattrone, G., and Capra, L. (2016). Analysing Volunteer Engagement in Humanitarian Mapping: Building Contributor Communities at Large Scale. *Cscw '16*, (January):108–118.
- Dormann, C. F., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., R. Peres-Neto, P., Reineking, B., Schröder, B., M. Schurr, F., Wilson, R., Dormann, C. F., Mcpherson, J. M., Arau, M. B., Davies, R. G., Kissling, W. D., Ohlemu, R., Peres-Neto, P. R., and Schurr, F. M. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30(5):609–628.

- Facebook (2023). Open mapping at facebook. <https://github.com/facebookmicrosites/Open-Mapping-At-Facebook/wiki/Available-Countries>. Accessed: 2023-04-12.
- Feldmeyer, D., Nowak, W., Jamshed, A., and Birkmann, J. (2021). An open resilience index: Crowd-sourced indicators empirically developed from natural hazard and climatic event data. *Science of the Total Environment*, 774:145734.
- Fila, M. and Štampach, R. (2022). The use of artificial intelligence for humanitarian mapping.
- Fraisl, D., Campbell, J., See, L., Wehn, U., Wardlaw, J., Gold, M., Moorthy, I., Arias, R., Piera, J., Oliver, J. L., Masó, J., Penker, M., and Fritz, S. (2020). Mapping citizen science contributions to the UN sustainable development goals. *Sustainability Science*, 15(6):1735–1751. Publisher: Springer Japan ISBN: 0123456789.
- Fritz, S., See, L., Carlson, T., Haklay, M. M., Oliver, J. L., Fraisl, D., Mondardini, R., Brocklehurst, M., Shanley, L. A., Schade, S., Wehn, U., Abrate, T., Anstee, J., Arnold, S., Billot, M., Campbell, J., Espey, J., Gold, M., Hager, G., He, S., Hepburn, L., Hsu, A., Long, D., Masó, J., McCallum, I., Muniafu, M., Moorthy, I., Obersteiner, M., Parker, A. J., Weissplug, M., and West, S. (2019). Citizen science and the United Nations Sustainable Development Goals. *Nature Sustainability*, 2(10):922–930.
- Gardner, Z., Mooney, P., Sabbata, S. D., and Dowthwaite, L. (2020). Quantifying gendered participation in openstreetmap: responding to theories of female (under) representation in crowdsourced mapping. *GeoJournal*, 85:1603–1620.
- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Glass, L.-M. and Newig, J. (2019). Governance for achieving the Sustainable Development Goals: How important are participation, policy coherence, reflexivity, adaptation and democratic institutions? *Earth System Governance*, 2:100031.
- Goldblatt, R., Jones, N., and Mannix, J. (2020). Assessing openstreetmap completeness for management of natural disaster by means of remote sensing: A case study of three small island states (haiti, dominica and st. lucia). *Remote Sensing*, 12:118.
- Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221.
- Goodchild, M. F. (2015). Two decades on: Critical GIScience since 1993. *Canadian Geographer*, 59(1):3–11.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27.
- Graells-Garrido, E., Lalmas, M., and Menczer, F. (2015). First Women, Second Sex: Gender Bias in Wikipedia. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media - HT '15*, pages 165–174, Guzelyurt, Northern Cyprus. ACM Press.

- Graham, M. and Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3):255–261. ISBN: 2043820613.
- Graham, M., Straumann, R. K., and Hogan, B. (2015). Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia. *Annals of the Association of American Geographers*, 105(6):1158–1178.
- Griffith, D. A., Chun, Y., and Li, B. (2019). *Spatial Regression Analysis using Eigenvector Filtering*. CRC Press.
- Griffith, D. a. and Peres-Neto, P. R. (2006). Spatial Modelling in Ecology: The Flexibility of Eigenfunction Spatial Analysis. *Ecology*, 87(10):2603–2613.
- Grinberger, A. Y., Minghini, M., Juhász, L., Yeboah, G., and Mooney, P. (2022a). OSM Science—The Academic Study of the OpenStreetMap Project, Data, Contributors, Community, and Applications. *ISPRS International Journal of Geo-Information*, 11(4):230.
- Grinberger, A. Y., Minghini, M., Yeboah, G., Juhász, L., and Mooney, P. (2022b). Bridges and Barriers: An Exploration of Engagements of the Research Community with the OpenStreetMap Community. *ISPRS International Journal of Geo-Information*, 11(1):54.
- Grinberger, A. Y., Schott, M., Raifer, M., and Zipf, A. (2021). An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. *Transactions in GIS*, page tgis.12746.
- Gueguen, L., Koenig, J., Reeder, C., Barksdale, T., Saints, J., Stamatiou, K., Collins, J., and Johnston, C. (2017). Mapping Human Settlements and Population at Country Scale from VHR Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(2):524–538.
- Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, 37(4):682–703.
- Haklay, M. and Weber, P. (2008). OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*, 7(4):12–18.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160):850–853.
- Harig, O., Hecht, R., Burghardt, D., and Meinel, G. (2021). Automatic delineation of urban growth boundaries based on topographic data using germany as a case study. *ISPRS International Journal of Geo-Information*, 10:353.
- Harley, J. B. (1989). Deconstructing the map. *Cartographica The International Journal for Geographic Information and Geovisualization*, 26:1–20.
- Hecht, R., Kunze, C., and Hahmann, S. (2013). Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *ISPRS International Journal of Geo-Information*, 2(4):1066–1091.

- Heipke, C. (2010). Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6):550–557.
- Herfort, B. (2023). Dataset: Investigating completeness and inequalities in openstreetmap - spatio-temporal analysis of global urban building data.
- Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., and Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports*, 11(1).
- Herfort, B., Li, H., Fendrich, S., Lautenbach, S., and Zipf, A. (2019). Mapping human settlements with higher accuracy and less volunteer efforts by combining crowdsourcing and deep learning. *Remote Sensing*, 11(15).
- Hicks, A., Barclay, J., Chilvers, J., Armijos, M. T., Oven, K., Simmons, P., and Haklay, M. (2019). Global Mapping of Citizen Science Projects for Disaster Risk Reduction. *Frontiers in Earth Science*, 7(September):1–18.
- Hoek, J. V. D., Friedrich, H. K., Ballasiotes, A., Peters, L. E. R., and Wrathall, D. (2021). Development after displacement: Evaluating the utility of openstreetmap data for monitoring sustainable development goal progress in refugee settlements. *ISPRS International Journal of Geo-Information*, 10:153.
- Huck, J. J., Perkins, C., Haworth, B. T., Moro, E. B., and Nirmalan, M. (2020). Centaur vgi: A hybrid human–machine approach to address global inequalities in map coverage. *Annals of the American Association of Geographers*, 111:231–251.
- Humanitarian OpenStreetMap Team (2019). Mapping 2021: HOT’s Strategic Plan 2019 – 2021. Technical report.
- Humanitarian OpenStreetMap Team (2023). fair: Ai-assisted mapping. <https://www.hotosm.org/tech-blog/hot-tech-talks-fair/>. Accessed: 2023-04-07.
- Iliadis, A. and Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, 3(2):205395171667423.
- Janowicz, K., Gao, S., McKenzie, G., Hu, Y., and Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, 34(4):625–636.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301).
- Jiao, J. (2018). Machine Learning Assisted High-Definition Map Creation. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, pages 367–373, Tokyo, Japan. IEEE.
- Johnson, I., Lemmerich, F., Sáez-Trumper, D., West, R., Strohmaier, M., and Zia, L. (2021). Global Gender Differences in Wikipedia Readership. *Proceedings of the International AAAI Conference on Web and Social Media*, 15:254–265.

- Jokar Arsanjani, J., Zipf, A., Mooney, P., and Helbich, M., editors (2015). *OpenStreetMap in GIScience*. Springer International Publishing, Cham. Series Title: Lecture Notes in Geoinformation and Cartography Publication Title: OpenStreetMap in GIScience: Experiences, Research, Applications Issue: JANUARY.
- Kampffmeyer, M., Salberg, A. B., and Jenssen, R. (2016). Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 680–688.
- Kathmandu Living Lab (2015). Nepal earthquake kathmandu living lab.
- Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3):262–267.
- Klotz, M., Kemper, T., Geiß, C., Esch, T., and Taubenböck, H. (2016). How good is the map? A multi-scale cross-comparison framework for global settlement layers: Evidence from Central Europe. *Remote Sensing of Environment*, 178:191–212.
- Kraak, M. J., Roth, R. E., Ricker, B., Kagawa, A., and Le Sourd, G., editors (2020). *Mapping for a sustainable world*. United Nations, New York. OCLC: 1269408111.
- Kuffer, M., Thomson, D. R., Boo, G., Mahabir, R., Grippa, T., Vanhuyse, S., Engstrom, R., Ndugwa, R., Makau, J., Darin, E., de Albuquerque, J. P., and Kabaria, C. (2020). The role of earth observation in an integrated deprived area mapping “system” for low-to-middle income countries. *Remote Sensing*, 12:982.
- Kwan, M.-P. (2016). Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge. *Annals of the American Association of Geographers*, 106(2):274–282.
- Li, H., Herfort, B., Huang, W., Zia, M., and Zipf, A. (2020). Exploration of OpenStreetMap missing built-up areas using twitter hierarchical clustering and deep learning in Mozambique. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166(November 2019):41–51.
- Li, H., Herfort, B., Lautenbach, S., Chen, J., and Zipf, A. (2022). Improving OpenStreetMap missing building detection using few-shot transfer learning in sub-Saharan Africa. *Transactions in GIS*, (March):1–22.
- Logar, T., Bullock, J., Nemni, E., Bromley, L., Quinn, J. A., and Luengo-Oroz, M. (2020). PulseSatellite: A Tool Using Human-AI Feedback Loops for Satellite Image Analysis in Humanitarian Contexts. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(09):13628–13629.
- Ma, D., Fan, H., Li, W., and Ding, X. (2019). The State of Mapillary: An Exploratory Analysis. *ISPRS International Journal of Geo-Information*, 9(1):10.
- MapBox (2023). Osm cha.
- Marco Minghini, Serena Coetsee, Levente Juhasz, Godwin Yeboah, Peter Mooney, and A. Yair Grinberger (2020). Editorial: OpenStreetMap research in the COVID-19 era. *Proceedings of the Academic Track at the State of the Map 2020*, pages 1–4.

- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54.
- Meta (2022). Map with ai. <https://mapwith.ai>. Accessed: 2022-06-01.
- Meyer, H. and Pebesma, E. (2022). Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications*, 13(1):2208.
- Microsoft (2023a). Global ml building footprints. <https://github.com/microsoft/GlobalMLBuildingFootprints>. Accessed: 2023-04-12.
- Microsoft (2023b). Road detections. <https://github.com/microsoft/RoadDetections>. Accessed: 2023-04-12.
- Miller, H. J. and Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, 80(4):449–461. Publisher: Springer Netherlands ISBN: 1070801496.
- Milojevic-Dupont, N., Hans, N., Kaack, L. H., Zumwald, M., Andrieux, F., de Barros Soares, D., Lohrey, S., Pichler, P. P., and Creutzig, F. (2020). Learning from urban form to predict building heights. *PLoS ONE*, 15(12 December):1–22.
- Milojevic-Dupont, N., Wagner, F., Nachtigall, F., Hu, J., Brüser, G. B., Zumwald, M., Biljecki, F., Heeren, N., Kaack, L. H., Pichler, P.-P., and Creutzig, F. (2023). EUBUCCO v0.1: European building stock characteristics in a common and open database for 200+ million individual buildings. *Scientific Data*, 10(1):147.
- Mocnik, F.-B., Zipf, A., and Raifer, M. (2017). The OpenStreetMap folksonomy and its evolution. *Geo-spatial Information Science*, 20(3):219–230. Publisher: Taylor & Francis.
- Moturi, A. K., Suiyanka, L., Mumo, E., Snow, R. W., Okiro, E. A., and Macharia, P. M. (2022). Geographic accessibility to public and private health facilities in kenya in 2021: An updated geocoded inventory and spatial analysis. *Frontiers in Public Health*, 10.
- Mulder, F., Ferguson, J., Groenewegen, P., Boersma, K., and Wolbers, J. (2016). Questioning Big Data: Crowdsourcing crisis data towards an inclusive humanitarian response. *Big Data and Society*, 3(2). Publisher: SAGE Publications Ltd.
- Murray, J., Sargent, I., Holland, D., Gardiner, A., Dionysopoulou, K., Coupland, S., Hare, J., and Atkinson, P. (2019). Opportunities for machine learning and artificial intelligence in a national mapping agency: a perspective on enhancing ordnance survey workflow. volume XXIV. International Society for Photogrammetry and Remote Sensing.
- Neis, P., Singler, P., and Zipf, A. (2010). Collaborative mapping and Emergency Routing for Disaster Logistics-Case studies from the Haiti earthquake and the UN portal for Afrika. *Geoinformatics Forum Salzburg*, (October):239–248.
- Neis, P. and Zielstra, D. (2014). Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap. *Future Internet*, 6(1):76–106. ISBN: 1999-5903.
- Neis, P., Zielstra, D., and Zipf, A. (2011). The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. *Future Internet*, 4(1):1–21.

- Neis, P., Zielstra, D., and Zipf, A. (2013). Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet*, 5(2):282–300.
- Neis, P. and Zipf, A. (2012). Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 1(3):146–165.
- Nievas, C. I., Pilz, M., Prehn, K., Schorlemmer, D., Weatherill, G., and Cotton, F. (2022). Calculating earthquake damage building by building: the case of the city of cologne, germany. *Bulletin of Earthquake Engineering*, 20:1519–1565.
- Nirandjan, S., Koks, E. E., Ward, P. J., and Aerts, J. C. (2022). A spatially-explicit harmonized global dataset of critical infrastructure. *Scientific Data*, 9(1):1–13.
- Oort, P. (2006). *Spatial data quality: from description to application*. PhD thesis, Wageningen Universiteit.
- OpenStreetMap Philippines community (2020). A Call to Correct Narratives about Geospatial Work in the Philippines.
- OSM Lab (2023). Show me the way.
- Palen, L., Soden, R., Anderson, T. J., and Barrenechea, M. (2015). Success & Scale in a Data - Producing Organization : The Socio - Technical Evolution of OpenStreetMap in Response to Humanitarian Events. *Chi 2015*, pages 4113–4122.
- Panek, J. and Sobotova, L. (2015). Community mapping in urban informal settlements: Examples from Nairobi, Kenya. *Electronic Journal of Information Systems in Developing Countries*, 68:1–13.
- Pavlovskaya, M. and Martin, K. S. (2007). Feminism and geographic information systems: From a missing object to a mapping subject. *Geography Compass*, 1:583–606.
- Poiani, T. H., Rocha, R. D. S., Degrossi, L. C., and Porto de Albuquerque, J. (2016). Potential of collaborative mapping for disaster relief: A case study of openstreetmap in the Nepal earthquake 2015. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, volume 2016-March, pages 188–197.
- Porto de Albuquerque, J., Anderson, L., Calvillo, N., Coaffee, J., Cunha, M. A., Degrossi, L. C., Dolif, G., Horita, F., Klonner, C., Lima-Silva, F., Marchezini, V., Martins, M. H. d. M., Pajarito-Grajales, D., Pitidis, V., Rudorff, C., Tkacz, N., Traijber, R., and Zipf, A. (2021). The role of data in transformations to sustainability: a critical research agenda. *Current Opinion in Environmental Sustainability*, 49(August 2020):153–163.
- Porto de Albuquerque, J., Herfort, B., and Eckle, M. (2016a). The Tasks of the Crowd: A Typology of Tasks in Geographic Information Crowdsourcing and a Case Study in Humanitarian Mapping. *Remote Sensing*, 8(10):859.
- Porto de Albuquerque, J., Herfort, B., and Eckle, M. (2016b). The tasks of the crowd: A typology of tasks in geographic information crowdsourcing and a case study in humanitarian mapping. *Remote Sensing*, 8(10).

- Quinn, J. A., Nyhan, M. M., Navarro, C., Coluccia, D., Bromley, L., and Luengo-Oroz, M. (2018). Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128):20170363.
- Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.-B., and Zipf, A. (2019). OSHDB: a framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards*, 4(1).
- Reckien, D., Creutzig, F., Fernandez, B., Lwasa, S., Tovar-Restrepo, M., Mcevoy, D., and Satterthwaite, D. (2017). Climate change, equity and the Sustainable Development Goals: an urban perspective. *Environment and Urbanization*, 29(1):159–182.
- Redi, M., Gerlach, M., Johnson, I., Morgan, J., and Zia, L. (2020). A taxonomy of knowledge gaps for wikimedia projects (second draft).
- Sachs, J., Lafortune, G., Kroll, C., Fuller, G., and Woelm, F. (2022). Sustainable Development Report. Technical report.
- Sachs, J. D., Schmidt-Traub, G., Mazzucato, M., Messner, D., Nakicenovic, N., and Rockström, J. (2019). Six Transformations to achieve the Sustainable Development Goals. *Nature Sustainability*, 2(9):805–814.
- Scheider, S., Ballatore, A., and Lemmens, R. (2019). Finding and sharing GIS methods based on the questions they answer. *International Journal of Digital Earth*, 12(5):594–613.
- Schiefelbein, J., Rudnick, J., Scholl, A., Remmen, P., Fuchs, M., and Müller, D. (2019). Automated urban energy system modeling and thermal building simulation based on OpenStreetMap data sets. *Building and Environment*, 149:630–639.
- Scholz, S., Knight, P., Eckle, M., Marx, S., and Zipf, A. (2018). Volunteered Geographic Information for Disaster Risk Reduction—The Missing Maps Approach and Its Potential within the Red Cross and Red Crescent Movement. *Remote Sensing*, 10(8):1239.
- Schott, M., Grinberger, A. Y., Lautenbach, S., and Zipf, A. (2021). The impact of community happenings in openstreetmap-establishing a framework for online community member activity analyses. *ISPRS International Journal of Geo-Information*, 10.
- Schröder-Bergen, S., Glasze, G., Michel, B., and Dammann, F. (2021). De/colonizing openstreetmap? local mappers, humanitarian and commercial actors and the changing modes of collaborative mapping. *GeoJournal*, 7.
- See, L., Georgieva, I., Duerauer, M., Kemper, T., Corbane, C., Maffenini, L., Gallego, J., Pesaresi, M., Sirbu, F., Ahmed, R., Blyshchyk, K., Magori, B., Blyshchyk, V., Melnyk, O., Zadorozhniuk, R., Mandici, M. T., Su, Y. F., Rabia, A. H., Pérez-Hoyos, A., Vasylyshyn, R., Pawe, C. K., Bilous, S., Kovalevskyi, S. B., Kovalevskyi, S. S., Bordoloi, K., Bilous, A., Panging, K., Bilous, V., Prestele, R., Sahariah, D., Deka, A., Nath, N., Neves, R., Myroniuk, V., Karner, M., and Fritz, S. (2022). A crowdsourced global data set for validating built-up surface layers. *Scientific Data*, 9(1). Publisher: Nature Research.

- Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., and Haklay, M. M. (2016). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 8816(June):1–29.
- Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M., and Quinn, J. (2021). Continental-Scale Building Detection from High Resolution Satellite Imagery. pages 1–15. arXiv: 2107.12283.
- Soden, R. and Palen, L. (2016). Infrastructure in the wild: What mapping in post-earthquake Nepal reveals about infrastructural emergence. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 2796–2807.
- Solís, P., Anderson, J., and Rajagopalan, S. (2021). Open geospatial tools for humanitarian data creation, analysis, and learning through the global lens of YouthMappers. *Journal of Geographical Systems*, 23(4):599–625. Publisher: Springer Science and Business Media Deutschland GmbH.
- Solís, P. and Zeballos, M., editors (2023). *Open Mapping towards Sustainable Development Goals*. Springer International Publishing, Cham. Series Title: Sustainable Development Goals Series.
- Specht, D. (2020). *Mapping Crisis: Participation, Datafication and Humanitarianism in the Age of Digital Mapping*. University of London Press.
- Sturrock, H. J., Woolheater, K., Bennett, A. F., Andrade-Pacheco, R., and Midekisa, A. (2018). Predicting residential structures from open source remotely enumerated data using machine learning. *PLoS ONE*, 13(9):1–10.
- Sui, D., Goodchild, M., and Elwood, S. (2013). Volunteered geographic information, the exaflood, and the growing digital divide. In *Crowdsourcing geographic knowledge*, pages 1–12. Springer.
- Sustainable Development Solutions Network (2015). Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a data revolution for the SDGs. *A report by the Leadership Council of the Sustainable Development Solutions Network*, (Version 7):160.
- Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data*, 4:2–5.
- Tracewski, L., Bastin, L., and Fonte, C. C. (2017). Repurposing a deep learning network to filter and classify volunteered photographs for land cover and land use characterization. *Geo-spatial Information Science*, 20(3):252–268.
- Ulbrich, P., Porto De Albuquerque, J., and Coaffee, J. (2019). The impact of urban inequalities on monitoring progress towards the sustainable development goals: Methodological considerations. *ISPRS International Journal of Geo-Information*, 8(1):1–18.
- UN OCHA Centre for Humanitarian Data (2023). The State of Open Humanitarian Data 2023. Technical report.
- United Nations (2014). A World that Counts: Mobilising the Data Revolution for Sustainable Development. *UN Data Revolution Report*.
- United Nations (2020a). Data Strategy of the Secretary-General for Action by Everyone, Everywhere. Technical report, United Nations.

- United Nations (2020b). The Sustainable Development Goals Report 2020. Technical report.
- United Nations Office for Disaster Risk Reduction (2015). Sendai Framework for Disaster Risk Reduction 2015-2030. Technical report.
- Vakalopoulou, M., Karantzas, K., Komodakis, N., and Paragios, N. (2015). Building detection in very high resolution multispectral data with deep learning features. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2015-Novem:1873–1876.
- van der Vaart, W. B. V., Lambers, K., Kowalczyk, W., and Bourgeois, Q. P. (2020). Combining deep learning and location-based ranking for large-scale archaeological prospection of lidar data from the netherlands. *ISPRS International Journal of Geo-Information*, 9:293.
- Vargas-Muñoz, J. E., Lobry, S., Falcão, A. X., and Tuia, D. (2019). Correcting rural building annotations in OpenStreetMap using convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147(November 2018):283–293.
- Vargas-Muñoz, J. E., Srivastava, S., Tuia, D., and Falcao, A. X. (2021). Openstreetmap: Challenges and opportunities in machine learning and remote sensing. *IEEE Geoscience and Remote Sensing Magazine*, 9:184–199.
- Vargas-Muñoz, J. E., Tuia, D., and Falcão, A. X. (2021). Deploying machine learning to assist digital humanitarians: making image annotation in openstreetmap more efficient. *International Journal of Geographical Information Science*, 35:1725–1745.
- Veselovsky, V., Sarkar, D., Anderson, J., and Soden, R. (2022). An Automated Approach to Identifying Corporate Editing. *Proceedings of the International AAAI Conference on Web and Social Media*, 16:1052–1063.
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., and Nerini, F. F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, 11:233.
- Watkinson, K., Huck, J. J., and Harris, A. (2022). Centaur VGI: An Evaluation of Engagement, Speed, and Quality in Hybrid Humanitarian Mapping. *Annals of the American Association of Geographers*, 112(8):2373–2392.
- Watkinson, K., Huck, J. J., and Harris, A. (2023). Using gamification to increase map data production during humanitarian volunteered geographic information (vgi) campaigns. *Cartography and Geographic Information Science*.
- Witt, R., Loos, L., and Zipf, A. (2021). Analysing the impact of large data imports in openstreetmap. *ISPRS International Journal of Geo-Information*, 10.
- World Bank (2018). *Information and Communications for Development 2018: Data-Driven Development*. Washington, DC: World Bank.
- Yeboah, G., Porto de Albuquerque, J., Troilo, R., Tregonning, G., Perera, S., Shifat Ahmed, S. A., Ajisola, M., Alam, O., Aujla, N., Azam, S. I., Azeem, K., Bakibinga, P., Chen, Y. F., Choudhury, N. N., Diggle, P. J., Fayehun, O., Gill, P., Griffiths, F., Harris, B., Iqbal, R., Kabaria, C., Ziraba, A. K.,

- Khan, A. Z., Kibe, P., Kisia, L., Kyobutungi, C., Lilford, R. J., Madan, J. J., Mbaya, N., Mberu, B., Mohamed, S. F., Muir, H., Nazish, A., Njeri, A., Odubanjo, O., Omigbodun, A., Osuh, M. E., Owoaje, E., Oyeboode, O., Pitidis, V., Rahman, O., Rizvi, N., Sartori, J., Smith, S., Taiwo, O. J., Ulbrich, P., Uthman, O. A., Watson, S. I., Wilson, R., and Yusuf, R. (2021). Analysis of openstreetmap data quality at different stages of a participatory mapping process: Evidence from slums in Africa and Asia. *ISPRS International Journal of Geo-Information*, 10(4).
- Yi, Z., Zurutuza, N., Bollinger, D., Garcia-Herranz, M., and Kim, D. (2019). Towards equitable access to information and opportunity for all: mapping schools with high-resolution satellite imagery and machine learning.
- Yi, Z.-F., Frederick, H., Mendoza, R. L., Avery, R., and Goodman, L. (2021). Ai mapping risks to wildlife in tanzania: Rapid scanning aerial images to flag the changing frontier of human-wildlife proximity. volume 2021-July, pages 5299–5302. IEEE.
- Yuan, J., Roy Chowdhury, P. K., McKee, J., Yang, H. L., Weaver, J., and Bhaduri, B. (2018). Exploiting deep learning and volunteered geographic information for mapping buildings in Kano, Nigeria. *Scientific Data*, 5:180217.
- Zhang, Y., Zhou, Q., Brovelli, M. A., and Li, W. (2022). Assessing osm building completeness using population data. *International Journal of Geographical Information Science*, 36:1443–1466.
- Zhou, Q., Wang, S., and Liu, Y. (2022a). Exploring the accuracy and completeness patterns of global land-cover/land-use data in openstreetmap. *Applied Geography*, 145:102742.
- Zhou, Q., Zhang, Y., Chang, K., and Brovelli, M. A. (2022b). Assessing osm building completeness for almost 13,000 cities globally. *International Journal of Digital Earth*, 15:2400–2421.
- Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., and Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36.
- Zielstra, D., Hochmair, H. H., and Neis, P. (2013). Assessing the effect of data imports on the completeness of openstreetmap - a united states case study. *Transactions in GIS*, 17:315–334.
- Zook, M., Graham, M., Shelton, T., and Gorman, S. (2010). Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. *World Medical & Health Policy*, 2(2):6–32.
- Zou, Z., Chen, K., Shi, Z., Guo, Y., and Ye, J. (2023). Object Detection in 20 Years: A Survey. *Proceedings of the IEEE*, 111(3):257–276.

Part II

Publications

The Evolution of Humanitarian Mapping within the OpenStreetMap Community

Benjamin Herfort, Sven Lautenbach, João Porto de Albuquerque,
Jennings Anderson, Alexander Zipf

04 February 2021

Abstract

In the past 10 years, the collaborative maps of OpenStreetMap (OSM) have been used to support humanitarian efforts around the world as well as to fill important data gaps for implementing major development frameworks such as the Sustainable Development Goals (SDGs). This paper provides a comprehensive assessment of the evolution of humanitarian mapping within the OSM community, seeking to understand the spatial and temporal footprint of these large-scale mapping efforts. The spatio-temporal statistical analysis of OSM's full history since 2008 showed that humanitarian mapping efforts added 60.5 million buildings and 4.5 million roads to the map. Overall, mapping in OSM was strongly biased towards regions with very high Human Development Index. However, humanitarian mapping efforts had a different footprint, predominantly focused on regions with medium and low human development. Despite these efforts, regions with low and medium human development only accounted for 28% of the buildings and 16% of the roads mapped in OSM although they were home to 46% of the global population. Our results highlight the formidable impact of humanitarian mapping efforts such as post-disaster mapping campaigns to improve the spatial coverage of existing open geographic data and maps, but they also reveal the need to address the remaining stark data inequalities, which vary significantly across countries. We conclude with three recommendations directed at the humanitarian mapping community: (1) Improve methods to monitor mapping activity and identify where mapping is needed. (2) Rethink the design of projects which include humanitarian data generation to avoid non-sustainable outcomes. (3) Remove structural barriers to empower local communities and develop capacity.

Reference— Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., and Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-82404-z>

1 Introduction

Drawing lessons from what has been called the "immeasurability" of the Millenium Development Goals (MDGs) (Attaran, 2005), much attention has been given towards indicators and frameworks to monitor the United Nation (UN) Sustainable Development Goals (SDGs) (United Nations, 2015). Early calls for a "Data Revolution for Sustainable Development" identified the need for new institutions, actors, ideas and partnerships to successfully monitor progress towards the SDGs (United Nations, 2014). The Sustainable Development Solutions Network (2015) has emphasized the role of the often-neglected geo-spatial and geo-referenced data to disaggregate information for almost every proposed indicator. Accordingly, new sources and partnerships are critical to fill data gaps for existing and yet to be developed indicators and might provide a complementary way to tackle the estimated gap of 1 billion US\$ in funding for national statistical offices (Sustainable Development Solutions Network, 2015). Consequently, it has

been identified that baseline geospatial data that should be provided by national agencies are often not accessible, not up-to-date or not available in standard format (Braunschweig et al., 2020; Altay and Labonte, 2014). Additionally, it is evident that effective humanitarian action and decision making rely on accurate, up-to-date spatial information (Braunschweig et al., 2020). More recently, the UN's Sustainable Development Goals Report 2020 surveys the devastating impacts of the COVID-19 pandemic for the SDGs and highlights the increased importance of "data innovations" to support sustainability goals by closing important gaps in data for monitoring and reporting on the SDG targets and indicators (United Nations, 2020).

It has been argued that several actions to achieve the SDGs (such as education, community engagement and community based problem solving, baseline research, planning and strategy development, allocation and coordination and improvement of public and private programs and services) can be informed by citizen generated data sets (Lämmerhirt et al., 2019). OpenStreetMap (OSM), also known as the Wikipedia of maps, constitutes a new open geographic database and a community mobilization initiative with potential to serve as a source of information for the called-for "data revolution". The potential of OSM to harness the power of new information and communication technologies in support of disaster risk reduction has been recognized by several authors (Aitsi-Selmi et al., 2016; Scholz et al., 2018; Neis et al., 2010). The past decade has seen the emergence of an ecosystem composed of volunteer mapping communities, corporations, governmental and humanitarian organizations which contribute to and use the open geographic database of OSM for various purposes. In this paper, we investigate the humanitarian mapping activity in OSM, predominantly organized through the web tool "HOT Tasking Manager". While initially the term 'disaster mapping' has been introduced to describe the "ability for volunteers to assist in disaster response situations via mapping and other spatial analysis" (Zook et al., 2010), since 2010 the broader term 'humanitarian mapping' has evolved. It refers to collaborative mapping in OSM (and other platforms such as MapSwipe or Ushahidi) for both humanitarian relief responses and humanitarian purposes in general (Palen et al., 2015; Dittus et al., 2016; Albuquerque et al., 2016). This activity is also referred to as remote mapping or digitization and consists of the generation of geographic data based on satellite imagery (Albuquerque et al., 2016). Since contributors do not need to physically be at the place of mapping, the efforts are collaborative and distributed to volunteers around the world. The HOT Tasking Manager (<https://tasks.hotosm.org/>) has been used to coordinate the mapping efforts of thousands of volunteers by various organisations such as the Humanitarian OpenStreetMap Team (HOT), American Red Cross (ARC), and Médecins Sans Frontières (MSF), as well as collaborative projects such as Missing Maps (<https://www.missingmaps.org/>) or YouthMappers (<https://www.youthmappers.org/>) that particularly align their efforts to the SDGs. Figure 1 provides context and motivation for our research by showing a chronology of key events in humanitarian mapping since 2010, the year HOT was founded in the aftermath of the devastating Earthquake in Haiti (Zook et al., 2010; Palen et al., 2015). It shows the immediate impact that post-disaster community "activations" have had on the contributions to OSM. Besides these punctual events, there are several additional factors with more long-term effects such as the introduction of new tools, the availability of open satellite imagery for mapping in OSM and the number of organizations involved. Furthermore, political frameworks set the policy agenda and form the institutional space in which local communities, national and international NGOs operate.

Previous work that analyzed humanitarian mapping in OSM has mostly focused on single events or relatively short time periods. For instance, there are detailed analyses of the mapping activity in OSM after the 2013 Typhoon Yolanda (Palen et al., 2015; Dittus et al., 2016), the 2014 West Africa Ebola response (Dittus et al., 2016), and the 2015 Nepal earthquake (Poiani et al., 2016). It has been shown that humanitarian remote mapping has undergone a series of changes in a socio-technological evolution between 2010 and 2013 (Palen et al., 2015). The 2015 Nepal earthquake has been identified as a turning point in the adoption of OSM by a wide range of organizations, leading to open-data practices becoming more common in humanitarian work and to the constitution of a community of practice (Soden and Palen, 2016). The Missing Maps project created in 2014 has shifted the focus of large-scale humanitarian mapping from disaster response to preparedness and preventative efforts (Dittus et al., 2016; Scholz et al., 2018). More recently, new actors have joined and existing mapping tools (such as the HOT Tasking Manager) are continuously improving with advanced technological approaches such as deep learning/artificial intelligence, which are now finding their way into humanitarian mapping workflows (Herfort et al., 2019; Li et al., 2020). For a more detailed review of the scientific literature on this topic please refer to Haworth et al. (2018).

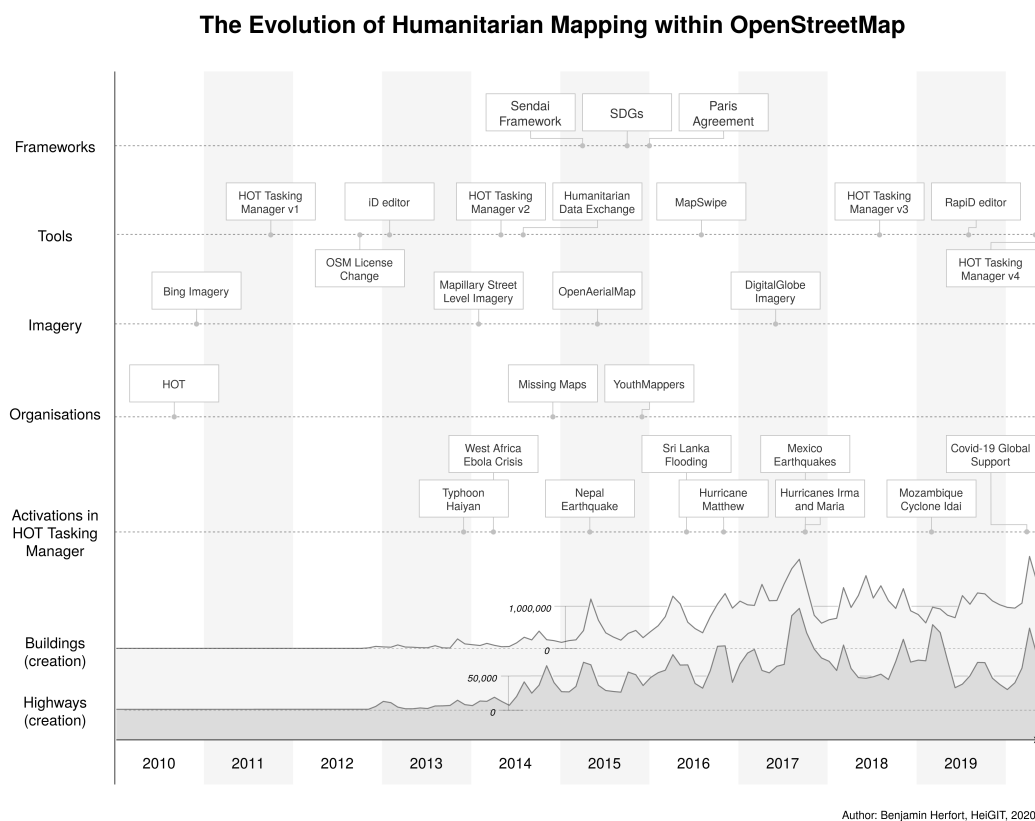


Figure 1: Sketch of the evolution of humanitarian mapping in OSM in regard to major disaster activations, the socio-technical development of the community and global political frameworks. The plots at the bottom show the number of buildings and highways added to OSM. Created using drawio 13.11.0 (<https://github.com/jgraph/drawio>).

A key underlying motivation behind the rich and multifaceted efforts of OSM humanitarian mapping has been to widen the availability of its geographic database by creating the "missing maps" needed so "everyone is counted, map data is accessible and used, everyone can engage and contribute to the map." (Humanitarian OpenStreetMap Team, 2019). This builds upon the acknowledgment that the OSM data in general is strongly biased, in part due to a much larger contributor basis in countries in the global North as a consequence of socio-economic inequalities and the digital divide (Sui et al., 2013). However, the question arises to what extent humanitarian mapping efforts have been able to widen the footprint of OSM to overcome the data gaps caused by socio-economic inequalities 10 years after what could be considered the first large-scale humanitarian mapping event in the aftermath of the 2010 Haiti Earthquake. How much has humanitarian mapping contributed to making the world-wide collaborative map more inclusive to support the overarching SDG goal of "leaving no one behind"?

In order to tackle these questions, this paper presents a comprehensive assessment of the evolution of humanitarian mapping in OSM. This analysis encompasses all humanitarian mapping projects organized through the HOT Tasking Manager since 2012 (start of the available data), enabling us — for the first time — to cast a longitudinal perspective on the intersecting effects of mapping efforts, socio-economic, and demographic characteristics. This analysis provides critical insights into the achievements of humanitarian mapping so far, by combining data about OSM humanitarian mapping with statistics on global population distribution and human development at the sub-national level. Only very recent technological advances to utilize OSM's full-history data for spatio-temporal data analysis on the global scale (Raifer et al., 2019) allow us to provide this novel perspective in our research. This spatio-temporal analysis is performed in the pursuit of the following research questions:

1. What is the effect of humanitarian mapping in the past decade on the global footprint of OSM in terms of socio-economic and demographic profile of the areas mapped?
2. How did OSM humanitarian mapping activity evolve over time on a global scale and in individual countries?

The remainder of this paper is organized as follows: The next section presents the results of our analysis of spatial footprint, temporal evolution and socio-economic characteristics of humanitarian mapping in OSM vis-à-vis general OSM mapping. This is followed by a discussion of our findings. The methodology is explained in the subsequent section and contains information on the data sets utilized and the statistical models applied.

2 Results

2.1 Scale and Spatial Footprint

Our results provide insight on the scale and spatial footprint of general mapping activity in OSM, which constitutes the baseline to evaluate humanitarian mapping efforts. As depicted in Table 1, all mappers (this term is often used to refer to the volunteers that edit the OSM database) have added 426.4 million buildings and 175.7 million highways in OSM between January 2008 and May 2020. Humanitarian mapping activities accounted for 60.5 million buildings (~14%) and 4.5 million highways (~3%). For buildings, changes to existing data in OSM, e.g. changing the shape of a building (geometry change) or changing its descriptive attributes (tag change) were observed less often in total and the proportion of humanitarian mapping varied between 4% - 12%. Highway mapping in OSM was characterized by frequent changes to existing data (geometry changes, tag changes)..

Mapping of buildings in OSM predominantly happened in Europe, North America, and Japan (see Fig. 2). Furthermore, there has been considerable building mapping activity in Nepal, East Africa, the Philippines, and Indonesia (Fig. 2 (a)). These were largely contributed through humanitarian mapping activities (Fig. 2 (b)). Besides this, extensive humanitarian mapping activity were observed in Central and Western Africa, whereas activity was much more spatially scattered for Central and Southern America and Central Asia.

For highway mapping in OSM (see Fig. 2 (c)), overall activity happened on a truly global scale in all inhabited parts of the world. Compared to the global reach of OSM mapping in general, humanitarian

	building contributions [millions]		highway contributions [millions]	
	all	humanitarian	all	humanitarian
added	426.4	60.5	175.7	4.5
tag change	91.3	3.6	151.9	0.5
geometry change	75.9	9.4	276.4	3.3
deleted	32.6	3.2	23.2	0.3

Table 1: Cumulative number of added, changed and deleted buildings and highways in OSM between 2008/01/01 - 2020/05/20 in regard to overall and humanitarian mapping activities.

highway mapping activities (see Fig. 2 (d)) focused on similar areas as the humanitarian building mapping activities and were extensive primarily in parts of Eastern and Western Africa. This map illustrates the observation that humanitarian highway mapping constitutes only a small fraction of overall highway mapping not only in terms of absolute numbers but also regarding its spatial distribution. Socio-economic data on the Sub-national Human Development Index (SHDI) and population distribution provide important variables to assess the scale and spatial footprint of overall and humanitarian mapping in OSM (see Fig. 2 (e) and (f)). This is explored in the next section.

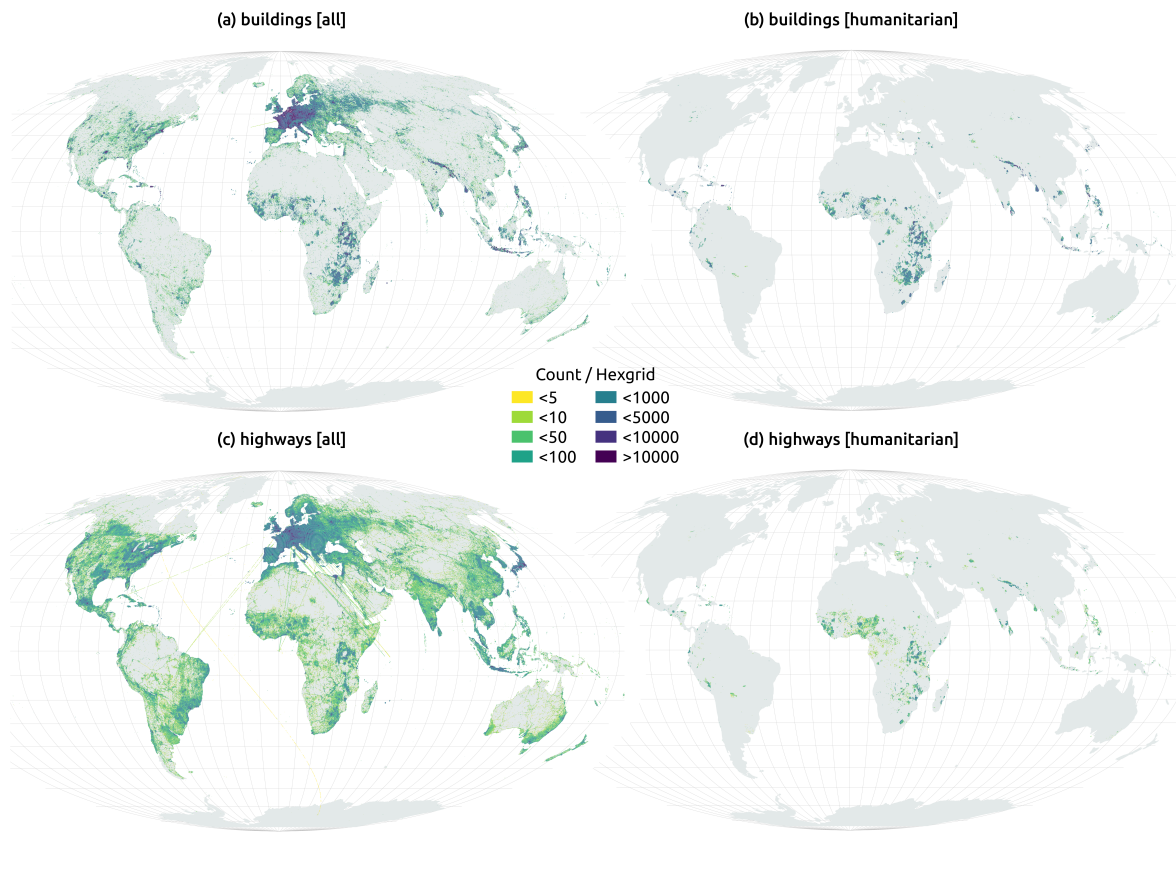
2.2 Socio-Economic and Demographic Characteristics

There was a sharp contrast between population distribution per SHDI (a) and mapping in OSM per SHDI for buildings (c) and highways (d) when considering the subnational level (Fig. 3). Whereas 20% of all humans lived in regions of very high human development, these regions accounted only for 60% of buildings and 65% of highways created in OSM. Regions with medium human development made up for 15% of added buildings and 11% of added highways in OSM in overall, but represented only ~36% of the global population. In regions with low human development we counted 10% of the global population, 11% of added buildings and 5% of added highways in OSM. Humanitarian mapping activities have been organized mainly in regions with an SHDI lower than 0.7. For humanitarian mapping, 83% of building creations and 92% of highway creations fell into this category.

The distinction between overall and humanitarian mapping activities has been underlined by the results of the Generalized Linear Model (GLM) regression both for buildings and in highways (Table 2). These results have been computed at the subnational level. In regard to buildings 12 respectively 8 spatial eigenvectors for overall and humanitarian contributions were included in the model, which underlines the complexity of the relationship. For highways the number of selected spatial eigenvectors were 16 for overall and 7 for humanitarian contributions. Spatial eigenvectors are additionally generated co-variables that account for the effect of spatial autocorrelation in the data (Griffith and Peres-Neto, 2006). They are based on the spatial weight matrix that defines neighborhood relationships between the observations. Without considering spatial autocorrelation coefficients were of same sign and order of magnitude (Table 2). The results revealed for both overall and humanitarian mapping an expected significant positive relationship between population (log scaled) and observed mapping in OSM for buildings and highways. Similarly, more buildings and highways have been added to OSM in regions that were impacted stronger by disasters (measured by the number of deaths caused by natural hazards per 100,000 inhabitants) for both overall and humanitarian mapping. The effect of the number of YouthMappers chapters was significant for the number of buildings but not for the number of highways for both overall and humanitarian mapping. Even for the number of buildings, the effect of the number of YouthMappers chapters was much weaker than for all other predictors.

The effect of SHDI on mapping activity differed clearly between overall OSM and humanitarian mapping activities: Considering overall mapping in OSM, we found a strong positive effect (more mapping in higher SHDI regions). For humanitarian mapping however, the results showed a negative effect (more mapping in lower SHDI regions). Similar, but less pronounced was the difference in regard to gender equality. Overall more buildings and highways have been mapped in regions with higher gender equality. Hence, gender inequality seemed to be a factor related to the prevention of mapping. But for humanitarian

Buildings and highways added to OSM between 2008-01-01 and 2020-05-20



Socio-Economic and Demographic Variables

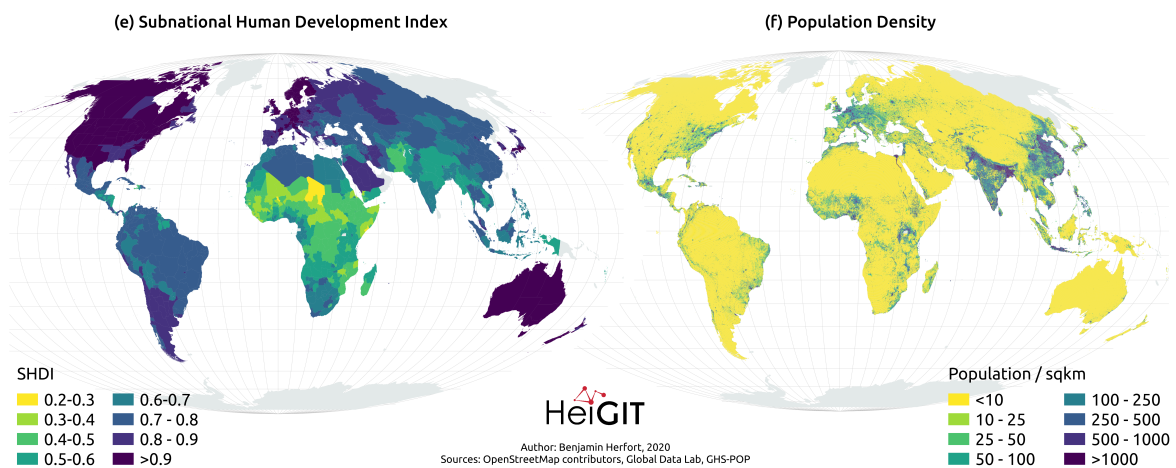


Figure 2: Spatial distribution of the number of (a)-(b) buildings and (c)-(d) highways added to OSM between 2008/01/01 - 2020/05/20 in regard to overall and humanitarian mapping activities. Further the maps depict the spatial distribution of (e) SHDI and (f) population density. Note that population count (and not density) is used in the statistical models. Created using QGIS 3.10.8 (<https://www.qgis.org/en/site/>).

mapping this effect was less pronounced and could only be estimated with a higher uncertainty. This might indicate that humanitarian mapping mitigates the negative effects of gender inequality on OSM mapping at least to some degree. Cellphone usage showed an ambiguous effect on mapping activity. For highways mapping its effect was very low and not significant at a level of 0.01. However, it seems that there was a stronger positive effect on humanitarian building mapping. In our model, cellphone usage was the variable which showed the highest collinearity with SHDI. This might have contributed to the observed higher uncertainty of the regression coefficient estimates. In total, our models explained 79% (buildings) and 91% (highways) of the observed deviance for overall OSM mapping activity. With regard to humanitarian mapping activity, we observed an explained deviance of 68% and 72%. Explained deviance for models without spatial eigenvectors were 60% and 83% for all OSM and 54% and 60% for humanitarian mapping. The increase in explained deviance shows that the spatial eigenvectors presumably captured effects of non available covariates with spatial structure that would otherwise lead to induced spatial autocorrelation in the regression residuals.

Analyzing the residuals of the GLM emphasizes the regional variations in humanitarian mapping activity. The residuals tell for which regions the association between response and predictors differed from the general relationship across all subnational units. The residuals have been derived for the subnational units and aggregated in a subsequent step to allow an interpretation on the country level. Hence, the analysis highlights most deviating countries (and not subnational units) where either other factors not captured in our analysis drive or prohibit mapping, or the effect of our set of predictors on mapping was considerably different to the global coefficients. The spatial eigenvector mapping captured some of the spatial pattern present in the data. Hence, residuals provide insights on deviations that were beyond what could be explained by the spatial structures in the data itself.

Figure 4 provides an overview on the residuals of the GLM in regard to humanitarian buildings and highways contributions for all countries with low and medium human development ($\text{SHDI} < 0.7$). Countries for which both residuals were negative showed stark, less-than-average humanitarian mapping contributions for buildings and highways. Among these countries were Pakistan, Somalia, Rwanda and Niger. For these regions we suspect the existence of factors which prevented mapping.

Zimbabwe and Uganda were part of the countries for which we observed humanitarian contributions slightly above average. For Uganda the residuals indicated the existence of moderately boosting factors, such as funding for mapping activity in Kampala provided from the Open Cities program as part of World Banks Open Data Resilience Initiative (Global Facility for Disaster Reduction and Recovery and The World Bank, 2016).

Finally, Nepal and Swaziland have been identified as outliers with unexpectedly high humanitarian contributions for buildings and highways. As depicted already this underlines the impact of the mapping in response to the 2015 Gorkha Earthquake in Nepal (Poiani et al., 2016). In Swaziland, extensive mapping has been organized by HOT during the Malaria Elimination Mapping Programme in 2016.

2.3 Temporal Evolution

This section draws attention to the temporal evolution of humanitarian mapping activity in OSM. Figure 5 provides monthly activity and trend-cycle for the number of (a) buildings and (b) highways added to OSM and further shows the number of volunteers in the HOT Tasking Manager that added data to OSM (mappers) and the number of volunteers that reviewed and corrected the work of others (validators).

The peaks in the time series plots revealed the impact of major disaster activations and organized mapping campaigns. During the mapping activations for 2013 Typhoon Haiyan and 2014 West Africa Ebola Crisis, up to 1,000 mappers added nearly 250,000 buildings and 40,000 highways to OSM per month. The mapping efforts after the 2015 Nepal Earthquake were supported by up to 1,900 mappers in the immediate aftermath of the disaster. In the month of the activation more than 1 million buildings and around 70,000 highways were added to OSM through the HOT Tasking Manager. The mapping activations for the 2017 Mexico Earthquake — as well as Hurricanes Irma and Maria around the same time — have resulted in unprecedented high numbers. More than 5,700 mappers added around 2 million buildings and 150,000 highways. Subsequent disaster activations, such as the response to 2019 Cyclone Idai in Mozambique showed fewer engagement. Only recently, the mapping activities during the COVID-19 pandemic reached the highest level from 2012-2017. In April 2020, more than 5,000 users added more than 2 million buildings and 120,000 highways.

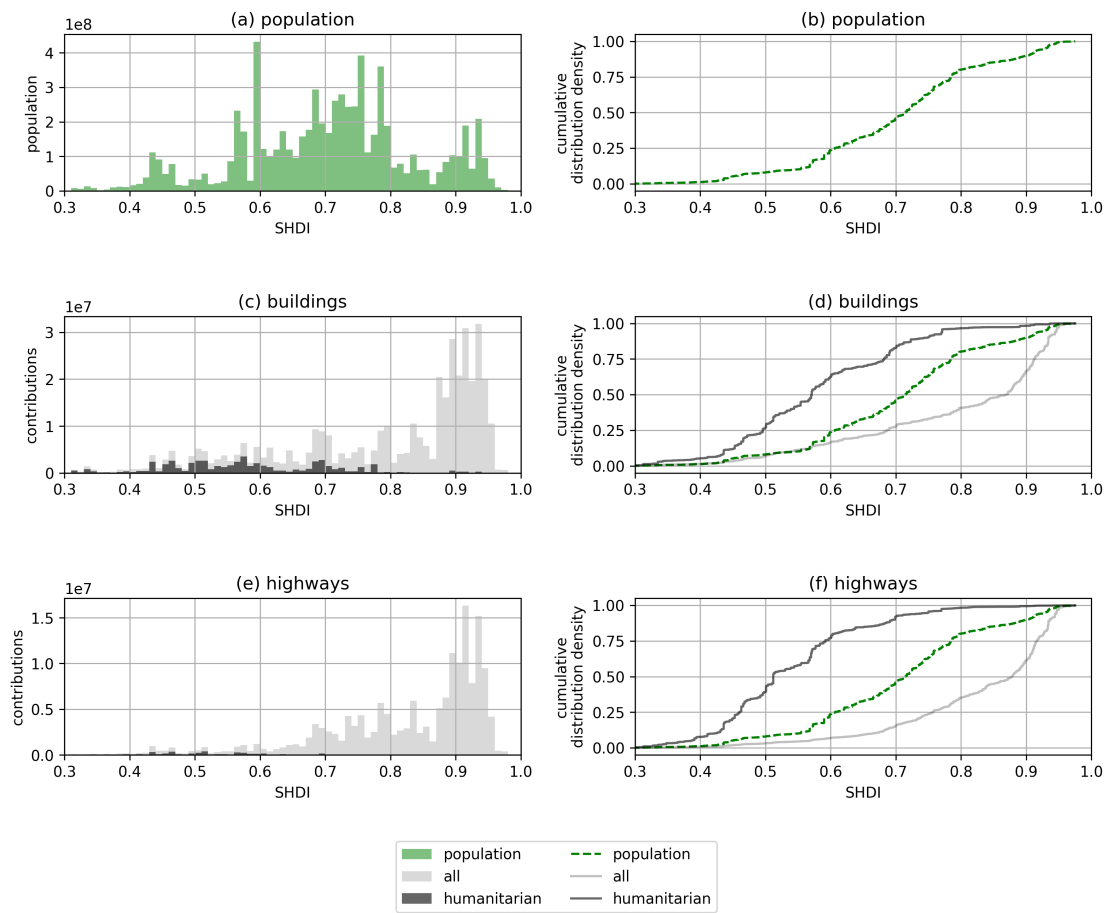


Figure 3: Frequency distribution and cumulative distribution density of (a)-(b) population and cumulative number of (c)-(d) buildings and (e)-(f) highways added to OSM between 2008/01/01 - 2020/05/20 in regard to overall and humanitarian mapping activities and Subnational Human Development Index. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

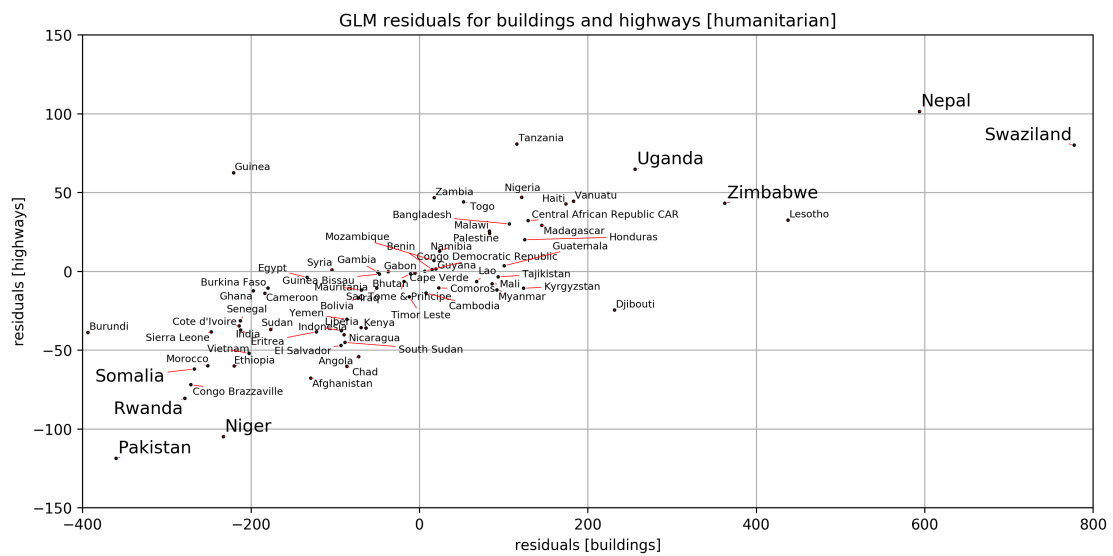


Figure 4: Distribution of GLM residuals for the number of buildings and highways added to OSM between 2008/01/01 - 2020/05/20 in regard to humanitarian mapping activities. The residuals have been derived for the subnational units and have been aggregated to the national level in a subsequent step. This plot depicts the values for all countries with medium and low human development (SHDI ≤ 0.7). Countries mentioned in the text are highlighted by increased font size. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

Buildings, SEVM	All			Humanitarian		
	Estimate	Std. err.	z value	Estimate	Std. err.	z value
(Intercept)	-12.33***	0.62	-19.79	-15.62***	1.22	-12.77
log(population)	13.24***	0.33	40.37	13.75***	0.64	21.56
shdi	0.41***	0.04	10.83	-1.13***	0.08	-14.29
cellphone users	0.11**	0.03	3.26	0.26***	0.06	4.06
gender equality	0.21***	0.03	7.30	0.15**	0.06	2.62
disaster deaths	0.39***	0.04	9.01	0.28***	0.05	5.51
youthmappers	0.03**	0.01	2.77	0.04*	0.02	2.16
	explained deviance: 0.79 Eigenvectors: 12			explained deviance: 0.68 Eigenvectors: 8		
Buildings, without SEVM	All			Humanitarian		
	Estimate	Std. err.	z value	Estimate	Std. err.	z value
(Intercept)	-10.93***	0.77	-14.12	-17.23***	1.20	-14.40
log(population)	12.56***	0.40	31.02	14.79***	0.62	23.88
SHDI	0.62***	0.05	13.72	-1.27***	0.07	-17.56
cellphone users	-0.003	0.42	-0.09	0.27***	0.05	5.14
gender equality	0.38***	0.03	11.29	0.34***	0.05	7.19
disaster deaths	0.58***	0.05	12.17	0.19***	0.05	3.57
youthmappers	-0.007	0.02	-0.43	0.03	0.02	1.62
	explained deviance: 0.60			explained deviance: 0.54		
Highways, SEVM	All			Humanitarian		
	Estimate	Std. err.	z value	Estimate	Std. err.	z value
(Intercept)	-11.78***	0.38	-31.11	-20.06***	1.43	-13.99
log(population)	12.32***	0.20	61.44	14.54***	0.74	19.65
SHDI	1.03***	0.03	35.46	-1.29***	0.09	-14.60
cellphone users	0.04*	0.02	2.04	0.03	0.07	0.056
gender equality	0.12***	0.02	7.32	0.15*	0.06	2.52
disaster deaths	0.21***	0.05	4.32	0.37***	0.05	7.17
youthmappers	-0.01	0.001	-1.05	0.04	0.02	1.86
	explained deviance: 0.91 Eigenvectors: 16			explained deviance: 0.72 Eigenvectors: 7		
Highways, without SEVM	All			Humanitarian		
	Estimate	Std. err.	z value	Estimate	Std. err.	z value
(Intercept)	-9.56***	0.43	-22.35	-23.11***	1.54	-15.00
log(population)	11.10***	0.22	49.49	16.33***	0.79	20.65
SHDI	1.10***	0.03	36.21	-1.56***	0.10	-16.20
cellphone users	0.07**	0.03	2.74	0.11	0.07	1.64
gender equality	0.14***	0.02	7.21	0.46***	0.06	8.18
disaster deaths	0.32***	0.05	6.34	0.28***	0.06	4.76
youthmappers	-0.02*	0.01	-2.07	0.003	0.03	0.14
	explained deviance: 0.83			explained deviance: 0.60		

Table 2: Results of Quasi Poisson Generalized Linear Regression Models with and without Spatial Eigenvector Mapping for cumulative number of buildings or highways added to OSM in regard to overall and humanitarian mapping activities. The spatial eigenvectors contribute to the explained deviance. For models without incorporation of spatial autocorrelation standard errors are too small, coefficients might be biased - these models are only reported for comparison. Regression coefficients are reported for standardized predictors to make the effect of the different predictors comparable. The only exception is population which has not been standardized due to the use of the log-transformation and since population is included as a covariate that needs to be controlled for but not of primary interest. Significance levels: * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.

Besides the monthly peaks related to disaster response efforts, humanitarian contributions in OSM have also increased in times without major disaster activations. The trend-cycle analysis revealed that humanitarian mapping has experienced sustained growth regarding monthly added buildings and highways from 2012 to mid-2017. This was supported by a growing number of monthly active mappers and validators participating in the HOT Tasking Manager. However, the trend-cycle analysis further depicted a consolidation phase since mid-2017. From this time on humanitarian building and highway mapping have shown stagnation or a slight downward trend. These trends were again supported by activity in the HOT Tasking Manager. Since mid-2017 there has been a stagnation in regard to number of mappers and a considerable decrease in number of validators. However, this downward trend might have changed again as the number of added buildings, mappers and validators increased since mid-2019.

The previous results show that there were considerable variations between the extent and quantity of overall and humanitarian mapping in OSM, as well as changes in regard to the global temporal evolution of humanitarian mapping. This section explores the similarity between the number of monthly

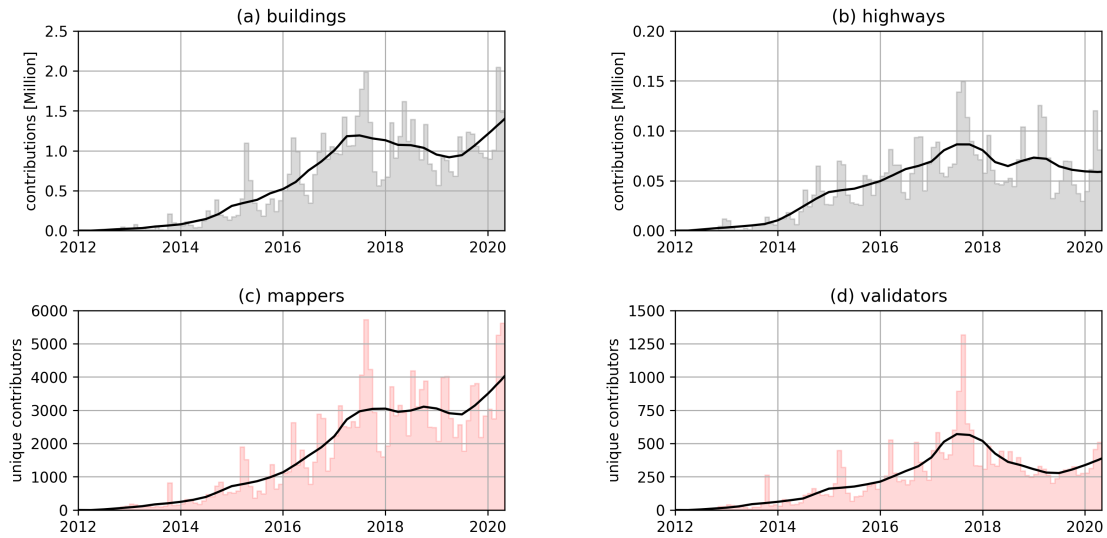


Figure 5: Monthly frequency distribution and trend cycle of humanitarian OSM contribution counts between 2012/01/01 - 2020/05/20 for (a) buildings and (b) highways and (c) mappers and (d) validators in HOT Tasking Manager. The line plots provide the trend line for each variable. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

added, changed and deleted buildings in OSM per country utilizing a dynamic time warping approach. The analysis has been performed for all countries with medium and low human development. Based on agglomerative hierarchical clustering (Legendre and Legendre, 2003) of the temporal evolution per country, we defined five groups which were part of three clusters: (a) steady trajectory, (b) alternating trajectory, (c) stepped trajectory. Figure 6 displays the cluster dendrogram and the cumulative distribution density for each group. Table 3 provides further insights on the cluster association of all countries with low and medium human development.

Most countries showed a steady trajectory of humanitarian building mapping in OSM over time (cluster a). The biggest group of countries within this cluster (a1) contained 25 countries for which mapping has followed a three stage process. The trajectories for Tanzania and Uganda were exemplary for this group. Firstly, building mapping has grown slowly in these regions over a longer time. In the second stage mapping activity increased heavily, but steadily over a time of several months. Finally, in the last stage mapping activity declined, but stayed at a higher level compared to the time before the peak. In the cumulative distribution density plot this results in a S-shaped graph reminding of a logistic curve.

The trajectories for 15 countries being part of group (a2) also showed continued and steadily increasing mapping activity over time. However, the cumulative distribution density plots for most of these countries did not level off. Currently, many of these countries appear to be approaching a peak in mapping activity, which some may have already reached by the end of 2019.

Not all countries show an unambiguous temporal pattern. For most of the 10 countries in group (b1) humanitarian building mapping has started relatively early between 2013 and 2015. However the mapping trajectory alternated between times of steady mapping, times of accelerated mapping and also times when little mapping happened. Mapping in Liberia and the Central African Republic could be described in this way. As depicted in the cluster dendrogram (see Figure 6) this group could be seen as a hybrid between (a) and (c).

The third cluster (c) depicts 16 countries for which the cumulative distribution density plot revealed a trajectory with either one single or multiple characteristic steps per country. For the 10 countries in group (c1), this represented multiple sequences of times with relatively high mapping activity (e.g.

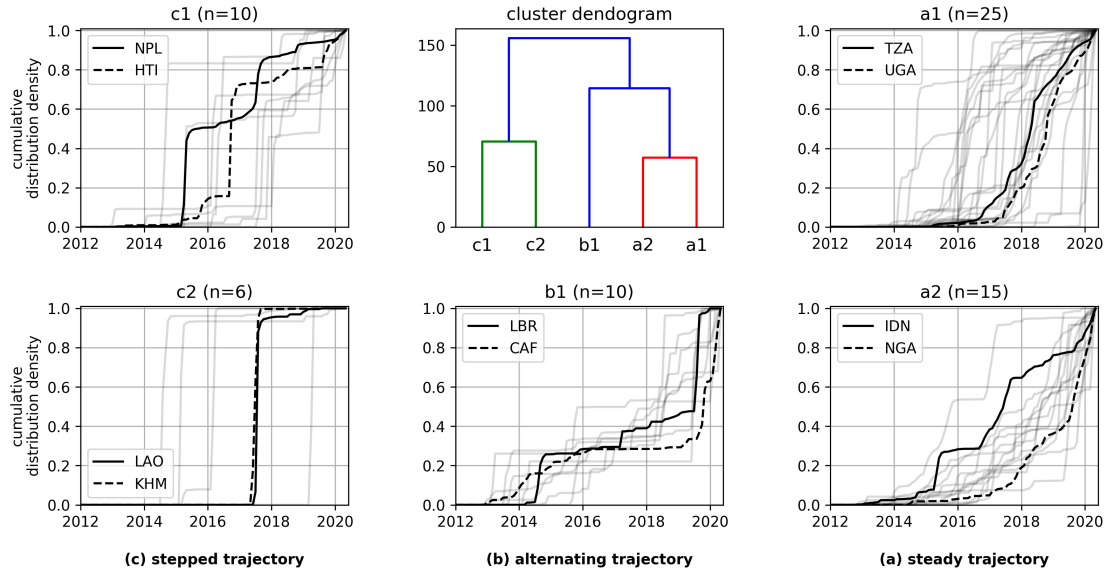


Figure 6: Cluster dendrogram and cumulative distribution density plots for monthly humanitarian OSM building mapping between 2012/01/01 - 2020/05/20. This plot depicts the values for all countries with medium and low human development index (SHDI <0.7). The groups are defined by a hierarchical clustering approach based on a distance matrix derived from dynamic time warping time series similarity analysis. For each group the trajectory for the two countries with highest overall mapping activity are highlighted. We used alpha-3 ISO 3166 codes to refer to the countries: TZA - Tanzania, UGA - Uganda, NPL - Nepal, HTI - Haiti, LAO - Laos, KHM - Cambodia, LBR - Liberia, CAF - Central African Republic, IDN - Indonesia, NGA - Nigeria. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

during a disaster activation) followed by times with lower mapping activity (e.g. after an activation). The humanitarian mapping contributions in Nepal and Haiti followed this temporal pattern. Note that the Haiti mapping activity here is in responses to the 2016 Hurricane Matthew mapping activation and does not include the mapping activity from 2010.

Countries that have been subject to large-scale, one-time mapping efforts form group (c2). In the cumulative distribution density plot, this is characterized by shapes with one single, big step. For these countries almost all data has been added to OSM within a single short-term period. Only very little mapping happened before and after that time span. Laos and Cambodia are part of this group.

3 Discussion

The long-term, global perspective of our analysis has — for the first time — characterized and quantified humanitarian mapping’s significant impact on the volume and spatial footprint of objects added to OSM. We showed that there still exists a clear mismatch between the distribution of mapping activity in OSM vis-à-vis the distribution of the global population on the Earth’s surface. The unprecedented detail of our analysis allowed us to identify distinct areas of concentrated mapping efforts and areas for which more mapping is still needed such as South America, Central America and Central Asia. The GLM models showed that socio-economic and demographic characteristics were able to explain a large part of this mismatch and that humanitarian mapping has helped change the spatial footprint of OSM globally in regards to the socio-economic profile of areas mapped.

Cluster	Temporal Evolution	Countries
(a)	(a1) steady - declining	Tanzania, Uganda, Zimbabwe, Mozambique, Bangladesh, Myanmar, India, Malawi, Mali, Guinea, Sierra Leone, Lesotho, Swaziland, Guatemala, Togo, Vietnam, Benin, Yemen, Djibouti, Niger, Ghana, Somalia, Tajikistan, Nicaragua, El Salvador
	(a2) steady - constant	Indonesia, Nigeria, Congo Democratic Republic, Zambia, Kenya, South Sudan, Burkina Faso, Madagascar, Chad, Honduras, Burundi, Cote d'Ivoire, Iraq, Kyrgyzstan, Ethiopia
(b)	(b1) alternating	Liberia, Central African Republic CAR, Cameroon, Senegal, Namibia, Gambia, Afghanistan, Sudan, Pakistan, Mauritania
(c)	(c1) stepped - multiple	Nepal, Haiti, Rwanda, Angola, Bolivia, Syria, Timor Leste, Congo Brazzaville, Bhutan, Egypt
	(c2) stepped - single	Laos, Cambodia, Palestine, Vanuatu, Comoros, Guyana

Table 3: Clusters of temporal trajectories of the monthly number of buildings added, changed and deleted in OSM per country between 2012/01/01 and 2020/05/20. The groups are defined by a hierarchical clustering approach based on a distance matrix derived from dynamic time warping time series similarity analysis. This table lists all countries with medium and low human development index (SHDI <0.7).

Furthermore, the analysis of the GLM residuals provided insights into the variability in humanitarian OSM contributions per country that could not be explained by socio-economic and demographic characteristics alone. This analysis makes it clear that there can be underlying factors that may promote (e.g. the occurrence of disasters) or block mapping (e.g. gender inequality), but which have not completely been captured in our analysis. These results have been refined by the analysis of the temporal evolution which enabled us to identify three groups with distinct profiles of OSM humanitarian mapping trajectories. This differentiation between (a) steady trajectory, (b) alternating trajectory and (c) stepped trajectory provides novel insights towards a typology of the temporal evolution of OSM contributions at the country level, complementing previous research analyzing trajectories of OSM communities (Barron et al., 2014; Barrington-Leigh and Millard-Ball, 2017).

In line with the latest research presented at the OSM community’s 2020 State of the Map Conference (Minghini et al., 2020), our results confirm that mapping activity is strongly affected by disasters and seems to be positively influenced by the existence of an established local OSM community, local offices of the HOT, as well as the focus of third party projects such as Open Cities Africa funded by the World Bank (Global Facility for Disaster Reduction and Recovery and The World Bank, 2016). However, the existence of YouthMappers chapters (Solís et al., 2018) didn’t show a significant effect in our models. Further research should investigate to what extent YouthMappers chapters might focus more on local mapping in contrast to remote mapping of buildings and highways. In contrast, we suspect that in some countries (e.g. Pakistan (Ali and Imran, 2020)), local restrictions that are not captured in our model such as legal constraints to surveying and mapping might have a strong negative impact on the mapping activity.

The increasing contribution of corporate mappers (Anderson et al., 2019) is another pressing field where a long-term monitoring and fine grained understanding of temporal and spatial components of mapping in OSM is needed. In this regard, recent technical developments towards machine learning based methods to extract buildings and roads from very high resolution satellite imagery that have been promoted by companies (such as Facebook and Microsoft) and scientists are showing their impact on OSM already today (Li et al., 2020; Chen and Zipf, 2019).

When it comes to the temporal evolution of OSM humanitarian mapping activity, our results could confirm the findings of other authors on the potential of collaborative mapping for disaster relief and the evolution it has gone through (Palen et al., 2015; Poiani et al., 2016; Dittus et al., 2016). But this analysis also shows that the growth of the humanitarian mapping community faced challenges as indicated by a phase of reduced activity which started in 2017 and seemed to have ended only recently in the year 2019. To the best of our knowledge, this has not been addressed by other authors so far. The stepped-trajectory temporal patterns on the country level put into question the long-term effects of some organized mapping campaigns towards making OSM more inclusive and democratic. Some researchers have put this absence of mapping activity into context with the diminishing feeling of urgency and crisis which have been observed in the aftermath of this major disaster activations (Liu et al., 2018).

The last months analyzed in this research fell into the time of the global COVID-19 spread. The overall temporal trend might indicate that this has contributed to increased mapping. However, investigating the further impact of COVID-19 on global OSM contributions had to remain out of scope for this research.

Unavoidably, our analysis comes with several limitations that need to be considered to put our findings into context. Our insights about humanitarian mapping in OSM only provide an incomplete picture which lacks an on-the-ground perspective and neglects other remote mapping tools, since we considered only the mapping that was organized through the HOT Tasking Manager. For instance, humanitarian mapping that has been organized by local residents on the ground is not considered here. This limitation is accompanied by the fact that our analysis only focused on two types of mapped objects (buildings, highways). Mapping in OSM comes with a much greater variety of potential map objects (e.g. health facilities, schools, water points), which can add particular value in comparison to other geographic data sets. We are aware of the fact that our definition of humanitarian mapping is therefore oversimplified and the results must be taken with a grain of salt. In many regions of the world there is no clear distinctive line between humanitarian and non-humanitarian mapping activities as the humanitarian and non-humanitarian OSM communities are not disjoint. Hence, not every contributor that used the HOT Tasking Manager is exclusively a “humanitarian” mapper, but might identify herself primarily as an OSM mapper or member of a local OSM community. Future work could expand this analysis with in-depth analyses of specific communities (e.g. using mixed qualitative-quantitative methods), as well as could investigate the role of other organized mappers such as cooperate mappers (Anderson et al., 2019) to disentangle the impact of the different sub-communities of OSM on mapping activities.

The observed explained deviance value in regard to overall highway mapping of 0.9 is very high and should be critically assessed as another potential limitation of our study. For both the distribution of buildings and roads a strong positive association with population density as well as with human development can be assumed. If one of the features would be mapped more completely, this is likely to lead to a stronger association compared to a feature that is less completely mapped. Roads have been the first feature mapped in many parts of the world and the road network has been estimated as 80% (Barrington-Leigh and Millard-Ball, 2017) complete at the global scale. A comparative completeness analysis for buildings is missing. However, the differences in model performance might hint towards a higher completeness for roads. In addition, one has to keep in mind the presence of strong spatial patterns that are reflected by the spatial eigenvectors which increase model performance. In addition to the removal of spatial autocorrelation and of the associated problems in regression coefficient estimation (Dormann et al., 2007) this presumably also captured the effect of missing spatial co-variables. The increase in explained deviance for highways by incorporating the spatial eigenvectors was only modest for all OSM (from 83% to 91%). For the other three models the increase in explained deviance was stronger (12%-19% absolute increase in explained deviance). This can be interpreted as that mapping of the OSM road network in countries with a high HDI was less influenced by prohibitive or stimulating factors.

Further limitations concern the information used from GHS-POP and SHDI which are only able to provide imperfect proxies for the actual population distribution and socio-economic profiles of regions. There are a few other data sets, such as WorldPop, that provide information on human population on the global scale (Tatem, 2017). Nevertheless, even such very high resolution datasets have similar limitations, i.e. the estimation of inhabited areas is biased towards urban areas (Klotz et al., 2016). As a consequence, settlements in rural areas are more likely to be left out in the global population data sets as well. The SHDI provides only a narrowed view on socio-economic development. Investigating the impact of the individual components of the SHDI (health, education, income) should be considered for future work. Information on cellphone users, gender equality, disaster deaths and YouthMapper chapters were only

available at the national scale. It has to be assumed that these vary as well between the subnational units. This has to be expected especially for disaster deaths. Adding further indicators to the analysis, such as the Multi Dimensional Poverty Index, could provide a way to better account for the structural effects that pose a challenge especially on poor and vulnerable populations (Alkire and Foster, 2011). In future work, we would like to investigate how this might reduce the uncertainty of the statistical models applied in this paper.

In conclusion, humanitarian mapping efforts of the previous decade have already made OSM more inclusive, contributing to diversify and expand the spatial footprint of the areas mapped. However, humanitarian organisations, local communities and other stakeholders involved should learn from the various challenges we have highlighted in this paper, in order to have sustained impact in the future. Based on our research results we want to highlight three important recommendations:

First: Improve methods to monitor mapping activity and identify where mapping is needed. At the current state for most organizations “monitoring” merely consists of counting the overall number of buildings and highways and usually neglects the spatial and temporal components of mapping activity. The humanitarian OSM community should therefore adopt monitoring methods that can more clearly address the question of “Where is mapping needed?”, instead of “How much mapping happened?”. The spatial analysis methods presented in this paper can offer a basis for starting addressing the former question.

Second: Rethink project design to avoid non-sustainable outcomes. We would recommend that the organizations that have been engaged in countries with one-time mapping efforts, e.g. in Laos or Cambodia, should be encouraged to assess how project could be designed to contribute to more sustainable outcomes. We argue that “putting countries on the map” should be seen as only the first step towards considering such countries effectively included, as real inclusion requires an active local OSM community with the capacity to generate, maintain and improve geographic data that reflects local perspectives in the long term. Especially in the fast developing domain of applying artificial intelligence and machine learning methods to humanitarian mapping purposes, we see a critical lack of local engagement by design. This might reproduce the very same one-time-mapping patterns that have been observed in the past as revealed by our analysis.

Third: Remove structural barriers to empower local communities and develop capacity. Future mapping efforts should promote local data generation methods that not only achieve high-quality data but also simultaneously empower local communities and support them to acquire new perspectives on their territories and development potentials (de Albuquerque and de Almeida, 2020). Focusing on empowerment goes beyond technically involving local communities in the map making process. In order to achieve this, humanitarian organisations and local OSM communities need to identify the structural barriers that exclude some social groups from participating in (humanitarian) OSM mapping and work with them to develop capacity to overcome those barriers.

Balancing the unequal spatial and temporal contribution patterns in OSM clearly goes beyond the population of a more comprehensive geographic data base. It is about building diverse and vibrant communities that can take part in the map making process and are empowered by the mapping results. We believe that such an inclusive OSM community can fuel the much needed data revolution and will prove its value to inform better political actions towards achieving the SDGs.

4 Methods

4.1 Data

4.1.1 OpenStreetMap History Data Set

Our analysis evaluated data contributed to OSM by utilizing the OSHDB (Raifer et al., 2019) a high-performance spatio-temporal data analysis platform for OSM full-history data. Following this approach enabled us to investigate all edits to the OSM database and we refer to these as contributions. The OSHDB distinguishes between different aspects of user contributions: Whenever a user adds a new object in OSM (e.g. draws a polygon, sets the key-value pair to `building=yes` and uploads to OSM) this is referred to as a creation in OSM. Users can modify existing objects in OSM. If the attribute description

of an object is adjusted (e.g. a user changes the value of `highway=unclassified` into `highway=residential`) or new attributes are added or existing ones are removed this is referred to as a tag change. Users can also modify the shape of existing objects in OSM. This so-called geometry change can be done by moving nodes (for all feature types node, way and relation) or by adding or removing nodes (for feature types way and relation). A change in the geometry will most likely cause a change (increase or decrease) in derived statistics such as length or area. Whenever an object is removed from the OSM database or the respective tag is removed from the object (e.g. "building" key is removed for a building, but the nodes are not deleted) this is referred to as a deleted object (deletion).

Our analysis covered contributions related to the following tag-value combinations and respective feature types in OSM: `building=*` (ways) and `highway=*` (ways). These feature types include all types of roads and paths as well as all types of buildings. For all objects we derived the number of contributions (count) and the payload (area in square meter for buildings, length in meter for highways). To allow for an unambiguous comparison between different contribution types (e.g. creation vs. tag change) and between different features (e.g. buildings vs. highways) we decided to report only on counts in the scope of this paper - instead of the common way of reporting the number of buildings and the length of highways.

The spatial unit of our analysis has been defined by a grid network of hexagons which has been generated using the DGGRID software (Sahr et al., 2003). All hexagons had the same size in terms of area and cover ~96 square kilometers (this refers to the so-called zoom level 12). Their spatial characteristics make hex grids an unit suited better for spatial analysis compared to other grid systems such as the Tile Map Service grid (which is handy to be displayed at web maps, but not so for deriving area based statistics).

4.1.2 HOT Tasking Manager Data Set

Our definition of humanitarian mapping activities utilized data from the HOT Tasking Manager platform which is online at tasks.hotosm.org. We were provided with a dump of the HOT Tasking Manager database by HOT covering all projects created within 2012-11 and 2020-05.

We refer to projects as defined by the HOT Tasking Manager. A project is usually set up by an organisation (e.g. American Red Cross) and defined by a number of tasks, which will be generated from a polygon geometry depicting the area of interest. Per project, project managers add information on which features (e.g. buildings, highways, waterways) should be mapped and which satellite imagery should be used for their digitization. Often the description will also provide contextual information for what purpose map data is collected.

Tasks are a central concept of the HOT Tasking Manager. For each project the overall area is split into many smaller areas (usually squares) which are defined as tasks. It's common that several users will work on a task (e.g. user A maps it and user B validates it), but two users will never work on the same tasks at the same time in the HOT Tasking Manager. Hence, after a user selected a task it is locked (for maximum 2 hours).

We measured the activity in the HOT Tasking Manager based on the concept of sessions, which we introduced for this research. A session has been defined by a user performing an action on a task. The action performed can be one out of the following which indicate the kind of change that the user induced: unfinished, mapped, validated, invalidated, split or bad imagery. For each session there was a start timestamp (the time the users selects the task to map in the HOT Tasking Manager) and an end timestamp (the time the user finished mapping and sets the status in the HOT Tasking Manager to either mapped or one of the others mentioned above). For each session we queried the OSHDB to calculate the OSM based statistics for buildings and highways contributed by the user for the given extent and time range.

4.1.3 Socio-economic and Demographic Datasets

Global population distribution played a central role in our research. We used the GHS-POP database (Freire et al., 2016) which relies on the Global Human Settlement Layer (GHSL) provided by JRC and a disaggregation of CIESIN's Gridded Population of the World (GPWv4.10). For the results presented here we used population information provided at a spatial resolution of 250 m for the year of 2015. We

selected this data set instead of the data available from WorldPop due to the fact that the methodology applied by the team behind World Pop uses various geospatial covariate layers derived from OSM data. This could have introduced an unintended bias when comparing that data set with OSM. Since the statistical analysis performed in this research was targeted at the sub-national level GHS-POP based on GPW provides a solid basis.

To characterize regions based on their socio-economic status we utilized the Subnational Human Development Database (Smits and Permanyer, 2019). This data set breaks down the national Human Development Index (HDI) based on data from statistical offices and aggregated census data and data derived from household surveys into their subnational counterparts. In our analysis we relied on the aggregated SHDI and did not further consider its individual components (education, standard of living, health).

We used information on the total number of deaths caused by disasters caused by natural hazards per 100,000 inhabitants provided by EM-DAT (International Disaster Database) (Guha-Sapir et al., 2016). We considered events which occurred for the years 2012 to 2020. The data is provided on the national level.

Gender equality has been selected as another co-variate in our analysis. Here we use the proportion of seats held by women in national parliaments (referring to SDG indicator 5.5.1). The data has been extracted from the United Nations' Global SDG Indicators Database (<https://unstats.un.org/sdgs/indicators/database/>).

Information on cellphone usage (mobile cellular subscriptions per 100 people) has been derived from World Bank's DataBank (<https://databank.worldbank.org>). The latest available information was used per country.

Finally, the number of YouthMappers chapters per subnational unit has been derived from the public chapter listing on their website (<https://www.youthmappers.org/chapter-listing>).

4.2 Statistical Analysis

4.2.1 Scale and Spatial Footprint

The spatial footprint of overall and humanitarian mapping activities is generated on the basis of the hexagonal grid described above. For each hexagon we derived the count of OSM contributions by intersecting the geometry of the OSM entity with the hexagon geometry. Humanitarian contributions were defined by the sessions from the HOT Tasking Manager and constitute a subset of the overall contributions in OSM.

The spatial footprint of OSM contributions for building creations and highway creations was visually assessed using a map representation of the results. The maps were based on the equal-area Mollweide projection to allow for non-biased visual comparison between all geographic regions. The same symbology was applied to all maps to facilitate the detection of relative and absolute differences between overall and humanitarian mapping and between buildings and highways.

4.2.2 Cumulative Distribution Density Analysis

Data from OSM and population has been aggregated to the subnational level to account for the spatial resolution of SHDI. We assess the cumulative distribution density of population and mapping in OSM in regard to four SHDI classes based on cutoff points defined by the United Nations Development Programme (UNDP) (United Nations Development Programme, 2019): low human development (SHDI < 0.550), medium human development (SHDI: 0.550 - 0.699), high human development (SHDI: 0.700 - 0.799), very high human development (SHDI > 0.800).

4.2.3 GLM and Spatial Eigenvector Mapping

The relationship between SHDI, gender equality, disaster deaths, cellphone usage, number of YouthMappers chapters and population in regard to overall and humanitarian mapping contributions was analyzed at the sub-national scale based on a quasi-Poisson Generalized Linear Regression Model (GLM) with Spatial Eigenvector Mapping (Griffith et al., 2019; Chun and Griffith, 2013; Griffith and Peres-Neto,

2006). Using a count regression approach was necessary to account for the characteristics of the response variables which are count data. To account for overdispersion in the mean-variance relationship a quasi-poisson model was used.

It has been shown that there are important advantages of this approach over any other spatial modeling approach applied to geographical problems in the presence of spatial autocorrelation (Griffith et al., 2019; Chun and Griffith, 2013; Griffith and Peres-Neto, 2006). For instance, it has been shown that trend surface analysis might provide reasonable solutions, but only when the sample area is approximately homogeneous and the spatial structure to be modeled is rather simple and global, e.g. a gradient (Griffith and Peres-Neto, 2006). The required use of a quasi-poisson approach ruled out approaches such as generalized least squares, spatial lag or spatial error models that are only suitable for linear models (Dormann et al., 2007). Regression analysis without spatial eigenvector mapping led to strong spatial autocorrelation in the residuals. The spatial eigenvector mapping successfully removed spatial autocorrelation in the residuals. The regression analysis was performed in R (R Core Team, 2020) using the packages `spatialreg` (Bivand and Piras, 2015), `spdep` (Bivand et al., 2013) and `sf` (Pebesma, 2018). We performed this analysis for building creations and highway creations in OSM. The spatial weight matrix for the Spatial Eigenvector Mapping was based on a k-nearest-neighbor approach using the 10 closest neighbors and using row-standardized weighting. We evaluate the model's performance by reporting estimates, standard error, z value and p value for each coefficient. Furthermore, the explained deviance is calculated to provide an estimate of the model's power to explain the observed variability in the data set.

We assessed contribution inequality in respect to building creations and highways creations on the national level for all countries with low and medium human development (SHDI <0.7). For doing so the sub-national data set utilized previously was aggregated on the national level by summing up OSM contributions and population numbers. SHDI on the national level was derived as the weighted average of the sub-national index values considering population as the weight.

4.2.4 Temporal Evolution

The analysis of the temporal evolution of humanitarian mapping activities was performed on a monthly basis based on a time series decomposition approach. The analysis relies on the `mstl` method implemented in R's forecast package (Hyndman and Khandakar, 2008). The time series is thereby decomposed into seasonal, trend and remainder components iteratively by loss smoothing (Cleveland et al., 1990). Multiple seasonal periods were thereby allowed. In this article we provide a visualization of the observed values and the derived trend for humanitarian OSM contributions and the number of distinct mappers and validators in the HOT Tasking Manager.

We assessed the temporal dimension of contributions by deriving the similarity between time series for the cumulative distribution density of humanitarian building contribution per country using dynamic time warping (DTW) (Salvador and Chan, 2007). We relied on the implementation of the algorithm in python's `fastdtw` library (<https://github.com/slaypni/fastdtw>). The similarity values were derived for all countries with low and medium human development (SHDI < 0.7). The analysis was complemented by an agglomerative hierarchical cluster analysis (Legendre and Legendre, 2003) which defined five groups of countries. The distance matrix was based on the euclidean distance between DTW based distances for humanitarian OSM contributions for buildings. The analysis was conducted based on scikit-learn's python implementation (Pedregosa et al., 2011) using the ward linkage criterion. Results were displayed for each cluster plotting the cumulative distribution density per country and highlighting the two countries with most overall contributions per group. As an addition to this the cluster dendrogram was plotted.

References

- Aitsi-Selmi, A., Murray, V., Wannous, C., Dickinson, C., Johnston, D., Kawasaki, A., Stevance, A. S., and Yeung, T. (2016). Reflections on a Science and Technology Agenda for 21st Century Disaster Risk Reduction: Based on the Scientific Content of the 2016 UNISDR Science and Technology Conference on the Implementation of the Sendai Framework for Disaster Risk Reduction 2015–20. *International Journal of Disaster Risk Science*, 7(1):1–29.

- Albuquerque, J., Herfort, B., and Eckle, M. (2016). The Tasks of the Crowd: A Typology of Tasks in Geographic Information Crowdsourcing and a Case Study in Humanitarian Mapping. *Remote Sensing*, 8(10):859.
- Ali, A. and Imran, M. (2020). The Evolution of National Spatial Data Infrastructure in Pakistan, Implementation Problems and the Way Forward. *International Journal of Spatial Data Infrastructures Research*, 15.
- Alkire, S. and Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95(7-8):476–487.
- Altay, N. and Labonte, M. (2014). Challenges in humanitarian information management and exchange: Evidence from Haiti. *Disasters*, 38(S1):50–72.
- Anderson, J., Sarkar, D., and Palen, L. (2019). Corporate Editors in the Evolving Landscape of Open-StreetMap. *ISPRS International Journal of Geo-Information*, 8(5):232.
- Attaran, A. (2005). An immeasurable crisis? A criticism of the millennium development goals and why they cannot be measured. *PLoS Medicine*, 2(10):0955–0961.
- Barrington-Leigh, C. and Millard-Ball, A. (2017). The world’s user-generated road map is more than 80% complete. *PLoS ONE*, 12(8):1–20.
- Barron, C., Neis, P., and Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, 18(6):877–895.
- Bivand, R. and Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, 63(18):1–36.
- Bivand, R. S., Pebesma, E., and Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, Second edition*. Springer, NY.
- Braunschweig, K., Eberius, J., Thiele, M., and Lehner, W. (2020). The State of Open Humanitarian Data.
- Chen, J. and Zipf, A. (2019). DeepVGI: Deep learning with volunteered geographic information. *26th International World Wide Web Conference 2017, WWW 2017 Companion*, (1):771–772.
- Chun, Y. and Griffith, D. A. (2013). *Spatial statistics and geostatistics: theory and applications for geographic information science and technology*. Sage.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). STL: A seasonal-trend decomposition. *Journal of official statistics*, 6(1):3–73.
- de Albuquerque, J. P. and de Almeida, A. A. (2020). *Modes of engagement: Reframing “sensing” and data generation in citizen science for empowering relationships*, pages 267–281. Manchester University Press, Manchester, England.
- Dittus, M., Quattrone, G., and Capra, L. (2016). Analysing Volunteer Engagement in Humanitarian Mapping: Building Contributor Communities at Large Scale. *Cscw ’16*, (January):108–118.
- Dormann, C. F., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., R. Peres-Neto, P., Reineking, B., Schröder, B., M. Schurr, F., Wilson, R., Dormann, C. F., Mcpherson, J. M., Arau, M. B., Davies, R. G., Kissling, W. D., Ohlemu, R., Peres-Neto, P. R., and Schurr, F. M. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30(5):609–628.
- Freire, S., MacManus, K., Pesaresi, M., Doxsey-Whitfield, E., and Mills, J. (2016). Development of new open and free multi-temporal global population grids at 250 m resolution. *Agile*, (June):6.

- Global Facility for Disaster Reduction and Recovery and The World Bank (2016). Open Data for resilience initiative (OpenDRI) — Policy Note and Principles. Technical report.
- Griffith, D. A., Chun, Y., and Li, B. (2019). *Spatial Regression Analysis using Eigenvector Filtering*. CRC Press.
- Griffith, D. a. and Peres-Neto, P. R. (2006). Spatial Modelling in Ecology: The Flexibility of Eigenfunction Spatial Analysis. *Ecology*, 87(10):2603–2613.
- Guha-Sapir, D., Below, R., and Hoyois, P. (2016). EM-DAT: the CRED/OFDA international disaster database.
- Haworth, B. T., Bruce, E., Whittaker, J., and Read, R. (2018). The good, the bad, and the uncertain: Contributions of volunteered geographic information to community disaster resilience. *Frontiers in Earth Science*, 6(November):1–15.
- Herfort, B., Li, H., Fendrich, S., Lautenbach, S., and Zipf, A. (2019). Mapping human settlements with higher accuracy and less volunteer efforts by combining crowdsourcing and deep learning. *Remote Sensing*, 11(15).
- Humanitarian OpenStreetMap Team (2019). Mapping 2021: HOT’s Strategic Plan 2019 – 2021. Technical report.
- Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3):1–22.
- Klotz, M., Kemper, T., Geiß, C., Esch, T., and Taubenböck, H. (2016). How good is the map? A multi-scale cross-comparison framework for global settlement layers: Evidence from Central Europe. *Remote Sensing of Environment*, 178:191–212.
- Lämmerhirt, D., Gray, J., Venturini, T., and Meunier, A. (2019). Advancing Sustainability Together? Citizen-Generated Data and the Sustainable Development Goals. *SSRN Electronic Journal*.
- Legendre, P. and Legendre, L. (2003). *Numerical ecology*. Elsevier.
- Li, H., Herfort, B., Huang, W., Zia, M., and Zipf, A. (2020). Exploration of OpenStreetMap missing built-up areas using twitter hierarchical clustering and deep learning in Mozambique. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166(November 2019):41–51.
- Liu, W., Dugar, S., McCallum, I., Thapa, G., See, L., Khadka, P., Budhathoki, N., Brown, S., Mechler, R., Fritz, S., and Shakya, P. (2018). Integrated participatory and collaborative risk mapping for enhancing disaster resilience. *ISPRS International Journal of Geo-Information*, 7(2):1–23.
- Minghini, M., Coetzee, S., Juhasz, L., Grinberger, A. Y., Mooney, P., and Yeboah, G., editors (2020). *Proceedings of the Academic Track at the State of the Map 2020*. Available at <https://zenodo.org/communities/sotm-2020>.
- Neis, P., Singler, P., and Zipf, A. (2010). Collaborative mapping and Emergency Routing for Disaster Logistics-Case studies from the Haiti earthquake and the UN portal for Afrika. *Geoinformatics Forum Salzburg*, (October):239–248.
- Palen, L., Soden, R., Anderson, T. J., and Barrenechea, M. (2015). Success & Scale in a Data - Producing Organization : The Socio - Technical Evolution of OpenStreetMap in Response to Humanitarian Events. *Chi 2015*, pages 4113–4122.
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1):439–446.

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830.
- Poiani, T. H., Rocha, R. D. S., Degrossi, L. C., and Albuquerque, J. P. D. (2016). Potential of collaborative mapping for disaster relief: A case study of openstreetmap in the Nepal earthquake 2015. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, volume 2016-March, pages 188–197.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.-B., and Zipf, A. (2019). OSHDB: a framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards*, 4(1).
- Sahr, K., White, D., and Kimerling, A. J. (2003). Geodesic discrete global grid systems. *Cartography and Geographic Information Science*, 30(2):121–134.
- Salvador, S. and Chan, P. (2007). Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580.
- Scholz, S., Knight, P., Eckle, M., Marx, S., and Zipf, A. (2018). Volunteered Geographic Information for Disaster Risk Reduction—The Missing Maps Approach and Its Potential within the Red Cross and Red Crescent Movement. *Remote Sensing*, 10(8):1239.
- Smits, J. and Permanyer, I. (2019). Data descriptor: The subnational human development database. *Scientific Data*, 6:1–15.
- Soden, R. and Palen, L. (2016). Infrastructure in the wild: What mapping in post-earthquake Nepal reveals about infrastructural emergence. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 2796–2807.
- Solís, P., McCusker, B., Menkiti, N., Cowan, N., and Blevins, C. (2018). Engaging global youth in participatory spatial data creation for the UN sustainable development goals: The case of open mapping for malaria prevention. *Applied Geography*, 98(January):143–155.
- Sui, D., Goodchild, M., and Elwood, S. (2013). Volunteered geographic information, the exaflood, and the growing digital divide. In *Crowdsourcing geographic knowledge*, pages 1–12. Springer.
- Sustainable Development Solutions Network (2015). Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a data revolution for the SDGs. *A report by the Leadership Council of the Sustainable Development Solutions Network*, (Version 7):160.
- Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data*, 4:2–5.
- United Nations (2014). A World that Counts: Mobilising the Data Revolution for Sustainable Development. *UN Data Revolution Report*.
- United Nations (2015). Transforming our world: The 2030 agenda for sustainable development. Technical report.
- United Nations (2020). The Sustainable Development Goals Report 2020. Technical report.
- United Nations Development Programme (2019). *Human development report 2019 : beyond income, beyond averages, beyond today: inequalities in human development in the 21st century*.
- Zook, M., Graham, M., Shelton, T., and Gorman, S. (2010). Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. *World Medical & Health Policy*, 2(2):6–32.

Investigating completeness and inequalities in OpenStreetMap: spatio-temporal analysis of global urban building data

Benjamin Herfort, Sven Lautenbach, João Porto de Albuquerque,
Jennings Anderson, Alexander Zipf

under review, submitted revised version 09 March 2023

Abstract

OpenStreetMap (OSM) has evolved as a popular dataset for global urban analyses, such as assessing progress towards the Sustainable Development Goals. However, many analyses turn a blind eye on the uneven spatial coverage of existing data. We employed a machine-learning model to infer the completeness of OSM building stock data for 13,189 urban agglomerations worldwide. For 1,848 urban centres (16% of the urban population), OSM building footprint data exceeds 80% completeness, but completeness remains lower than 20% for 9,163 cities (48% of the urban population). Although OSM data inequalities have recently receded, partially as a result of humanitarian mapping efforts, a complex unequal pattern of spatial biases remains, which vary across various human development index groups, population sizes and geographic regions. Based on these results, we provide recommendations for data producers and urban analysts to deal with the uneven coverage of OSM data, as well as a framework to support the assessment of completeness biases.

Reference— Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., and Zipf, A. (2023). Investigating completeness and inequalities in OpenStreetMap: spatio-temporal analysis of global urban building data. Under Review. Preprint available at: <https://doi.org/10.21203/rs.3.rs-1913150/v1>

1 Introduction

Between 2001 and 2018 the urban population growth and built-up area expansion accelerated, especially in large cities in the low- and lower-middle-income countries (Sun et al., 2020). Global urban land expansion is expected to grow rapidly in the next 20 years (Boo et al., 2022). Buildings constitute one of the most important physical elements of urban settlements (Hecht et al., 2013). However, little is known on a consistent basis about building inventories worldwide; a spatially detailed survey of the distribution and concentration of the building stock does not yet exist (Esch et al., 2022). Improving the systematic monitoring of the global urbanization process is a requirement for achieving the United Nation’s Sustainable Development Goals (SDGs), e.g. “urban” SDG 11 (“Make cities and human settlements inclusive, safe, resilient and sustainable.”), whilst especially in low-income countries the data are usually scarce (Sun et al., 2020).

Although building data are usually maintained by national statistics offices, they suffer from an estimated gap of \$1 billion USD globally in funding, and consequently, baseline geospatial data that should be provided by these agencies are often not accessible, not up-to-date or not available in standard formats (Braunschweig et al., 2020; Altay and Labonte, 2014). Tackling building data scarcity requires moving beyond insufficient traditional data sources to utilizing new, non-traditional sources for measuring the SDGs (Fritz et al., 2019; Sustainable Development Solutions Network, 2015).

Building data is an essential asset in global urban analyses for assessing progress towards a number of important urban goals. For instance, SDG Indicator 11.3.1 (“Ratio of land consumption rate to population growth rate”) would directly benefit from building footprint data. However this indicator is currently

mainly based on easily available remote sensing data, e.g. World Settlement Footprint (WSF) (Esch et al., 2022), which exacerbates the monitoring of structural changes such as changes in floorspace per capita or re-densification trends. The monitoring of SDG Indicator 11.1.1 ("Proportion of urban population living in slums, informal settlements or inadequate housing") would benefit from an analysis of building blocks and street networks considering their spatial relations, such as density and neighbourhood relations (Brelsford et al., 2018). SDG Indicator 11.7.1 ("Average share of the built-up area of cities that is open space for public use for all, by sex, age and persons with disabilities") can be partly derived from Sentinel-2 satellite data, but this data source falls short for distinguishing public from private green spaces as it fails to capture fine-grained urban structures due to its 10 metres resolution (Ludwig et al., 2021). Global Building Morphology Indicators could allow us to quantify the form of urban areas and enable novel analyses comparing morphological parameters across cities which go much further than existing approaches which are often based on aggregated population data (Biljecki and Chow, 2022). In addition to these examples of urban research using global building footprint data, Table 1 provides recent policy examples and practical applications of global urban analysis using building footprint data.

It has been shown that open data communities — such as OpenStreetMap (OSM) — are already contributing to filling existing data gaps (Fritz et al., 2019; Herfort et al., 2021; Scholz et al., 2018). OSM is now used widely for applications such as web maps and navigation services and OSM data has been used in domains such as urban planning (Milojevic-Dupont et al., 2020), SDG monitoring (Van Den Hoek et al., 2021), disaster management (Feldmeyer et al., 2021; Scholz et al., 2018), public health (Bhatia et al., 2018; Sturrock et al., 2018; Yeboah et al., 2021), as well as for supporting crisis response during the COVID-19 pandemic (Marco Minghini et al., 2020). In addition to contributions by individual volunteers (mappers), there is an intensifying trend of organized corporate and humanitarian mapping communities contributing to OSM (Anderson et al., 2019; Herfort et al., 2021). So far, corporations have predominantly focused on mapping roads in OSM, but they are also interested in building data, as testified by current efforts towards automated identification based on machine learning, such as the Microsoft Building open dataset (Microsoft, 2022). As these datasets have been generated outside OSM but released with an OSM-compatible format, the two data sources could be feasibly combined (Meta, 2022a). In contrast, data on roads have been predominantly directly edited within OSM by corporations, as they require significantly more effort to conflate, needing to maintain referential integrity and a topological network to operate properly.

However, particular attention needs to be paid to data inequalities and biases when OSM data is utilized in global urban studies or to derive global data products to inform policy and decision making. Since the OSM data is generated by volunteers who are not evenly distributed around the globe, their mapping efforts in OSM have been strongly biased towards the Global North (Sui et al., 2013; Neis et al., 2013), even if the humanitarian mapping community has contributed to spread efforts towards some cities in the Global South to some extent (Herfort et al., 2021).

A common data quality requirement for many research and policy applications based on building footprint data is the completeness of the building stock for analysis purposes. This is particularly important for comparative analyses that seek to discern global urban patterns, such as for instance to derive a "global" dataset of critical infrastructure (Nirandjan et al., 2022), or to use "big data" for comparing urban morphology across the globe (Boeing, 2021). When unaccounted for, spatial bias in completeness can lead urban analysts and researchers to draw general conclusions which are only valid for well-represented (well-mapped) areas (Meyer and Pebesma, 2022). Completeness of building stock data is particularly important for ensuring equitable and fair decision making based on OSM data for the policy and practice applications of Table 1 and, as such, has a direct importance for the overarching principle of the SDGs of "leaving no one behind".

While previous studies that assessed OSM vis-à-vis authoritative data were able to provide detailed insights on the completeness of OSM in selected cities (Hecht et al., 2013; Brovelli and Zamboni, 2018), their results and methods cannot be transferred easily to other regions for which reference datasets are either missing or unavailable. To overcome dependencies on sparsely available administrative datasets, proxy data that is globally available such as remote sensing data (e.g. Nighttime Light, built-up-area, Sentinel 2 derived spectral indices) (Goldblatt et al., 2020) or population data (Zhang et al., 2022) have been suggested as a potential resource to assess and predict OSM building completeness. Nevertheless, despite the widespread usage of OSM building footprint data for urban analyses and policy making, an investigation of their completeness on the global scale has not been conducted so far.

Table 1: Policy and practice examples and applications of global urban analysis using building footprint data.

Domain	Stakeholder	Description
Urban Sustainable Development (SDG 11)	UN Habitat	Building footprints and street network layout enable the detection of informal settlements via topological analysis, which can be used to inform infrastructure extensions with a focus on low and middle income countries (Brelsford et al., 2018). UN Habitat is using this for global monitoring of the SDG Indicator 11.1.1: "Proportion of urban population living in slums, informal settlements or inadequate housing" (UN Stats, 2023a).
	UN Habitat	UN Habitat is using global building footprints to assess the spatial structure, size and population distribution of city blocks to improve walkability and access to the public transport system. This is related to global monitoring of the SDG Indicator 11.2.1: "Proportion of population that has convenient access to public transport, by sex, age and persons with disabilities" (UN Stats, 2023b).
	UN Habitat	Building footprints in combination with other features such as roads and specific land use types can be used to model land consumption rates. This is used by UN Habitat for global monitoring of the SDG Indicator 11.3.1: "Ratio of land consumption rate to population growth rate" (UN Stats, 2023c).
Public Health (SDG 3)	WorldPop	Building footprints enable a precise disaggregation of population datasets at granular spatial scale to estimate where people live at the household level. This is done in the WorldPop programme and has had numerous application cases in support of public health interventions (Boo et al., 2022).
	Clinton Health Access Initiative	In the Clinton Health Access Initiative Malaria elimination campaign, using OSM data field teams were able to get a full overview of the buildings they needed to cover for Malaria interventions such as spraying campaigns (Clinton Health Access Initiative, 2017).
	Médecins Sans Frontières (MSF)	During the 2014 Ebola response in Guinea MSF used building footprints, roads and other infrastructure from OSM to identify and support affected communities for the first time on a large scale. As of 2023 OSM is the reference geographical dataset for most of MSF operations on the ground (Lüge, 2014).
Disaster Risk Reduction and Climate Adaptation (SDG 13 and Sendai Framework)	World Bank Open Data for Resilience Initiative (OpenDRI)	Building footprints are needed to estimate exposure and vulnerability to climate-related hazards and disasters when conducting infrastructure and network risk assessment at large scales, including national and globally. OpenDRI created building footprint data in OSM and engaged local government, civil society, and the private sector in disaster risk reduction strategies with a focus in sub-saharan Africa, Asia and more recently Latin America and Caribbean (Bank, 2020).
	RiskFactor	FloodFactor uses building footprints (from OSM and other sources) to help project past, present, and future flood risk at the household level across the United States. Building footprints are utilized in a similar way to predict fire risk (FireFactor) and heat risk (HeatFactor) (Risk Factor, 2023).
	Pacific Disaster Center	The DisasterAWARE platform uses many data layers from OSM (including building footprints) to support nongovernmental and governmental organizations worldwide when planning disaster risk reduction strategies and developing integrated early warning and decision support system. (Humanitarian OpenStreetMap Team, 2021)
Humanitarian Action and Emergency Response	Copernicus Emergency Mapping Service (EMS)	Building footprints and roads are used in Copernicus EMS rapid disaster response maps and data products to provide information about affected population and infrastructure. This supports first responders in their specific crisis management tasks (European Commission, 2023).
	UN Department of Operational Support, UN Global Service Center	UN Mappers is using OSM building footprints to enrich topographic and operational data in support of UN peacekeeping and humanitarian missions worldwide (UN Mappers, 2023).
	Humanitarian OpenStreetMap Team (HOT), UN OCHA	HOT mobilizes online volunteers around the world to map buildings, roads and other features in OSM in response to disasters. Daily updated extracts from OSM are provided to first responders through UN OCHA's Humanitarian Data Exchange (HDX) platform (UN OCHA, 2023).

To fill this gap, this article investigates building data completeness and inequality in OSM for 13,189 urban centres around the world, which are home to an estimated population of 3.5 billion people (about 50% of the global population). Our spatio-temporal analysis pursues the following two research questions:

1. What is the completeness of OpenStreetMap building data in the context of global urban analysis applications?
2. How unequal is urban OpenStreetMap building data distributed within the space of a city, across continents and on the global level?

First, we propose a machine-learning regression method based on a random forest to assess OSM building completeness within 13,189 urban centers (as defined by the European Commission (Florczyk et al., 2019)). We utilize an extensive collection of open building data from commercial and authoritative sources as training data and utilize OSM full-history data for spatio-temporal data analysis on the global scale (Raifer et al., 2019). The model further relies on information obtained from remote sensing data (land cover, population distribution, night time lights), Subnational Human Development Index (SHDI), and urban road network density as predictors. Second, this paper builds upon the extensive methodological skill set developed to investigate urban segregation and transfers it towards analysing geographic inequalities within OSM. This allowed us – for the first time – to present a comprehensive assessment of the evolution of urban OSM building completeness which encompasses all data contributed to OSM since 2008.

2 Results

2.1 Urban OSM Building Completeness

Our results reveal that for 1,848 cities (14% of the analysed amount) OSM building footprint data exceeded 80% completeness. In total, these cities were home to a population of 492 Million people (16% of the global urban population). Contrary, our results show that for 9,163 cities (69% of the analysed amount and home to 48% of the global urban population) OSM building footprint data did not reach 20% completeness. Our analysis found a global average urban OSM building completeness of 24% per urban center (see Table 2). Relatively high completeness was estimated for Europe & Central Asia (71%) as well as for North America (64%). Completeness values lower than the global average were observed for the regions Latin America & Caribbean (20%), East Asia & Pacific (20%), Middle East & North Africa (12%), and South Asia (9%). The completeness value for East Asia & Pacific was strongly influenced by the fact that urban centers in China were hardly mapped, very likely because mapping in OSM is prohibited by law. Sub-Saharan Africa completeness (30%) was slightly higher than the global mean.

We found that organized humanitarian mapping activities in urban centers contributed an average of about 10% of the building footprints globally (see Table 2). However, humanitarian contributions were focused on specific regions, especially in Africa where more than 50% of all building edits in Sub-Saharan Africa were related to organized humanitarian mapping activities. Overall, organized humanitarian mapping activities were expectedly associated with lower subnational human development index values, in line with previous findings (Herfort et al., 2021). We generally found corporate mapping activity to constitute less than 2% of all building edits globally (and only about 0.1% in urban centers), a significant difference in participation from corporate mappers editing nearly 20% of the global road network as previously found (Anderson et al., 2019).

Distinguishing urban centers by SHDI also revealed dramatic differences in the temporal trajectories of completeness (see Figure 1 (b)). In general, urban centers in regions with very high SHDI had the highest levels of mapped building completeness. Surprisingly, however, there was no positive correlation between SHDI and completeness. The completeness in low SHDI urban centers was higher than the completeness of urban centers with high SHDI. Our results suggest that this was due to the positive impact of organized humanitarian mapping activities since 2015, especially on urban centers located in low and medium SHDI regions (see Table 2).

The size of the urban centers measured by population was positively correlated to completeness (see Table 2), albeit the differences were not as pronounced as for world regions or SHDI classes. OSM

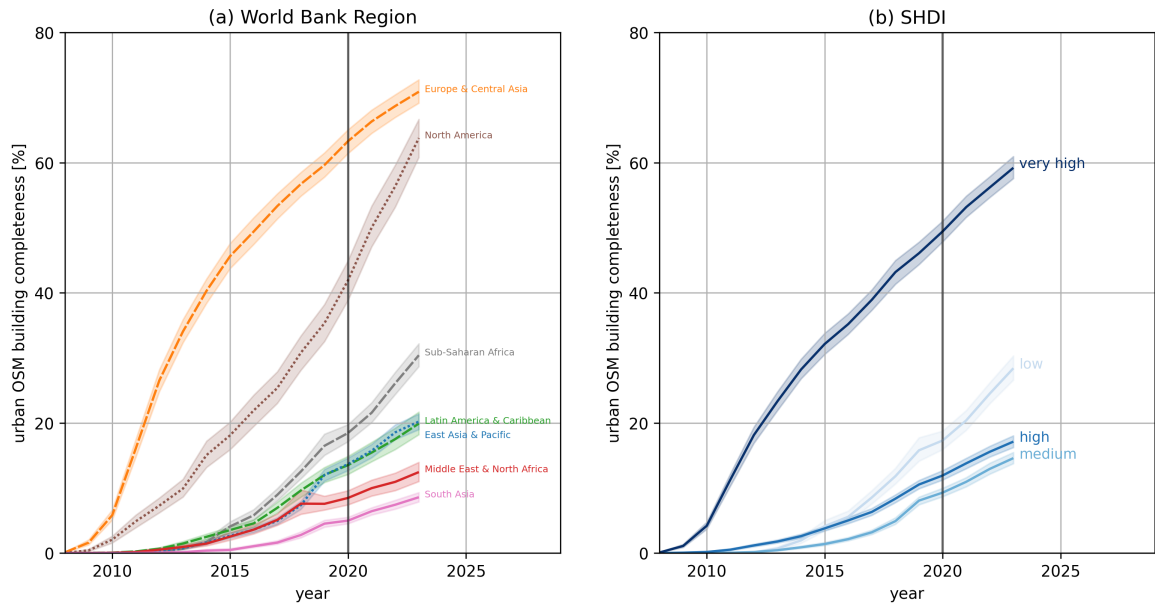


Figure 1: Temporal evolution of average urban OSM building completeness by (a) world regions and (b) Subnational Human Development Index (SHDI) group. Completeness was derived by aggregating building area predictions based on a Random Forests model and annual OSM building area per urban center. The shaded areas represent the 95% confidence interval for each line. SHDI classes were based on cut-off points defined by the United Nations Development Programme (United Nations Development Programme, 2019): low human development (SHDI < 0.550), medium human development (SHDI: 0.550 - 0.699), high human development (SHDI: 0.700–0.799), very high human development (SHDI > 0.800). OSM data from 2008-01-01 to 2023-01-01. Building area predictions were based on explanatory data for 2020. Therefore, the uncertainty of building completeness estimates rises with increasing distance to 2020. This is not reflected in the confidence bands as this additional uncertainty is hard to quantify. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

Table 2: OSM building completeness in urban centers on the global scale and grouped by world regions, Subnational Human Development Index class and city size class measured by population. Completeness was computed as the average of the individual OSM building completeness values per urban center. Humanitarian mapping and corporate mapping were quantified by their share on the overall map data. SHDI classes were based on cut-off points defined by the United Nations Development Programme (United Nations Development Programme, 2019): low human development (SHDI < 0.550), medium human development (SHDI: 0.550 - 0.699), high human development (SHDI: 0.700–0.799), very high human development (SHDI > 0.800). City size classes were based on population thresholds defined by OECD (OECD, 2016): small urban areas (50k–200k), medium-size urban areas (200k–500k), metropolitan areas (500k–1.5M), large metropolitan areas (>1.5M).

	n	Completeness [%]	Humanitarian Mapping [%]	Corporate Mapping [%]
Global	13,189	24	10.0	0.1
World Regions				
East Asia & Pacific	3,068	20	13.4	0.2
Europe & Central Asia	1,351	71	1.1	0.1
Latin America & Caribbean	1,073	20	15.2	0.4
Middle East & North Africa	901	12	16.9	0.2
North America	378	64	0.2	<0.1
South Asia	3,997	9	17.8	<0.1
Sub-Saharan Africa	2,421	30	51.1	0.1
Subnational Human Development Index				
Low	2,289	28	52.3	0.1
Medium	4,960	15	30.3	0.3
High	3,883	17	15.8	0.2
Very High	1,967	59	2.1	<0.1
City Size by Population				
Small Urban Areas	10,930	23	10.1	0.1
Medium-Size Urban Areas	1,348	32	9.5	<0.1
Metropolitan Areas	563	37	9.4	0.1
Large Metropolitan Areas	287	41	10.4	0.1

building data in large metropolitan areas were considerably more complete compared to small urban areas. However, the temporal evolution of urban building completeness showed very similar patterns for urban centers regardless of their population (see Figure 9 in Supplementary Material section).

The spatial distribution of building completeness across urban centers shows a strong regional variability across that global trend: numerous cities in any region were mapped with a very high completeness regardless of the overall completeness or mapping activity in that region (see Figure 2). For instance, within Africa, we found urban centers in Egypt and Ethiopia with particularly low OSM building completeness, whereas cities in Tanzania, Uganda and western African countries achieved much higher completeness. Similarly, building completeness values in Indonesia and the Philippines were notably higher than for other countries of Southeast Asia. In contrast, most urban centers in India and China were hardly mapped with regard to building footprints. Strikingly, the spatial distribution of OSM building completeness for urban centers was characterized by spatially clustered patterns at various scales.

The uneven building completeness between urban centers was also indicated by a global Gini coefficient of 0.8. This characteristic was observed across all regions and was most pronounced in South Asia and Sub-Saharan Africa (c.f. Figure 3 a). The temporal evolution of the Gini coefficient indicated that both globally and regionally, OSM building data distribution has become slightly more even over time. In contrast, we find that global spatial inequality in OSM building completeness sharply increased between 2008 and 2014 (c.f. Figure 3 b). During that time—although overall OSM building completeness became more even (as measured by the Gini coefficient)—mappers favoured cities near already well-mapped cities. We also find that until 2014, the expansion of OSM mapping to distant and un-mapped regions (likely to be located in the Global South) did not happen at a significant scale.

Nevertheless, since 2014, Moran’s I as a measure of global spatial autocorrelation declined from 0.71 till 0.56 as of 2023. This indicated that the spatially clustered completeness pattern became less intense, albeit still clearly visible. Combined with the decrease of the Gini coefficient in the same period, our results suggests OSM building data in 2023 was much less segregated in both evenness and clustering

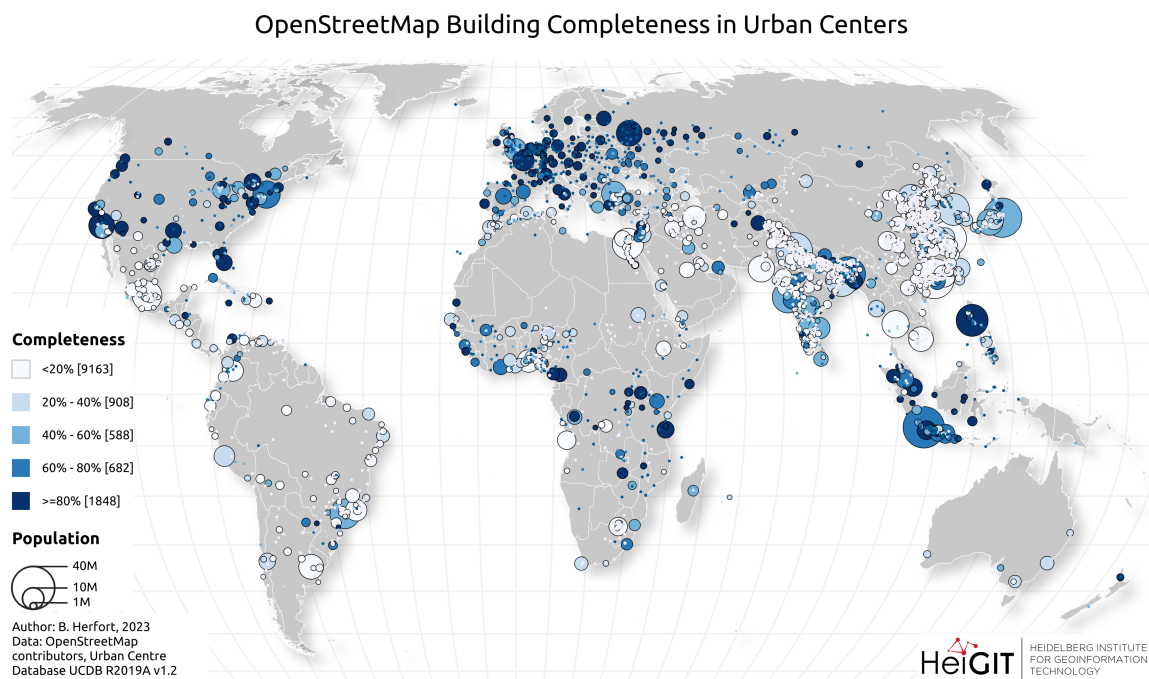


Figure 2: Spatial distribution of OSM building completeness in 13,189 urban centers. For each class the overall number of urban centers is reported in the squared bracket. For an interactive web map visualization visit https://hex.ohsome.org/#/urban_building_completeness. OSM data as of 2023-01-01. Created using QGIS 3.28.3 (<https://www.qgis.org/en/site/>).

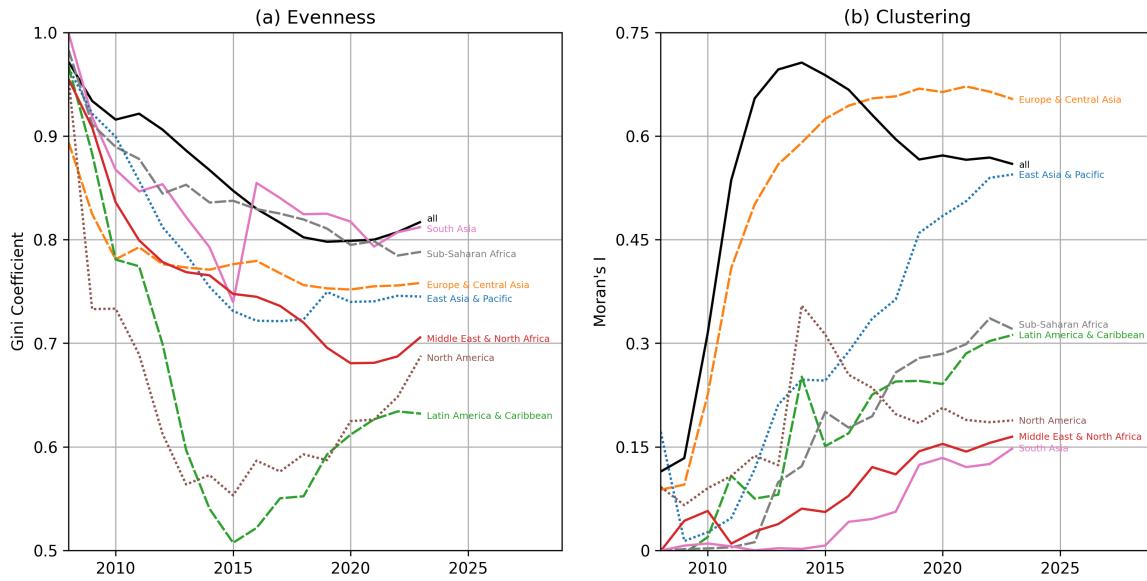


Figure 3: Non-spatial and spatial inequality measures of completeness. Temporal evolution of (a) evenness and (d) clustering of urban OSM building completeness per world region. Moran's I measures spatial autocorrelation, positive values indicate spatial clustering. Values for Moran's I in practice often range between -0.5 and 1.15 with zero indicating absence of global spatial autocorrelation. OSM data from 2008-01-01 to 2023-01-01. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

compared to the state-of-the-map in 2014. Attention should be paid to the fact that Gini coefficient and Moran's I have been stagnating on the global level since 2019, which might indicate a shift in mapping behaviour due to restrictions caused by the COVID-19 pandemic. However, since 2021 Gini coefficient started to increase again, implying a tendency towards a more uneven distribution of building mapping across urban centers on the global level.

In support of the global pattern, we also found spatial autocorrelation within regions to increase steadily over time regardless of overall map completeness. Europe & Central Asia reveal a moderate spatial clustering (Moran's I: 0.23) in 2010, but a very high spatial clustering (Moran's I: 0.65) in 2023 (see Figure 4 (a) and (b)). In 2023, urban centers with high completeness in Europe & Central Asia were surrounded by other urban centers with high completeness and this effect was much stronger than in 2014. An analogous process was observed in all other regions, e.g. as shown for Sub-Saharan Africa in Figure 4 (c) and (d). The only exception to this is North America, where spatial clustering of building completeness has decreased since 2014.

2.2 Intra-Urban OSM Building Completeness

This section builds upon the intra-urban spatial heterogeneity of OSM building completeness estimated at a resolution of 1x1 kilometers. Here we will refer to each 1x1 km section of the map as a "grid cell". These grid cells were used to calculate the Gini coefficient and Moran's I for individual urban centers. Based on both those indicators and OSM building completeness, urban centers were classified into three different types (and dividing into two additional sub-types) utilizing an agglomerative clustering approach (c.f. Figure 5 and c.f. Figure 6). Figure 7 shows the temporal evolution of OSM building completeness per urban center and highlights the trajectory for each cluster representative. The spatial distribution of urban centers based on the intra-urban inequality (Figure 11 Supplementary Materials section) shows a similar spatial pattern as Figure 2.

Urban centers of type (1) (see cluster dendrogram in Figure 6) showed very low completeness and

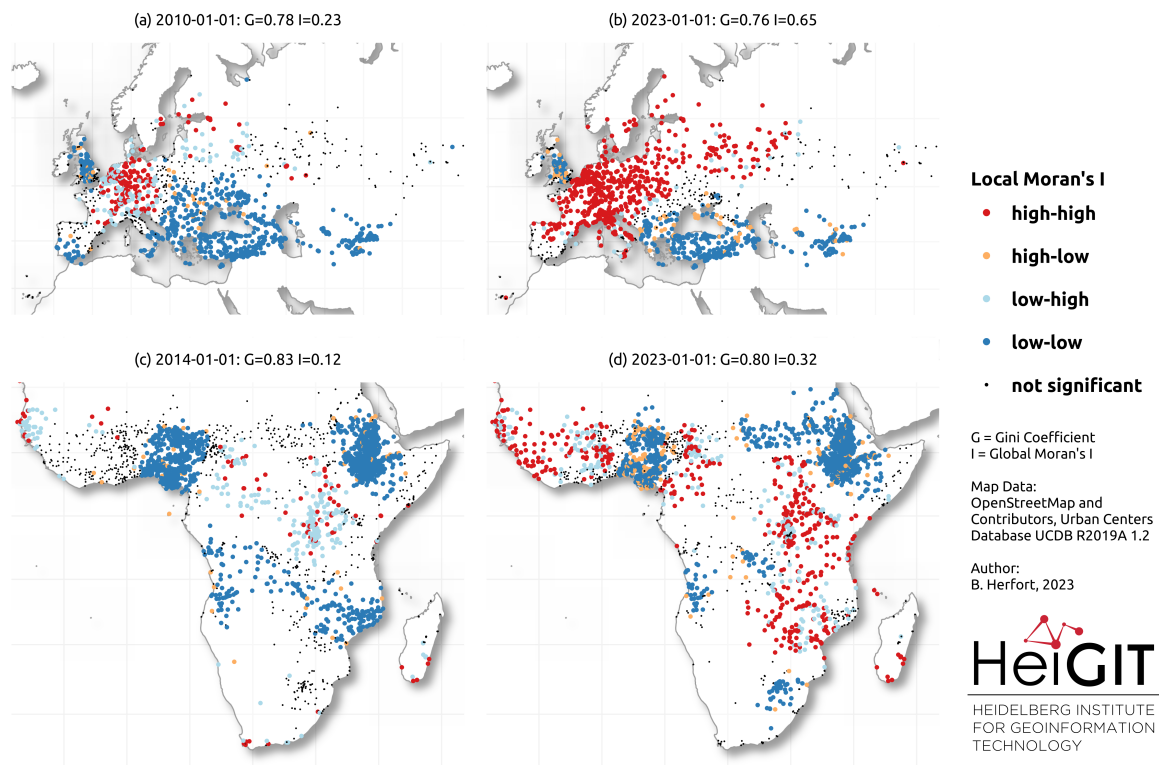


Figure 4: A comparison of local spatial autocorrelation of completeness at two points in time for urban centers within (a)-(b) Europe & Central Asia and (c)-(d) Sub-Saharan Africa. Each urban center was classified according to whether its building completeness value was above (high) or below (low) the global mean and if the weighted mean across its neighbors was above or below the global mean. Based on this four quadrants are defined: high-high (HH), low-high (LH), low-low (LL) and high-low (HL). High-high describes clusters of high completeness values, low-low describes clusters of low completeness values while low-high and high-low indicate spatial outliers in the sense that the completeness value of the urban area was unexpected in their neighborhood. Significance levels were adjusted for multiple testing. For each region and point in time we provide the Gini coefficient (G) and Moran's I for the region shown in the sub-plot. Created using QGIS 3.28.3 (<https://www.qgis.org/en/site/>).

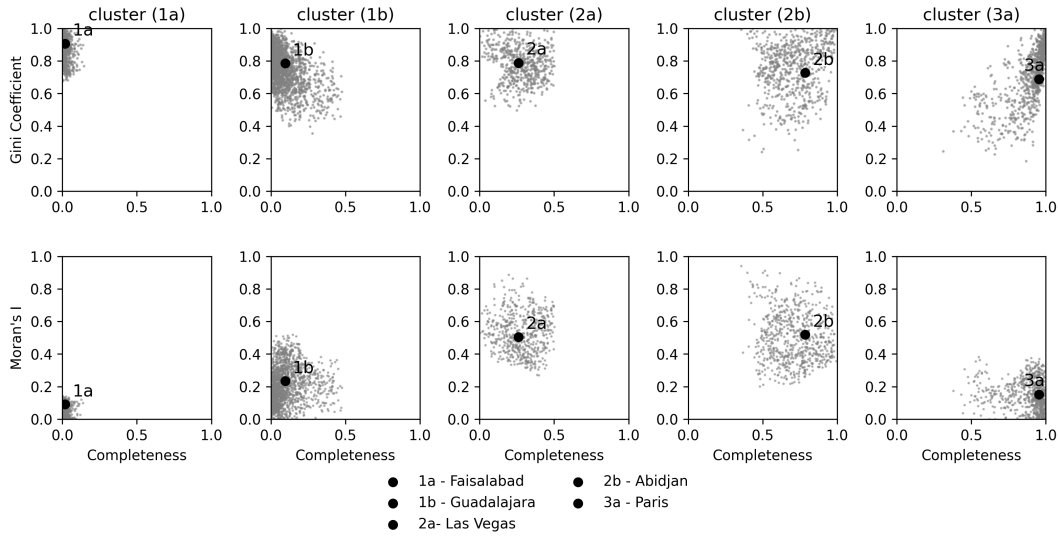


Figure 5: Agglomerative clustering of urban centers based on OSM building completeness, Gini coefficient G and Moran's I . Each point represents a single urban center with a minimum area of 25 square kilometers. Smaller urban centers were ignored as Gini coefficient and Moran's I could not be reliably estimated. For each of the clusters a single representative example was selected out of the 4,647 urban centers considered in this analysis. OSM data as of 2023-01-01. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

could be further distinguished into subtypes (1a) and (1b) respectively. Urban centers of type (1a) could appear as white spots on the map, more aptly described as the "unmapped" cities and are characterized by a high Gini coefficient and very low Moran's I value. There was no particular spatial pattern defining where the small number of eventually mapped buildings would be located within the city. Among 1,692 urban centers, the city of Faisalabad, Pakistan is shown in Figure 6 (1a) as an example of this category. Urban centers of type (1b) could be considered hardly mapped as well, but there were often a few grid cells which have been mapped with a much higher completeness in regard to buildings. Such mapped grid cells were not distributed randomly, but tended to cluster spatially as indicated by a higher Moran's I value. The urban agglomeration of Guadalajara, Mexico (see Figure 6 (1b)) exemplifies that there were several distinct mapping hot spots surrounded by a larger number of unmapped grid cells.

The second group of urban centers is characterized by relatively high Moran's I and moderate to high evenness covering cities with wide range of completeness. For urban centers of type (2), the overall completeness value hardly reflected the local completeness values. The urban centers of type (2a) such as Las Vegas, USA (see Figure 6) could be considered a divided city from the perspective of mapped building in OSM. This pattern is characterized by a few spatially clustered neighbourhoods which are mapped very well, whereas large parts of the city remain unmapped. Also for urban centers of type (2b) a highly segregated spatial distribution of OSM building completeness was observed. For these cities, such as Abidjan, Ivory Coast, there exist large blocks of completely mapped grid cells, however still entire neighbourhoods of the city are missing from the map.

Finally, urban centers with the highest overall completeness and very low spatial clustering were most likely to get classified as type (3). As shown in Paris, France (see Figure 6 (3a)), almost all parts of the city may be considered completely mapped. Only a few grid cells remained unmapped and these are often not strongly spatially clustered. Figure 7 highlights that for the case of Paris between 2010 and 2013 almost 80% of the entire building stock was added to OSM. As opposed to type (1), the urban centers of type (3) could be described as the "well-mapped" cities and represent some finality of mapping building footprints in OSM.

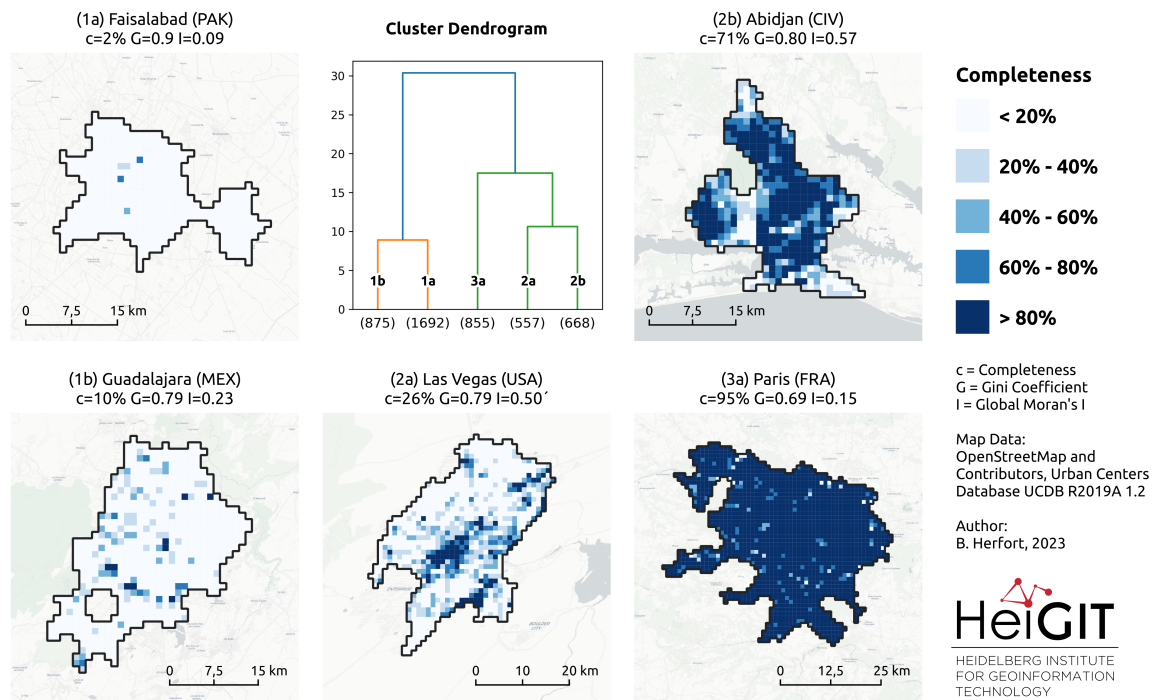


Figure 6: Spatial distribution of intra-urban OSM building completeness for selected urban centers. For each urban center we report on overall OSM completeness c , Gini coefficient G and Moran's I . Cell size is always 1x1 kilometers for any urban center. The clusters are the same as in figure 5. The number of urban centers in each cluster is indicated in the dendrogram. For an interactive web map visualization visit https://hex.ohsome.org/#/urban_building_completeness. OSM data as of 2023-01-01. Created using QGIS 3.28.3 (<https://www.qgis.org/en/site/>) and Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

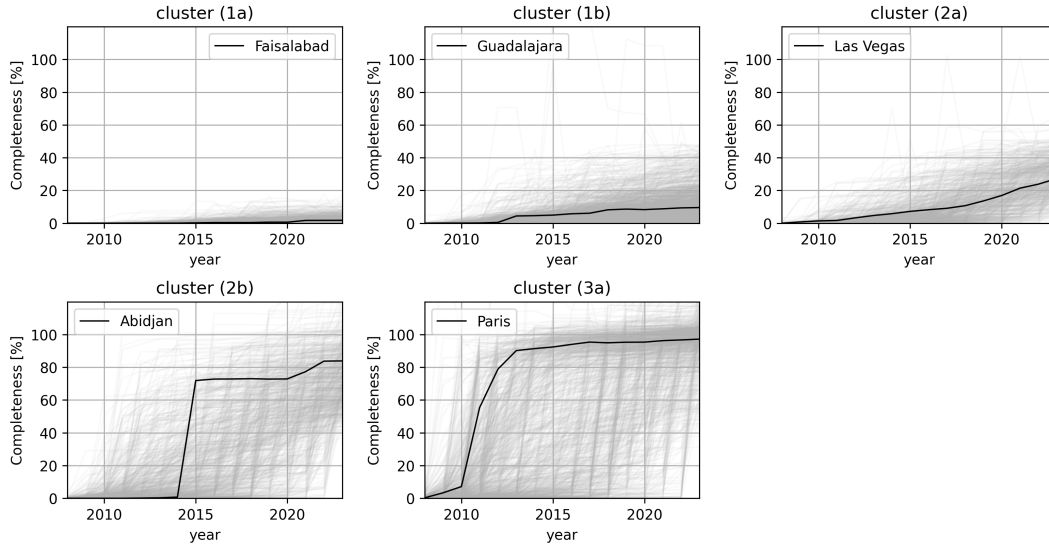


Figure 7: Temporal evolution of urban OSM building completeness per cluster and for selected urban centers. The clusters are the same as in figure 5. OSM data from 2008-01-01 to 2023-01-01. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

3 Discussion

Mapping efforts of communities in OSM over the previous decade have made OSM a unique global database of building footprint data, which is accessible to all with no licensing costs. Besides the well-examined Global North - Global South bias in OSM, our results demonstrate that the OSM building stock as of 2023 showed a much more spatially diverse pattern than previously considered, which was shaped by regional, socio-economic and demographic factors across several scales. The results confirm the similar, albeit less dichotomous, global completeness pattern of land use land cover (LULC) information from OSM (Zhou et al., 2022a) and stand in slight contrast to the relatively higher and more evenly distributed completeness for OSM’s road network (Barrington-Leigh and Millard-Ball, 2017). This highlights the complexity of OSM mapping activities and the challenge for urban analyses based on OSM data. Thus, we want to provide two important recommendations for OSM data users and producers to pick the best strategies to cope with OSM’s uneven spatial coverage and to promote sustained impact of future mapping efforts.

First, the unequal patterns of completeness we evidenced make it essential for urban analysts to assess the potential negative impact of missing data, i.e., OSM data users should investigate if the intended urban analysis is subject to spatial bias caused by OSM’s uneven spatial coverage at multiple scales. To support this assessment, we provide an open dataset (Herfort, 2023) resulting from this study with estimated completeness maps for 13,189 urban agglomerations worldwide using a 1x1 kilometer grid, which enables the assessment of variations within and across urban centres.

We encourage researchers and practitioners to further develop this datasets and the other data quality assessment methods which may help to quantify the highly uneven geographies of participation in communities such as OSM or as observed in Wikipedia (Graham et al., 2015). Global studies and global frameworks will benefit from these approaches, as researchers will be able to draw more robust conclusions and avoid misleading recommendations for decision makers once the biases in OSM’s coverage are known and can be accounted for. This will further increase the reliability and usefulness of using OSM data for global urban analysis such as monitoring progress towards the SDG’s (Van Den Hoek et al., 2021), as

well as for humanitarian activities, including disaster risk reduction (Scholz et al., 2018).

Second, the global community of OSM data producers will be able to use our completeness maps as a guidance to plan where future mapping activities should take place to improve coverage so that “no one is left behind” as encouraged by the SDGs. By combining the completeness maps with socio-demographic characteristics of the areas of interest, it will be possible to ensure a fair and balanced selection of target geographic areas to reduce existing data inequalities within OSM on the global, regional or intra-urban level. Our open dataset about urban OSM building completeness can inform this process. In light of the critical challenges to finance high-quality data systems for addressing inequalities in SDG monitoring in both low and middle income countries (Ulbrich et al., 2019)—despite heightened demand (Sachs et al., 2022)—the creation and usage of OSM data could be promoted further as a cost-effective alternative. Data generation with OSM not only allows filling the current data gaps essential to monitor progress, but can also be an pathway for equitable urban transformations by empowering local communities to have a voice and benefit from the data production process (Porto de Albuquerque et al., 2021). The cultural openness and social nature of OSM is a clear strength to achieve transparency about existing inequalities and how to address them, especially in comparison with building footprint datasets that are derived using proprietary, black-box machine learning approaches for which bias and fairness are often still unknown (Mehrabi et al., 2021).

However, our analysis also comes with unavoidable limitations that need to be considered. One major limitation is that the work presented here only investigated buildings mapped in OSM within urban centers. Even if our study encompasses about 50% of the global population which lives in urban agglomerations, one should be careful to transfer our findings to rural areas. Researchers have shown that there is a tendency of OSM data to be of higher quality in cities (Hecht et al., 2013). On the other hand, rural areas that have been the target of humanitarian mapping campaigns will likely have higher completeness than surrounding areas (Anderson et al., 2018).

The results of machine learning models, such as the random forest model used by us to derive building area predictions, are conditioned by potential biases present in the training data or biases that arise from the algorithms (Mehrabi et al., 2021). For the authoritative data utilized to train the model, a high geometric accuracy was assumed, however, these datasets might be outdated depending on the publisher’s update cycle. The quality of building footprint data from Microsoft (c.f. Supplementary Material Table 7) showed a sufficient recall of more than 80% for urban centers in all regions. However, self-declared recall values by Microsoft have been considerably lower for some regions. This indicates that recall might be lower in rural areas, which have not been included into our study. Nevertheless, the remaining biases of the Microsoft buildings dataset for urban areas are also reflected in the results reported here and might lead to extra low building area predictions and consequently over-estimated OSM building completeness. We were also not able to quantify the uncertainty for countries with a large number of urban centers (e.g. China) for which training data was not available. Nevertheless, other authors have highlighted that rapidly urbanizing cities, for instance in China (Cervero and Day, 2008), are “mimicking [suburbanisation] trends and patterns of the post-World-War-II United States”. As such, we do not expect severe structural deviations in respect to the modelled relationships between explanatory variables and building area prediction for these countries. The low feature importance for the geographic world region (i.e., variable “World Bank Region Code”) seems to be in line with that assumption.

To be transparent about model uncertainty, we reported on the model performance utilizing a spatial cross validation procedure (c.f. Table 4). While the completeness estimation performed well with a global r^2 score of 0.9, slightly higher uncertainty was observed for Sub-Saharan Africa. However, this may constitute a starting point for local communities and researchers to design local completeness models that can overcome the limits of the global modelling approach used in this study, even if these local models might not be easily transferred to other regions.

Our findings extend the general pattern of urban OSM building completeness as of 2020-01 (Zhou et al., 2022b) a) by highlighting the temporal evolution of mapping activity in conjunction with other events such as the COVID-19 pandemic, b) by considering a spatially much more balanced and extensive training data set from various sources and c) by investigating the global pattern and consequences of inequalities in completeness. In addition, geospatial data quality is comprised of dimensions beyond measuring completeness (Oort, 2006). For some sectors, such as public health programs, assessment of completeness is only the first step, and information on building usage is also required, but often only

available for a small subset of buildings (Sturrock et al., 2018). It has been shown for land cover and land use (LULC) in OSM that the spatial pattern for completeness and accuracy are not necessarily the same (Zhou et al., 2022a). Whereas urban OSM building completeness and OSM LULC completeness show similar global trends on the national level, OSM building data accuracy needs to be further investigated, e.g. towards its potential utilization as training and/or validation samples in machine learning models. A more detailed analysis of the temporal trajectories for urban centers would facilitate this investigation and might reveal to what extent data quality dimensions, temporal evolution of mapping activity and inequality measures are mutually dependent.

Future work should further investigate the potential of a harmonious ensemble dataset that combines the best of OpenStreetMap buildings with additional building coverage from deep learning based datasets such from Microsoft Buildings (Microsoft, 2022). We have performed the direct comparison between OSM and Microsoft buildings and report the findings in the Supplementary Materials section (Table 5 and Figure 10). Ultimately, OSM buildings and MS buildings represent two very different types of data, and the major advantage of OSM is that it is continuously verified by human editors. This is the reason why not all of the Microsoft buildings are in OSM, as the Microsoft buildings represent a different type of dataset: algorithmically extracted from aerial imagery using automated methods. For some places, the building footprints in OSM might already come from an imported dataset, but they need to be accepted by a human-in-the-loop process, for instance using the mapwith.ai editor (Meta, 2022b) or similar tools. A large direct import of Microsoft building data into OSM seems unlikely at the moment, however it has been shown for the road network that imports can increase contributor activity especially for already engaged mappers and an interplay of data imports and updates by contributors could improve OSM data significantly (Witt et al., 2021; Zielstra et al., 2013).

The OSM community started its journey in the early 2000's in western European cities and this history is still clearly visible in the unbalanced spatial distribution of map data in 2023. Nevertheless, there are numerous successful examples of local mapping communities that are overcoming structural barriers which exclude others from participating in OSM. We believe that by empowering these communities, OSM will further evolve into the most comprehensive open geographic data base which is needed to help achieve the SDGs and ensure equitable and sustainable urban futures.

4 Methods

4.1 Building Data

The analysis was carried out for 13,189 urban centers on a global scale. To delineate our study areas we used the Global Human Settlement Layer Urban Centred Database (GHS-UCBD) which has been developed by the European Commission (Florczyk et al., 2019). Accordingly urban centers have been characterized as "high-density clusters of contiguous grid cells of 1 km² with a density of at least 1500 inhabitants per km² and a minimum population of 50000" (Florczyk et al., 2019). Each urban center was spatially disaggregated using a 1x1 kilometer grid based on the equal-area Mollweide projection. The grid adopted the same structure utilized by the raster datasets of the GHS-UCBD. For each of the resulting 665,641 grid cells we aggregated both the reference data sets (if available) and the datasets utilized as predictors in the model.

For each grid cell we derived the overall OSM building footprint area in square kilometers using the `osmapi` API which relies on the OSHDB framework for spatio-temporal analysis of OSM history data (Raifer et al., 2019). We included buildings which have been mapped in OSM as of 2023-01-01. We considered all OSM objects that have been tagged with `'building=*'` and were of the geometry type `'Polygon'`.

As no single reference building data set on the global scale exist, we combined a set of external datasets (c.f. Table 6) which have been obtained either from authoritative or commercial sources building upon the great work by Biljecki et al. (2021). For each grid cell we derived the overall building footprint area in square kilometers for each of the reference datasets by intersecting the grid cells and building footprints and then summing up the corresponding surface area of all building (parts) per grid cell. In case two reference datasets were available, e.g. from Microsoft and an authoritative source, we considered the information from the authoritative dataset. In total these reference datasets covered 6,633 urban centers

(404,982 grid cells) among 162 countries. For some regions data was not available for the entire country, but only selected cities. The data obtained from Microsoft was derived using a deep learning based building detection approach (Microsoft, 2022).

The Geo-Wiki built-up reference dataset (See et al., 2022) has been utilized to assess the suitability of the Microsoft building footprint dataset for our OSM completeness modelling approach. The Geo-Wiki campaign visually assessed very high-resolution satellite images of 50 K sample locations for the presence of built-up surfaces containing any building with a roof using a crowdsourcing approach (See et al., 2022). All Geo-Wiki grid cells intersecting the urban center geometries for which Microsoft building data was available at the city level were considered in the analysis. Precision and recall for the Microsoft building footprints have been derived on the 10x10 meters Geo-Wiki grid level (see Supplementary Material Section Table 7). Grid cells were defined as true positives (TP) if at least one Microsoft building footprint intersected with a Geo-Wiki grid cell labelled as 'Built-up'. True negatives (TN) were defined as Geo-Wiki grid cells labelled as 'Not built-up' that did not intersect with any Microsoft building footprint. Accordingly we defined false positives (FP) as grid cell labelled as 'Not built-up' which intersected with at least one Microsoft building and false negatives (FN) as the reverse. Finally, we also provide precision and recall for the Microsoft building footprints as self-declared by Microsoft in the Global ML Building Footprints GitHub repository (see Supplementary Material Section Table 8).

4.2 Explanatory Variables

As explanatory variables we used the following datasets (see Table 3): the Global Human Settlement Layer Population (GHS-POP) is provided by the European Commission (Freire et al., 2016) and based on a disaggregation of CIESIN's Gridded Population of the World (GPWv4.10). To characterize regions based on their socio-economic status we utilized the Subnational Human Development Database (Smits and Permanyer, 2019). In our analysis we relied on the aggregated SHDI and did not further consider its individual components (education, standard of living, health). Information on night-time lights was obtained as the annual average of 2020 and aggregated by summing up all values per grid cell (Elvidge et al., 2021). Land cover information at 10 meter resolution was utilized from the ESA WorldCover 2020 dataset which has been derived from Sentinel-1 and Sentinel-2 data (Zanaga et al., 2021). For each grid cell we derived the overall area per land cover class in square kilometers.

From OSM we extracted the road network length in kilometers per grid cell for main roads. The main roads were selected using the `ohsome API` (Raifer et al., 2019) using the following filter: 'highway in (primary, primary_link, secondary, secondary_link, tertiary, tertiary_link, unclassified, residential)'. We included data which have been mapped in OSM as of 2023-01-01. We have investigated the spatial variations in the completeness of OSM road data by utilizing an intrinsic quality assessment approach following the 'mapping saturation' methodology proposed by Rehrl and Gröchenig (2016). Accordingly, the completeness of the road network was estimated for each urban center and aggregated by world region (see Table 9). The intrinsic completeness measure revealed that for a small, but still decent share of the urban centers in South Asia (10.5%), Middle East & North Africa (10.2%) and Sub-Saharan Africa (5.9%) road network mapping could be considered only in the initial stage. For these urban centers it is very likely that the majority of the road network was not (yet) completely mapped.

The data values for ESA WorldCover, GHS-POP and VNL are subject to uncertainty, as they reflect the situation for the year 2020, albeit they are utilized in a model to assess OSM building completeness for a time range from 2008 to 2023. The model will be most suited to predict the building stock of 2020 and not the past building stock, nor the correct building stock for 2023. As a consequence, the analysis will slightly overestimate completeness for regions which have seen rapid urbanisation since 2020, but might also underestimate completeness for timestamps before 2020.

Initially, we considered additional explanatory variables (e.g. permanent water bodies, fossil fuel consumption, OSM railway length, OSM amenity count), but these have been disregarded as their feature importance in the Random Forest model turned out to be very low (<0.02).

4.3 Building Area Prediction Model

We used a Random Forest (RF) regressor (Breiman, 2001) to predict the overall building area in square meters per 1x1 kilometer grid cell using the covariates described in the section above. Building complete-

Table 3: Summary statistics and the impurity based feature importance (Gini importance) for explanatory variables utilized in the machine learning model. World Bank Region code is a categorical variables, thus summary stats are not reported.

Variable	Min	Mean	Median	SD	Max	Feature Importance
GHS-POP 2020 (Freire et al., 2016)	0.0	4484	2750	6104	615757	0.08
SHDI 2019 (Smits and Permanyer, 2019)	0.276	0.77	0.77	0.126	0.985	0.04
Average Night-Time Lights 2020 (Elvidge et al., 2021)	0	121.375	78.0	231.79	71509	0.03
WorldCover 2020 "built-up" [sqkm] (Zanaga et al., 2021)	0.0	0.403	0.376	0.27	1.005	0.68
WorldCover 2020 "tree cover" [sqkm] (Zanaga et al., 2021)	0.0	0.228	0.149	0.228	1.008	0.04
WorldCover 2020 "sparse vegetation" [sqkm] (Zanaga et al., 2021)	0.0	0.083	0.045	0.108	1.003	0.04
OSM road network length 2023 [km]	0.0	9.106	8.133	6.453	57.635	0.06
World Bank Region Code [categorical]	-	-	-	-	-	0.03

ness is not directly predicted, but inferred in a second step using this prediction and the corresponding surface area of all OSM buildings per grid cell. There are various applications of Random Forest regressors for producing spatial predictions in general (Hengl et al., 2018) and to estimate building completeness in particular (e.g. in Haiti, Dominica and St. Lucia (Goldblatt et al., 2020)). RF constitutes a non-spatial approach to spatial prediction as sampling locations are ignored during the calculation of the model parameters (Hengl et al., 2018). Hence, in this study we initially also considered generalized additive models (GAMs) as an explicit spatially aware approach (if smooths of coordinates are included) which has been used for geospatial modelling e.g. in the domains of geomorphology (Goetz et al., 2011) or for the analysis of social media data (de Albuquerque et al., 2015). Nevertheless, our results revealed that RF outperformed the GAM approach and we decided to utilize the RF implementation in the Python package 'scikit-learn' (Pedregosa et al., 2011).

To evaluate the performance of the proposed building area prediction we adopted a spatial cross validation approach based on k-means clustering. Especially for large scale mapping studies, such as this work, but also in the domain of ecological modelling data are almost always spatially autocorrelated and a spatially explicit assessment of machine learning models is required (Ploton et al., 2020). Due to spatial autocorrelation in the observations (data from nearby locations will not be independent) training samples and validation samples cannot be randomly selected as this would lead to overly optimistic error estimates (Roberts et al., 2017). Spatial blocking of samples, e.g. through k-means clustering, decreases this spatial dependence and provides more realistic performance scores (Roberts et al., 2017; Brenning, 2012). Our spatial cross validation blocks were derived using a 20-fold k-means clustering based on scikit-learn's python implementation (Pedregosa et al., 2011).

We investigated the performance of our model in respect to building area prediction for the 1x1 km grid and in respect to OSM building completeness prediction for the urban centers using the following indicators: r^2 score, explained variance, mean squared error (MSE) and mean absolute error (MAE). On the grid level, MAE describes the average of the absolute differences between predicted building area and reference building area obtained from authoritative sources or Microsoft's Global ML Building footprints in square kilometers. Accordingly, MSE describes the average of the squared differences between predicted building area and reference building area per grid cell. On the urban centers level, MAE describes the average of the absolute differences between predicted building completeness and reference building completeness. Accordingly, MSE describes the average of the squared differences between predicted building completeness and reference building completeness per urban center. These measures were calculated based on scikit-learn's python implementation (Pedregosa et al., 2011).

To report global performance, we first estimated the model performance metrics for each of the seven regions (see Table 6). As each of these regions contained different numbers of samples, we computed the global scores from the weighted average of the regional scores using the total number of samples per region as the weight. Table 4 provides the performance scores. Overall the OSM building completeness model performed with a mean absolute error of 0.067 and achieved a r^2 score of 0.84 and explained variance of 0.85. In addition, we checked for spatial clustering in the residuals of the OSM building completeness prediction utilizing Moran's I as a measure of spatial autocorrelation (Cliff and Ord, 1981). The residuals were not distributed entirely random across space, but nevertheless only showed a slight tendency to cluster (Moran's I: 0.29).

For all regions the distribution of raw residuals resembled a normal distribution (c.f. Figure 8). The histogram of raw residuals revealed that for samples located in Sub-Saharan Africa and North America

Table 4: Global and regional model performance measures based upon 20-fold spatial cross validation.

region	n	r2	exp var	MSE	MAE
Building Area Prediction (1km Grid)					
Global	403,357	0.74	0.74	0.0025	0.034
East Asia & Pacific	48,389	0.72	0.73	0.0028	0.038
Europe & Central Asia	76,714	0.73	0.73	0.0019	0.029
Latin America & Caribbean	51,792	0.70	0.70	0.0047	0.047
Middle East & North Africa	31,748	0.77	0.77	0.0027	0.037
North America	99,064	0.70	0.70	0.0012	0.025
South Asia	55,743	0.84	0.84	0.0022	0.030
Sub-Saharan Africa	39,907	0.67	0.67	0.0024	0.035
OSM Building Completeness Prediction (Urban Centers)					
Global	6,553	0.9	0.9	0.012	0.055
East Asia & Pacific	483	0.93	0.93	0.009	0.050
Europe & Central Asia	1,103	0.85	0.85	0.016	0.089
Latin America & Caribbean	874	0.88	0.88	0.011	0.044
Middle East & North Africa	663	0.90	0.90	0.005	0.024
North America	367	0.92	0.92	0.009	0.067
South Asia	1,660	0.88	0.88	0.007	0.024
Sub-Saharan Africa	1,403	0.81	0.83	0.034	0.091

the distribution was slightly negatively skewed and had a weak tendency to predict too low completeness values for urban centers. For some regions, e.g. East Asia & Pacific or Sub-Saharan Africa, we observed heteroscedasticity in the distribution of residuals for urban centers. This means that the higher the predicted completeness value, the larger the residual and thus the uncertainty of our model. Still, our predictions are conservative in the sense that they are rather underestimating completeness for urban centers.

4.4 Urban OSM Building Completeness

For each urban center we calculated the OSM building completeness using the area ratio method (reference building area / OSM building area) which has been applied by several other researchers in the context of urban areas (Fan et al., 2014; Hecht et al., 2013). We did not consider the building count, but building area instead, to account for the high sensitivity to disparities in modelling when using unit based completeness measures (Hecht et al., 2013), which is especially important when using a wide range of different building datasets as in this study. For instance, some datasets, such as OS OpenMap Local, model terraced housing as a single polygon, whereas in OSM these buildings are usually mapped as individual geometrical features. Whereas this would result only in a slight difference in the surface area of all buildings per grid cell, the building count can vary dramatically and completeness might be overestimated.

First, we obtained the overall predicted building area by summing up the values for all grid cells per urban center. OSM building completeness per urban center was derived annually by computing the ratio of OSM building area versus predicted building area. We report on the average monthly OSM building completeness for urban centers globally and distinguished this score further by world region, SHDI class and city size class by population. For the spatial aggregation we used the regions defined by World Bank and will refer to them also as 'world regions' in the manuscript. In addition, 95% confidence intervals have been calculated for each time series. SHDI classes were based on cut-off points defined

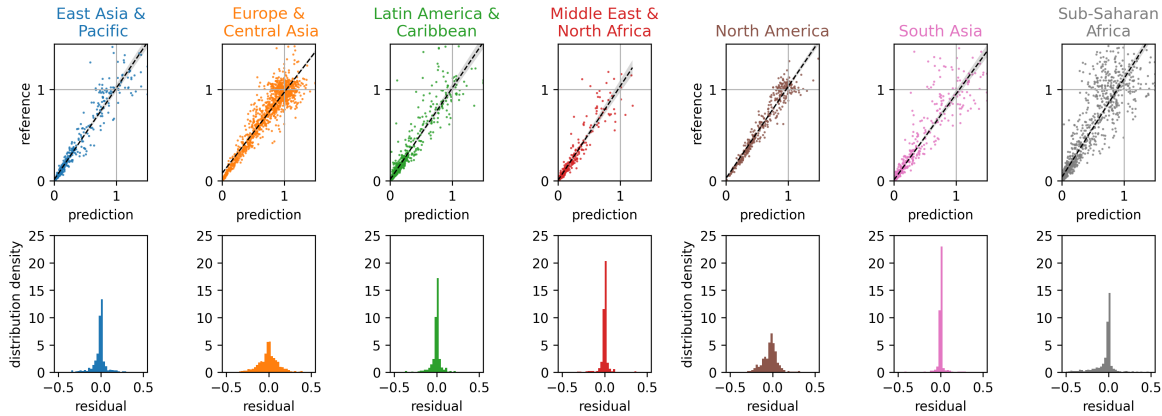


Figure 8: Scatterplot of raw residuals in regard to predicted OSM building completeness and distribution density of raw residuals on the urban centers level. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

by the United Nations Development Programme (United Nations Development Programme, 2019): low human development ($\text{SHDI} < 0.550$), medium human development ($\text{SHDI}: 0.550 - 0.699$), high human development ($\text{SHDI}: 0.700 - 0.799$), very high human development ($\text{SHDI} > 0.800$). City size classes were based on population thresholds defined by OECD (OECD, 2016): small urban areas (50k–200k), medium-size urban areas (200k–500k), metropolitan areas (500k–1.5M), large metropolitan areas ($> 1.5\text{M}$).

We investigated the impact of humanitarian mapping through the HOT Tasking Manager and corporate mapping by Apple, Meta, MapBox, Microsoft and Kaart on overall completeness and inequality measures. OSM contributions have been considered as humanitarian mapping activities following the approach developed by Herfort et al. (2021) which utilizes information obtained from a HOT Tasking Manager database dump. Corporate mapping activities were identified by OSM user ID, expanding on the approach presented in Anderson et al. (2019) by using a mapper’s self-disclosed corporate affiliation in their OSM user bio instead of relying on out-of-date lists on the OSM wiki (OpenStreetMap Contributors, 2022). According to these two approaches, the contributions of each OSM user were first grouped into either “humanitarian”, “corporate” or “other” and in a second step according to world region, SDHI class and city size class by summing up the building area added to OSM per map edit. Based on this information, we derived the share of humanitarian map edits and corporate map edits on the overall OSM building data considering the building area (and not building count) as the base unit.

Several measures have been adopted to describe the temporal evolution of inequality in urban OSM building mapping on the global scale and per world region. This analysis has been conducted for annual snapshots from 2008-01-01 up until 2023-01-01. The Gini coefficient has been utilized to derive the degree of evenness of urban OSM building completeness following an approach proposed by Massey and Denton (1988) to study residential segregation. Analogous to their approach, the Gini coefficient was derived from the Lorenz curve, which plots the cumulative proportion of observed OSM building area against the cumulative proportion of “missing” building area (difference between OSM building area and predicted building area) across urban centers, which are ordered from smallest to largest proportion of observed building area. The Gini coefficient constitutes a non-spatial measure of segregation which provides insights on the evenness dimension, but does not allow conclusion about the spatial structure. Pysal’s ‘segregation’ package has been utilized to calculate the annual Gini coefficient from 2008-01-01 up until 2022-01-01 (Rey and Anselin, 2007).

Moran’s I (Cliff and Ord, 1981) has been selected as a measure of global spatial autocorrelation of urban OSM building completeness. A high Moran’s I value describes situations where urban centers and their neighbors showed similar high (or low) values of completeness. It’s values are not strictly bound by the interval $[-1, 1]$. The range depends on the largest and the smallest eigenvalue of the spatial weight matrix used, but frequently ranges from -0.5 to 1.15 (Griffith et al., 2019). A Moran’s I value close to zero indicates a spatially random pattern, where the completeness of an urban center was not correlated

to the completeness of its neighbours. Spatial autocorrelation has been proposed as an explicitly spatial indicator of segregation covering the dimension of clustering (Morrill, 1991; Massey and Denton, 1988). Following this approach a high degree of clustering describes a spatial structure where areas with OSM building mapping are contiguous and closely packed, creating a single large block of "mapped" urban centers. In contrast, a low level of clustering implies that the observed OSM building stock is widely scattered around the globe (or within regions) (Massey and Denton, 1988). Moran's I relies on the definition of the spatial weight matrix - here, it was defined based upon the centroid of each urban center using a distance band threshold of 5 degree and an inverse distance weighting with a power of 1. Weights were row standardized. The neighborhood definition led on average to 367 neighbours per urban centers and 13 urban centers were classified as 'islands' for which no neighbours were identified.

We used the same spatial weight matrix to calculate local spatial autocorrelation (Local Moran Statistics (Anselin, 1995)) for Europe & Central Asia and Sub-Saharan Africa to compare spatial inequalities within these regions between two timestamps. Pysal's 'esda' package was utilized to calculate the global and local Moran's I statistics from 2008-01-01 up until 2022-01-01 (Rey and Anselin, 2007).

4.5 Intra-Urban OSM Building Completeness

To ensure a sufficient sample size, we calculated both inequality measures (Gini coefficient, Moran's I) for the intra-urban assessment only for urban centers with a minimum area of 25 square kilometers respective 25 data points. For the resulting 4,722 urban centers 'local' OSM building completeness was derived using the area ratio method described above for each 1x1 km grid cell. The Gini coefficient and Moran's I global spatial autocorrelation of the 'local' OSM completeness were calculated per urban center as described in the previous section. The spatial weight matrix has been defined based upon the Queen contiguity graph. As such grid cells that share at least a vertex were considered as neighbours. The weight matrix was row-standardized.

The investigation was complemented by an agglomerative hierarchical cluster analysis of urban centers considering evenness and clustering within each city. The number of clusters has been selected based on the hierarchical structure of the full dendrogram (see Figure 6) and by investigating the Variance Ratio Criterion (Calinski-Harabasz Index) and Silhouette Coefficient for various number of clusters (see Supplementary Material Section Table 10). This analysis revealed that a number of two or three clusters would be optimal when only considering the clustering performance scores. We decided for three main clusters (1-3), but also report sub clusters for 1 (a, b) and 2 (a, b) to allow for a more fine grained distinction of urban centers, especially for those with low or medium completeness.

The distance matrix was based on the euclidean distance between OSM completeness, Gini coefficient and Moran's I. Since all variables considered in this analysis already showed a similar range of values between 0-1 normalization was not necessary. The analysis was conducted based on scikit-learn's python implementation (Pedregosa et al., 2011) using the ward linkage criterion. Results were displayed for each cluster using scatter plots.

Cluster representatives were selected by first calculating the average values for Moran's I, Gini coefficient and completeness across all urban centers per cluster. Based on these cluster centroids the euclidean distance to each sample was derived. Among the 15 samples closest to the cluster centroid one representative was selected per cluster.

Data Availability

Source data are provided with this paper. The full set of data for training and running the machine learning model and the final results presented in all figures and maps are available on Figshare: <https://www.doi.org/10.6084/m9.figshare.22217038>.

Code Availability

All Python code and Jupyter notebooks necessary to calculate the geospatial statistics, create maps and derive figures are available in this GitHub repository:

<https://doi.org/10.6084/m9.figshare.22241776>.

References

- Altay, N. and Labonte, M. (2014). Challenges in humanitarian information management and exchange: Evidence from Haiti. *Disasters*, 38(S1):50–72.
- Anderson, J., Sarkar, D., and Palen, L. (2019). Corporate Editors in the Evolving Landscape of Open-StreetMap. *ISPRS International Journal of Geo-Information*, 8(5):232.
- Anderson, J., Soden, R., Keegan, B., Palen, L., and Anderson, K. M. (2018). The crowd is the territory: Assessing quality in peer-produced spatial data during disasters. *International Journal of Human-Computer Interaction*, 34:295–310.
- Anselin, L. (1995). Local Indicators of Spatial Association - LISA. *Geographical Analysis*, 27(2):93–115.
- Bank, W. (2020). Open cities africa final report. <https://opendri.org/resource/open-cities-africa-final-report/>. [Online; accessed 06-March-2023].
- Barrington-Leigh, C. and Millard-Ball, A. (2017). The world’s user-generated road map is more than 80% complete. *PLoS ONE*, 12(8):1–20.
- Bhatia, A., Mahmud, A., Fuller, A., Shin, R., Rahman, A., Shatil, T., Sultana, M., Morshed, K. A., Leaning, J., and Balsari, S. (2018). The Rohingya in cox’s bazar: When the stateless seek refuge. *Health and Human Rights*, 20(2):105–122.
- Biljecki, F., Chew, L. Z. X., Milojevic-Dupont, N., and Creutzig, F. (2021). Open government geospatial data on buildings for planning sustainable and resilient cities. preprint.
- Biljecki, F. and Chow, Y. S. (2022). Global building morphology indicators. *Computers, Environment and Urban Systems*, 95:101809.
- Boeing, G. (2021). Spatial information and the legibility of urban form: Big data in urban morphology. *International Journal of Information Management*, 56(October 2019):102013.
- Boo, G., Darin, E., Leasure, D. R., Dooley, C. A., Chamberlain, H. R., Lázár, A. N., Tschirhart, K., Sinai, C., Hoff, N. A., Fuller, T., Musene, K., Batumbo, A., Rimoin, A. W., and Tatem, A. J. (2022). High-resolution population estimation using household survey data and building footprints. *Nature Communications*, 13(1):1–10.
- Braunschweig, K., Eberius, J., Thiele, M., and Lehner, W. (2020). The State of Open Humanitarian Data. Technical Report January.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45:5–32.
- Brelsford, C., Martin, T., Hand, J., and Bettencourt, L. M. A. (2018). Toward cities without slums: Topology and the spatial evolution of neighborhoods. *Science Advances*, 4.
- Brenning, A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5372–5375.
- Brovelli, M. and Zamboni, G. (2018). A New Method for the Assessment of Spatial Accuracy and Completeness of OpenStreetMap Building Footprints. *ISPRS International Journal of Geo-Information*, 7(8):289.
- Cervero, R. and Day, J. (2008). Suburbanization and transit-oriented development in china. *Transport Policy*, 15:315–323.

- Cliff, A. D. and Ord, J. K. (1981). *Spatial processes: models & applications*. Taylor & Francis.
- Clinton Health Access Initiative (2017). Improving the impact of indoor residual spraying in southern africa through data-driven approaches. <https://www.clintonhealthaccess.org/blog/improving-impact-indoor-residual-spraying-southern-africa-data-driven-approaches/>. [Online; accessed 06-March-2023].
- de Albuquerque, J. P., Herfort, B., Brenning, A., and Zipf, A. (2015). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science*, (February):1–23.
- Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F. C., and Taneja, J. (2021). Annual time series of global viirs nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing*, 13(5):1–14.
- Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., Marconcini, M., Roth, A., and Zeidler, J. (2022). World Settlement Footprint 3D - A first three-dimensional survey of the global building stock. *Remote Sensing of Environment*, 270(January):112877.
- European Commission (2023). Copernicus emergency management service - mapping. <https://emergency.copernicus.eu/mapping/copernicus-emergency-management-service>. [Online; accessed 06-March-2023].
- Fan, H., Zipf, A., Fu, Q., and Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28(4):700–719.
- Feldmeyer, D., Nowak, W., Jamshed, A., and Birkmann, J. (2021). An open resilience index: Crowdsourced indicators empirically developed from natural hazard and climatic event data. *Science of the Total Environment*, 774:145734.
- Florczyk, A. J., Melchiorri, M., Orbane, C., Schiavina, M., Maffenini, M., Politis, P., Sabo, S., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D., and Zanchetta, L. (2019). *Description of the GHS Urban Centre Database 2015*. Number February.
- Freire, S., MacManus, K., Pesaresi, M., Doxsey-Whitfield, E., and Mills, J. (2016). Development of new open and free multi-temporal global population grids at 250 m resolution. *Agile*, (June):6.
- Fritz, S., See, L., Carlson, T., Haklay, M. M., Oliver, J. L., Fraisl, D., Mondardini, R., Brocklehurst, M., Shanley, L. A., Schade, S., Wehn, U., Abrate, T., Anstee, J., Arnold, S., Billot, M., Campbell, J., Espey, J., Gold, M., Hager, G., He, S., Hepburn, L., Hsu, A., Long, D., Masó, J., McCallum, I., Muniafu, M., Moorthy, I., Obersteiner, M., Parker, A. J., Weissplug, M., and West, S. (2019). Citizen science and the United Nations Sustainable Development Goals. *Nature Sustainability*, 2(10):922–930.
- Goetz, J. N., Guthrie, R. H., and Brenning, A. (2011). Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology*, 129(3-4):376–386.
- Goldblatt, R., Jones, N., and Mannix, J. (2020). Assessing OpenStreetMap Completeness for Management of Natural Disaster by Means of Remote Sensing: A Case Study of Three Small Island States (Haiti, Dominica and St. Lucia). *Remote Sensing*, 12(1):118.
- Graham, M., Straumann, R. K., and Hogan, B. (2015). Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia. *Annals of the Association of American Geographers*, 105(6):1158–1178.
- Griffith, D. A., Chun, Y., and Li, B. (2019). *Spatial Regression Analysis using Eigenvector Filtering*. CRC Press.
- Hecht, R., Kunze, C., and Hahmann, S. (2013). Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *ISPRS International Journal of Geo-Information*, 2(4):1066–1091.

- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B., and Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 2018(8).
- Herfort, B. (2023). Dataset: Investigating completeness and inequalities in openstreetmap - spatio-temporal analysis of global urban building data.
- Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., and Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports*, 11(1).
- Humanitarian OpenStreetMap Team (2021). Data-driven disaster risk management philaware. <https://www.hotosm.org/projects/data-driven-disaster-risk-management-philaware/>. [Online; accessed 06-March-2023].
- Ludwig, C., Hecht, R., Lautenbach, S., Schorcht, M., and Zipf, A. (2021). Mapping public urban green spaces based on openstreetmap and sentinel-2 imagery using belief functions. *ISPRS International Journal of Geo-Information*, 10.
- Lüge, T. (2014). Gis support for the msf ebola response in guinea in 2014. [Online; accessed 06-March-2023].
- Marco Minghini, Serena Coetzee, Levente Juhasz, Godwin Yeboah, Peter Mooney, and A. Yair Grinberger (2020). Editorial: OpenStreetMap research in the COVID-19 era. *Proceedings of the Academic Track at the State of the Map 2020*, pages 1–4.
- Massey, D. S. and Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, 67(2):281–315.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6).
- Meta (2022a). Daylight map distribution. <https://daylightmap.org/>. Accessed: 2022-06-01.
- Meta (2022b). Map with ai. <https://mapwith.ai>. Accessed: 2022-06-01.
- Meyer, H. and Pebesma, E. (2022). Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications*, 13(1):2208.
- Microsoft (2022). Microsoft building footprints. <https://github.com/microsoft/GlobalMLBuildingFootprints>. Accessed: 2022-06-01.
- Milojevic-Dupont, N., Hans, N., Kaack, L. H., Zumwald, M., Andrieux, F., de Barros Soares, D., Lohrey, S., Pichler, P. P., and Creutzig, F. (2020). Learning from urban form to predict building heights. *PLoS ONE*, 15(12 December):1–22.
- Morrill, R. L. (1991). On the measure of geographic segregation. *Geography Research Forum*, 11:25–36.
- Neis, P., Zielstra, D., and Zipf, A. (2013). Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet*, 5(2):282–300.
- Nirandjan, S., Koks, E. E., Ward, P. J., and Aerts, J. C. (2022). A spatially-explicit harmonized global dataset of critical infrastructure. *Scientific Data*, 9(1):1–13.
- OECD (2016). *OECD Regions at a Glance 2016*.
- Oort, P. (2006). *Spatial data quality: from description to application*. PhD thesis, Wageningen University.
- OpenStreetMap Contributors (2022). Openstreetmap wiki - organized editing activities. https://wiki.openstreetmap.org/wiki/Organised_Editing/Activities. Accessed: 2022-06-18.

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830.
- Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., and Péliissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, 11(1):1–11.
- Porto de Albuquerque, J., Anderson, L., Calvillo, N., Coaffee, J., Cunha, M. A., Degrossi, L. C., Dolif, G., Horita, F., Klonner, C., Lima-Silva, F., Marchezini, V., Martins, M. H. d. M., Pajarito-Grajales, D., Pitidis, V., Rudorff, C., Tkacz, N., Traijber, R., and Zipf, A. (2021). The role of data in transformations to sustainability: a critical research agenda. *Current Opinion in Environmental Sustainability*, 49(August 2020):153–163.
- Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.-B., and Zipf, A. (2019). OSHDB: a framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards*, 4(1).
- Rehrl, K. and Gröchenig, S. (2016). A Framework for Data-Centric Analysis of Mapping Activity in the Context of Volunteered Geographic Information. *ISPRS International Journal of Geo-Information*, 5(3):37.
- Rey, S. J. and Anselin, L. (2007). PySAL: A Python Library of Spatial Analytical Methods. *The Review of Regional Studies*, 37(1):5–27.
- Risk Factor (2023). <https://riskfactor.com/>. [Online; accessed 06-March-2023].
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929.
- Sachs, J., Lafortune, G., Kroll, C., Fuller, G., and Woelm, F. (2022). Sustainable Development Report. Technical report.
- Scholz, S., Knight, P., Eckle, M., Marx, S., and Zipf, A. (2018). Volunteered Geographic Information for Disaster Risk Reduction—The Missing Maps Approach and Its Potential within the Red Cross and Red Crescent Movement. *Remote Sensing*, 10(8):1239.
- See, L., Georgieva, I., Duerauer, M., Kemper, T., Corbane, C., Maffenini, L., Gallego, J., Pesaresi, M., Sirbu, F., Ahmed, R., Blyshchyk, K., Magori, B., Blyshchyk, V., Melnyk, O., Zadorozhniuk, R., Mandici, M. T., Su, Y. F., Rabia, A. H., Pérez-Hoyos, A., Vasylyshyn, R., Pawe, C. K., Bilous, S., Kovalevskyi, S. B., Kovalevskyi, S. S., Bordoloi, K., Bilous, A., Panging, K., Bilous, V., Prestele, R., Sahariah, D., Deka, A., Nath, N., Neves, R., Myroniuk, V., Karner, M., and Fritz, S. (2022). A crowdsourced global data set for validating built-up surface layers. *Scientific Data*, 9.
- Smits, J. and Permanyer, I. (2019). Data descriptor: The subnational human development database. *Scientific Data*, 6:1–15.
- Sturrock, H. J., Woolheater, K., Bennett, A. F., Andrade-Pacheco, R., and Midekisa, A. (2018). Predicting residential structures from open source remotely enumerated data using machine learning. *PLoS ONE*, 13(9):1–10.
- Sui, D., Goodchild, M., and Elwood, S. (2013). Volunteered geographic information, the exaflood, and the growing digital divide. In *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, volume 9789400745872, pages 1–12. Springer.

- Sun, L., Chen, J., Li, Q., and Huang, D. (2020). Dramatic uneven urbanization of large cities throughout the world in recent decades. *Nature Communications*, 11(1).
- Sustainable Development Solutions Network (2015). Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a data revolution for the SDGs. *A report by the Leadership Council of the Sustainable Development Solutions Network*, (Version 7):160.
- Ulbrich, P., De Albuquerque, J. P., and Coaffee, J. (2019). The impact of urban inequalities on monitoring progress towards the sustainable development goals: Methodological considerations. *ISPRS International Journal of Geo-Information*, 8(1):1–18.
- UN Mappers (2023). Un mappers. <https://mappers.un.org/>. [Online; accessed 06-March-2023].
- UN OCHA (2023). Humanitarian data exchange (hdx). <https://data.humdata.org/dataset?q=building>. [Online; accessed 06-March-2023].
- UN Stats (2023a). Sdg 11.1.1 indicator metadata. <https://unstats.un.org/sdgs/metadata/files/Metadata-11-01-01.pdf>. [Online; accessed 06-March-2023].
- UN Stats (2023b). Sdg 11.2.1 indicator metadata. <https://unstats.un.org/sdgs/metadata/files/Metadata-11-02-01.pdf>. [Online; accessed 06-March-2023].
- UN Stats (2023c). Sdg 11.3.1 indicator metadata. <https://unstats.un.org/sdgs/metadata/files/Metadata-11-03-01.pdf>. [Online; accessed 06-March-2023].
- United Nations Development Programme (2019). *Human development report 2019 : beyond income, beyond averages, beyond today: inequalities in human development in the 21st century*.
- Van Den Hoek, J., Friedrich, H. K., Ballasiotes, A., Peters, L. E. R., and Wrathall, D. (2021). Development after Displacement: Evaluating the Utility of OpenStreetMap Data for Monitoring Sustainable Development Goal Progress in Refugee Settlements. *ISPRS International Journal of Geo-Information*, 10(3):153.
- Witt, R., Loos, L., and Zipf, A. (2021). Analysing the impact of large data imports in openstreetmap. *ISPRS International Journal of Geo-Information*, 10.
- Yeboah, G., de Albuquerque, J. P., Troilo, R., Tregonning, G., Perera, S., Shifat Ahmed, S. A., Ajisola, M., Alam, O., Aujla, N., Azam, S. I., Azeem, K., Bakibinga, P., Chen, Y. F., Choudhury, N. N., Diggle, P. J., Fayahun, O., Gill, P., Griffiths, F., Harris, B., Iqbal, R., Kabaria, C., Ziraba, A. K., Khan, A. Z., Kibe, P., Kisia, L., Kyobutungi, C., Lilford, R. J., Madan, J. J., Mbaya, N., Mberu, B., Mohamed, S. F., Muir, H., Nazish, A., Njeri, A., Odubanjo, O., Omigbodun, A., Osuh, M. E., Owoaje, E., Oyeboode, O., Pitidis, V., Rahman, O., Rizvi, N., Sartori, J., Smith, S., Taiwo, O. J., Ulbrich, P., Uthman, O. A., Watson, S. I., Wilson, R., and Yusuf, R. (2021). Analysis of openstreetmap data quality at different stages of a participatory mapping process: Evidence from slums in Africa and Asia. *ISPRS International Journal of Geo-Information*, 10(4).
- Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., and Tsendbazar, O. (2021). ESA WorldCover 10 m 2020 v100. Technical report.
- Zhang, Y., Zhou, Q., Brovelli, M. A., and Li, W. (2022). Assessing OSM building completeness using population data. *International Journal of Geographical Information Science*, pages 1–24.
- Zhou, Q., Wang, S., and Liu, Y. (2022a). Exploring the accuracy and completeness patterns of global land-cover/land-use data in OpenStreetMap. *Applied Geography*, 145(June):102742.
- Zhou, Q., Zhang, Y., Chang, K., and Brovelli, M. A. (2022b). Assessing osm building completeness for almost 13,000 cities globally. *International Journal of Digital Earth*, 15:2400–2421.

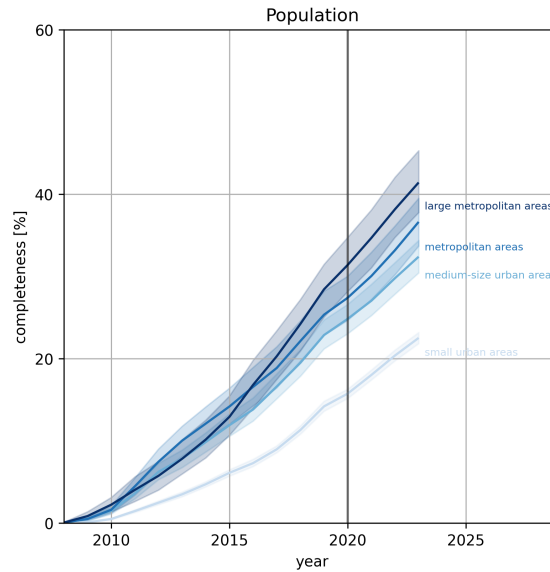


Figure 9: Temporal evolution of average urban OSM building completeness by population. Completeness was derived by aggregating building area predictions based on a Random Forests model and monthly OSM building area per urban center. The shaded areas represent the 95% confidence interval for each line. OSM data from 2008-01-01 to 2023-01-01. Created using Matplotlib 3.3. in Python 3.7.5 (<https://www.python.org/>).

Zielstra, D., Hochmair, H. H., and Neis, P. (2013). Assessing the effect of data imports on the completeness of openstreetmap - a united states case study. *Transactions in GIS*, 17:315–334.

Acknowledgements

The authors would like to thank the OSM contributors for their inspiring work, as well as Matthias Schaub, Levi Szamek, Rafael Troilo and Clemens Langer for their great help with building the computational framework for the analysis. B.H. and S.L. were supported by the Klaus Tschira Stiftung.

Appendix - Supplementary Material

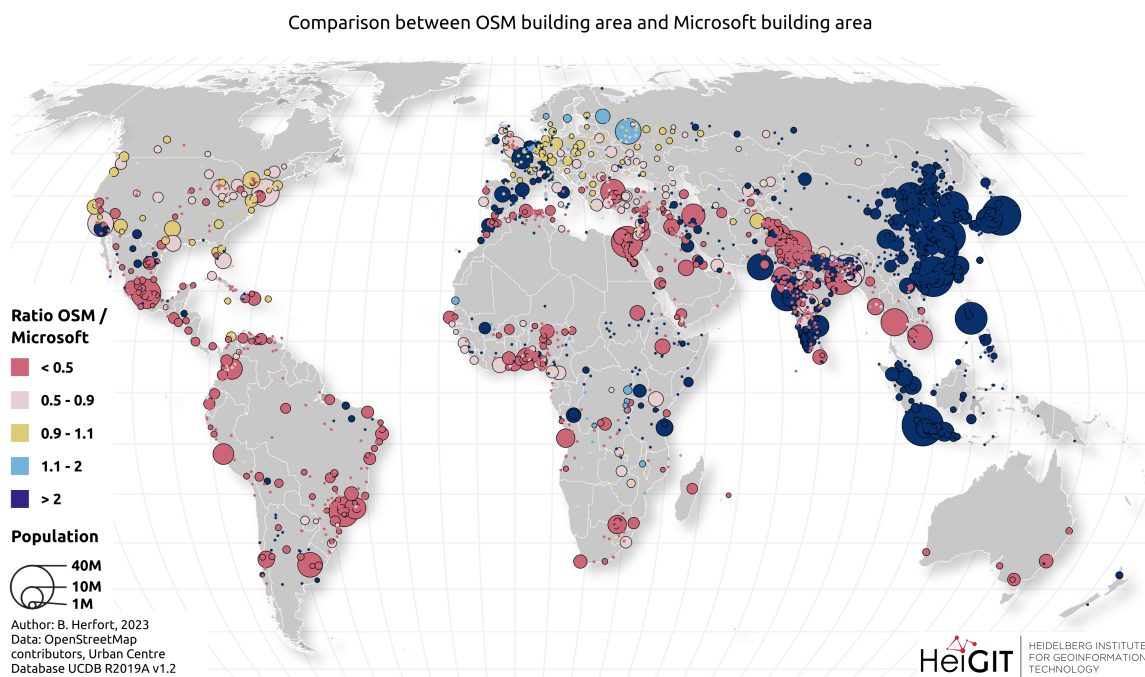


Figure 10: Spatial distribution of urban centers based on the ratio between OSM building area and Microsoft building area. This comparison only entails Microsoft building footprints made available through the Global ML Building Footprints dataset as of 2023-01-15. OSM data as of 2023-01-01. Created using QGIS 3.28.3 (<https://www.qgis.org/en/site/>).

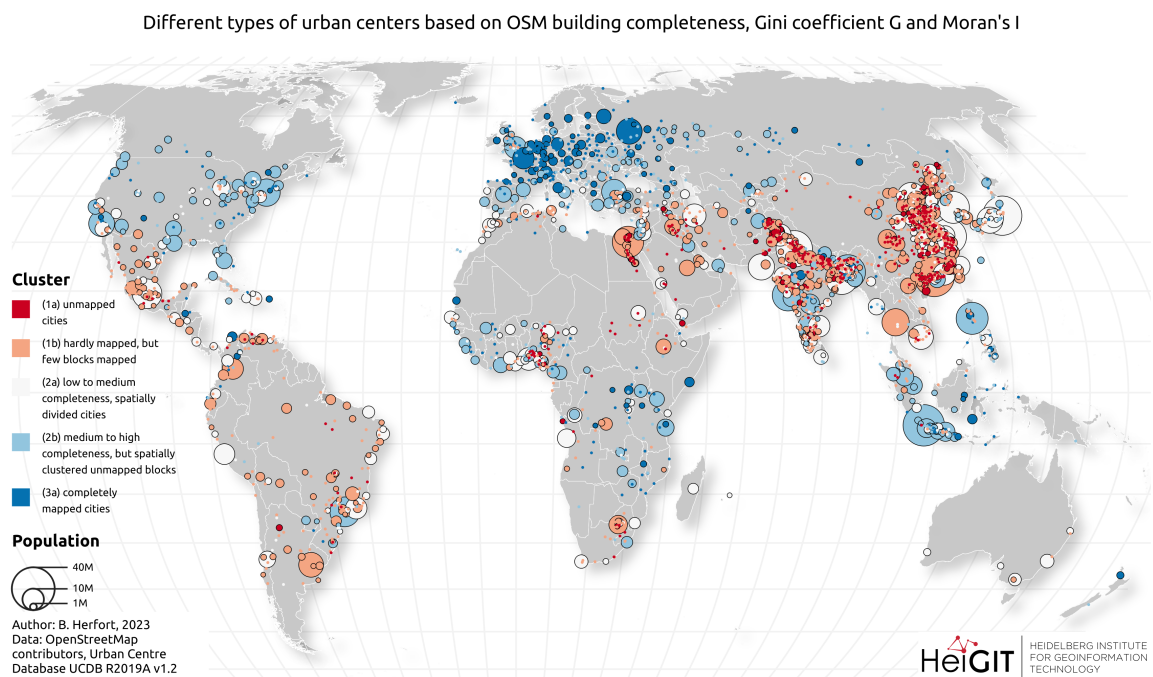


Figure 11: Spatial distribution of agglomerative clustering of urban centers based on OSM building completeness, Gini coefficient G and Moran's I . Each point represents a single urban center with a minimum area of 25 square kilometers ($n= 4,647$). Smaller urban centers were ignored as Gini coefficient and Moran's I could not be reliably estimated. OSM data as of 2023-01-01. Created using QGIS 3.28.3 (<https://www.qgis.org/en/site/>).

Table 5: Comparison between OSM building area and Microsoft building area in urban centers on the global scale and grouped by world regions, Subnational Human Development Index class and city size class measured by population. SHDI classes were based on cut-off points defined by the United Nations Development Programme (United Nations Development Programme, 2019): low human development (SHDI < 0.550), medium human development (SHDI: 0.550 - 0.699), high human development (SHDI: 0.700–0.799), very high human development (SHDI > 0.800). City size classes were based on population thresholds defined by OECD (OECD, 2016): small urban areas (50k–200k), medium-size urban areas (200k–500k), metropolitan areas (500k–1.5M), large metropolitan areas (>1.5M). This comparison only entails Microsoft building footprints made available through the Global ML Building Footprints dataset as of 2023-01-15. OSM data as of 2023-01-01.

	n	OSM Buildings [sqkm]	Microsoft Buildings [sqkm]	Ratio OSM / Microsoft
Global	13,189	34,531	50,371	0.69
World Bank Regions				
East Asia & Pacific	3,068	7,828	2,983	2.62
Europe & Central Asia	1,351	9,839	9,259	1.06
Latin America & Caribbean	1,073	2,198	9,815	0.22
Middle East & North Africa	901	1,156	4,727	0.24
North America	378	9,122	12,347	0.74
South Asia	3,997	1,971	6,366	0.31
Sub-Saharan Africa	2,421	2,417	4,874	0.50
Subnational Human Development Index				
Low	2,289	1,109	1,862	0.60
Medium	4,960	4,123	11,519	0.36
High	3,883	7,425	18,122	0.41
Very High	1,967	21,810	24,376	0.89
City Size by Population				
Small Urban Areas	10,930	7,151	13,185	0.54
Medium-Size Urban Areas	1,348	4,907	8,018	0.61
Metropolitan Areas	563	6,361	9,931	0.64
Large Metropolitan Areas	287	16,111	24,757	0.65

Table 6: Reference datasets used for training the machine learning model.

Dataset Name	Country ISO A3 codes	Urban Centers	Grid Cells
East Asia & Pacific			
Microsoft Building Footprints	AUS, KHM, LAO, MMR, MNG, PRK, THA, VNM	332	19,963
GSI Basic Map Information Buildings	JPN	108	22,342
NSDI Continuous Numerical Topographic Map Building Data	KOR	37	5,258
LDS NZ Building Outlines	NZL	8	941
Europe & Central Asia			
Microsoft Building Footprints	ALB, ARM, AUT, AZE, BEL, BGR, BIH, BLR, CHE, CYP, CZE, DEU, DNK, ESP, EST, FIN, FRA, GBR, GEO, GRC, HRV, HUN, IRL, ISL, ITA, KAZ, KGZ, LTU, LUX, LVA, MDA, MKD, MNE, NLD, NOR, POL, PRT, ROU, RUS, SRB, SVK, SVN, SWE, TJK, TKM, TUR, UKR, UZB	1,042	69,954
OS OpenMap Local	GBR	135	10,845
IGN BD TOPO Bâtiments	FRA	71	6,292
GUGiK BDOT10k Budynki	POL	48	3,221
NGR Basisregistraties Adressen en Gebouwen (BAG)	NLD	38	2,687
Hausumringe Nordrhein Westfalen, Sachsen, Berlin	DEU ^a	24	4,203
CUZK Budovy (BU)	CZE	12	717
Estonian Land Board Buildings	EST	2	130
Latin America & Caribbean			
Microsoft Building Footprints	ARG, BHS, BLZ, BOL, BRA, BRB, CHL, COL, CRI, CUB, CUW, DOM, ECU, GTM, GUY, HND, HTI, JAM, MEX, NIC, PAN, PER, PRI, PRY, SLV, SUR, TTO, URY, VEN	877	52,135
SP Mapa Digital da Cidade	BRA ^a	3	887
Quito Gobierno Abierto Construcciones	Ecuador ^a	2	340
Construcción. Bogotá D.C	Colombia ^a	1	415
BA Data Tejido Urbano	ARG ^a	1	187
Middle East & North Africa			
Microsoft Building Footprints	ARE, DJI, DZA, EGY, IRN, IRQ, ISR, JOR, LBN, LBY, MAR, MLT, OMN, PSE, QAT, SAU, SYR, TUN, YEM	666	31,948
North America			
Microsoft Building Footprints	CAN, USA	367	100,210
South Asia			
Microsoft Building Footprints	AFG, BGD, BTN, IND, LKA, MDV, NPL, PAK	1,677	56,421
Sub-Saharan Africa			
Microsoft Building Footprints	AGO, BDI, BEN, BFA, BWA, CAF, CIV, CMR, COD, COG, COM, ERI, ETH, GAB, GHA, GIN, GMB, GNB, GNQ, KEN, LBR, LSO, MDG, MLI, MOZ, MRT, MUS, MWI, NAM, NER, NGA, RWA, SDN, SEN, SLE, SOM, SSD, SWZ, TCD, TGO, ZAF, ZMB, ZWE	1,449	40,534

^aBuilding data did not cover the entire country.

Table 7: Precision and recall for the Microsoft building footprints by comparing to Geo-Wiki built-up surface validation dataset. The Geo-Wiki campaign visually assessed very high-resolution satellite images of 50 K sample locations for the presence of built-up surfaces using a crowdsourcing approach. In the Geo-Wiki dataset built-up area is defined as an area containing any building with a roof.

Region	Urban Centers	Geo-Wiki Cells	Recall [%]	Precision [%]
East Asia & Pacific	24	2,352	86.0	87.3
Europe & Central Asia	144	11,531	83.5	86.0
Latin America & Caribbean	91	8,768	83.4	89.8
Middle East & North Africa	55	5,094	84.4	81.7
North America	119	17,373	86.4	88.3
South Asia	70	7,526	84.8	73
Sub-Saharan Africa	161	23,588	81.2	83.3

Table 8: Precision and recall for the Microsoft building footprints as self-declared by Microsoft in the Global ML Building Footprints GitHub repository. The evaluation metrics are computed on a set of building polygon labels for each region. Values might differ across the region.

Region	Recall [%]	Precision [%]
Africa	70.9	94.4
Caribbean	76.8	92.2
Central Asia	79.5	97.2
Europe	85.9	94.3
Middle East	85.4	95.7
South America	78.0	95.4
South Asia	76.7	94.8

Table 9: Aggregated OSM road network completeness estimation per world region based on the mapping saturation approach. Mapping activity in urban centers was classified into a "start", "growth" and "saturation" phase. Saturation indicates that most of the mapping of the road network has been completed. During the growth phase the majority of features have been added.

Region	Start	Growth	Saturation
East Asia & Pacific	0.022	0.665	0.302
Europe & Central Asia	0.003	0.145	0.847
Latin America & Caribbean	0.004	0.238	0.752
Middle East & North Africa	0.102	0.504	0.384
North America	0.013	0.016	0.947
South Asia	0.105	0.553	0.334
Sub-Saharan Africa	0.059	0.478	0.460

Table 10: Silhouette Coefficient and Calinski-Harabasz Index for various number of clusters. Higher Silhouette Coefficient and Calinski-Harabasz scores relate to a model with better defined clusters.

clusters	Silhouette Coefficient	Calinski-Harabasz Index
2	0.47	4,903
3	0.46	5,026
4	0.39	4,566
5	0.30	4,393
6	0.29	4,125
7	0.29	3,930
8	0.29	3,789

Mapping Human Settlements with Higher Accuracy and Less Volunteer Efforts by Combining Crowdsourcing and Deep Learning

Benjamin Herfort, Hao Li, Sascha Fendrich,
Sven Lautenbach and Alexander Zipf

31 July 2019

Abstract

Reliable techniques to generate accurate data sets of human build-up areas at national, regional and global scales are a key factor to monitor the implementation progress of the Sustainable Development Goals as defined by the United Nations. However, the scarce availability of accurate and up-to-date human settlement data remains a major challenge, e.g., for humanitarian organizations. In this paper, we investigated the complementary value of crowdsourcing and deep learning to fill the data gaps of existing earth observation (EO) based products. To this end, we propose a novel workflow to combine deep learning (Deep-VGI) and crowdsourcing (MapSwipe). Our strategy for allocating classification tasks to deep learning or crowdsourcing is based on confidence of the derived binary classification. We conducted case studies in three different sites located in Guatemala, Laos, and Malawi to evaluate the proposed workflow. Our study reveals that crowdsourcing and deep learning outperform existing EO based approaches and products such as the Global Urban Footprint. Compared to a crowdsourcing-only approach the combination increased the quality (measured by Matthew's correlation coefficient) of the generated human settlement maps by 3 to 5 percentage points. At the same time, it reduced the volunteer efforts needed by at least 80 percentage points for all study sites. The study suggests that for the efficient creation of human settlement maps, we should rely on human skills when needed and rely on automated approaches when possible.

Reference— Herfort, B., Li, H., Fendrich, S., Lautenbach, S., and Zipf, A. (2019). Mapping human settlements with higher accuracy and less volunteer efforts by combining crowdsourcing and deep learning. *Remote Sensing*, 11(15). <https://doi.org/10.3390/rs11151799>

1 Introduction

Currently, 55% of the world's population reside in urban areas and especially in low-income and lower-middle-income countries rapid urbanization is expected between now and 2050 (United Nations Department of Economic and Social Affairs Population Division, 2018). The Sustainable Development Goals (SDGs) (United Nations, 2015) and the Sendai Framework for Disaster Risk Reduction (SFDRR) (United Nations Office for Disaster Risk Reduction, 2015) both highlight the relevance and increasing need of up-to date information on the spatial distribution of human settlements. For instance, humanitarian organizations cannot help people, if they cannot find them. Consequently, reliable techniques to generate accurate data sets of human settlements at national, regional and global scales are crucial in manifold domains such as disaster management, habitat and ecological system conservation, public health monitoring.

Earth observation (EO) using satellites already provides data for a broad range of purposes such as disaster assessment, forestry or crop land monitoring and land use/land cover classification. Recently,

remote sensing technologies have been successfully employed to derive information on human settlements at regional to global scale. Accuracy and completeness of EO derived human settlement data sets have been improved a lot in the last 15 years. Current data sets, which have been made available recently, include the Global Human Settlement Layer (GHSL) (Pesaresi et al., 2013), the Global Urban Footprint (GUF) data set (Esch et al., 2013) and the High Resolution Settlement Layer (HRSL) (Tiecke et al., 2017). However, these data sets still show great variations for different regions and geographic settings (Klotz et al., 2016). Especially rural areas and non-solid building structures are still disregarded or under-represented in these data sets.

Several researchers highlight the potential of crowdsourcing to collect information on human settlements and to complement the data that is produced using satellite imagery (Gueguen et al., 2017; Albuquerque et al., 2016; Hachmann et al., 2017). Additionally, humanitarian organizations start using new methods from Citizen Science and Volunteered Geographic Information (VGI), to gather information on the spatial distribution of human settlements (Scholz et al., 2018). However, the quality and the reliability of those methods and resulting data sets remain major concerns, which are extensively discussed in current research (See et al., 2016). Spatial varying data quality and the lack of reference data with sufficient quality still constitute barriers in using VGI data in general and for humanitarian purposes or in disaster management in particular (Hecht et al., 2013).

The mapping of human settlements tends to be done either from an earth observation perspective or from a citizen science position. However, a tighter integration of both approaches has the potential to derive improved data sets that presumably outperform existing one (Gueguen et al., 2017).

In this article we present a novel workflow to overcome the scarce availability of accurate and up-to-date human settlement data sets (see Figure 1). Our proposed workflow combines two methods: (1) object detection and classification using deep learning algorithms (DeepVGI (Chen and Zipf, 2017)) and (2) crowdsourced mapping of human settlements by volunteers (MapSwipe (Herfort et al., 2017)). To combine both methods we propose a task allocation strategy (3) that chooses classification labels either from DeepVGI or MapSwipe. Moreover, we investigate whether our proposed methodology helps to produce better maps faster with respect to the following research questions.

- RQ1: How good are crowdsourcing (MapSwipe) and deep learning (DeepVGI) with respect to generating human settlement maps in comparison to existing EO based approaches?
- RQ2: Which spatial and non-spatial characteristics of misclassifications are accompanied by applying the DeepVGI approach?
- RQ3: What is the added value of the proposed task allocation strategy with respect to performance and effort?

The remainder of this paper is organized as follows: Section 2 provides background information on techniques for deriving information on human settlements utilizing either deep learning or crowdsourcing. The study areas and data sets are described in Section 3. We present our overall methodology in Section 4 and show results in Section 5. Section 6 discusses our findings and makes suggestions for future work, whereas Section 7 draws conclusions.

2 Background: Mapping human settlements utilizing crowdsourcing or deep neural networks

Previous research shows the strengths and shortcomings of existing global settlement data products such as the Global Human Settlement Layer or the Global Urban Footprint data set (Klotz et al., 2016; Esch et al., 2017). In this section we review to what degree VGI and citizen science approaches and techniques based on deep learning have been applied to tackle the known challenges.

Citizen science projects and VGI tools are widely used to collect information on human settlements. The OpenStreetMap (OSM) platform played a central role in generating map data during several international disasters, e.g., after the 2010 Haiti earthquake (Zook et al., 2010) and after the 2015 Nepal earthquake (Degrossi, 2016). Additionally, OpenStreetMap is used also in disaster preparedness and disaster risk reduction activities, for instance, by the organizations of the Missing Maps project (Scholz

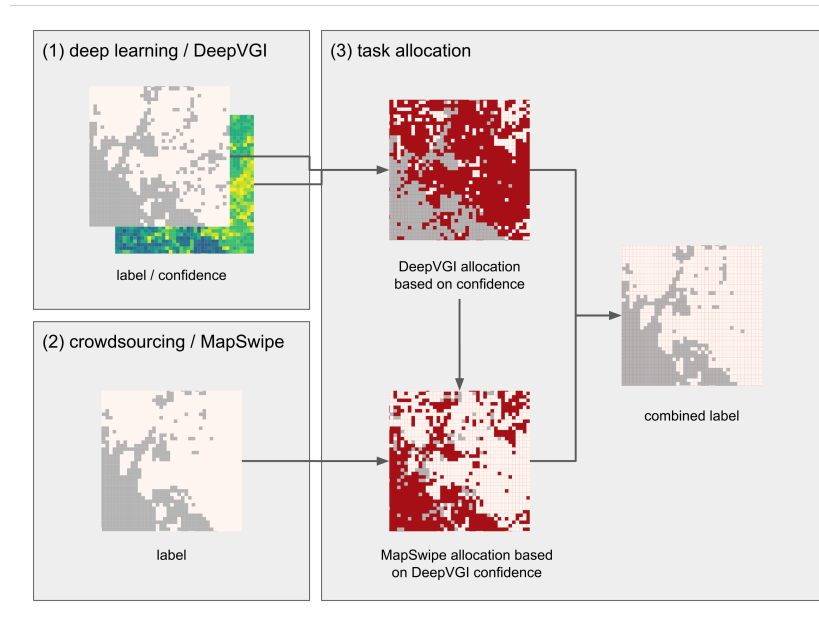


Figure 1: Proposed workflow to combine deep learning and crowdsourcing methods: Combined labels are obtained, by choosing labels either from DeepVGI or MapSwipe based on the confidence of the DeepVGI labels.

et al., 2018). Regarding urban planning, Crooks et al. (2015) show how various user-generated data sets (GPS trajectories, social media data) enrich our understanding of urban form and function from a bottom up perspective.

The increased usage of volunteered geographic information (VGI) has been accompanied by discussions on the quality of data sets, which have not been produced by experts (Goodchild and Li, 2012; Ballatore and Zipf, 2015). The research on VGI data quality reveals that spatial heterogeneity, e.g., regarding completeness of building footprints, remains a major challenge on different geographic scales (Hecht et al., 2013). Fan et al. (2014) confirm a high completeness for building footprint features in an urban area in Munich, Germany. For crowd-sourced classification of human settlements in Madagascar and South Sudan Herfort et al. (2017) conclude, that disagreement between users is not randomly distributed in space but rather clustered, indicating that reliability of information varies spatially. Additionally, Comber et al. (2016) show for land cover mapping, that the quality of VGI data sets is influenced by differences between user groups, which are a potential source of error and uncertainty.

In addition to the above VGI-related work, several authors have investigated the potential of deep learning technology in various satellite image processing tasks including human settlement mapping (see Zhu et al. (2017) for a comprehensive overview). In Jean et al. (2016) a convolutional neural network is used to distinguish urban areas, non-urban area, roads and water in optical satellite imagery for predicting poverty in Nigeria, Tanzania, Uganda, Malawi, and Rwanda. In Li et al. (2019), a pre-trained neural network is employed to estimate large scale OSM missing built-up areas in Tanzania. Regarding land cover mapping, several authors propose workflows based on deep neural networks with a focus on urban areas (Kampffmeyer et al., 2016; Tracowski et al., 2017). Furthermore, building footprint extraction based on deep learning has been a central research topic in the last years (Vakalopoulou et al., 2015; Yuan et al., 2018).

Although deep learning shows promising results with respect to object detection in images in general, analyzing data quality of geographic approaches (e.g., using geographic data sets such as satellite imagery and building footprints) remains a huge challenge. The transferability of deep learning models constitutes a key challenge towards global scale data products, e.g., for human settlements. For instance, Yuan et al. (2018) highlight difficulties in extracting footprints for buildings in rural areas which differed significantly from the footprints presented in the training set. Missing benchmark data sets tailored to remote sensing

tasks make it difficult to compare the growing number of deep learning algorithms (Zhu et al., 2017). Similar to the factors contributing to spatial heterogeneity of VGI data sets, deep learning approaches are vulnerable to changes in input factors such as atmospheric scattering conditions, intraclass variability, culture-dependent characteristics and a limited number of training samples (Zhu et al., 2017).

In addition to studies focusing either on deep learning or on VGI approaches in isolation to generate information on human settlements, some studies have successfully combined both approaches. A study by Gueguen et al. (2017) produces regional- and country-scale population distribution maps from very high resolution satellite imagery based on the detection of village boundaries by a deep neural network and a crowdsourced validation of the results. The study reports benefits from combining the high recall of automated methods with the high precision of human validators. By combining data from multiple crowdsourcing projects in an active learning framework for convolutional neural networks Chen et al. (2018) address incompleteness and spatial heterogeneity of input training samples regarding road and building mapping. Their results show a promising avenue how deep learning can be utilized to improve VGI data. However, the small sample size used for validation hinders conclusions on the transferability of their findings. Vargas-Muñoz et al. (2019) investigate the quality of OSM data in study sites in Zimbabwe and Tanzania using a deep learning approach. Their approach is able to detect missing building footprints and misalignment in the OSM data, but the validation data sets contain only 1000 buildings per site, which again casts doubts on the transferability of the proposed approach.

Previous work has shown how citizen science, VGI and deep learning can contribute to improve large scale geographic data sets on human settlements. Those methods help to produce data sets desperately needed for monitoring urban growth, sustainable development, disaster risk reduction and many other applications. Nevertheless, researchers have also revealed that spatial heterogeneity is a key issue, which needs to be addressed to understand and enhance data quality. Regarding VGI data sets, spatial heterogeneity is expressed by regional difference in data completeness and varying data quality due to diverging user experience. Considering deep learning approaches, spatial heterogeneity can be interpreted as the difficulty to transfer models from one region to another and to provide training samples which incorporate the geographic properties for all object structures or characteristics and regions.

We propose a workflow to combine the strengths of recent machine learning algorithms and crowdsourced data production by means of a confidence-based task allocation strategy. Our work is guided by the common hypothesis that humans rarely identify something as a building which is not a building, but tend to miss some objects. Furthermore, deep learning approaches will miss fewer buildings at the cost of detecting also a number of objects which are not buildings. By bringing together those two research streams we aim at producing human settlement data sets which are both more complete and precise.

3 Description of the study areas and data sets

3.1 Study areas

We investigated our combined mapping workflow at three study sites: (a) Guatemala, (b) Laos and (c) Malawi. The study sites covered an area between 675 to 2,700 square kilometers (see Table 1). For Guatemala and Laos training regions were slightly bigger than testing regions, whereas for Malawi the opposite was the case. Regarding the testing sites, all case studies showed an imbalanced proportion of "no building" and "building" tiles. This imbalance was very strong towards "no building" tiles for Guatemala and Laos (87%), and less marked for Malawi (56%).

All study regions have been mapped by OpenStreetMap users in response to requests by the Humanitarian OpenStreetMap Team (HOT), The Netherlands Red Cross and the Clinton Health Initiative. Our analysis was based on the tile level. In this study a tile corresponded to the definition applied by tile map services (TMS) at zoom level 18 (Maso et al., 2010). Hence, each study regions was made of several thousand tiles. A tile covered around 0.02 square kilometers (0.15×0.15 kilometers) with slight variations depending on geographic latitude. Table 1 summarizes the details about the training and testing areas for each study site.

The Guatemala region was made up by several different rather contrary regions. The northern and western areas were slopes of volcanic mountains covered with dense forests, whereas urbanized areas and agricultural land covered the valleys. Most settlements were part of a compact city with only few

buildings lying within farmland. The Laos region was characterized by dense forested areas covering more than 90% of the total area and by a few smaller settlements along the road network. Besides those villages, many buildings were distributed sparsely over the entire area accompanied by smaller patches of agricultural land. For Malawi larger cities were missing. The region was characterized by many smaller villages and intensive agricultural land use. Only a very small fraction of the area was covered by forests.

3.2 Data sets

This section describes the four datasets we used to evaluate the performance of our workflow. The reference data set was derived from OpenStreetMap and depicts the presence of human settlements in a satellite imagery tile. Building footprint geometries were directly obtained from the OpenStreetMap database using the Overpass API by filtering for all OSM ways tagged with the key "building" for each study site. To ensure the validity of the OSM data, we intentionally selected study areas for which the mapping was organized through the HOT Tasking Manager tool and a validation had taken place. This validation procedure ensured the precision of the OSM data set. More important, humanitarian mapping experts carefully validated all tiles for which no human settlements have been mapped in OSM, but a positive result was given either in MapSwipe or predicted by the deep neural network. Following this additional manual validation approach ensured the completeness of the reference data set.

MapSwipe yielded results on the tile level. In MapSwipe a tile was also called a task. For each individual tile at least three different users contributed a binary label ("building", "no building"). These results have been aggregated using majority voting. Furthermore, for each task the proportion of building labels (MapSwipe score) on the total number of labels has been generated. For example, if two out of three volunteers classified a tile as "building", the aggregated label would be considered as "building" (MapSwipe score: 0.66). Further details on the MapSwipe data model can be found in Herfort et al. (2017).

The Global Urban Footprint has been derived fully automatically from TanDEM-X and TerraSAR-X radar images with 3 m ground resolution by the German Aerospace Center (DLR) (Esch et al., 2013). The imagery has been collected between 2011 and 2012. In this paper, we used the binary GUF settlement mask with a spatial resolution of 0.4", which corresponds to a ground resolution of around 12 meters at the equator.

The High Resolution Settlement Layer maps human settlements derived from high resolution satellite imagery (0.5 meters) by the Connectivity Lab at Facebook. The data has been produced for 18 countries using deep neural networks (Tiecke et al., 2017).

4 Methodology

The workflow we propose in this paper addresses the challenge of combining deep learning and crowd-sourcing to generate high-quality human settlement maps. Section 4.1 explains the DeepVGI method to automatically classify satellite imagery tiles into "building" and "no building". Section 4.2 explains the data quality evaluation procedure. In Section 4.3 we present the procedure to analyze spatial and non-spatial characteristics of misclassifications of the DeepVGI method. Finally, Section 4.4 investigates the proposed task allocation strategy and how the combined use of MapSwipe and DeepVGI affects performance and volunteer efforts.

Table 1: Study Sites Characteristics

		Guatemala	Laos	Malawi
Training	Area	929.0 km^2	1556.3 km^2	265.7 km^2
	Tiles	42833	72360	12408
Testing	Area	745.5 km^2	1136.6 km^2	410.3 km^2
	Tiles	34181	52796	19272
No Building Tiles		87%	87%	56%

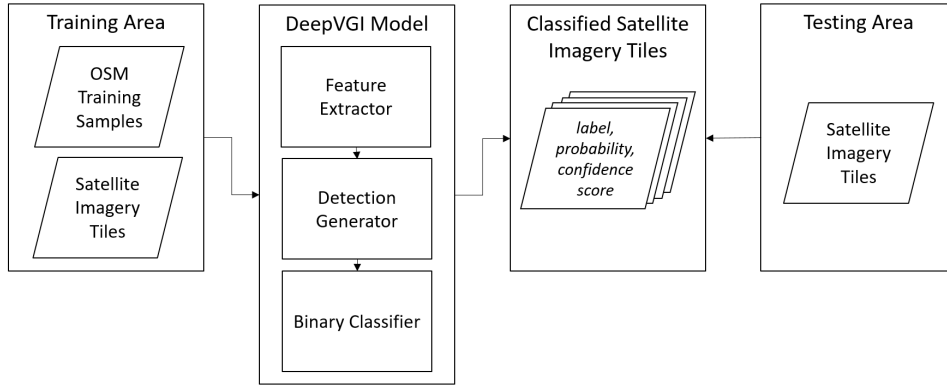


Figure 2: DeepVGI Workflow: The DeepVGI model is trained using building footprint sample from OpenStreetMap and satellite imagery tiles from Bing. For each tile in the testing area the model generates label, probability and confidence score.

4.1 Data preparation

We employed the DeepVGI method presented in Figure 2 for classifying satellite imagery tiles into "no building" and "building" classes. The DeepVGI building detection model consisted of three parts: feature extraction, object detection and binary classification.

Based on Single Shot Detection (SSD) networks (Liu et al., 2016) the model extracted heterogeneous features from either the base network or from extra layers. This enables SSD networks to better handle complex objects (such as buildings) of diverse scale and shape. For doing so, SSD networks apply the concept of tiling into default boxes so that specific feature maps learn to be responsive to particular scales of the objects (Liu et al., 2016).

The object detection generated a set of predicted building bounding boxes together with the corresponding probability scores. The implementation of the SSD network was based on the programming language Python 3.6 and the deep learning library Tensorflow (Abadi et al., 2015). For the initialization of the SSD parameters, a pre-trained network based on the Microsoft COCO data set (Lin et al., 2014) has been employed, which reported a mAP^{-1} (mean Average Precision) of 24. The maximum training epochs has been set to 60,000, and the initial learning rate is set to 0.0004 with a momentum of 0.9. The pre-trained network is available at the Tensorflow detection model zoo (Huang et al., 2017).

The object detection generated up to 50 bounding boxes per tile, however the majority of those bounding boxes did not represent buildings and were associated with very low probabilities. Since we were interested in a binary classification of a tile t into "no building" and "building" we only selected the highest bounding box score $P_{\max}(t)$ for each tile. From the training data set derived a classification threshold θ . Tiles with $P_{\max}(t)$ below θ were most probable to belong to the "no building" class. Vice versa tiles with $P_{\max}(t)$ above θ were most probable to belong to the "building" class. Based on θ we generated the label ("no building" and "building") for each tile in the testing data set.

Finally, we derived a confidence score δ_t for each tile by computing the absolute difference between the binary threshold θ and (deep learning) probability $P_{\max}(t)$. This was used as a proxy for how confident we can be in binary classification into either "building" or "no building". Taking the absolute value eased the visual interpretation. However, situations where the highest bounding box score was lower than the threshold (potential false negatives) could not be distinguished from those where the bounding box score was higher than the threshold (potential false positives) from δ_t alone. Hence, the analysis of the confidence score provided insights into accuracy of the DeepVGI approach, but not towards its specificity or sensitivity.

To train the DeepVGI model, building footprint samples from OpenStreetMap and satellite imagery tiles from Microsoft Bing were employed. In our experiment, satellite imagery tiles were collected by requesting a tile map service (TMS) from Microsoft Bing at zoom level 18. This corresponded to a spatial resolution of the displayed image of roughly 0.6 meters per pixel, as measured at the equator. The size of all image tiles was 256×256 pixels.

4.2 Overall Performance Evaluation

We evaluated our method by investigating the quality of the produced results against the reference data (see Section 3.2). Initially we derived the proportions of false negatives (FN), false positives (FP), true negatives (TN), and true positives (TP). To address the imbalance of building and no building labels in our study areas, we utilized the following metrics: specificity (TNR), sensitivity (TPR), and Matthews correlation coefficient (MCC). We further derive the accuracy (ACC). The statistics were computed as shown in Equation (1) through Equation (4). TNR, TPR and ACC are restricted between 0 and 1. MCC is in essence a correlation coefficient between the observed and predicted binary classification. It is bound between -1 and 1. For all statistics higher values indicate a better model fit (Luque et al., 2019).

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (1)$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (3)$$

$$\text{MCC} = \frac{(\text{TP} \times \text{TN}) - (\text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}} \quad (4)$$

Whereas ACC was mainly used to compare the results with works of other authors, MCC gives a more reliable indicator on quality for imbalanced data sets (Luque et al., 2019). TNR and TPR were utilized to investigate how well the analysed methods identify positives ("building" class) and negatives ("no building" class). This provided insights on the strengths and weaknesses of each method. Other metrics commonly applied for machine learning performance assessment such as F1 score or precision were not considered since they are highly biased for imbalanced data sets (Luque et al., 2019).

Additionally, we generated a map representation of the confusion matrix for each method and study site to spot spatial pattern in the false negatives and false positives.

4.3 Spatial and Non-Spatial Characteristics of Misclassifications

After investigating the overall performance of all methods, we conducted a detailed analysis of spatial and non-spatial characteristics of misclassifications of the DeepVGI approach. For doing so, we generated kernel density distribution plots of the confidence scores δ_t for (a) all tiles, (b) "building" tiles and (c) "no building" tiles. The implementation of the kernel density functions was based on the programming language Python, version 3.6 and the scipy library, version 1.1.0 (Jones et al., 01). The kernel bandwidth was set by applying Scott's rule (Scott, 2015). We generated a map representation of the distribution of δ_t to analyze the spatial characteristics of the confidence in the presence of buildings. This map should be interpreted together with the maps showing the distribution of false negatives and false positives.

Furthermore, we generated conditional density plots to visualize the conditional distribution of accuracy ACC in respect to δ_t and to compare the performance of the DeepVGI and MapSwipe approaches.

We tested if tasks with lower δ_t had a higher probability of being wrong by utilizing a logistic regression model for the DeepVGI and MapSwipe approaches, using δ_t as the predictor and Y as the response. Y was defined as "0" for wrong classifications (e.g., DeepVGI label and reference label were not the same) and "1" for correct classifications (e.g., DeepVGI label and reference label were the same). For the logistic regression we report on regression coefficient, standard error, significance, and McFadden's pseudo-r-squared values (McFadden et al., 1973). Those were computed based on the programming language Python 3.6 and the statsmodels library, version 0.9.0 (Seabold and Perktold, 2010).

4.4 Performance of Task Allocation Strategy

Finally, we proposed a task allocation strategy based on δ_t and compared the approach to a random allocation of tiles between DeepVGI and MapSwipe. The task allocation strategy defines for which tasks it is better to rely on results being produced by crowdsourcing and for which it is preferable to use the DeepVGI workflow. First, tiles were sorted ascending by δ_t . We generated the labels based on the fusion

of both approaches by choosing a proportion α of tiles which should be labeled by the crowd (we will refer to this as crowd proportion). For the remaining tiles we assigned the label of the DeepVGI method. Due to this design tiles with the lowest confidence were allocated to the crowd first.

We generated 100 set of the combined labels for each study site by choosing a crowd proportion α between [0,1] and adopting a step size of 0.01. For each α we further derived 250 random combinations of MapSwipe and DeepVGI. For each α we investigated the performance of both task allocation strategies in terms of accuracy ACC, Matthews correlation coefficient MCC, specificity TNR and sensitivity TPR (described in Section 4.2). The results were visualized in a graph depicting performance in relation to the crowd proportion α .

5 Results

5.1 Overall Performance Evaluation

This part of the results section describes the performance of different methods to create human settlement data sets. For all study sites we analyzed data from crowdsourcing (MapSwipe), deep neural networks (DeepVGI) and existing EO products (GUF). The HRSL data set was only available for the Guatemala and Malawi study sites.

For all three case studies MapSwipe performed best with respect to accuracy ACC (91% - 96%) and Matthews correlation coefficient MCC (80% - 83%) (see Table 2). The MapSwipe approach was characterized by high specificity TNR (99% for all study sites) and intermediate sensitivity TPR (75% - 82%). The spatial representation of the confusion matrices indicated that MapSwipe was able to correctly depict most "no building" tiles. Nevertheless, for the MapSwipe approach clusters of false negatives were observed, e.g., in the north-western part of the Guatemala study site (Figure 3) or in the south-eastern part of the Malawi study site (Figure 5). Thus, MapSwipe users were more likely to miss buildings (false negatives), than to map something as a building, which actually is not a building (false positives).

Regarding accuracy ACC the DeepVGI approach reached similar results to MapSwipe (91% - 96%), however the differences between those approaches became slightly more distinct when considering the less biased MCC (74% - 84%). The DeepVGI approach achieved lower specificity TNR compared to MapSwipe and GUF (95% - 97%), but higher sensitivity TPR (81% - 89%). Hence, false positives were expectedly the major concern for the DeepVGI approach. The visual interpretation of the confusion matrix maps confirmed higher spatial concentrations of false positives. For instance, clusters of false positives were present in the central and southern part of the Guatemala study site (Figure 3) or in the north-western part of the Malawi study site (Figure 5).

GUF was characterized by lower ACC (58% - 92%) and MCC (18% - 60%) compared to MapSwipe or DeepVGI. Regardless the very high specificity TNR (>99% for all study sites), the data set mapped only parts of the buildings as described by low sensitivity TPR (6% - 44%). The huge number of false negatives was the major drawback of the GUF approach. The map representation of the results for Guatemala (Figure 3) indicates that only major settlements were captured. For the rural study sites in Laos (Figure 4) and Malawi (Figure 5) the GUF approach generated many false negatives and thus missed most buildings.

The HRSL approach ranked below MapSwipe and DeepVGI in terms of accuracy ACC (86% - 90%) and MCC (69% - 73%). This approach reached the best results regarding sensitivity TPR (94% - 96%), but achieved only moderate specificity TNR (80% - 89%). Similar to the DeepVGI approach, the HRSL suffered from a high number of false positives. This is also depicted in the maps for Guatemala (Figure 3) and Malawi (Figure 5). It seems that false positives were located at the edges of correctly identified building tiles and along the road network.

5.2 Spatial and Non-Spatial Characteristics of Misclassifications

This section focuses on the analysis of the characteristics of misclassified tiles of the DeepVGI approach.

For all study sites, the distribution density of "no building" tiles showed a clear peak, whereas the distribution of "building" tiles was much more flat. For instance for the Guatemala study site, Figure 6 (center) depicts that the distribution density of "no building" tiles peaked for a confidence score of around

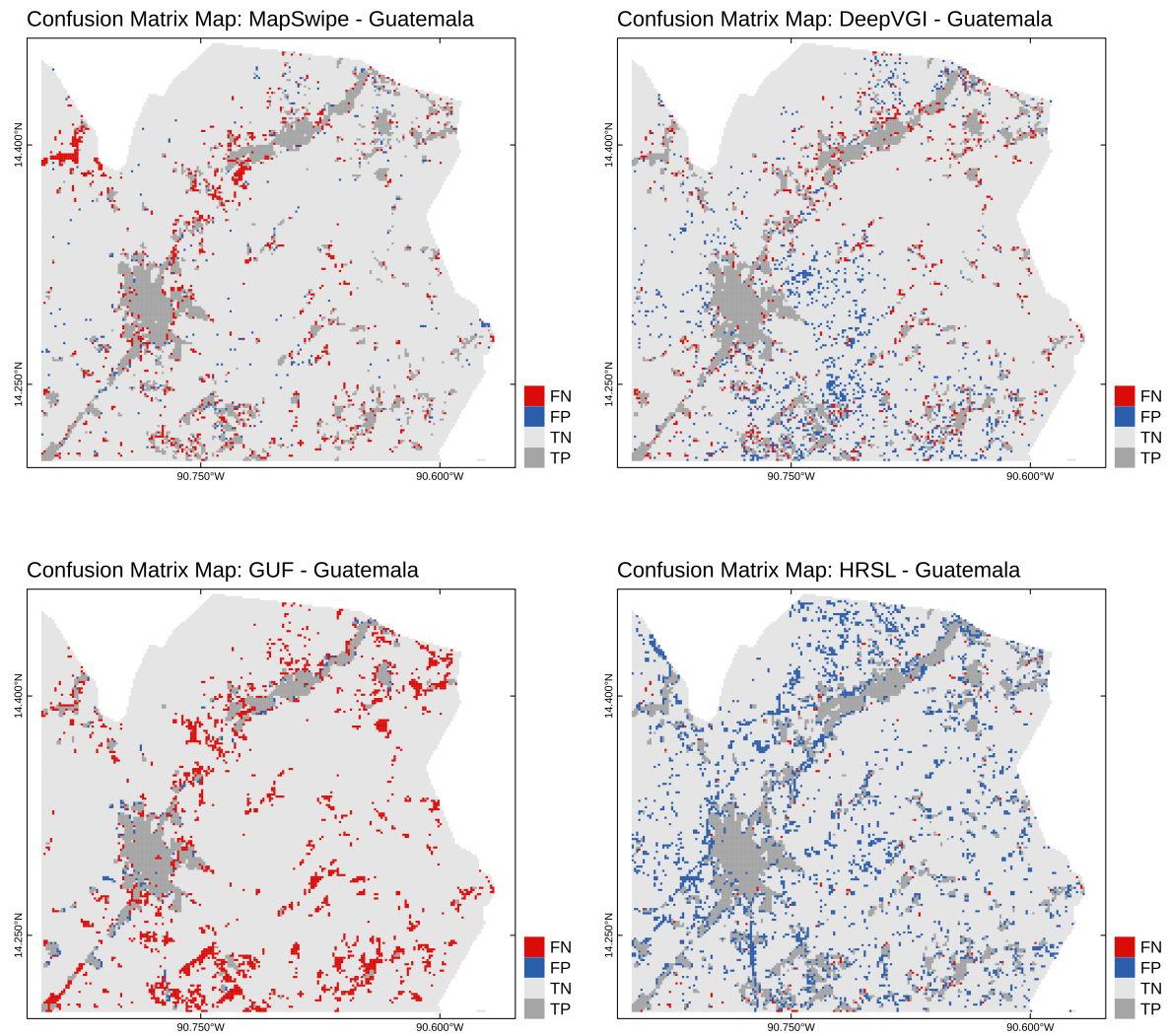


Figure 3: Map representation of the confusion matrix for Guatemala: Each map shows the spatial distribution of correct "building" (TP) and "no building" (TN) classifications for a specific method in the testing area. Incorrect classifications are split into false positives (FP) and false negatives (FN). Each pixel corresponds to a single task / satellite imagery tile.

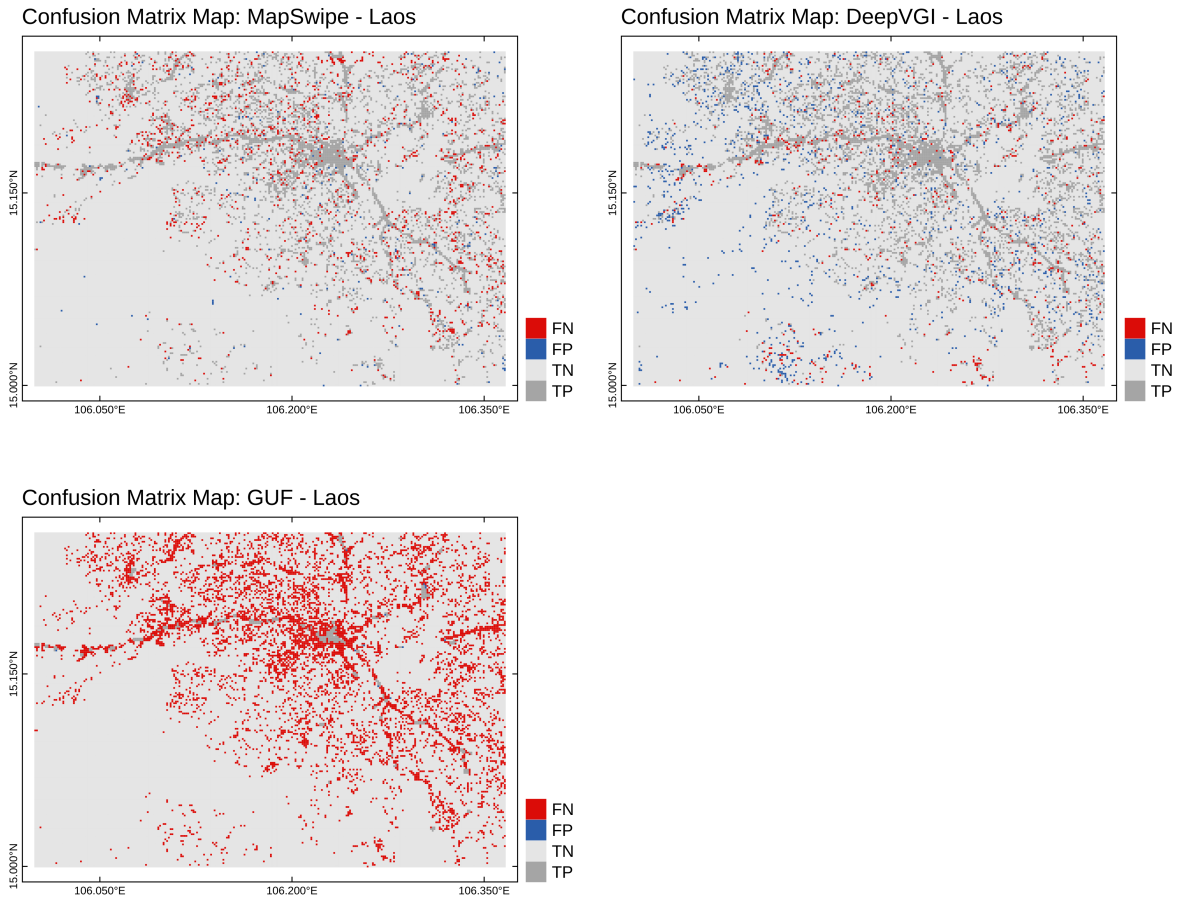


Figure 4: Map representation of the confusion matrix for Laos: Each map shows the spatial distribution of correct "building" (TP) and "no building" (TN) classifications for a specific method in the testing area. Incorrect classifications are split into false positives (FP) and false negatives (FN). Each pixel corresponds to a single task / satellite imagery tile.

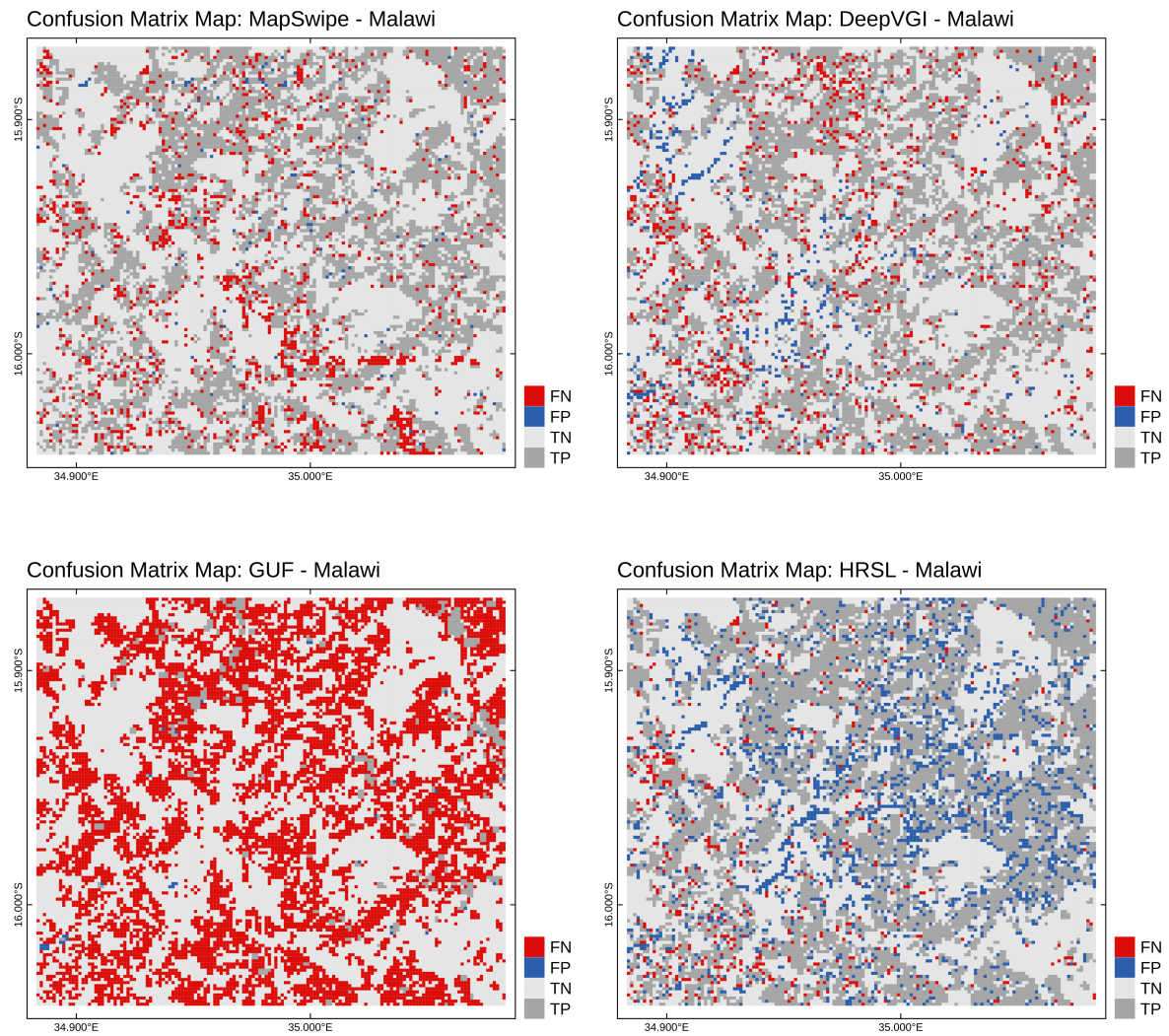


Figure 5: Map representation of the confusion matrix for Malawi: Each map shows the spatial distribution of correct "building" (TP) and "no building" (TN) classifications for a specific method in the testing area. Incorrect classifications are split into false positives (FP) and false negatives (FN). Each pixel corresponds to a single task / satellite imagery tile.

Table 2: Performance of different methods to generate human settlement data sets

		Guatemala	Laos	Malawi
TNR	MapSwipe	0.99	0.99	0.99
	DeepVGI	0.96	0.97	0.95
	GUF	0.99	0.99	0.99
	HRSL	0.89	-	0.80
TPR	MapSwipe	0.74	0.79	0.82
	DeepVGI	0.81	0.89	0.85
	GUF	0.44	0.06	0.07
	HRSL	0.96	-	0.94
ACC	MapSwipe	0.96	0.97	0.91
	DeepVGI	0.94	0.96	0.91
	GUF	0.92	0.87	0.58
	HRSL	0.90	-	0.86
MCC	MapSwipe	0.80	0.85	0.83
	DeepVGI	0.74	0.84	0.81
	GUF	0.60	0.22	0.18
	HRSL	0.69	-	0.73

0.25. "Building" tiles distributed equally between confidence scores of 0.0 and 0.45. Very high confidence scores (>0.3) are observed only for "building" tiles. At the same time, for very low confidence scores (<0.15) both "no building" and "building" tiles were present.

The conditional density plots (Figure 6, blue axis) revealed the tendency that the accuracy of the DeepVGI approach increased with higher confidence scores. For all three study sites, accuracy increased steadily from around 50% - 60% to more than 95% when raising the confidence score from 0.0 to around 0.35. There was no such clear trend for the conditional density of the accuracy for the MapSwipe approach. MapSwipe's accuracy ranged between 80% and 95% and indicated no dependency from the confidence score. Additionally, the comparison of the conditional density plots from DeepVGI and MapSwipe underlined that tiles that were relatively easy for DeepVGI (high confidence scores) were on average not easy for MapSwipe users. Vice versa, tiles that were difficult for the DeepVGI approach (lower confidence scores), were on average not particularly more difficult for MapSwipe users. For example for the Guatemala study site, the DeepVGI approach provided more accurate results than MapSwipe for tiles with a confidence score higher than 0.25. For tiles with a confidence score below 0.2 MapSwipe reached higher accuracy.

The spatial distribution of the confidence score revealed highest confidence scores for major settlements for all study sites. The most uncertain results were located in the areas with a mixed land-use, e.g., agricultural land and settlements. Furthermore, uncertain results seemed to be clustered, e.g., for the Laos study site (Figure 6) in most of the southern part of the study area and in a smaller area in the northern part.

The indications from the conditional density plot were confirmed by the results of the logistic regression analysis (Table 3): as depicted by the regression coefficient and the corresponding standard error and significance with increasing confidence score, probability of correct classifications increased significantly. However, as indicated by the low pseudo-r-squared values of (0.109 - 0.210) the confidence score explained only a fraction of all misclassified tiles. The results further confirmed that the confidence score had no explanatory power with respect to the accuracy of the MapSwipe approach (McFadden's Pseudo-r-squared of -0.211 to -0.130).

5.3 Combination of Crowdsourcing and Deep Learning

Figure 7 shows performance and effort in respect to crowd proportion. For all study sites allocating 10% to 20% of the tiles to MapSwipe (raising the crowd proportion from 0.0 to around 0.1 - 0.2) resulted in an overall performance increase in respect to accuracy ACC and Matthew's correlation coefficient MCC. Reducing the volunteer efforts to one fifths (labor reduction of 80 percentage points) resulted in a

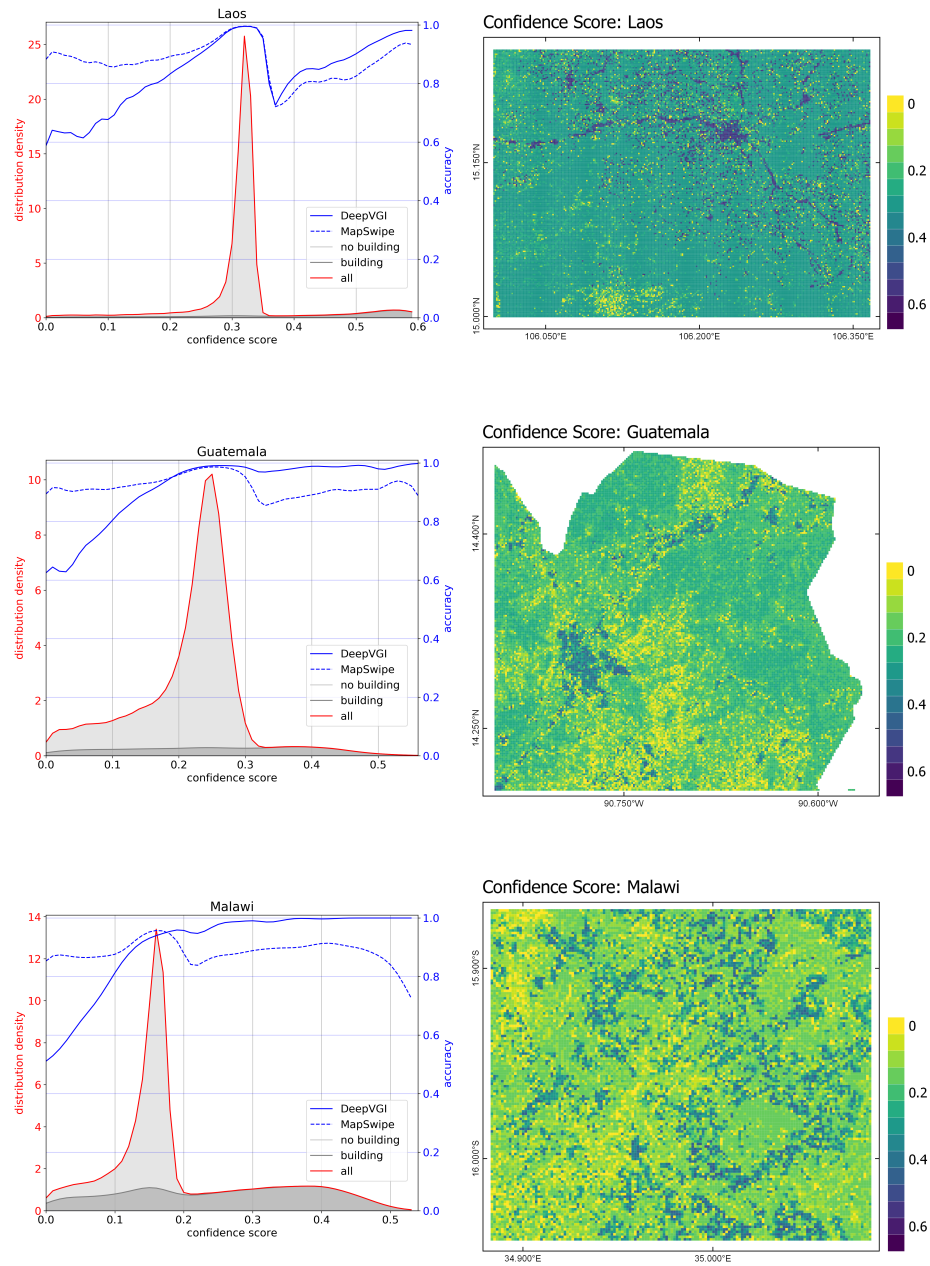


Figure 6: Spatial and Non-Spatial Distribution of the Confidence Score and Conditional Density of Accuracy for Laos, Guatemala and Malawi. The figure provides insights on the kernel density distributions of the confidence scores of "all" (red line), "no building" (light gray) and "building" (dark gray) tiles for the different study regions (red axis). The same plot shows the conditional distribution of accuracy (blue axis) for DeepVGI (solid blue line) and MapSwipe (dashed blue line). The map depicts the spatial distribution of the confidence scores.

Table 3: Logistic Regression analysis

		Guatemala	Laos	Malawi
DeepVGI	Coefficient	16.262	11.164	16.291
	Standard Error	0.165	0.086	0.210
	Significance	0.0***	0.0***	0.0***
	McFadden's Pseudo-r-squared	0.235	0.109	0.194
MapSwipe	Coefficient	15.224	10.454	11.676
	Standard Error	0.149	0.078	0.149
	Significance	0.0***	0.0***	0.0***
	McFadden's Pseudo-r-squared	-0.13	-0.154	-0.211

performance gain of 3-5 percentage points measured as MCC in all regions. For all study sites, this was caused mainly due to an increase in TNR (compared to the DeepVGI-only approach). For Guatemala and Laos, TPR remained stable, whereas for Malawi TPR increased as well (compared to the DeepVGI-only approach).

Vice versa, allocating 10% to 30% of the tiles to DeepVGI (decreasing the crowd proportion from 1.0 to around 0.7 - 0.9) also resulted in an overall performance gain. For all study sites, a gain in TPR (compared to the MapSwipe-only approach) was observed the more tiles have been allocated to DeepVGI. For Guatemala TNR remained stable, whereas for Laos and Malawi TNR decreased slightly at the same time.

The performance of the combined approach did not change, when allocating 20% to 70% of the sorted tiles to MapSwipe. For crowd proportions between 0.2 and 0.7 ACC, MCC, TNR and TPR remained mainly stable at a higher level compared to DeepVGI-only or MapSwipe-only approaches. For instance for the Guatemala study site, MCC gained around 10 percentage points (75% to 85%) compared to the performance of the DeepVGI-only approach by allocating 30% of the tiles to MapSwipe and 70% to DeepVGI. At the same time this resulted in an increase in MCC of around 5 percentage points (80% to 85%) and labor reduction of 70 percentage points compared to the performance of the MapSwipe-only approach. The MCC for a crowd proportion of 0.7 varied only slightly and reached around 84%.

When investigating the results of the random task allocation (see Figure 7) no such effects were observed. An increase in crowd proportion resulted in an improved TNR and decreased TPR with an uniform gradient for all study sites. Overall, ACC and MCC improved slightly with a homogeneous slope when allocation tasks from DeepVGI to MapSwipe. However, no performance gain was observed compared to the MapSwipe-only approach.

6 Discussion

6.1 Overall Performance Evaluation

The crowdsourcing approach by MapSwipe generated the most accurate human settlement maps for all case study sites with Matthew's correlation coefficient between 80% to 85%. As expressed by the differences in specificity TNR and sensitivity TPR between the MapSwipe approach and the DeepVGI approach, our results supported the common hypothesis that humans rarely identify something as a building which is not a building, but tend to miss buildings. In contrast, our results indicate that deep learning approaches tend to miss miss fewer buildings at the cost of falsely detecting also a range of objects which are not buildings. In comparison to GUF, crowdsourcing and deep learning based approaches demonstrated an improvement in data quality for our case study sites.

The performance of MapSwipe seemed to be consistent with previous findings from Albuquerque et al. (2016), where an accuracy of 89%, a sensitivity of 73% and a precision of 89% are achieved for a very similar crowdsourced classification task for a study site in South Kivu (Democratic Republic of the Congo). For the case of automated village boundary detection Gueguen et al. (2017) report an average precision of around 70% and a sensitivity of around 84%. Yuan et al. (2018) compare a deep learning based approach against GUF and GHSL for Kano city (Nigeria) and reach similar results. Their building

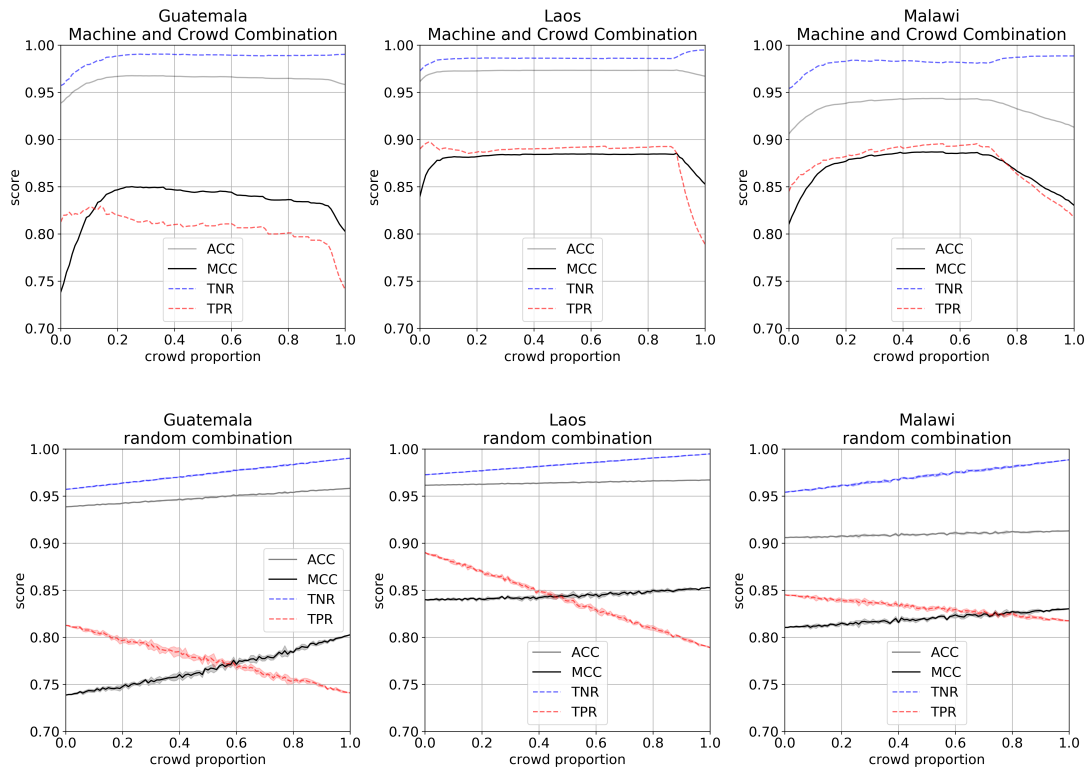


Figure 7: Data quality and effort in respect to crowd proportion and task allocation strategy for a combined MapSwipe-DeepVGI methods. The upper side of the figure represents the performance of the confidence score based task allocation strategy, whereas on the bottom side the mean and standard deviation of the performance of 250 random allocations between DeepVGI and MapSwipe are shown. Performance is measured by accuracy (ACC), Matthew’s correlation coefficient (MCC), sensitivity (TPR) and specificity (TNR). The x-axis of the plot shows the crowd proportion. A crowd proportion of 0.0 (left) refers to DeepVGI. Consequently, the performance for a crowd proportion of 1.0 (right) refers to the performance of MapSwipe. For a crowd proportion of 0.3, 30% of the results are obtained from MapSwipe (these tiles refer to the DeepVGI tiles with the lowest confidence) and 70% from DeepVGI.

extraction algorithm performs with a precision of 72% and a sensitivity of 70%. Their approach slightly improves the GUF, but significantly outperforms the Global Human Settlement Layer. Klotz et al. (2016) show that GUF and GHSL significantly increased the completeness and precision of global build-area maps in comparison to previous low-resolution products such as MOD500 or GLOBEC. Nevertheless, the authors also point out quantifiable weaknesses in rural areas, which could be confirmed by our study results as well. Whereas the GUF was of moderate quality for the sub-urban Guatemala study site, its weakness for the rural study sites in Laos and Malawi was immense.

Nevertheless, our results also suffered from limitations in our data sets and methods applied. Due to the imbalance of "no building" and "building" tiles, the accuracy ACC reported was biased towards identifying "no building" tiles correctly. This imbalance was strong for Guatemala and Laos, but less pronounced for Malawi. Whereas TNR and TPR show no bias, Matthew's correlation coefficient is biased as well (but not as strong as accuracy) (Luque et al., 2019). This reduces the comparability of our results with the findings from other studies with less imbalanced data.

In this study we did not investigate the effects of the imbalance on the training procedure of the DeepVGI approach. Furthermore, we decided to utilize a very specific network architecture and pre-trained model (SSD based on COCO data set, see Section 4.1). Our two-step approach (object detection first, then binary classification) also introduced further uncertainties. Whereas our results seemed to be consistent with the findings from other studies, further research is necessary to fully understand the impact of the data preparation on performance. For instance, Tiecke et al. (2017) present a computer vision method to create population maps from satellite imagery with a very high resolution and provide further insights on potential systematic errors. The authors highlight the problem related to repetitive errors, such as the misinterpretation of large rocks, boats or mountain ridges as buildings. Analyzing the potential sources of systematic errors would be of great benefit for our study to understand spatial clusters of false positives. New advances in machine learning research might produce architectures and training data sets which suit better to the specific use case of mapping human settlements. To reduce uncertainties in our approach, future studies should compare different architectures and training data set characteristics also with regard to imbalanced classes.

Uncertainties were also present for the results of the MapSwipe approach. Using majority aggregation to generate binary labels from the individual user classifications favoured higher specificity TNR, whereas choosing another method might have promoted higher TPR. The drawbacks of majority agreement are well described by Salk et al. (2016). The confusion matrix maps showed that wrong classifications were not randomly distributed, but revealed spatial pattern. For MapSwipe this has already been confirmed in Herfort et al. (2017). For our study, this implies that individual user behaviour, geographical characteristics of the surrounding of building features might be major causes of wrong classifications limiting the transferability of this approach. A more detailed analysis is necessary to understand the factors which drive the quality of crowdsourcing and its implications for human settlement maps.

Satellite imagery quality is another major concern for both crowdsourcing and deep learning approaches. Better satellite imagery (e.g., in terms of resolution), might strongly influence the performance of MapSwipe and DeepVGI. Our current study was limited to satellite imagery tiles at zoom level 18. For the regions analysed in this study, satellite imagery tiles with a higher image resolution at zoom level 19 were not available from Bing Maps. Nevertheless, new earth observation satellites such as WorldView3 would potentially provide sufficient imagery data for this zoom level. The scarce availability of up-to-date satellite imagery, emphasizes another drawback of MapSwipe and DeepVGI: the quality of the human settlement maps is closely tied to the structures visible in the satellite images. In situations where settlement pattern change rapidly, e.g., due to forced displacement, on the ground data is irreplaceable unless up-to-date satellite data becomes available.

6.2 Spatial and Non-Spatial Characteristics of Misclassifications

Our analysis provided insights into the spatial and non-spatial characteristics of misclassified tiles of the DeepVGI method. For all study sites the conditional density plots and logistic regression analyses revealed a significant correlation between confidence score and accuracy. Additionally, the results showed no such correlation for the MapSwipe data set. This indicates that DeepVGI and MapSwipe tended to detect different tiles with different characteristics at varying accuracy. We interpreted this as potential complementary value of both approaches.

However, the limited explanatory power of accuracy for imbalanced data sets needs to be considered for our study (especially for Guatemala and Laos). Hence, the increase in accuracy mainly depicted the correlation between specificity TNR and confidence score. The design of the confidence score (using absolute values, see Section 4.3) hampered the differentiation of false positives and false negatives. Whereas this simplification turned out to be beneficial for the logistic regression analysis, using a more sophisticated method, e.g., relying on a quadratic function, would have reduced the bias introduced due to class imbalance. Furthermore, higher confidence scores might be also related to the number of detected buildings per tile. In our approach, the confidence score was based only on the most probable building detection and did not consider multiple detections per tile. This approach increased uncertainties for tiles, for which only few buildings were located, e.g., in rural areas.

The results of the logistic regression analysis highlighted that confidence score contributed to the probability of tiles being misclassified, but only to a minor fraction. For the case of land cover mapping with a focus on urban areas Kampffmeyer et al. (2016) provide similar findings and show that pixels with low uncertainty are more likely to be classified correctly. As in our study, areas of class boundaries were a cause of wrong classifications. But, our study showed as well that the majority of misclassified tiles had a different cause not captured by our design. Our methods fell short especially in differentiating the confidence of "no building" classifications, which however constituted the majority of tasks in our study areas.

Further research is needed to expand our understanding of the systematic errors underlying our approach. For example, we would investigate to what degree deep learning based methods are able to map various building types, e.g., in relation to shape and size or to characteristics related to ethnic or social groups.

6.3 Combination of Crowdsourcing and Deep Learning

Combining the MapSwipe and DeepVGI methods using the confidence score based task allocation strategy increased performance by around 3 - 5 percentage points measured by MCC (compared to the MapSwipe-only approach). At the same time, the approach reduced the volunteer efforts to one fifth (labor reduction of 80 percentage points). Our results suggest that the task allocation strategy helped to exploit the complementary value of a sensitive method (DeepVGI) and a specific method (MapSwipe) and would improve the existing crowdsourcing approach MapSwipe uses.

For a similar set up but limited geographic scope Chen et al. (2018) show that a combination of results from machine learning and crowdsourcing can result in an labor reduction of 85 percentage points and achieves a similar accuracy. Gueguen et al. (2017) report similar findings in regard to semi-supervised village boundary mapping and are able to improve the precision of automated data extraction by introducing a crowdsourced validation step.

The combination of crowdsourcing and deep learning showed promising results for our selected case studies, but limitations of the presented workflow have to be considered. First, the individual performance of both methods might have a great impact on the performance of the combined approach. Our approach was able to improve results because of the complementary characteristics MapSwipe (high TNR) and DeepVGI (high TPR) hold. Due to the uncertainties of crowdsourcing and deep learning (discussed in Section 6.1) these differences between the two methods might be less pronounced or even reversed in other regions.

Considering the needs of humanitarian organizations (e.g., getting information on human settlements for which no other data sets exist) our approach could be used in real applications. However, the current workflow was not able to provide a clear estimation which crowd proportion would generate optimal results beforehand. Project managers organizing humanitarian mapping campaigns would still need to adjust the proportion of tasks mapped by the crowd manually, e.g., in respect to the given time frame and complexity of the mapping task.

We tested the workflow for three rather diverse study sites, nevertheless a more detailed investigation of the influence of geographic characteristics is necessary. Quantifying the differences between the study sites (e.g., in respect to land cover) would be a first step towards contextualizing the results. Together with an analysis of the quality of the satellite imagery and quality of the crowdsourced classifications this would help to understand for which study site characteristics a combination might result in better data.

7 Conclusions

Human settlement maps produced by crowdsourcing (MapSwipe) or deep learning (DeepVGI) showed large overlaps and for most areas both methods generated results with a similar accuracy. In general, both methods outperformed existing EO based products such as the Global Urban Footprint in terms of MCC and TPR. The proposed confidence score indicator helped to explain misclassified tiles of the DeepVGI method and revealed the complementary value of DeepVGI and MapSwipe. Combining crowdsourcing and deep learning by applying the proposed task allocation strategy facilitated the complementary values of both methods and provided a promising extension to the existing crowdsourcing approach MapSwipe incorporates.

Further research needs to validate these findings also for other study regions and various settlement types and shapes and contextual features such as vegetation and land use. The large amount of finished MapSwipe projects provides an obvious starting point for such an extended geographical analysis.

Our study focused on the binary classification of satellite imagery tiles to map human settlements. Due to the structure of the MapSwipe results, the analysis was limited to the tile level. Future research should overcome this drawback and investigate human settlement classifications at a more fine grained resolution and/or move on to investigating automatically generated building footprint geometries. Initial research in this direction has been conducted by Vargas-Muñoz et al. (2019), however national and regional investigations are necessary.

Reaching the targets of the sustainable development, planning disaster responses more efficiently and reducing the vulnerability of people at risk before disasters occur will remain challenges for the upcoming decades. The proposed combined use of satellite data, deep learning technology and citizen-based observations showed great potential to contribute to those efforts and future applications should consider the lessons learned from this research. However, to take fully advantage of the new opportunities there is a need to further understand the technical and non-technical challenges that come with them. The presented approach might help to identify data quality issues during or immediately after an object has been mapped to provide real-time or near real-time feedback for mappers. But, understanding and communicating the quality of automatically generated results and ensuring that tools and data are open and accessible are the very basis for this.

Integrating machine learning techniques into existing crowdsourcing applications will also create an increased need for technical know-how for project managers and data users. From the citizen science and VGI projects perspective, introducing these new techniques might also lead to an increased demand for experienced validators, which are already few in numbers nowadays. Whereas we envision to use these new tools to improve overall data quality and availability, they also constitute a new potential source of bias introduced into data sets such as OpenStreetMap. This bias might also be caused by class imbalances, which were present also in our study.

Taking all this into account, our results endorsed that for the creation of human settlement maps, we should rely on automated approaches (e.g., machine learning) when possible, but rely on human skills (e.g., citizens science and crowdsourcing) when needed.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
- Albuquerque, J., Herfort, B., and Eckle, M. (2016). The Tasks of the Crowd: A Typology of Tasks in Geographic Information Crowdsourcing and a Case Study in Humanitarian Mapping. *Remote Sensing*, 8(10):859.
- Ballatore, A. and Zipf, A. (2015). A Conceptual Quality Framework for Volunteered Geographic Information. In Fabrikant, S. I., Raubal, M., Bertolotto, M., Davies, C., Freundschuh, S., and Bell,

- S., editors, *Spatial Information Theory: 12th International Conference, COSIT 2015, Santa Fe, NM, USA, October 12-16, 2015, Proceedings*, pages 89–107. Springer International Publishing, Cham.
- Chen, J., Zhou, Y., Zipf, A., and Fan, H. (2018). Deep Learning From Multiple Crowds: A Case Study of Humanitarian Mapping. *IEEE Transactions on Geoscience and Remote Sensing*, PP:1–10.
- Chen, J. and Zipf, A. (2017). DeepVGI: Deep Learning with Volunteered Geographic Information. *WWW '17 Companion: Proceedings of the 26th International Conference Companion on World Wide Web*, (1):771–772.
- Comber, A., Mooney, P., Purves, R. S., Rocchini, D., and Walz, A. (2016). Crowdsourcing: It matters who the crowd are. The impacts of between group variations in recording land cover. *PLoS ONE*, 11(7):1–19.
- Crooks, A., Pfoser, D., Jenkins, A., Croitoru, A., Smith, D., Karagiorgou, S., Efentakis, A., Crooks, A., Pfoser, D., Jenkins, A., Croitoru, A., Smith, D., Karagiorgou, S., Efentakis, A., and Lamprianidis, G. (2015). Crowdsourcing urban form and function. *International Journal of Geographical Information Science*, 29(5):720–741.
- Degrossi, L. C. (2016). Potential of Collaborative Mapping for Disaster Relief : A Case Study of Open-StreetMap in the Nepal Earthquake 2015 Potential of Collaborative Mapping for Disaster Relief : A Case Study of. (JANUARY).
- Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S., and Strano, E. (2017). Breaking new ground in mapping human settlements from space – The Global Urban Footprint. *ISPRS Journal of Photogrammetry and Remote Sensing*, 134:30–42.
- Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenbock, H., Muller, A., and Dech, S. (2013). Urban footprint processor-Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. *IEEE Geoscience and Remote Sensing Letters*, 10(6):1617–1621.
- Fan, H., Zipf, A., Fu, Q., and Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28(4):700–719.
- Goodchild, M. F. and Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1:110–120.
- Gueguen, L., Koenig, J., Reeder, C., Barksdale, T., Saints, J., Stamatiou, K., Collins, J., and Johnston, C. (2017). Mapping Human Settlements and Population at Country Scale from VHR Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(2):524–538.
- Hachmann, S., Jokar Arsanjani, J., and Vaz, E. (2017). Spatial data for slum upgrading: Volunteered Geographic Information and the role of citizen science. *Habitat International*, pages 1–9.
- Hecht, R., Kunze, C., and Hachmann, S. (2013). Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *ISPRS International Journal of Geo-Information*, 2(4):1066–1091.
- Herfort, B., Reinmuth, M., de Albuquerque, J. P., and Zipf, A. (2017). Towards evaluating crowdsourced image classification on mobile devices to generate geographic information about human settlements. *Proceedings of the 20th AGILE*.
- Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., and Others (2017). Speed/accuracy trade-offs for modern convolutional object detectors. In *IEEE CVPR*, volume 4.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301).

- Jones, E., Oliphant, T., Peterson, P., et al. (2001–). SciPy: Open source scientific tools for Python. [Online; accessed *today*].
- Kampffmeyer, M., Salberg, A. B., and Jenssen, R. (2016). Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 680–688.
- Klotz, M., Kemper, T., Geiß, C., Esch, T., and Taubenböck, H. (2016). How good is the map? A multi-scale cross-comparison framework for global settlement layers: Evidence from Central Europe. *Remote Sensing of Environment*, 178:191–212.
- Li, H., Herfort, B., and Zipf, A. (2019). Estimating OpenStreetMap Missing Built-up Areas using Pre-trained Deep Neural Networks. *Proceedings of the 22nd AGILE*.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 740–755. IEEE.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., and Berg, A. C. (2016). SSD: Single shot multibox detector. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9905 LNCS, pages 21–37.
- Luque, A., Carrasco, A., Martín, A., and de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91:216–231.
- Maso, J., Pomakis, K., and Julia, N. (2010). OpenGIS web map tile service implementation standard. *Open Geospatial Consortium Inc*, pages 4–6.
- McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.
- Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M., Kemper, T., Lu, L., Marin-Herrera, M. A., Ouzounis, G. K., Scavazzon, M., Soille, P., Syrris, V., and Zanchetta, L. (2013). A global human settlement layer from optical HR/VHR RS data: Concept and first results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(5):2102–2131.
- Salk, C. F., Sturn, T., See, L., and Fritz, S. (2016). Limitations of Majority Agreement in Crowdsourced Image Interpretation. *Transactions in GIS*, 00(00):n/a–n/a.
- Scholz, S., Knight, P., Eckle, M., Marx, S., and Zipf, A. (2018). Volunteered Geographic Information for Disaster Risk Reduction—The Missing Maps Approach and Its Potential within the Red Cross and Red Crescent Movement. *Remote Sensing*, 10(8):1239.
- Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.
- Seabold, S. and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.
- See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H.-Y., Milčinski, G., Nikšič, M., Painho, M., Pödör, A., Olteanu-Raimond, A.-M., and Rutzinger, M. (2016). Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *ISPRS International Journal of Geo-Information*, 5(5):55.
- Tiecke, T. G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E. B., and Dang, H.-A. H. (2017). Mapping the world population one building at a time.

- Tracewski, L., Bastin, L., and Fonte, C. C. (2017). Repurposing a deep learning network to filter and classify volunteered photographs for land cover and land use characterization. *Geo-spatial Information Science*, 20(3):252–268.
- United Nations (2015). Transforming our world: The 2030 agenda for sustainable development. Technical report.
- United Nations Department of Economic and Social Affairs Population Division (2018). World Urbanization Prospects: The 2018 Revision. Technical report.
- United Nations Office for Disaster Risk Reduction (2015). Sendai Framework for Disaster Risk Reduction 2015-2030. Technical report.
- Vakalopoulou, M., Karantza, K., Komodakis, N., and Paragios, N. (2015). Building detection in very high resolution multispectral data with deep learning features. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2015-Novem:1873–1876.
- Vargas-Muñoz, J. E., Lobry, S., Falcão, A. X., and Tuia, D. (2019). Correcting rural building annotations in OpenStreetMap using convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147(November 2018):283–293.
- Yuan, J., Roy Chowdhury, P. K., McKee, J., Yang, H. L., Weaver, J., and Bhaduri, B. (2018). Exploiting deep learning and volunteered geographic information for mapping buildings in Kano, Nigeria. *Scientific Data*, 5:180217.
- Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., and Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36.
- Zook, M., Graham, M., Shelton, T., and Gorman, S. (2010). Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. *World Medical & Health Policy*, 2(2):6–32.

Part III

Declarations



**Eidesstattliche Versicherung gemäß § 8 der Promotionsordnung für die
Naturwissenschaftlich-Mathematische Gesamtfakultät der Universität Heidelberg / Sworn
Affidavit according to § 8 of the doctoral degree regulations of the Combined Faculty of
Natural Sciences and Mathematics**

1. Bei der eingereichten Dissertation zu dem Thema / *The thesis I have submitted entitled*

*Mapping what is not mapped: Investigating questions of representation and data quality in
OpenStreetMap from the perspectives of humanitarian and machine learning-assisted mapping*

handelt es sich um meine eigenständig erbrachte Leistung / *is my own work.*

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe
Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene
Inhalte als solche kenntlich gemacht. / *I have only used the sources indicated and have not made
unauthorised use of services of a third party. Where the work of others has been quoted or
reproduced, the source is always given.*
3. Die Arbeit oder Teile davon habe ich ~~wie folgt~~ bislang nicht ¹⁾ an einer Hochschule des In- oder
Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt. / *I have not yet/have
already¹⁾ presented this thesis or parts thereof to a university as part of an examination or degree.*

Titel der Arbeit / *Title of the thesis:*.....

Hochschule und Jahr / *University and year:*.....

Art der Prüfungs- oder Qualifikationsleistung / *Type of examination or degree:*.....

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich. / *I confirm that the declarations made
above are correct.*
5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen
oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. / *I am aware of the importance of
a sworn affidavit and the criminal prosecution in case of a false or incomplete affidavit*

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts
verschwiegen habe. / *I affirm that the above is the absolute truth to the best of my knowledge and that
I have not concealed anything.*

Heidelberg, den

Ort und Datum / *Place and date*

.....
Unterschrift / *Signature*

¹⁾ Nicht Zutreffendes streichen. Bei Bejahung sind anzugeben: der Titel der andernorts vorgelegten Arbeit, die
Hochschule, das Jahr der Vorlage und die Art der Prüfungs- oder Qualifikationsleistung. / *Please cross out what is
not applicable. If applicable, please provide: the title of the thesis that was presented elsewhere, the name of the
university, the year of presentation and the type of examination or degree.*

1. Publikation/Publication:

Vollständige bibliographische Referenz/Complete bibliographic reference:

Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J. & Zipf, A. The evolution of humanitarian mapping within the OpenStreetMap community. Sci. Reports 11, DOI: 10.1038/s41598-021-82404-z (2021).

2. Erst- oder gleichberechtigte Autorenschaft/First or equal authorship: Ja/Yes ☒ Nein/No ☐3. Veröffentlicht/Published ☒ Zur Veröffentlichung akzeptiert/Accepted ☐

Q1/Q2*:

Ja/Yes ☒ Nein/No ☐

*SCImago Journal Rank (SJR) indicator

Im Erscheinungsjahr oder im letzten verfügbaren Vorjahr/In the year of publication or the last prior year available: 2021: 1.005

Eingereicht/Submitted ☐ Noch nicht eingereicht/Not yet submitted ☐**4. Beteiligungen/Contributions****

Contributor Role	Doktorand/in/ Doctoral student	Co-Autor/in 1/ Co-author 1	Co-Autor/in 2/ Co-author 2
Name, first name	Benjamin Herfort	Sven Lautenbach	Joao Porto de Albuquerque
Methodology	☒	☒	☒
Software	☒	☐	☐
Validation	☐	☐	☐
Formal analysis	☒	☐	☐
Investigation	☒	☒	☐
Resources	☐	☐	☐
Data Curation	☒	☐	☐
Writing-Original Draft	☒	☐	☐
Writing-Review&Editing	☒	☒	☒
Visualization	☒	☒	☐
Supervision	☐	☒	☒
Project administration	☐	☐	☐
Funding acquisition	☐	☐	☐

**Kategorien des CRediT (Contributor Roles Taxonomy, <https://credit.niso.org/>)

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind/I confirm that all declarations made above are correct.

Unterschrift/Signature

Doktorand/in/Doctoral student

Co-Autor/in 1/Co-author 1

Co-Autor/in 2/Co-author 2

Betreuungsperson/Supervisor

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind und dass die selbstständigen Arbeitsanteile des/der Doktoranden/in an der aufgeführten Publikation hinreichend und signifikant sind/I confirm that all declarations made above are correct and that the doctoral student's independent contribution to this publication is significant and sufficient to be considered for the cumulative dissertation.

Alexander Zipf

Name/Name

Unterschrift/Signature

Datum/Date

1. Publikation/Publication:

Vollständige bibliographische Referenz/Complete bibliographic reference:

Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., and Zipf, A. (2023). Investigating completeness and inequalities in OpenStreetMap: spatio-temporal analysis of global urban building data. Under Review at Nature Communications. Preprint available at: <https://doi.org/10.21203/rs.3.rs-1913150/v1>

2. Erst- oder gleichberechtigte Autorenschaft/First or equal authorship: Ja/Yes ☒ Nein/No ☐

3. Veröffentlicht/Published ☐ Zur Veröffentlichung akzeptiert/Accepted ☐

Q1/Q2*:

Ja/Yes ☒ Nein/No ☐

*SCImago Journal Rank (SJIR) indicator

Im Erscheinungsjahr oder im letzten verfügbaren Vorjahr/In the year of publication or the last prior year available: 2021: 4.846

Eingereicht/Submitted ☒ Noch nicht eingereicht/Not yet submitted ☐

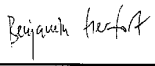
4. Beteiligungen/Contributions**

Contributor Role	Doktorand/in/ Doctoral student	Co-Autor/in 1/ Co-author 1	Co-Autor/in 2/ Co-author 2
Name, first name	Benjamin Herfort	Sven Lautenbach	Joao Porto de Albuquerque
Methodology	☒	☒	☒
Software	☒	☐	☐
Validation	☐	☐	☐
Formal analysis	☒	☐	☐
Investigation	☒	☐	☐
Resources	☐	☐	☐
Data Curation	☒	☐	☐
Writing-Original Draft	☒	☐	☐
Writing-Review&Editing	☒	☒	☒
Visualization	☒	☐	☐
Supervision	☐	☒	☒
Project administration	☐	☐	☐
Funding acquisition	☐	☐	☐

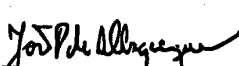
**Kategorien des CRediT (Contributor Roles Taxonomy, <https://credit.niso.org/>)

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind/I confirm that all declarations made above are correct.

Unterschrift/Signature


Doktorand/in/Doctoral student


Co-Autor/in 1/Co-author 1

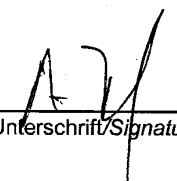

Co-Autor/in 2/Co-author 2

Betreuungsperson/Supervisor:

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind und dass die selbstständigen Arbeitsanteile des/der Doktoranden/in an der aufgeführten Publikation hinreichend und signifikant sind/I confirm that all declarations made above are correct and that the doctoral student's independent contribution to this publication is significant and sufficient to be considered for the cumulative dissertation.

Alexander Zipf

Name/Name


Unterschrift/Signature

5.5.23
Datum/Date

1. Publikation/Publication:

Vollständige bibliographische Referenz/Complete bibliographic reference:

Herfort, B., Li, H., Fendrich, S., Lautenbach, S. & Zipf, A. Mapping human settlements with higher accuracy and less volunteer efforts by combining crowdsourcing and deep learning. Remote. Sens. 11, DOI: 10.3390/rs11151799 (2019).

2. Erst- oder gleichberechtigte Autorenschaft/First or equal authorship: Ja/Yes ☒ Nein/No ☐3. Veröffentlicht/Published ☒ Zur Veröffentlichung akzeptiert/Accepted ☐

Q1/Q2*:

Ja/Yes ☒ Nein/No ☐

*SCImago Journal Rank (SJR) indicator

Im Erscheinungsjahr oder im letzten verfügbaren Vorjahr/In the year of publication or the last prior year available: 2019: 1.422

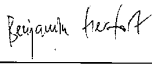
Eingereicht/Submitted ☐ Noch nicht eingereicht/Not yet submitted ☐**4. Beteiligungen/Contributions****

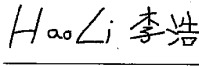
Contributor Role	Doktorand/in/ Doctoral student	Co-Autor/in 1/ Co-author 1	Co-Autor/in 2/ Co-author 2
Name, first name	Benjamin Herfort	Hao Li	Sven Lautenbach
Methodology	☒	☒	☐
Software	☒	☒	☐
Validation	☐	☐	☐
Formal analysis	☒	☐	☐
Investigation	☒	☒	☐
Resources	☐	☐	☐
Data Curation	☒	☐	☐
Writing-Original Draft	☒	☒	☐
Writing-Review&Editing	☒	☒	☒
Visualization	☒	☐	☐
Supervision	☐	☐	☒
Project administration	☐	☐	☐
Funding acquisition	☐	☐	☐

**Kategorien des CRediT (Contributor Roles Taxonomy, <https://credit.niso.org/>)

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind// confirm that all declarations made above are correct.

Unterschrift/Signature


 Doktorand/in/Doctoral student


 Co-Autor/in 1/Co-author 1


 Co-Autor/in 2/Co-author 2
Betreuungsperson/Supervisor:

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind und dass die selbstständigen Arbeitsanteile des/der Doktoranden/in an der aufgeführten Publikation hinreichend und signifikant sind// confirm that all declarations made above are correct and that the doctoral student's independent contribution to this publication is significant and sufficient to be considered for the cumulative dissertation.

Alexander Zipf

Name/Name


 Unterschrift/Signature

5.5.23

Datum/Date

