universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

The Anthropogenic Factor of Natural Disasters

Determination of potential correlations between soil sealing alongside the Danube River and the extent of the 2013 flood event using geographic citizen science and traditional methods for improving official disaster management

verfasst von / submitted by

Mario Zussner, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2017 / Vienna 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 856
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium / Master Programme Kartographie und Geoinformation UG2002
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Wolfgang Kainz

Table of Contents

List of Abbreviations

CS	Crowdsourcing
DEM	Digital Elevation Model
DMC	Disaster Management Cycle
DRR	Disaster Risk Reduction
EC	European Commission
EEA	European Environment Agency
FAO	Food and Agriculture Organization of the United Nations
FEMA	Federal Emergency Management Agency
FQA	FotoQuest Austria
FQG	FotoQuest Go Europe
GIS	Geographic Information System
IFRC	International Federation of Red Cross and Red Crescent Societies
IIASA	International Institute for Applied Systems Analysis
LUCAS	Land Use and Cover Area frame Sample
MYA	Million Years Ago
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
OSM	OpenStreetMap
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SVC	Supervised Classification
UNISDR	United Nations Office for Disaster Risk Reduction
USGS	United States Geological Survey
VGI	Volunteered Geographic Information

List of Figures

List of Tables

Abstract

The frequency of natural disasters increased measurably in recent years. As a result of the growing world population and the trend towards urbanisation, the effects on the environment and the society are becoming more serious. The investigation of vulnerability, risks and hazards should therefore be an essential part of disaster preparation and response, as well as spatial planning strategies. The aim of this master thesis is to address some of these elements with an investigation of anthropogenic influences on the extent and severity of natural disasters. This should be achieved by determining potential correlations between the sealing of the soil alongside the Danube River in Austria and the extent of the 2013 flood events. The analyses of this thesis are based on the geographic citizen science project FotoQuest Austria (FQA), conducted by the International Institute for Applied Systems Analysis (IIASA). The results of this campaign are processed and prepared in order to produce a suitable map of sealed surfaces in the research area. For determining the quality and accuracy of this crowdsourced data, the same analyses are conducted with traditionally generated open data. Although the results of the FQA campaign show good agreements with comparable established land cover datasets, the quantity of data is still matter for improvement. Recommendations for improving data quality are provided, so that geographic citizen science evolves into a reliable source of information for urban planning and disaster management.

Die Häufigkeit von Naturkatastrophen hat in den letzten Jahren stark zugenommen. Aufgrund der steigenden Weltbevölkerung und Verstädterung, werden die Auswirkungen auf die Umwelt und die Gesellschaft immer gravierender. Die Untersuchung von Angriffsflächen, Risiken und Gefahren sollten daher integraler Bestandteil von Katastrophenmanagements und der Raumplanung sein. Das Ziel dieser Masterarbeit ist es einige der genannten Elemente zu bearbeiten, indem die menschlichen Einflüsse auf die Ausbreitung und Schwere von Naturkatastrophen untersucht werden. Erreicht werden soll dies mittels der Bestimmung möglicher Korrelationen zwischen der Bodenversiegelung entlang der Donau und den Überflutungsflächen des Hochwassers von 2013. Die Analysen dieser Arbeit basieren auf dem Citizen Science Projekt FotoQuest Austria (FQA), durchgeführt vom Internationalen Institut für Angewandte Systemanalyse (IIASA). Diese Ergebnisse werden so bearbeitet, dass eine passende Karte aller versiegelten Flächen im Untersuchungsgebiet entsteht. Um die Qualität und Genauigkeit der Daten festzustellen, werden die gleichen Analysen auch mit traditionell erstellten freien Daten durchgeführt. Auch wenn die Ergebnisse des FQA Projekts gute Übereinstimmungen mit etablierten vergleichbaren Datensätzen aufweisen, könnte die Datenmenge noch weiter erhöht werden. Weitere Empfehlungen für die Verbesserung der Datenqualität werden am Ende dieser Arbeit bereitgestellt, damit sich die geographischen Bürgerwissenschaften (Citizen Science) in eine verlässliche Informationsquelle für die Raumplanung und das Katastrophenmanagement entwickeln können.

Acknowledgments

Right at the beginning I want to thank my supervisor Univ.-Prof. Dipl.-Ing. Dr. Wolfgang Kainz, whose calm and reflected comments and suggestions contributed a lot to this master thesis. His exemplary function and working methods surely had influences on my scientific orientation and on my academic future's path. I owe further thanks to all other members of the department of Geography and Regional Research at the University of Vienna and my fellow students, who supported me throughout the years.

My very grateful thanks also go to all the participating staff members at the International Institute for Applied Systems Analysis (IIASA) at Laxenburg. This non-governmental research organisation kindly provided me with most of the original data and contributed valuable inputs throughout all phases of this thesis. In this context, I want to acknowledge particularly the ever-present help and support by IIASA team members Christoph Perger MSc. and Dr. Ian McCallum. Without them, writing this thesis would have been a more difficult and lengthier process.

In addition to that, I also owe further thanks to Dr. Erik Formann, Mag. Michael Holzapfel, DI Ingo Niederbichler, DI Georg Puchner, DI Bernd Winkler and especially DI Christoph Hackel, which helped me during the acquisition of additional data.

Finally, I want to offer my regards and thanks to my family and friends, who offered me encouragement and mental support whilst writing this thesis, although they did not see me very often in the past few months.

1. Introduction

"A prudent person foresees the danger ahead and takes precautions.
The simpleton goes blindly on and suffers the consequences"

Book of Proverbs 27:12, Old Testament

For merely 4.5 billion (thousand million) years, planet Earth has been shaped by dynamic and random processes. Every part of it, we now observe and describe, was created purely by this unbiased system which is now called *nature*. It triggered all geophysical events, just like climatic warming and cooling, earthquakes, floods, landslides, sea-level risings, volcanic eruptions and wildfires. The "only" victims of these events were the prevailing flora and fauna [ALC-02]. For millions of years, different kinds of lifeforms developed and became extinct due to predators, pandemics or serious changes in their environment. This could comprise amongst others acid impact on the oceans, long glacial periods (e.g. snowball-earth), but also external sources of danger just like meteorite strikes. According to several scientific sources, there has been five major and several minor extinction events, which led to a severe mass disappearance of animals and plants [ALR-08]. The five most mentionable events are:

~ *The Ordovician-Silurian extinction* happened approximately 450 million years ago (mya) and eliminated about 85% of all Ordovician species [BRT-16d]. This makes it the second most deadly extinction event in history (of which science currently knows about). The catalyst is yet to question but theories suggest sudden glaciation, volcanic activities or a gamma-ray outburst.

~ *The Devonian extinctions* occurred about 360 mya and were a series of several global extinctions [BRT-16a] that focussed on the maritime lifeforms. The latest of these events was responsible for the elimination of more than 50% of the Earth's species. Possible reasons are rapid climate changes, excessive sedimentation, the loss of nutrient matter or a meteorite strike.

~ *The Permian-Triassic extinction* took place about 250 mya and is the deadliest of all the listed events. It is responsible for the elimination of approximately 95% of marine species and about 70% of terrestrial species. The duration differs (dependent on the study) between 200,000 and 15 million years [BRT-16e]. Possible causes comprise of temperature crises, changes in the carbon cycle, methane-hydrate gasification, volcanic events or the inevitable impact event.

~ *The Triassic-Jurassic extinction* occurred around 200 mya ranks fourth in the most deadly extinctions. About 75% of all species demised. Due to this event, the dinosaurs were able to become the most dominant animal on the planet [BRT-16b]. Possible reasons are considered in theories about sudden climate changes, the release of great amounts of CO_2, massive volcanic activities or meteorite strikes.

~ *The Cretaceous-Paleogene extinction event*, formerly known as KT-extinction, took place about 65 mya and is the most famous of all the mentioned events. It was responsible for eliminating about 80% of all living species [BRT-16c]. One of the well-known consequences of this event is the extinction of nearly all dinosaurs on Earth. This enabled the development of smaller mammals, which led to the evolution of modern humans. Theories suggest that the main reason for this extinction was an impact event on the Mexican peninsula Yucatan.

It is therefore quite obvious that there had always been different kinds of natural disasters in Earth's history. These major or minor catastrophes kept on threatening the existing lifeforms and either eliminated them or forced them to assimilate or relocate. This fascinating research area shows the importance to know, how the early species coped with these events. Nonetheless, the main focus of this master thesis lies on the effects of disasters on modern day humans.

Only a few millions ago, the first hominids evolved in Africa. The ancestors of today's apes and humans developed in parallel, until approximately three mya the first bipedal "Homo" appears in eastern Africa. The Homo habilis was the first of our ancestors who was able to create stone tools and use them. The Homo erectus was the next step of modern day humans and the modified physique enabled it to cover great distances, even beyond the African continent. The Homo heidelbergensis appeared about 600,000 years ago and developed a similar body structure to the Homo sapiens. All this time in the evolution of this species, the hominids had to adapt to changes in landscapes, climate or food availability. They invented new ways of hunting, gaining food, creating tools and shaping the natural scenery. And in this period, the human presence transformed normal geophysical events (e.g. earthquakes, floods, landslides, storms, wildfire, etc.) into natural disasters [ALC-02]. This transformation "occurred simultaneously with the appearance of the human system, when human beings began to interact with nature […]" and when the "[…] evolution of humans left behind the age in which only nature existed" [ALC-02]. Catastrophic events do happen eventually, but the disaster follows because of anthropogenic interaction or rather the lack of it [WIS-01]. From this period on, every geophysical event or meteorite strike would have had an impact on human societies and their development. In this time of human and natural evolution, the dynamics of natural hazards evolved.

Due to this very specific development, it is clear why certain events are called disasters. But what are the reasons for a disaster? To answer this question, it is necessary to know the definition of the term first. In certain religions or cults, catastrophic events are responses from the gods and the stars. This kind of view is not as unrealistic as it seems, because the word "disaster" derives from the Greek term for "bad star or planet" and refers to an unfavourable constellation of celestial bodies [ETY-16a]. In scientific terms, a disaster is "a sudden, calamitous event that seriously disrupts the functioning of a community or society and causes human, material, and economic or environmental losses that exceed the community's or society's ability to cope using its own resources", according to the International Federation of Red Cross and Red Crescent Societies [IFR-16]. The most commonly used term is *natural disaster*, but catastrophic events can be natural, man-made or a combination of both. The geophysical and anthropogenic influences in the disaster extent, damage or frequency however are matter of the following second chapter.

The number of catastrophic events which qualify as a natural disaster has risen over the last few decades and the need for a comprehensive disaster management grew, because every emergency response agency knows: *after the disaster is before the next disaster*. Because of this recurring theme, the concept of pre-event planning for post-event recovery has been developed in the late 1980ies. This led to the use of the Disaster Management Cycle (DMC) in disaster risk management. It comprises phases like *disaster response, recovery, mitigation, risk reduction and preparedness*.

This master thesis will focus on the phase of flood risk reduction, so possible future consequences of such catastrophic events may be abated. In this case, not only coastal areas are affected because of the rising sea level, but also river basins all over the world. The flood volume and the extent of damage caused are also increasing as a further consequence. The research of this thesis should offer valuable clues to the extent and distribution of endangered areas in case of a flood in the research area, the Danube River in Austria. The last bigger spate in Austria and Central Europe took place in June 2013 and was caused by heavy rainfalls. Only due to the improved protection facilities after the catastrophic flood from 2002, the extent of the damage was able to be kept within a certain bearable limit.

In the last decades, more and more people moved to cities and to coastal areas. This increases the chances to be more affected by catastrophic flood events. It is important to develop faster and better ways to receive suitable information during, after and before the event to reduce the risk for the next time. With the help of modern remote sensing technologies and voluntarily generated geoinformation, realised by the International Institute for Applied Systems Analysis (IIASA), it should be possible to determine the extent of vulnerable regions in the research area due to their level of soil sealing and flood protection. Citizen science and VGI should be presented as fast and economic alternatives to generate and assess information. There is no professional training needed and its usability can easily be adapted to different kinds of catastrophes. The combination of these approaches, remote sensing technologies and different sources of additional data can facilitate the autochthonous methods of coping with disasters.

Nobody can protect oneself absolutely from disaster and because of that, it should be logical, that every scientific branch that can contribute their knowledge and tools to improve risk reduction and disaster response, should do so. In the following chapters, the author of this thesis will investigate faster ways of detecting the location of possible flood hazard zones and how this information can possibly be used to reduce the risk by implementing this data into urban planning policies.

1.1. Problem Description and Goals of the Thesis

Based on this thematic background, the main objective of the thesis is to highlight the usability of geographic information science in the field of disaster management and risk prevention. The scientific passion of this thesis' author is to show, how the latest developments in modern remote sensing techniques can improve current disaster management and urban planning, together with innovative methods of data generation. We are currently facing a time, where the global climate is changing measurably and the frequency of natural disasters is rising. The occurrences and effects of such catastrophic events are often threatening highly populated areas and are even capable of influencing the economic and social development of an affected region [BEL-14]. Poorer countries just like Georgia for instance, owe about 70% of their human casualties and about 65% of their economic loss to climate-induced catastrophes [KOR-11]. Therefore it is an essential need for every disaster management agency or rescue service to be able to work with adequate data and technologies, because this facilitates the execution and the coordination of rescue missions.

Data which is remotely sensed and evaluated can provide a valuable source of information at each stage of disaster management. This information can also provide experts and policy makers with objective data sources for making decision and adapting urban planning strategies [JOY-09]. Remote sensing via satellite-based systems and airborne photogrammetry developed intensively in the past few decades. A number of sensors are already currently available (e.g. thermal or optical sensors, synthetic aperture radar) and can be used during various disaster operations. This technology has the potential to generate, prepare and provide data to "[…] assist risk reduction initiatives through identification of hazard zones […]" in a faster and more accurate way [JOY-09]. In combination with common geographic information systems it offers the possibility to accelerate the data generation and to reduce any given inaccuracies [HUY-14]. These methods of acquiring and assessing data should support research which mainly concentrates on the detection of affected areas in case of a disaster and possible damages. The main objective of this master thesis focusses on the detection of vulnerable areas close to the Danube River, due to the imperviousness of the ground alongside of it.

Due to all sorts of climatic changes and anthropogenic interferences, the number of flood events affecting densely populated areas rises every year. It seems inevitable for concerned decision makers to adapt their land use regulations and their urban planning policies to these global changes. One approach to tackle these challenges is to investigate possible linkages between the degree of soil sealing in the research area and the correlation to the extent and damage of recurring flood disasters. The sealing of soils can be described as the "permanent covering of the land surface by buildings, infrastructure or any impermeable artificial material" [PIS-15]. The use of impervious materials near waterbodies or rivers is increasing in developed countries. The consequences can vary from increasing number of flood events, loss of flora and fauna or pollution. This increase of flood risk is the reason, why there is room for investigation with technologies relating to geographic information systems, remote sensing and the cartographic toolkit.

Along the mentioned research possibilities, there are several sub-questions that occur during the analyses. Questions like: if there is a linkage, how does it affect the damages? Are there other possible ways to investigate these questions? To what extent can citizen science and VGI complement satellite or photo assessments? This last issue also leads straight to the next problem, which will be discussed in the course of this thesis. To be able to investigate the possible linkages between soil sealing and the extent of the flood event in 2013, a lot of information has to be acquired. Traditional research could either focus on a very small area or contents oneself with very imprecise examinations of bigger areas.

One possible solution is to outsource the task to lots of volunteers, which take their place as human sensors and contribute the required data, albeit not being scientifically trained in this field of research. This specific method is called crowdsourcing. Referring to its use in disaster management and risk prevention it is defined as "using the power of the internet and social media to "virtually" harness the power of individuals and bring them together in support of a disaster" [RIC-16]. If there is a huge amount of data involved, the assessment and interpretation of this information can take ages with common remote sensing techniques. Crowdsourcing and its geographic derivatives VGI (volunteered geographic information) and citizen science can produce, assess and interpret loads of information through non-expert volunteers, which take the place of sensors, software or qualified professionals. The motivation for such a cooperation could be incentives like money, vouchers, co-authorships in publications or in case of a disaster, the altruistic wish to help those who suffered.

This thesis wants to point out possible linkages between soil sealing and the extent and damages of a flood in urban areas, so planning agencies and policy makers can possibly adapt their regulations. Another aim of this thesis is to point out the anthropogenic factor of flood disasters in the research area and how it can be mitigated. Possible weaknesses in current flood risk management and flood protection may also be addressed as well as the benefits of using geographic citizen science to assess the situation and help improving the availability of information for the responsible authorities.

Crowdsourcing and its derivatives can be an enrichment not only for policy makers and disaster response agencies, but also for the assessment of possible risk factors and potential adaptions through planning officers. The expected results of this master thesis should help local authorities and other official bodies with their urban planning strategies in densely populated areas. These findings should offer a useful contribution for the improvement of flood risk management or risk reduction, and point out the strengths and weaknesses of the research approach.

1.2. Research Questions and Hypotheses

To solve the listed problems and realise the quite ambitious objectives, several separated analyses have to be done. To facilitate these tasks, the main higher-level question had to be split up into numerous sub-questions and hypotheses. The main title of this master thesis is based on the general research question and can be extensively specified as:

> *"The anthropogenic factor of natural disasters – can data from citizen science projects determine possible correlations between sealed surfaces alongside the Danube River and the extent and damage of the flood disaster in 2013? How can these results improve official urban planning strategies and disaster management?"*

Answering this extensive thematic block requires small analytic steps, which are based on the following working questions. They will be answered and adapted step by step in the course of this thesis:

> *"How can new technological achievements in the field of remote sensing, image processing and data acquisition (e.g. geographic citizen science) contribute valuable improvements in the field of disaster response and flood risk management?"*
>
> *"Are there possible linkages measurable between the imperviousness of the soil and the flood extent during the 2013 inundation of the Danube River in Austria?"*
>
> *"If there is a strong linkage between soil sealing and the flood extent, how does it affect the resultant economic, environmental and social damage?"*
>
> *"Which methods to investigate possible linkages were used by other studies? Did they succeed to analyse the imperviousness of the soil with satellite data or are ground photos more accurately?"*
>
> *"To what extent can crowdsourcing and its derivatives citizen science and volunteered geographic information complement traditional satellite or ground photo assessments?"*
>
> *"Which additional data is necessary in addition to the FotoQuest Austria data to measure and quantify the link between (impervious) land cover and flood extent? Is there open-data available?"*
>
> *"How difficult is it to integrate and obtain additional factors such as elevation and the use of (portable or permanent) flood protection into the analyses?"*
>
> *"Which are the best ways to ensure the quality and the accuracy of the contributed information?"*

"Can geographic citizen science comply with the data quality of traditional land cover classification methods?"

"Which method is the best way to investigate the potential correlations between soil sealing and flood extent: the relatively new geographic citizen science or the more traditional data processing methods?"

"What are possible fields of application for the results of the analyses? How could it help local authorities to manage urban planning strategies in densely populated areas?"

"How can policy makers and rescue agencies benefit from the analysed results before, during and after a flood event referring to disaster response? What are the possible improvements referring to risk prevention or preparation?"

To check the research question and its sub-questions on their validity, the following hypotheses were formed. According to the Merriam-Webster dictionary [WEB-16], a hypothesis is a "tentative assumption made in order to draw out and test its logical or empirical consequences" and it is an elementary part of an analysis. A hypothesis is based on former experiences and is not satisfyingly proven. In this case, the hypotheses reflect the most important parts of the given research question, which hopefully will be confirmed or verified as a result of this thesis:

"There are significant correlations between the sealing of the ground alongside the Danube River and the resultant extent and damage of the flood in 2013."

"Geographic citizen science and volunteered geographic information can complement the assessment of satellite imagery and ground photos."

"The quality of crowdsourced data is sufficient and accurate enough to produce meaningful and significant results."

"The level of soil sealing derived from ground photos is correlated with the flood extent, flood depth and the duration of the event."

"The combination of remote sensing technologies and volunteered geographic information can facilitate and accelerate the process of data generation, preparation and assessment."

"Geographic citizen science and crowdsourcing in general, are helpful new methodical approaches for the acquisition and interpretation of land cover data, which will become more important in the near future."

"Although citizen science is an interesting approach, traditional methods are still more accurate."

1.3. Structure of the Thesis

The structure of this thesis can be divided into four big parts: thematic background, data acquisition and methods, data analysis and the presentation and interpretation of the results. The first chapter focusses on an introduction and discusses the problems and scientific objective, as well as the research questions, hypotheses and structure of the following chapters.

The *second chapter* comprises the theoretical part about the field of research. It will provide the necessary background of basic knowledge about disaster management, floods and flood protection, soil imperviousness and current research, urbanisation as well as information about crowdsourcing and its derivatives geographic citizen science and volunteered geographic information.

The *third chapter* deals with data preparation and the used methodology. All aspects of the data acquisition will be closely examined. Methodology covers all the technology, software and techniques which will be used throughout the analyses. The quality management and accuracy assessment ensure the correctness of the information by comparing it to established datasets. The performance of an automated supervised classification in the research area, based on satellite imagery, is the basis for further analyses and comparisons throughout the course of the thesis.

Chapter four includes the actual data analysis. There will be investigations of the soil imperviousness alongside the Danube in the research area, a comparison of the different classification approaches from the FotoQuest Austria campaign (field studies and ground photos) and the supervised classification (assessment of satellite imagery) and the use of traditionally generated data like elevation and permanent or portable flood protection from official bodies in order to determine potential correlations between the sealing of the soil and the extent of the 2013 flood event.

The *fifth chapter* presents the results of the analyses and the interpretation. The strengths and weaknesses of the methodical approach will be identified and clearly summarised. Improvement opportunities for the use of geographic citizen science and recommendations or suggestions of possible applications in the field of urban planning and disaster management will be highlighted and discussed briefly before closing with a short conclusion.

After a final summary in the *sixth chapter*, the *seventh chapter* will complete this master thesis with an extensive listing of all the data sources and published or online resources, used throughout the following pages.

2. Thematic Introduction

This master thesis consists of a wide variety of complex thematic blocks, which are necessary for understanding the full significance, entanglements and interdependences of the possible correlations between the imperviousness of the soil and the extent and damage of the 2013 flood event. But first it is advisable to list a few definitions of standard terms and phrases, who may vary in their meaning and usability, although they are sounding quite similar. The concept of disaster has already been explained in the introduction and will be further discussed in the corresponding chapter about disaster management and risk prevention.

2.1. Definitions

Hazard

According to the United Nations Office for Disaster Risk Reduction (UNISDR) [SDR-17b], the phrase *natural disaster* is somewhat unsuitable and they would prefer the use of *natural hazards* instead. They claim that disasters are consequences of natural hazards and they are measured by the effect they have on human society. This aspect will be discussed in the following chapter about disaster management. However, hazard is often defined as something that is dangerous [CAM-17c]. It is also an accumulation of “elements of the physical environment, harmful to man and caused by forces extraneous to him” [BUR-78]. It basically means, that every geophysical event is just a natural occurrence and only turns into a hazard, when it threatens human life or property [GRV-01]. Hazards can be subdivided into various types, according to the Central Board of Secondary Education [VAS-08]:

~ *Geological hazards:* dam burst, earthquake, landslide, mass movement, tsunami, volcanic eruption.

~ *Water & climatic hazards:* cloudburst, desertification, drought, flood, hailstorm, heat & cold waves, landslide, tornado and hurricane, tropical cyclone, (sea) erosion, snow avalanche, tsunami.

~ *Environmental and biological hazards:* deforestation, disease, environmental pollutions, food poisoning, human and animal epidemics, pest attacks, pest infection, weapons of mass destruction.

~ *Chemical, industrial and nuclear accidents as well as other accident related hazards.*

Nonetheless, some types are missing, which can also be seen as hazards. This comprises social elements like behavioural changes, civil unrest, famines, terrorism, war and environmental elements like soil degradation [WHI-01].

Risk

Terms like risk, hazard or disaster often get confused with each other and therefore repeatedly misused. In the normal common parlance, risk is defined as the possibility that something bad might happen, or that someone or something creates a hazard [WEB-17c]. The concept of risk in the field of disaster and hazard management however can be seen as a “function of the probability of the specified natural hazard event and vulnerability of cultural entities” [CHA-94]. Just like the definition of hazards, risks only occur when people or their property are involved. Therefore the level of risk consists of the economic value of the property, the degree of vulnerability and the type of hazard [VAS-08]. There is an easy example to clarify the distinction between hazard and risk. Two individuals are crossing the Atlantic Ocean, where one travels in a big liner and the other one in a small rowing boat. “The main hazard (deep water and large waves) is the same in both cases, but the risk (probability of drowning) is very much greater for the person in the rowing boat” [FER-03]. Moreover, an earthquake is always a hazard, but the risk is higher when it appears on land than on the ocean floor. A flood risk can consequently be specified as the “combination of the probability of a flood event and of the potential adverse consequence for human health, the environment, cultural heritage and economic activity associated with a flood event” [ALF-15].

Vulnerability

The term vulnerability has already been used throughout this thesis several times and is normally used when something is open to an attack or to damage, or when someone is capable of being emotionally or physically wounded [WEB-17d]. It can be paraphrased as the susceptibility or the potential exposure to danger. In the context of disaster and hazard management it is the concept of describing the “extent to which a community, structure services or geographic area is likely to be damaged or disrupted by the impact of particular hazards, on account of their nature, construction and proximity to hazardous terrains or a disaster prone area” [VAS-08]. Due to the nature of this concept, some groups of the population are more susceptible to danger than others, because of their age, class, disability, ethnicity, gender, wealth, or (rather important in these troubled times) the immigration status. Similar to the definitions of risk and hazard, vulnerability also implicates a correlation between natural events and society [WHI-01]. As one of the key concepts in disaster risk reduction and climate change adaption, vulnerability is often subdivided into its economical, physical, political and social components, as well as into the capacity of resisting, coping with and recover from disasters. Vulnerability can be seen as the summation of exposure, resilience and resistance. Although being so important, there is yet no clear consensus of the final definition in the scientific world. The concept of vulnerability has emerged in the past few decades [WHI-01] and is still under permanent change and adaptation.

Capacity

The term capacity is mostly referred to as the "maximum amount or number that can be contained or accommodate" [WEB-17a] and will be used in this thesis primarily in connection with the volume of water bodies and the molecular level of soil science. The more thematic definition however describes the concept of capacity as the combination of resources and strengths of a community which allows them to prevent, prepare for, cope with, mitigate or recover quickly from a catastrophic event [VAS-08] and strengthen its resilience. Capacity can be subdivided into its attitudinal, economic, physical and social elements [MNE-17]. The economic capacity includes the economic situation and development of an affected area as well as the income and wealth of individuals. Wealthier persons have the possibility of building their houses in a safer place, out of stronger materials and are therefore more likely to recover faster from a disaster. Physical capacity comprises the available communications technology, constructions and engineering skills, equipment and infrastructure of a community. The social capacity sums up the relations and interactions between the people of the affected community during and after a disaster event. Attitudinal capacity deals with a collectivistic approach, where some people may fight against certain strategies due to their cultural or religious belief.

Resilience

Resilience is not just the ability of a system to adjust to or recover from changes [WEB-17b], but also one of the most important concepts in the field of disaster management and risk reduction. The term resilience was first coined by Sir Francis Bacon in 1625 [EPB-13] in one of his books about natural history. In the late 19th century *resilience* became the meaning of surviving the "application of force by resisting it with strength […] and absorbing it with deformation" [EPB-13] in the field of industrial steel processing. In the last centuries, resilience evolved from a legal term to scientific methods, mechanics, psychology, anthropology, ecology, until it reached the field of risk reduction and climate change adaption. Nowadays the term defines the capability of a certain system (e.g. communities), which is exposed to a specific hazard, to absorb, resist and recover from the changes and consequences "in a timely and efficient manner, including through the preservation and restoration of its essential basic structure and functions" [ALE-13] through disaster and risk management. In short, it simply takes a closer look whether an affected area is able to absorb a disaster or not [JOE-14]. This means, that a community or society has to feature an optimal mixture of resistance to hazardous events and the capability to adapt to them [UCL-13] in order to be fully resilient. Resilience is a multi-faceted concept, which tries to encourage researchers to investigate the separation of dynamic adaption and static resistance [ALE-13].

2.2. Disaster Management and Risk Prevention

One of the main reasons why the author chose a topic, which is closely related to disaster management and risk prevention, is the scientific research done by Professor David Alexander. He is the head of the Institute for Risk and Disaster Reduction at the University College London and gave a lecture about the key principles of emergency and disaster management at the Department of Geography and Regional Research in October 2015. For him, the most common source of human misery are poverty, disease epidemics, conflicts, displacements and disasters. He pins down the beginning of academic disaster research (or “disasterology” as he calls it) with the Halifax cargo ship explosion (USA, Fig.1) and the Silvertown ammunitions-factory explosion (UK, Fig.2), both in 1917 [EPB-17].

This implicates, that the modern era of studying disasters would celebrate its 100th anniversary in 2017. Both tragedies were serious disasters, but only after the Halifax explosion a serious systematic analysis of the disaster, its consequences and social regularities, had been carried out by the Canadian Samuel Henry Prince [SCN-88]. Today, barely someone remembers the similar British disaster.

Fig.1: Halifax after the cargo ship explosion [MAS-16]

Fig.2: Silvertown after the ammunition factory explosion [TEE-14]

From then on, the management and organisation of “resources and responses for dealing with all humanitarian aspects of emergencies” [IFR-17] slowly evolved and improved. Professor Alexander once stated, that the theory of disaster management enables people, to make sense out of chaotic situations or higher levels of complexity. This is necessary in order to fully understand the phenomenon of disaster [YTB-13]. It has already been explained in the former chapter, that disasters and hazards are no natural events and that they materialise through human existence. Geophysical occurrences like earthquakes, floods or volcanic eruptions are only trigger mechanisms. Human society’s choice to put itself at risk, is really accountable for the disaster [YTB-13]. The intervention of humans in ecological systems can turn a natural phenomenon into a disaster. Interactions with their environment can lead to an increase in the frequency of natural hazards and even to the creation of new ones [ADP-17].

If the presence and interactions of humans create the concept of natural disasters, then why are they called *natural* instead of *anthropogenic*? How can the nature be blamed for catastrophic events, which are created, triggered and amplified by the people? Is it to ease ones conscience because of poor planning, management or preparedness that took place in previous times? Whatever the initial cause for this terminological development was, nowadays natural disasters comprise environmental, meteorological and topographical events, and man-made disasters could occur due to industrial, technological and warfare incidents. Today, mainly local governments, official emergency agencies or international organisations are responsible for disaster management. Preventing and responding to disasters however is such complicate complex, which cannot be only in the remit of public services, but also have to be addressed by the broader public as well. Every citizen exposes himself or herself to the risk of getting affected by disasters, and risk reduction or prevention is therefore everyone’s responsibility. It should be everybody’s duty to contribute his or her share of collecting and assessing data in order to mitigate current impacts or even prevent future catastrophes from happening.

2.2.1. Disaster Management and Competence

A disaster is the profound disturbance of a community's functional capability "due to hazardous events interacting with conditions of exposure, vulnerability and capacity" which can cause effects on human life or property, and damages the environment or economy [SDR-17b]. Bearing in mind the definitions of hazard and vulnerability, the assessment of what qualifies as a disaster is culturally influenced as well. Gravley [GRV-01] therefore states, that a disaster is the start of a dangerous event which leads to a serious "damage or loss as perceived by the inflicted people". Scale, frequency and velocity of a disaster are dependent on social perception as a result of that. Disastrous events can test the capacity and resilience of an affected community and point out their weaknesses and potential for improvement.

The following Figure (Fig.3) shows some of the most common phrases used in disaster management and this master thesis as a word-cloud. In order to plan, organise and apply specific measures to improve the response and recovering from disasters within a community, disaster management has to be implemented.

According to the Federal Ministry of the Interior in Austria [BMI-17], disaster management refers to the "entirety of all coordinated measures in the areas of disaster prevention, disaster preparedness, disaster response, and recovery after disasters, including the ongoing evaluation of the measures taken in these areas" and are part of the official civil protection mechanisms. Disaster management can therefore be referred as the main official body of policy and governmental decisions that comprises the pre- and post- planning and responding in case of a disaster. There is also emergency management, which may overlap with the field of disaster management (and the terms are often used in a similar way) but mostly concentrates on a smaller scale and may not lead to a profound disturbance of a community's functional capability or the deterioration of human health.

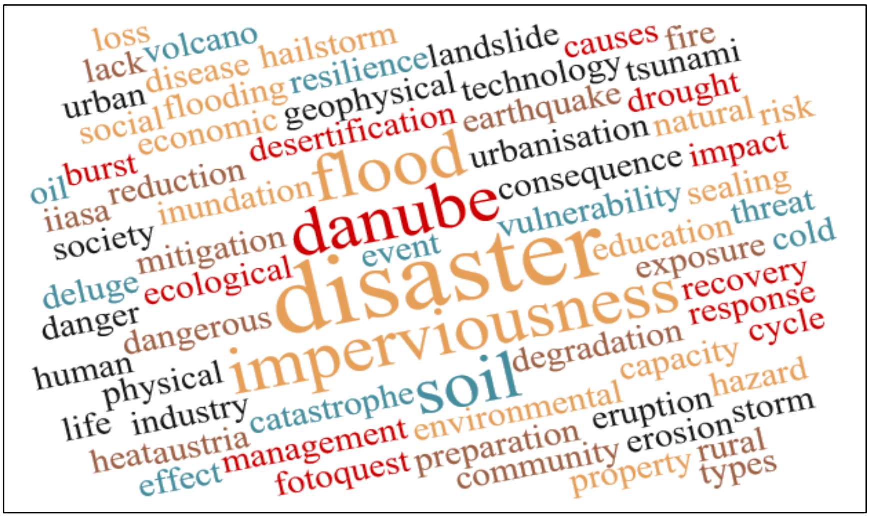


Fig.3: Word-cloud generated from terms and phrases associated with disaster

One of the earliest (semi-realistic) examples of managing disasters has to be the biblical tale of the Noachian Deluge. In this case, Noah has been warned of an upcoming flood and has time to *prepare* for the disaster (building an ark), *mitigate* the effects on biodiversity (collecting two of each species) and *recover* from the flooding (landing on Mt. Ararat) [COP-06]. This may seem like a questionable example, but it shows that the different phases of the Disaster Management Cycle are valid since the earliest days of human civilisation. In every advanced civilisation, some form of adapting to potential hazards occurred.

The modern version of disaster management however “did not emerge until the mid-20th century” [COP-06]. Mainly as a response to certain disasters, affected governments started to take over the role of responding to disasters and trying to prevent them. This may have happened due to “advances in warfare technology” and in “response to the threat posed by air raids” in the two World Wars [COP-06]. Some of today’s disaster management agencies can trace their roots back to civil defence units, which were implemented during or after World War II (e.g. Algeria, Canada, France, United Kingdom and United States of America) [COP-06].

Whenever such a disastrous event happens, a fast and effective way of coordinating and executing rescue missions in the affected area are vital to safe as many lives as possible. In many cases, laws and regulations that regulate the legislation and responsibility of disaster management, can be quite complex and antiquated. The legal framework concerning the response to, recovery from, and preparation for disasters differs from country to country. In some countries (e.g. India), different ministries are responsible for different types of hazards. This kind of decentralised responsibility can lead to conflicts of competence and delays, without the backing of a supervised institutional framework.

The US-American Federal Emergency Management Agency (FEMA) for instance, was created in 1978 (as a derivative of the Federal Civil Defence Act of 1950) because of the huge losses of life, property and infrastructure due to numerous disaster events in the past. They now use a framework for integrated emergency management, which clearly regulates the responsibilities and functions of all three governmental levels (local, state and federal) [MCL-85], which has been adapted constantly.

In addition to that, there are also a number of private and professional disaster management organisations nowadays, which offer their services to local or municipal governments in order to adapt and improve current management plans and train local rescue authorities. They help to adjust their planning and strategies to the different occurrences of disaster (e.g. droughts, earthquakes, epidemics or pandemics, floods, landslides, storms, volcanic eruptions, wildfires) and to the different phases of the Disaster Management Cycle. Nevertheless, managing disasters forms just one facet of the whole circular process, another one would be the reduction of disasters and risks.

2.2.2. Disaster Management Cycle

As already established, disasters are the result from combining the concepts of hazard, vulnerability, insufficient capacity and risk reduction [VAS-08]. One of the most important statements in the early period of pre-event planning for post-event recovery can be quoted as “after a disaster is before the next disaster”. In order to take all possible measures, the traditional approach of only *responding* to disasters had to be improved and extended [JOY-09]. Disaster Management now comprises also risk assessment and reduction, emergency management planning and the implementation of early warning systems. Now the three sections of catastrophic events can be identified as before, during and after a disaster. The Disaster Management Cycle (DMC) then combines every potential measure which can be taken during all of these sections [VAS-08] and forms the four phases: *response*, *recovery*, *mitigation* and *preparation*. These initiatives can overlap, interlock or merge with each other.

The separation into four phases is the most common version of the Disaster Management Cycle today. There are however several variations of how to manage disasters in different countries, based on their different cultural, political, social or topographical demands. The current four phase management cycle is a modern approach to highlight, plan and organise the vast range of initiatives which take place surrounding a disastrous event. It aims to reduce the negative effects on human life and property as well as on infrastructural facilities and social functionality.

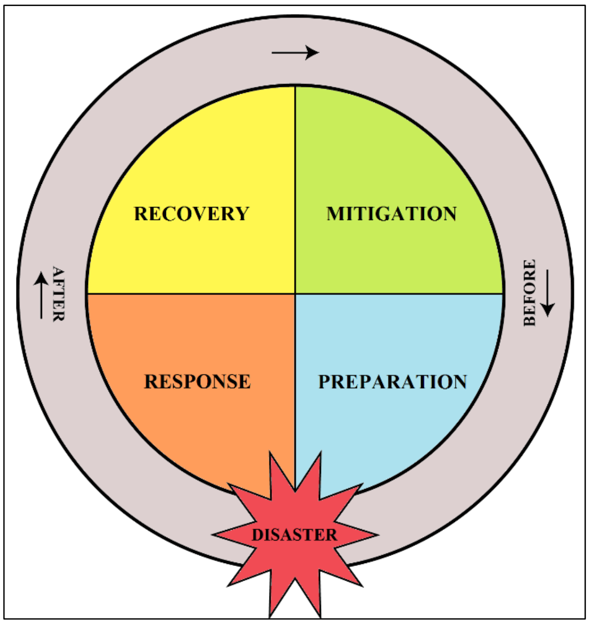


Fig.4: The Disaster Management Cycle

According to the illustration of the DMC (Fig.4), everything starts with a triggering event, in this case a *natural* disaster. Afterwards follow the four phases of disaster management:

- *Response*: is the immediate reaction to a disaster in order to minimise the effects on a community and its properties, and to offer the essential assistance. At first, early warning systems should inform the broader public about a potential forthcoming disaster. Next, the situation and needs have to be assessed by coordination managers to ensure the smooth flow of the search and rescue of affected persons and of first aid measures [FEM-16], to provide a certain disaster relief. Other short-term activities include evacuation and protection of damaged buildings, bracing dams and riverbanks, supply of temporary emergency shelters or transportation as well as the prevention of disease outbreaks [JOY-09]. One of the main tasks of immediate disaster response is the provision of basic medical or health resources and treatment, as well as the assurance of food supply and clean water.

- *Recovery*: The next phase, after the disaster and its effects were brought under control, is the recovery and restoration of the affected community, social functionality and security, as well as damaged buildings and infrastructure [BEN-16]. The main focus lies on the recovery of health and social services or economic, environmental and infrastructural systems [FEM-16] in order to allow people to return back in their normal lives. Temporary housing, financial support, psychological and medical treatment have to be provided by the governing institutions or other organisations quite rapidly. In the course of developing and reconstructing new housings (for people to repatriate) and infrastructure, first steps for a damage assessment and cause-analysis have to be taken in order to facilitate disaster risk reduction and prevention for the next time.

- *Mitigation*: The third phase of the Disaster Management Cycle comprises all efforts which try to permanently minimise and reduce the negative effects of the next disaster. It includes the elimination of “the likelihood or the consequences of a hazard” [COP-06] and the limitation of certain impacts on human lives and economic development. Disaster mitigation consists of the identification and reduction of hazards, the analysis and reduction of risks and vulnerability as well as prevention activities. Other mitigation aspects are the adaption and improvement of spatial planning strategies or building regulations and zoning.

- *Preparation and Prevention*: The fourth phase of the DMC is often referred to as the phase of preparedness and readiness. The idea is to set up programmes and plans to coordinate multisectoral activities and measures, which try to “strengthen the technical and managerial capacity of governments, organizations, and communities” [GDR-17]. It also includes “equipping people” with skills and tools and enable them to help affected people in need during and after the next disaster [COP-06]. Preparedness comprises the development and adaption of certain standards and legal regulations, the building of managing capabilities, the upgrade of communication systems and educational programmes and the coordinated interaction with the media [SGA-14].

It is obvious, that these four phases of disaster management are closely related and interconnected. They cannot be separated and are ideally integrated into planning processes and strategies [JOY-09].

Some studies try to refine the disaster management process by implementing remote sensing techniques to each phase of the cycle. Researchers from New Zealand considered every aspect of disaster management and incorporated remote sensing into all activities, which reduce risks and increase the preparedness in order to improve disaster management and citizen participation [JOY-09]. This approach facilitated the provision of spatial, spectral and temporal information significantly.

Sometimes, however, the concept of relying on one's own version of the Disaster Management Cycle is subject of discussion, especially after the phases of response and recovery have failed. In certain cases, disaster management can fail and therefore lead to more catastrophic consequences than necessary. The degree of preparedness can decide whether an event becomes a disaster or not. One example of failed preparedness and poor disaster response was one of the most extreme storm events in modern history. Hurricane Katrina represented one of the deadliest and most expensive natural disasters in the USA [NHC-11]. During and after the hurricane, poor federal flood protection, incapable coordination and management, as well as slow and delayed evacuation processes led to heavy criticism of the Federal Emergency Management Agency (FEMA). Some emergency agencies were deliberately slowed down, only to wait for the petty bureaucratic process to begin. Private initiatives and voluntary help (e.g. evacuation helpers, water suppliers and communication services) were turned down, because they were not officially authorised by the head of FEMA.

In the humoristic "Daily Show", Jon Stewart commented these controversial actions by asking "What should FEMA have done? Perhaps the answer can be found on their website. This chart, clearly depicting the agencies responsibilities in the event of a disaster. It begins with a response to a disaster, leads to recovery, mitigation, risk reduction, prevention, preparedness… and ends up back in disaster! In truth, FEMA did exactly what they said they were going to do" [CCE-15]. This may seem like a sarcastic view on a federal agency's failure, but it also highlights the problem of the communication between governmental policy or decision makers and the broader public. The illustration (Fig.5) may wrongly implicate, that every form of response by FEMA will lead to the next disaster, but the events surrounding Hurricane Katrina could give the impression that it may be true. This whole episode of disaster management failure indicates that the top-down approach will never work without public participation. Potential forms of public participation could comprise of geographic citizen science, volunteered geographic information or voluntary help and assistance.

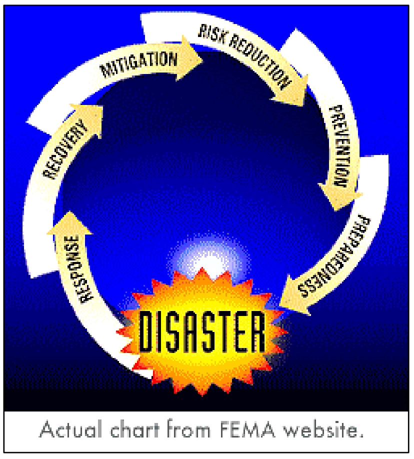


Fig.5: FEMA Disaster Management Cycle 2005 [PBZ-05]

2.2.3. Current Situation of Disaster Management

Disaster Management is an essential area of responsibility for every sovereign nation. Sometimes the extent and severity of a catastrophic event goes beyond the response capability and management capacity of a government. In these cases, international aid is required. Until the early 1990s, disaster response was mostly spontaneous, individual and often chaotic [COP-06], which led to a general demand for an institutionalised process. This international version of disaster management is nowadays coordinated by the United Nations Office for Disaster Risk Reduction (UNISDR) [SDR-17a], which organises discussion forums and legal frameworks, but also campaigns for resilient cities, sustainable development, educational programmes and public awareness. The newest framework for international disaster management declared three strategic objectives to reduce human, environmental and economic loss through disasters [SDR-15]:

- effective integration of disaster risk reduction into spatial planning and official policies, enhancing the importance of the four phases of the Disaster Management Cycle [COP-06]
- elaboration and improvement of management capacities and mechanisms at all governmental levels, focussing on the establishment of resilience within an affected community [COP-06]
- systematic integration “of risk reduction approaches into design and implementation of emergency preparedness, response, and recovery programs” [COP-06]

However, throughout the past few decades there were a number of occasions, where international disaster management functioned badly or failed totally. The events referring to Hurricane Katrina were quite an illustrative example for poor leadership quality on a national level during and after a disaster. The next sections however, concentrate on bad or missing disaster management on an international level, where the global community of states failed to respond in an effective and quick way.

One famous examples for inadequate disaster management took place in Guatemala in 1976. An earthquake with a magnitude of 7.5 on the Richter scale caused approximately 23,000 casualties and an economic damage of nearly 18% of the nation’s gross domestic product [WBG-13]. The first tremor occurred in the early morning of February 4th and many of the fatalities died in their collapsing houses, which barely met the building regulations or planning strategies. Narrow streets and several aftershocks increased the number of deaths. Another contributing factor was the long wait for the local government to call for international help. This led to a chaotic and uncoordinated private disaster response, which provoked further deaths and injuries. International aid then started slow and insufficiently. In the aftermath of the Guatemalan earthquake, first requests were made for an institutionalised framework of national and international disaster management, which led to the foundation of UNISDR in 1999.

Another more recent example for poor international disaster management were the Mozambique floods in 2000. Heavy rainfalls lasted for several weeks and caused serious inundations of urban and rural areas. These events resulted in more than 700 casualties. The total economic loss of the south-east African country was estimated to exceed 2 billion (thousand million) US-Dollar, which corresponded to approximately 40% of the nation's GDP at that time [WBG-16]. The first international aid (from the neighbouring country South Africa) arrived nearly two weeks after the first cities were deluged [COP-06]. At that time, every donation appeal of the United Nations Organisation was ignored by the international community of nations. It took more than three weeks for the first international aid workers to arrive and provide assistance [COP-06]. One form of support for the government of Mozambique at that time, was the postponement or cancellation of debt payments by wealthier nations [REL-00].

Sometimes, poor disaster management is caused by a lack of preparedness and information, or by the lack of interest by the international community. Some of the reasons could be geographical isolation or missing media coverage. Wealthier nations sometimes also suffer from a certain kind of donor fatigue [COP-06], or the tiredness of constantly donating money to disaster zones. The Mozambique flood in the year 2000 is one of the most famous examples of a situation "in which the international community has been accused of sitting idly by as hundreds of people died" [COP-06].

The lack of preparedness due to missing information is one of the cornerstones of this thesis and one aim is to point out possible fields of application for crowdsourced data and public involvement, regarding pre-disaster urban planning strategies and post-event disaster management. In order to mitigate the effects of a prevailing and upcoming disaster more effectively, the community's hazards or vulnerability have to be assessed and re-evaluated shortly after an event [WIL-07]. Based on the research findings, there is a potential for adapting local regulations of buildings, land use, reconstructions or zoning. The collaboration of governmental levels, nongovernmental and private organisations, academic institutions, media representatives and the broader public is thereby an essential factor.

Some research institutes already collaborate with governmental bodies on a wider scale. The Cascading Disaster Research Group at the University College London [UCL-17] for instance, started a project with the official office of resilience in London, for the purpose of integrating potential progressive disasters in the city's response and preparedness strategies and contingency planning. The main goals are the assurance of urban electricity supply and the implementation of the theory of cascading disaster.

Other ways to contribute crucial information in the course of a disaster are private and non-governmental initiatives like Ushahidi [USH-17]. This platform was created during a political dispute in Kenya in 2007 to visualise geospatial information. Soon afterwards it became a valuable tool for crowdsourcing information during events like the Haiti earthquake in 2010 or the floods in Australia in 2011. Ushahidi and OpenStreetMap (OSM) are the main pioneers of collaborative mapping and will be referred to and relied on throughout this master thesis.

2.2.4. Risk Prevention and Reduction

One of the international strategies, which were implemented before UNISDR was founded, demanded from its member states that the prevention of disasters and risks have to be considered as one of disaster management's major principles [COP-06]. The interconnected prevention and reduction of risks should be an integral part of official spatial planning strategies and policies, no matter which governmental level they belong to.

The prevention of disasters and risks comprise all efforts and measures to eliminate possible risks, in order to hinder future disaster damages and losses. Preventing risks is realised by reducing the exposure and vulnerability of an affected area to the minimum. This could comprise the construction of dams or retaining walls, the regulation of planning and zoning, as well as earthquake-resistant buildings [SDR-17c]. The latter could be implemented by specific and sufficient training of civil engineers, adaption of building materials to local needs and appropriate scheduling [HUY-14].

As some risks cannot be prevented entirely, the next best aim is to reduce it to the smallest possible extent. As already stated, risk may be described as the function of hazards and vulnerability. In this case, reducing disaster risks is defined as a decreased effect of future catastrophic events on communities and the prevention of new risks [SDR-17c]. Because of the importance of this concept, distinguishable approaches have emerged, like disaster risk reduction and disaster risk management. Several policies and frameworks were developed globally to meet the addressed requirements (e.g. International Strategy for Disaster Reduction). The United Nations Development Programme even states, that there are strong linkages between anthropogenic development and disaster risk: "Any development activity has the potential to either increase or reduce disaster risk" [UND-04]. In future scenarios, the concepts of disaster risk reduction, climate change adaption and human mobility will converge and become one vast and essential research field [ALE-16]. Organisations like UNISDR are therefore constantly working on the adaption of their international disaster and risk management frameworks, so that they comply with the prevailing situations. One example is the Sendai Framework [SDR-17d] for 2015-2030. Its reason for existence is based on the need for understanding "disaster risk in all its dimensions of exposure, vulnerability and hazard characteristics" [SDR-15b]. The four main priorities of the framework are: correct conception of disaster risk, support of risk governance in order to manage disaster risk, reduction of risks for building resilience and improve disaster preparedness and response.

The prevention and reduction of risks have to be applied to specific sub-types like flood risk and flood prevention, which will also be of major importance for the final analyses in this master thesis. The next chapter therefore focusses on the explanation of floods, their causes, historic events, vulnerable areas, effects and consequences, as well as potential prevention and portable or permanent protection.

2.3. Floods and Flood Protection

Since there is water on the Earth, there have been flood events. The Earth is called the “blue planet”, because about 71% of its surface is covered with water in all its states of aggregation. This comprises rivers, lakes, glaciers, icecaps, saltwater oceans, water vapour, soil moisture as well as the flora and fauna [USG-16]. Water is a unique substance, because it appears in nature in all three states (liquid, solid, gaseous). All lifeforms that ever existed on this planet, from the first single-celled one up to the more sophisticated multicellular lifeforms, needed water to evolve and survive. But where did this special and vital compound come from?

The origin of water on the Earth is still a matter of investigation. The most acknowledged theory says, that after the formation of planets in this solar system about 4.5 billion (thousand million) years ago, the main fraction of the water volume was added by giant impactors [FRA-01]. A small fraction of water is believed to be added already during the formation of the planet by water-rich bodies. Some other studies disagree with this theory and state that the water already arrived during the formation of the planet [PHY-14]. Based on the waters chemical fingerprint, the deuterium-to-hydrogen ratio, scientists suggest “a common source of water” for the Earth and the Moon, which both finally formed after the (still hypothetical) giant impact of the protoplanet Theia and Proto-Earth around 4.3 billion years ago [NAT-13]. In summary, the origin of the Earth’s water is either due to collisions with asteroids and comets long after the planets formation or due to the actual planet-forming-process [PHY-14].

No matter which theory turns out to be the most plausible one, the amount of water provided is quite extraordinary. It is now considered that there are roughly about 1.4 billion (thousand million) cubic kilometers of water on the Earth. If all this water would form a sphere, its diameter would be around 1400 kilometers. The following figure (Fig.6) shows all the water on the Earth as perfect spheres compared to the planet and the continents [USG-16]. The largest sphere with its diameter of 1400 kilometers represents all the water available. This roughly equates the linear distance between Vienna (Austria) and Valletta (Malta). The second sphere with about ten million cubic kilometers represents all the fresh liquid water which is available in the rivers, lakes, swamps or in the ground. Its diameter is about 273 kilometers and would reach from Vienna to Bad Gastein in air-line distance. The third sphere with only 93,000 cubic kilometers represents all the utilisable fresh water from rivers and lakes. The diameter of this sphere is about 56 kilometers and approximates the linear distance between Vienna and Bratislava (Slovakia).

The largest share in the global water distribution is taken by the saline water of oceans, seas and bays with about 96.54%. On second place follows water in solid state (glaciers, snow, icecaps) with 1.74%. Groundwater is on third place with 1.69%. All other water sources share the remaining 0.03% [USG-16].

As mentioned at the beginning of this chapter, flood events do happen since the existence of water on this planet. Floods are therefore one of the oldest natural phenomena in the world, but what actually qualifies as a flood? According to the Cambridge Dictionary [CAM-16a] a flood is a "large amount of water covering an area that is usually dry". The Council of the European Union explains floods in their directives [EUC-07] as the "temporary covering by water of land not normally covered by water" which includes "floods from rivers, mountain torrents, Mediterranean ephemeral water courses, and floods from the sea in coastal areas, and may exclude floods from sewerage systems". So if a usually clear defined water body escapes its traditional confines, it can be called a flood event.

The word "flood" derives from the old English word "flōd", has common roots with the German "Flut" and refers to an inundation or an overflowing body of water. Similar sources could be the Sanskrit "plavate" or the Greek "plynein", which both can be translated as navigate, wash or swim [ETY-16b].

There are various causes of floods, different types and temporal extents, as well as numerous effects and consequences of such an event. The following paragraphs will provide a short summary of the most important elements and properties.

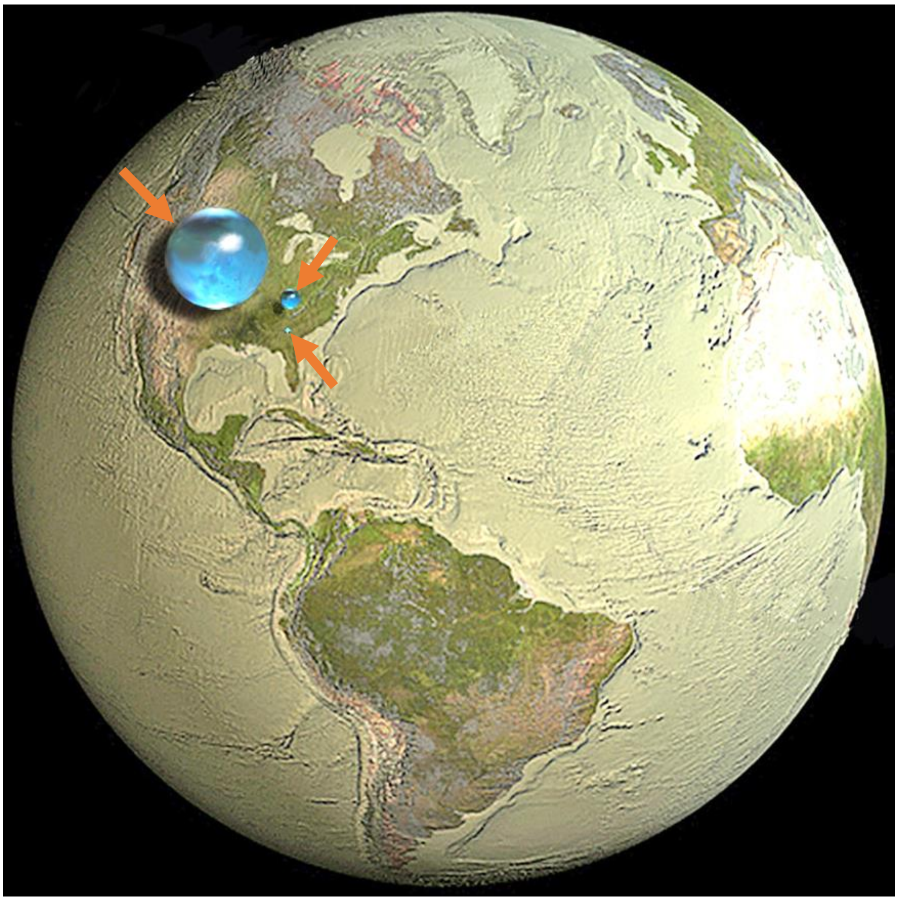


Fig.6: All the water on the Earth displayed as giant spheres [USG-93]

2.3.1. Causes of Flooding

No matter what it is called, every flood begins with a corresponding hazard. Nearly every place on Earth is exposed to this kind of danger. Even the most remote deserts can be threatened by a deluge of biblical dimensions. There are a number of possible causes and certain factors, which can contribute to such a flood event:

~ The most common reason is *heavy rainfall or precipitation* in general. If the existing watercourses do not have the capacity to convey excess water or the predominant soil type cannot absorb the available water quantity, normally dry land areas will be inundated [AGA-17a]. The duration and intensity can thereby vary intensely.

~ Another possible cause for flooding is a *strong wind* (e.g. storm, hurricane, cyclone etc.) in coastal areas, which can lead to a storm surge. This kind of rising water is a change in the water level caused by heavy winds. A storm tide is the combination of a surge and the astronomical tide [NWS-16]. This flood type can occur wherever strong winds are blowing onshore. Just a few months ago, there has been the strongest storm surge in the Baltic Sea since 2006, which affected numerous cities on the German and Danish coast [FAZ-17].

~ *River overflows* occur when the capacity of the river bed is lower than the available water volume. When there is more water than usual upstream, the floodplains will be inundated [ESC-10]. A floodplain is an "area of flat land near a river that is often flooded when the river becomes too full" [CAM-16b]. The reasons therefor can be heavy rainfalls, dam failure or unusual high amount of ice and snow melting.

~ Just like mentioned before, another possible reason could be a *dam failure*. A dam is a block or wall across a river bed, to stop the flow and accumulate a certain amount of water for creating a reservoir or generating hydroelectric power. Possible causes for failure could be bad construction standards, excessive water volume, erosion, human error, sabotage, geological movement or other natural disasters. One of the most devastating events was the failure of the Chinese Banqiao Reservoir dam in 1975, which led to a human loss of more than 170,000 people [PEO-05].

~ In most regions, the air temperature rises in spring and the *snow layers* start to melt. The surface runoff fills up the groundwater reservoirs and provides vital drinking water. Sometimes the temperature stays low and the snow does not melt in early spring. If it gets warmer and warmer in late spring, the ice and snow start to melt at the same time the rainfalls get heavier. This could cause snowmelt floods (in addition with reinforcing factors) just like the Alberta Flood in 2013 [AWP-17].

~ A quite drastic cause for a flood event is a *tsunami*. The name derives from the Japanese word for "harbour wave" and is defined as "an extremely large wave caused by a violent movement of the earth under the sea" [CAM-16c]. This movement can be triggered through earthquakes, volcanic activities, mass movement or, on rare occasions, a meteor strike. After the triggering incident happened, the waves do not increase their height dramatically in the deeper ocean. They start building up when arriving in the shallower water of the coastal area [NOS-16]. The most commonly known tsunamis happened 2004 in the Indian Ocean, with a total loss of approximately 250,000 people, and the 2011 Tohoku catastrophe near Japan with a total loss just shy of 15,000 people and the tsunami causing a major nuclear accident at the Fukushima power plant [GOT-11].

~ The last possible cause of a flood on this list is a *high tide* coinciding with higher than normal river levels [AGA-17a]. The tide is the regular rising and falling in the sea-level due to lunar and solar gravitational forces. If this well-known phenomenon coincides with an increased river gauge, there is an enhanced possibility for a flooding in the estuary area.

In addition to the mentioned possible causes of flood events, there are some factors, which can also contribute largely to the general flood hazard [AGA-17a]. These factors comprise amongst others [AGA-17a; BBC-16]: the capacity of water bodies, collateral weather conditions, concrete drainage basins in urban areas, duration and intensity of rainfalls, lack of vegetation, spatial distribution, steep-sided channels, surrounding topography and of course land cover.

Quite obvious, the contributing factor this thesis is most interested in, is the land or ground cover as well as the imperviousness and the sealing of the soil, but more information about this later. The following picture (Fig.7) from a small European town illustrates the consequences of such recent flood events quite drastically. It was caused by heavy and long-lasting rainfalls all over the continent and nearly all forms of soil or land cover were no longer visible. Besides the economic loss, the social and human impacts were immense and unmeasurable.

Fig.7: Example for high water in Europe [NEW-16]

2.3.2. Flood Types

There are several possibilities for classifying flood types, which are based on the origins of the flood, the type of flooded area or on the rate of spread. This subsection discusses just a few of the most important ones, which are commonly known. The following classification relies primarily on the publications of the National Severe Storm Laboratory (NSSL), which is a department of the US-American National Oceanic and Atmospheric Administration (NOAA) [NSS-16].

~ *Flash flood*: the most indicative characteristics of a flash flood is the short period of time between the triggering event and the inundation (normally less than six hours). Possible causes for such rapid flooding are heavy rainfalls on not absorbable ground, snow melting, dam failure or tropical storms. According to the NSSL, flash floods “[…] are usually characterized by raging torrents after heavy rains that rip through river beds, urban streets, or mountain canyons sweeping everything before them” [NSS-16]. Flash floods are a serious hazard, because they normally occur surprisingly and with little warning. “As a result of these events the drainage system has insufficient capacity or time to cope with the downpour” [AGA-17b] (see Fig.8).

Fig.8: Flash flood in Kenya [GEO-17]

Fig.9: Sight of a river flood in California in 2011 [IAM-13]

~ *River flood*: this type of flooding takes place near rivers, streams or other flowing watercourses. The inundated regions are mostly low-lying flat areas and often spread over many kilometers (Fig.9). Possible causes are rainfall from tropical storms, stationary thunderstorms or a combination of rainfall and snow melt. River floods develop at a slower rate than flash floods and could last for many days or even weeks. Because of this slow on-set flood type, relevant warnings can be issued well in advance.

~ *Coastal and estuarine flood*: similar to river floods, coastal or estuarine floods usually affect adjacent low-lying areas. The inundation is often caused by "higher than average high tide and worsened by heavy rainfall and onshore winds" [NSS-16] or by storm surges striking the coastal area. An estuary is the waterbody of a river mouth and its boundary to the sea or river is based on the salinity. The most serious types of coastal floods are tsunamis or tropical storms.

Figure 10 represents the aftermaths of such an event in Deep Bay, Hong Kong.

Fig.10: Coastal flooding of Deep Bay in Hong Kong [GET-17]

Fig.11: Devastation after a storm surge [CRH-10]

~ *Storm surge*: just like mentioned before, a storm surge is one of the most "devastating hydrodynamic features of tropical oceans that strike low-lying areas […]" [PAU-17]. They are caused by heavy winds or tropical storms, which increase the usual water level in a coastal area and are therefore closely related to the coastal flood. Storm surges can be compounded of low atmospheric pressure, storm winds, co-occurring waves, heavy rainfalls and partly of the rotation of the Earth [NSS-16].

Figure 11 shows the Bolivar Peninsula after a serious storm surge.

~ *Inland flooding*: According to the National Severe Storms Laboratory [NSS-16], an inland flooding "occurs when moderate precipitation accumulates over several days, intense precipitation falls over a short period, or a river overflows because of an ice or debris jam or dam or levee failure". In contrast to the other types of flood events, the inland floods do not occur exclusively near coastal or fluvial areas. It can result when the surficial water volume cannot be conveyed through the capacity of natural und anthropogenic drainage systems [CRT-16].

The following picture (Fig.12) was taken in the USA after Hurricane Floyd in 1999.

Fig.12: Inland flooding after the Hurricane Floyd [CRT-16]

Fig.13: Urban flooding of the Ijssel River [UFL-12]

~ *Urban floods*: this flood type is a "naturally-occurring hazard", which is characterized by the fact, that it only affects cities or other populated areas. The urbanised nature of this battered regions is the thematic centre of this type. The main cause of an urban flooding is the undersized capacity of the local water drainage systems. As a result of the sealing of the soil, hardly any kind of precipitation can trickle into the ground. The more land becomes urbanised, the higher is the resultant flood risk [WEA-16]. An exemplary urban flood event is depicted in Figure 13, showing an inundation of the Ijssel River in the Netherlands.

Thanks to the increased human mobility and the ongoing migration into the cities worldwide, the number of people who are in danger of suffering flooding is increasing rapidly. A possible solution to prevent the growing urban flood risk is an implementation of integrated water management. This kind of "inter-disciplinary and intersectoral integration of the components of urban water, is a necessary approach for achieving results in line with sustainable urban development" [TUC-07].

2.3.3. Vulnerable Areas

Just like any other natural hazard, floods often result in human fatalities, damaged property and economic or social loss [YAL-04]. All the different flood types accounted for nearly 43% of the global weather-related events between 1995 and 2015 (Fig.14) and affected approximately 2.3 billion people [REV-16a]. In financial terms, flood-prone areas show a 1% chance of being inundated every year [FDF-17]. As indicated in previous chapters, some areas are more flood-prone than others and have an increased flood risk. Therefore they feature a higher physical and social vulnerability. This concept is regarded as the function of exposure, resilience and resistance. It is formed through economic, environmental, political and social processes, which determine the degree of disruption by natural hazards [SAD-17]. The measurement and assessment of vulnerability is challenging and not always universally usable. There are a number of models and indices that try to evaluate flood-prone areas (e.g. FVI [IHE-17]), but they often depend heavily on the affected society's cultural development and the prevailing topography.

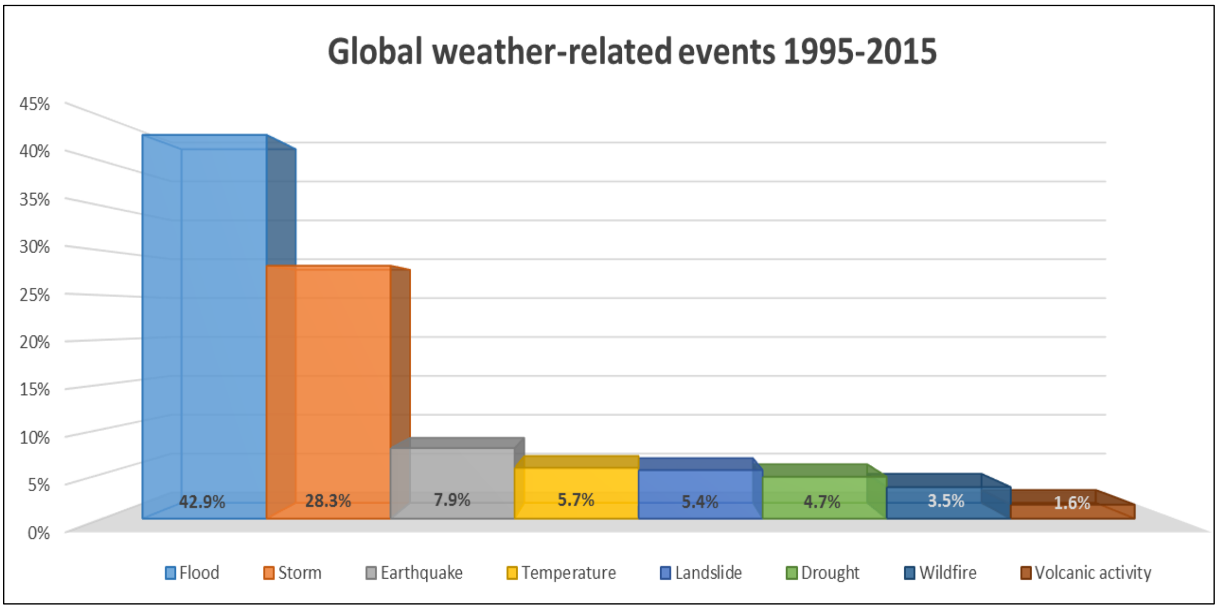


Fig.14: Different types of global natural disasters by type from 1995 to 2015 (author's illustration based on [REV-16a])

In simple terms, low-lying areas near rivers or other waterbodies are more susceptible to floods than remote regions on a higher altitude. Only, it is way more complex than that. Many factors contribute to the extent, frequency and damage of flood events, especially in urban and densely populated areas. However, it is necessary to be able to detect, measure and assess flood risks and vulnerability because of their potential in improving decision making and strengthening a community's resilience. Some established methods comprise the traditional deterministic modelling approach (physically based, linked to damage assessment) and the more recent parametric approach (uses existing information for assessing vulnerability) [BAL-13]. Even though these complex methods offer a quick and comparable insight into an area's flood vulnerability, the limited level of detail and the high formal complexity prevent the widespread usability in the field of policy making and urban engineering [BAL-13]. Other ways of supporting spatial planning and disaster management, in order to decrease vulnerability and increase resilience, have to be developed and applied.

The increasing frequency of flood events and the resulting damage of social functionality and urban infrastructure indicate, that the implementation of flood risk management into local management strategies has not progressed fast enough in the past decades [BUC-16]. Having said this, adapting or changing the ancient human settlement habit is not a fast or easy process. For thousands of years, people always erected housings near rivers, lakes or at the coast. The close proximity to these waterbodies guaranteed fresh water supply and potential trading routes. For this reason, nearly every modern major city is located near some form of waterbody. Insufficient spatial planning and an increasing rate of migration into these cities (thanks to urbanisation and the associated sealing of the soil) raise the general vulnerability and lower the capacity to cope with floods.

The exposure to rising sea-levels and frequent inundations due to climate change threaten millions of people worldwide (Tab.1). The densely populated Netherlands are most vulnerable due to extensive land reclamation in the past. The construction of flood protection measures is just a temporary solution and will become more and more complicated and expensive in the future [EUC-17]. Referring to the absolute values of an exposed population, China is far ahead with more than 50 million people affected [CLC-14]. This major Asian country does not appear in the top five of the following table (Tab.1), because the percentage of primarily vulnerable population in relation to the total population is only about 4%. In this context, the most vulnerable Chinese cities are Guangzhou, Shenzhen, Tianjin, Zhanjiang and Xiamen [LSC-13]. Severe flood events could therefore have profound effects and consequences on the millions of affected citizens, their property or on the economic performance of an entire nation.

Rank	Country	Exposed Population (Rounded Values)	Percentage of National Population Exposed
1.	Netherlands	7,800,000	47 %
2.	Vietnam	23,400,000	26 %
3.	Thailand	8,200,000	12 %
4.	Japan	12,800,000	10 %
5.	Myanmar (Burma)	4,700,000	9 %

Tab.1: Five of the most flood-prone countries, based on percentage of exposure of national population [CLC-14]

2.3.4. Potential Effects and Consequences

The possible impacts of flood disasters on nature and humans cover a wide range of short- and long-term effects. They are depending heavily on external influences like the topography of the affected region, the flood extent and duration, as well as the community's vulnerability. The flood hazard increases with higher degrees of population density. Urban areas with their permanently sealed surfaces and higher rates of surface water runoff are therefore exposed to a higher flood risk. The following paragraphs summarise some of the most negative effects on urban social, economic and environmental systems. There are however also some positive effects on the affected community after a flood event.

Social systems comprise of personal integrity and social functionality, which are the foundations of human civilisation and culture. Depending on the severity of the event, public health care and medical aid are struggling to function properly. This shortage can intensify the transmission of infections and communicable diseases like cholera, dengue, hepatitis, malaria, typhoid and all kinds of other fevers [WHO-17]. The devastation of infrastructure and farmland can disrupt the supply of clean water and food, which can additionally lead to dehydration and famine. There are also various psychological consequences regarding the consternation or shock as well as sudden displacement and destroyed livelihoods. Floods can therefore issue serious problems with mental health, mainly represented in developing countries [CRA-12].

In a capitalistic system, the economic aspect is always in the major focus of post-event assessment. The most obvious consequences of a flood event is the damaged infrastructure, like buildings, roads, railways or ports. These disruptions cause further restraints such as succumbed motorised private transport, the interruption of energy and resource supply, the absence of tourists and the potential contraction in demand. Additional costs emerge for disaster response and the following reconstructions. NOAA reports, that in 2005 (the year of Hurricane Katrina) the estimated economic loss through flood damages was the highest in the USA since records began, with approximately 55 billion US-Dollars [NWS-17].

The effects on environmental systems are as versatile as on economic or social systems. This includes the potential loss of agricultural land and crops, the salinisation of soil in coastal areas, the degradation and leaching of fertile soil, the distribution and accumulation of toxic substances or decreased biodiversity. A further consequence could be damaged nuclear power plants, like in the Japanese Fukushima-Daiichi nuclear disaster in 2011 [SCE-16]. This event demonstrated the importance of proper building regulations and sufficient provision with basic information for policy makers or urban planning strategies.

The number of positive effects of flood disasters is quite low. The main argument is, that the high water deposits a lot of nutrients on the floodplains, which then work as a natural fertiliser. A common example is the annual Nile flooding, which was the basis of Egyptian culture for thousands of years.

2.3.5. Preventing Floods and Flood Protection

Unless humanity is able to control precipitation and build insurmountable river engineering structures, it is still necessary to prepare for floods and try to protect the population from the rising water. There are a couple of aspects that could be considered in flood risk management: preparedness, prevention, protection and emergency response. Some are easy to realise, and some need much more research done.

One way to prepare for disasters is the implementation of appropriate flood risk management. This includes every measure and activity that aims to reduce the possibility and consequences of an inundation [EUC-16a]. Only responding to a flooding and trying to defend oneself on a local level is not enough in today's society. The "increasing availability of remotely sensed data" [PEN-05] is one of the main drivers for the detection and prediction of upcoming flood disasters. This includes the analysis of satellite imagery and aerial photography, the Synthetic Aperture Radar (SAR), as well as the generation and interpretation of large datasets by untrained volunteers (VGI and citizen science). Thanks to these technological improvements, hydrologic models and flood scenarios can be developed and investigated in a more effective way.

There are however some analytical challenges, which lead to a challenging complexity of flood risk management [COL-15]. If flood risk management fails and every effort of preventing a disastrous event is ignored by the government, a catastrophic situation may be imminent. When the different levels of governmental decision makers argue about the areas of responsibility, the citizens are often the sufferers [BBC-17]. The solution could comprise the implementation of a practicable legal framework and the integration of the affected population in the process of data generation and decision making.

To link the early stage of preparedness to the more immediate phase of prevention, the adaption of early warning systems is an essential task in today's flood risk management. People have to be aware of the risks or the upcoming dangers. Disaster risk reduction (DRR) should be one of the first steps to a safer community. After that, follow the phases of preparation and prevention. Flood warning systems provide adequate tools for predicting and communicating future inundations. They require a comprehensive system to develop and substantial investments or resources (e.g. financing the acquisition of information) [PAP-15]. The ideal situation would be a transboundary system, where every participating nation contributes their data. An example could be the European Flood Awareness System [JRC-16].

Other methods to prevent or reduce the impact of floods, comprise stricter building regulations and forward-looking urban planning strategies. The construction of residential complexes on floodplains, unstable grounds or low-lying areas increase flood risks and raise the human and economic loss in case of a disaster. Another possible preventive measure is the sophisticated planting of trees and deep-rooted plants. A sustainable and careful approach concerning the control of land use, farming, forestry and river regulations could be the key to a successful flood prevention. It is therefore necessary for the different official bodies to communicate with the population and offer enough educational material to assure public acceptance for every planned measure.

If all activities to prepare for and prevent the flood disaster have failed or were not strong enough, protective measures will go into action. This mainly constructional actions try to control the extent and damage of floods by preventing them from reaching populated or other valuable areas. They increase the resilience and decrease the vulnerability of urban and rural regions. New technological advancements allow the local authorities to install permanent and mobile flood protection, as well as temporary barrier systems. Long-term solutions include coastal defence walls, breakwaters, retaining walls, floodwalls, dams and levees, retention basins, flood diversion channels, and river straightening or channelisation. Short-term protection systems comprise different forms of panel barriers, rigid or flexible flood barriers, filled tubes, permeable or impermeable filled containers [EFR-02], sandbags, flood gates or portable aqua-fences. All these efforts are based on the aspiration to protect human life, the economy as well as urban properties and infrastructure.

In this time of urbanisation, more and more people live near the coast or big rivers and in order to protect themselves from rising sea-levels or storm surges, new ways of barrier systems have to be developed and implemented. London and Venice offer famous examples of how these monumental tasks can be realised. The Thames Barrier in London is operational since 1984 and "one of the largest movable flood barriers in the world" [THB-17]. It is more than half a kilometre wide and consists of ten giant steel gates, which prevent storm surges from the North Sea to inundate most of London's metropolitan area (Fig.15).

The MOSE project in Venice [MOS-17], Italy, is one of the most interesting structural measures currently under construction (Fig.16). Its purpose is to protect the pile dwellings of Venice against the high tide by using a system of electromechanical tide gates, which span more than 1.6 kilometres over four inlets and safeguarding the Adriatic lagoon behind.

Fig.15: The Thames Barrier in London [NBI-16]

Fig.16: The Italian MOSE project in Venice [NVZ-15]

2.3.6. Historic Events and Floods in the Research Area

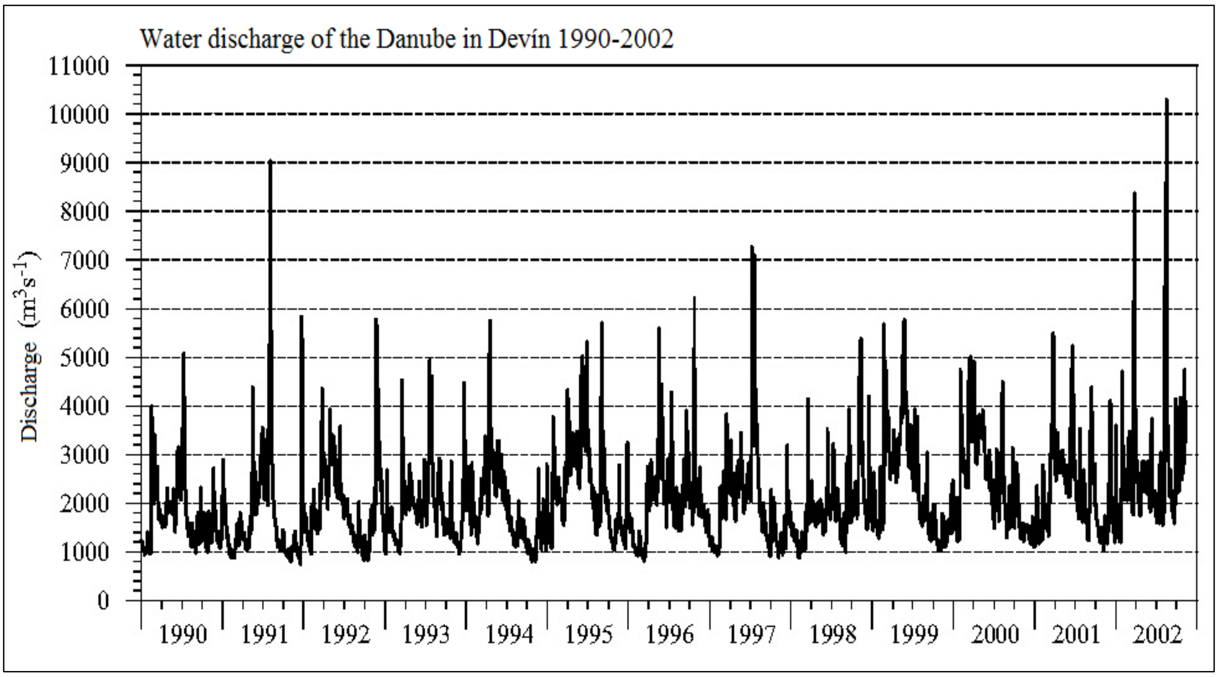
Mentioning the historic flood events will help putting the analysed results in a specific context. It shows that natural disasters and adapting to the resultant altered conditions are as old as human existence. It all starts with the earliest written records about the global flood myth. Various legends about a great flooding exist in different cultures all over the world. Traditionally there is some kind of divine punishment for the disobedient people involved. Some of the most famous examples are the Epic of Gilgamesh and the biblical story about the Noachian Deluge [CRG-15]. Thousands of years ago, when these tales were written down, people did not fully understand the scientific aspect beyond a disaster and interpreted the flood event as a will of their gods. Back then, religion often acted as the main source of explanations when the common knowledge was insufficient. Every type of natural disaster was declared as a divine discontent.

However, ancient disaster myths are occasionally based on real historic events. Theories suggest, that some of the flood myths were based on one specific event, the filling of the Black Sea basin about 7500 years ago. The so called *Black Sea deluge hypothesis* states, that “a massive inundation of the Pontic basin had a profound impact on the culture of prehistoric humans, forcing large-scale migration out of the affected area and creating a basis for the biblical Great Flood legend” [YAH-07].

Another illustrative example of a historic flood event is the eruption of the *Thera volcano* in Minoan times around 3500 years ago. The massive outburst of the Santorini (or Thera) volcano triggered a series of tsunamis, which could have caused the demise of the Minoan civilisation [BRU-08].

An impressive example of a devastating European Flood is the Great Drowning or *Grote Mandrenke* of 1362 on the coast of today’s Germany and the Netherlands. The inundation of the Dutch Wadden Sea in North Frisia may have caused up to 100,000 fatalities [GOR-15].

Only in 2013, a major flooding occurred amongst others in the Austrian Danube basin. Its peak discharge was at the capital Vienna [GON-16]. Some other affected cities experienced water levels as high as the previous record holding flood of 1501 [BLO-13]. The Danube River inundates its surrounding floodplains quite frequently, mainly due to extensive precipitation. A serious number of bigger floods took place in the past decades (1899, 1954, 2002, 2005, and 2013) which caused a lot of material damages alongside the course of the river and its tributaries [BLO-13]. The Danube is described as the most “international river basin in the world, draining water from 19 nations, forming the international boundary for 8 of these, passing through 10 countries and 4 capitals from source to mouth and covering approx. 10% of Continental Europe” [FEL-16]. This makes the Danube also a very dangerous place to live around, because if the frequency and extent of floods increase, considering that the severity of extreme weather events is also raising, the consequence for human health and property are going to be quite bad. An example for the average water discharge of the Danube River in the Slovakian province of Devín (Graph 1), 30 kilometres east of Vienna, shows the seasonal variations within the data. The peak in 2002 with a discharge of more than 10,000 m^3s^{-1}, fits perfectly with the already mentioned *bigger floods* of the Danube in the past decades. Although not perfectly visible in the illustration (Graph 1), there is a small trend towards an increase of the average discharge recognisable.



Graph 1: Exemplary water discharge of the Danube River in Devín between 1990 and 2002 [VVB-17]

The implementation of sophisticated flood protections measures [BLO-13] and the integration of transboundary flood risk management planning strategies [GON-16] over the past 20 years, already contributed a lot to the local disaster risk reduction and to the decrease of the extent of flood damage.

There is however still room for improvement in the Austrian Danube basin. Outdated flood management regulations, low cooperation between neighbouring regions, partially high vulnerability, alpine topography and a lack of precise scientific knowledge by policy makers provide still an opportunity for improvement [GON-16].

2.4. Soil-Sealing in a Time of Urbanisation

The research goal of this thesis is to find possible linkages between the imperviousness of the ground alongside the Danube River in Austria thanks to soil sealing, and the correlation to the extent and potential damage of the 2013 flood disaster. It is essential for the further understanding to investigate the subject of soil and its importance during a flood event and also in people's everyday life.

The derivation of the word soil is apparently from the Latin "solum", which means the bottom or the ground. Put in highly simplified terms, soil is per definition the top layer of the surface of the Earth, where plants can grow in [CAM-16d]. It consists of rocks, minerals, fluids, gases, organisms in combination with decayed organic matter [FDF-16]. This part of the Earth's crust is responsible for filtering and storing potable groundwater, offering habitats for countless organisms, storing raw materials and vital substances (e.g. carbon or nitrogen) as well as producing food, feed, textile fibres and even certain types of fuel [FAO-17a].

Soil is also the main foundation for the development of human civilisation and the construction of all cities. Every building in this world is based on some type of soil. Due to its slow creation, the Earth's soil can be considered "as a non-renewable resource" [EUC-16b]. According to figures from the Food and Agriculture Organization of the United Nations (FAO), it takes about 2000 years to create ten centimetres of fertile soil, which therefore contributes further to the importance of soil in general. Soil is an underrated substance and is characterised by the composition of different kinds of layers or horizons, which are clearly distinguishable from each other (Fig.17). This differentiability is due to different types of processes and transformations of the original parental material, the bedrock. According to some sources, the composition of a healthy soil sample is about 45% minerals (e.g. sand, slit, clay), 25% water, 25% air and other gases and 5% organic matter, which supports optimal plant growth [FAO-85]. In addition to that, there is a considerable amount of lifeforms in all shapes and sizes. The several horizons (Fig.17) can be consolidated into four main categories:

~ *C-horizon:* The parental matter or substratum combines solid bedrock and already weathered raw material. Due to the lack of roots and organic matter [FAO-85], it is not important for agriculture.

~ *B-horizon*: This illuviation zone contains aluminium, clay, organic matter and even iron accumulate to form a mineral subsurface layer. It offers a different colouring and the parental rock gets modified by chemical and physical processes [HYP-17].

~ The topmost layer is the *A-horizon* and is made up in large parts of organic matter. This eluvial horizon is one of the most important parts of the soil, because most of the plant root growth takes place in these 20 to 30 centimetres of mineral matter. The plants get most of their nutrient matter from the topsoil. Additionally, most of the soil's lifeforms live in this layer, where it depletes the substances, which are then accumulated in the B-horizon [HYP-17]. Due to its intensive exposure to agricultural preparation (ploughing), the differentiation to the layers can be fuzzy [FAO-85].

~ In nature-oriented areas, there is also the possibility of a surficial organic layer, the *O-horizon*. This deposition of organic matter consists largely of less decomposed plants. This few centimetres thick mulch provides a fair amount of nutritive substances for the flora and fauna in the soil.

~ There are variations, which can be adapted to different types of soil, like organic horizons (H, L, O), and mineral horizons (A, B, C, S, G, E, Y), which offer less than 30% of their weight to organic substances [HYP-17].

The explanation of all processes of soil formation fills entire books and, for this reason, only the most important parts were chosen for this chapter. Depending on the soil texture, different types of soil can be classified (Fig.18) and it is determined by different proportions of compounds like clay, sand and silt. Lighter soil which consists mostly of sand is coarse-grained, with a grain size between 0.5 and 1 mm. Loamy soils are medium textures soils, where the silt-percentage is highest and the grain size differs between 0.002 and 0.5 mm [FAO-85]. The finest soil texture consists mostly of clay particles. They are nearly impossible to distinguish and the grains are smaller than 0.002 mm.

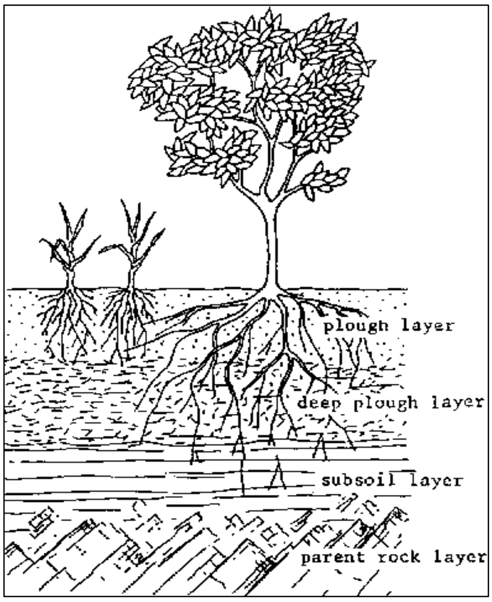


Fig.17: Profile of average soil layers [FAO-85]

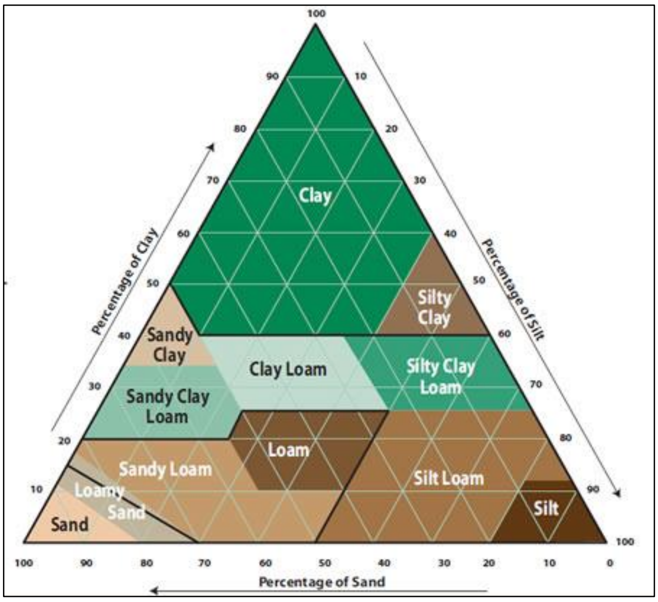


Fig.18: Soil types based on three major compounds [FAO-85]

The characteristics of soil can be classified into biologic, chemical or physical properties. In order to lead over to its connection with soil imperviousness and flood events, the main focus lies on the physical property of hydrological balance:

According to the FAO [FAO-17b], the *biological properties* can be divided into the nitrogen cycle and the carbon cycle. The nitrogen cycle comprises the circulation of nitrogen in its various forms through the soil. It includes denitrification, fixation, mineralisation and nitrification [FAO-17b]. The carbon cycle on the other hand, is a process where carbon is changing and moving through the air and the soil, which includes the atmosphere, biosphere, geosphere, hydrosphere and the pedosphere [COR-16]. This current exchange of carbon is as important as the water cycle for the development and support of life on Earth. According to British scientists [EVA-14], the understanding of these processes is “central to analysis of the geographical consequences of environmental change” considering “climate change, water security or flood risk hazard”. Deforestation for instance is a major anthropogenic interference with carbon storage possibilities in the ecosystem, to name but one reason for soil degradation.

Another possibility to characterise soil is after its *chemical properties*. These include the base saturation, cation-exchange-capacity, the pH-level of soil and plant nutrients, as well as the availability of compounds like organic carbon, calcium carbonate, nitrogen, salts, sodium and sulphate [FAO-17g].

Some of the soils *physical properties* have already been mentioned. These comprise colour, consistency, density, porosity, structure, temperature and texture. The texture and the structure of the soil affect processes like the aeration, the heat flow, the root growth of plants, the susceptibility to erosion and water movement [FAO-17c]. The physical properties of soil include furthermore the availability and characteristics of soil water, the circulation and interaction of water within the soil. Due to the different types of flood disasters, it is important to shed some light on the connections between soil, water content, water potential and its implications on flood risk.

The known soil does not only consist of mineral and organic matter. Their particles vary in shape and size but they do not fit perfectly together like jigsaw pieces. This void spaces are called pores and are mostly filled with different compositions of gases, fluids or living material like insects or plant roots. If the soil is dry, the pores are mainly filled with gases like air. If it precipitates or the ground gets irrigated, the void spaces get filled up with water. The soil moisture content therefore “indicates the amount of water present in the soil” [FAO-85]. The rate of infiltration (seeping velocity) relies heavily on the type and structure of the soil. Water trickles faster into granular sand for example, than into compact clay. The velocity of water movement is called hydraulic conductivity [FAO-17c] and it depends on the coarseness of the soil. Sand is more coarsely grained and features larger pores, clay on the other hand is more finely granulated and offers therefore smaller pores [FAO-85]. Another reason for changes in the infiltration rate is the already available soil moisture content. When the ground is saturated (all pores filled with water), the seeping velocity decreases and water is starting to pond. Plants also need air to grow effectively. When the pores are just filled with water, plants will suffer and after a long “wet” period, die back presumably. If the superficial precipitation or irrigation stops, the saturation decreases through processes like drainage, evaporation, percolation, runoff or plant transpiration. If the area drains through gravitational forces, it reaches a level which is called the field capacity [FAO-17c]. In this state, the smaller pores are filled with water and the larger pores with water and air, which is considered to be ideal for plant growth [FAO-85]. The main mechanisms which are responsible for the soil water movement are capillary action, gravitational forces and osmosis. Deeper areas in the soil, which are constantly saturated, form the groundwater table and its depth differs mostly due to changes in topography.

The knowledge of soil moisture and its variations is not only necessary for the study of soil and plant growth, it also affects "the evolution of weather and climate over continental regions" [CLM-17]. According to NASA [CLM-17] it is vital to have accurate information on soil moisture in order to improve weather prediction, agricultural crop growth and climate models. Their measurements estimate the water content in the upper sheet of topsoil, its dwell time and its effects on the weather and the climate. Other ways of soil moisture estimations are performed by remote sensing instruments like airborne and spaceborne radar and scatterometers. The measurement of soil moisture can be used to assess flood risks and to predict flash floods in a given area. The variation of soil moisture is also considered "as the most important soil factor for rapid runoff and flash flooding" [GRI-16].

A good approach to get information about the different soil types and their classifications, are FAO soil maps and databases [FAO-17d]. They comprise global and national maps, as well as regional maps about certain projects. An example of a national map from Austria is the BORIS Soil Information System [EAA-17]. Its main aim is to provide "harmonised soil data of Austria based on a specific data quality management" and offers information about soil types, geology, vegetation, land use and hydrologic balances. With this tool it is possible to investigate the research area and the soils behaviour in case of increased water availability. Unfortunately there is no possibility to compare the development of the data (such as land use and hydrological balance) over a longer time period.

From a legal point of view, the German Federal Soil Protection Act and Ordinance [BJV-17] states that the main function of soil is supporting the livelihood and habitat for animal, plants, humans and other soil organisms. Another natural function deals with its importance in the water cycle and the nutrient cycle. Soil is defined as a decomposition and neutralisation medium because of its absorbing-, converting-, filtering- and buffering-properties. Furthermore, it also offers functions, e.g. as location for agriculture, forestry, settlement, recreation, infrastructure and as a source of resources [BJV-17].

Regarding to soil loss and degradation, any impairment of these soil functions is considered as harmful soil change and shall therefore be prosecuted (at least in the ideal case). But where is the border between normal urban development and damaging soil alterations? What is considered as sealing of the ground and permanent loss of soil? To shed some light on this problems, the next chapter copes with social issues like urbanisation, land grabbing, soil degradation and soil imperviousness.

2.4.1. Urbanisation and its Implications

Soil sealing is not only connected with *urbanisation*, but is also one of the major consequences of this phenomenon. According to the Cambridge Dictionary [CAM-17b], urbanisation can be defined as the "process by which more and more people leave the countryside to live in cities". The Organisation for Economic Co-operation and Development (OECD) also specifies it as a process, by which "a large number of people becomes permanently concentrated in relatively small areas, forming cities" [OEC-03]. As a conclusion, urbanisation causes urban growth and rural depopulation.

Urbanisation is often linked to some modern concepts like industrialisation or modernisation in general, but its roots can be traced back thousands of years to the beginnings of human settlements and civilisations. Before people started to become sedentary, almost everyone had to migrate from one place to another, due to the seasonal availability of water and fertile soil. About 6500 years ago, the evolution of agricultural capabilities and techniques facilitated the establishing of permanent settlements in Mesopotamia. According to Elmqvist [ELM-13], this development can be determined as one of the earliest examples of "ecosystem management for enhanced productivity". This leads to a higher food availability and therefore a constant increase in population.

The most fertile and crowded regions have always been near rivers and coastal areas, but these coastal and river proximities are rather vulnerable regions and can be both a blessing and a curse. Emerging civilised life "was shaped by two conflicting factors" [AHE-14], the unpredictability and flood risk of nearby rivers, and the fertility due to the deposition of fertile mud after a flood event (e.g. the annual flooding of the Nile and the resulting development of the Egyptian civilisation). As a result, fertile land attracted an increasing number of people, who were willing to settle down. The higher flood risk however required an early form of environmental management.

The most widespread type of settlement of that time was the farming village. Due to sustainable farming, small communities and "flexibility in the sources of subsistence", this settlement type became the most enduring and prevalent form of ancient urbanised areas [ELM-13]. As communities became more and more complex, the social order and interaction changed. The specialisation of crafts and the construction of monumental buildings by the ruling elite (to impress and intimidate the people) are first indicators for the development of early cities [ELM-13]. These early cities are defined by their social complexity, diversity, interdependency and sophisticated infrastructure. This enabled the society to develop and increase their population size. Although cities were dependent on settlements in their close proximity to safe food availability and supply, the population shift from rural to urban areas continued to gain in strength. This subsequently led to a dissemination of social achievements, like class structure, "hierarchical territorially-based government" [ELM-13] and an organised legal system.

The evolution from small and loose gatherings of people to villages and cities benefited from the improvements in agriculture (e.g. animal breeding, irrigation, ploughing) and the development of trade and science. This encouraged people to settle down instead of wandering around from one fertile spot to the next. It is for this reason that human housings developed from hide tents, mud brick buildings and timber structures to rather permanent stone or brick buildings and later concrete constructions.

Industrialisation and modernisation accelerated the velocity of urbanisation in the 18th century dramatically. The Industrial Revolution led to an increase in agricultural productivity and the improved food supply led to an accelerated growth of the population. The following figure (Fig.19) illustrates the 400 years' time span (from 1300 to 1700), where the global population size increased by 300 million [TAN-94]. In the beginning of the Industrial Revolution, it only took about 100 years (from 1700 to 1800) to gain another 300 million people. Within the next 50 years, the population size increased by another 300 million (1.2 billion in 1850). Another 50 years later, in 1900, the global population was estimated to be around 1.6 billion. It seems obvious, that industrial and agricultural developments were the main drivers of this demographic explosion (current global population approx. 7.5 billion [WOM-17]). The industrialised production of agricultural goods enabled former farm workers to take employments in the booming industrial cities. This migration necessitated affordable accommodation and improved infrastructure, like public transport systems.

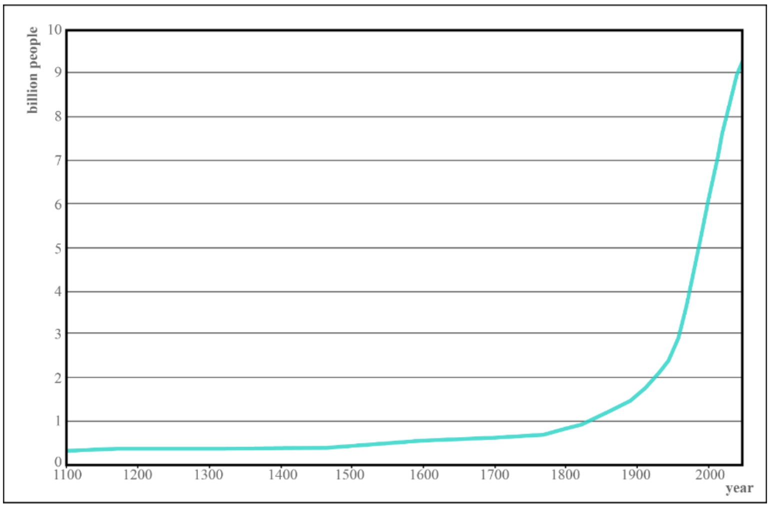


Fig.19: Increase of the world's population in the past few hundred years (based on [UND-14])

In 1800 around three percent of all people lived in towns and cities [BAR-12]. Today approximately 50% of the Earth's population live in cities or otherwise urbanised areas. The United Nations Organisation estimates, that this percentage rises up to 66% by the year 2050, which means that only one third of all people would live in rural areas [UND-14] (Fig.20). This may be due to notable urbanisation benefits like better health care, better chances in finding qualified and well-paid jobs, improved infrastructure, higher living standards and social amenities as well as more leisure facilities.

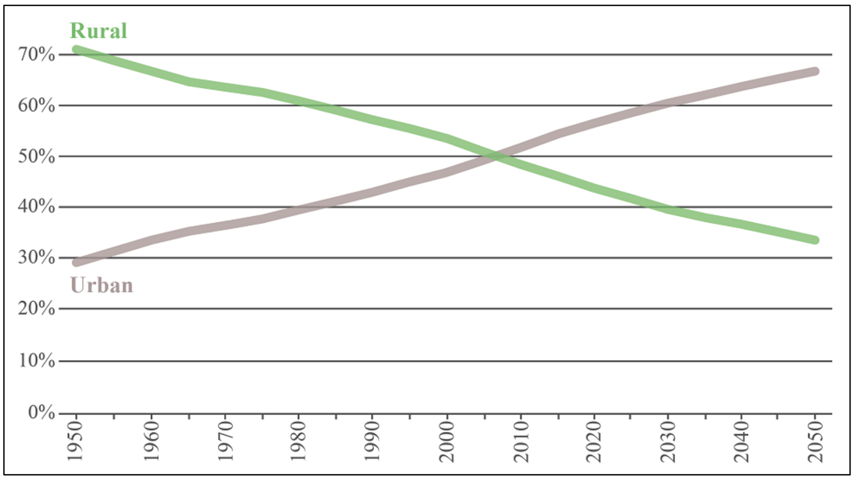


Fig.20: Global percentages of rural and urban population from 1950 to 2050 (based on [UND-14])

The most serious consequences of these developments are economic, social and environmental effects:

The impacts of *economic effects* can be positive and negative. The development of cities and infrastructure during the Industrial Revolution led to a significant reduction of transportation. For this reason, the commuting to the workplace became easier and worker could take advantage of better job opportunities all over the city. Cities are also important to the global economy. About 80% of the world's Gross Domestic Product (GDP) is generated in urban areas [MCK-16], with a predicted increase in the next 20 years. To satisfy the requirements of a functional community, cities have to maintain and update their infrastructure constantly. On the other hand, house prices were raised due to increased demand. The economic development in cities is mostly based on a limited number of branches and could therefore be vulnerable. Large urban areas also require a strong migration of workers to maintain a high level of economic growth. The "slowing population growth and plateauing urbanization" [MCK-16] in a few of the world's largest cities are responsible for decreased economic growth and bleaker future prospects.

Positive *social effects* include the exchange of knowledge as well as benefitting from cultural diversity. Urban areas offer an increased tolerance for religious attitudes and more educational possibilities. Urbanised areas are also a melting pot of culture and knowledge. The exchange of experiences and ideas lead to intellectual stimulation and an expansion of common knowledge. Negative social effects deal with a shortage and price increase of accommodation, due to overcrowding and therefore a higher demand. The prospect of work in the cities can also lead to an accumulation of low skilled workers and consequently a potential emergence of slums and class segregation [DOC-12]. Due to a higher population density, traffic volume increases measurably. Possible unemployment of low skilled workers can result in poverty, hopelessness, stress, alcoholism, drugs and crime [DOC-12].

Positive *environmental effects* of urbanisation comprise the sustainable use of natural resources within urban areas. Due to sustainability concepts and commodity prices, it is common nowadays to build products using as little resources as possible. Even new housing is replacing idly buildings. There are however far more negative environmental effects listed in corresponding scientific studies. With more population comes more consumption of durable goods and increased non-recyclable waste. This allows diseases to spread within urban communities, which can also be fuelled by the anthropogenic pollution of air and water. The contamination of the atmosphere and waterbodies can affect “the health and quality of life of the urban population” quite heavily [TOR-04]. Other problems include the formation of urban heat islands, high consumption of energy and non-renewable resources as well as the increase of damages due to bad urban planning. The urbanisation of former open grassland results in a loss of habitats, followed by a significant loss of biodiversity. Another negative effect is soil degradation in urban and rural areas, which could also lead to a severe food shortage and potential land grabbing.

The following Figure (Fig.21) from the latest World Urbanization Prospect displays the global growth rates of urban agglomerations. It is obvious, that the fastest growing conurbations can be found primarily in Western and Central Africa, but also in Central America and South and Southeast Asia. The slowest growing urban agglomerations are in the *old* world with western character, like Europe, Eastern USA and Japan. The map also classifies the affected cities, but the growth rate unfortunately does not distinguish between growth and potential shrinking rates.

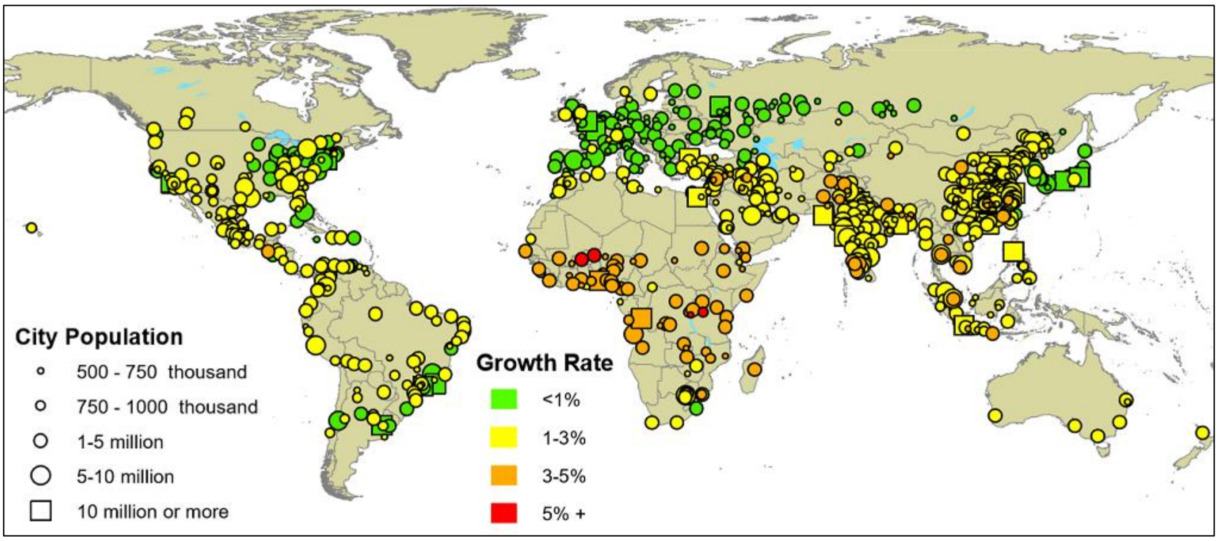


Fig.21: Growth rates of urban agglomerations worldwide based on size class in 2014 [UND-14]

The next Figure (Fig.22) shows the percentages of urban areas and urban agglomerations worldwide. It illustrates the percentage of people, who lived in urban or urbanised areas in 2014. The colour scale offers five classes from a low urban percentage in cyan (0-20 %) up to a very high percentage in brown (80-100 %). Although graphically questionable, it is noticeable that mostly industrialised countries like Argentina, Australia, Brazil, Canada, the USA, Saudi-Arabia or the Scandinavian countries show the highest shares of urban agglomerations. Some surprisingly high values can be found in countries like

Western Sahara or Gabon. The lowest percentages of urban areas are mainly in developing countries in Africa and Southeast Asia like Niger, South Sudan, Ethiopia, Uganda, Nepal or Papua New Guinea.

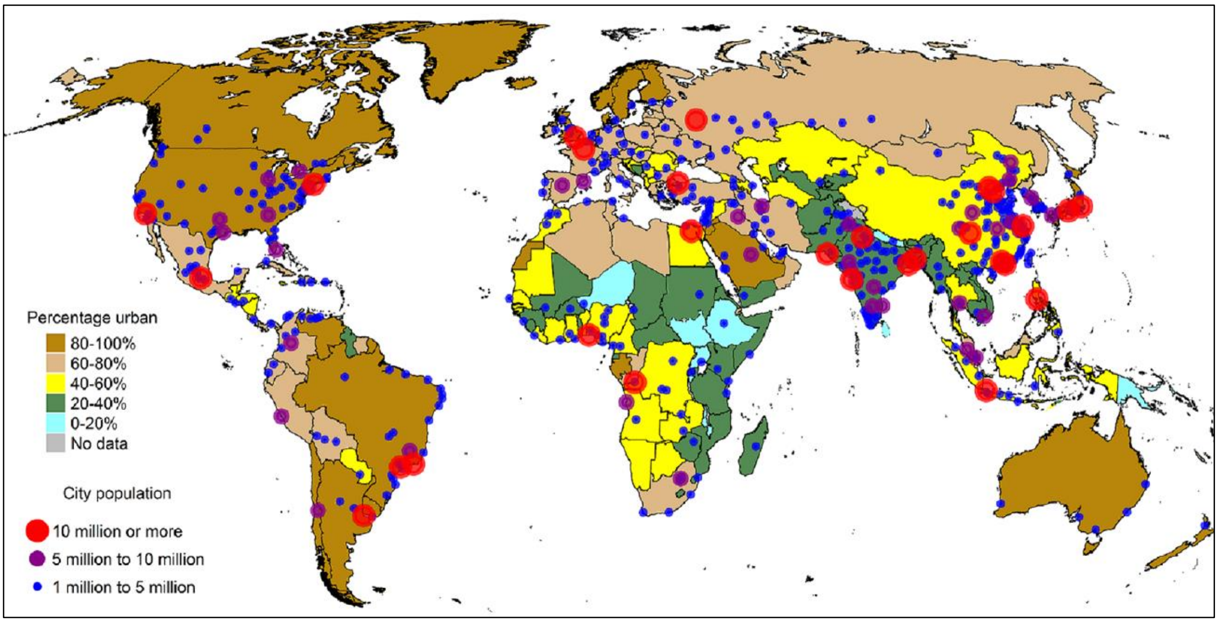


Fig.22: Percentage of global urban and urban agglomeration areas based on size class in 2014 [UND-14]

Urbanisation is also constant companion in the research area. The Danube is the longest and most important river in Austria. The Danube valley and its surrounding landscape forms the core area in Austria and Central Europe [AUF-17]. There are more people living around the riverbanks of the Danube, than along any other watercourse in the country (about half the population). According to the Austrian Hailstorm-Insurance [HAG-17], every day the area of an average Austrian farm (approx. 200,000 m²) is sealed permanently and therefore lost. This means that in the ten seconds it took the reader of this thesis to read the previous sentence, around 32 m² of soil has been covered. If this speed continues, there will be no agricultural land left in Austria in the year 2200 and soil will be reduced only to its limited carrier functions. Every year, 0.5% of Austrian soil is covered with concrete. One of the related challenges is the insufficient use of already built up properties (residential or commercial buildings), which lie idle in the country. The problem with this kind of overdevelopment is, that new properties are built on green fields or just outside the city boundaries, instead of using brownfields.

The Austrian proportion of urban and rural areas (Fig.23) developed very differently from the global proportion (Fig.20) in the same period of time (1950-2050). Unlike the global development, the percentage of urbanised areas in Austria has been very constant over the last 70 years (approx. 65%). The United Nations estimate a slow increase of urban areas from 2020 to 2050 (+10%). The percentage of rural areas will therefore decrease to the same extent (approx. -10%).

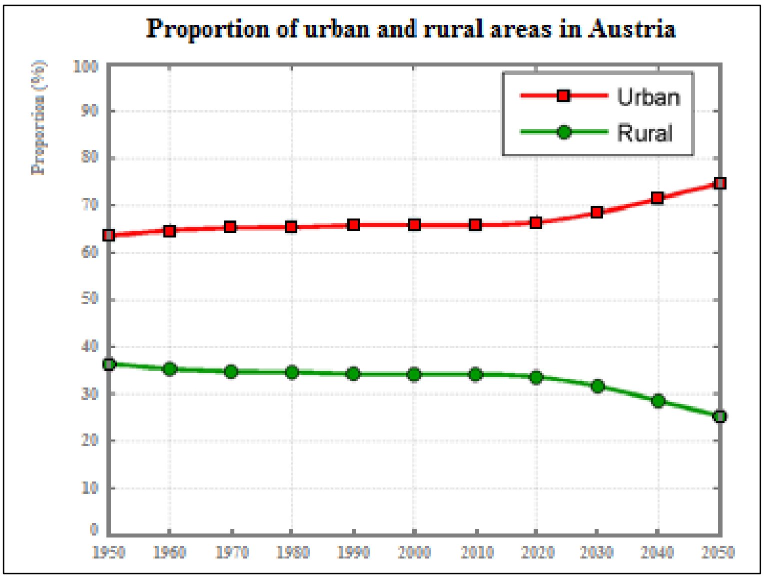


Fig.23: Proportion and estimation of urban and rural areas in Austria between 1950 and 2050 (based on [UND-14])

The following figure (Fig.24) illustrates the percentage of the already sealed areas in Austria in the year 2012. The sealing is mostly prevailing in urban centres and the biggest cities in the country (Innsbruck, Salzburg, Linz, Vienna or Graz). On a closer inspection, the river course of the Danube is clearly perceivable between Linz and Vienna, due to a higher percentage of sealed areas alongside of it. This perception depicts the potential vulnerability of densely populated urban areas in Austria very well. It demonstrates quite figuratively the need to investigate the affected regions for risks and hazards.

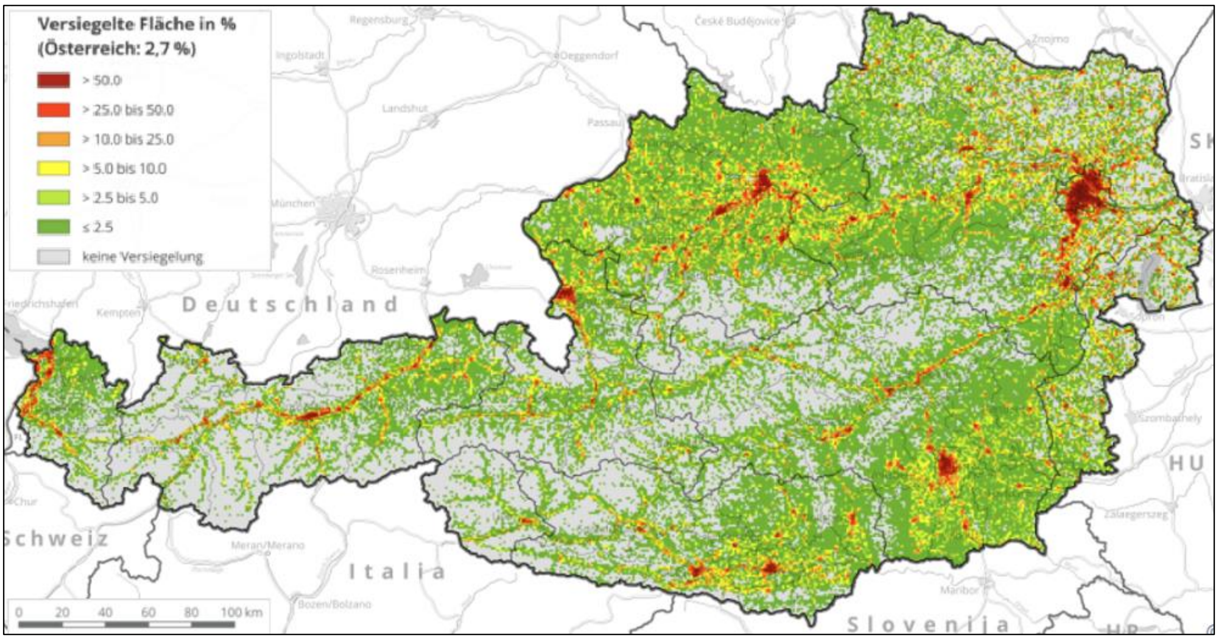


Fig.24: Percentage of sealed area in Austria 2012 [OER-12]

2.4.2. Degradation and Sealing of Soil

The sealing of soil is one of the various types of soil degradation, which can be defined as "a change in the soil health status resulting in a diminished capacity of the ecosystem to provide goods and services for its beneficiaries" [FAO-17e]. Degradation can be the deterioration of soil through different causes and can lead to a decrease in soil productivity and fertility because of human interaction. The decline of soil quality is mainly amplified by climatic conditions, agricultural mismanagement and industrial exploitation. This can result in waste and fallow land, which is unusable for agricultural usage and threaten the cultivation of crops and consequently the food availability. The causes of soil and land degradation can be very versatile and could cause numerous hazards. They can comprise physical, chemical and biological processes and anthropogenic activities:

- *Water erosion* is one of the most serious causes of global soil degradation and it is a "threat to the provision of food supply and security" [OLL-16]. Intense rainfall, steady water flows and floods can wash away the top layers of the ground and accumulated materials such as clay, sand and organic substances [OMA-15] (Fig.25). The severity of water erosion is also depending on the climate, susceptibility of soil (due to density, permeability or structure), slope gradient (the steeper, the more vulnerable), topography and vegetation and its degree of rootedness [OMA-15].
- *Wind erosion* has similar effects on the top layer of the soil as water erosion (Fig.26). This physical process is intensified by dry or arid ground. It is one of the driving factors of degradation and desertification. The effect of Aeolian erosion is increased by improper agriculture, deforestation and urbanisation. The magnitude of wind erosion is also composed of climatic conditions, susceptibility and roughness of soil or land cover [OEH-14]. Soil particles are transported depending on their size by saltation, surficial creeping or suspension and can lead to damaged crops, lower moisture holding capacity and public safety hazards [OMA-15].

Fig.25: Typical form of soil erosion by water [NTU-14]

Fig.26: Erosion by wind on agricultural soil [OEH-17]

- *Deforestation* is the "process in which all the trees in a large area are cut down" [CAM-17a]. More complicated however, are the causes and consequences on soil degradation and the environment in general. About 31% of the Earth's land cover are forests and woodlands [WWF-17] and they are largely responsible for the global production of oxygen and the sequestration of carbon dioxide. Forests protect endangered wildlife, vegetation and the soil they are rooted in. Every year approx. 13 million hectares of woodland are deforested [FAO-17f], which equals the size of countries like Nicaragua or Greece [CIA-17]. Exposed soil is much more vulnerable to all sorts of erosion and the degradation rate of mineral soil, organic matter and humus raises distinctly. Deforestation also favours the flood risk in affected areas, because the protective land cover has been removed for plantations, the soil erodes, the water absorbing capacity decreases and, in combination with insufficient drainage systems, the flood hazard increases as a further consequence.
- Another cause is *tillage erosion*. Extensive or improper cultivation practices (extensive grazing, farming on slopes, fertilisation, irrigation, monocultures) are the main reasons for the "progressive down-slope movement of soil, causing severe soil loss" [OMA-15]. It facilitates water erosion and due to deep ploughing, the amount of vulnerable soil has been seriously increased. Conventional farming offers not only room for economic and environmental improvement, but also potential negative effects on soil erosion rates and water quality [NRC-10]. Soil erosion from agricultural land is even one of the most important driving factors of deteriorated water quality. Soil quality indicators provide information about productivity of animals, plants and about air or water quality [SIO-17]. Value changes indicate considerable transformations, which could implicate soil degradation.
- Additional contributing factors can be due to *chemical* (e.g. acidity, salinisation or lack of nutrients) or *biological processes* (e.g. reduced activity of microbial flora) [BIO-17].

All these causes have severe impacts on the global environment. The yields of agricultural produce are reducing, food availability decreases and soils are getting permanently infertile. The excessive use of fertilizers also has a severe impact on soil and water quality. Economic losses because of soil erosion costs every year around 70 Dollar (66 Euro) per person [USD-01]. One of the most important consequences of soil degradation is the increase in natural disasters (or rather *anthropogenically influenced*). The number of droughts, famines, floods, mud flows and pollution is on the rise, resulting in increasing environmental, economic and social losses and damages.

Soil sealing as an anthropogenic factor is the permanent covering of land with impervious materials like concrete, pavers, stones or tarmac. It is mainly a result of the development and expansion of existing or new infrastructure, private and commercial buildings as well as other facilities. The sealing of the ground is therefore "closely related to land take or land consumptions, which is understood as the development of open areas […] into built-up areas" [ART-15]. Urbanised areas and cities in general are especially characterised by extensive areas of sealed surfaces, "limiting the supply of ecosystem services"

[ART-15]. Although the population growth is no longer rising dramatically, the rate of sealed urban spaces is still increasing. In the EU, every year an area as large as the City of Berlin is transformed into an urban environment (approx. 890 km²) and half of it totally and permanently sealed [EUC-16c].

The sealing of the soil can be caused by different triggers like anthropogenic factors, but soil can also be naturally quite impervious. If the texture is very fine grained (e.g. silts, clay or loam), water cannot penetrate deeper soil layers because of higher resistance to potential water flows [CET-11]. The sealing of the soil's surface "often affects fertile agricultural land, puts biodiversity at risk, [and] increases the risk of flooding and water scarcity" [EUC-16d]. The current condition of soil sealing (Fig.27) is especially poor in Northern Africa, quite poor in Europe and Asia and fair on the American continent. Good conditions can be found in sub-Saharan Africa, Australia and the southwest Pacific region.

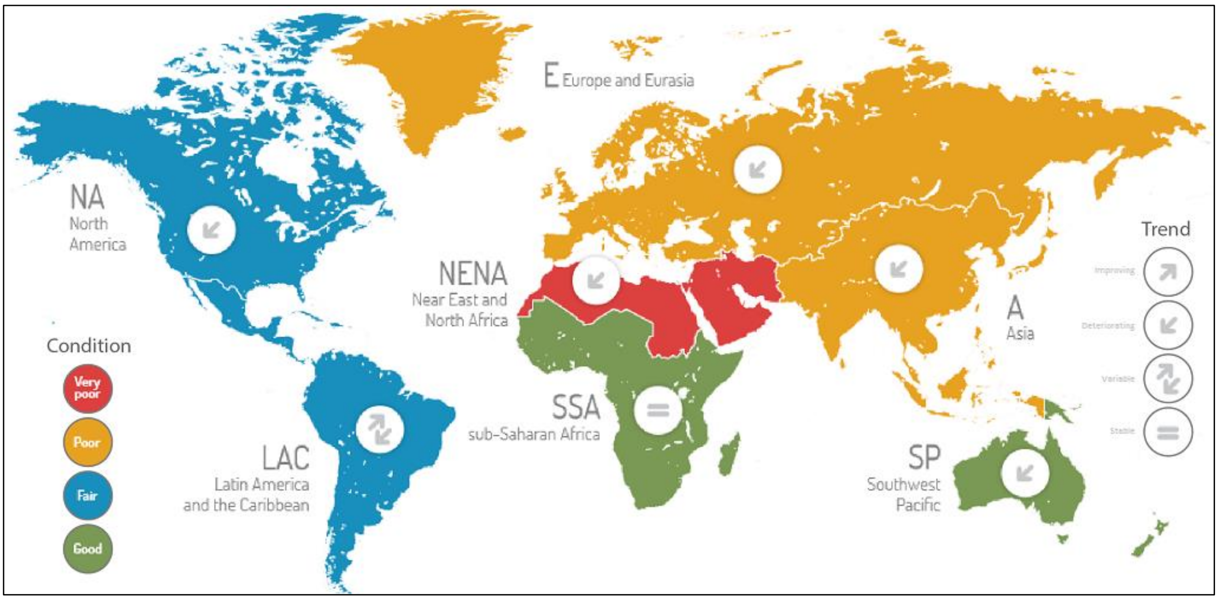


Fig.27: Current condition of soil sealing worldwide, based on edited classifications by FAO [FAO-16]

Soil however is an "essential, finite and non-renewable natural resource" [EUC-16c], which offers so many functions and serves a lot of purposes. Soil should therefore be sustainably managed. Every permanent covering of soil is very hard to reverse and consequently results in the definitive loss of soil [IAS-13]. It is therefore essential to use this resource reasonably and sustainably. Thoughtful soil management and spatial planning can lead to a more gentle use of natural resources and to an improvement of future environmental prospects [EUC-12]. This subject is often a reason for dispute between economic, social and environmental objectives of a community, which include commercial, industrial, infrastructural, residential and recreational areas within an urban environment [ART-14]. It is now up to decision-makers and urban planning agencies to overcome these challenges.

The changes due to urbanisation "threaten the European landscape and lead to fragmentation of natural habitats and ecological corridors" [EUC-12]. Taking appropriate measures is quite expensive and therefore often ignored. This ignorance has serious impacts on urban development. The continuous sealing of open surfaces results in impacts on the ecosystem services and the loss of the soil's biological functions. This leads to the enabling of urban heat islands and heat effects in cities, the loss of natural carbon sequestration capacities, the reduction of filtering and purification capabilities, the loss of food and fibre production, the endangering of biological diversity, reduced quality of living and a decrease in the soil's water absorption capacities [FAO-16] and therefore an increased flood risk. Normal soil can store huge amounts of water and precipitation. Extensive sealing can prevent the absorption completely and increase risks. The prevention of water infiltration also alters the soil hydrology significantly. If surface areas are covered, less water is stored in the different soil layers and available to plants. This increases the probability of droughts and the need for excessive fertilisation and irrigation.

Other impacts are increased run-offs from impervious surfaces, altered water quality, straightening and concreting of river beds or interfering with the sediment budget [EEA-16]. These effects of soil sealing can all influence the flood risk, frequency and extent. In order to satisfy the increased need for agricultural and construction land, new fertile areas need to be found to guarantee a sufficient food supply. More and more vulnerable space (floodplains, flood zones or river banks) are then converted into farmland or construction sites. This decreases the amount of freely available areas and increases the conflict potential.

Throughout history, territorial disputes have always been a usual way to secure and expand the ruler's claim to power. Wars and colonisation were the inevitable consequences. Nowadays, the strong international economic ties makes it highly unlikely that a developed country attacks another developed country in order to expand their territory. Other ways of "conquering" new territory have to be found. The uncertainty about the sufficient supply of energy, food and water has triggered so-called *transnational land acquisitions*. This phenomenon, also known as land grabbing, is the occupation of mostly agricultural land in poorer and less developed countries by other nations or companies. The aim of these acquirers is to buy huge areas of land in poor countries in south-east Asia or Africa [EEA-16] and invest in the cultivation of different crops, which are then used to produce food or biofuel for the investing nations. This kind of extensive agricultural business also goes hand in hand with water grabbing, which has serious impacts on the local soil hydrology and on the rural poor.

In order to prevent or reduce this conflict potential, the degradation and sealing of soil has to be mitigated. The best ways to restore and regenerate soil to fulfil its natural functions are imposing planning restrictions, providing planning guidance, protection of fertile arable soil, reinstalling greenbelts and de-sealing of covered areas, regenerating and reusing of already sealed waste land and brownfields as well as supporting ecologic building technologies (e.g. permeable materials and surfaces) [EUC-12]. All these approaches require adequate land cover information, which this master thesis will try to provide for the research area.

2.4.3. Current Situation at the Danube River

In the whole of Europe, soil and its functions are considered as a matter of course and very often no further thoughts are wasted on potential implications [JRC-17]. Consequences on agriculture and ecosystems have already been discussed sufficiently. Degradation and sealing of soil however could also influence the sediment budget and dynamics of rivers and the navigability of inland waterways [JRC-13]. One of the most important European waterways is the Danube. With the help of the Rhine-Main-Danube Canal, various types of cargo can be transported from the North Sea to the Black Sea. As a result of that, a number of private and official initiatives were started in the past few years, in order to improve the quality of planning strategies based on reliable scientific data. Time, effort and money had to be invested to collect high-quality and relevant information. This kind of investment should motivate people to interact sustainably with their surrounding environment, as well as contribute largely to the communication with scientific initiatives like citizen science.

The following figure (Fig.28) displays the average annual density change of sealed areas in Europe between 2006 and 2009. The colour scheme is improvable but it is clear to see where the most urbanised areas are. The biggest increase of sealed areas were in urban centres like Germany, Northern Italy and the Benelux countries. In Austria however, the biggest noticeable change has taken place alongside the Danube River. Although not quite as distinctive as in Northern Italy for example, it clearly contrasts with the rest of the Austrian territory.

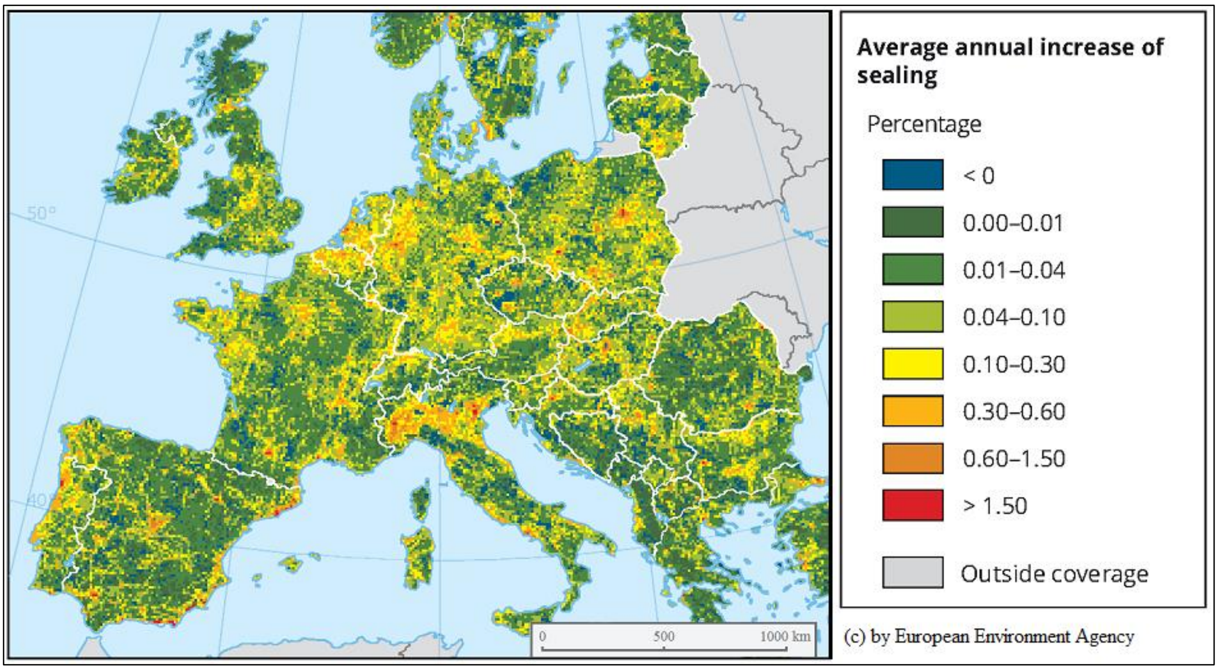


Fig.28: Display of the average annual density change of impervious or sealed soil between 2006 and 2009 in Europe ([EEN-17] edited)

This matches the information about urbanisation in Austria in the corresponding chapter. The next figure (Fig.29) outlines the historical importance and development of the Danube River. The illustration shows the changes of land cover within the watershed of the Danube River in Central and South East Europe over almost 1000 years. The data is based on current research results and reconstructed historical records. This scientific method is often used to indicate the degree of deforestation and the loss of vegetation cover [GIO-12]. Red areas represent the areas where natural vegetation (forests, meadows) has been removed and converted into cultivated areas or settlements. Green surfaces stayed naturally overgrown. In 1000 AD, most of the changes took place in higher populated areas which are today known as Linz, Vienna, Brno or Bratislava. 850 years later however, the scenario has changed dramatically. A big share of the investigated area has lost its natural vegetation. The fractions which have not been affected by the transformation are mainly concentrated in mountainous regions. It is not extractable from the data, which regions had turned into farmland, permanently covered or sealed areas.

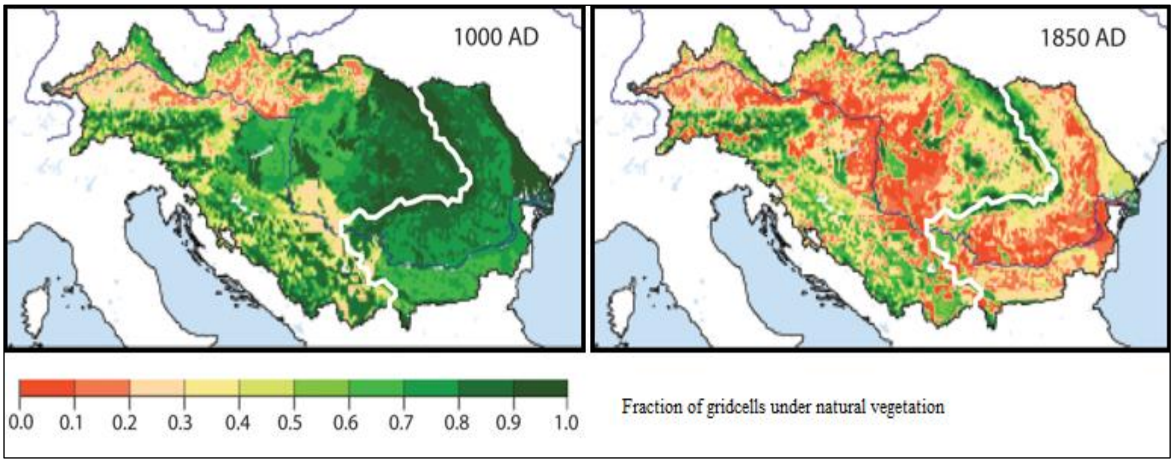


Fig.29: Model of land use change in the watershed of Danube River ([GIO-12] edited)

The river meadows alongside the Danube regulated the water balance in times of peak discharges or low water levels for centuries. With the regulation and straightening of watercourses, this volatile equilibrium got seriously disrupted. In the early days of adjusting river courses and basins in Lower Austria (around 1875), the number of extreme water levels rose up nearly sevenfold [JUN-14]. This led to further interferences with the river ecology. In the beginning of the 20th century, measurable effects of this often unthoughtful constructions comprised of increased flood peaks and accelerated inundations. Did it took flood waves on the Danube 54 hours to travel from Ybbs (Lower Austria) to Vienna in 1954 (around 95 kilometres in linear distance [LUF-17]), it sped up to just 16 hours in 2011 due to various river construction works [JUN-14]. In the last few years, protests rose up to stop the modification and concreting of river beds in order to improve navigability.

To get a more detailed picture about the current situation in this area, further research has to be done.

2.4.4. Comparable Research

There are currently numerous ways of detecting covered land areas. The use of satellite information and aerial imagery is the most common technique to obtain accurate and suitable data on the type of land cover and the extent of sealed soils. This offers valuable sources of geospatial data for defining the ecological implications of urbanisation [MIL-14]. The gathered information is then processed, analysed, evaluated and integrated into models and land cover data sets. Land cover is by definition the "observed (bio-) physical cover on the earth's surface" [FAO-00] and comprises of free surface areas (e.g. grassland, forests, waterbodies) and sealed areas (e.g. buildings, streets). Mainly based on assessing remote sensing products, these land cover datasets are used as the basis for further scientific research, like hydrologic modelling or the detection of climatic changes [TSE-15]. Due to different sensor types, mode of operation and scope, they disagree in various ways. Some of them are listed below:

~ *GLC* (Global Land Cover Network) was initiated by the European Commission and receives its data from the SPOT4 satellite systems, which focus on long-term observation of vegetation [CNE-04].

~ *GlobCover* refers to a programme by the European Space Agency (ESA) and receives its data from the Medium Resolution Imaging Spectrometer on board of the Environmental Satellites [DUE-05].

~ *MODIS* (Moderate-Resolution Imaging Spectroradiometer) is a part of the US-American Terra Satellite [MOD-17]. Its main purpose is to support geophysical research and short-term observations.

~ *Copernicus* is an observation programme which was initiated by ESA and the European Commission, and is based on Sentinel satellites and in-situ data [CPC-17]. The main thematic focus of *Copernicus* lies on the atmosphere, climate, emergencies, land, marine and security services.

~ *CORINE* (Co-Ordinated Information on the Environment dataset) was initiated by the European Community to collect standardised and comparable information on environmental changes across Europe [EPA-17]. In 2006, information about artificially sealed surfaces were integrated.

~ *LUCAS* (Land Use Cover Area frame Sample) is an "in-situ sample of land use and land cover carried out across EU member countries" [BAY-16]. Its aim is to create a geospatial dataset based on standard samples by surveyors and is used for improving official urban planning.

At closer inspection of the available satellite imagery and aerial photographs, it is quite obvious that urban areas look distinctively different from forests or other unsealed green areas. Therefore a lot of studies are dealing with the subject of detecting and measuring soil sealing with the aid of remote sensing technologies. They often use the mentioned land cover datasets as reference systems to assess the analysed data's accuracy. There is however hardly any research available on possible linkages between sealed soils and the extent or damage of recurring flood events using crowdsourced data. Most of the studies focus either on using satellite images to detect soil sealing and its changes [GAR-13] or improving the estimation of impervious surfaces [XUR-17]. Only very few also focus remotely on the connection between sealing of the soil and flood risks [PIS-15], and even fewer include crowdsourcing methods. Some of the comparable studies and their research methods will now be discussed shortly:

~ The first presented paper concentrates on the measurement of the soil sealing extent near Madrid (Spain) [GAR-13]. A number of Landsat TM datasets, recorded between 1989 and 2010, were used to estimate the loss of arable land due to soil sealing. Using the usual method to process and analyse remote sensing data, the main results were the benefit of free access to high resolution satellite data, the facilitated detection of changes in land cover and sealed surfaces (> 400 km² in 20 years), and the advantage of semi-automatic classification compared to automatic.

~ Chinese scientists stated, that the "accuracy for impervious surface estimation is insufficient due to high diversity of urban land cover types" [XUR-17]. Their approach to solve this issue was to use improved tools for the image processing (e.g. pan-sharpening), whereas this thesis should demonstrate the benefit of voluntarily generated information to analyse and assess satellite-derived data. Similar research activities used dual polarisation Synthetic Aperture Radar (SAR) instead and claimed it is possible the best remote sensing method to detect impervious surfaces [ZHA-16].

~ The next comparable study has already been quoted in this thesis and deals with the assessment of the soil sealing impact on increased flood peaks and volumes in the plains of Emilia-Romagna in Italy [PIS-15]. It is one of the few studies, where the measurement of soil sealing was not the key issue but its connection with increased flood frequency and extent. They compared estimated inundation volumes and compensatory volumes for flood detention between 1976 and 2008, in order to characterise the impact of soil imperviousness on drainage networks in semi-rural areas [PIS-15].

~ Another study focusses on the quantification of impervious surfaces in order to investigate its impact on flood processes [DUS-15]. It points out, that the few studies referring to imperviousness of soil and its connection with flood events, frequently offer contradictory results. It seems that there may be substantial disagreements within the research findings. However, by using hydrologic modelling systems, specific indicators were developed to quantify the impact of sealed surfaces. The results confirmed their theories and suggested that "cities should consider an increase in flood discharge under future IS (note: impervious surface) scenarios" [DUS-15] when planning new flood protection.

~ The last presented study deals with the use of mobile technology and gamification to crowdsource data about land cover [BAY-16]. It is closely related to this master thesis, because it used similar data generated by the FotoQuest Austria campaign (which was initiated by the International Institute for Applied Systems Analysis in 2015). The aim of the study was to compare crowdsourced data, collected by untrained citizens, to established land cover and land use datasets, set up by experts (LUCAS). Although the agreement was considered not high enough (~70%), the operational capability of the mobile application for planning authorities to gather information is definitely there.

2.5. Citizen Science and Volunteered Geographic Information (VGI)

The next important thematic block discusses the crowdsourcing approach, its useful derivatives and the usability of this method in modern system analyses. This topic has been the author's field of research right from the beginning of his bachelor programme and was one of the main drivers for choosing the overarching research question. The all-embracing concept in question is called crowdsourcing. This compound word consists of the terms *crowd* and *outsourcing*, which was first named so by the US-American Journalist Jeff Howe in 2006 [HOW-06]. It comprises, amongst other things, the quite frequently developed derivatives crowdfunding, micro-tasks, citizen science and volunteered geographic information (VGI).

Crowdsourcing is often defined nowadays as the outsourcing of specific tasks to a group of non-experts, in order to solve a certain problem in a faster and more economical way. This method should facilitate the traditional way of generating data by replacing technical sensors with human volunteers. Due to the common phrase *wisdom of the crowd*, every task is repeatedly validated by different participants and therefore provides a quality control and protection from deliberate misconduct or statistical outliers. Motivating incentives could comprise monetary, scientific or even altruistic aspects. After analysing about 30 of the existing definitions in the literature, Spanish scientists tried to provide an integrated and universally usable definition for the crowdsourcing approach, which should cover all possible associated initiatives [EST-12]:

> *"Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. […] The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken."*

In many ways, the concept of crowdsourcing existed for centuries, before the distinctive name has been coined for the first time. The use of crowdsourced information and ideas can already be found in the early years of the Industrial Revolution where crowds or communities were used as a way to solve social or scientific problems.

At the beginning of the 18th century, navigating in the stormy oceans was challenging and dangerous. Knowing where the ship was, the determination of the exact longitude and latitude, which course to set and what time it was in relation to the departure or arrival point were the essential tasks for the ship's crew [WIT-15]. Without these information, the safe arrival at the destination could not be guaranteed. The accurate measurement of time was and still is a vital part of naval navigation. Today most of the technical operations are done by computers and atomic clocks that are incorporated into satellite systems. Back then, it all relied on the observation of eclipses and the comparison of the moon's relation to the stars or local meridians [WIT-15]. These methods unfortunately were quite complex and prone to error. Another approach included the accurate time at the arrival or departure point, but in those days robust and precise clockwork mechanisms could not be built in a portable form. Life at sea was pretty rough and the technical mechanisms had to resist cold or hot weather, saltwater and violent movement due to strong waves. For this reason, the British Government offered a prize of 20,000 British Pounds (GBP) in 1714 for the solution of this *Longitude Problem* [BET-93]. The prize money was so high (the equivalent of several millions today), because no one of the responsible government board really thought that it was possible to create. After a number of weird or unrealisable suggestions, the carpenter and clockmaker John Harrison came up with his first marine timekeeper in the 1730s. Dedicating his remaining life to this problem, Harrison was finally declared winner of the Longitude Prize in the 1770s and was still able to present the first accurate marine chronometer. Thus proofed, that the creation of such a portable and robust timekeeper was indeed possible [BET-93]. He did not only win the first kind of crowdsourcing competition, but also revolutionised global naval navigation.

Another milestone in the implementation of crowdsourcing into social and scientific initiatives is the Toyota Logo Contest in 1936. The Japanese car manufacturer wanted to update its company logo and received more than 27.000 suggestions [ROY-15]. The winning entry combined three Japanese characters, which even led to the renaming of the company from the former *Toyoda* to the more illustrative *Toyota*. The final design of the Sydney Opera House was also based on a similar crowdsourcing contest. More than 230 ideas were sent in from 32 countries and led to this fantastic icon, the whole world knows and recognises [ROY-15].

Wikipedia is a modern phenomenon. Founded in January 2001, it is a freely accessible online encyclopaedia and one of the most visited websites in the world [AYE-08]. It covers every possible topic or subject and shows no thematic limitations. According to Wikipedia itself [WIK-17], there were more than 40 million articles available in 293 language editions in the year 2016. The platform is based on voluntarily contributed articles, which are peer-reviewed before they get published. It serves as one of the most successful examples in the history of crowdsourcing worldwide, on a par with OpenStreetMap, YouTube, FoldIt or Waze (Fig.30).

As already indicated, the development and accessibility of the Internet was one of the main drivers for the explosive growth of crowdsourcing platforms and initiatives. More and more aspects of life, society or culture are transformed in the exponentially expanding worldwide network of information [CHO-17]. The present time is often referred to as the *Digital Age* or the *Age of Information*. Some predictions of the future even go so far as proposing the dismantling of nation states and the implementation of a digital citizenship, but this is all still up in the air. Since the development and broader global availability of the Internet however, the whole methodical and administrative structure of crowdsourcing has changed considerably. Thanks to this global network, the accessibility of potential volunteers has exponentiated. Since the first humans existed, they had to be physically close to each other in order to form a crowd. Since the introduction of a world wide web, people from all over the world can be considered as a crowd, if they all work on the same project. This progress also supported the recent emergence of crowdsourcing platforms. As a result, companies and organisations can publish their specific problems in the form of contests and rely on the cooperation of the participating community. The major stakeholders of a crowdsourcing process are the data producers (public or private organisations, untrained volunteers), developers (GIS experts, programmers or software developers) and consumers (broader public, companies, educational institutions or local governments) [BRO-13]. Throughout the last few years, a number of serious projects confirmed the usability of the configuration of these different stakeholders in various subject areas. Some of them could be useful in the course of this thesis.

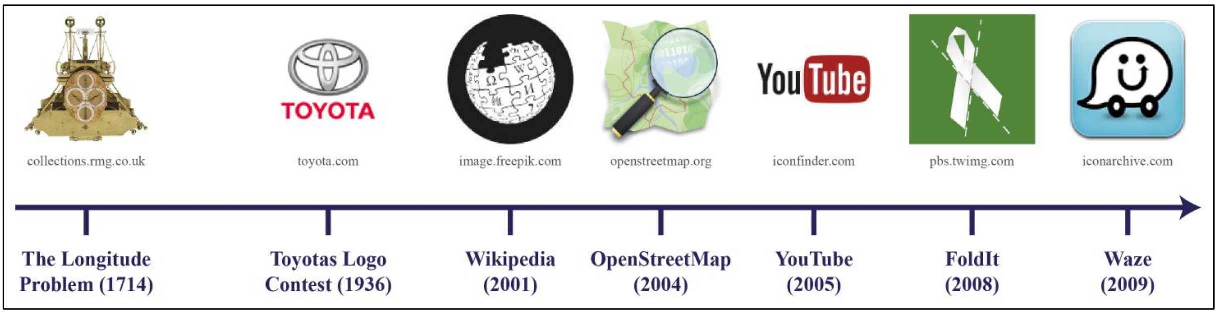


Fig.30: Brief timeline of the milestones in crowdsourcing history

The application of crowdsourcing is depending heavily on remote sensing technology and further processing by geographic information systems (GIS). After the data acquisition via satellite systems or photogrammetric imagery, the next step would be the initial analysis and assessment of the generated information, in order to create a supportive data base for the application in disaster management or urban planning. Crowdsourcing and its derivatives geographic citizen science and volunteered geographic information are very versatile approaches and can therefore be used in the field of disaster management, emergency response or flood risk reduction, in combination with social media technology. The next few paragraphs examine the use of these approaches in recent disasters:

~ *Haiti*: The disaster response and management operations during the Haiti earthquake in January 2010 were very difficult and dangerous ventures [CAL-16]. An earthquake with a magnitude of seven on the moment magnitude scale (M_W) hit the western part of Hispaniola and caused an estimated 100,000 to 200,000 casualties. Thankfully, a lot of human lives and properties were saved by private and official crowdsourcing operations [RIC-16]. These special crowdsourcing operations collected and assessed messages of affected Haitians, to coordinate the effective and fast transport of adequate and sufficient aid to people in need [RIC-16]. Even international disaster response initiatives relied upon the information collection via social media services. This data was successfully used to "coordinate knowledge and action between cooperating response agencies" [YAT-11] which also included the United States armed forces.

~ *Colorado*: Another similar example with a different kind of disaster, were the extensive wildfires in the state of Colorado. In 2012 and 2013 a series of massive fires destroyed forests on a large scale. The 2013 wildfires killed two people, demolished nearly 500 homes, burned down around 57 million square metres of wood and claimed an economic damage of more than 400 million US-Dollar [RIC-16]. Due to these cleared areas, floods occurred later the same year and destroyed even more housings. However, the constructive cooperation between disaster management, Homeland Security and the public by interviews and email communication showed the benefits of crowdsourcing in disaster management [RIC-16].

~ *Fukushima*: In 2011, the so called Tōhoku-earthquake hit the eastern coast of Japan with a magnitude of 9 M_W and triggered a tsunami, which lead to the already mentioned Fukushima-Daiichi nuclear disaster. Japanese engineers used the crowdsourcing approach during emergency response operations [RIC-16] and after the immediate event. An initiative was launched, where volunteers could measure and submit radiation values from all over the country. This provided a freely available and up-to-date radiation map of Japan [HIR-13].

2.5.1. Volunteered Geographic Information

One integral part of crowdsourcing is VGI. This stands for volunteered geographic information and can be considered as the combination of public participation and geographic information. The VGI-approach has developed in a time, where the collection, processing, assessment and publication of spatial data saw substantial changes due to improved performance and accessibility of the Internet [SUI-13]. Extensive growth of user generated spatial information has triggered a number of similar or identical phenomena, e.g. neogeography, geoweb, maps 2.0, citizen mapping and of course volunteered geographic information [LIN-15]. Different names for almost interchangeable concepts. They can be defined as “the breaking down of the traditional distinctions between expert and non-expert, in the specific context of the creation of geographic information, since all of the traditional forms of expertise can now be acquired through the use of technology” [GOO-10]. All these approaches can therefore be summarised as the generation of georeferenced data by mostly voluntary untrained non-experts, with the help of appropriate digital platforms and tools (API, KML, geotags), which can then be integrated into geospatial analyses, in a faster and cheaper way.

Until the 1990s, the traditional way of collecting and producing specific geospatial data included the assessment of maps and surveying results by official authorities or companies. The resultant products were published as paper copies and made available through a system of retailers [GOO-10]. The basis data and acquisition methods stayed reserved for experts only. The modern era of collecting data started in the early years of the 1990s thanks to new information technologies. This included the wider accessibility of personal computing systems, the shutdown of selective availability of global satellite systems (GPS) and the rapid development of the Internet. Special services like Keyhole or Google Earth respectively and tools like API (application programming interface) enabled everyone with a personal computer and Internet access to “gain the ability to make maps from acquired data, and to employ the kinds of cartographic design skills previously possessed only by trained cartographers” [GOO-10], whether they had any previous experience or not.

The world is becoming smaller and smaller as a result of the fast digitalisation of people’s life, but the need for georeferenced spatial data is on the increase. This kind of information can be of considerable commercial and social value for public organisations or profit-oriented companies [FEI-13]. Nearly every public community, governmental authority or private corporation has its own spatial data infrastructure, which focusses on processing and analysing geographic information. A lot of part-time university lecturers even work in a commercial GIS department as their main profession. Traditional methods of collecting data have always been a long process, carried out by trained experts. User-generated content facilitates the challenges and reduces the costs significantly. This enables smaller businesses with lower budgets to compete with transnational corporations or global players. Some of the highlights in crowdsourcing history owe their existence to these possibilities (Fig.30). However, the generated and processed data has to offer a specific characteristic in order to be useful for private or public entities: currency, positional and qualitative accuracy, reliability and resolution [FEI-13]. Ways to control the validity of these characteristics are deterministic methods like accuracy or quality assessment and tests like the root-mean-square error or the Cohen's kappa coefficient.

When the quality of a certain spatial dataset depends on so many components (and more, e.g. time), all collected geographic information may "be subject to uncertainty" [GOO-10]. When initiatives like flood risk management, urban planning or emergency response depend on user-generated data, uncertainty has to be reduced to a minimum. Traditional methods rely on the knowledge of highly trained experts, which are in charge of the whole process. Volunteered geodata has to persist a couple of control procedures. The quality assurance methods of VGI comprise, amongst others, the attribute accuracy, positional accuracy, lineage, logical consistency and the completeness of present data [ALI-14]. The quality also depends on the existence of a distinctive classification system or the motivation and experience of the volunteers [ALI-14]. To guarantee at least a certain level of correctness and conformity (by the increasing number of participants within crowdsourcing projects), some kind of thematic introduction or training material has to be provided.

The qualities and opportunities of volunteered geographic information also reach higher governmental levels nowadays. Creating a faster and more effective way of communicating with citizens also contributes to a greater governmental transparency [JOH-13]. It offers the potential of an increase in direct democracy, or citizens as decision makers, and mutual governance. The use of people as social and economic sensors can lead to a leaner and less bureaucratic administration, which matches with current developments of downsizing local authorities in western countries [JOH-13]. Relying on human sensors provides not only faster and more frequent information, but also reduces the costs for additional employees or resources. Another benefit from this two-way cooperation between citizens and the government is the fact that, thanks to frequent polls and joint projects, economic or social decisions match more with people's needs and are therefore more likely to be broadly accepted throughout the community [JOH-13].

Even though this approach sounds very reasonable, there are a number of obstacles to overcome before VGI initiatives become an omnipresent phenomenon in governmental policy making and planning. Some reservations concern data privacy, "organizational and cultural challenges, technological issues, and issues involving the scaled and interconnected nature of governance" [JOH-13]. There are already "normal" GI-systems implemented into official administrations, but there are four substantive areas where there are still application possibilities. This includes citizen-oriented transit information, citizen relationship management, citizen-volunteered geographic information and citizen participation in planning and decision making [GAP-10]. The latter aspect can also be extended to urban planning and disaster management in this case.

Disaster management and risk reduction can benefit from the implementation of VGI and associated initiatives. Before, during and after an emergency, people can write messages or send tweets with georeferenced information about the situation and accumulate large amounts of valuable geodata. This data then can be implemented in every phase of disaster management (preparedness, response, etc.). Although there are some limitations of this method, which have to be considered [CAM-14], projects like *Ushahidi* or the *Humanitarian OpenStreetMap Team* [HOT-17] have proven their usefulness in emergency cases. VGI is based on rapidly developing technology and transforming social media, and if it is able to adapt to future changes, the role of volunteered geographic information in human society will continue to grow [CAM-14].

2.5.2. Citizen Science

The second derivative of crowdsourcing which will be discussed in this chapter, is citizen science. It is hard to find a distinctive definition for the term, because it includes different concepts and is perceived differently in scientific literature. Some authors call it the "longest running of VGI activities" [HAK-13] and some locate the approach more in a non-geographical environment [DON-14]. It can possibly be defined as the participation of untrained volunteers in scientific research and the simplification of scientific communication between the public and trained academics. This could include different levels of engagement, like the active intellectual contribution of information or just the passive supply of their tools and resources [SEE-16]. It aims to empower people's thirst for knowledge in a wide range of possible fields of application. Another major purpose of such an initiative is to broaden people's mind and to improve their cognitive capacities and skills through "democratic research" [SEE-16].

The role of the participating citizens can comprise the initial problem definition, the elaboration of possible research objectives and in the final data processing and analysis. To deepen the public understanding, the participation of voluntary contributors in the final discussion of the results is essential. Just like the VGI approach, everybody can take part and get a *scientific citizenship*. There are however similar cultural reservations about quality assurance and the missing experience or professionality of the contributors. These problems regarding data quality have been extensively addressed by the scientists. Some of the used methods include the provision of adequate formal training, a close supervision of the contributions, the constant control based on topical literature, constant checks for consistency with professional contributions and control points, conclusive queries about the volunteer's confidence and reliability as well as adapting the scientific tasks as simple and error-resistant as possible [RIE-14]. These measures should assure a certain level of data integrity and weaken some of the social concerns about using untrained citizens as sensors.

There are four consecutive levels of participation and engagement referring to citizen science initiatives, which acts as a formal framework to organise and define distinctive citizen science projects [HAK-13]:

- *Level 1 – Crowdsourcing*: citizens act as sensors, includes volunteered computing
- *Level 2 – Distributed Intelligence*: citizens act as interpreters, including volunteered thinking
- *Level 3 – Participatory Science*: participation of citizens in problem definition and data acquisition
- *Level 4 – Extreme Citizen Science*: or collaborative science, includes problem definition, collection and analysis of data (plus final discussion with researchers)

As noted before, citizen science does not have to be exclusively about geographic issues, but it can be. This possibility raises the question, where the difference in comparison to volunteered geographic information is. The data contribution by untrained volunteers and non-experts is one of the most common features of those two approaches. One of the major differences however, is already retrievable from the concepts' names: citizen science focusses on all levels of public engagement in *scientific* initiatives, whereas VGI comprises more fields of application, like private activities (e.g. recreation, vacation) or the already mentioned immediate disaster response. For this reasons, not all citizen science can be considered as geographic and not all volunteered geographic information is scientific [DON-14]. One specific subtype of citizen science is geographical citizen science, which shares even more similarities with VGI. It can be defined as the intersection between scientific research, public participation and geographic information (Fig.31). Haklay [HAK-13] states that geographic citizen science focusses heavily on the contribution of georeferenced data as a scientific activity, but because of blurred boundaries between all crowdsourcing aspects, this categorisation cannot be seen exclusively.

Finally it should be mentioned, that there have been a number of commonly known examples of citizen science projects over the last two decades. In 1999, the passive SETI@home project (Search for Extra-Terrestrial Intelligence) was released by Berkeley University, in order to use unused computational capacity from volunteers for analysing radio telescope data [HAN-10]. Years later, similar projects were initiated, like Rosetta@home and its successor project FoldIt (2005/2008, volunteers fold proteins in various forms online), Stardust@home (2006, analysis of 1.6 million microscope images to find traces of cosmic radiation) or Galaxy Zoo (2007, classifying the shape of galaxies and detecting new ones, with data from Hubble telescope) [HAN-10]. All these examples confirm the variety of possible *scientific* applications and the reason for existence of the citizen science approach.

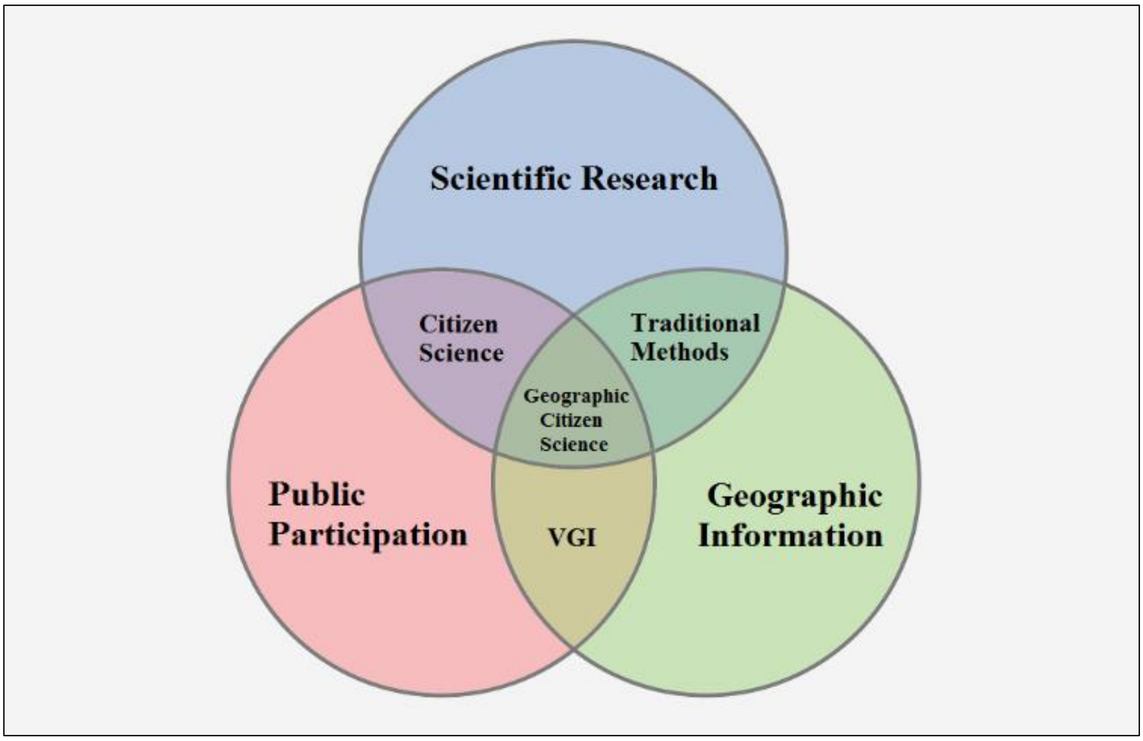


Fig.31: Thematic intersections between public participation, scientific research and geographic information (author's illustration, based on [NOD-14])

2.5.3. Usability in Modern Analyses

Some of the mentioned examples already indicate or confirm the value of volunteered geographic information and geographic citizen science, used in modern data analyses. The usability is depending on the research objective, the data contributors and the final users [HRD-13]. The significance of geographic information in everyday's life however, cannot be doubted. Tourists need navigation systems or maps, captains and pilots rely on navigation charts, pupils learn about Earth's variety with atlases and disaster response authorities need accurate information to organise fast and effective recovery efforts [GOO-08].

Speed, effectivity and saving of expenses are some of the major keywords for the dynamic development of crowdsourcing, VGI and citizen science over the last few years. Smaller initiatives with lower budgets are now more competitive than before, which increases the cultural and scientific diversity of a society. There are however some common drawbacks of these approaches, which include data quality and civic reservation about the reliability of information, that has been contributed by untrained volunteers. Till now, everyone trusted maps and charts, because mapmakers and cartographers were qualified experts in their professional area. Nowadays, with facilitated ways of creating and distributing maps over the Internet, data reliability and quality is often questionable.

As the author is well aware, this thesis tries to answer a lot of questions about soil sealing, flood extent, crowdsourcing and disaster management. One of the main objectives however, is the demonstration of the wide array of possibilities which CS and in particularly geographic citizen science have to offer. To avoid quality issues and ensure the usability of the analysed data, the assessment of accuracy and accompanying quality management will be integrated in the data preparation process.

The open access to information is not always feasible due to various reasons. Either there is no social demand for democratic science or there is a substantial lack of information. In his bachelor thesis, the author wrote about missing geospatial cropland information in Myanmar, which is one of the major requirements for assessing food security by humanitarian organisations. The paper, which was based on the results of this analysis, stated that the combination of remote sensing techniques and VGI is an "appropriate tool for gathering up-to-date land cover information of areas that are, for example, not easily accessible or restricted", although the agreements with various established land cover datasets were rather moderate [ALB-14].

Another indication for the usability of the data used in this master thesis, is the publication of a research paper, based on similar data by IIASA. Based on earlier generated FotoQuest data, Bayas et al. [BAY-16] focussed on the application of gamification and mobile technology for crowdsourcing in-situ data about land cover and land use. The following map [Fig.32] shows the original data from the first FotoQuest campaign and compares it to an established dataset. Red points represent the validated FQA points, green points represent the available LUCAS points. Back then, half of the collected information had to be excluded due to unsuitability, lack of quality or uncertainty [BAY-16] The results however confirmed the benefit of the approach, but also pointed out the moderate agreement with established datasets (70-80%). However, the project was considered a success and has been extended all over Europe. The data basis used in this thesis is a result of this spatially and temporally extended project.

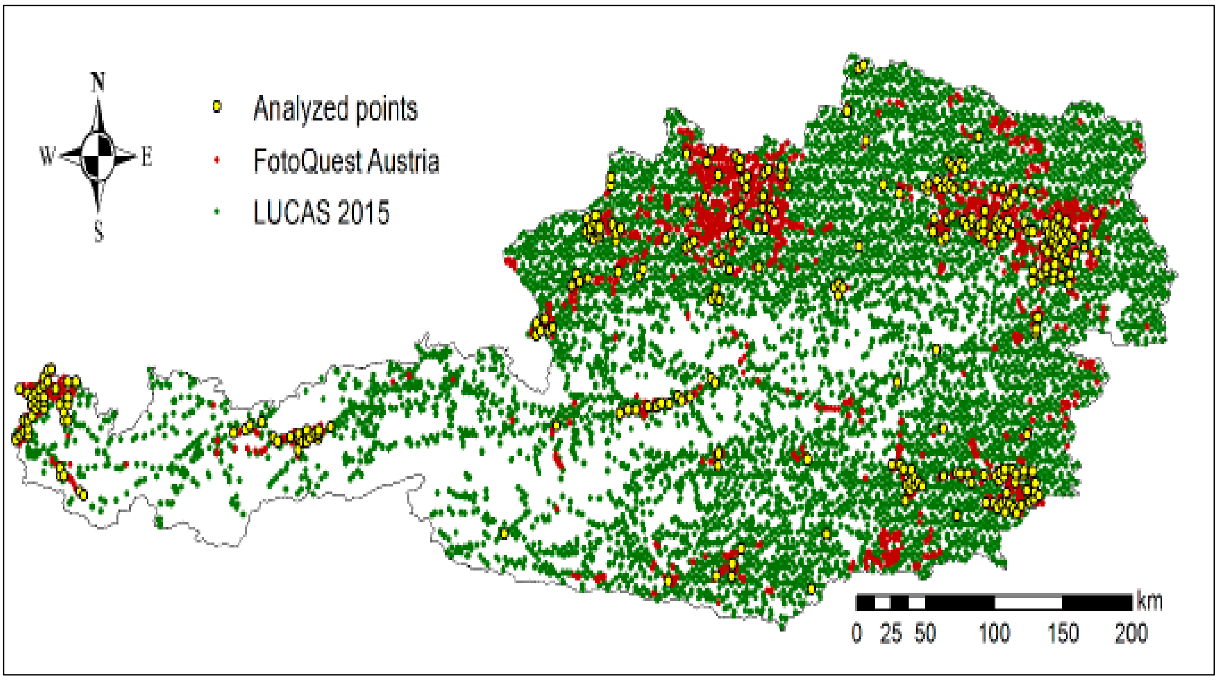


Fig.32: Original FQA data compared to the established LUCAS dataset from 2015 [BAY-16]

More on the two land cover classification initiatives FQA and LUCAS can be found in the corresponding chapters. This paragraph however, concludes the thematic introduction. The following chapters deal with the process of data collection, pre-processing, accuracy assessment, alternative methods, quality management and assurance as well as the required methodology and technology. Afterwards, the analysis procedures and the final results will be explained and presented.

3. Data Preparation and Methodology

After providing a fundamental theoretical framework about integral parts of this master thesis (concept definitions, disaster management and risk reduction, flood events, characteristics and appropriate protection, soil sealing and urbanisation, as well as crowdsourcing and its derivatives VGI or geographic citizen science), it is time for the methodical and practical part. This chapter includes the determination of the research area, the methods and structure of the data acquisition (FotoQuest Austria, LUCAS, flood extent, flood protection, elevation model and building density), the pre-processing and preparation of the data and a statistical evaluation (general statistics and accuracy assessment).

3.1. Determination of the Research Area

As the main title has already revealed, the main focus of this master thesis is on the anthropogenic influence on natural disasters alongside the Danube River in Austria. There are several reasons, why this area has been chosen, including geographical proximity, adequate topography, urbanisation and the availability of data.

~ *The first reason for choosing the research area was the proximity to the author's residential area in Vienna*. The author lives in Vienna for nearly ten years and has acquired some good local knowledge about the research area and its environment. This facilitates the basic understanding for the problems and challenges of the area. Whenever there are problems or incomprehensibilities within the data, additional information can be acquired through field trips, which could be scheduled at very short notice.

~ *The second reason was the fact, that the Danube is a watercourse in an urban area*. One of the main objectives of this master thesis is the sealing of the soil and the potential correlation to flood extents. Considering that, it would be quite handy if there is some kind of floodable watercourse near an urban or densely populated area, with a certain history or frequency of flooding, Thanks to the internationality of the Danube River and its importance for major industrial districts, the research area provides the perfect combination for investigating the quite specific research question. Additional information about urbanisation alongside the Danube is available in the corresponding chapter about the Danube River (2.4.3. Current Situation at the Danube River).

~ *The third reason was that the provided data was only available in Austria*. Thanks to a former collaboration with the IIASA during the author's bachelor study, it was planned from the start, that this institute would be the perfect partner for writing on this master thesis. One of their current projects was the FotoQuest Austria campaign and it was possible to receive some data. This information was good enough for investigating the author's research objectives. As the name of the project already indicated, the main focus of the FQA campaign was on the Austrian territory. This facilitated the search for an appropriate research area, which combined watercourses running through urbanised areas. With the help of IIASA's successor project FotoQuest Go, it will be possible to extend the research to other waterbodies and coastal areas all over Europe in the next few years.

~ *The fourth reason for choosing the research area was the denser distribution of data points alongside the Danube*. Probably the major reason for choosing the Danube River near Vienna was the apparent high point density within the FQA data. The following illustration (Fig.34) shows the concentration of data points in the Northern part of Upper Austria and alongside the Danube in Lower Austria. The higher concentration enabled the author to analyse more complex issues, with a higher chance of getting accurate results.

~ *The fifth reason was the higher degree of applicability and practicality for the society.* It may sound harsh, but research in densely populated areas may possibly be more "important" than in sparsely populated rural areas. The intensified research in large cities could have more positive consequences to a larger amount of people, because of the higher structural applicability of its results and the following adaption of official planning strategies and disaster management.

With these reasons and arguments in mind, the author first narrowed the research area down, from the whole of Austria to just the three states with direct connections to the Danube River: Upper Austria, Lower Austria and Vienna (Fig.33).

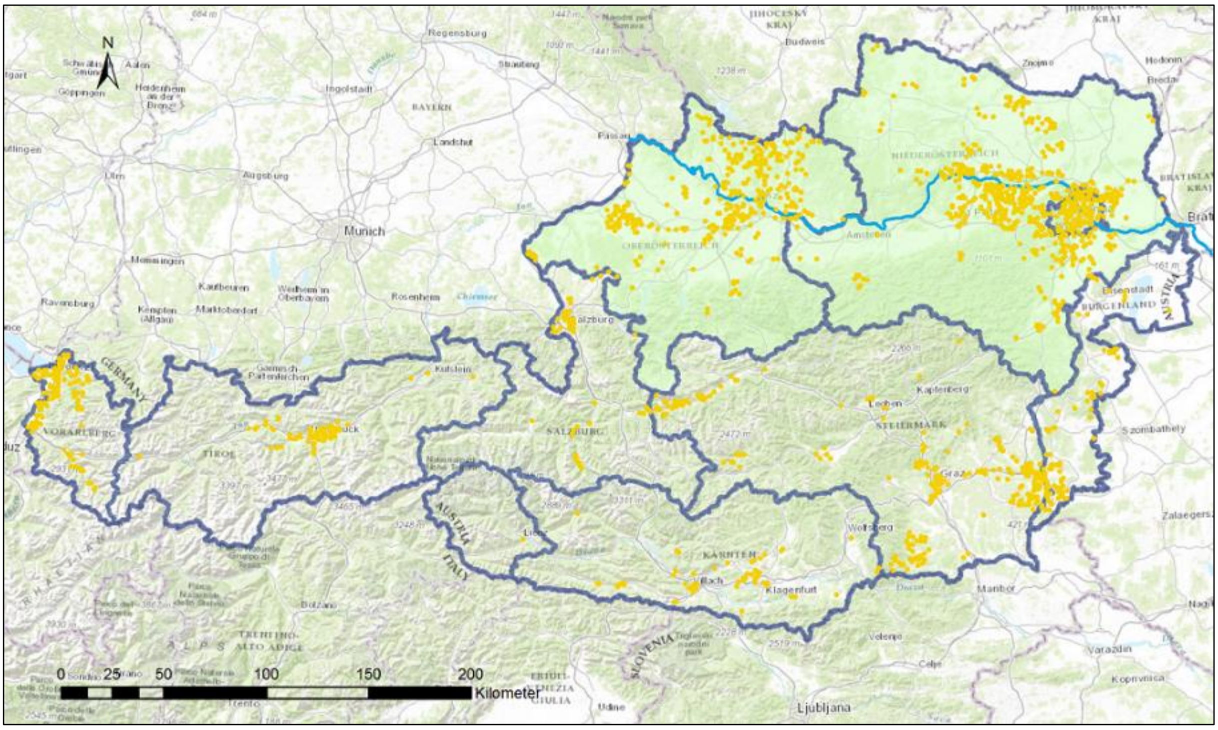


Fig.33: Course of Danube River in Austria and the affected provinces with yellow data points

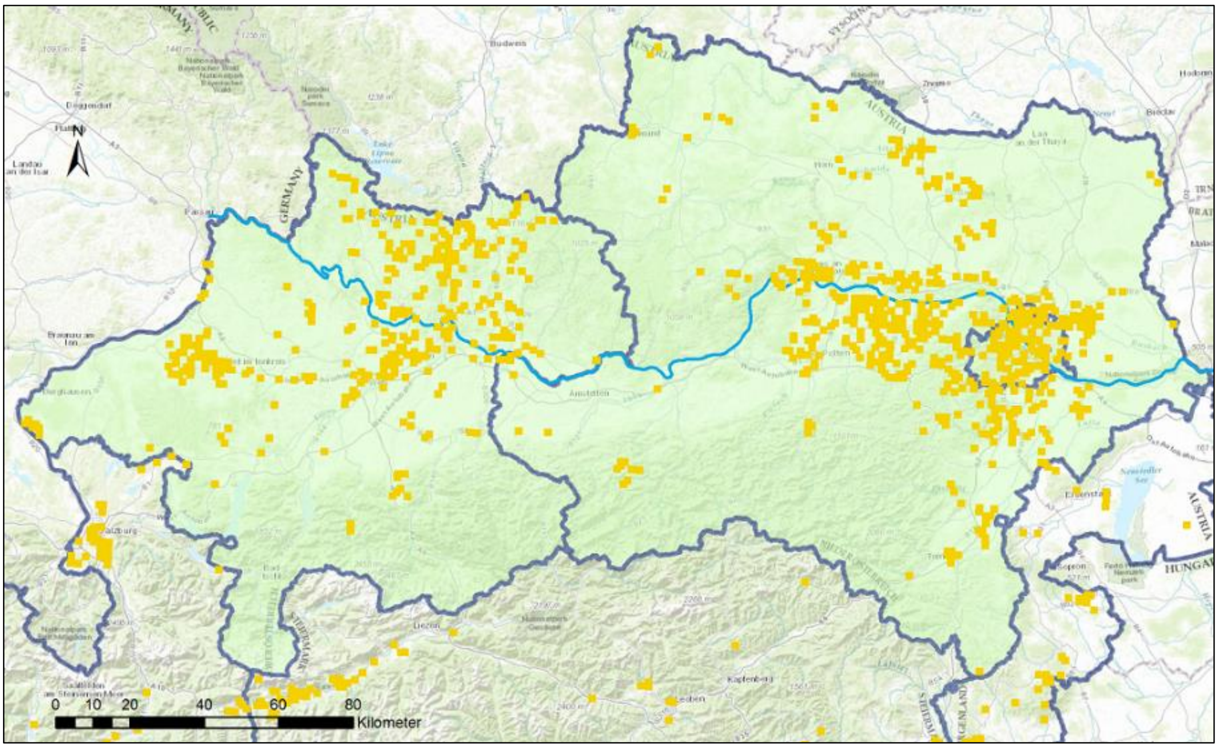


Fig.34: Distribution of collected validations in the affected Austrian provinces in 2015

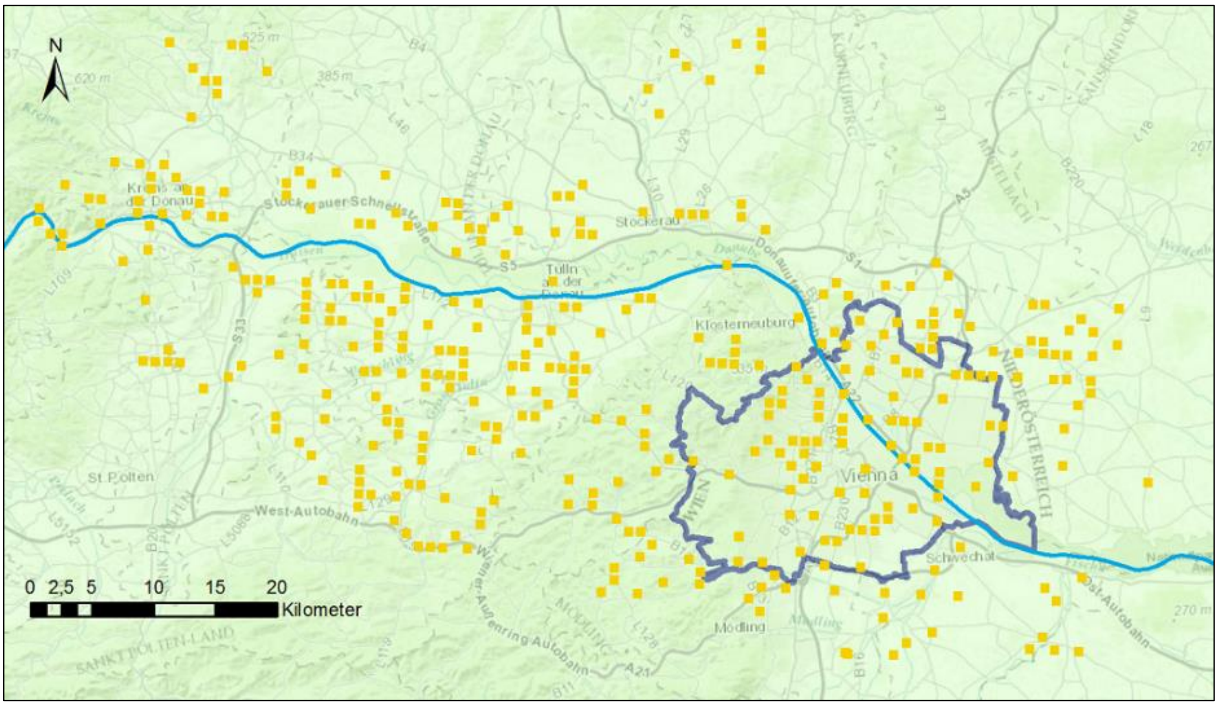


Fig.35: Data of the selected research area together with its surrounding environment in 2015

As mentioned before, there are two clusters of data points near the Danube, in the northern part of Upper Austria and in central Lower Austria. The previous chart (Fig.34) clearly depicts this situation and shows the higher data point density in and around the city of Vienna. The Austrian capital is not only one of Europe's leading centres for arts, crafts, music and science, but it is also one of the few European cities with more the one million inhabitants. With almost 1.9 million people living in an area the size of approx. 414 square kilometres, Vienna features a population density of more than 4500 people per square kilometre [STA-17] and can therefore be considered as a densely populated area.

All this made it the perfect testing ground for new and advanced scientific research methods, like crowdsourcing and citizen science projects. The final determined research area (Fig.35) eventually extends from Krems an der Donau in the West to Breitstetten in the East, and from Ernstbrunn in the North to Mödling in the South. This spans over an area of approximately 4000 square kilometres (90 x 50 kilometres) and includes many various landscapes, like sparsely populated areas in the Vienna Woods, agricultural land, small towns as well as the bustling city of Vienna.

Even within the boundaries of the Austrian capital, there are wide ranges of land cover types or land use prevailing. The following chart (Fig.36) illustrates the degree of soil sealing in the urban morphological zones (UMZ) of Vienna. Although the cartographic visualisation is improvable and the class sizes differ inconsistently, it is possible to detect, where the most sealed and impervious areas are within the city limits. The densely populated inner quarters are also the most sealed spaces (red colour). Areas which only show a low degree of sealed surfaces are either parts of the Vienna Woods, the Danube itself or parks and recreation areas. There are however many heavily sealed areas which are in close proximity to the Danube River and its canals.

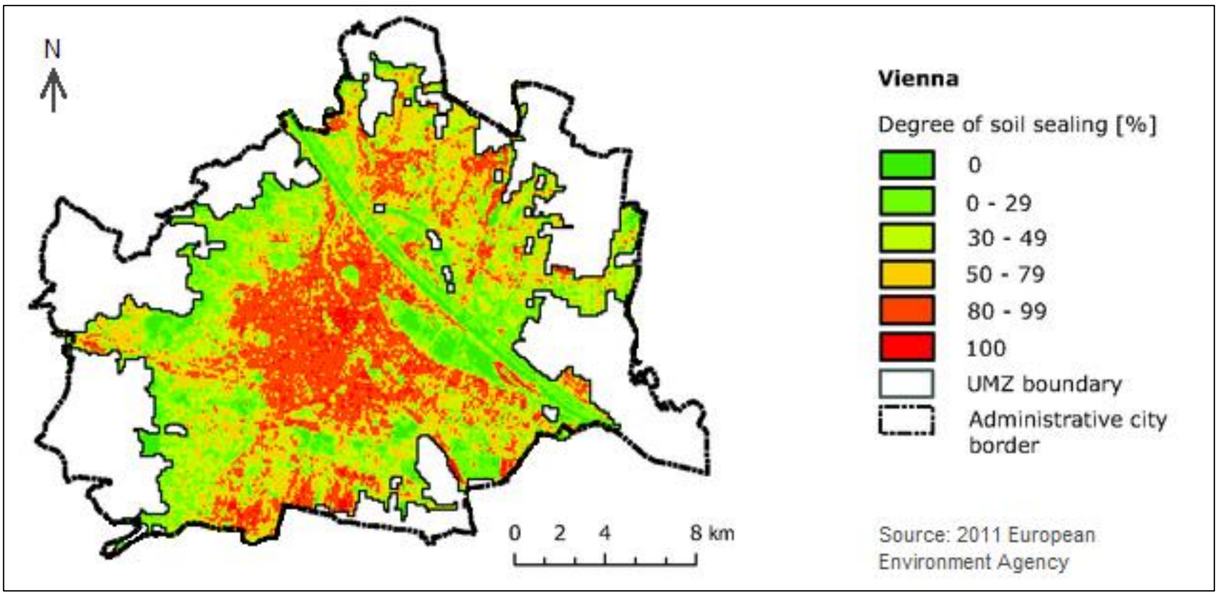


Fig.36: Degree of soil sealing in Vienna in 2011 based on the urban morphological zones [EEN-11]

3.2. Data Acquisition

The acquisition of data is always one of the integral parts of crowdsourcing projects or geographic citizen science campaigns. The ascertainment of specific research questions and areas, as well as the development of appropriate tools, are surely some of the most challenging components of such a scientific undertaking.

For these reasons, the focus of the next chapters is on the description of the data acquisition campaign FotoQuest Austria (conducted by IIASA) and comparable initiatives. Information about the development of the freely available application will be provided, as well as the individual stages of data contribution via untrained volunteers or users. The different available land cover classes (on which the final analysis is based on) and the land use classes will be discussed, followed by a short exemplary overview over all contributed data in tabular form.

The next subchapter focusses on the Land Use and Cover Area Samples survey (LUCAS), which was chosen as the ground-truth for the accuracy assessment. Although there are a number of established land cover datasets available, LUCAS and its classification scheme was used as a model during the design process of the FQA project. As a result of that, there are nearly identical class types or sample points, which may contribute positively to the data quality and accuracy.

In order to prove the point, that geographic citizen science data is capable of determining possible correlations between soil imperviousness and flood extents, there is a need for details about high water lines, the extent of past flood events and a reliable damage estimation. This challenging data search will focus on the major inundations in the Danube region in 2002 and 2013, because these two floods had been examined quite comprehensively by various official and commercial institutions. The Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT) is the federal administration for all waterways and inland shipping, and has conducted numerous investigations of recent flood events. The most frequent and accurate data is available for the flooding in 2013. The author fortunately managed to acquire data about flood discharge areas, the accretion and erosion of sediments and the effects of protective measures and damages.

The last subchapter of the data acquisition process discusses portable and permanent flood protection in the research area. These structural defence measures include permanent installations like flood barriers, dams or protective walls, and portable facilities like filled plastic tubes, sandbags or aqua fences. The data was kindly provided by the Federal Ministry for Transport, Innovation and Technology. This official administration has long-term experiences in collecting flood data, conducting projects in the area, collaborating with numerous relevant offices and even with predicting future developments or events and taking respective measures.

3.2.1. The IIASA FotoQuest Austria Project

The FotoQuest Austria Project (FQA) was initiated by the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria, and the internationally active Zurich Insurance Group. IIASA is a leading scientific research facility near Vienna and was founded in 1972 to investigate sustainable solutions for natural, human, social and technological systems [IIA-16]. The research programs include ecosystem services and global health, hydrological systems, global pollution and air quality, sustainable energy systems, as well as risk and resilience studies. The FQA project is part of the Ecosystems Services and Management Program and aims at German-speaking volunteers who want to participate in a scientific adventure and explore the countryside. The main objective is to collect geographic data for sustainable spatial planning, which could oppose the increasing land consumption and permanent soil cover. The biggest prevailing problem today is the inadequate supply of in-situ data in affected areas. Changes in land use or cover and the degree of urbanisation are therefore difficult to measure in a proper way. Scientists from IIASA tried to create a database for land cover and land use information, which exceeds the quality of other established databases in order to contribute to climate change research, resource management, flood risk reduction and spatial planning. The relevance and urgency of these specific research areas have already been described sufficiently in the thematic introduction.

The FQA project started in July 2015 (parallel to similar campaigns like Cropland Capture and Picture Pile) and is defined as one of the flagship citizen science projects done by the IIASA research team. However, in the corresponding chapter about crowdsourcing and its derivations, the differences between citizen science and volunteered geographic information (VGI) have been discussed as detailed as possible. The conclusion of this comparison was, that not all citizen science projects use geographic information, and not every VGI project can be considered as scientific. The FQA project on the other hand, comprises all three determining domains: scientific research, public participation and geographic information (Fig.31). Based on these definitions, FotoQuest Austria has to be declared as a *geographic citizen science* project, because of its intersection with all characteristic attributes and combination of the citizen science approach with volunteered geographic information.

In order to collect and analyse any geographically relevant information, some kind of scientific initiative had to be created. FQA is only the first in a series of similar campaigns and it is based on a freely available application, which can be used on personal computers, tablets or smartphones. The platform is currently just supported by the two operating systems Android and iOS. As already indicated, the FotoQuest app was programmed to suit the Geo-Wiki tool and certain web mapping services [BAY-16]. It also uses a programming interface for exchanging data in the background, without affecting the user interface of the tool, and open source databases for storing and maintaining the collected spatial information [BAY-16].

The collected data of the FQA project was processed and rectified at first, and then made available to the public via the Geo-Wiki online tool, which provides a number of established land cover datasets and applications or campaigns, together with their own products like hybrid global land cover maps [GWO-17]. Some of this information has already been used for the author's bachelor thesis on using volunteers for improving traditional land cover products.

The comprehensive information can be used for sustainable spatial planning or, as raw data, for this master thesis. The main profiteer is the ordering customer and financier Zurich Insurance Group, because the results will be analysed and implemented by the company's flood resilience alliance.

The following figure (Fig.37) shows the original FotoQuest logo and illustrates the initial focus on Austrian territory. With the implementation of FotoQuest Go, the initiative has been extended onto the whole European continent. In order to scientifically measure the impacts of human activities and climate change on the carbon sequestration of forests, soils and wetlands, reliable and frequently updated information is vital. Benefitting from the Pokémon Go hype, IIASA launched the new and extended FotoQuest Go application, which is usable all over Europe. The best and most accurate contributors can win material prizes as special incentives. One of the consequential effects of this project is the education of people about environmental problems and hazards, as well as raising the awareness on potential disasters and risks. The campaign shares some aspects with the Geocaching game, where people use GNSS to go on treasure hunts around the countryside. In case of FotoQuest Go however, participants do not find small gifts, but take pictures and answer questions about the surrounding environment.

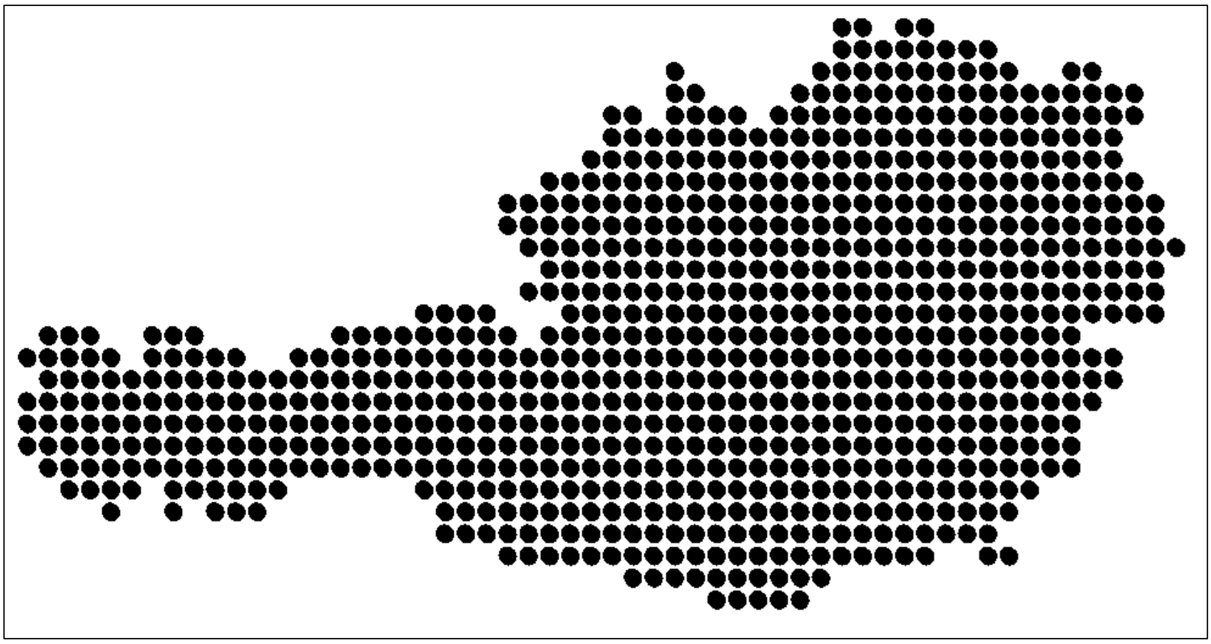


Fig.37: Initial FotoQuest Austria logo by IIASA [CSN-15]

The functionality and organisation of FQA and its succeeding projects will be discussed on the following pages. The first step for the player is to download the application on a personal computer or mobile device and register with a specifically created account. When using the programme for the first time, an overview map appears and shows selectable locations for potential validations in different colours. A screenshot of the application's user interface (in German) is illustrated in the following Figure (Fig.38). Blue points stand for places which have not been handled yet, yellow points marks the multiple validation of this location (up to five times) and red points symbolise that more than five people have already visited the location. The less a place has been validated, the more points can be earned. The maximum point value is 100 for an unvisited location. This should encourage the actual players to explore the country and visit new and unattended validation points more frequently. Figure 39 shows all the available quests in a closer local area with their respective points achievable. When the player is closer than 50 metres to the location, the quest can be started and the appropriate button appears.

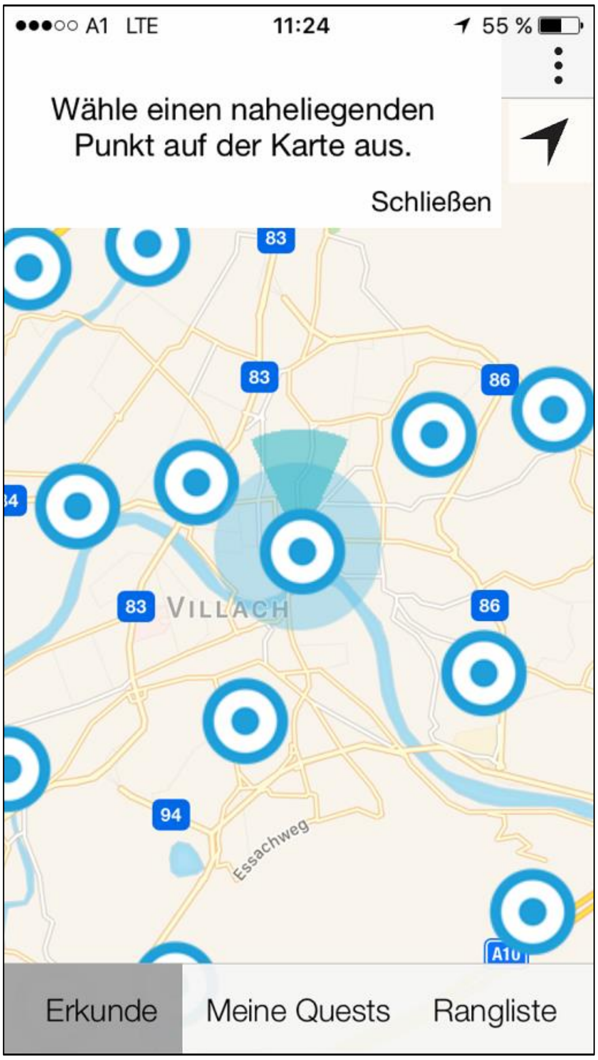


Fig.38: Location of available quests in the area

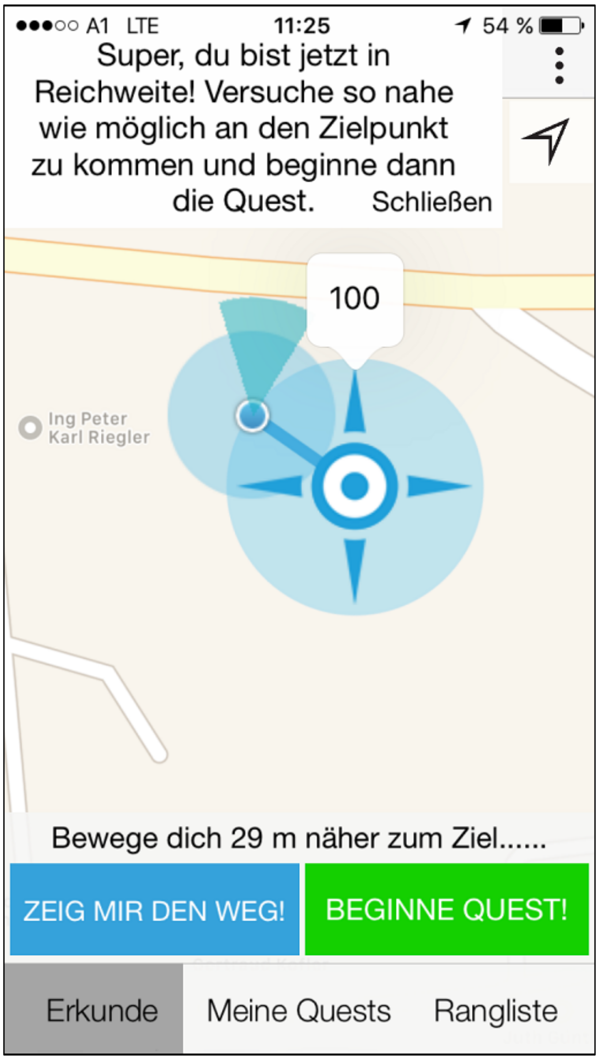


Fig.39: Quest in a closer local area with achievable points

When the quest has started, the player has to choose whether the point is accessible or not, and if the point is visible from the distance or not. If the location is inaccessible but visible, the player has to take a picture in the direction of the point in order to complete the quest [BAY-16]. If the player has access to the point and the spot is clearly visible, he or she will be asked to take pictures in all four directions North, East, South and West (Fig.40). The FQA application uses the built-in compass of the smartphone in order to ensure the right direction. If the arrow is green, the agreement is good enough and the picture can be taken (Fig.41). If the arrow turns red, the direction is too far off the correct direction and taking the picture is not possible. The line within the camera mask marks the horizon (Fig.41), which helps the player to photograph one third of the sky and two thirds of the prevailing landscape (if necessary). After all four directions are done, the player is asked to go five steps backwards and take a picture of the correct location of the spot on the ground.

Fig.40: Introduction for taking the right pictures

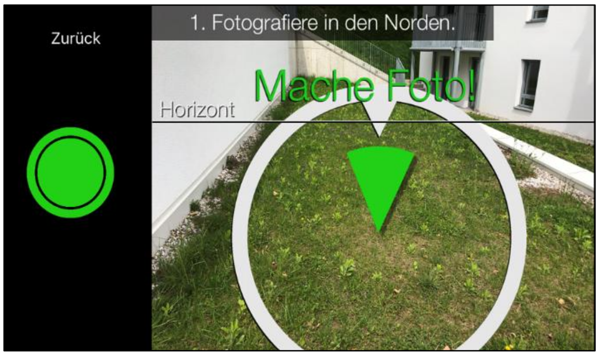


Fig.41: Player is asked to take a picture in northern direction

If all pictures were taken correctly, the player now has to classify the prevailing land cover and land use with the help of a number of predefined options. Based on the established LUCAS classification system, there are eight main categories: artificial landscape, cropland, woodland, shrubland, grass land, bare land and lichens, water areas and wetland. Every category has a number of subcategories, which are listed in the following table (Tab.2). Additionally to the given options, there is always the possibility of choosing an *unknown* land cover. After selecting the appropriate land cover type, the player is "asked to select a radius, which provides an indication of how homogeneous the land cover is around the point location" [BAY-16]. The intervals range from smaller than 1.5 metres, to smaller than 10 metres, smaller than 25 metres and bigger than 25 metres. The smaller the range is, the more heterogeneous is the surrounding environment. The larger the patch of the specific land cover is, the more homogeneous is the landscape. The next step is to select the appropriate class of land use. The categories were adopted from the LUCAS hierarchical nomenclature and comprise of 14 possible classes (Tab.3).

Land Cover	
Artificial Landscape	built-up areas
(A)	artificial non built-up areas
	other built-up areas
Cropland	cereals
(B)	root crop
	industrial crop
	dry pulses, vegetables, flowers
	fodder crop
	fruit trees
Woodland	broad-leaved woodland
(C)	coniferous woodland
	mixed woodland
Shrubland	sparse tree cover
(D)	shrubland without tree cover
Grassland	sparse tree cover
(E)	grassland without tree cover
	re-vegetated surfaces
Bare Land and Lichens	rocks and stone
(F)	sands
	lichens and moss
	other bare land
Water Areas	standing water body
(G)	running water body
	glacier, permanent ice
Wetlands	swamp
(H)	peat bog
	salt marsh

Tab.2: Classification of prevailing land cover (adapted translation from German interface)

Land Use
agriculture
forestry
aquaculture and fishing
mining and quarrying
production primary sector
energy production
industry and manufacturing
transport
water supply and waste treatment
construction
information service
recreation, leisure, sport
residential area
natural unnused area

Tab.3: Classification of prevailing land use

When all the classifications are done, the player can finish the quest and upload the validated data and pictures (Fig.42). This could either happen immediately after the validation or at a later time. The player receives a confirmation if the upload has been successful and if the information has been saved (Fig.43).

The next screen offers an overview about all completed quests and the collected points (Fig.44). All the achieved points are recorded in a ranking list, where the players can compare themselves to each other. These rankings are subsequently used in combination with the contributed data quality, to determine the final placement of the players and potential qualification for material or financial prizes.

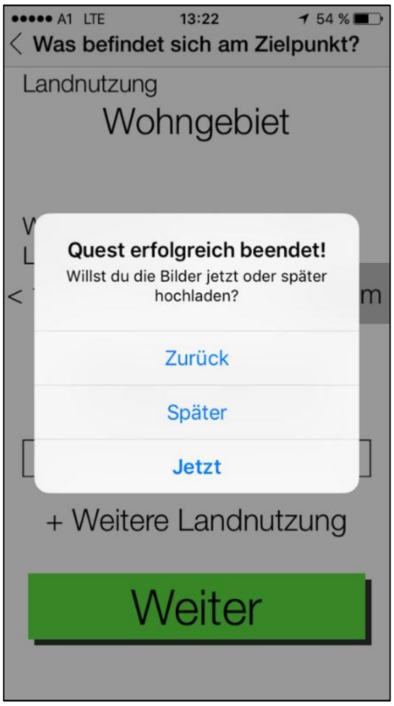


Fig.42: Finalisation of the quest

Fig.43: Successful upload of the data

Fig.44: Completed quests and points

Throughout the FotoQuest Austria campaign, additional information has been collected about the ways and methods the application could be improved. All these ideas were evaluated and implemented in the follow-up project, called FotoQuest Go (FQG). There are some differences between the first FQA campaign and the succeeding FQG. The main screen has been refreshed and updated. The icons and font style have been newly designed, but offer the same functions as before. One of the main differences between the two projects is, that the newer FQG does not offer a drop-down menu with all the predefined categories anymore, but asks the player a series of questions about possible land cover types and the user has to choose *yes* or *no*.

The most important difference however, is the expansion of the potential research area from Austria to the whole of Europe. This means that the range of possible players increases from 8.5 million in the country to more than 500 million on the whole continent.

The original FotoQuest Austria campaign started in July 2015 with the launch of the mobile application for Android and iOS. In order to introduce the project to the broader public in Austria, the campaign has been advertised in television and radio broadcasts. There were also strategically placed ads in newspapers and social media channels, always in combination with short demonstration videos and material or financial incentives (smartphones and tablet computers). The campaign formally stopped in September 2015 and therefore ran only for three months. Some players continued until December 2015 and collected an extra amount of information and doubled the originally intended running time. This is one example of the possibilities, citizen science projects can provide: (more or less) large amount of data in short period of time.

The basis data for this master thesis was extracted at the initial target date. More than 2300 quests were played in the course of three months. More than 200 individual users visited over 1800 unique locations. Around 12,000 pictures have been taken, which were mainly geo-tagged (the GPS-coordinates were recorded). An estimated three-quarters of all players used the Android platform on their mobile devices, around one quarter used the version available for Apple's iOS platform. Additional and more specific statistics follow in the corresponding chapter (3.3.1. General Statistics).

point_id	lucas_id	x	y	userid	fotoquest_item_id	platform	device_id	skip_reason	score	visible	timestamp	validation_id
728	42882682	9,564398	47,250709	8	935	8	123456789	reasonlikenotreachable/p	100	t	2015-07-22 13:42:18.405921+02	6532
746570	41283104	7,249058504	51,0146282	11646	2654	10|6|1.0|eed	eeda0c06598b0858		1	t	2015-11-07 14:46:28.487012+01	19215
728	42882682	9,564398	47,250709	2714	2185	7	65e15e222188738d		81	t	2015-09-18 17:43:32.762169+02	15679
728	42882682	9,564398	47,250709	2714	2185	7	65e15e222188738d		81	t	2015-09-18 17:43:32.762169+02	15681
728	42882682	9,564398	47,250709	2714	2185	7	65e15e222188738d		81	t	2015-09-18 17:43:32.762169+02	15683
746570	41283104	7,249058504	51,0146282	11646	2654	10|6|1.0|eed	eeda0c06598b0858		1	t	2015-11-07 14:46:28.487012+01	19216
728	42882682	9,564398	47,250709	2714	2185	7	65e15e222188738d		81	t	2015-09-18 17:43:32.762169+02	15680
746570	41283104	7,249058504	51,0146282	11647	2653	10|7|1.0|a75	a75ba273858b6353		2	t	2015-11-07 14:38:58.258039+01	19208
728	42882682	9,564398	47,250709	2714	2185	7	65e15e222188738d		81	t	2015-09-18 17:43:32.762169+02	15682
728	42882682	9,564398	47,250709	10960	2184	5|8.4.1|1.03|	7EAB5253-3F05-4DB3-A25	noskip	90	t	2015-09-18 17:40:37.885187+02	15673
746570	41283104	7,249058504	51,0146282	11647	2653	10|7|1.0|a75	a75ba273858b6353		2	t	2015-11-07 14:38:58.258039+01	19209
728	42882682	9,564398	47,250709	10960	2184	5|8.4.1|1.03|	7EAB5253-3F05-4DB3-A25	noskip	90	t	2015-09-18 17:40:37.885187+02	15672
747967	41642770	7,895985281	48,0234619	19	2652	10|6|1.0|123	123e9308788965		2	t	2015-11-06 17:51:47.363694+01	19201
728	42882682	9,564398	47,250709	10960	2184	5|8.4.1|1.03|	7EAB5253-3F05-4DB3-A25	noskip	90	t	2015-09-18 17:40:37.885187+02	15674
728	42882682	9,564398	47,250709	10960	2184	5|8.4.1|1.03|	7EAB5253-3F05-4DB3-A25	noskip	90	t	2015-09-18 17:40:37.885187+02	15676
728	42882682	9,564398	47,250709	10960	2184	5|8.4.1|1.03|	7EAB5253-3F05-4DB3-A25	noskip	90	t	2015-09-18 17:40:37.885187+02	15675
747967	41642770	7,895985281	48,0234619	19	2652	10|6|1.0|123	123e9308788965		2	t	2015-11-06 17:51:47.363694+01	19202
731	42902680	9,590939	47,232809	2714	2182	7	65e15e222188738d	Hindernis	81	f	2015-09-18 17:16:14.711132+02	15660
728	42882682	9,564398	47,250709	8	935	8	123456789	reasonlikenotreachable/p	100	t	2015-07-22 13:42:18.405921+02	6531
731	42902680	9,590939	47,232809	2714	2182	7	65e15e222188738d	Hindernis	81	f	2015-09-18 17:16:14.711132+02	15661
731	42902680	9,590939	47,232809	2714	2182	7	65e15e222188738d	Hindernis	81	f	2015-09-18 17:16:14.711132+02	15659
731	42902680	9,590939	47,232809	2714	2182	7	65e15e222188738d	Hindernis	81	f	2015-09-18 17:16:14.711132+02	15662
731	42902680	9,590939	47,232809	2714	2182	7	65e15e222188738d	Hindernis	81	f	2015-09-18 17:16:14.711132+02	15663
731	42902680	9,590939	47,232809	2714	2182	7	65e15e222188738d	Hindernis	81	f	2015-09-18 17:16:14.711132+02	15664

Tab.4: Exemplary section of the provided FotoQuest Austria data

The previous table (Tab.4) summarises some exemplary sections of the collected data by the FQA campaign. It offers information, amongst others, about the location, date and time the user was on his quest, potential obstacles and reasons why a certain sample point had to be skipped, if the location was visible or not and which kind of smartphone the corresponding volunteer used. Some of the data fields are listed and explained on the following page:

~ The *sample point ID* offers its own individual identification number for every potential point. A lot of the collected FQA data points share the same coordinates with the established LUCAS dataset.

They were chosen on purpose to provide a comparability between crowdsourced data and information collected by experts. Additional information and the corresponding analyses will follow in the course of this master thesis.

- *X and Y coordinate*s describe the actual location of the validated sample point.
- The *User ID* points out all the individual users and is used for quality checks and the awarding of financial and material prizes for the best and most accurate contributors.
- Information about *Platform* and *Device ID* provides details about the contributor's mobile device.
- *Skip Reason* defines the reason, why a point could not be validated properly and has been skipped. This includes bad GPS-connectivity, private properties or inaccessibility.
- Just like mentioned before, the *Score* is one of the foundations for the awarding of monetary and material prizes. The higher the score for one validation is, the less volunteers have already validated the individual sample point.
- *Visible* describes the visibility of a point and provides details if the sample point is accessible and assessable or not. The two values are true ("t") and false ("f").
- The *Timestamp* offers information about the exact time, the location was validated.
- *Comment*s provide additional information about the validated sample point, whether it was accessible during the visit or not, or if there is an immediate change in land cover detectable.
- *Legend Item ID* and *Legend Item Name* are two of the most important columns of the dataset, because they comprise of the information about land cover and the classification scheme.
- The values of *Radius* provide information about the accessibility of a certain location and from how far the ground pictures had to be taken or the land cover class had to be chosen.
- *Media Item Name* explains from which project the information comes from (e.g. FotoQuest Austria).
- Just like mentioned before, every picture is stored at one of the IIASA's online servers. With the help of the listed *Photo URL*, it is possible to check individual points and change a questionable classification if necessary. This feature was used for the assessment of the quality management and issues.
- The field *photo transaction* provides information of the time the pictures had been uploaded.

3.2.2. LUCAS – Land Use and Cover Area Samples

Before the FotoQuest Austria data is intersected with information about the 2013 flood extent or discharge area, the quality of the collected data has to be guaranteed. If volunteers contribute to scientific research, there is always the fear of substandard data quality. Their missing educational background and training can pose a risk to the validity of the results. Most of the reasons and arguments were discussed in the chapter about crowdsourcing and geographic citizen science.

In order to be sure, that the present data is reliable and as accurate as possible (and more or less usable), an accuracy assessment is one of the most important parts of the data processing. This includes the comparison of two datasets, in order to assess the crowdsourced classification's validity and accuracy. The first dataset are the results of the FQA campaign, and the second one is an established land cover or land use dataset, which serves as the ground-truth.

As stated before in the corresponding chapter about soil sealing in a time of urbanisation (*2.4.4. Comparable Research*), there are various established land cover datasets currently available for Europe and other regions (e.g. GLC, GlobCover, Copernicus, MODIS, CORINE etc.). One of them is the LUCAS dataset, conducted by the European Union (EU-28). The acronym stands for *Land Use / Cover Area frame Survey* and executing operators state that the collected data provides perfectly "harmonised information for studying a range of socioenvironmental challenges, such as land take, soil degradation or biodiversity" [LUC-16] for EU member countries. It is considered to be probably the only official dataset with in-situ data for the EU's territory [BAY-16].

The first campaign of its kind was held in 2001, in order to provide an overview over the predictable crop yield for the European Commission, based on a reduced number of participating member countries [LUC-16]. The remarkable success, the increasing amount of collected information and the frequent use by the governments and decision makers, led to a reconfiguration of the project and its methodology. From now on, a three-year interval was implemented and the main focus was on land use and land cover instead of agricultural issues [LUC-16]. The new character of the survey was introduced in 2006 and from then on, every survey is carried out in two steps. In the first phase there are systematic samples, with points spaced two kilometres apart from each other all over the European territory, which lead to more than 1.1 million individual points [BAY-16]. Each of these generated points is then photo-interpreted by experts and classified into eight predefined classes, which are very similar to the classification system, used by the FQA campaign (Tab.2). The following Table (Tab.5) shows these classes, used by the LUCAS land cover classification system. This system has been adapted and extended over the past one and a half decades, in order to become comparable to other international classification systems (e.g. United Nation's FAO and other European information systems) [LUC-16].

Land cover			
A00	ARTIFICIAL LAND	A10	Built-up areas
		A20	Artificial non built-up areas
B00	CROPLAND	B10	Cereals
		B20	Root crops
		B30	Non-permanent industrial crops
		B40	Dry pulses, vegetables and flowers
		B50	Fodder crops (mainly leguminous)
		B70	Permanent crops: fruit trees
		B80	Other permanent crops
C00	WOODLAND	C10	Broad-leaved woodland
		C20	Coniferous woodland
		C30	Mixed woodland
D00	SHRUBLAND	D10	Shrubland with sparse tree cover
		D20	Shrubland without tree cover
E00	GRASSLAND	E10	Grassland with sparse tree/shrub cover
		E20	Grassland without tree/shrub cover
		E30	Spontaneously re-vegetated surfaces
F00	BARE LAND AND LICHENS/MOSS	F10	Rocks and stones
		F20	Sand
		F30	Lichens and moss
		F40	Other bare soil
G00	WATER AREAS	G10	Inland water bodies
		G20	Inland running water
		G30	Coastal water bodies
		G50	Glaciers, permanent snow
H00	WETLANDS	H10	Inland wetlands
		H20	Coastal wetlands

Tab.5: LUCAS – Land cover classification system [LUC-16]

As it is clearly visible in the table (Tab.5), most of the different classes and subclasses are nearly identical to those, used during the FotoQuest Austria campaign. The IIASA project is therefore deeply interwoven with the LUCAS survey and wants to maintain an unlimited comparability and mutual utilisation of standardised methodologies and data collection processes. With a view to simplify the technical procedures and a horizontal hierarchy, a lot of subclasses have been summarised and generalised throughout the course of the campaign.

However, the second phase of this land cover survey comprises of a derivation of fewer sample points, which are then visited and evaluated by trained experts (field surveyors) in the course of a field survey [BAY-16]. “The stratified sample is selected independently in each NUTS level 2 region fixing precision targets for the estimates of the main land cover classes” [LUC-16]. The threshold of accessibility in this field study varies between 1000 and 1500 metres. When the surveyor arrives at the sample point, he or she has to follow a set of predefined protocols and written instructions, to ensure the consistency of the collected data. These tasks are also quite similar to the FotoQuest Austria project: every sample point has to be photographed in all four cardinal directions and on the ground [BAY-16]. Afterwards, the surveyor has to take notes about the prevailing land cover and land use on the sample point and the proximate surrounding, based on the mentioned hierarchical classification systems. In addition to that, the surveyor has to walk alongside a 250 metre long transect in an easterly direction, in order to register any given change in land cover and land use [BAY-16]. If necessary, even soil samples had to be taken in ten percent of all visited points. These samples can provide information about soil types and textures, sequestrated organic carbon and even trends in soil degradation [LUC-16]. All these necessary instructions, explanations and optional examples for the smooth conduct of the project were summarised by Eurostat in a 74-page long technical reference document [EUS-13].

This extensive field work led to an amount of more than 273,000 samples all over Europe, which have been surveyed by 750 experts in the year 2015 [BAY-16]. Around 8800 sample points were surveyed on Austrian territory in the course of the 2015 survey of the LUCAS project, using the mentioned classification system (Tab.5) [BAY-16]. The LUCAS land use classification scheme however (Tab.6), uses only 14 categories, which are quite similar to the 15 FQA land use classes (Tab.3).

Land Use	
U110	agriculture
U120	forestry
U130	aquaculture and fishing
U140	mining and quarrying
U210	energy production
U220	industry and manufacturing
U310	transport, communication networks, storage, protective works
U320	water and waste treatment
U330	construction
U340	commerce, finance, business
U350	community services
U360	recreation, leisure, sport
U370	residential
U400	unused

Tab.6: LUCAS – Land use classification theme [LUC-16]

When choosing an established dataset as a ground-truth in order to assess the accuracy of a crowdsourced dataset, it is important to know how accurate the reference data actually is. Comparing FQA data with the results of LUCAS is only advisable, when the quality of the latter data is sufficient and reliable itself. There are a number of security mechanisms installed, to ensure the demanded data quality. During the collection campaign, an automated quality test checked the contributed information for completeness and consistency [LUC-16]. When the data in the field has been validated and transmitted to the respective data storage, every submitted sample was reviewed and checked by trained experts in the regional or central offices [LUC-16]. After these two steps of quality assurance, an independent controller checks all the contributed information again and:

~ controls the accuracy and adherence of all LUCAS requirements for 36% of the samples and
~ checks the first 20% of validated points of each surveyor for systematic errors [LUC-16].

The main reason why the author chose LUCAS as the comparable ground-truth in this thesis is obviously the affiliation between the two compared datasets and the similarity between their respective land cover classification schemes. Major parts of the FQA campaign's structure are intentional derivatives of the Land Use and Cover Area Samples. This facilitates the basic understanding and the subsequent analysis in the course of this master thesis. Other land cover datasets may have more contributed samples, but their classification scheme or original assignment differs more from the crowdsourced FQA campaign.

Another reason for choosing LUCAS as a comparable ground-truth is the fact, that it is an established dataset, conducted by a highly respected supranational organisation. The combination of experts working for the European Union, the European statistical office Eurostat and the Joint Research Centre (JRC), ensures a perfect inter-comparability with other databases, reliable scientific working methods and standardised and harmonised surveying methodologies.

There are however other comparable datasets available, which have been mentioned before in this thesis. These include, amongst other examples, the Climate Change Initiative (CCI), the Global Land Cover Network (GLC), Globcover, GeoCover, Copernicus, MODIS and CORINE.

Today there are a number of published studies, which also use the LUCAS dataset for their research and as a basis for their analyses or estimations. Gallego [GAL-08] points out the scientific usability of the LUCAS dataset for European crop area estimations, whereas Ballabio [BLL-16] maps the physical properties of European topsoil on a continental scale for instance. Other exemplary uses for the LUCAS data are the estimation of soil organic carbon content in the European Union (key word: soil samples) [PAN-13] or the estimation of soil erodibility, using the point survey data [PAN-12]. The LUCAS dataset is also used for numerous economic, environmental and social projects within the European Commission. These include the Directorate-Generals for Agriculture and Rural Development, Environment, Enterprise and Industry or for Climate Action [LUC-16]. Even the European Environment Agency (EEA) uses this data as their base data for analysing climate change and biodiversity.

More about the actual comparison of the FotoQuest Austria campaign with LUCAS data and the quality assurance, follows in the corresponding chapter about accuracy assessment and quality management (3.3.2. Accuracy Assessment with LUCAS Data).

3.2.3. Flood Extent or Flood Discharge Areas

Finding precise and detailed data about the flood extent and damage alongside the Danube River was a challenging task. Although there is various research of this area available about flood events, floodplains and damages, a lot of the information was quite vague, rough or on a small-scale. Some studies focused on major floods in 2002 and 2013, but basis geodata (e.g. flood attack lines) is rarely accessible for the research area. There was a lot of information available about flood flow rates, risk zones, flood depths, retention areas and flood attack lines. This data, provided by the Danube Reference Data and Services Infrastructure (DRDSI), focussed unfortunately only on rivers and other watercourses in the state of Upper Austria, where the data density of the geographic citizen science project FotoQuest Austria was too low for further investigations.

As mentioned above, some studies were focussing on the Danube floods of 2002 and 2013. Even though these two flood events were eleven years apart, they are definitely connected to each other, due to similar water levels and precipitation quantities. As a consequence from the 2002 inundations, a number of initiatives and projects were conducted, which focussed on the mitigation of and preparation for the next flood event. A comprehensive survey of the weaknesses and weak points of flood protection and response in vulnerable areas led to a series of construction measures, in order to improve permanent and portable flood protection. These initiatives and structural improvements alongside the Danube and its numerous inflows seemed to be a success, proven only eleven years later.

In May and June 2013, a complex system of low pressure areas formed in the eastern Mediterranean, following a long winter and a slightly delayed wet spring. The already saturated soil contributed significantly to the extent and spreading speed of the following inundation [LFW-15] in seven European countries (Austria, Czech Republic, Germany, Hungary, Poland, Slovakia and marginally in Switzerland). In addition to that, the amount of precipitation was way above-average. The following figure (Fig.45) illustrates the daily and total sums of rainfall in Austria between May 29th and June 4th 2013. On average, the daily precipitation was between 90 and 175 mm, with the Tyrolean village of Niederndorferberg leading the record with a daily total of 175 mm [LFW-15].

The total amount of precipitation for the duration of the event was up to 365 mm. The large amount of rainfall in this short period of time only happens once in 500 years according to official statistics [LFW-15]. The most severely affected areas were Vorarlberg, Tyrol, Salzburg, Upper Austria, Styria and Lower Austria. In Germany, most of the rainfall concentrated in southern and eastern parts of the country (e.g. Bavaria, Baden-Württemberg, Saxony, etc.). This resulted in an increased precipitation quantity in the catchment area of the Danube River, which is also depicted in the following figure (Fig.59). Nearly all tributaries of the Danube swelled up and intensified the effect on the river even more.

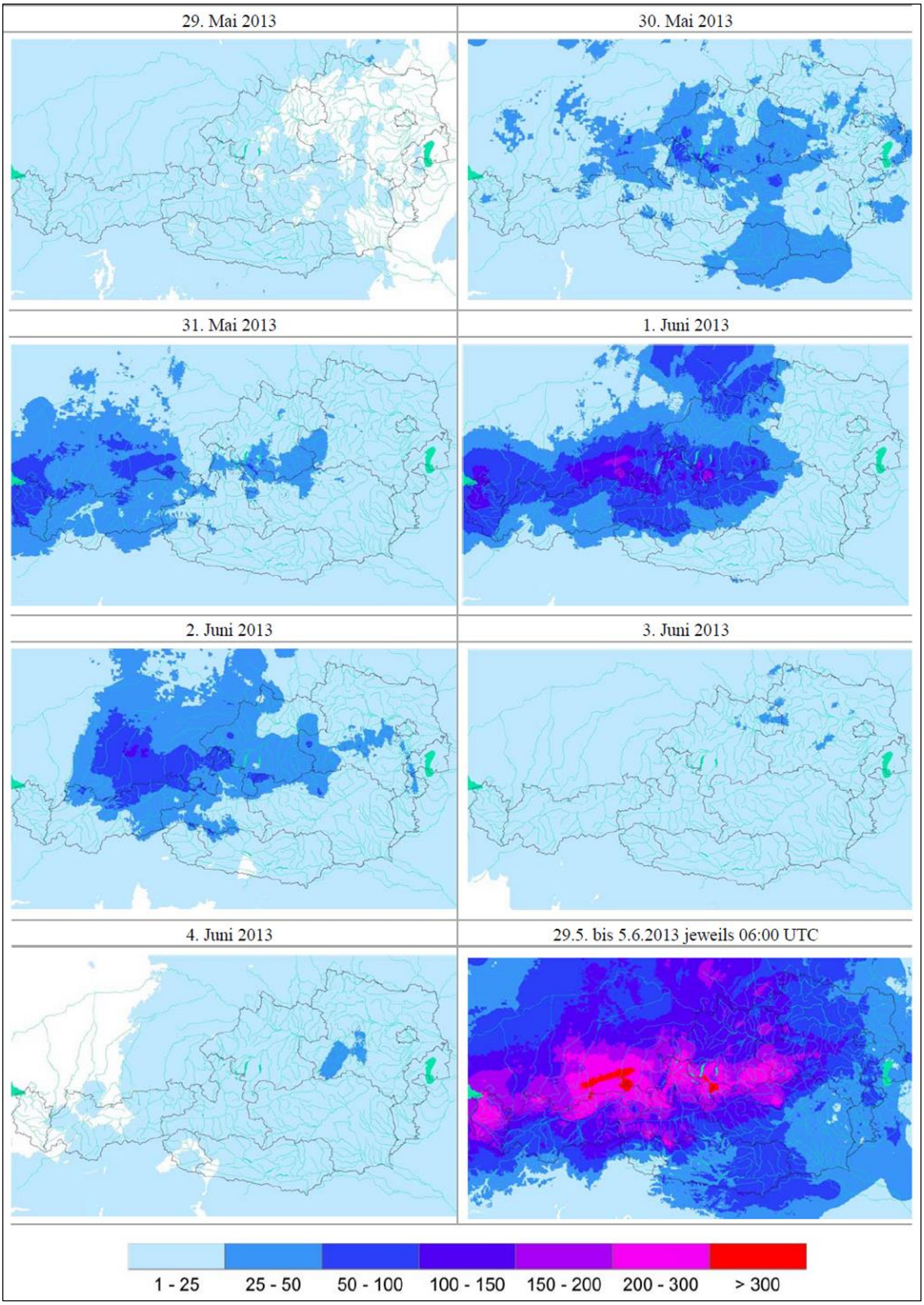


Fig.45: Daily sums and total sum of precipitation in Austria between May 29th and June 4th 2013 [LFW-15]

The combination of saturated soils, large amount of precipitation and the short period of time of the whole event resulted in a flooding with an annuality of 300 years. This value means that such an event is quite rare and happens only once in 300 years, according to recorded statistics. In today’s official spatial planning and disaster management, this information is usually used to examine past events and try to predict the extent of floods for different annualities. In most of the cases, participating agencies or surveyors focus on events, which statistically happen every 30, 100 or 300 years. These extents are designated as HQ30, HQ100 and HQ300, because high water is abbreviated with “H” and the flow quantity with “Q”. There are also other rates available, but not as frequent as the three rates mentioned.

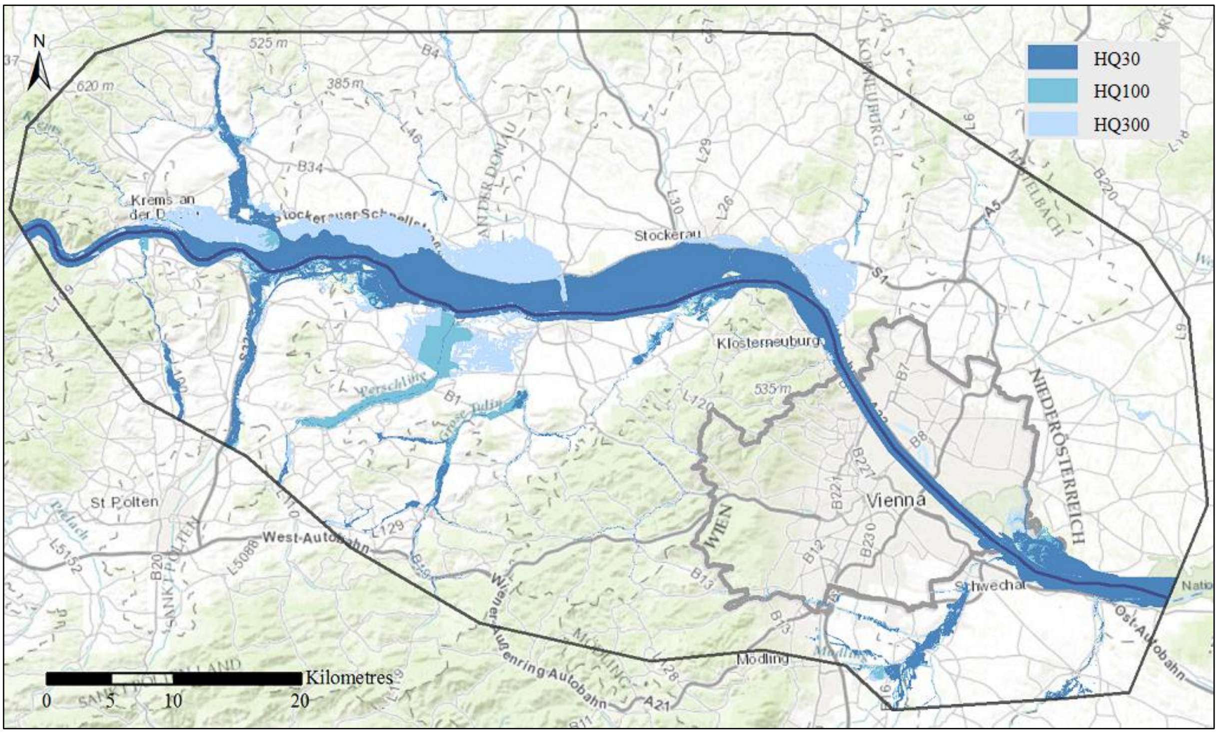


Fig.46: Predicted flood areas for 30-, 100- and 300-year floods in the research area

The map above (Fig.46) shows these predicted flood areas for inundations with an annuality of 30, 100 and 300 years in the research area. In case of a HQ30 event (dark blue), large parts north of the Danube will be flooded between Krems and Stockerau. Looking at satellite imagery however, it is obvious that these areas are mostly undeveloped or uncultivated and a flood would not cause severe damage. It appears that policy makers are well aware of the threat of a HQ30 and have therefore excluded these areas from further urban development. Anyway, the extent of a HQ100 differs not that much from a HQ30. Looking at the tables with all potential water levels, the differences are only between 50 and 70 centimetres and can be restrained by the same flood protection measures most of the time. There are only a few less altered tributaries which would overflow during a HQ100. In case of a HQ300 however, large areas which are developed and cultivated would additionally be inundated. This would include towns and villages like Korneuburg, Stockerau, Absdorf, Zwentendorf and Krems. Thousands of people would be at risk and large areas of vulnerable agricultural land and forests could be destroyed. Interestingly enough, the city of Vienna only shows flooded areas around the Danube Island (*Donauinsel)*, but nowhere else. This matter of fact indicates high quality flood protection (mobile or permanent) and sufficient urban planning strategies in this densely populated area.

All this information leads to the questions, whether all this predicted areas were actually inundated in the 2013 floods and how the distribution of water was influenced by the degree of sealed surfaces. This information can be retrieved by investigating the actual flood discharge areas of the Danube River.

As already mentioned before, the 2002 flooding in Central Europe caused severe economic, environmental and social damages in the affected regions. These events required extensive investigations in potential causes, reinforcing factors, social and structural neglect, disaster recovery situations, bad spatial planning policies or insufficient protection measures. Official bodies like the Federal Ministry for Transport, Innovation and Technology (BMVIT) or the provincial governments of Upper and Lower Austria invested time, money and effort into hydrological and structural analyses of this catastrophe. New methods and technologies were developed, to improve all four phases of the disaster management circle (response, recovery and restoring order, mitigation and preparation for the next event). Because of this efforts, all sorts of economic, environmental and social damages were reduced during the substantial flooding in 2013. This is rather astonishing, considering the fact that the latter catastrophe had higher flow rates, water levels and partly exceeded even the HQ300 predictions [LFW-15]. After this event, the same official bodies conducted new projects and investigations on how effective the implemented measures were and where there is still room for improvement.

Some results of these investigations will be used for the final analysis in this master thesis about potential correlations between soil sealing and flood extent. For instance, one of the most important datasets contains the flood discharge areas of the Danube River during the inundations in 2013. These polygonal representations cover every location which has been flooded by the overflowing Danube. The following map (Fig.47) illustrates these flood discharge areas and the two different methods, which were used to receive the data. The first thing that is noticeable, is the fact that the major parts of the inundations in 2013 happened on the northern river bank, between Krems and Stockerau. Only a few areas on the southern river bank around Klosterneuburg and Schwechat show slight gluts. Possible reasons like topography, river training measurers or permanent flood protection systems will be discussed in the following chapters.

However, it is also depicted in the following map (Fig.47), that there are two distinctive classes of flooded areas. This is a result of two different methods used for modelling the specific discharge areas (interpretation of aerial imagery and additional in-situ surveys).

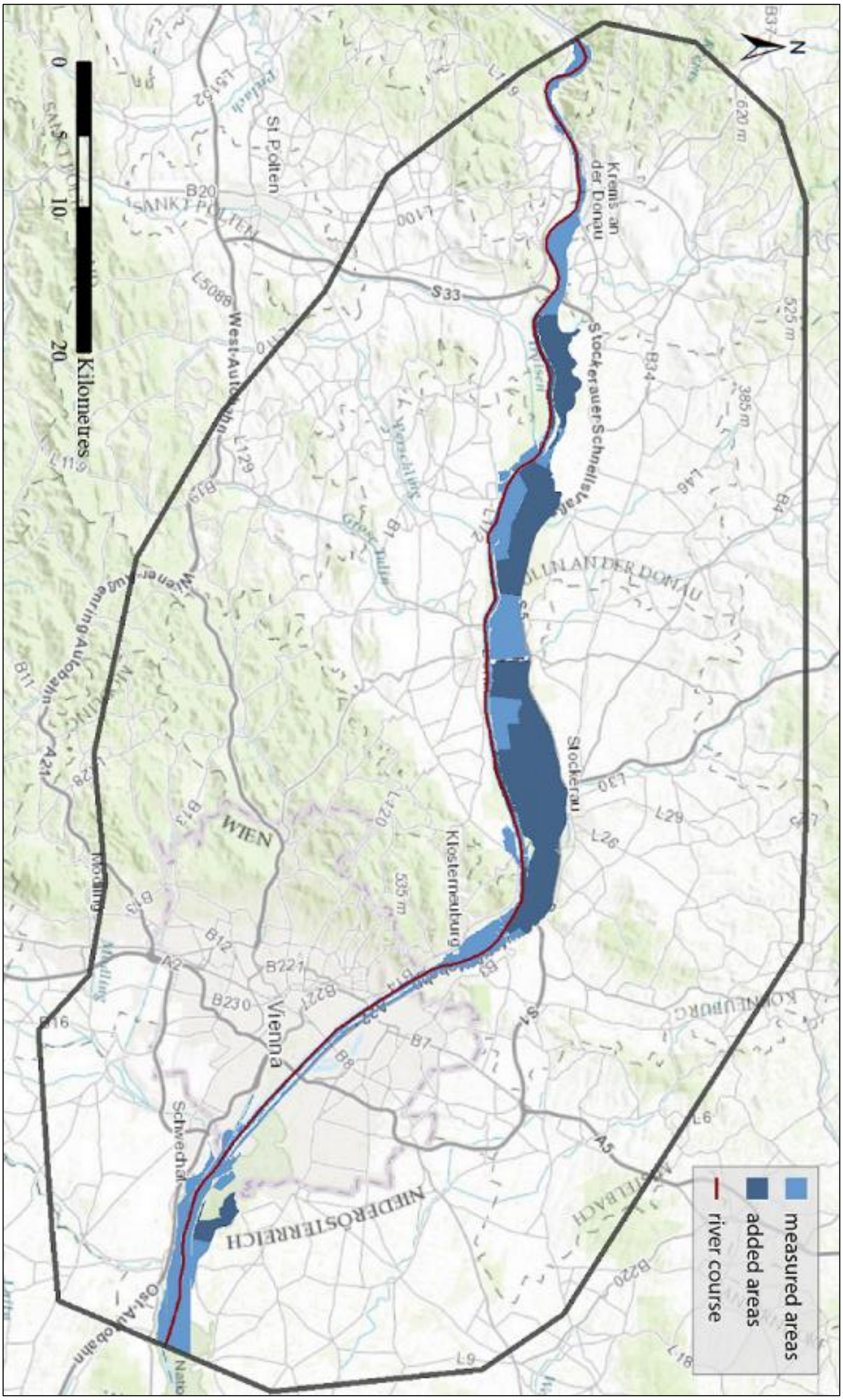


Fig.47: Flood discharge areas derived from aerial photography (measured) and from external sources (added) in 2013

The first and more traditional method is based on the interpretation of georeferenced aerial photographs or orthophotos. Right after the initial flood event, private companies and the Austrian armed forces were commissioned to take pictures of the inundated areas with high resolution cameras and according to a predefined pattern. The photographs then were orthorectified and transmitted to the corresponding agencies (e.g. BOKU, BMVIT etc.). The large amount of data was processed in a GIS environment and formed the basis for further digital interpretation. Every inundated area was covered with polygons, representing the official flood discharge areas. The following picture shows a comparison of the digitally evaluated flood discharge areas on the left and the actual orthophoto on the right (Fig.48). This work step was carried out for every affected region alongside the Danube and its tributaries in Austria.

Fig.48: Comparison of geodata evaluation and aerial photograph of Greifenstein in 2013 ([VIT-15], edited)

The other method of receiving the required data is based on additional external data. This external contributions include field surveys, bridge sensors, electricity supplier (e.g. Verbund) and measurements of local authorities or disaster relief teams. All these gathered information was checked for data quality, added to the GIS project and adapted by trained experts. The results for the research area can be seen in the previous map (Fig.47). Inundated areas which were measured from orthophotos are coloured light blue and areas derived from additional external data are depicted dark blue. The latter category apparently covers the major part of the flooded area, especially undeveloped zones. Other ways of getting similar data would include automated interpretation of up-to-date satellite imagery, which were too imprecise in this case.

The section displayed in the following map (Fig.49) shows the actual combined flood discharge area of 2013 in comparison to the extent of the predicted HQ300 event. In some parts of Upper Austria and Lower Austria, the discharge areas equalised or even exceeded the predicted extent of a 300-year event. There are however only very small patches in the research area, which were inundated alongside the approximately 107 kilometres of the Danube River and not disclosed in the HQ300 plan. A few larger areas (yellow polygons) show the predicted floodplains of a 300-year event, which were not inundated during the 2013 floods, but in 2002. Most of the measuring stations recorded equal or higher flow rates and water levels during and after the 2013 events, but the flood extent was significantly reduced compared to the 2002 events. This is mainly based on the improved and newly created flood protection. However, the major differences between the discharge areas and the HQ300 extent, which are retrievable from the dataset (Fig.49), may be due to permanent infrastructure just like the Stockerauer express road. Investigations on a large scale map confirm the assumption that the road and its foundations function as an artificial dam, which retains high water levels just below a HQ300 event.

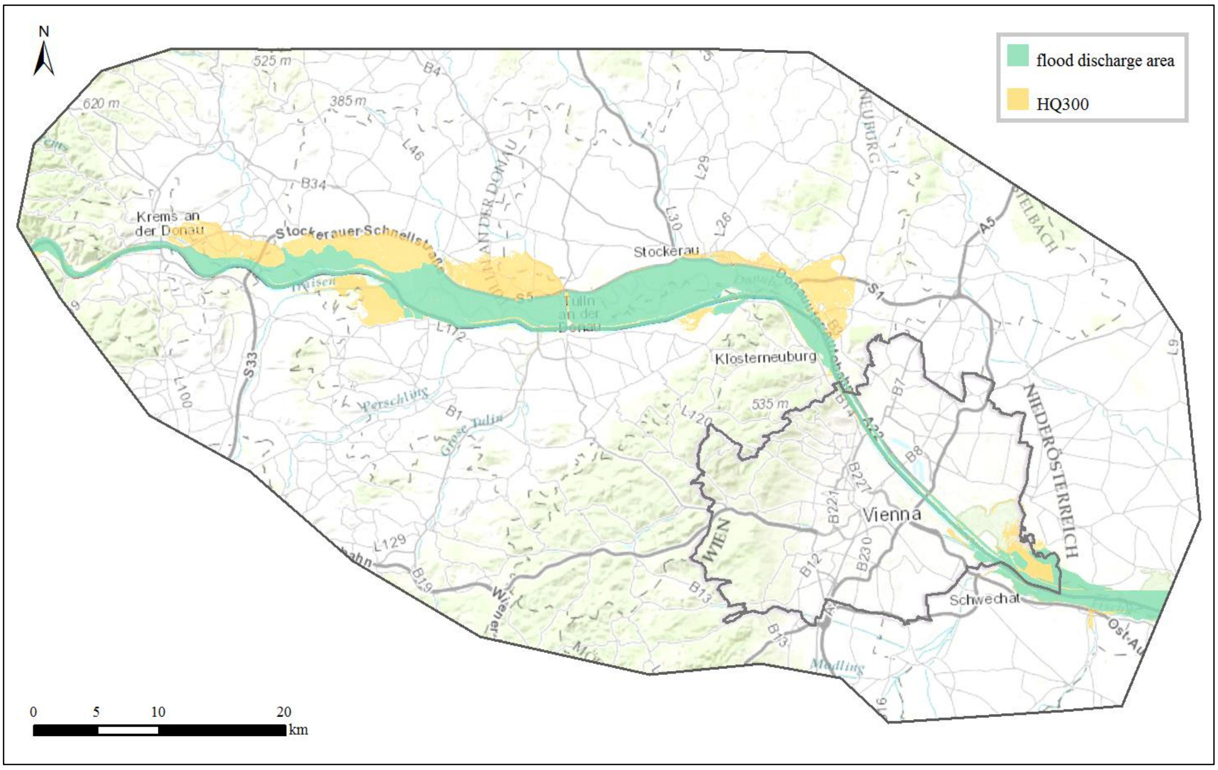


Fig.49: Combined discharge area of the flood in 2013 compared to the HQ300 model

3.2.4. Additional Factors and Information

In order to be able to conduct a supporting analysis about the relationship between the imperviousness of soil and the 2013 flood extent, which focusses on the explanation of the calculated correlations from the FQA data and the supervised classification, further information has to be gathered and added. In the following chapter, all the additional important data will be introduced and discussed shortly. This includes general data about the research area, data from official governmental bodies, internet platforms or information from other professional sources.

The *general data* section consists mostly of already familiar information about the extent of the research area (which is based on the distribution of FQA points), the borders of the two affected Austrian states Vienna and Lower Austria, as well as the course of the Danube River. Other features, which will be used during the following analyses, include:

~ *points of the FQA campaign*, which comprise the distribution of the FotoQuest Austria unique sample points in the research area and their corresponding land cover classifications (Fig.61)

~ *reclassified FQA points*, in order to illustrate the unique sample points in the research area which represent the attributes sealed (1) or unsealed (0) (Fig.69)

~ *reclassified LUCAS points* for the illustration of comparable sealed (1) and unsealed (0) areas

~ *an interpolated map* (IDW) of all the sealed surfaces in the research area based on FQA points, with a threshold for sealed or unsealed areas with 0.5 (Fig.72)

~ *the distribution of sealed surfaces* in the research area, based on the comprehensive automated supervised classification (SVC) of millions of picture elements or pixels (Fig.68)

Another important part of the general data is the Digital Elevation Model (DEM). A DEM is a digital representation of a certain area's topography and commonly represented as a raster dataset. This grid of squares consists of numerous pixels, which feature different grey tones. These grey scales contain the height information, ranging from white (higher areas) to black (lower areas). The elevation model for the research area is illustrated in the following mapping (Fig.50).

The basic data was acquired through the Earth-Explorer platform [EEX-17], which is a product of the United States Geological Survey (USGS). It offers various datasets like aerial imagery, declassified Keyhole satellite imagery, global land cover situations and trends, Landsat data, radar and LIDAR information as well as vegetation monitoring for nearly every location on the planet. In this case, the data for the Digital Elevation Model was retrieved from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), which is an US-American and Japanese co-production.

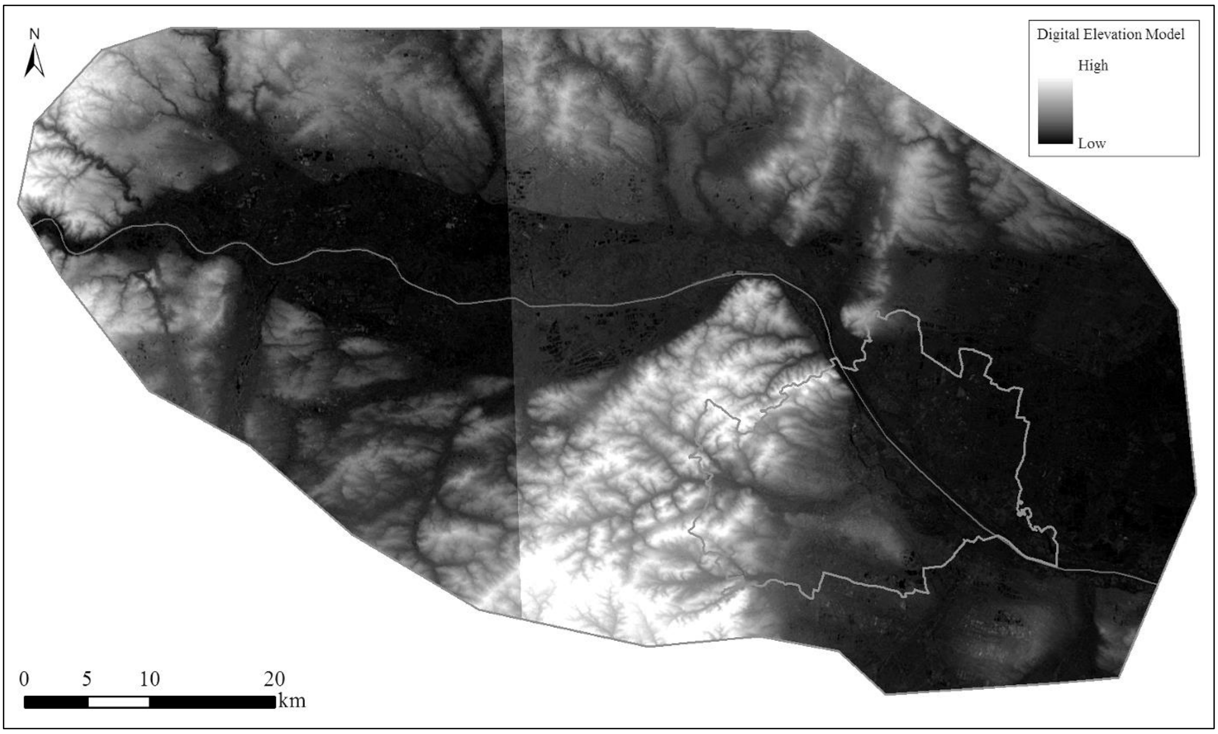


Fig.50: Digital Elevation Model (DEM) of the research area

The section of the DEM, which will be used in the following analyses, is illustrated in the figure above (Fig.50). Other sources for additional information were, amongst others, the Federal Ministry for Transport, Innovation and Technology in Austria (BMVIT) and the Provincial Government of Lower Austria (NOE). This section of the collective dataset contains information, which has already been mentioned in the previous chapter, and some new and important data about the potential water depths in case of a flood event in the research area. Additional information includes [VIT-15]:

~ the predicted average *water depths* of inundated surfaces in the research area alongside the Danube River in case of a 30-year event (HQ30), provided by BMVIT

~ the predicted average *water depths* of inundated surfaces in the research area alongside the Danube River in case of a 100-year event (HQ100), provided by BMVIT

~ the predicted average *water depths* of inundated surfaces in the research area alongside the Danube River in case of a 300-year event (HQ300), which is shown in the following figure (Fig.51)

~ the actual *flood discharge area or flood extent* of the 2013 event, based on the combination of aerial photograph interpretation and local surveys by official bodies (Fig.47)

~ the predicted *flood extent* in case of a 30-year flood (HQ30) in the research area alongside the Danube River (Fig.46), based on the provincial government of Lower Austria (NOE)

~ the predicted *flood extent* in case of a 100-year flood (HQ100) in the research area alongside the Danube River (Fig.46), based on the provincial government of Lower Austria (NOE)

~ the predicted *flood extent* in case of a 300-year flood (HQ300) in the research area alongside the Danube River (Fig.46), based on the provincial government of Lower Austria (NOE)

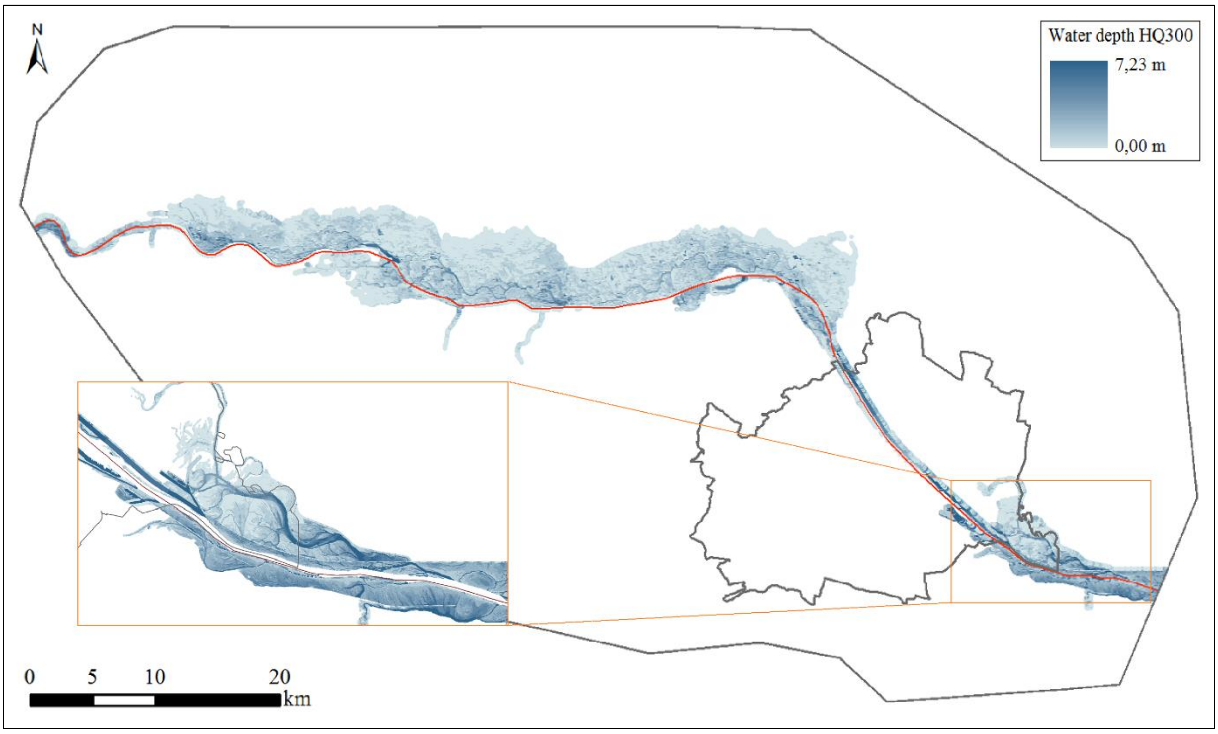


Fig.51: Average water depths in the research area in case of a 300-year flood event (HQ300)

Another major source for additional data is the Electronic Navigational Chart (ENC). It is basically an electronic map of all navigable water bodies, for navigating through oceans, seas, lakes, rivers and along coastlines. Provided by the Donau River Information Services (DORIS), the ENC includes traffic regulations, the locations of fairways, buoys, restricted areas, signs, structures or obstacles [ENC-17]. Main advantages of this electronic chart are the frequent data updates, different levels of detail and additional information about individual objects [ENC-17]. This thesis only uses a selection of reasonable and enhancing datasets, including:

~ *hydraulic structures* like dykes, levees, floodwalls, fences, groins, ground sills, training walls or revetments (if existent), some examples are shown in the following map (Fig.52)

~ additional information about *built-up areas* near the Danube River, which contain a certain amount of buildings, roads, rails or other forms of infrastructure, in this case as *polygon*-features

- additional information about *built-up areas* near the Danube River, which contain a certain amount of buildings, roads, rails or other forms of infrastructure, in this case as *point*-features
- a collection of *bridges*, tunnels, piers, pillars and other overhead obstructions [ENC-08]
- other forms of exceptional navigational structures just like *anthropogenic dams or barriers*, in order to raise the water level or as a form of flood prevention [ENC-08], as shown in the map (Fig.52)
- every integral part of a *lock system*, such as locks, lock basins, docks, walls and gates (Fig.52)
- natural *dunes, walls or ridges*, which were retained or improved by experts in order to keep potential high water within the original river course [ENC-08], also illustrated in the following map (Fig.52)
- the courses of *non-navigable rivers, side arms or tributaries* of the Danube River, which may or may not be connected to the actual riverbed of the Danube (Fig.52)
- additional information about the normal average *water depth* of the Danube River (Fig.52)
- and polygonal information about the *dry land area* surrounding the water bodies as an opposition to the river courses and tributaries [ENC-08], also retrievable from the following map section (Fig.52)

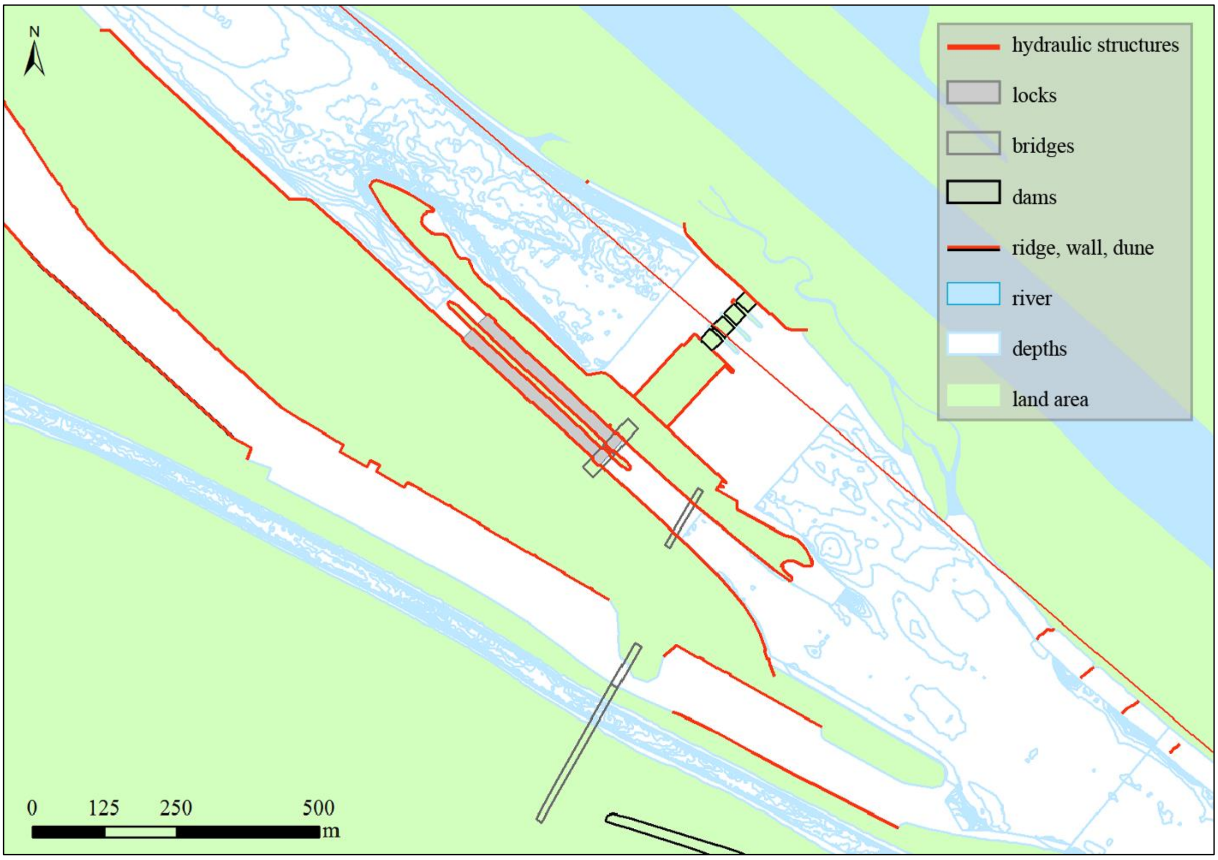


Fig.52: Used elements of the Electronical Navigation Chart (ENC) for inland shipping on the Danube River (selection)

In addition to these three big complexes listed before (general data, federal agencies and governments, Electronic Navigational Chart), the financial, social and environmental damage in case of a flood event is also worth mentioning at this point. In 2002 alone, the disastrous flooding in Central Europe resulted in a property damage of an estimated 15 billion Euro, from which three billion Euro worth of damage was observed in Austria alone [VIT-15]. About eleven years later, after the flood event in 2013, damages of *only* 870 million Euro were estimated in Austria, although there were higher flow rates than in 2002 [VIT-15].

The following map (Fig.53) offers an overview about a selection of different types of damages, after the 2013 flood event. First, there are official numbers of damaged buildings in the research area. As it appears, most of the registered damages included more than 50 buildings, in the western part of Krems and near Klosterneuburg. Other types of damages include structural impairments, mostly of local infrastructure, agricultural land, hydraulic structures and other facilities. Another problem was the recorded sediment shift. The Austrian part of the Danube River showed a higher amount of sediment erosion (a minus of 12,350,000 cubic metres) after the 2013 flood, than sediment accretion (a plus of 6,589,000 cubic metres). This redistribution of masses during the inundation, also had a massive influence on the high water levels and the flood extent alongside the whole Danube River course.

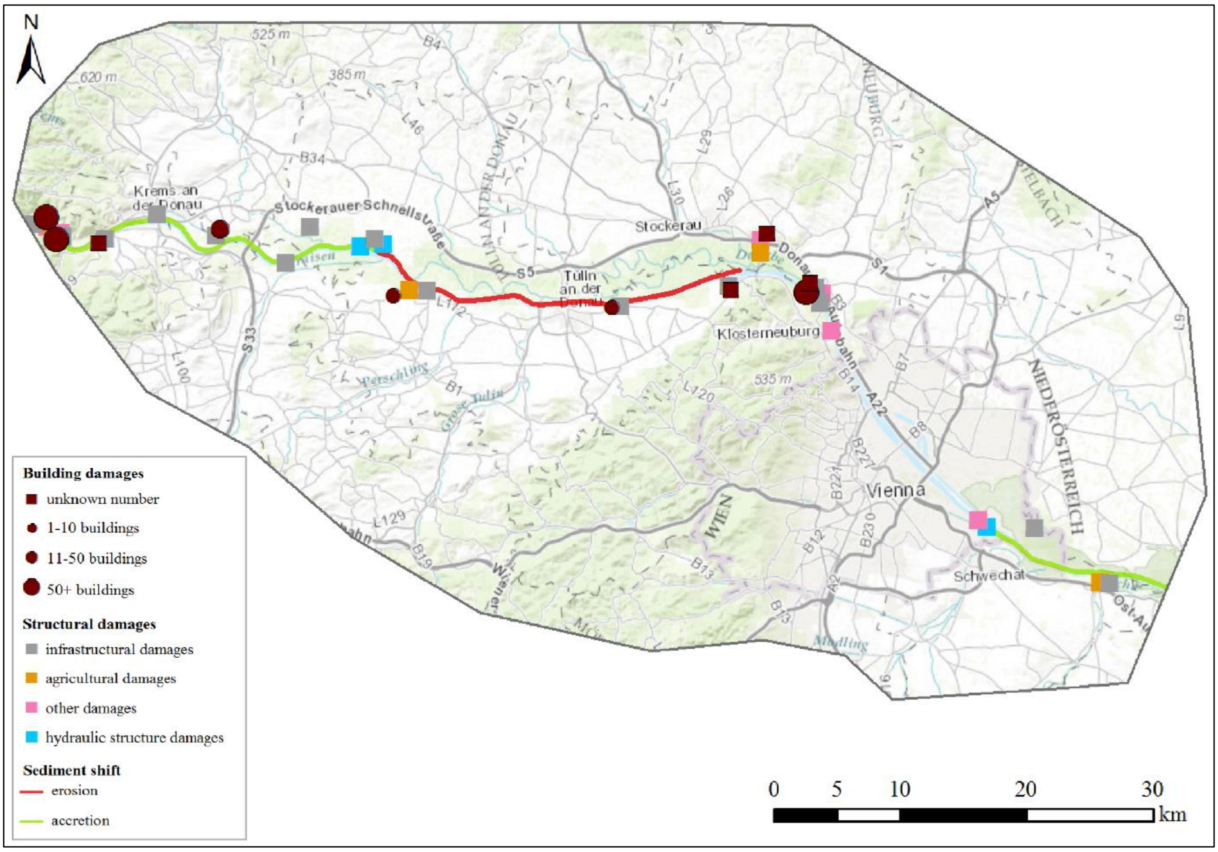


Fig.53: Different types of damages due to the 2013 flood event and resulting sediment shifts in the research area

Because of this sediment shift, some areas were inundated, although the flood prediction for a 300-year event (HQ300) did not disclose them as vulnerable (and vice versa). There are however other influencing factors, which may or may not affect the distribution of flood discharge areas and the propagation speed.

In order to investigate the correlations between soil sealing and flood extent, based on the FQA data and the automated supervised classification, open data from corresponding platforms like OpenStreetMap was imported. With the appropriate plugin for QGIS, it was possible to select OSM data for the research area, containing all the information available. All required features like building footprints and the prevailing street network were extracted and saved as polygon and polyline shapes. After clipping the shapefiles with a 4500 metre buffer polygon, it covers the same specified area, which will be investigated in chapter *4.3. Soil Sealing and the Correlation to Flood Extent*. The following map section (Fig.54) shows a part of the research area near Korneuburg with the OSM building footprints and the street network. In addition to that, the river course, other water areas and the discharge area of the 2013 flood event were also incorporated. This will be the data base for general area statistics and the comparison to the results of the potential correlations, based on FQA data and the supervised classification.

All these freely available datasets, presented over the previous pages, will be used in chapter *4.3.2.* about the traditional approach of investigating the potential relationships between soil sealing alongside the Danube and the flood extent (using open data and unrestrictedly accessible governmental platforms).

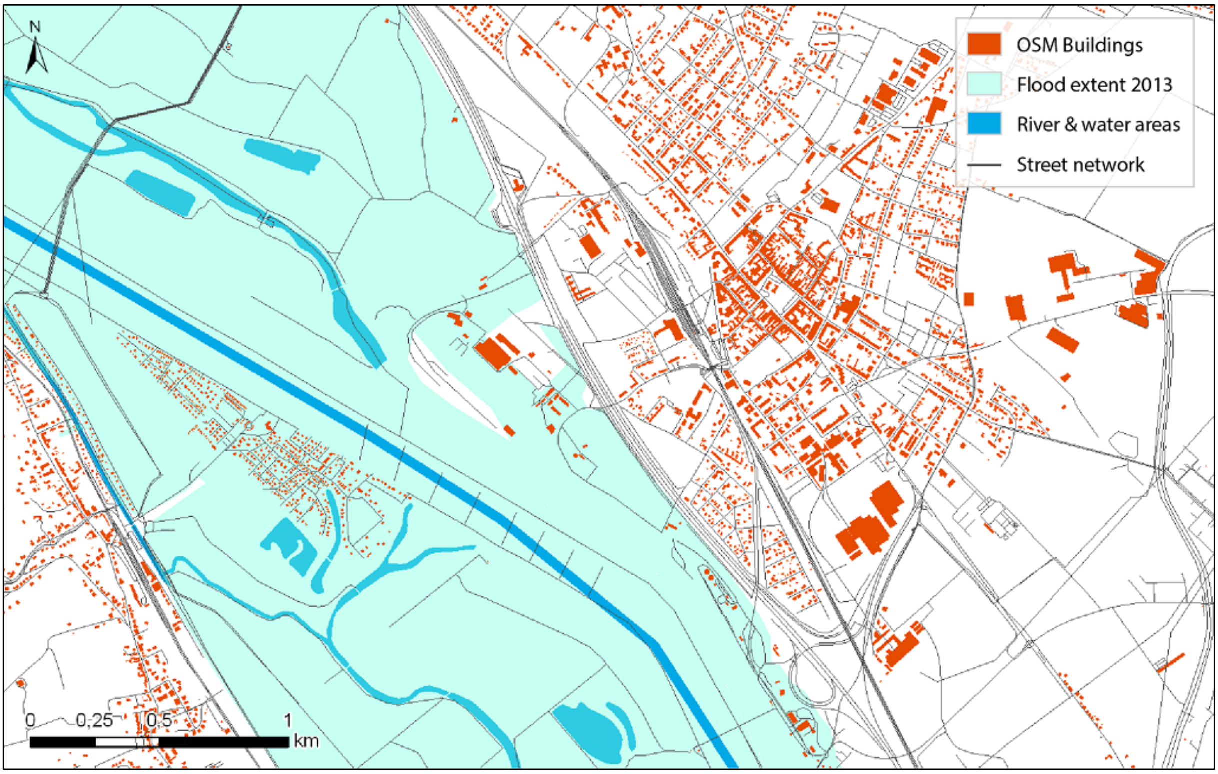


Fig.54: Section of Korneuburg with building footprints and traffic infrastructure from OpenStreetMap

3.3. Data Pre-Processing

The next part of this master thesis focusses on the pre-processing and preparation of the data. Since the analysed information was generated through volunteers during a geographic citizen science campaign, several quality and consistency checks have to be carried out, before finally analysing the data. Starting with a closer look at the general statistics of the raw and processed data, the chapter continues with an accuracy assessment, with LUCAS 2015 data as comparable ground-truth. The chapter about quality management and issues provides an insight into the problems and challenges that occurred during the data preparation. A supervised classification from satellite imagery provides a reasonable comparison to the FQA classification in the research area. The discussion of the applied methodology and technology concludes the chapter about data pre-processing and preparation.

3.3.1. General Statistics

The following tables (Tab.7, Tab.8) provide general information about the data used in this master thesis. This data has been separated into three parts: the original raw data, the processed data and the information concerning only the research area. The original dataset consists of more than 19,900 rows of information. This includes five rows for each validation with the same information, because every picture which has been taken in the five directions (North, South, West, East and on the Spot) created one row in the dataset. After filtering the information just for the spot, less than 3000 rows of information remained, because for calculating the general statistics and the accuracy assessment, only one data row per validation was enough. The next step was the elimination of every unusable point, which could not be attended or accessed due to inaccessibility or invisibility.

After all these processes were performed, *only* 1945 validations (Tab.7) for 1483 individual sample points were left (Tab.8). After choosing the research area alongside the Danube River, the number sank to 592 validations for 393 unique sample points. This is nearly a third of the processed data and only a quarter of the original raw data. This may be caused by a higher number of inaccessible or private locations.

During the pre-processing of the data, the number of individual users and contributors sank from originally 208 people, to 207 and finally 80 people in the research area (Tab.7). This implies a decrease of more than 60%. Similar developments continue, when investigating the average validations per user (Tab.7). The arithmetic mean is the sum of all validations divided by the number of users. The original data offers a value of 11.4 validations per user, whereas the processed data only offers 9.39 validations. Focussing only on the research area however, the value sinks to an average of 7.4 validations per user.

This means that the number of individual validations is actually quite low, although there are a number of users playing the FQA application. The middle value of the dataset (median) offers a constant value of two validations per user (Tab.7). This calculation eliminates statistical outliers and provides more information of how much the individual user actually contributed. The difference between these two values may be due to some strong and motivated individual volunteers, which may or may not be closely connected to the project.

The maximum validations per user in the original data are 501. Considering the 2372 total validations, this means that more than 21% of all information was contributed by only one person. After processing the data, the maximum validations sank to 207 (10.6%). The highest amount of validations per user in the research area is 140, which implies that more than 23% of all the information was contributed by one person. This could cause either positive (consistency) or negative (error susceptibility) effects on further analyses. The minimum amount of validations per individual user is always one (Tab.7).

User Validations	Original Total Data	Processed Total Data	Research Area
Total Validations	2372	1945	592
Individual Users	208	207	80
Validations per User (arithmetic mean)	11.40	9.39	7.40
Validations per User (median)	2	2	2
Maximum Validations per User	501	207	140
Minimum Validations per User	1	1	1

Tab.7: General statistics about the voluntary contributions by the Users / Players

The next table (Tab.8) focusses on the validated samples. Originally there were 1849 unique sample points, which were narrowed down to 1483 after the data processing. In the research area there are only 393 sample points left, which were validated 592 times. This indicates an arithmetic mean of 1.51 validations per sample, which is actually an increase from the original data. All original information offer an arithmetic mean of 1.28 validations per sample point. A value of 1.51 means that every unique sample point has been validated one and a half times. Such a value is actually rather low and leaves quite a lot of room for errors. As a result of that, additional quality assurances like accuracy assessments are necessary, which will be discussed in the following chapter.

The median value offers only one validation per sample (Tab.8). To ensure a good and reliable result, the value should be at least around three or four validations per sample point as a minimum. However, the maximum number of validations per sample in the research area is ten. This point has been visited ten times by the users and can therefore be clearly designated. The minimum amount of validations per sample is only one.

Validated Samples	Original Total Data	Processed Total Data	Research Area
Total Validations	2372	1945	592
Unique Sample Points	1849	1483	393
Validations per Sample (arithmetic mean)	1.28	1.31	1.51
Validations per Sample (median)	1	1	1
Maximum Validations per Sample	11	11	10
Minimum Validations per Sample	1	1	1

Tab.8: General statistics about the validated samples

To provide an insight into the actual statistics of validations per sample, the next Figure (Fig.55) illustrates the number of validations per sample point using the data for the research area only. It is quite obvious, that most of the locations were validated only once. With a share of 73.54%, nearly three-quarters of the sample points were visited only once (289 of 393 unique locations). Only 15% were visited twice (59 locations). Approximately 5.34% were validated three times (21 locations), 2.80% four times (eleven locations) and 2.04% were validated five times (eight locations). The lowest amount of validations per sample are six times (0.51%, two locations), eight times (0.51%, two locations) and ten times (0.25%, one location). No sample point has been visited seven or nine times. In summary it can be said, that almost 90% of all unique sample points, were validated only once or twice.

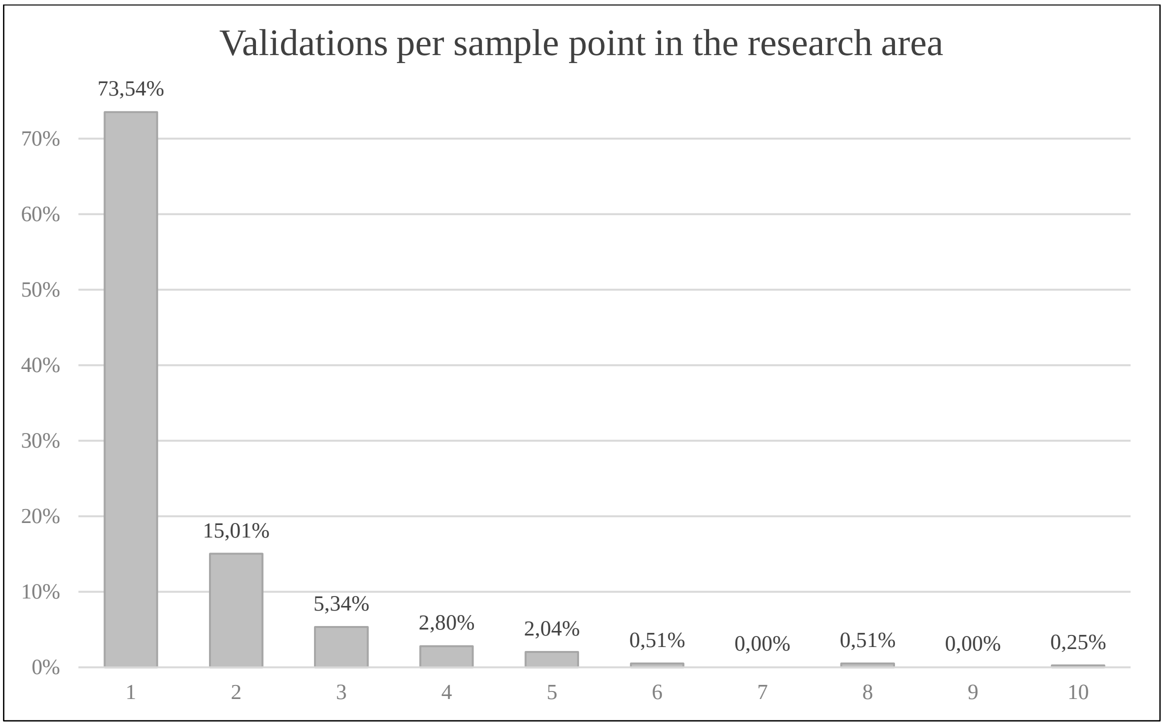


Fig.55: Number of validations per sample point using the data of the research area

Another point of view for looking at these statistics is illustrated in the following figure (Fig.56). This chart shows the number of validations per user on the basis of data from the research area, divided into five classes. The main reason for presenting this diagram, is to illustrate how often the 80 individual users contributed information for the FQA project. As already mentioned before, a large number of users only contributed one validation to the dataset. Exactly 46.25% of all volunteers only validated once, which means that 37 out of 80 individuals did not want to take part in the campaign anymore, after their first try. Some of the causes were hopefully evaluated by IIASA team members, before they launched their successive FotoQuest Go campaign for the whole of Europe.

The next class comprises all the users which contributed from two to ten validations to the FQA dataset (Fig.56). The value of 40% equals 32 individuals who validated this often. Another class represents all the users, which validated between eleven and 20 times (two individuals). Only 5% contributed between 21 and 30 validations (four individuals), and 2.5% between 31 and 40 (two individuals). Only three players can be considered as *power users*, because they each validated more than 40 times (3.75%). In summary, more than 86% of the volunteers (or 69 individuals) only contributed less or equal to ten times to the FotoQuest Austria campaign. Compared to the crowdsourcing project, the author already worked on for his bachelor thesis, these values are substantially smaller. In 2014, extracts of this similar crowdsourcing project have been published [ALB-14], which showed an average of nearly 293 validations per user (arithmetic mean) or 205 respectively (median). The results of the analyses were quite promising, even if the producer's and user's accuracy were capable of improvement.

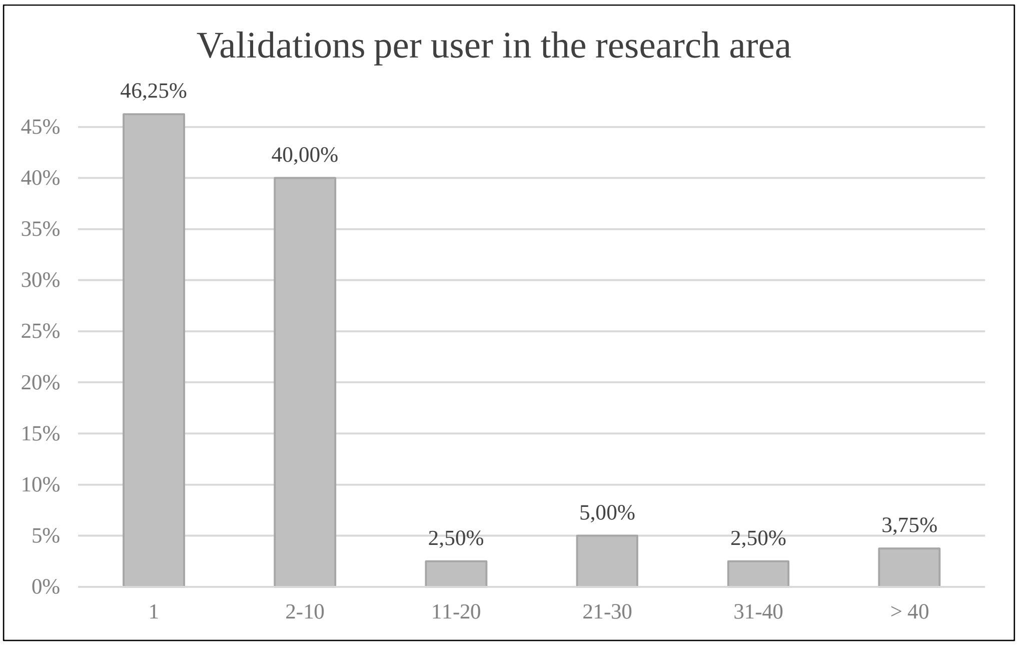


Fig.56: Number of validations per user using data from the research area

Another point that can be discussed, is the reason why so many sample points could not be validated properly and were therefore skipped by the players. This statistic is only available for the original raw data, because the processed data and the information for the research area only consists of useful and valid data, which has been validated properly. There were 792 omissions in total during the 2015 campaign. The following pie chart (Fig.57) illustrates the three main classes of reasons, why the sample points were skipped. The classes are based on inaccessibility, private property and technical problems during the validation. Some original examples of the main skip reasons are:

~ *Inaccessible:* field, highway, construction site, crop land, dense forest, airport, golf course, in the water, maize field, military compound, obstacle, trees, transformer station, vineyards, fence…

~ *Private property:* private property, on private ground.

~ *Technical problems:* bad GPS signal.

The main problem with 36.36% was that the sample point lay on private property (288 skips). Not far behind with 34.97% were technical problems, mainly due to bad GPS connectivity (277 skips). Inaccessibility was the third most common reason for skipping the sample point. With 28.66%, there were a lot of agricultural, forestal and infrastructural obstacles in the way (227 skips).

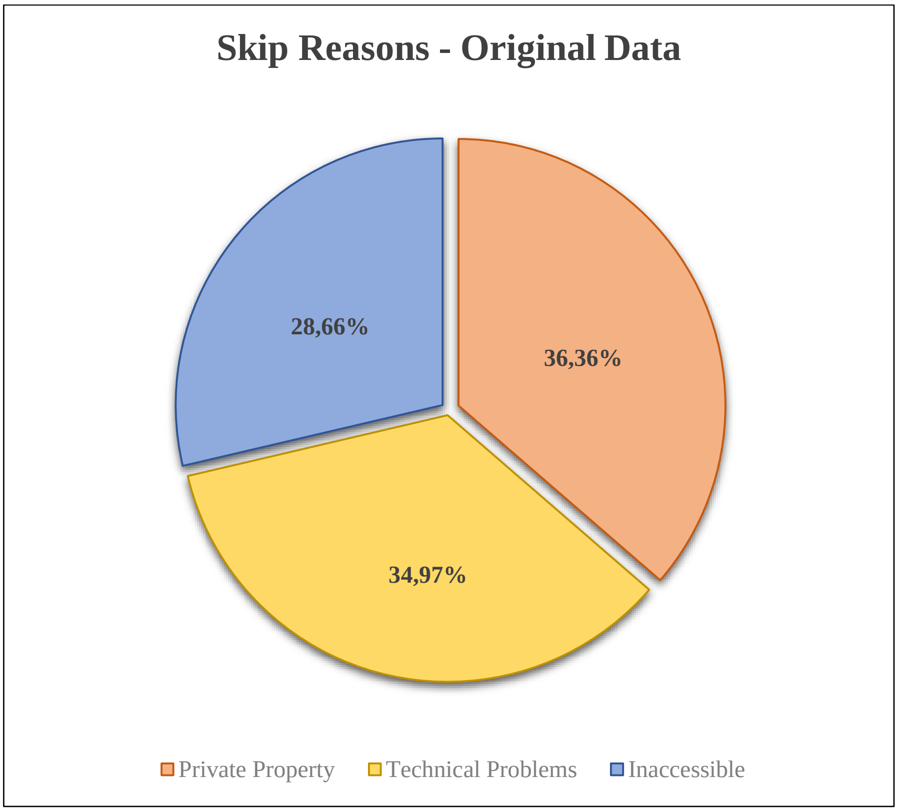


Fig.57: Different classes of skip reasons during the FQA campaign

The most important part of the general statistics is the distribution of the different validated classes in the research area. The following bar chart (Fig.58) displays the percentage of all different class types prevailing in the research area, on basis of the FQA dataset. This information is the basis of further analyses, concerning possible correlations between soil imperviousness and the extent and damage of recurring flood events. Artificial Landscapes, Cropland, Woodland, Shrubland and Grassland are of particular interest in the following cases.

The research area encompasses an area of approximately 4000 square kilometres (Fig.35) and offers the living environment for more than two million people. However, according to the data, only a quarter of the area was classified as artificial landscape. Around 27.95% of the nearly 400 unique sample points were categorised as class A. The highest share of all the classifications refer to cropland (class B) as the prevailing land cover. With 46.67% nearly half of the area was considered as agricultural or farm land. This means, that nearly three-quarters of all the visited sample points were one of these two classes.

The remaining six classes of the FQA classification scheme shared the remaining one quarter of the validations. Woodland (C) offers a share of 11.03% and Shrubland (D) only 2.82%. Grassland (E) is the fourth most common land cover type in the research area with 7.18%, whereas Bare Land and Lichens (F) only cover 1.03%. The Water areas (G), including the Danube River, cover more than 3% and surrounding Wetlands (H) around 0.26%.

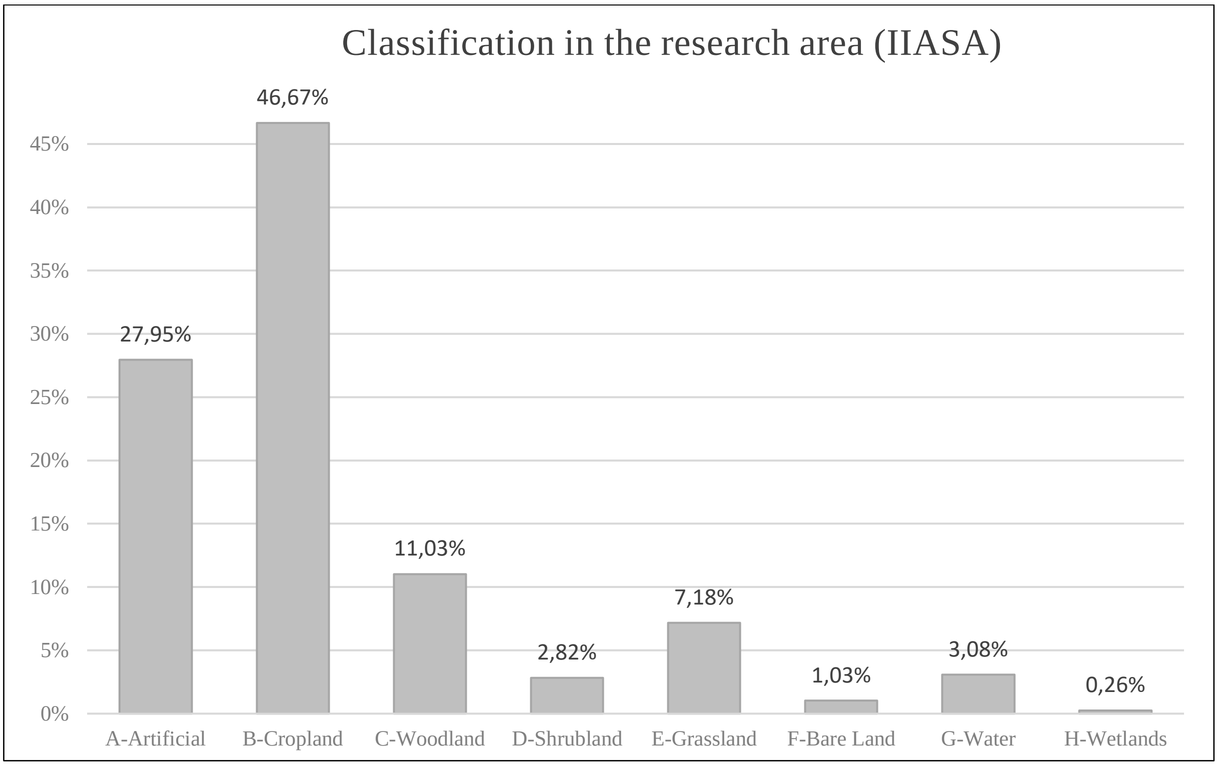


Fig.58: Percentage of the different IIASA class types prevailing in the research area (n=390)

If we consider that the research area encompasses approximately 4000 square kilometres, then nearly 1120 square kilometres would be artificially sealed or impervious, 1870 square kilometres cropland and 850 square kilometres of areas with a prevailing vegetation (woodland, shrubs, grass).

The next chart offers similar information (Fig.59). It shows the percentage of the eight land cover class types prevailing in the research are, based on the comparable LUCAS dataset. In this case however, only 143 unique sample points are usable. Due to the different number of sample points, comparing these two figures should be considered with caution. Looking at the two graphs however, there is a similar trend derivable. They both show similar peaks and lows for the classification types, although the percentages differ a bit. The Artificial Landscape (A) offers 23.08%, which is only 4.87% less than in the FQA dataset. According to the LUCAS survey, nearly 40% of the research area is covered with farm land (B). This is approximately 6.80% less than the FQA campaign suggests. The land cover type Woodland (C) offers 11.89%, which is quite similar to the FQA value. Shrubland (D) with 4.90% and Grassland (E) with 16.08% nearly double their values in the LUCAS survey. This is one of the most considerable differences between these two land cover datasets. Bare Land (F) covers around 0.70% of the research area. Water areas (G) with 3.50% and Wetlands (H) with 0% almost stay the same compared to the FQA data. Although there are some remarkable differences in the share of the different land cover classes, the main distribution and trend can be treated as similar and therefore comparable.

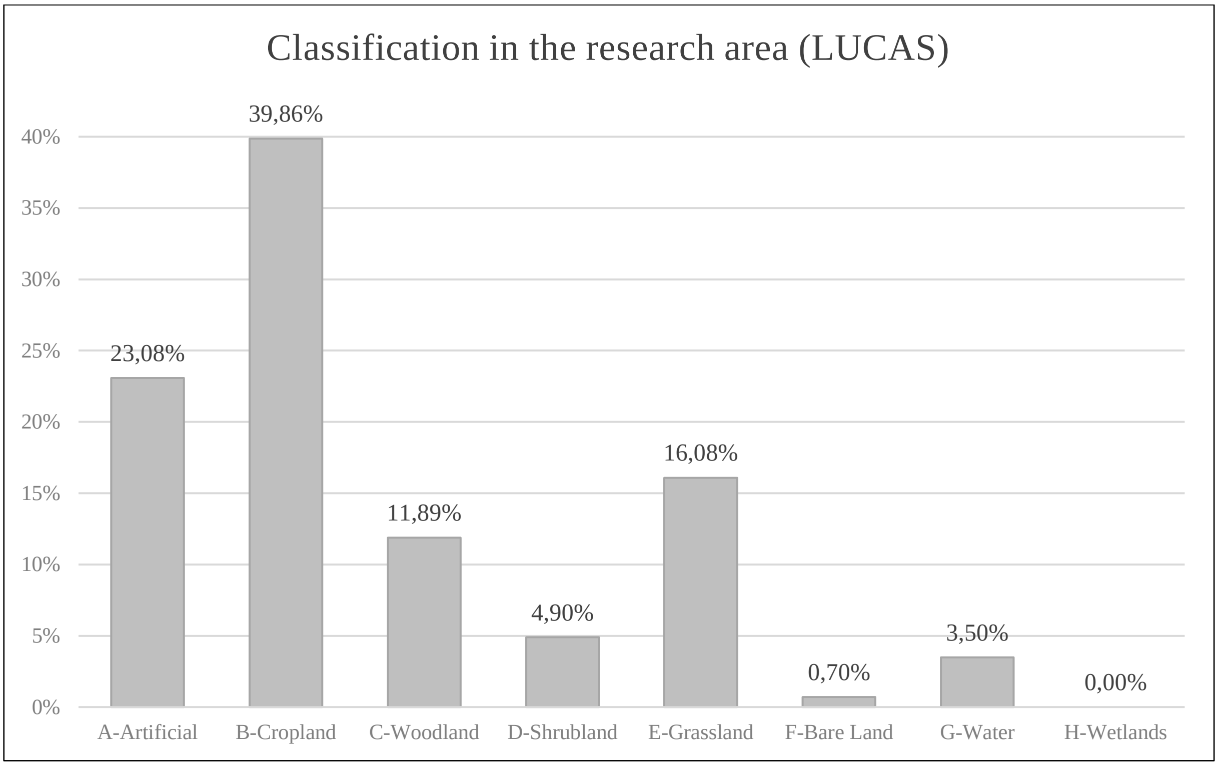


Fig.59: Percentage of the different LUCAS class types prevailing in the research area (n=143)

The distribution of different land cover types is not surprising for a European country like Austria. The research area comprises a wide array of different landscapes, including the metropolis Vienna, the mighty Danube River, parts of the Vienna Woods and large areas of agricultural land. The differences are resulting from the different amount of validations and unique sample points.

As already indicated before, most of these sample points were visited only once during the FQA campaign, which also reflects in the score statistics. Every first validation of a certain point awarded the participating user 100 points for their account. If the player was the second visitor of the same unique sample point, he or she just got 90 points for the validation. This system was designed to offer an incentive for the exploration of new and remote places, in order to generate a wide spectrum of validations. However, the more often a location was visited, the fewer points the user received (e.g. 81, 72, 65, 59, 53, 43 … 10, 2, 1). This surely did not encourage the players to visit an already validated location and therefore decreased the data quality.

By looking at the percentages of the most frequent scores (Fig.60), it is possible to take an educated guess about the data's quality. If nearly three-quarters of all the scored points stand for the one-time validation of a location, the resulting quality can be considered questionable. Just like mentioned before, the more often a unique point was validated, the better it was for the resulting data quality. Further research about the quality and reliability of the information has to be done.

More than 72% of all validations were first-time and last time visits of a unique location. About 11.70% of the analysed locations generated 90 points, or were therefore visited twice. Approximately 16% of the unique locations scored less than 90 points. The average score in the research area is approximately 89.32. Less points per average would have been an indicator for better data quality and reliability in this case. Anyway, not every one of these sample points offered a value in the dataset. This means, that these numbers cannot be compared to the former statistics straightaway. They are just mentioned for the purpose of getting an idea, what one of the reasons for a possible below-average data quality might be.

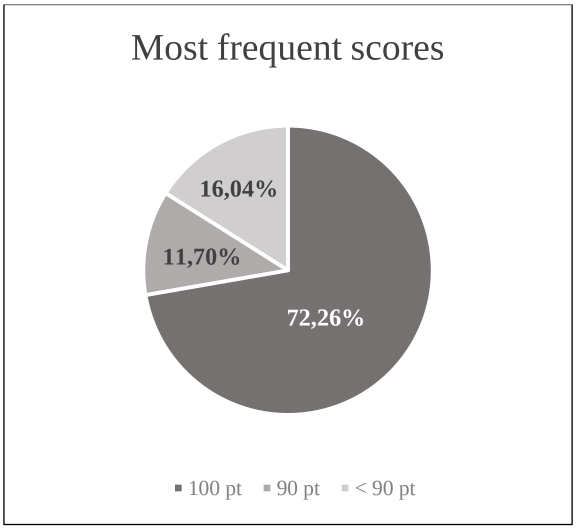


Fig.60: Most frequent scores during the FQA 2015 campaign

This chapter concludes with a chart, showing the spatial distribution of all the remaining 393 validated land cover classifications, contributed during the FotoQuest Austria campaign (Fig.61). The values for validations per user or validations per sample point from the last few pages, indicated that there is great potential for the use of crowdsourced data, but only if there is a sufficient amount available. The main problem of the FQA dataset is still the improvable spatial distribution of classified sample points all over the country. The campaign obviously did not appeal to a wider public, which led to a considerable amount of one-time users.

However, as the following chapter about the accuracy assessment will show, there is still enough information with sufficient data quality available, in order to carry out suitable analyses for this master thesis. The following map about the distribution of all classifications (Fig.61), shows the fundamental basis on which the coming calculations will rely on. It shows all eight land cover types and their unique locations. The colours correspond to the ones used in the preceding tables about the FQA class types. It is obvious, that most of Vienna was classified as artificial landscape (grey dots, containing transport infrastructure, buildings and otherwise sealed surfaces), as well as Stockerau, Korneuburg, Traismauer and other relevant cities and villages. A visual interpretation and estimation of the classifications, overlaying satellite imagery, seem to confirm their credibility and reliability. To make sure, that the crowdsourced information about this area is as accurate as possible, the next chapters focus on accuracy assessments, quality management and issues within the data.

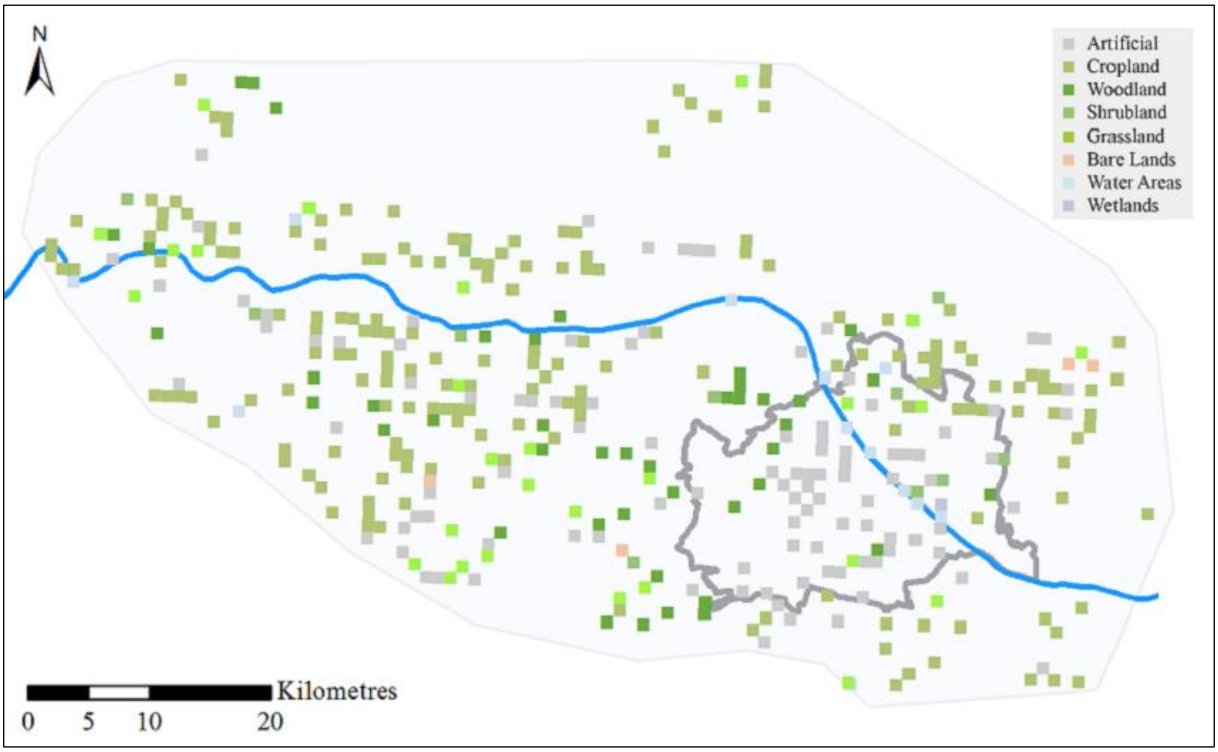


Fig.61: Distribution of all land cover classifications based on the FQA data

3.3.2. Accuracy Assessment with LUCAS Data

Whenever someone is working with crowdsourced data, quality appears to be one of the major issues of this method. Untrained volunteers contribute most of the geographic information via citizen science projects or mobile applications. Crowdsourcing initiatives always require a conceptual framework, which ensures the consistency, quality and reliability of the generated information. Several ways of guaranteeing a certain level of data credibility and user motivation, in the course of a crowdsourcing or citizen science project, have been discussed in the corresponding chapter about the crowdsourcing approach and its geographic derivatives.

A well established and elaborate way of estimating the correctness of crowdsourced data is the accuracy assessment. This mathematical method provides information about the quality of classification results by comparing them to a certain ground-truth. The main conclusion of the accuracy assessment is the degree of how much the voluntarily generated data represents the real world [MKA-07]. In this case, the LUCAS dataset was chosen as the ground-truth. This method is only appropriate, when the ground-truth is accurate and reliable. The main contracting authority of LUCAS was the European Commission and the survey was conducted by trained experts all over the continent. This should vouch for a certain degree of reliability. In general, however, if established global land cover datasets (GLC, CORINE, MODIS, LUCAS etc.) are compared to each other, they only match in approximately 80% of all cases. This points out the base-uncertainty, which dominates the scientific use of land cover information.

The LUCAS dataset from 2015 was used in a number of research projects in the past two years and very often delivered solid and reliable results. As a result of that, it was chosen as an appropriate ground-truth or reference data (Producer) for assessing the FQA dataset (User). Not every unique sample point of the FotoQuest campaign had an associated LUCAS point, due to different sample sizes or sampling methods and missing classifications. That is why out of the 393 unique FQA points, only 143 could be compared to corresponding LUCAS points. There were however enough sample points to compile a decent assessment, which provides statistical information about the overall, user's and producer's accuracy.

All eight different land cover types were compared to each other in the subsequent figure. The accuracy assessment is usually displayed as a confusion or error matrix (contingency table) and all following explanations are based on results of the following confusion matrix (Tab.9). The abbreviations of the classes used in this matrix are as follows:

A ...... Artificial Landscape
B ...... Cropland
C ...... Woodland
D ...... Shrubland
E ...... Grassland
F ...... Bare Land and Lichens
G ...... Water Areas
H ...... Wetlands

The overall accuracy is the division of correctly classified sample points by the total amount of sample points (Tab.9). In this case, all the matching numerical values, which have been highlighted in the appropriate colours (29, 52, 11, 1, 5, 0, 4, 0), were divided by the sum of all validated sample points (n=143). The result is 71.33% (or 102 out of 143 compared points) and therefore not as high as the author had wished. It can be considered as a viable agreement, but approximately three times out of ten occasions, the classification does not correspond to the reference dataset (LUCAS). As indicated before, one of the main problems of this value is, that there is no further information whether the disagreement is evenly distributed between the different land cover types, or if some classes offer an excellent agreement and some others do not agree at all. As a result of that, it is necessary to calculate the values for the much more informative user's and producer's accuracy.

The user's accuracy provides information about the correctness of the data from a user perspective. In this case, it is represented by the FQA dataset. The user's accuracy corresponds to the disagreement of commission or inclusion, which is why this method is also referred to as commission error [PDX-12]. It provides details about the probability of a crowdsourced classification, representing the land cover type in the reference dataset or the real world [MKA-07]. The arithmetic way is defined as follows:

$$User's\ Accuracy = \frac{Number\ of\ correctly\ classified\ sample\ points}{Row\ totals}$$

A perfect agreement of 100% can be found for water areas (class G). All four classifications coincide with the reference data (Tab.9). However, this value is quite inconclusive, because of the small amount of validations. The second highest value of user's accuracy can be found for the woodland class (C). A very good agreement of about 84.62% indicates, that eleven out of the 13 classifications done by the FQA volunteers represents the real world (or rather the reference dataset) correctly. This means by implication, that the commission error is just 15.38%. The next higher percentage amounts to exactly 80% for the cropland class (B). This very good agreement of 52 out of 65 classifications, leads to a commission error of 20%. With 61.70%, only 29 out of 47 classifications were validated correctly in the artificial landscape (A). This rather moderate agreement implies a commission error of 38.30%. Shrubland (D) and Grassland (E) both only have a user's accuracy of 50%, which means that only half of all classifications done by volunteers correspond to the IIASA dataset (generated by trained experts). This is equivalent to a moderate agreement. The classifications for bare land (F) and wetlands (H) both offer a value of 0%. This means, that there was no agreement between the FQA campaign and the LUCAS survey and that both commission errors are 100%. However, the small amount of validations (two in class F) or the absence of any classification at all (class H) put these *low* values into some perspective. Overall it can be said, that the accuracy from the user's perspective is improvable and partly based on the low amount of comparable data. A similar land cover project was conducted by the author during his bachelor thesis, which offered values for the user's accuracy between approximately 75% and 80%.

The producer's accuracy on the other hand provides information from a producer's perspective. The reference data in this case is the same LUCAS dataset from 2015. This method calculates the probability of a certain point in the real world (or reference dataset) complying with the same land cover class in the crowdsourced dataset. The producer's accuracy therefore corresponds with the disagreement of omission (exclusion) and is called omission error [PDX-12]. The arithmetic way is defined as follows:

$$Producer's\ Accuracy = \frac{Number\ of\ correctly\ classified\ sample\ points}{Column\ totals}$$

The highest value for the producer's accuracy with 91.23% can be found within the cropland classifications (B) (Tab.9). This almost perfect agreement means that 52 out of 57 validations in the real world (reference dataset) also represent the same class on the crowdsourced data (FQA). The omission error in this case would be 8.77%. The next highest value is 87.88% for artificial landscapes (A). About 29 out of 33 classifications are labelled accordingly and therefore offer an omission error of only 12.12%. With 80% producer's accuracy, water areas (G) seem to have a substantial agreement. Unfortunately, there were only five classifications in total and therefore this value can be considered as not conclusive enough. Woodland classifications (C) however only offer a producer's accuracy of 64.71%, which can be described as a moderate agreement. Just eleven out of 17 classifications define the probability for the sample points of the reference data to be labelled accurately in the FQA dataset. The omission error is a substantial 35.29%. Grassland (E) has only a slightly fair agreement with 21.74% (omission error 78.26%). Out of 23 classifications, only five of them seemed to be correct. Even lower is the slight agreement for Shrubland (D) classifications with 14.29%. Only one of seven is represented correctly. Just like before, bare land (F) and wetlands (H) only offer values of 0%. This may again be due to the inconsiderable low amount of available classifications.

Overall Accuracy **71.33%**		**LUCAS** (producer)								Sum of validated sample points	User's Accuracy
		A	B	C	D	E	F	G	H		
IIASA FQA (user)	A	**29**	3	3	2	10	0	0	0	47	61.70%
	B	3	**52**	1	3	5	1	0	0	65	80.00%
	C	0	1	**11**	1	0	0	0	0	13	84.62%
	D	0	0	0	**1**	1	0	0	0	2	50.00%
	E	1	1	2	0	**5**	0	1	0	10	50.00%
	F	0	0	0	0	2	**0**	0	0	2	0.00%
	G	0	0	0	0	0	0	**4**	0	4	100.00%
	H	0	0	0	0	0	0	0	**0**	0	0.00%
Sum of reference area		33	57	17	7	23	1	5	0	143	
Producer's Accuracy		87.88%	91.23%	64.71%	14.29%	21.74%	0.00%	80.00%	0.00%	Cohen's Kappa: **0.60**	

Tab.9: Confusion matrix of the first level land cover classifications

Summarised it can be said, that the three highest values of the producer's accuracy offer substantial agreements and are therefore acceptable and suitable for the whole analysis. However, four of eight classes only have values between zero and 22%, which is way too low for a proper accuracy assessment. Just like mentioned before, one of the major reasons for this situation may be the low amount of potentially comparable classifications. Out of more than 2000 unique FQA sample points all over Austria, only 393 remained in the research area. However, only 143 of them offer accordant locations with its *maternal* LUCAS dataset.

The last indicator for determining the degree of agreement between the reference data (producer) and the results of the classification matrix, is Cohen's kappa. This statistical estimate for inter-rater agreement, takes the potential agreement by chance into account [ALB-14]. It reflects the differences between the "actual agreement and the agreement expected by chance" [PDX-12]. The method is considered to be more robust and reliable than just the percentages of user and producer accuracies. The main formula (Fig.62) is quite complex and it is not necessary to describe the arithmetic way exactly at this point. For the sake of completeness however, Cohen's kappa can be characterised as the observed accuracy minus the agreement by chance, divided by one minus the agreement by chance:

$$\kappa = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} \cdot x_{+i})}$$

Fig.62: Equation for calculating Cohen's Kappa (Κ-value) [AAS-13]

Comparing the reference dataset and the geographic citizen science dataset FQA, the calculated Cohen's kappa value is 0.6013. This indicates only a moderate agreement between these two datasets. In other words, there is only a 60% better agreement than just by chance alone. This reflects similar results of the author's findings in the course of his bachelor thesis. A crowdsourced dataset of cropland cover in Myanmar was compared to three established land cover datasets (GLC2000, MODIS2005, GlobCover 2005) and on hybrid map by IIASA. The values back then only were between 0.4 and 0.5, without exception. This was most likely because of the already mentioned disagreement within these datasets [ALB-14].

When focussing only on the three major land cover classes (artificial landscape – A, cropland – B, woodland – C), the value for Cohen's kappa rises significantly, from a moderate agreement of 0.6 up to a substantial agreement of 0.81. This shows that, the more classified sample points are available in the dataset, the higher the degree of agreement might be. These three land cover classes comprise of more than 85% of all classifications and also offer the highest significant values for the user's accuracy as well as the producer's accuracy. The following diagram juxtaposes both these accuracies for each of the eight land cover classes (Fig.63).

It is obvious that the values for the user's accuracy differ widely from the producer's accuracy. The closest data convergences can be found for all classifications concerning cropland (B) with a difference of just 11% and woodland (C) with 19.91%. The highest discrepancies on the other hand are between the user's accuracy and producer's accuracy for shrubland (D, 35.71%) and grassland (E) with 28.26%.

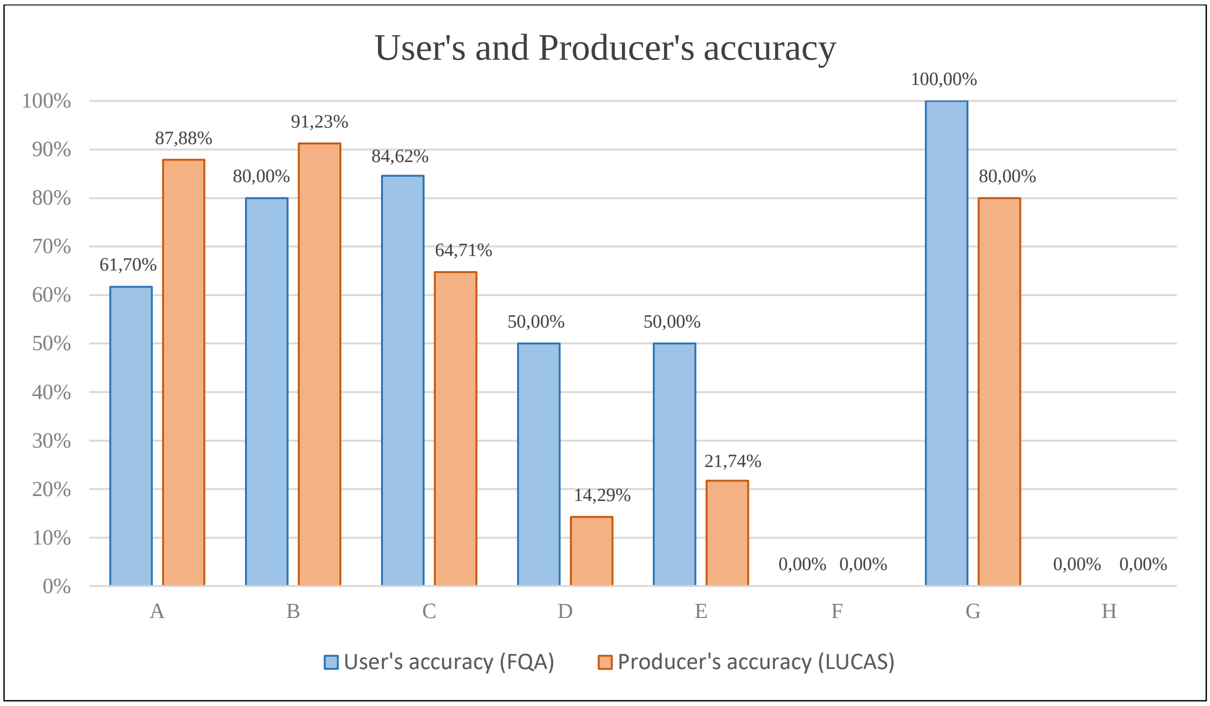


Fig.63: User's and Producer's accuracy of each class

The collective conclusion of this accuracy assessment can be described as indecisive. The overall accuracy (71.33%) indicates a rather good agreement, which indicates the worth of the crowdsourced dataset for this master thesis. However, if someone only relies one this value, it may lead to an incorrect sense of accuracy [PDX-12]. Some of the individual values in both the user's accuracy and producer's accuracy, indicate a worryingly low degree of agreement. The main problem seems to be the low amount of FQA data. Not only the number of unique sample points (n=393) is too low, but also the amount of corresponding reference data for the comparison (n=143). Some other possible explanations for the moderate accuracy are potential mistakes in the crowdsourced classifications or errors in the reference dataset, which was used as the ground-truth. As already implied, a lot of established land cover datasets show considerable disagreements up to 20%. This may also contribute additionally to the moderate data accuracy. Some ways of improving the data and prevent errors in future projects include the selection of more appropriate reference datasets, facilitate the classification process during the in-situ validation and incentivise an increased data generation, so that more usable data could be generated.

A paper by some of the IIASA project team members also covered the comparison for FQA data and LUCAS data [BAY-16]. They focussed however on the complete dataset, covering sample points all over the Austrian territory, and the different levels of hierarchy in both land cover (LC) and land use (LU) classifications. Their overall accuracy is with around 69% quite similar to the 71.33% of this thesis. It is even more than two percentage points lower. This could mean, that the volunteers performed even better in the research area as in the whole of Austria. From this point of view, the data quality on which this thesis is based, may not be as mediocre as it seems. When comparing both user accuracies (FQA) and producer accuracies (LUCAS) for the five major land cover classes of the research area (Danube) and of Austria (Fig.64), it is quite obvious that in almost four out of five cases, the accuracies within the research area (diagonally descending bars) were considerably higher as all over the country (diagonally ascending bars). This could indicate an immanently higher accuracy of this master thesis' data.

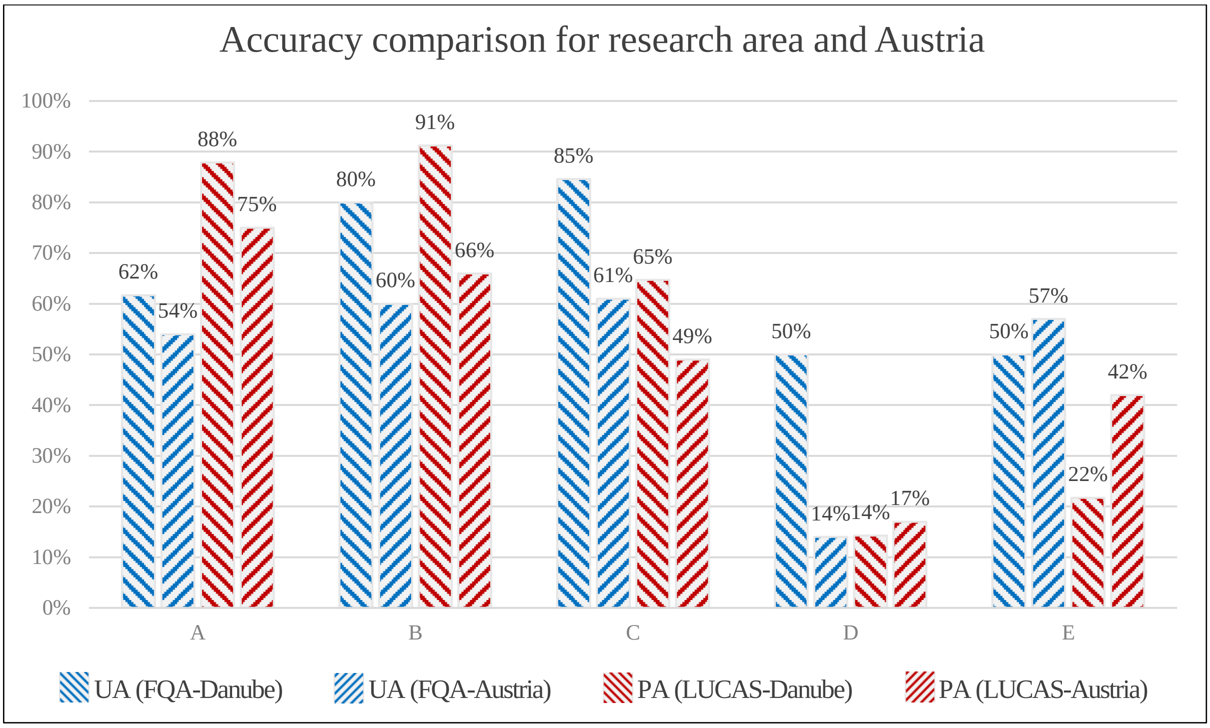


Fig.64: Comparison of both user's accuracy (UA) and producer's accuracy (PA) for the five largest classes of the FQA dataset and the LUCAS 2015 dataset in the research area (Danube) and all over Austria

The published paper by IIASA also compared the lower levels of hierarchy for land cover and land use classifications [BAY-16]. Land cover subclasses came with much more detailed information about the different classes and the prevailing ground cover. The overall accuracies however turned out to be much lower with approximately 37% for second level land cover and around 23% for the third level. This may be caused by a lack of training for the volunteers on identifying different crop types and tree species.

This low amount of accuracy was also one of the major reasons why for this thesis only the more accurate first level land cover classifications were chosen.

After all these general statistics and different methods of calculating accuracy assessments, it would be quite convenient to illustrate the participating sample points on a chart. The following map (Fig.65) shows all the 143 unique locations, used in the previous accuracy assessment between the FQA dataset and the LUCAS 2015 survey results. All green squares stand for total agreement between the classifications of the two datasets and all red square represent total disagreement. As indicated before, 71.33% of all locations were classified consistently, which resulted in 102 (green) agreements. However, around 28.67% of all comparable sample points were classified inconsistently. This resulted in the remaining 41 (red) disagreements.

As the *disagreement map* shows (Fig.65), there is no conspicuous distribution of the two classes. There is apparently no form of clustering or dependency on a specific type of land cover class prevailing. In fact, the distribution of agreements and disagreements appear to be evenly across the whole research area, covering the city of Vienna and its surroundings, as well as the bigger part in the West.

There is however one concern, which has been extensively discussed in this chapter about the accuracy assessment: the number of usable sample points for the comparison. However pleasing the disagreement map and the distribution of the points may look, the significance of the expressed information could be improved by introducing more comparable LUCAS points in this urban area by the European Commission. Nearly 400 sample points were implemented in the FQA campaign by IIASA, but only approximately 36% of them offer a corresponding counterpart in the LUCAS dataset for a comparison.

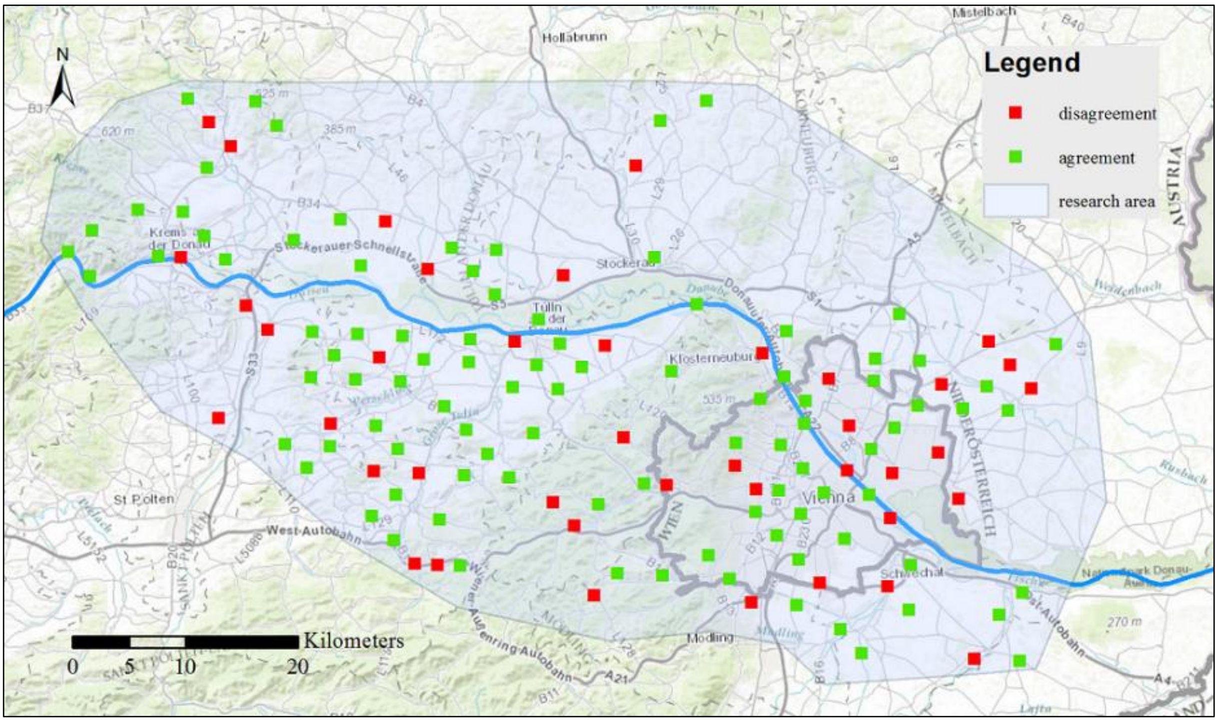


Fig.65: Map of the research area showing agreements and disagreements from the accuracy assessment

This expandable number of comparable data also has effects on further analyses. The following map (Fig.66) shows an interpolation of the previous disagreement map (Fig.65). It was created to provide a general view of the potential situation concerning the disagreements of the two datasets in the research area. The interpolation method was the Inverse Distance Weighted Interpolation (IDW), which predicts possible values for the unclassified areas between the sample points by measuring existing values and providing nearer points with a higher weight. Normally, this method is used in cases, were enough sample points provide a good basis for the calculation, therefore it would be standard for analyses like this. As a result of the rather *manageable* amount of validated unique sample points, the following map is not as significant for the overall analyses as it could have been with a higher density of individual sample points.

The basic message of the map is that the possibility is quite high, that near the points, which disagree in the accuracy assessment, there is also a certain amount of uncertainty. This could be a result of the inaccessibility of certain areas, private properties or a bad connectivity to telephone networks or global navigation satellite systems (GNSS). The map (Fig.66) indicates, that there are no specific topographic or anthropogenic reasons, why the points of disagreement are distributed this way. The corresponding patches do not show signs of a significant cluster or even distribution. Some of the possible explanations, why there are disagreements at all, were discussed extensively in the corresponding chapter. The map should only provide a short thematic overview.

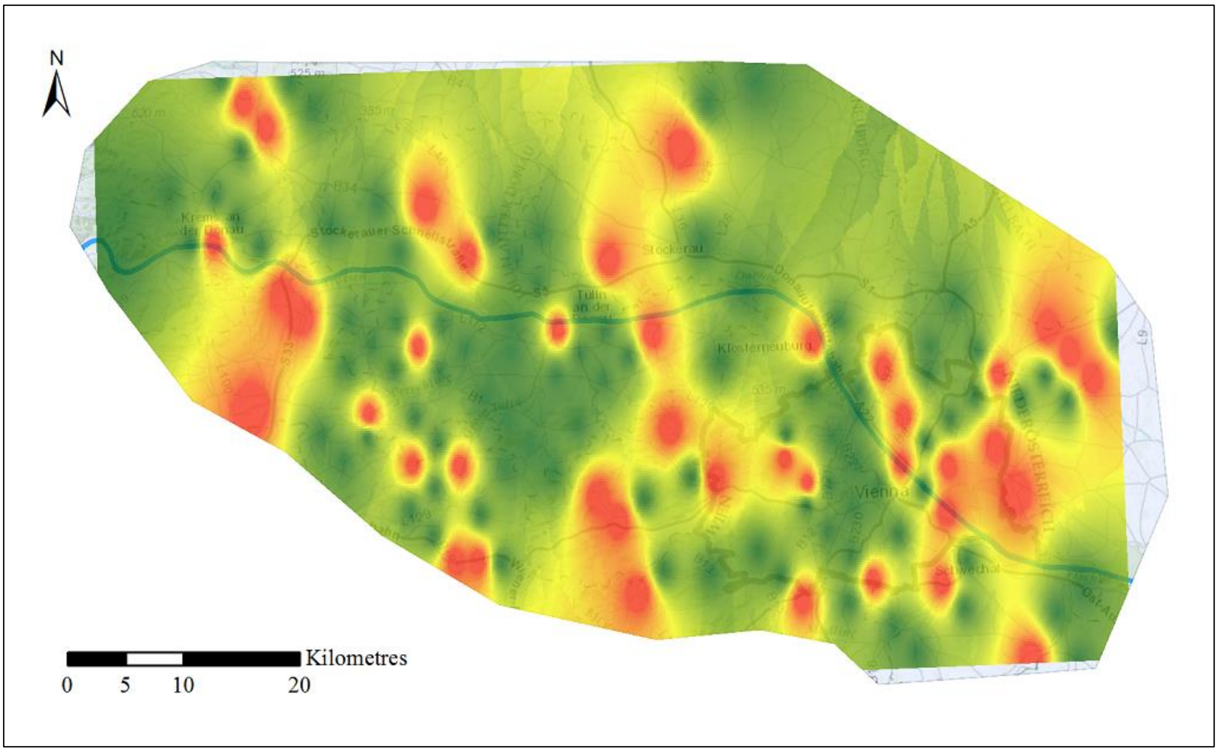


Fig.66: IDW interpolation of the disagreement map in the research area

3.3.3. Quality Management and Issues

After discussing the results and implications of the accuracy assessment between the crowdsourced FQA data and the established LUCAS dataset, there is now the opportunity for dealing with general quality management and accompanying issues. The following paragraphs discuss some of the problems and challenges which appeared in the previous data preparation process and decreased the amount of suitable data in the research area significantly. This includes the classification quality of the individual users, inconsistency within the classifications, some black sheep among the volunteers, changes of classifications throughout the project and uncertainties because of borderline cases.

The previous accuracy assessment indicates that just about 71% of all the compared sample points were classified correctly in the Fotoquest Austria campaign (if LUCAS data is accepted as the impeccable ground-truth). Some causes, like the moderate amount of comparable samples, were already mentioned in the last chapter. Another contributing factor for the rather viable agreement however, may be the inconsistency within some of the validations. Although nearly three-quarters of all sample points were visited only once, some of the repeatedly validated points revealed a collection of partial diametrically opposed classifications. A way of ensuring the correctness of the data, was the random check of 80 sample points. These test samples represented approximately 20% of the complete dataset in the research area. The ID number and coordinates of these test samples were extracted and classified by the author, without the knowledge of the user's classifications in the FQA campaign. This should provide an unbiased approach to evaluate the data's quality and happened either through a local field trip or the interpretation of aerial photographs with the help of two established web mapping services (Google Maps, Bing Maps). Fortunately, the inconsistent sample points only accounted for approximately five percent of all validations and were re-evaluated by the author. If one point was visited five times and offered four different land cover classes, there has either been a change in land cover or the validations were conducted incorrectly. The re-evaluation was executed the same way as the test samples had been classified: via field trips or the interpretation of aerial photographs.

In the FQA dataset, there was also a column which contained the internet links to the ground photos, which the users had to take of the prevailing land cover. In the course of classifying the test samples, some of the pictures were used as a support for an easier identification, in the case of an uncertain interpretation of aerial photographs. However, a few of those pictures were apparently taken in someone's home office. The first thought was, that maybe the locations of some sample points were coincidently in the place of buildings, but this was not the case. The associated classifications showed different class types, from cropland to water areas. After checking other validations, potential LUCAS points and the mandatory coordinate check, with the help of web mapping services, the classifications appeared to be absolutely correct. It seems that this particular user did not actually visit the location of the sample point, but carried out this task at his or her home and uploaded pictures of the home office instead of the prevailing land cover (Fig.67). Because the FQA dataset also included the coordinates of the mobile device during the data upload and some of the pictures were very informative, it was quite easy to figure out the address and name of the *culprit*. However, because the classifications were all accurate, there is no need to expose this black sheep or exclude the contributed information from the dataset. Anyway, this was certainly a questionable and awkward surprise during the data preparation process.

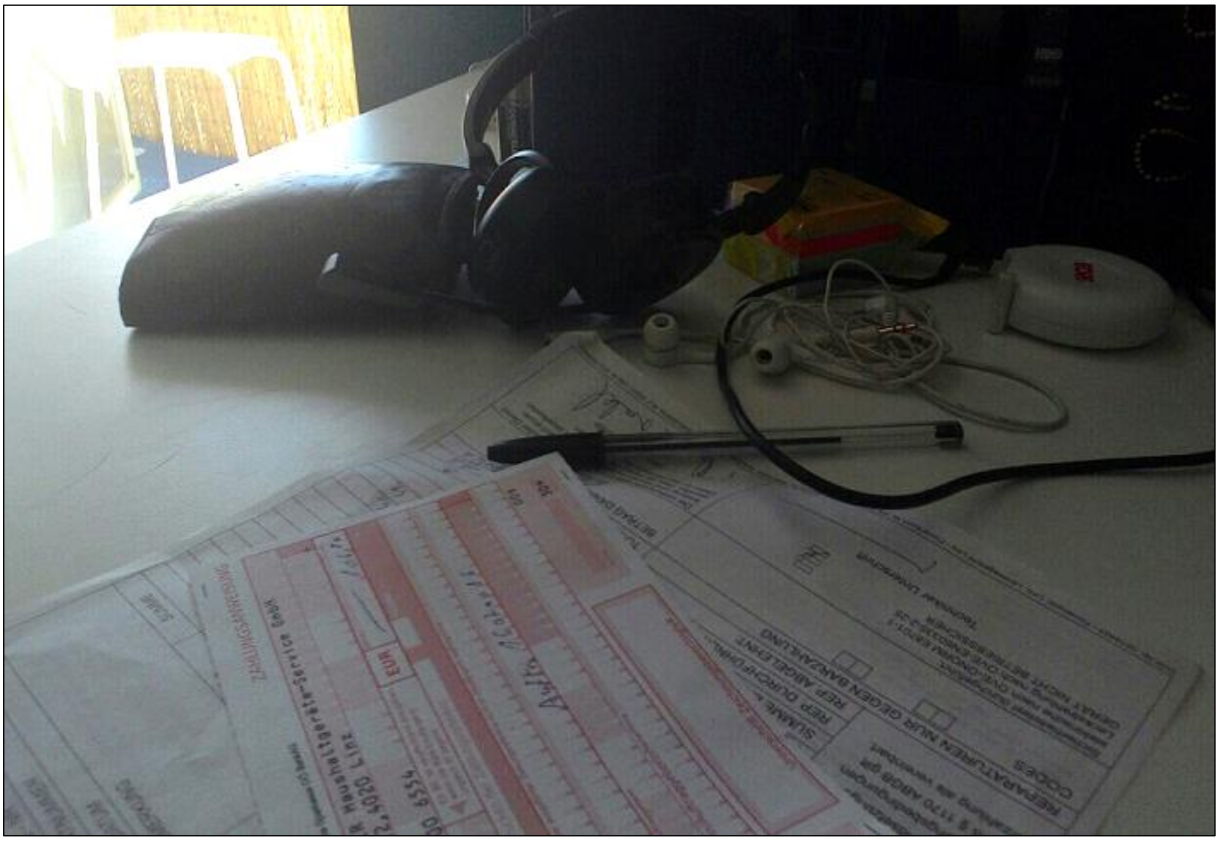


Fig.67: Example for wrongfully uploaded pictures during the FQA campaign

After all the test samples were classified and inconsistent validations were checked and, if necessary, re-evaluated by the author, all the contributing users were tested for their reliability (for obvious reasons). As indicated before, many of the 80 individual users visited only one location and then retreated from this geographical citizen science project. Unfortunately, they did not leave any feedback about their personal reasons. The analysis revealed, that a lot of one-time users offered poor data quality, whereas users with more contributions offer a normal or even high quality. In summary, it seems that most of the inconsistent and inaccurate information during the campaign was contributed by one-time users. This uncovers an interesting weak point of the citizen science project and has to be addressed before launching one of the successor projects.

Another possible explanation for a moderate data quality and some of the raised issues, could be the transformation of locations throughout the duration of the project. It is possible, that some of the multiple classifications of the same sample point are different, although they were classified correctly. It is imaginable that a certain area was grassland (class E) at the beginning of the campaign, a construction site throughout (class F or A) and a building or otherwise sealed area (class A) at the end of the campaign. These comprehensible and understandable discrepancies had to be addressed and adjusted, which happened during the phase of data preparation.

A similar case of undermining the data quality, without sinister intentions, can be found when borders or transition areas were involved. Uncertainties can occur whenever the sample point lies literally on the borderline. Some users may choose one classification, other users may choose the other possibility. This is based on personal experience, natural effects and coincidence. Examples for such situations are areas at the side of the road, transitions between cropland and shrubland or woodland respectively, or even bridges over water areas.

All these listed problems and challenges explain most of the issues with the data quality. However, some of the mentioned quality management methods helped to create a useful and mostly reliable dataset, which forms the basis for all upcoming analyses. As a result of the data pre-processing, the preparation, the quality management and the accuracy assessment, nearly 400 unique sample points with their mostly correct classifications (overall accuracy is nevertheless around 71%) are usable and may remain as a sufficient and legitimate data basis for this master thesis. There will be however a supporting investigation of the research question with traditional methods and based on official and freely available open data.

3.3.4. Supervised Classification from Satellite Imagery

Another way of managing and ensuring data quality, is comparing the FQA data from IIASA with other methods of classifying land cover. This could enable a direct and reasonable comparison as the analyses go on. In this case, the author chose a supervised classification of Landsat 8 satellite imagery. This simple and (relatively) fast method of image classification enables researchers to determine and interpret superficial phenomenon on the Earth's surface. This quantitative analytical method is basically an aggregation of unique picture elements (pixels) in order to feature a specific type of land cover in the investigated image. The workflow was as follows:

The first step was the generation of suitable data, on which the classification processes rely upon. In this case, data from the Landsat 8 satellite was chosen. This eighth instalment from the Landsat Data Continuity Mission is a collaboration between the US-American National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS) [NAS-13]. The satellite offers eleven spectral bands, ranging from the shortwave coastal and aerosol band (0.433-0.453 µm) to the long wavelength infrared band (11.5-12.5 µm). The satellite imagery was provided georeferenced.

After downloading the considerable amount of image data (with a resolution of mainly 30 metres), the raster data was added to the open-source GIS software QGIS. This programme is not only a freely available GIS, but also offers a lot of additional plugins, which facilitate complex tasks. In this case, the author chose the Semi-Automatic Classification Plugin (SCP), developed by Italian environmental engineer Luca Congedo. It offers the possibility of loading appropriate Landsat, Sentinel-2, ASTER or MODIS data in order to perform either supervised or unsupervised classifications.

These two types of classification have different workflows and differ in length. The unsupervised classification does not need human supervision and is based on cluster-grouping algorithms. Such an analysis results in numerous clusters of similar valuation, whereat there is a high possibility, that the same land cover classifications are grouped in separate clusters. This requires subsequently a human identification.

The method used in this thesis however is the supervised classification with spectral angle mapping. This is one of the more commonly used technique for classifying satellite imagery in the field of remote sensing. The trained researcher has to collect training samples for each of the different land cover classes, evaluate them and change them if necessary. Training samples are small polygons created manually, which cover only one class type within a narrow interval of wavelength values. The software then uses these samples to create signature files on average and determine land cover for the whole research area. These training samples, or regions of interest (ROI), are specified with an overall macro-class identification (MC ID) and subcategories called class identification (C ID). This allows the user to perform a rough classification (e.g. cropland, woodland…) of the surface, or a more detailed classification with subcategories (e.g. maize, wheat, birch wood, mixed wood, conifer forest etc.). Based on these training samples, eight land cover signature files were created. The next step, after testing and adjusting (e.g. threshold) the signature file with small-scale temporary classifications, was the application of the classification for the entire image. The output-raster and its land cover classes were coloured in the same eight shades like in the illustrations before. The following figure (Fig.68) shows the result of the supervised classification in the research area, compared to the unique LUCAS points.

An accuracy assessment with LUCAS data as the ground-truth was conducted, similar to the accuracy assessment of the FQA data. The overall accuracy of the supervised classification (SVC) is about 74.83%, when comparing it to the 143 LUCAS points. This is a good agreement, but there is still room for improvement. There are however a number of influencing factors, which potentially affected this value. The already reduced cloud cover was about 5%, bare lands and artificial areas scarcely had similar spectral values and the setting of the classification's thresholds was rather a compromise for receiving the best possible result. Compared to the overall accuracy of the FQA data (71.33%), the supervised classification still offers a higher value (+3.5%).

A comparison of these two methods (land cover classification with the help of untrained volunteers and ground photos, or based on an automated supervised classification of satellite data by trained experts) follows in chapter *4.2 Comparison of Geographic Citizen Science with Automated Supervised Classification*. In this part, advantages, disadvantages, challenges and similarities will be discussed.

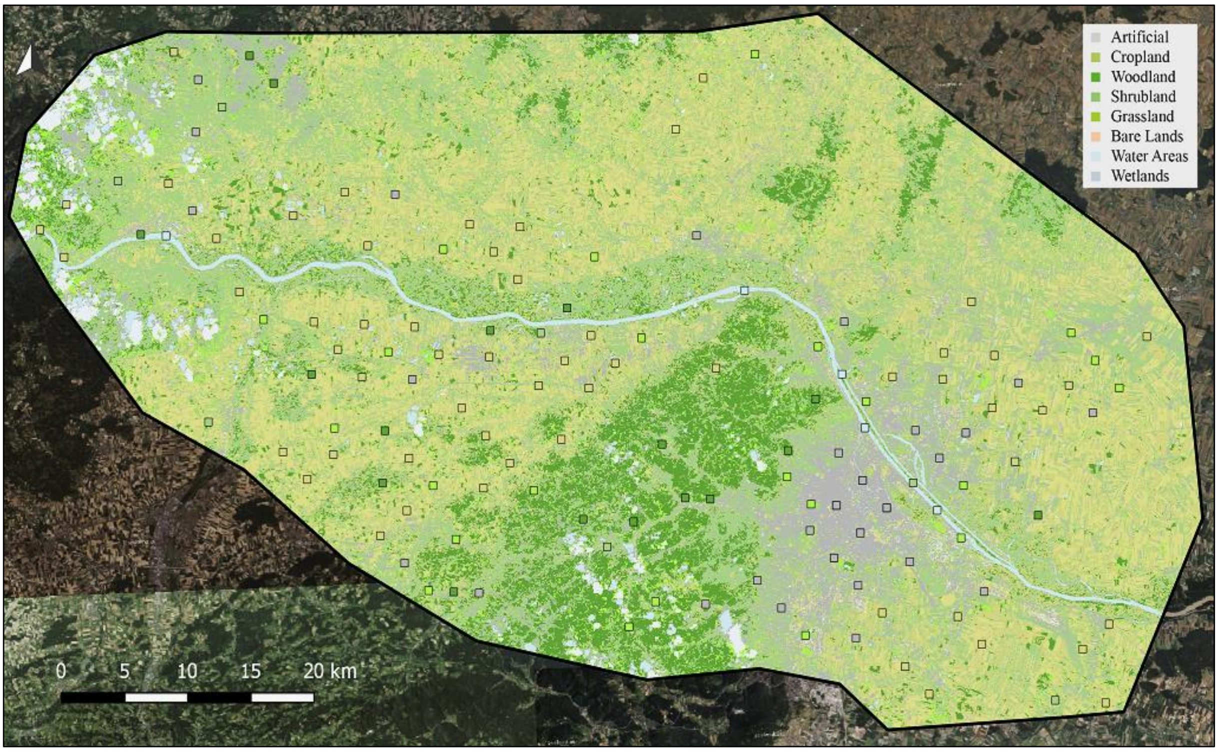


Fig.68: Supervised classification of the research area, based on Landsat 8 satellite imagery with unique LUCAS points

3.4. Methodology and Technology

The first chapter of this master thesis provided an introduction into the objectives of this master thesis and the challenges that come with it. The overarching research question was specified as the determination of potential correlations between soil sealing alongside the Danube River and the extent of the 2013 flood event using geographic citizen science and traditional methods. The main motive was to investigate the capability of the crowdsourcing approach in the field of disaster management and urban planning.

The thematic introduction was based on an extensive literature research, which included resources like scientific journals, papers, articles and online encyclopaedias. This provided the required expertise in order to fully understand the different topics or the connections and interactions between flood events, the sealing of the soil, natural disasters and disaster management, as well as the crowdsourcing approach and its geographic derivatives citizen science and volunteered geographic information.

The third chapter focussed on the acquisition of data with the geographic citizen science project FotoQuest Austria, the LUCAS survey, as well as open data and traditionally generated geoinformation. Data processing provided the foundation for further research and led to the statistical investigation of the FQA data. In order to provide a certain amount of quality, an accuracy assessment with established datasets were conducted and nearly every usable sample point in the research area was checked manually by the author. The creation of an automated supervised classification from satellite imagery provides further opportunities to compare and assess the FotoQuest data.

The following chapter about the final data analysis should answer the major research questions and provide the information for investigating the potential fields of applicability for the used approaches. The methodology and technology, which will be used for these analyses, are grouped into the following categories:

- ~ *Secondary research*: Just like indicated before, secondary research comprises of searching for appropriate scientific literature about the wide range of discussed topics, collecting journals, books and papers, reading and condensing the information in order to compile the most relevant parts of the published resources. This process took longer as the whole data collection. In addition to that, numerous reliable online resources contributed additional and often up-to-date information about these topics. Secondary research provided not only the thematic introduction, but also offered an insight into current research and comparable problems or challenges of the same research direction.
- ~ *Data collection*: The empirical process of collecting the data was extensively described in the corresponding chapter about the FotoQuest Austria campaign. FQA by IIASA focussed on an application for mobile devices like smartphones or tablet computers. Based largely on the

classification and distribution of the LUCAS survey, non-expert volunteers could classify the selected sample points by choosing from predefined categories and taking ground photos. The Land Use and Cover Area Samples (LUCAS) survey on the other hand was conducted by the European Union. Their sample points were validated by trained experts by interpretation of aerial and satellite imagery, as well as by visiting a smaller amount of training samples in the course of a field trip.

Similar to this method, flood discharge areas from 2013 were surveyed by interpreting aerial photography and additional measurements from local authorities. These areas are one of the foundations of the following investigations. Additional information was either obtained from official open data portals (e.g. digital elevation model from Earth-Explorer, electronic navigational chart ENC, building footprints and infrastructure from OpenStreetMap) or provided by federal ministries or provincial governments (e.g. water depths, flood predictions, flood discharge areas from BMVIT).

The creation of an additional land cover classification dataset was conducted in order to check and compare the geographic citizen science data FQA constantly. The automated supervised classification is based on Landsat 8 satellite imagery. The semi-automatic classification plugin in QGIS facilitated the classification of more than four million pixels, using the same eight land cover types as for the FQA campaign and the LUCAS survey.

~ *Statistical treatment*: The data generated during the FotoQuest Austria campaign was almost predestined for quantifying it in numbers and statistics. The previous chapter about the general statistics focussed on the obvious statistical relations between the different data fields. Information about total validations, number of participants, validations per user, validations per individual sample point and reasons for skipping certain locations are just a few examples of the statistical treatment. The accuracy assessment is also a statistical calculation, which helped to describe the accuracy of the citizen science data compared to the established land cover survey LUCAS. Most of these calculations were performed by using Excel software.

In the following chapter about data analysis, some more advanced statistical methods will be applied. It is planned to compare the geographic citizen science data from the FQA campaign with the results of the automated supervised classification (SVC), by juxtaposing the different class type percentages of both classification methods. In order to detect and calculate the prevailing agreements and disagreements of sealed and unsealed surfaces, calculations will be performed by using the raster calculator in QGIS. This should provide a visual and statistical overview over concordances and discrepancies of sealed and unsealed soil between FQA data and the SVC.

The determination of potential correlations between the sealing of the soil along the Danube and the extent of the 2013 flood event is one of the major research issues in this master thesis. There are some appropriate and easy to use methods available. Suitable software for this investigation could be ArcGIS, QGIS, Crimestat, Excel and SPSS. Both methods of classifying sealed surfaces (FQA and SVC) will be compared to the flood discharge area of 2013. Potential statistical methods could include Band Collection Statistics, Standard Correlation, Pearson's method, Covariance, R-squared, Moran's I or Ripley's K.

~ *Additional analysis methods*: In order to investigate the calculated correlations, there will be a traditional approach of generating information about soil sealing. This will include the use of open data like OpenStreetMap and freely available data from official bodies or governmental agencies. The analysis of this information should indicate whether there is a positive or negative correlation between soil sealing and the flood extent, or if there is no correlation at all. By analysing the building density and the road density inside and outside the flooded area in combination with the average terrain elevation, it should be possible to find out whether the geographic citizen science approach is accurate, inaccurate or not suitable at all in this case.

~ *Software technology*: The standard programmes for analysing spatial information in this case are ArcMap 10.5 and ArcGIS Pro by Esri. They will be used for nearly every process like clipping, converting, interpolating, analysing or calculating indices etc., and of course data visualisation. The free software QGIS 2.18.9 is another tool for analysing and processing geoinformation. One of the major purposes of this software in this thesis is the automated supervised classification, with the help of the semi-automatic classification plugin SCP. Another useful feature of QGIS is the OSM extension, which allows the author to download OpenStreetMap data for the research area. For adapting and improving the generated maps and visualisations, the drawing software Adobe Illustrator provides a number of refined image processing tools. For calculating spatial correlations, Excel and Crimestat are leading software solutions. Crimestat is a spatial statistics software, which is mostly used for the investigation of criminal cases and their respective localisation. The programme is not only capable of providing valuable information for law enforcement agencies around the world, but can also be used for the detection of potential spatial correlations.

~ *Justification*: The main reason for choosing these methods of data acquisition goes back to the author's bachelor thesis, which was about crowdsourcing and volunteered geographic information. Its aim was to describe a new method of improving traditional land cover products by using volunteers to generate comparable datasets. Since then, the author stayed in contact with the International Institute of Applied System Analysis (IIASA), which is also the initiator of the FotoQuest Austria campaign. In order to justify this approach, potential strengths and weaknesses will be addressed and dealt with. Potential issues, restrictions and limiting conditions will be discussed and suggestions for improvement will be contributed. The reasons for choosing the methods and software for analysing the data are based on the author's study experience, the results from the bachelor thesis and the methods or techniques used in scientific papers and articles, which focus broadly on the same topical issue.

4. Data Analysis

After a comprehensive thematic introduction into the fields of disaster management, floods and flood protection, soil sealing and urbanisation, as well as crowdsourcing and its derivatives geographic citizen science and volunteered geographic information, followed the more statistical chapter about data acquisition, preparation, interpolation, accuracy assessment and data quality management. These sections are the necessary foundations for further analyses and lead up to the more technical and analytical fourth chapter about *Data Analysis*.

The upcoming pages focus on the investigation of the actual soil imperviousness in the research area alongside the Danube River, a brief comparison of land cover classifications generated by untrained volunteers, with the help of field trips and photographs, and a supervised classification of satellite imagery, as well as the implementation of additional factors like elevation models or permanent and portable flood protection. The final analyses investigate the potential correlations between soil sealing and the extent of recurring flood events according to geographic citizen science and traditional methods.

4.1. Investigation of the Soil Imperviousness in the Research Area

Undoubtedly the most important dataset in this master thesis, the FQA data addresses the distribution of sealed and unsealed regions in the research areas. This information is the backbone of the following analyses and this subchapter pays special attention to the transformation process of all unique classifications and their eight individual land cover classes, to a comprehensive dataset which only focusses on sealed and unsealed areas. The following paragraphs illustrate the process of turning a few hundred rows of information into a land cover map, the interpolation of unclassified areas and the different degrees of soil sealing. A final discussion of the challenges of this method will conclude this chapter.

The first step to a soil sealing map (or impervious soil map) of the research area is to turn the tabulated results from the FQA campaign into an appealing visualisation with all land cover classifications. The map (Fig.61) has been displayed previously in the section about general statistics. All 393 classified sample points represent one of eight different land cover classes. For the following analyses however, only two classes are required, because only the information whether a certain location is sealed or unsealed matters. Class A, or Artificial Landscape, already sums up every sealed location. All the other classes however, symbolise amongst others Cropland (B), Woodland (C), Grassland (E) or Water Areas (G). This classes have to be combined into one single class, representing unsealed areas. The following data processing can only operate with values between 0 and 1. This is why new attribute fields were created, where the value *1* represented sealed areas and value *0* represented unsealed areas. The following map (Fig.69) illustrates these two classes for every unique sample point. The red dots represent all areas where artificially sealed surfaces prevail and green dots all of the unsealed surfaces. The basemap by OpenStreetMap provides the possibility of a superficial and visual assessment of the reliability and usefulness for the now two different classes. It seems that the majority of the sealed surfaces were classified correct, as they appear to be near settlements or other infrastructural facilities. If anything, the sparse distribution and low amount of sample points prevented an even more significant map of the research area.

With this adaption of the original data, it was possible to perform an interpolation in order to estimate all the locations, which were not classified during the FQA campaign. Current GIS software like ArcGIS or QGIS offer a number of geostatistical tools, which are suitable for different analyses or types of data. The most common interpolation methods used in spatial analysis are the Inverse Distance Weighted Interpolation (IDW) or Empirical Bayesian Kriging. IDW uses a weighted average, where close sample points are more likely to show the same class type, than sample points who are further away. This makes it the most appropriate of the available interpolation methods for the analysis. In the name of balance (and for ensuring the best possible results), every available interpolation method in ArcGIS and QGIS was tested and compared to each other. Unsurprisingly, IDW showed the best and most significant result and was therefore chosen as the basis for further analyses.

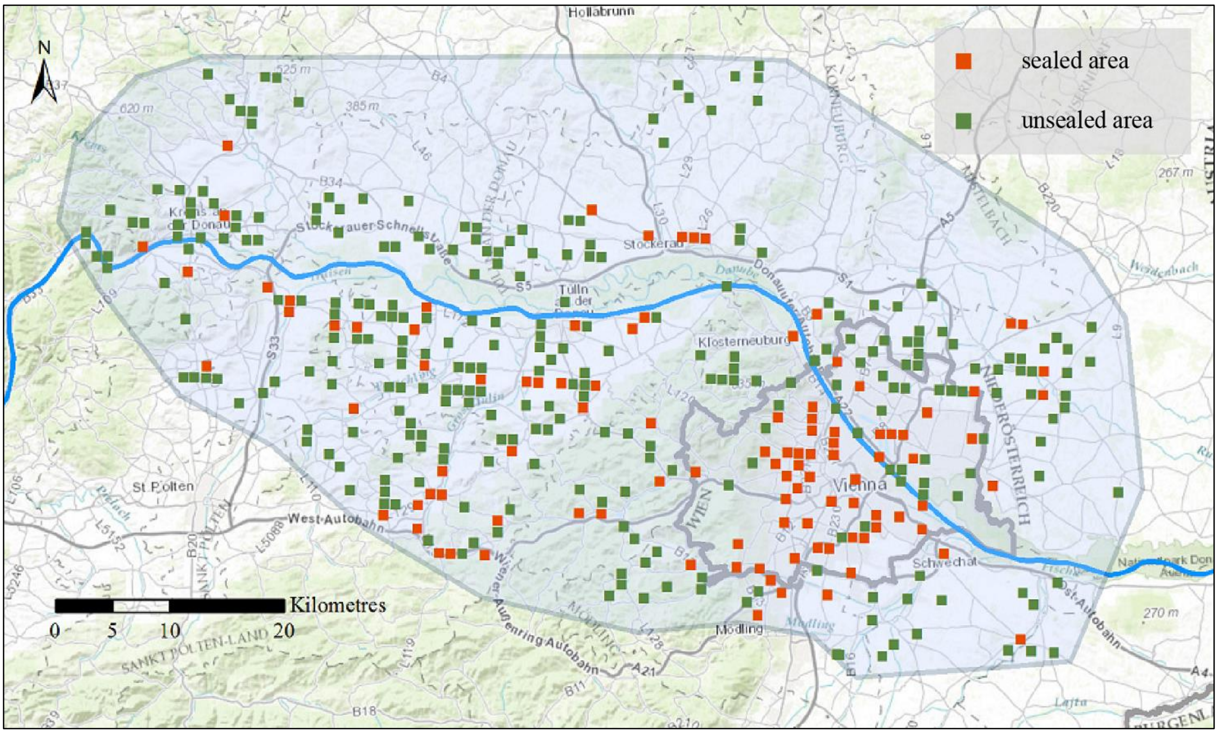


Fig.69: Illustration of all sealed and unsealed classifications of the FQA campaign in the research area

The outcome of this geostatistical analysis is pictured in the following map (Fig.70). The values range from 0 or unsealed area to 1 or completely sealed area. With this type of interval, normally a single-colour scale would be the adequate fit, because the map illustrates the interpolated degree of soil sealing in the research area from very low to high. However, to emphasise the difference between the two main classes, the author used a bipolar colour scheme, ranging from green to red. All the reddish areas offer a high possibility of sealed surfaces. The orange and yellow areas offer a moderate degree of sealing and the green areas are either slightly sealed or even unsealed.

The illustration indicates, that in densely classified areas (e.g. Vienna), there seem to be more accentuated and distinctive interpolated values, than in areas where only a few sparsely distributed sample points were classified (e.g. north of the Danube River). Patches like at Stockerau and its surroundings have quite blurry and simplified boundaries, which may be an indicator for too few or too rough basic information. This yellow areas symbolise an area of uncertainty, where the interpolation method calculated values around 0.5. This would implicate a moderate degree of soil sealing and as a further consequence, this areas could be sealed or unsealed in equal shares. As mentioned before, interpolation methods estimate all unclassified sample points, based on their surroundings. This implies, that the more basic data is existent, the better and more valid are the following interpolation estimations. In this case however, the Inverse Distance Weighted Interpolation used for the following calculations is still the most accurate option currently available.

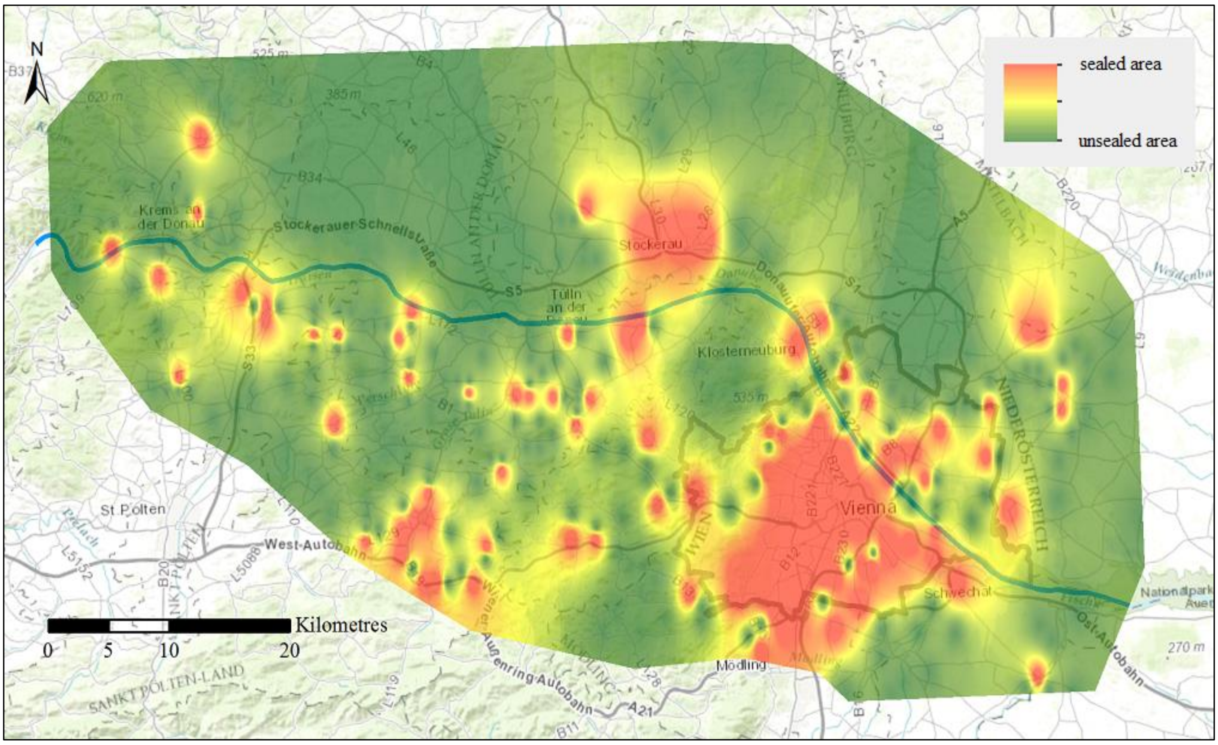


Fig.70: IDW interpolation of sealed and unsealed areas (or interpolated degree of soil sealing)

The following illustration (Fig.71) shows an enlarged section of the previous interpolated map (Fig.70) and covers the city of Vienna with its surroundings. The interpolation overlays satellite imagery of the area with labels, in order to identify city districts and smaller villages. It is obvious, that the built-up area within the city limits have been calculated quite well (at least southwest of the Danube). Also close and frequented suburbs like Klosterneuburg and Korneuburg were classified as sealed within the interpolation. There are however a few disagreements between interpolated map and the satellite imagery. Numerous smaller villages have not been visited during the campaign and therefore do not show up on the soil sealing map. This could be mended by revisiting these points, validate them in the field with the appropriate application and add them to the dataset. This additional effort would require further users or more time and therefore has to be rejected.

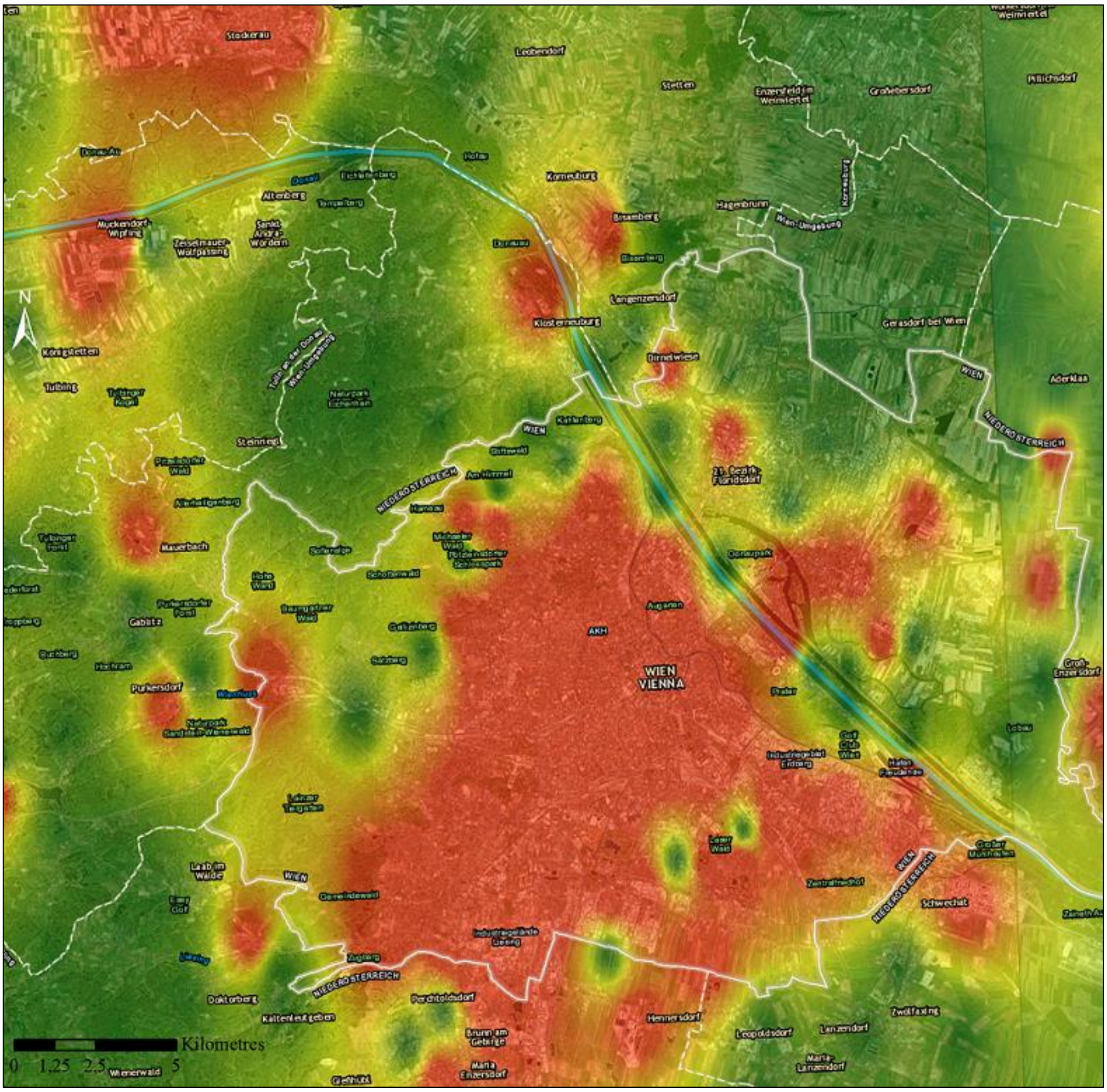


Fig.71: Using the interpolated degree of sealed surfaces as an overlay over satellite imagery with labels as a comparison

In order to receive a distinctive map of sealed and unsealed surfaces (based on the IDW interpolation), the values of the interpolated degrees of soil sealing have to be classified into two classes again. This can be done by adapting the thresholds and building two value ranges. Setting up the right threshold is challenging and requires a certain amount of consideration. In this case, the threshold was set up on 0.5, which means that all areas with a value range from 0 to 0.5 are now unsealed areas and all areas from more than 0.5 to 1 are sealed areas. The higher the estimated value was, the higher was the probability, that the surface was actually sealed or impervious. The following map of the research area (Fig.72) shows the resulting distribution of sealed and unsealed areas alongside the Danube River.

The author is well aware, that this method with interpolating and estimating values contains a certain degree of uncertainty, but in this case it is still the best option available for guessing unknown values. When comparing the following map (Fig.72) to the underlying basemap, it is apparent, that only bigger settlements have been classified correctly, but most of the smaller villages and most of the transport infrastructure (streets, railways, etc.) were not taken into account. This implies that the number of visited sample points in this area should have been higher for a more detailed map.

However, the information about the distribution of sealed surfaces in the research area will be used as the basis for the final analysis. In addition to that, the results of the automated supervised classification, which was conducted by the trained author, will also be analysed further at a later stage, in order to provide a scientific comparison to the results of the geographic citizen science campaign.

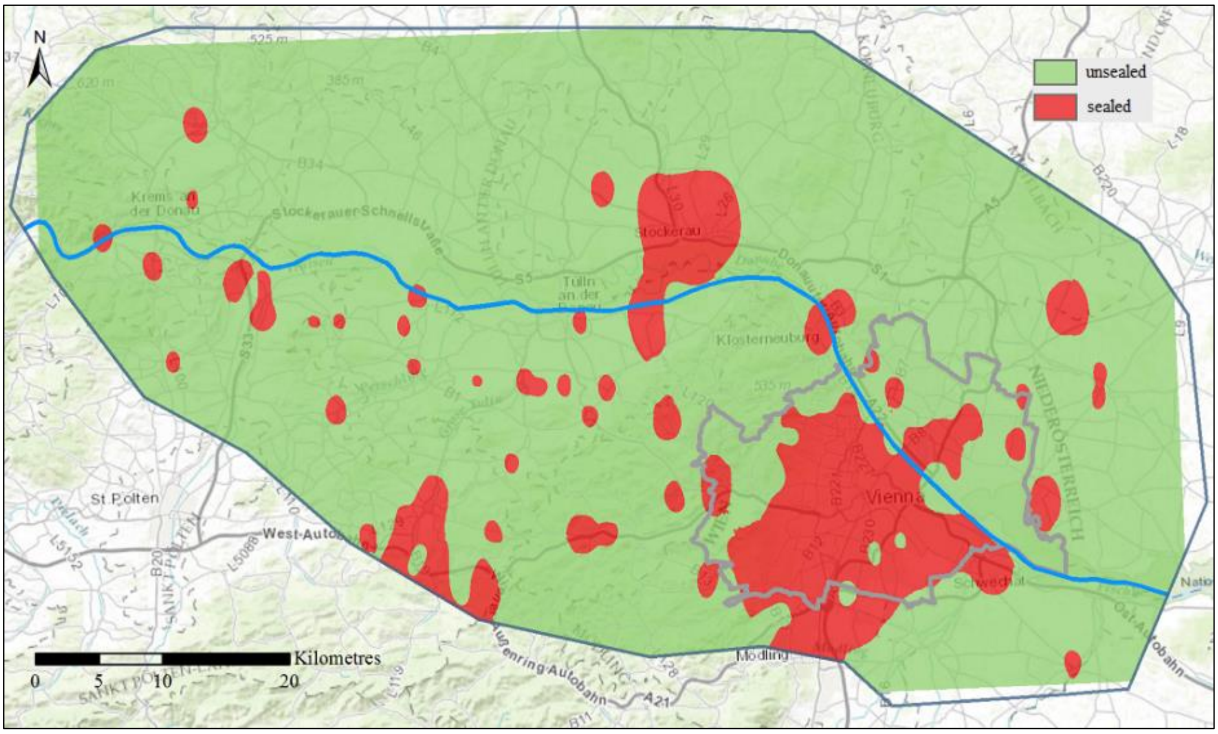


Fig.72: Sealed and unsealed surfaces in the research area, based on the interpolated degree of soil sealing

4.2. Comparison of Geographic Citizen Science with Automated Supervised Classification

The previous chapter focussed on the process of creating a map which only identifies sealed and unsealed areas, based on the data received from the geographic citizen science campaign FotoQuest Austria. With the help of the right interpolation method and the adjustment of thresholds, the previous map (Fig.72) was created. The resulting raster then was reclassified in ArcGIS, with only two land cover classes *0* (unsealed, values from 0 to 0.5) and *1* (sealed, values from > 0.5 to 1) left. This was the first step for the following comparison of the FQA classifications with the results of the automated supervised classification. The latter raster map was prepared in a similar way, but first a short analysis of the two raster maps' land cover classifications.

One of the previous charts (Fig.68) shows the results of the supervised classification conducted by the author. Every pixel was classified by the software, which resulted in exactly 4,138,273 small and unique validated points. Compared to the FQA dataset with its 393 unique sample points, the automated version seems to be a sublime method of determining land cover. There is however only an increase of 3.5% in overall accuracy, with LUCAS as ground-truth (74.83% vs. 71.33%). This implies, that the more extensive supervised classification is only marginally better, than the voluntarily collected FQA data. Another way of interpreting this information is, that there are not enough comparable sample points of the FQA campaign and the LUCAS survey. This definitely compromises the results.

The percentages of the eight different class types from the FQA campaign and the supervised classification (SVC) are compared in the following bar chart (Fig.73). There are a few obvious differences between some of the land cover classifications.

Minor discrepancies can be found for class H (Wetlands). The difference is only around 1.1%, but is mostly caused by incorrect classified clouds in the SVC data. The Landsat 8 satellite imagery showed a minor impairment due to occasional cloud cover in the north-western part of the research area. Similar challenges occurred for distinctive land cover types like Water Areas (class G, difference about 0.8%) and Woodland (class C, difference less than 1%).

There are also some considerable discrepancies retrievable from the statistical evaluation, especially concerning land cover classes A (Artificial Landscape) and F (Bare Lands) or classes B (Cropland), D (Shrubland) and E (Grassland). Already during the creation of training samples for the supervised classification, it became apparent that some wastelands (or bare lands) had similar spectral values as the artificially sealed landscapes. Even the increase of training samples and the adjustment of the spectral thresholds improved this phenomenon only moderately. The percentage of artificial areas of the FQA campaign with 27.95% is considerably higher than from the supervised classification with 8%. The difference is almost 20%. Looking at the values of bare land, it is quite obvious that there is nearly the same difference, only with a different sign. In this case, the FQA classification only shows a value of

1.03% and the supervised classification a value of 19.99%. When adding these two land cover classes with their similar spectral values together, the discrepancy would only be 0.99% (FQA: 28.98%, SVC: 27.99%). These results imply that in the case of Landsat 8 satellite imagery, artificial landscapes have a similar spectral radiance as bare areas, and also that the 393 unique sample points from the FQA campaign are no real match for the more than four million classified pixels from the automated supervised classification. The results also provide a small outlook on the following analysis of potential correlations.

The final minor nonconformity concerns the land cover classes B (Cropland), D (Shrubland) and E (Grassland). Their situation is quite similar to the previous one with artificial and bare landscapes. It is apparent that the percentages show differences between 3.73% and 14.2%. Again, because of their similar spectral values, the vague Semi-Automatic Classification Plugin had its difficulties do distinguish the various forms of (more or less green) vegetation. When combining the percentages of these three unsealed class types for both classification methods, they suddenly seem to be consistent and show a gap of just 0.34% (FQA: 56.67%, SVC: 56.33%). As a result of these three class types being unsealed areas anyway, they do not influence the following analyses and therefore only show the differences between geographic citizen science and the supervised classification of satellite imagery.

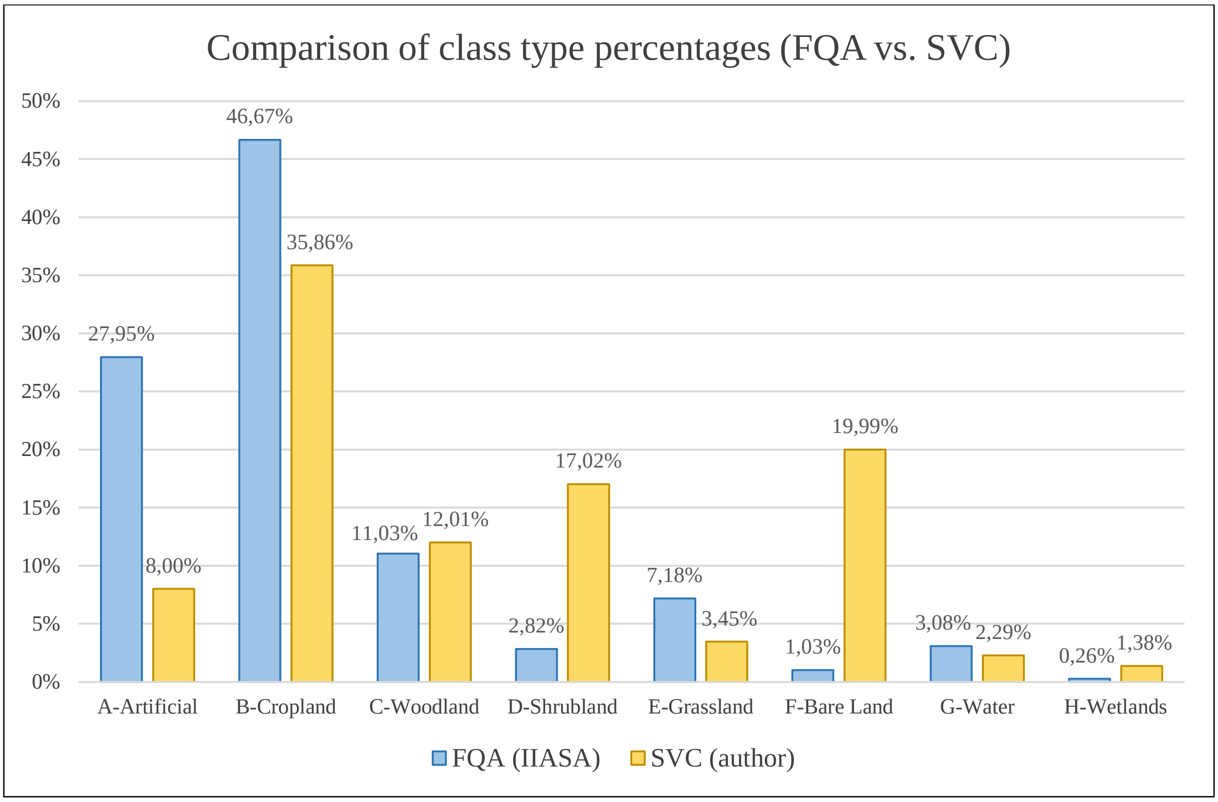


Fig.73: Comparison of class type percentages from FQA data and supervised classification (SVC)

Potential reasons for this phenomenon could be the different sample sizes of both methods, the algorithms of the supervised classification, the residual cloud cover, image resolution and the difference of about six months between the FQA campaign in 2015 and the satellite imagery of early 2016.

The previous paragraphs showed, to a certain extent, the usability of the voluntarily generated land cover information and a comparison with more established methods, like the automated supervised classification (SVC). For the following determination of possible correlations between soil sealing and flood extent and damage, there is only the use for information about whether an area is sealed or unsealed. It is therefore necessary, to transform the eight land cover classes of the SVC in only two required categories. As a result of the comprehensively and completely classified research area, there is no need to perform any form of interpolation. Before that, most of the cloud cover, which sometimes showed similar spectral values as sealed areas, had to be reduced with the corresponding GIS tools. Remaining patches can be found in the north-western corner of the research area (Fig.74). After that, all eight land cover classes had to be reclassified with the ArcGIS tool of the same name. Similar to the results of the FQA data, artificial landscapes (class A) were reclassified as *1* or sealed, and every other land cover type (classes 2 to 8) was reclassified as *0* or unsealed. The resultant map is illustrated as a semi-transparent overlay in the following figure (Fig.74). All the individual sealed pixels were coloured red and all the unsealed pixels were coloured green. It is apparent that most of the research area is unsealed. This covers woodland, cropland, bare land, grassland, shrubland and all the wet areas.

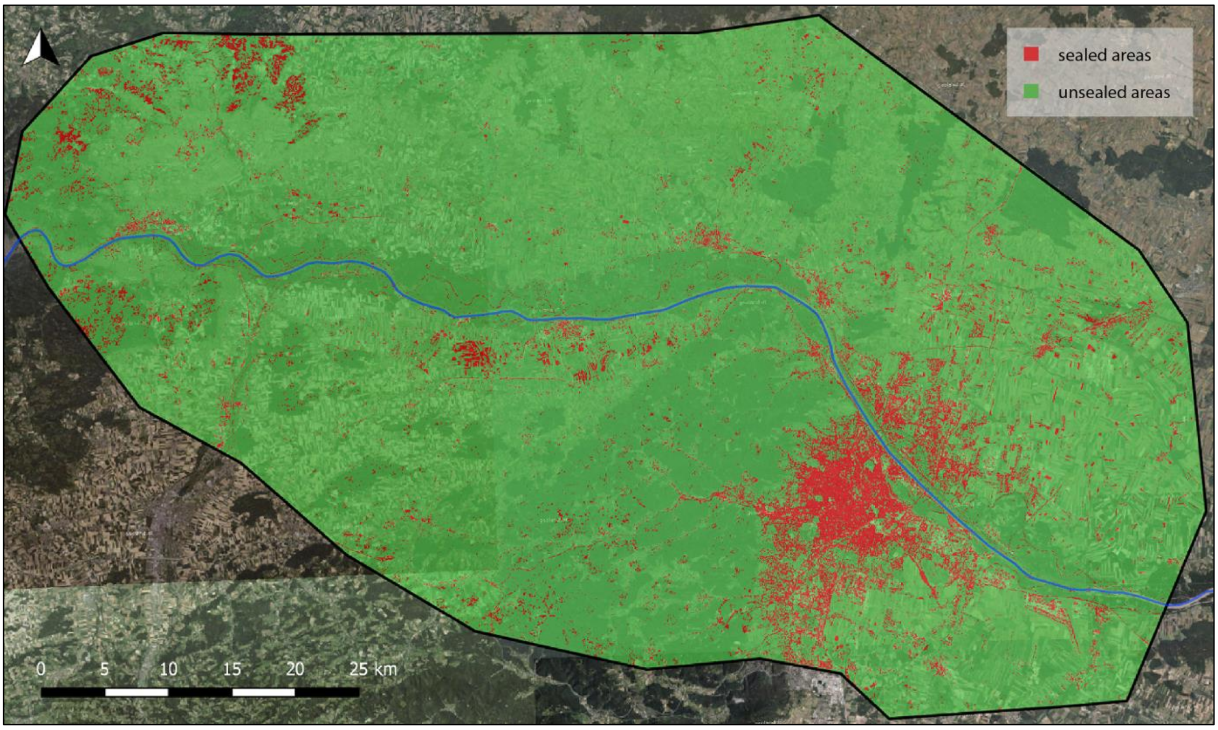


Fig.74: Semi-transparent map of sealed and unsealed areas, based on supervised classification of satellite imagery

The sealed surfaces on the other hand are mostly sparsely distributed in the research area. Bigger patches represent towns and villages, the biggest patch of all covers the city of Vienna. Other smaller dots and patches represent other forms of artificially sealed areas just like roads, bridges and other infrastructural and industrial constructions. Some minor remains of cloud cover were unfortunately inevitable. This spatial distribution of sealed areas, can now be compared to the results of the FotoQuest campaign.

The following two illustrations (Fig.75, Fig.76) offer a visual comparison of the two generated soil sealing maps. The first figure (Fig.75) shows a direct comparison of the interpolated degree of soil sealing, based on the FQA campaign, with the supervised classification's degree of soil sealing in the research area (SVC). As indicated before, interpolations are only mathematical estimations. This method calculated possibilities for the degree of soil sealing for areas, where no classification took place during the campaign. The more classified unique sample points, the better is the interpolated map. As a result of this rough estimations, subtle features and finer characteristics, like open roads or retaining walls, were not detectable. The results of the supervised classification however, offer finer and more detailed classifications. Artificial structures like individual buildings, roads, walls or even the international airport near Vienna are clearly detectable, although the interpolated map may classifies these areas as unsealed (Fig.75). On the downside, there are some minor sealed areas in the countryside according to the supervised classification, which were exaggerated to a certain extent by the interpolation method. These areas may compromise the investigation of potential correlations between soil sealing and flood discharge areas alongside the Danube. In addition to that, there are also the uncertain yellow areas, which may or may not be sealed. Most of these uncertain areas disappear after setting the thresholds and leaving just the regions with the highest probability of soil sealing (Fig.76). These uncertain intermediate areas show only low amount of overlap with the results of the supervised classification, which is why the threshold for the reclassification of the interpolated FQA map was chosen (Fig.75). This could already be an indicator, that for a more substantial and complete investigation of potential correlations, it may be necessary to conduct additional analyses with officially available and open data, in order to be able to provide reliable information about the prevailing situation. Throughout the data preparation, assessment and general statistics, there were already some minor and major indicators that the FQA dataset may not be as significant as traditional datasets from official bodies.

The next figure (Fig.76) is quite similar to the previous one (Fig.75) and shows a direct comparison of the now adapted interpolation map based on the FQA campaign, with the sealed areas according to the supervised classification. After setting the thresholds, only areas from the citizen science campaign with the highest probability of being sealed are left, sealed surfaces from the supervised classification stay the same. Only unsealed areas were excluded from this illustration. It is already noticeable in the map, that there are some agreements as well as disagreements within the research area and between these two methods of classifying soil sealing. All the possible combinations of agreement or disagreement, will be discussed on the following pages.

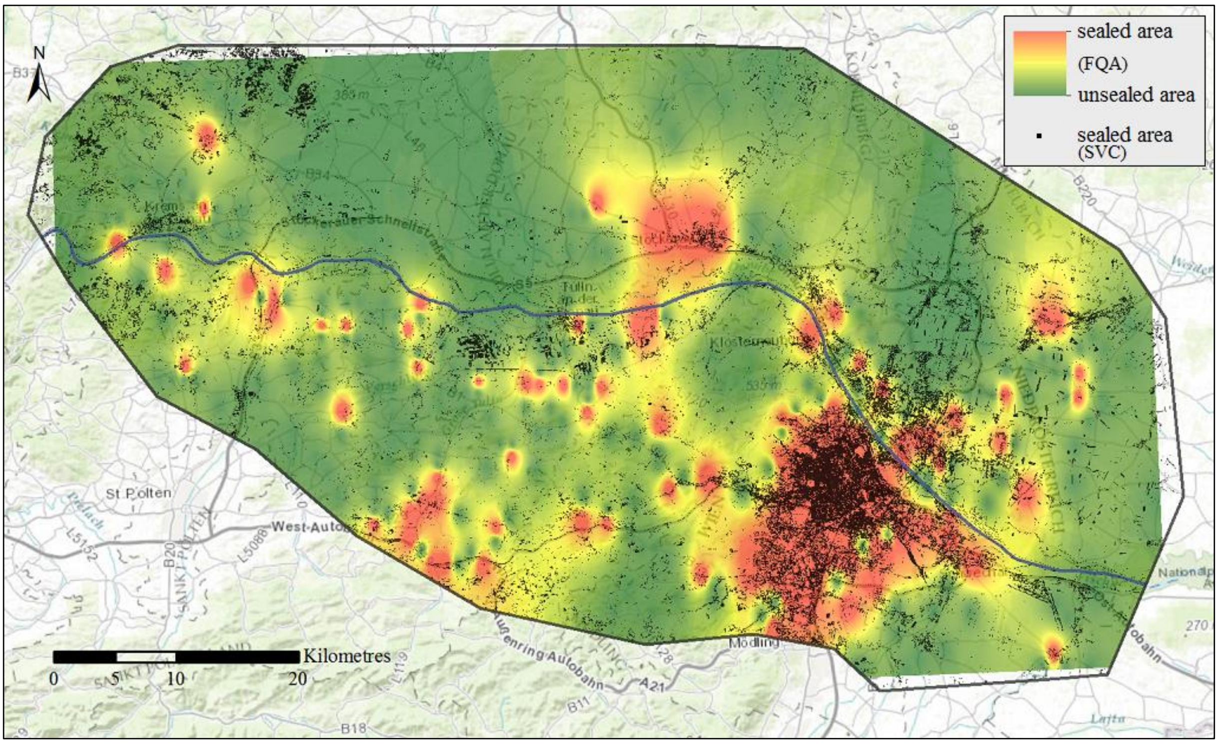


Fig.75: Direct comparison of the FQA campaign's degree of soil sealing and the supervised classification's degree of soil sealing (SVC) in the research area

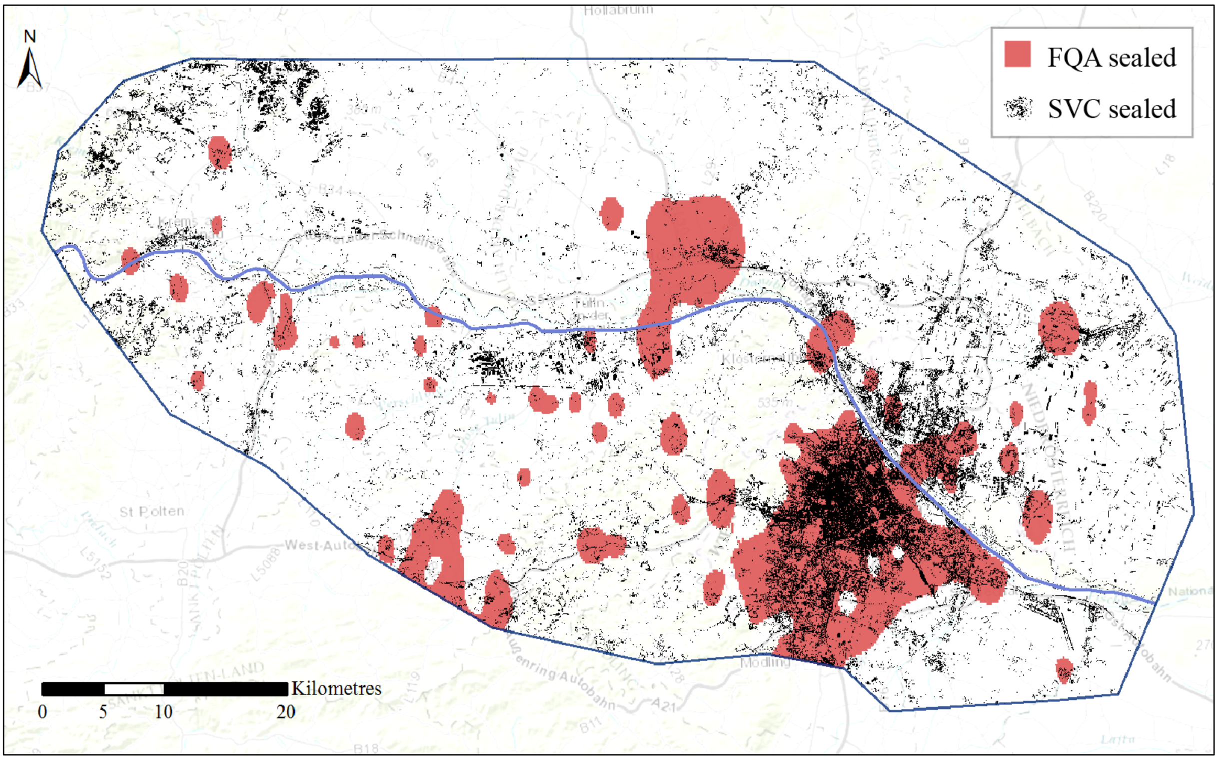


Fig.76: Direct comparison of the adapted interpolated FQA map of sealed areas with the map of sealed areas based on the supervised classification (SVC)

The next step is to compare the two different methods of classifying sealed or unsealed areas (geographic citizen science and automated supervised classification) not only visually, but also with a more scientific approach. In this case, all possible combinations of sealed or unsealed areas according to FQA and SVC. The exact calculation process and some additional statistics will be explained later in this chapter. The following map of the research area (Fig.77) points out all the areas of agreement (in green) and all areas of disagreement (in red) between the interpolated map and the results of the supervised classification. Although there are some substantial and coherent areas of disagreement, most of the research area shows a level of agreement. Later in this chapter, the general statistics of this comparison will offer values of approximately 85% for agreement and about 15% for disagreement. This indicates a substantial and very strong agreement. It means that in more than eight out of ten cases, the classified areas of both soil sealing maps were validated correctly or at least consistently. Based on the initial visual interpretation, this high percentage of almost 85% comes as a bit of a surprise. The consistent areas are mostly unsealed areas in the countryside and the sealed city centre of Vienna. Most of the areas which show disagreement include the exaggeratedly sealed areas of the FQA interpolation. Best examples for this issue are Stockerau in the centre of the research area, the southern outskirts of Vienna and the few remaining cloud patches in the north-western part of the map (Fig.77). However, there are some silver linings when comparing the FQA campaign to the automated supervised classification, at least statistically. The next map (Fig.78) depicts an even more detailed situation of the agreements and disagreements.

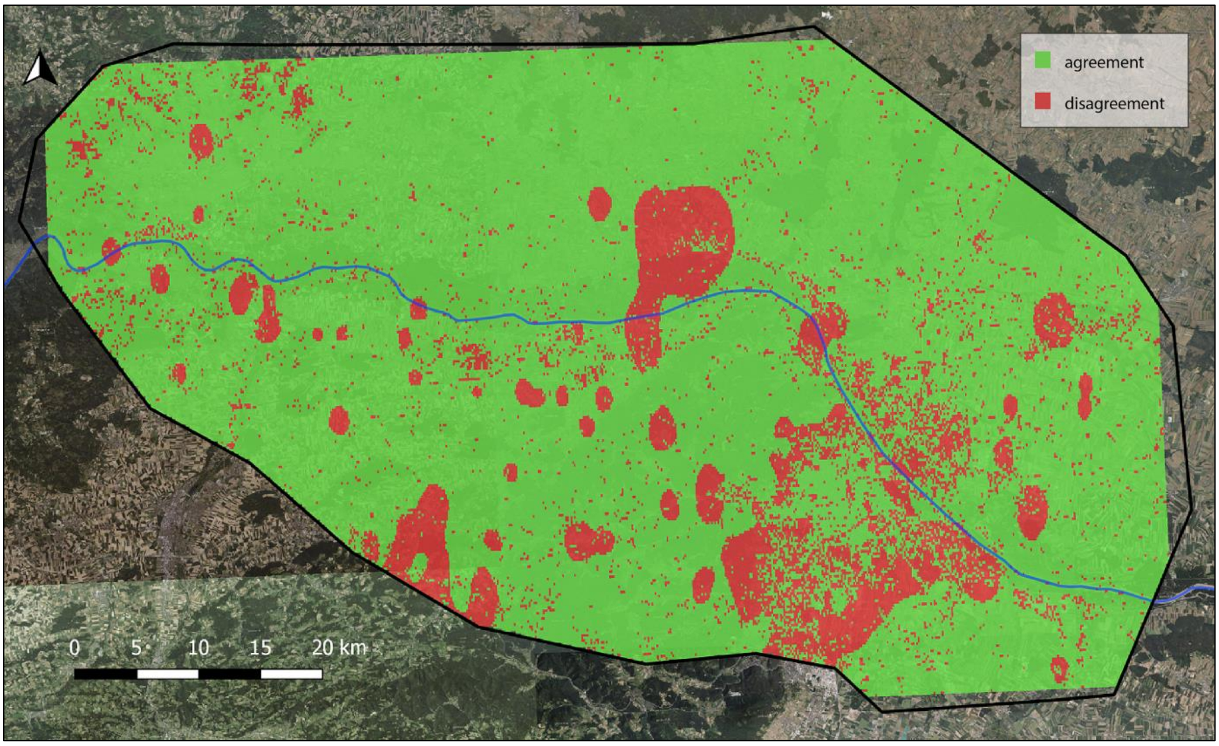


Fig.77: Map of all classification agreements and disagreements between the interpolated map and supervised classification

The following map of the research area (Fig.78) is a more detailed version of the previous map (Fig.77) and shows all possible combinations of agreement and disagreement between sealed and unsealed areas, based on the interpolated map and the supervised classification (SVC). These four combinations show:

~ sealed areas, which were classified consistently in both datasets,
~ unsealed areas, which were classified consistently in both datasets,
~ sealed areas from the interpolated map, which were not accordingly classified in the SVC, and
~ unsealed areas from the interpolated map, which were not accordingly classified in the SVC.

Consistently classified areas are depicted in light green (agreement unsealed) or dark green (agreement sealed), and inconsistently classified areas are shown in light red (disagreement unsealed) and dark red (disagreement sealed). Most consistently sealed areas can be found in Vienna and in the centres of the bigger interpolated (and dark red) patches. This is an indicator that, although the basic situation in these areas indicate sealed surfaces, most of the interpolated sealed areas were oversized (dark red patches). All light red areas represent sample points, which were classified as sealed during the supervised classification, but not in the interpolated FQA map. This includes obvious linear structures like roads or the airport, and polygonal structures like smaller villages. However, the distribution of rightly sealed areas, which are close to the Danube River, seem to be very rare in this map (Fig.78). This raises doubts, if the results of the interpolated soil sealing map are as adequate as the results of the SVC.

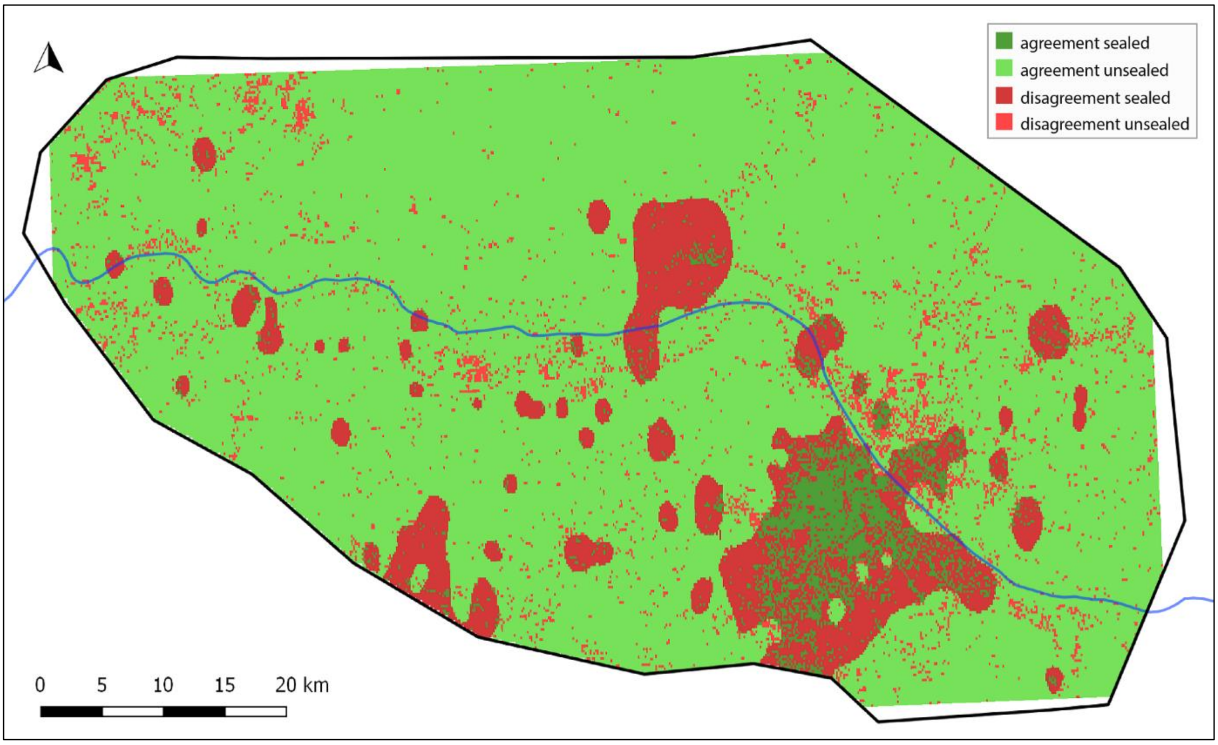


Fig.78: Different combinations of agreement and disagreement of sealed and unsealed areas

In order to be able to provide the previous illustrations, it was necessary to calculate all the possible combinations of agreement and disagreement. To achieve this aim, it was necessary to type the four coded combinations into the raster calculator. The first row of the following equation (Fig.79) yields all sample points, which are sealed ("1") in the interpolated map ("fqa_idw") and unsealed ("0") in the SVC map ("svc"). The second row shows all areas where there are sealed surfaces according to FQA as well as SVC. In the third row only unsealed areas will be displayed. The fourth and last row of the equation yields every sample point which is unsealed according to FQA and sealed according to SVC.

```
( ( ( "fqa_idw@1" = 1 ) AND ( "svc@1" = 0 ) ) * 1 ) +
( ( ( "fqa_idw@1" = 1 ) AND ( "svc@1" = 1 ) ) * 2 ) +
( ( ( "fqa_idw@1" = 0 ) AND ( "svc@1" = 0 ) ) * 3 ) +
( ( ( "fqa_idw@1" = 0 ) AND ( "svc@1" = 1 ) ) * 4 )
```

Fig.79: Formula used to calculate the four possible combinations of agreement and disagreement

This rather simple calculation is a helpful method of revealing some general statistics for the two compared classification methods. The comparison of geographic citizen science (FQA campaign) with the automated supervised classification (SVC) conducted by the author, is not only an integral part of this thesis, but also offers an insight into the particularity and usability of voluntarily generated geodata. The visualisation of the results of the calculation (Fig.79) has been presented on the previous pages (Fig.77, Fig.78). The statistical output follows in the next few paragraphs.

Calculating the four possible combinations of agreement or disagreement of sealed or unsealed areas also resulted in the following histogram (Fig.80). It illustrates the distribution of all four possible combinations. Every row of the calculation produced a class, which summarises the frequency or count of affected sample points. In total there were 124,526 comparable picture elements or pixel sample points. The first class (sealed in FQA, unsealed in SVC) consists of 13,366 pixels, which equals a share of approximately 10.73%. The second class (sealed in FQA and SVC) offers 4257 pixels or 3.42%. The third class (unsealed in FQA and SVC) is the biggest of them all and consists of 101,141 pixels or 81.22%. The fourth class (unsealed in FQA and sealed in SVC) shows 5762 pixels or 4.63%. Most important for this analysis are the classes two and three, because they are the two classes which offer total agreement between the interpolated FQA map and the supervised classification (SVC).

As indicated before, areas which were classified consistently represent the majority of the comparable pixels. About 3.42% of the total research area were accordingly classified as sealed and 81.22% of the research area were consistently classified as unsealed. These agreements equal 84.64%, which leaves a disagreement of 15.36%. This indicates a substantial and strong match of the two differently generated soil sealing maps. This is actually not bad at all, but the significance of it should not be overrated.

One final note on the general statistics of this comparison, is the immense decrease of usable pixels of the supervised classification. At the beginning there were more than 4.1 million unique classified picture elements, but only 124,526 of them were comparable to the interpolated soil sealing map. This represents a drop of approximately 97%. Only 3.01% of the originally classified SVC points were suitable and applicable for the comparison to the interpolated geographic citizen science data from the FotoQuest Austria campaign. The significance of this low percentage cannot be estimated at this point, but in order to ensure a reliable and informative comparison of these two different methods, more comparable sample points would have been better.

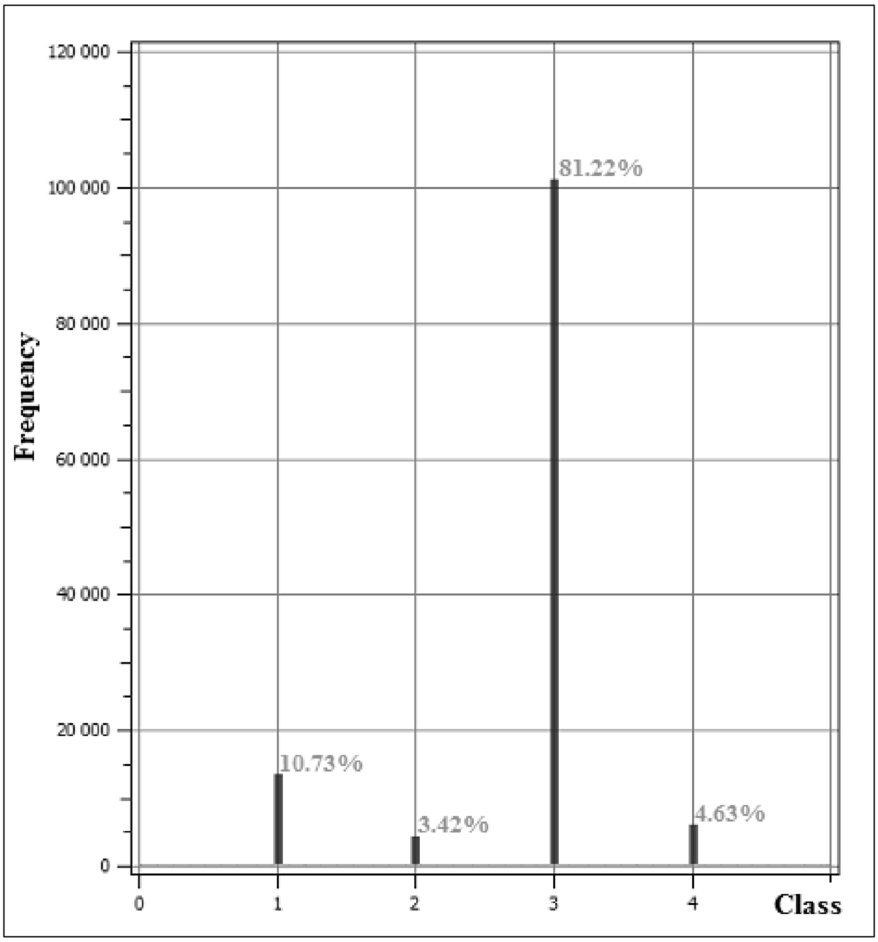


Fig.80: Distribution of all the different combinations

4.3. Soil Sealing and the Correlation to Flood Extent

Until now, all the aspects of the acquisition and preparation of the data, the statistics, accuracy, as well as the soil imperviousness in the research area and the comparison of the geographic citizen science campaign FQA with the automated supervised classification were investigated and interpreted. However, probably the most important part of this thesis is the attempt to determine any potential correlations between the classified soil sealing alongside the Danube River and the extent of the 2013 floods. In the course of writing the past chapters, the author realised that the generated data from the FotoQuest Austria project may not be sufficiently detailed enough, to determine potential correlations precisely. For this reason, the following chapter was divided into two subchapters. The first uses the well-known citizen science results in comparison with supervised classifications and the second sub-chapter focusses on the traditional approach of data generation, implementation and analysis. The aim of all these methods is to determine possible connections between soil sealing in the research area and the discharge area of the 2013 flood event. The traditional approach however uses data from official bodies like federal agencies, planning offices, surveyors or private institutes. This information includes changes in portable or permanent flood protection, which were installed after the great flood in 2002.

4.3.1. FotoQuest Austria and Supervised Classification

The final analysis of this thesis aims to detect and investigate the connections between the distribution of inundated areas in June 2013 and the degree of sealed surfaces in the research area. As a result of the manageable amount of classified sample points during the FQA campaign, missing areas were interpolated and are therefore only estimations, based on their neighbouring surroundings. This led to large circular and drop-shaped patches of sealed areas, which are distributed around the region (Fig.81). A visual comparison of sealed surfaces according to the processed and interpolated FQA data with the official discharge areas of the 2013 flood did not really indicate any correlations between both datasets. Some of the interpolated sealed areas would have been inundated during the 2013 flood (e.g. Stockerau, Klosterneuburg, Korneuburg), but most of the impervious patches are distributed far from the discharge areas. This reduces the usable amount of data from the 2015 FotoQuest Austria campaign once again.

The city of Vienna with its high density of sealed areas is also not really informative, because due to extensive flood protection measures after the 2002 flood event, there were almost no inundations in this area during the more recent 2013 flood event. Although there is a high degree of soil sealing in Vienna detectable, there was no inundation worth mentioning. This area therefore does not contribute any useful information to the analysis of potential correlations.

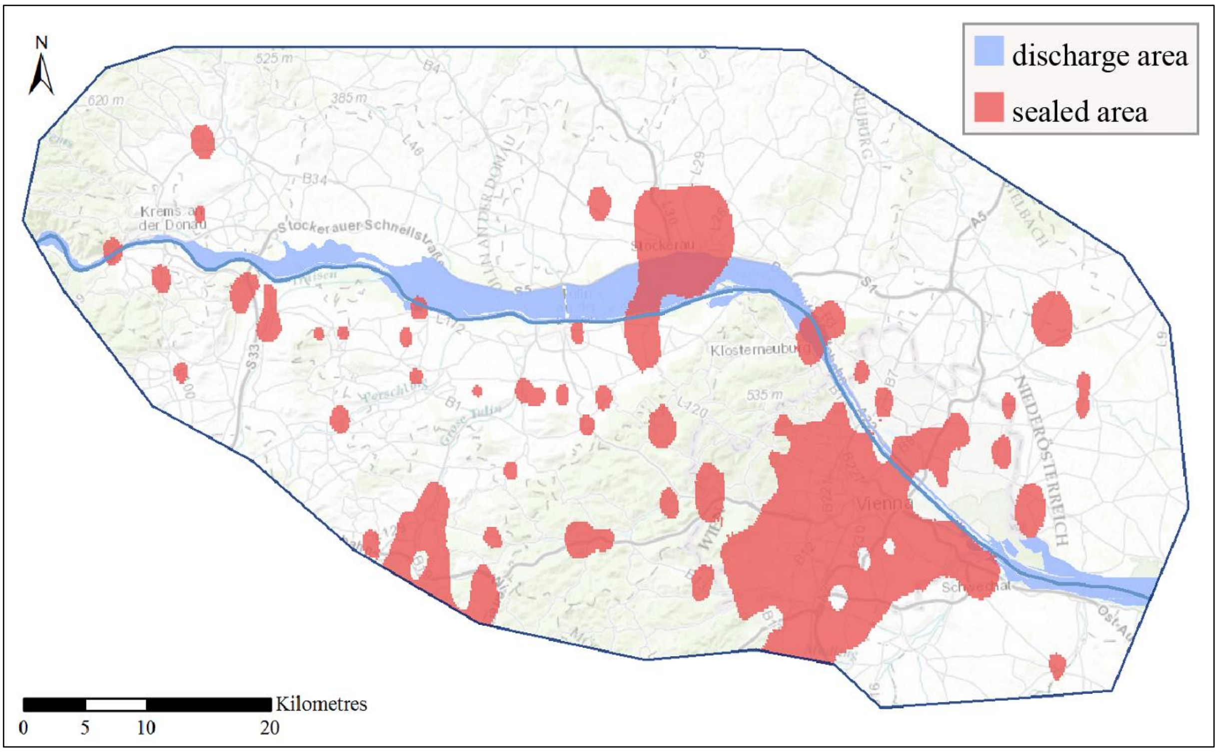


Fig.81: Visual comparison of sealed areas according to FQA interpolation with 2013 flood discharge areas

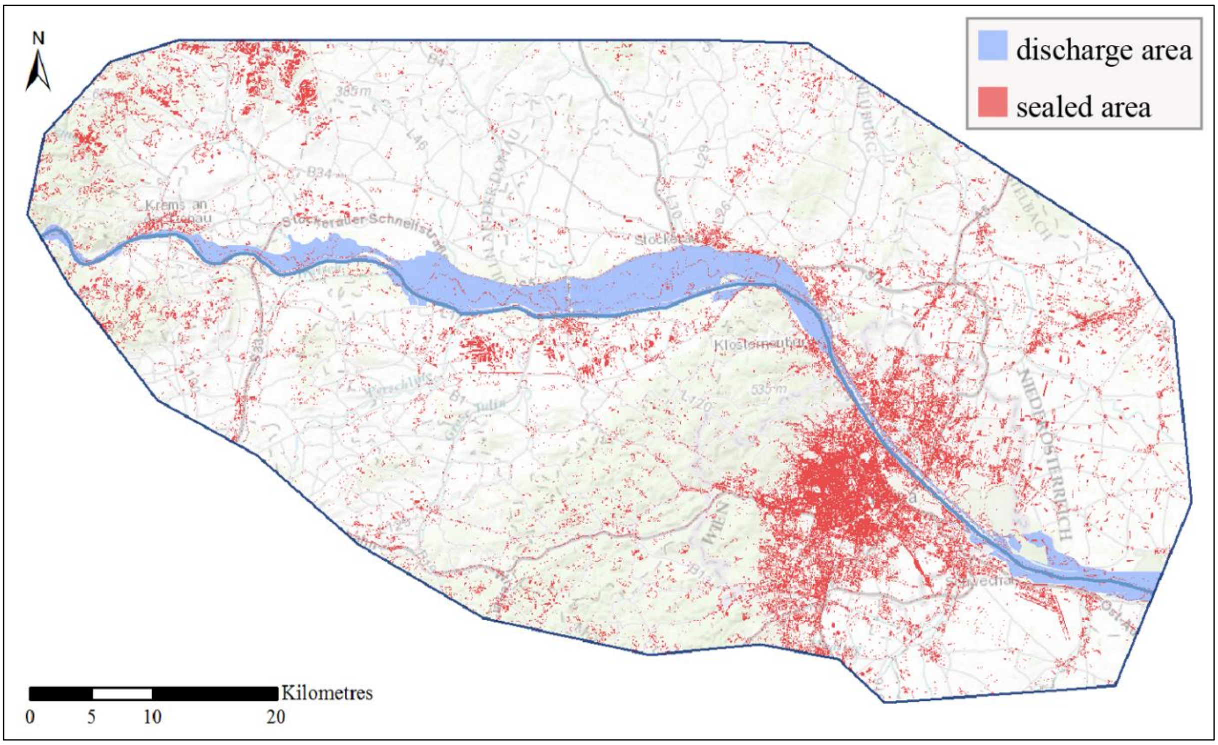


Fig.82: Visual comparison of sealed areas according to the supervised classification with 2013 flood discharge areas

The visual comparison of sealed areas, which are based on the automated supervised classification, with the discharge areas from the 2013 flood event results in a much clearer picture of potential correlations (Fig.82). Towns like Krems an der Donau in the western part of the research area and Stockerau in the centre, are clearly located outside the discharge area. This may be due to the prevailing topography and partly due to former experience and intelligent spatial planning. Comparing this situation to the predicted flood discharge areas (HQ30, HQ100, HQ300; Fig.46), it is apparent that the inundated areas in 2013 mostly coincide with the nearly identical 30-year and 100-year flood events. This led to stricter regulations for spatial planning and restrictions for the land-use plan.

The next figure (Fig.83) shows a magnified section of the previous map, where the sealed area is based on the supervised classification and is compared to the inundated areas. It is quite obvious that most of the affected towns and villages, in this case Stockerau in the north and Klosterneuburg or Korneuburg in the eastern part, are located outside the discharge area. This indicates that, at least in parts of the research area, there is a sense for vulnerability to recurring flood events. In the flood discharge area, there are mostly linear features detectable, which indicate the presence of infrastructural constructions like streets, railways or other connecting routes. The few detached dots, which were classified as artificially sealed surfaces, indicate a very sparse distribution of residential or commercial buildings. These facts are unfortunately not receivable from the FQA data (Fig.81), which can be seen as another indicator that the voluntarily generated information is not as accurate as desired or requested.

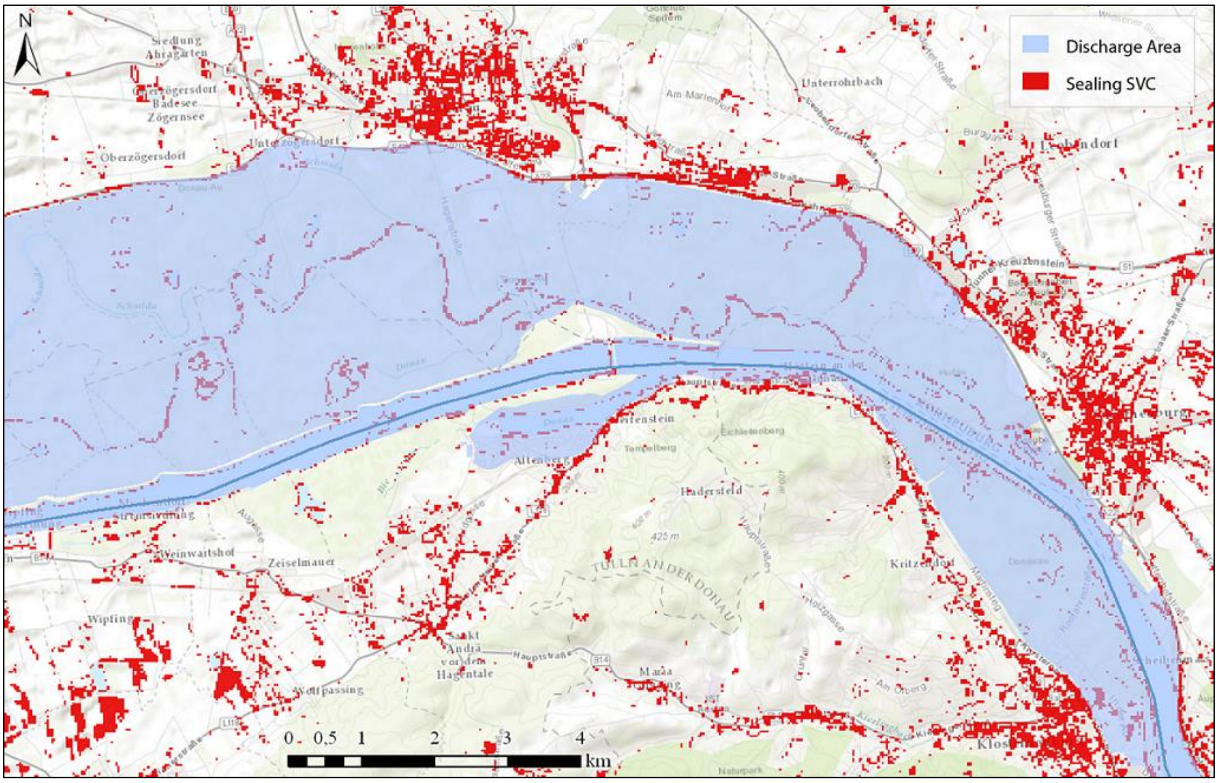


Fig.83: Partial discharge area of the 2013 flood event compared to sealed surfaces based on the supervised classification

In order to receive a conclusive statement about the correlations of flood discharge areas and soil sealing based on the FQA campaign, spatial analyses have to be conducted. There are only a few GIS programmes available, which analyse spatial correlations between two datasets of this kind. The most common software is either ArcGIS (ArcMap or ArcGIS Pro), QGIS or Crimestat.

One simple method of calculating correlation statistics include the spatial analyst tool *Band Collection Statistics* in ArcGIS. This tool calculates general statistics about raster datasets and, amongst others, also outputs a simple correlation matrix. When analysing the flood discharge area (which was converted into a raster-file for this assignment) and the interpolated soil sealing map, based on the citizen science campaign FQA, the resultant correlation matrix offers a positive value of just 4.14%. This percentage indicates only a low positive correlation between the discharge area of the 2013 flood and the sealed surfaces. It basically says, that soil sealing all over the research area only contribute a small percentage to the spatial distribution of inundations. Artificially sealed soils (e.g. streets or buildings) are therefore none of the main drivers of the spatial expansion of flood discharge areas, according to this analysis. However, this value is a result of a number of factors. Firstly, interpolated values are always estimations. The missing spots were estimated, based on the few hundred usable classifications in the research area. And secondly, both raster datasets cover the whole research area and not only the relevant areas alongside the Danube River. As a result of that, every sealed or unsealed location is compared to the vast area, which was not inundated in 2013 due to elevation or distance to the river course. This compromises the resultant value of the calculation significantly. As a further consequence, the next step was to create a buffer around the river course, which includes every flood discharge area and the most significant sealed surfaces of the FQA campaign. The same procedure will be performed for the automated supervised classification.

In order to focus on the more important areas alongside the Danube River, it is essential to clip the interpolated map of sealed surfaces (FQA) to a certain extent. This requires a constant buffer around a specific feature. The used input feature is the digitalised course of the river Danube. Based on this linear feature, a buffer with 4500 metres was calculated. It took some trial and error, but a buffer with this extent covers the most important sealed areas and the whole discharge area. After successfully creating a buffer with full side lines and round end types, it was used as a clip feature for the interpolated map of sealed areas. The output feature now only consists of the sealed areas, which appear only in the buffer area (Fig.84). This reduces the unnecessary values, which have no real effect on the core of the research.

With the former result of the former *Band Collection Statistics* in mind (4.14%), it is now time to conduct the same analysis with the clipped data. The result shows a positive correlation of almost exactly 15%. This means that if investigating only the affected areas alongside the Danube River (Fig.84), the correlation more than triples its value. Although 15% is still very low and considered no correlation worth mentioning, it shows that focussing on the river course and increasing the data accuracy would help to improve the usability of the interpolated FQA dataset. With only this one analysis done by now, it has to be said that the information generated during the geographic citizen science campaign FQA is not accurate and precise enough to determine any positive or negative correlations between soil sealing and the extent of discharge areas during the 2013 flood event. There are however still other methods of measuring the potential correlations.

Another factor who could have influenced the low values of the analyses is the fact, that the government of Lower Austria is one of the most innovative and diligent official bodies, when it comes to install permanent and portable flood protection (e.g. barriers, retention basins and walls). The largest part of the area is in the state of Lower Austria and was therefore investigated after the extensive flooding in 2002. A lot of protective measures have been taken since then, in order to decrease the vulnerability. When there is heavy protection and clever spatial planning, there is a low impact in case of a disaster. The results of this protection measures then could be one of the most influencing factors on these correlation analysis.

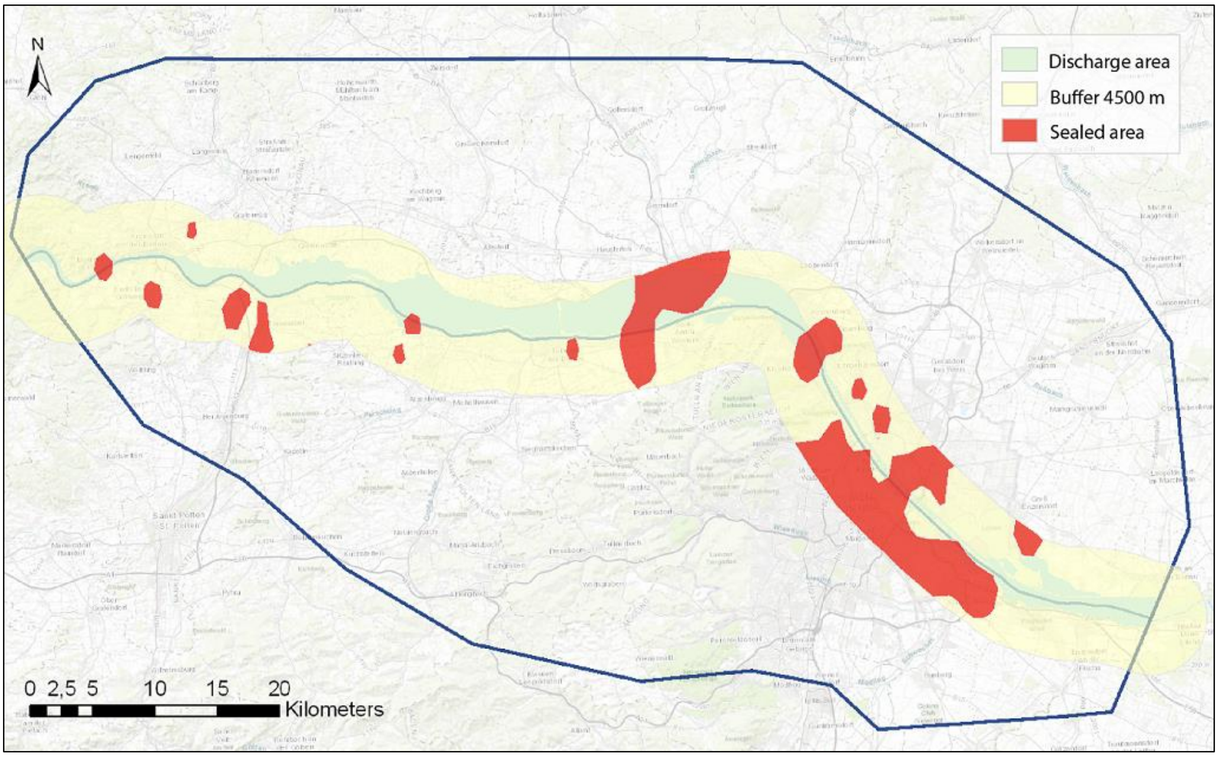


Fig.84: Clipping a 4500 metre buffer area along the Danube with the interpolated soil sealing map from the FQA campaign

In order to continue the investigation to what extent geographic citizen science can complement traditional satellite imagery assessments, the results of the automated supervised classification have to be analysed the same way as the results of the FQA campaign. Based on the experiences from the analyses of the previous data and in order to more comparable, the raster dataset from the supervised classification was clipped with the 4500 metre buffer as well. After converting all the required datasets, both analyses could be conducted. First, the total extent of the supervised classification (SVC) was compared to the flood discharge area. The result is quite different to the result from the interpolated FQA map. There is apparently a negative correlation with a value of -3.15%. A negative correlation basically says, if the affected areas were sealed, the inundation was actually retained to a certain extent. This may be a sign for working protective measures (e.g. permanent flood barriers). The main difference between the correlation values of the supervised classification and the interpolated FQA data, may be due to the raw and vague original citizen science data, which unfortunately did not distinguish between all the different types of sealed surfaces (e.g. streets, buildings or walls). More information about protective measures and river training measures follow in the next chapter about the traditional approach of data generation and preparation.

However, after clipping the soil sealing map, based on the supervised classification, with the 4500 metre buffer, the following raster analysis revealed a negative correlation of -11.44%. This is also more than three times the value, when focussing only on the buffer area alongside the Danube (Fig.85). Nonetheless, it is also a percentage almost not mentionable in terms of statistical correlations. The negative value is, just like mentioned before, an indicator that sealed surfaces in the buffer area actually limited the extent of flood discharge areas. It does not, however, indicate why and if the sealed soil is covered with flat or not interfering streets or three-dimensional obstacles like buildings.

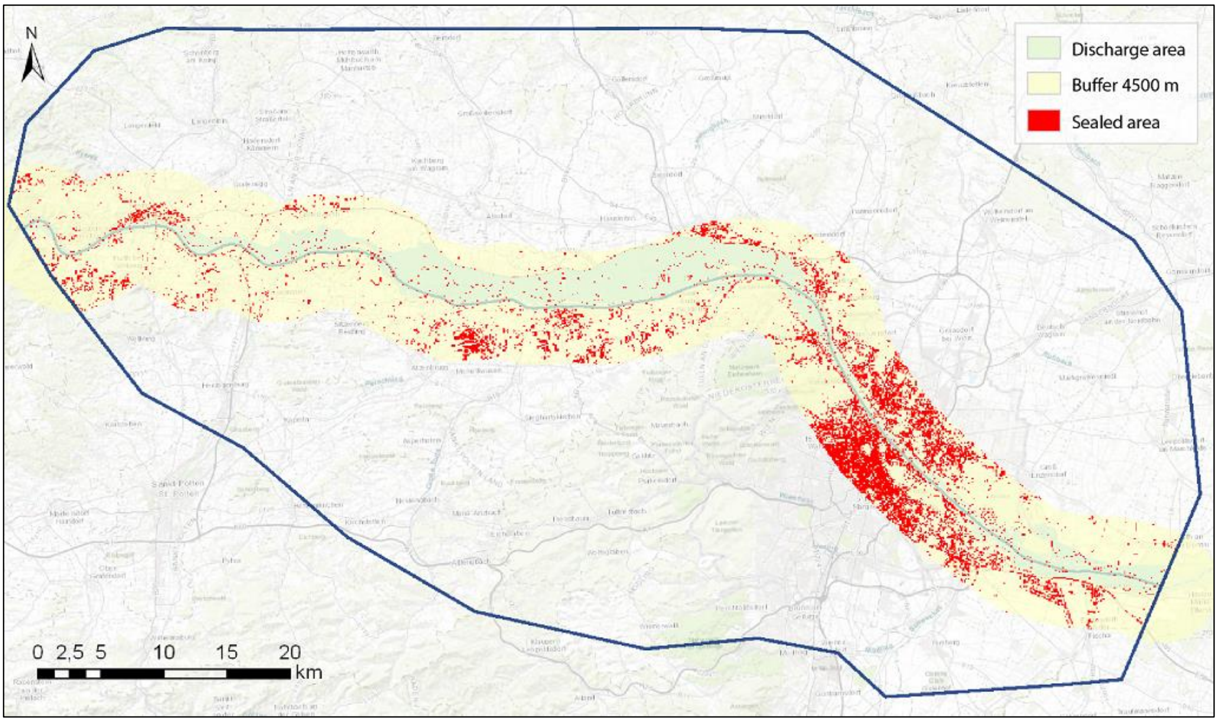


Fig.85: Clipping a 4500 metre buffer area along the Danube with the soil sealing map from the supervised classification

The conclusion of analysing the rather simplified correlation matrices of the raster datasets is, that there are no real common features concerning potential correlations between sealed surfaces and the extent of floods. The following table (Tab.10) summarises all values of the calculated correlations and their changes. It shows the quite remarkable changes when clipping the investigated areas to a smaller extent, based on a 4500 metre buffer alongside the Danube River. The positive correlation of the interpolated geographic citizen science data increases by 10.86% and the negative correlation of the supervised classification decreases by 8.29%. This shows that for a reliable investigation of potential correlations, it is necessary to have an accurate data basis and a reasonable spatial definition. Areas that are far away from the main point of interest may influence the analyses with a number of unrelated and unconnected factors or excluding criteria (e.g. topography or elevation).

Correlation summary	Total research area	Clipped with buffer	Change
Interpolated sealed surfaces (FQA)	4.14%	15%	+10.86 %
Supervised classification (SVC)	-3.15%	-11.44%	-8.29 %

Tab.10: Correlations based on the raster band collection statistics in total and clipped with a 4500 m buffer

The following figure (Fig.86) shows one of the areas, where the most intersections between the flood discharge area and the interpolated sealed soils (FQA) are. It could be another indicator of how imprecise the citizen science data is compared to the satellite imagery based supervised classification.

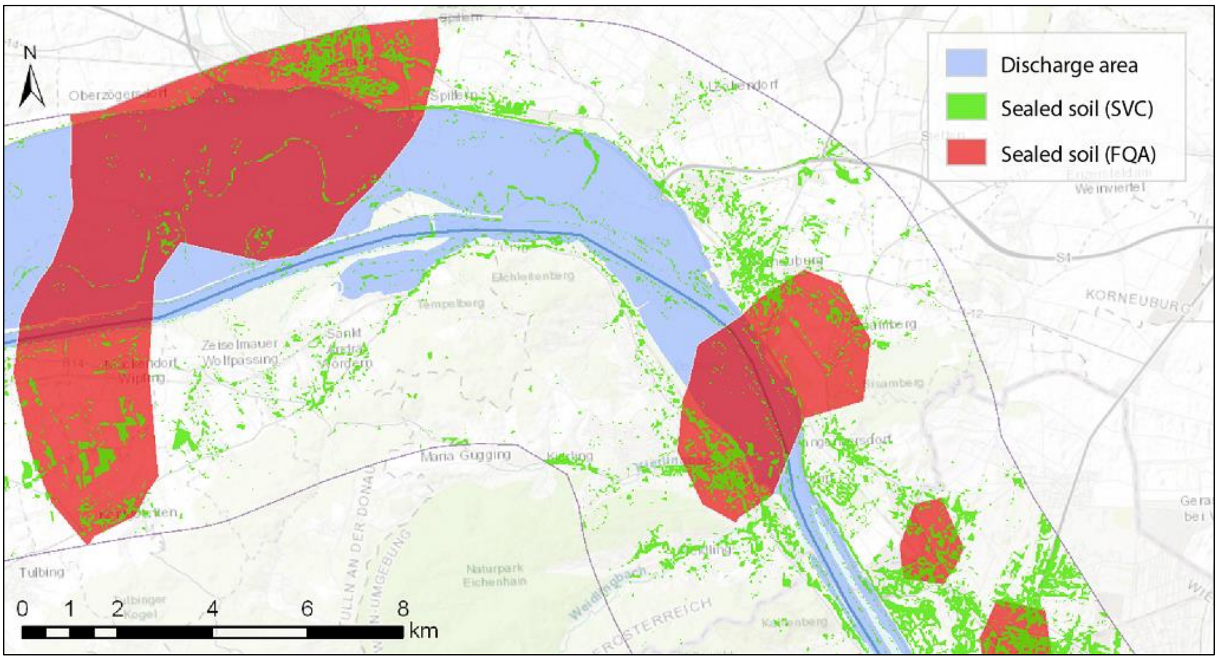


Fig.86: Magnified buffer area with the most frequent intersections between the discharge area and the FQA soil sealing

After an extensive further preparation of the FQA data in the buffer area and a number of classifications and calculations, the final raster was converted into a corresponding point feature. With this processed information, another correlation analysis was conducted. By refining and sharpening the original interpolated FQA data, the positive correlation value increased up to 19.72%. This is approximately 4.72% more than before (Tab.10). It is now quite apparent that by further processing the citizen science data, it is possible to raise the calculated values considerably and strengthen the correlations. In the field of statistics however, all values beneath around 30% are considered to be uncorrelated. The results of the data processing was visualised in the following map (Fig.87). It shows all the possible combinations of sealed or unsealed surfaces with flooded or unflooded areas during the 2013 event in the buffer area, alongside the Danube River. The point feature class was the basis for this calculation of the correlation, based on FQA data.

Relying on the former calculation results, there is quite a strong indication that there are either no real correlations prevailing or the voluntarily generated information from the geographical citizen science campaign FQA is not precise enough in order to determine any potential correlations between soil sealing and the extent of the flood event. As a result of this uncertainty, the next chapter focusses at some of the most significant locations of the research area and discusses the possible reasons for missing correlations.

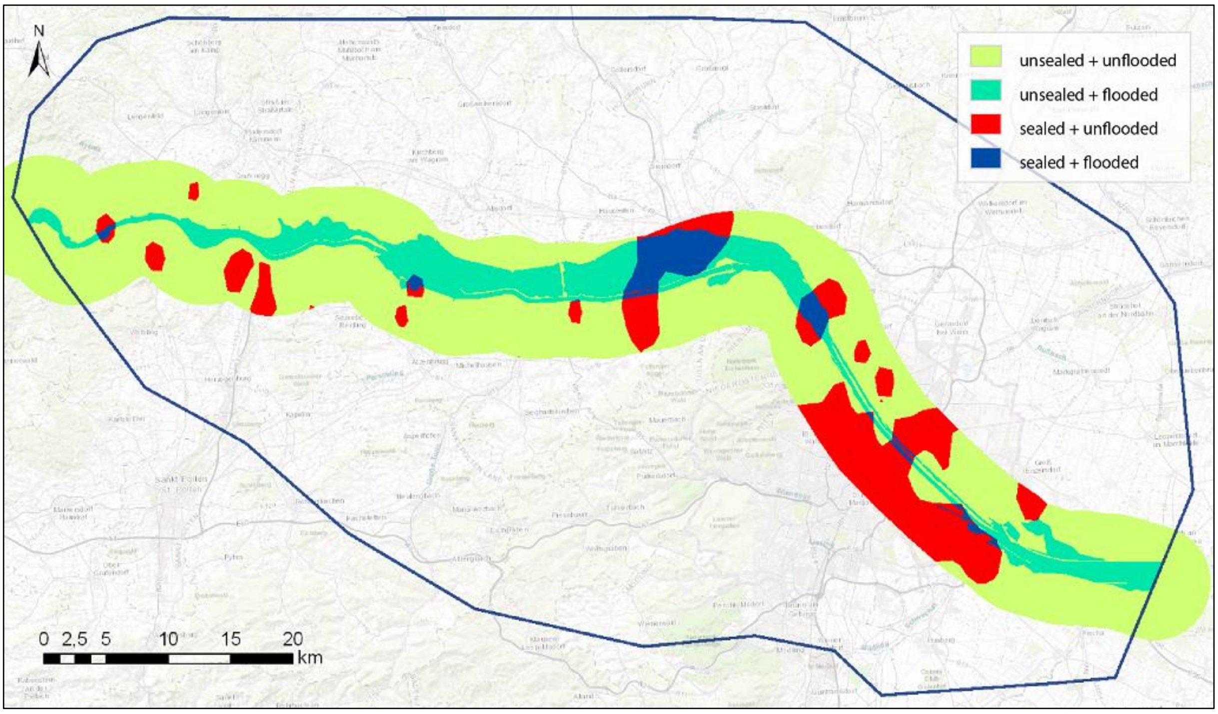


Fig.87: All possible combinations in the buffer area that were used as data basis for the calculation of correlations (FQA)

There are however some other ways of investigating potential correlations between soil sealing and flooded areas. Correlation is defined as the relationship or connection between at least two variables. The resultant values can indicate if there is a causal link between two or more facts or not. There are a number of correlation coefficients, which investigate similar complex issues but with different methods. The following paragraphs provide a brief summary of some of the most common statistical concepts.

In addition to the previous correlations with ArcGIS Pro software (Tab.11), the standard methods of investigating statistical relationships can be calculated with spreadsheet software like Microsoft Excel. After converting the FQA raster dataset into a point shapefile, calculating the corresponding data fields (sealing = S, flooding = F) with binary values (no = 0, yes = 1), the attribute table was saved as a DBF-table. To calculate the correlation, the arithmetic means for the two variables S (sealing) and F (flooding) were computed. These means were subtracted from the original variables to form the new variables S’ (sealing [0/1] – mean) and F’ (flooding [0/1] – mean), which work as the basis for the following formula, with *r* as the coefficient of the correlation:

$$r = \frac{\sum(S' * F')}{\sqrt{\sum S'^2 * \sum F'^2}}$$

The coefficient *r* is the summation of the product of S’ and F’, divided by the square root of the product from the summation of all S’ squared with the summation of all F’ squared. This complex sounding formula returned a value of 1.39% for the data from the FQA campaign (Tab.11). In the world of statistics this means, that there is apparently no spatial correlation between artificially sealed surfaces and the flood extent. The same calculation for the supervised classification (SVC) resulted in a negative value of -12.12%, which is also no indicator for a strong relationship between the two variables. The similar Pearson method also returned the exact same values.

The R-squared method measures linear relationships between dependent and independent variables and is the square of the standard correlation. In this case the value for FQA is 0.02%, which indicates quite strongly that there is no linear relation between soil sealing and flood extent. Although the value of 1.47% for the SVC is a bit higher, it also indicates no linear connection between the two variables.

Covariance is an indicator to which extent the two variables change together. Similar to the previous calculations, FQA shows a very weak positive relationship (0.21%) and SVC shows a weak negative relationship (-1.54%). This still indicates, that there is no real connection between the two variables.

Correlations	Total Raster Area	4500 m Buffer of Raster Area	Point Shapefile Standard Corr.	Point Shapefile R Squared	Point Shapefile Covariance
FQA	4.14%	15.00%	1.39%	0.02%	0.21%
SVC	-3.15%	-11.44%	-12.12%	1.47%	-1.54%

Tab.11: Summary of all correlations based on different calculation methods and spatial extent

4.3.2. Traditional Approach with Open Data

Already sensing that the calculation results of the FQA campaign may not be as significant and informative as possible, the author chose to perform a number of analyses with traditionally generated geodata and freely accessible open data. The results of the correlation calculation in the previous chapter showed, that there is either no sufficient correlation between the sealing of the soil alongside the Danube and the extent of the 2013 flood event, or that the data is not suitable (or precise) enough to determine any connections. Other findings of the previous investigations are the unexpected differences between the values from the FQA data and the SVC. One indicates negligible positive correlations and the other also very low negative correlations. Although the values were not high enough for a significant correlation, the different signs provoke to do further research.

The main aim of the present chapter is to find out whether there is actually a positive correlation, a negative correlation or no correlation at all. If there are other indicators, that potential connections between the extent of the 2013 flood and the imperviousness of the soil could exist, taking a look at additional data like the digital elevation model, water depths, flood discharge areas, hydraulic structures, built-up areas and permanent or portable flood protection could show some of the possible reasons and causes for this situation. A brief examination of exemplary locations should provide an answer to whether the citizen science data (FQA) or the supervised classification (SVC) is right, or if both methods are inaccurate.

This analysis is based on data which is freely obtainable from corresponding internet platforms of official agencies or upon request. Open data like building footprints and road networks were taken from the OpenStreetMap project. This is not only one of the most famous examples for a successful crowdsourcing approach, but also one of the main pioneers of collaborative mapping worldwide. Projects like the Humanitarian OpenStreetMap Team have already proven their usability in the field of disaster management and risk reduction.

Other data was generated in a more traditional way. The following analyses of building densities and road densities inside and outside of flooded areas, are also based on the previously used flood discharge areas, provided by the Federal Ministry for Transport, Innovation and Technology (BMVIT). The investigations after the big flood in 2002, led to an optimal interaction between the interpretation of remote sensing information by trained experts in a GIS environment and the measurements of qualified surveyors on site.

However, both this data resources were used to investigate the density distributions of buildings and roads. These steps were taken to find possible reasons for the correlation values of the soil sealing datasets FQA and SVC. If the following results are significant and informative enough, they may confirm or contradict the previous findings of the correlation computation or at least give a hint about their validity and accuracy.

In order to achieve these high expectations, the OSM data had to be comprehensively prepared. To be fully comparable, all polygon and line features were clipped with the previously determined 4500 metre buffer around the Danube River course. The building footprints were presented as polygon features and after some extensive processing, they were ready to measure the densities. The dataset was split into an area which was flooded in 2013 and into an unflooded area. The following table (Tab.12) shows, that the building density *inside* the flooded areas is 2.18%, whereas the building density *outside* the inundated areas is considerably higher with 29.56%. These values obviously indicate that the concentration of buildings is lower in flooded areas than outside. This may be an indicator that there are already working spatial planning guidelines existent in these regions. The density of the OSM road network seems to confirm this statement. Inside the flooded areas the road density is only 2.03% and outside it increases on a value of 24.68%.

With the results from the calculation of the correlation in mind, it is more or less clear to see, that there is some kind of connection between the degree of artificially sealed surfaces (e.g. buildings and roads) and the extent of the 2013 flood event. Put into even simpler words, if the area was inundated, there were hardly any sealed surfaces. If the area was unaffected by the flooding, the density of sealed surfaces increased significantly.

	Building density *inside* flooded areas	Building density *outside* flooded areas	Road density *inside* flooded areas	Road density *outside* flooded areas
OSM	2.18%	29.56%	2.03%	24.68%

Tab.12: Density of buildings and roads in flooded and unflooded areas based on OSM data

This not only indicates a slight negative correlation, but also contradicts all calculated correlation values for the citizen science data FQA (Tab.11). These values were very low and had a positive sign, which indicates no real detectable relationship between soil sealing and flooded areas, when focussing on the FQA data. The results of the measured building and road densities however, indicate that there is rather a measurable negative correlation. This means that the more sealed surfaces are prevailing in a certain area, the less inundations were measurable. It seems that a higher degree of artificially sealed surfaces in the buffered area (based on OSM building footprints and road networks) slows down the distribution of flood discharge areas or stops it at all.

The standard calculation of the correlation value for the combination of OSM building footprints and road networks (forming a simplified dataset of sealed surfaces in the buffer area) revealed a relatively strong negative correlation of -31.14% (buildings only: -8.90%; roads only: -33.09%). Although implicating a rather moderate negative correlation, this reliable and established dataset therefore confirms the previous indications from the building and road densities. A low density of sealed surfaces (OSM) *inside* flooded areas and a high density *outside* flooded areas stand for a negative correlation, where obviously a high degree of artificially sealed surfaces prevent the distribution of flood discharge areas. The negative correlation value of -31.14% also somehow indicates, that the soil sealing map from the supervised classification may be more accurate and valid than the FQA data, because this dataset derived from satellite imagery already offered low negative correlation values for nearly every calculation method (Tab.11). It is therefore possible to say (very cautiously), that the degree of preciseness of the geographic citizen science project FQA may not be high enough for answering the quite detailed research question. There is however the consideration that after the extensive data processing, all the data from OpenStreetMap also have a certain range of variation.

Similar to the previous calculations of building and road densities, the same method was used for the two different classification methods FQA and SVC. The results are listed in the following table (Tab.13). Like indicated before, the soil sealing density inside the flooded areas for SVC is 3.65% and outside these areas 13.26%. Although maybe not as distinctive as the OpenStreetMap results (Tab.12), there is still a similar tendency noticeable.

The findings for the FQA data on the other hand, confirm the discrepancy, which was detected earlier on (Tab.10). According to the analysis, the sealing density inside flooded areas with 18.69% is even higher than the density outside flooded areas (17.34%). Again, these points imply a very low positive correlation. It basically says, that there is a higher concentration of buildings in regularly inundated areas, than in unaffected regions. A possible explanation for the rather high value for the sealing density inside flooded areas (18.69%), could be the considerable uncertainty within the data due to the performed interpolation. This mathematical method is more or less a rough estimation of sample points and areas, which have not been validated throughout the data collecting campaign in 2015.

	Sealing density *inside* flooded areas	Sealing density *outside* flooded areas
FQA	18.69%	17.34%
SVC	3.65%	13.26%

Tab.13: Soil sealing density in flooded and unflooded areas from FQA and SVC data

The following figures (Fig.88-91) represent the visual representation for the line and point densities of the OSM building footprints and road networks, inside and outside the flood discharge area from 2013. They are based on special analysis tools from the ArcGIS model range, like line density and point density. The basis data was appropriately prepared and converted if necessary.

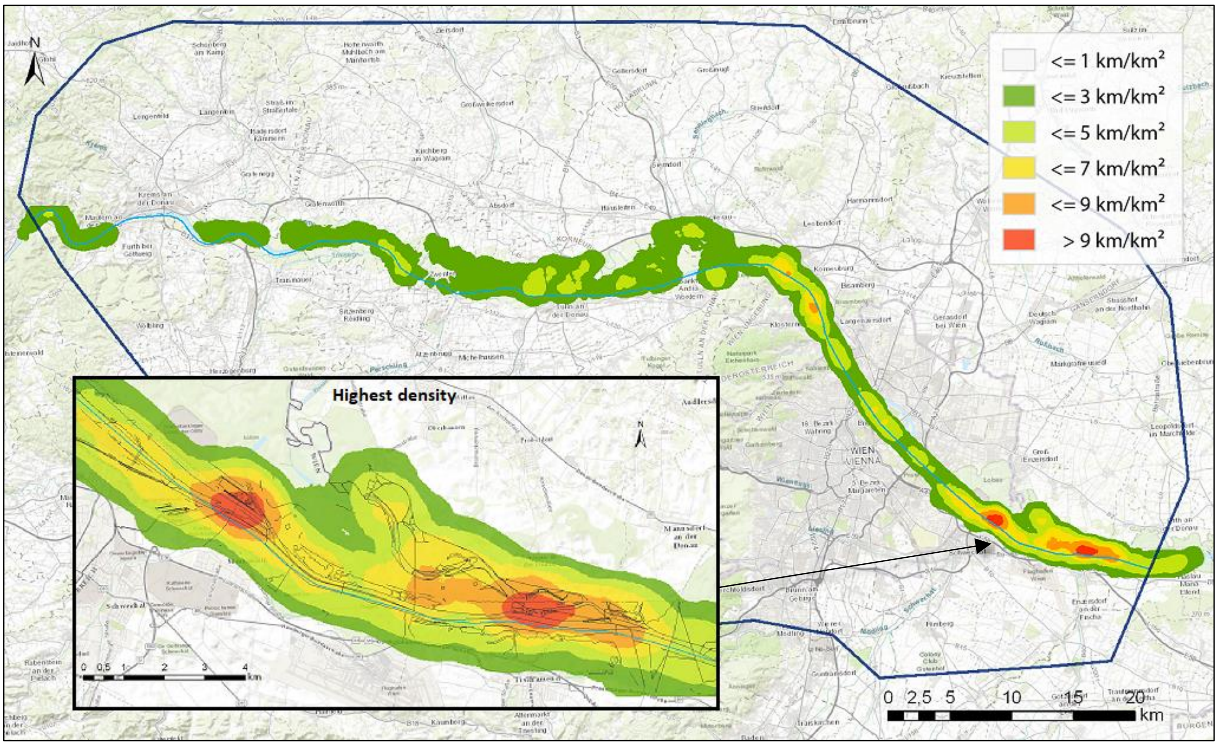


Fig.88: Line density from the OSM road network in flooded areas with magnified peak area

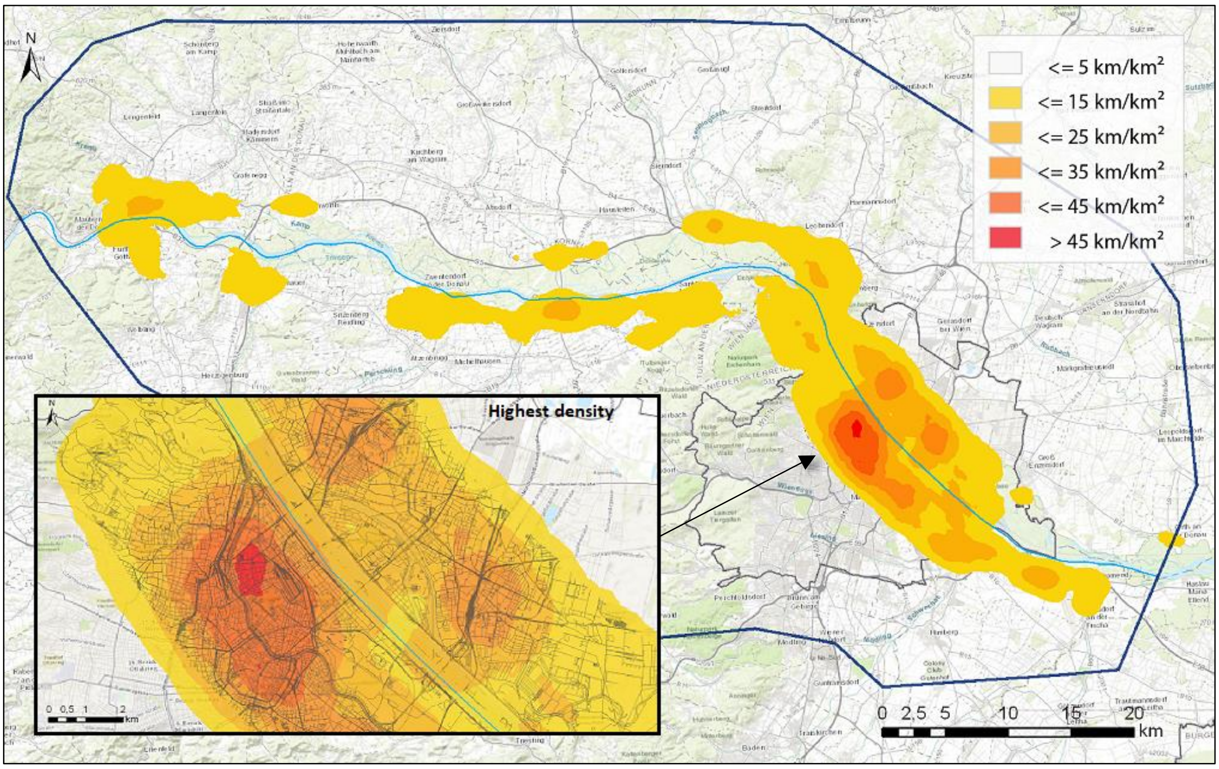


Fig.89: Line density from the OSM road network in unflooded areas with magnified peak area

The first illustration (Fig.88) represents the line density from the OSM road networks within flooded areas. It is based on the line features of the road network alongside the Danube River. The used unit is kilometres per square kilometre (km/km²), because it visualises the amount of road kilometres per surface unit inside the flood discharge area. The lowest values between one and five road kilometres per square kilometre can be found between Krems an der Donau, in the western part of the research area, and Stockerau in the central part. The hotspots with the highest density of roads inside flooded areas are near Klosterneuburg and Korneuburg, but mainly in the south-eastern part of the research area, along the A4 motorway, the Vienna Airport and Schwechat.

The next map on the previous page (Fig.89) is quite similar to the prior map. It shows the line density from the OSM road networks, but in unflooded areas during the 2013 event. The illustration differs however from the first density map (Fig.88) insofar, that there are different class sizes used in either map. Inside the flooded areas the low density required only small classes with low values, outside the flooded areas however, the road network density increases substantially and requires bigger classes with higher values. Furthermore, areas with a density of less than five km/km² are invisible for the sake of clarity. The lowest values can be found in the western parts of the research area, with outliers in Krems and Tulln. The highest values are naturally in the city of Vienna (considering the buffer area only).

The first illustration on the following page (Fig.90) shows the point density from the OSM building footprints inside flooded areas. As a result of the missing polygon density tool in ArcGIS' Spatial Analyst Toolbox, all the buildings had to be converted into points. This method enables the user to produce a more generalised raster. As a result, the unit is points per square kilometre (pts/km²). The illustration (Fig.90) shows the few buildings in flooded areas as coloured patches. Low densities can be found all over the research area, with some of them in the western part near Tulln and in south-east Vienna. The highest values are near Klosterneuburg and Korneuburg. The building density ranges there between 100 and more than 250 pts/km². This fact is not surprising, because in the crowded centre of Vienna there were hardly any inundations during the 2013 flood event at all.

The next map (Fig.91) shows the point density from the OSM building footprints in unflooded areas. Similar to the analysis of the road densities, the values between flooded and unflooded regions varied significantly. The low values of building densities inside inundated areas required only an interval of zero to 250 pts/km², whereas the much higher building density outside inundated areas required an interval of zero to more than 1200 pts/km². Areas with less than 200 pts/km² were hidden for the sake of clarity. The illustration points out the location of higher building densities, which can be equated with the location of towns like Krems, Furth, Traismauer, Tulln, Wördern, Stockerau, Klosterneuburg, Korneuburg, Fischamend and of course Vienna. The highest density values can be found in the metropolitan area of Vienna and the just mentioned towns and villages. Smaller villages with a lower building density were not shown in this illustration.

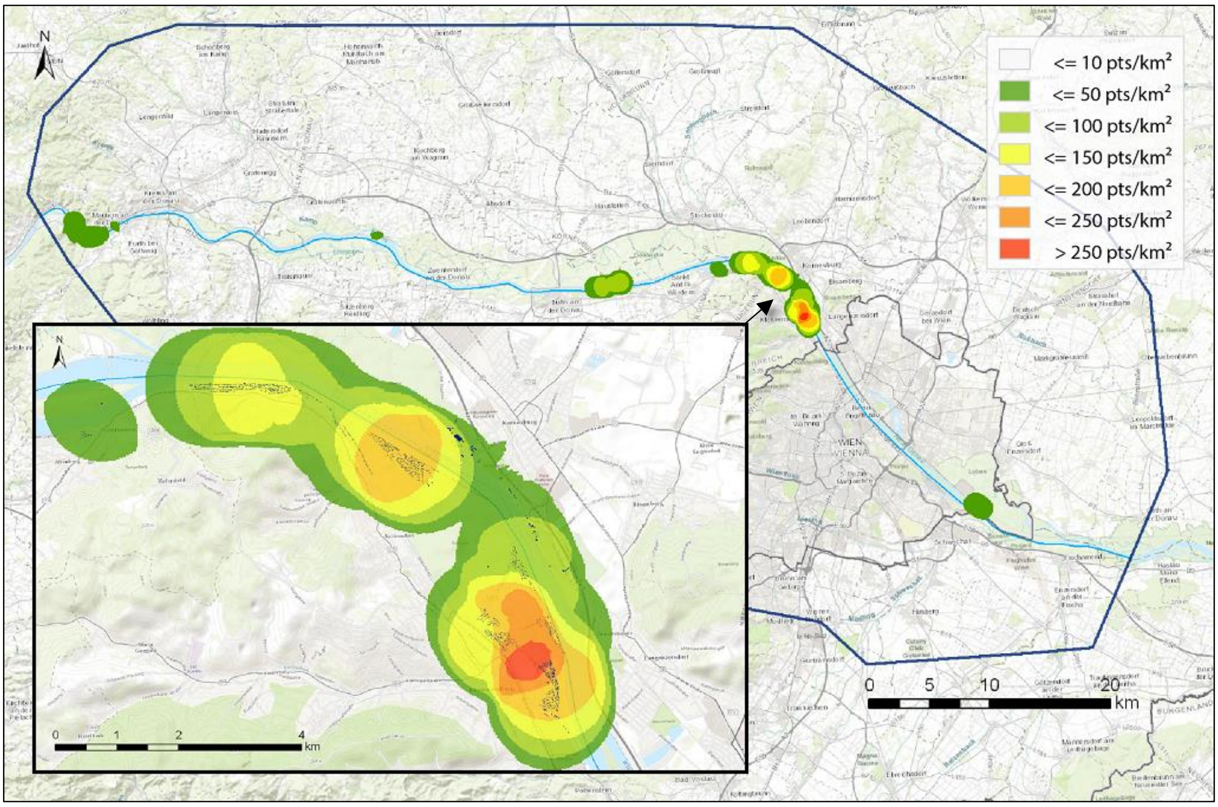


Fig.90: Point density from the OSM buildings in flooded areas with magnified peak area

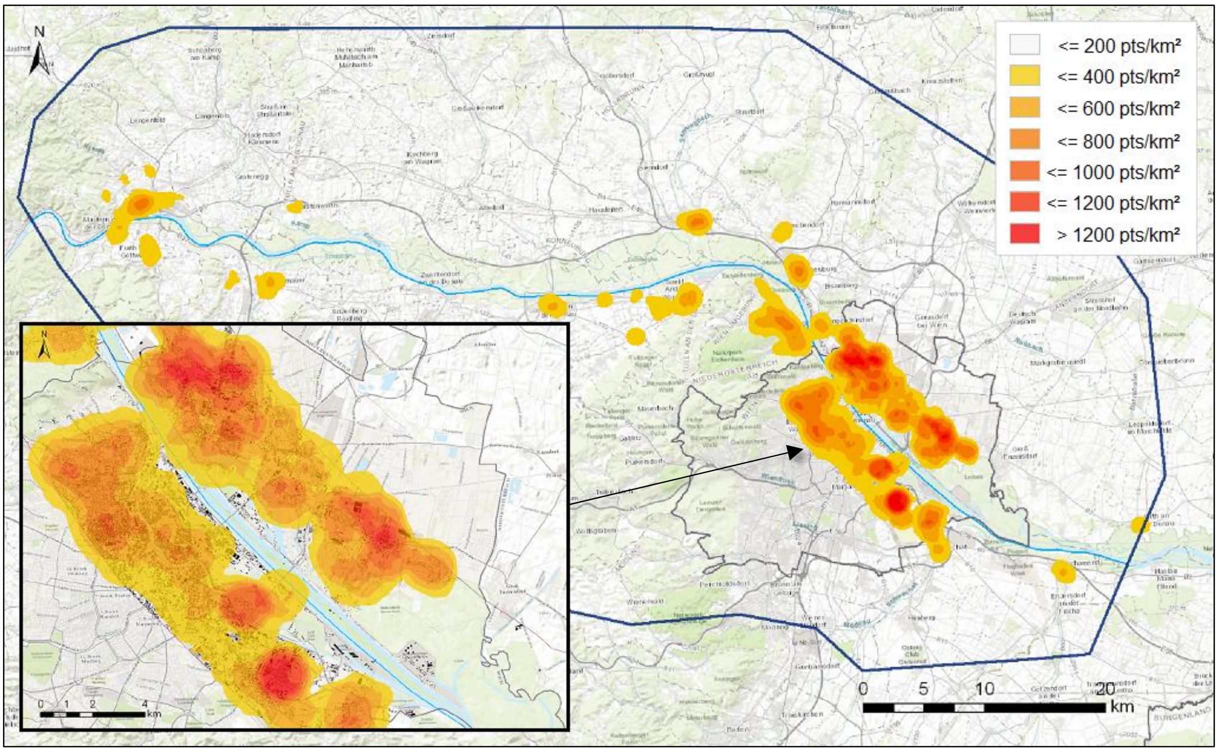


Fig.91: Point density from the OSM buildings in unflooded areas with magnified peak area

The next step to investigate the correlations between artificially sealed surfaces and the flood extent of 2013 is to investigate the influence of portable and permanent flood protections and other river control measures. There is currently no central database, which contains the location of all constructions, which could influence the distribution of flooded areas in the region. It was therefore necessary to collect all the possible information about barriers, dams, hydraulic structures, locks, ridges or walls. One of the main sources for this required data is the national Electronic Navigational Chart (ENC).

The following pages contain of some exemplary meaningful sections from the research area along the Danube River. The first of these maps (Fig.92) shows the area around Freudenau harbour in the south-eastern part of Vienna. There are several flood protection measures observable, like dams, locks and hydraulic structures, which include harbour walls and piers. The central area is apparently heavily protected by the appropriate structures. As a result, the main harbour buildings were safe and not inundated in the 2013 flood event. This is an example for a successful combination of an extensive investigation after the 2002 event and the resultant river training measures. These actions that took place after the last flood, helped to stop the river of inundating the bordering sealed areas.

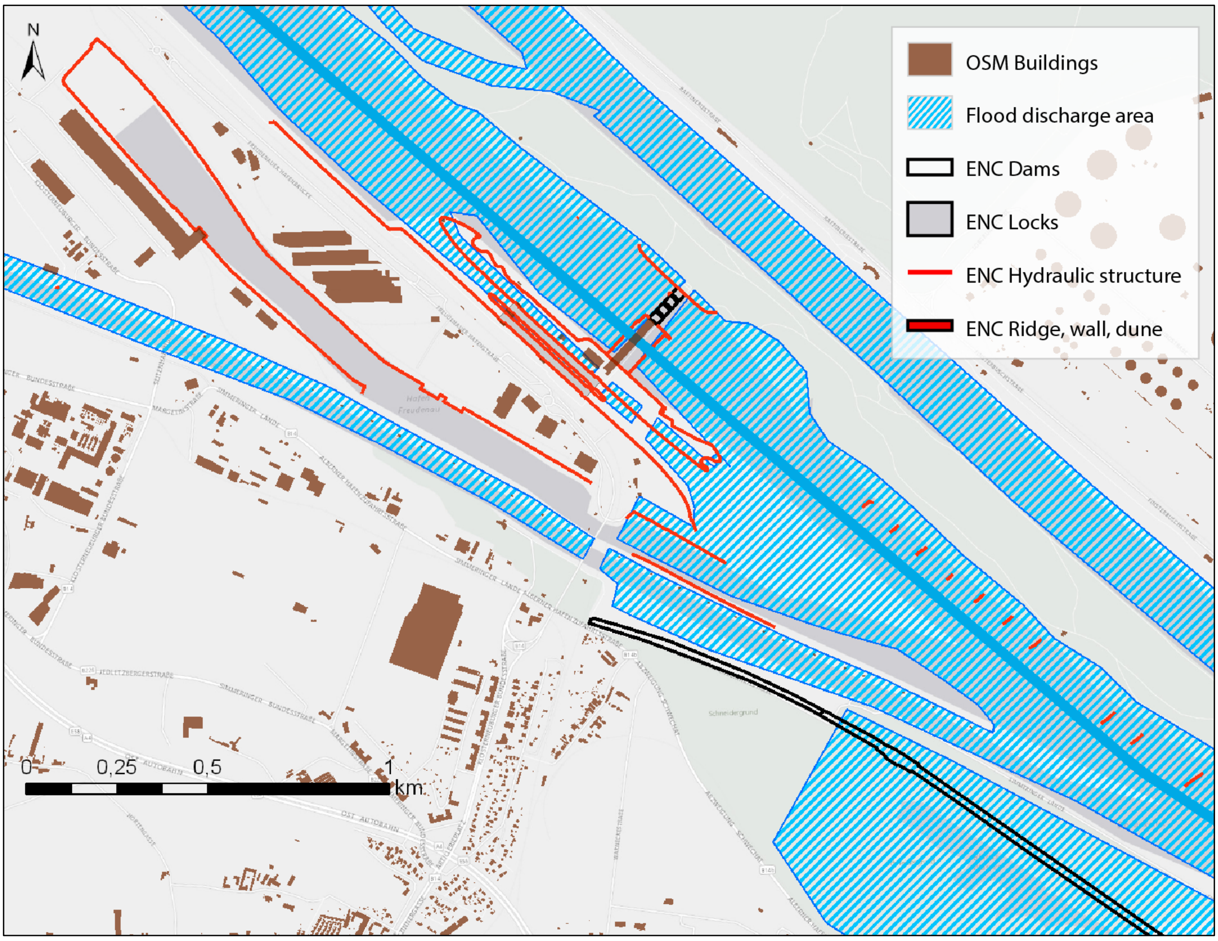


Fig.92: Exemplary section of the Freudenau harbour with protective structures from the Electronic Navigational Chart

One less successful example is not far from the previous map section. The following illustration (Fig.93) shows the Alberner harbour and the oil terminal Lobau, which are also in the south-eastern part of Vienna. On the left side, the Alberner harbour has clearly been flooded in 2013, although there were constructions present, which should have restrained the flood distribution. Several buildings were inundated and damaged. There is however a second line of defence, which apparently did hold back the water masses. These dams saved several other buildings from damage. At least some measures worked.

On the right sight of the map (Fig.93), parts of the oil terminal and several industrial buildings on the small peninsula were inundated in the 2013 floods. There were also some retaining walls and dams present, which did not prevent the buildings from being damaged. According to the data from the Electronic Navigational Chart (ENC), the protection measures were not surrounding the whole area, but only a few exposed locations. Another interesting factor is the long dam on the right, which apparently prevented the area in the east from being flooded, but the water found a way further south and inundated the region nonetheless.

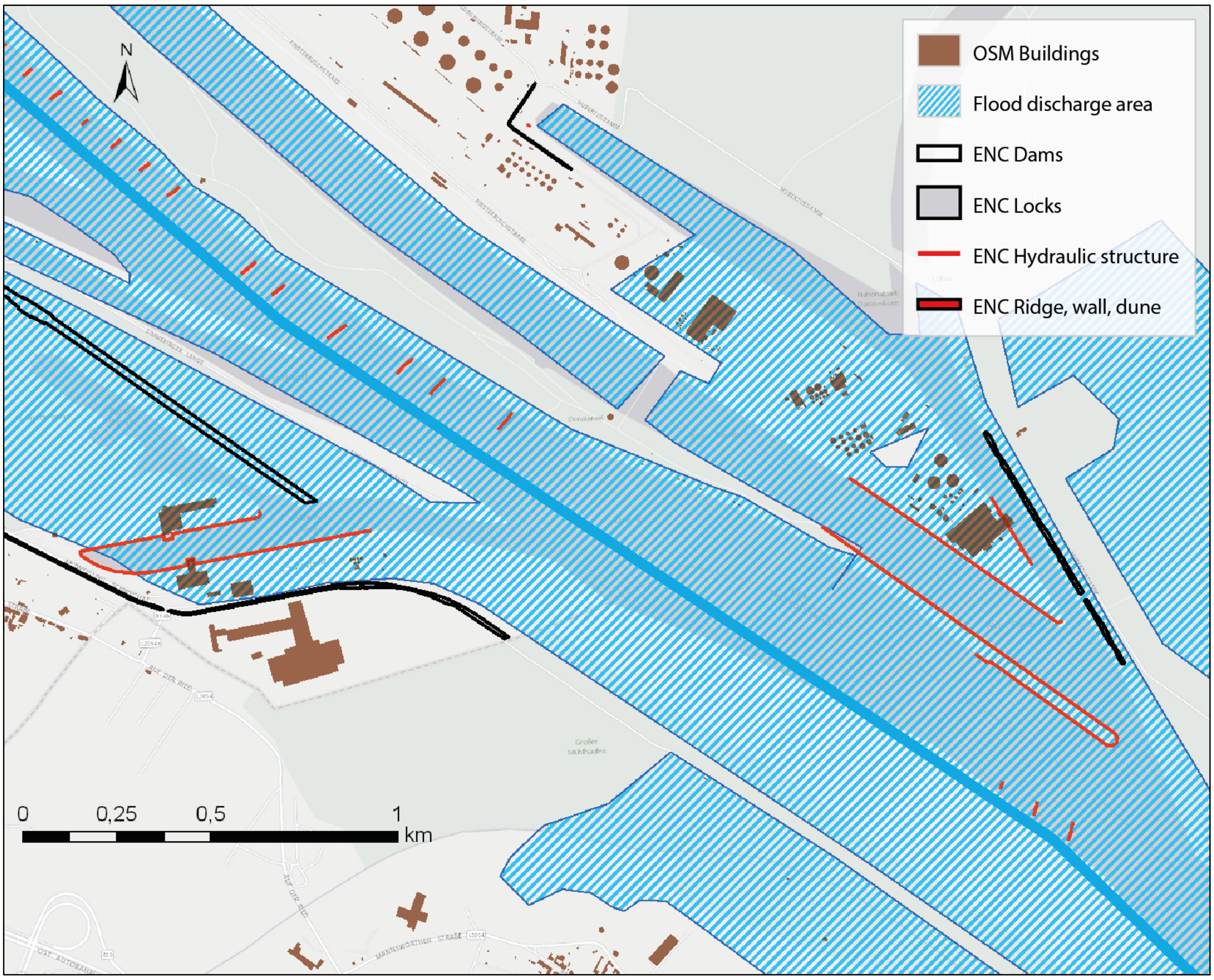


Fig.93: Exemplary section of the harbours Albern (left) and Lobau (right) with protective structures from the ENC

After an example of working (Fig.92) and non-working flood measures (Fig.93), the next map illustrates one of the major discharge areas of the 2013 flood event (Fig.94). The area is south of Stockerau, in the central part of the research area. Some of the building footprints of the town can be seen in the northern part of the map. The illustration shows that large areas were inundated, but almost no building was affected or damaged. This is almost certainly a result of functioning urban planning, because these areas are prone to flooding. Even in a predicted 30-year event (HQ30), these areas will be flooded. This is why most of these areas are meadow landscapes. Nearly no sealed surfaces can be found in these regions. In the southern part of the map section, there are hydraulic structures and retaining walls observable. These measures seem to have worked during the flood, but thanks to a tributary further down the river, large areas behind the wall were inundated nonetheless. There is however a more interesting question suggesting itself: Why is the town of Stockerau not inundated? There are no recorded flood protection measures or other constructions. Some of the reasons could be that the town is too far off the river course or that the local express road may work as a dam. Another possibility could be the prevailing elevation of the terrain. The influencing factor of the ground elevation will be discussed briefly on the next pages.

Fig.94: Exemplary section of one of the major discharge areas south of Stockerau with some flood protection measures

One of the factors which were not integrated in the previous analyses is the ground elevation. Every area has its own topographical characteristics and the required information for the interpretation was obtained from the Earth-Explorer data portal. The digital elevation model is a digital representation of an area's terrain. The basic DEM raster is based on ASTER data and retrievable from the Earth-Explorer platform, hosted by the United States Geological Survey. This data will be processed and functions as the basis for a short visual interpretation.

After loading the digital elevation models (the research area contains two separate DEMs), both raster datasets were converted into a Triangular Irregular Network (TIN). A TIN displays the surface morphology of the prevailing terrain by creating a number of triangles (via points and edges), which use the digital elevation model as data basis. The altitude can be classified into a customisable range of colours. In this case, the minimum height of the area is represented in dark green and the highest altitudes in dark red. With the help of this TIN it is possible, to illustrate the research area and try to find connections between the prevailing elevation and discharge areas from the 2013 flood event.

The next map (Fig.95) shows the total research area, modelled by numerous triangular units, and the flooded areas. Focussing on the biggest inundations in the central part of the map, it is apparent that there is no significant change in elevation, which would explain why the flood distribution was stopped. Possible reasons that towns like Stockerau (Fig.94) were not inundated in 2013, could be the distance to the river course or working flood protection. Although there were no protection measures recorded in the previously used database, it is no guarantee that there are no such measures implemented. Focussing on the surrounding terrain, the water volume and the average predicted water depth (Fig.51), it seems that there were river training measures and bank protection installed in recent history. All the major discharge areas are in a relatively flat area with no change in elevation to stop them prematurely. Even downstream, in the south-eastern part of the research area, there is no significant alteration. Vienna however, with its sophisticated river regulations and walls, has no recorded inundation, no matter what the terrain was. Elevation, it seems, has no significant influence on the flood distribution. Even the statistics only indicate a very low negative correlation.

There is however a small, almost not recognisable area, which shows some influence of the local terrain on the flood extent (Fig.96). In the most western part of the research area, not far from Krems an der Donau, there is a river bend, which prevented the water masses from inundating the surrounding environment. Large hills up to 900 metres are very effective ways of stopping a flood and directing it into another direction. After the terrain levelled out, the Danube did not inundate the flat plains, but was restricted by the usual flood protection measures. Another one of the few examples of elevation influencing the flood extent, can be seen around Klosterneuburg. The last foothills of the Wienerwald prevented the Danube briefly from inundating areas south of the river course. In summary, the elevation of the research area seemed to have almost no noticeable influences on the flood extent in 2013.

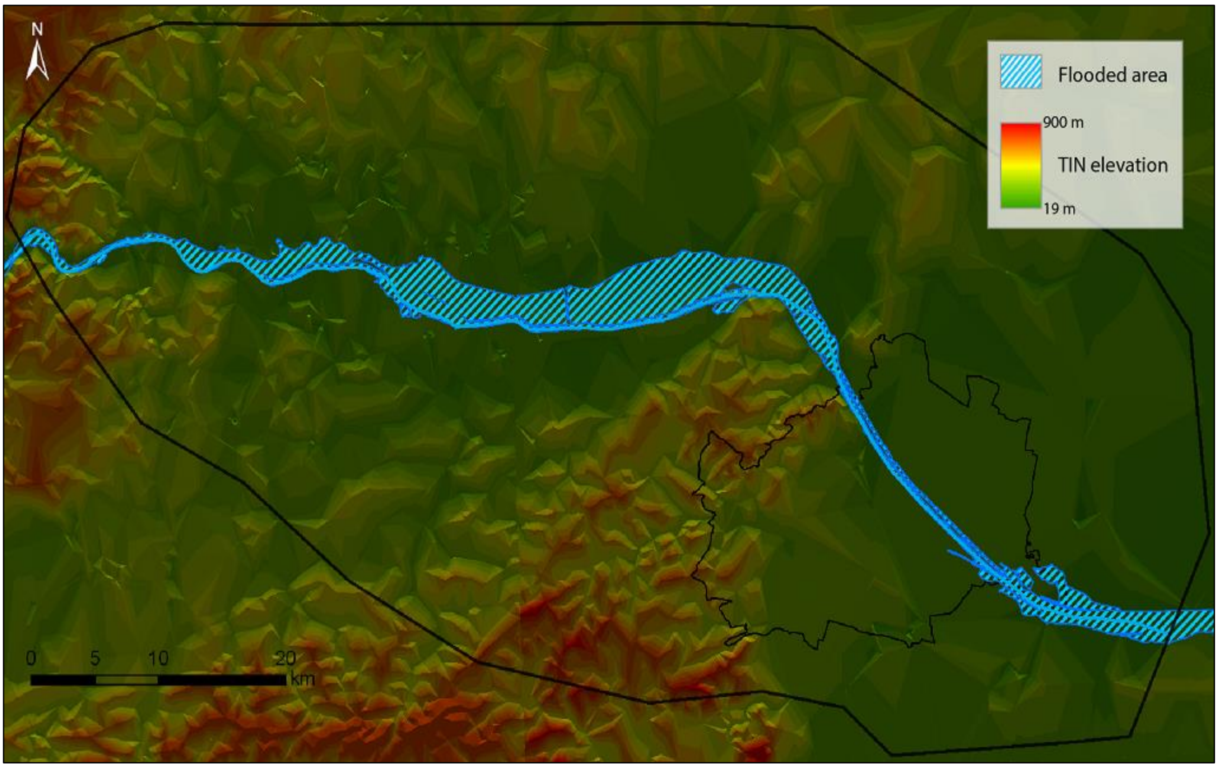


Fig.95: Triangular Irregular Network (TIN) based on the DEM with the flood discharge area as an overlay

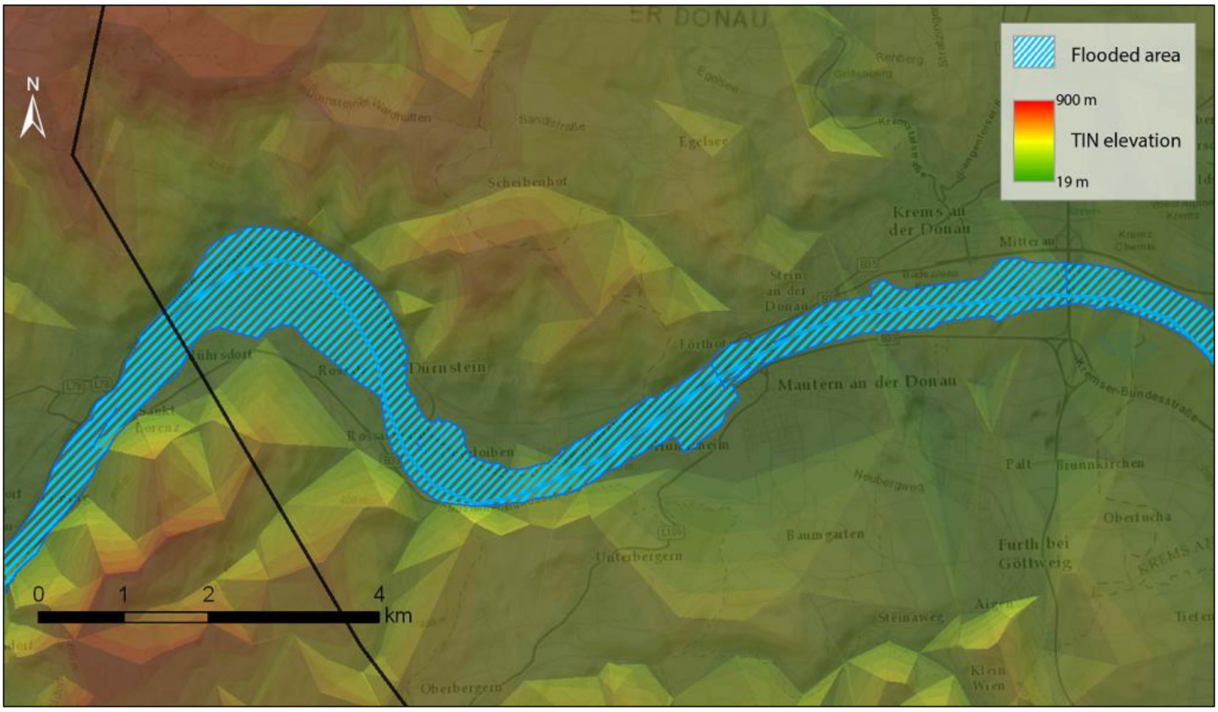


Fig.96: Section of the western part of the research area with the flood discharge area following the terrain (TIN)

5. Discussion and Interpretation of the Results

The main aim of this master thesis was to determine potential connections between the extent of the 2013 flood event and the sealing of the soil alongside the Danube. This task was set up to investigate the usability of geodata, which was produced by untrained experts in the course of a citizen science project. Another important goal was to point out the applicability of this information in the field of urban planning and disaster management. Taking a closer look at all the findings of the past few chapters, it is hard to tell if these aims were achieved sufficiently or not. The following paragraphs will focus on the interpretation of the results. They will also examine the strengths and weaknesses of the approach and the used data, as well as the possible applicability of the method for official bodies. The conclusion should mark the end of the practical and analytical part of this thesis and shall include deductions, implications, recommendations, predictions for the future and the author's personal opinion.

5.1. Interpretation of the Findings

The author of this thesis has stressed throughout this thesis that the voluntarily generated data from the FotoQuest Austria campaign is not precise enough for the determination of correlations between sealed soils and the extent of the 2013 floods. Compared to traditional methods of land cover classification like supervised classifications or open data from OpenStreetMap, the citizen science approach FQA did not perform as well as the author anticipated before choosing the research question. However, the author may argue that this method still has its advantages. This could include a quick and cheap land cover classification of a whole country or even the continent. Former research during the author's bachelor thesis has shown, that the geographic derivatives of the crowdsourcing approach are able to cover a substantial share of a country's land area and classify their degree of settlement so sufficiently, that even official agencies were interested in the findings. The geographic citizen science method therefore still is an appropriate and fast tool for the collection of large-scale land information. The usability on small-scale areas like the research area along the Danube River may not be as high as required yet. With a higher number of motivated volunteers, the reliability of this kind of information for a limited space could definitely be increased.

The accuracy assessment of the FQA data was compared to established datasets like LUCAS and the automated supervised classification. The comparison with LUCAS as ground-truth revealed a moderate or viable agreement of 71.33%, which is not as high as the author expected. The assessment of other established land cover datasets showed that they also have an agreement of just about 80%. So although being nearly 10% down, the citizen science approach of detecting land cover performed quite well, considering the low amount of multiply validated sample points. The values of the calculated user's accuracy ranged from 50% to 100%, although the most significant values with the highest amount of sample points ranged between 61.70% and 84.62%. These values indicate that the accuracy from the user's perspective are still improvable and even marginally lower than the values from a similar bachelor's project. From the producer's perspective however, the most significant values ranged from 21.74% up to 91.23%. Most of the values offer some substantial agreement and are within the normal range. The last indicator for the data's accuracy was Cohen's kappa. All land cover classes combined resulted in an agreement of approximately 60%. This means that just six out of ten classifications were better than by chance alone. Focussing on the three major land cover classes however, resulted in an agreement of 81%, which means that the more validations were available, the higher the accuracy became. As a result of these analyses it can be said, that if the initial FQA campaign would have attracted more voluntary contributors, the resultant data would have been more accurate and reliable. It is therefore not wrong to say, that the geographic citizen science approach does not fail to produce accurate data due to its internal structure, but mostly because of the lack of willing participants.

Even the investigations of the FQA dataset alone revealed helpful insights, which can be used in order to improve any form of follow-up projects and their data quality. In the research area, 592 validations of 393 individual sample points were recorded. This means that on average, every point was visited only one and a half times. In order to provide a certain amount of data quality, every sample location should have been visited at least three times. Nearly three-quarters of all points were visited only once. Already at this stage in the course of the master thesis, the author knew that the data quality would probably not be sufficient enough for answering the quite detailed research question. One of the reasons for these demotivating statistics is the coarse distribution of the initial sample points throughout the country. This is a result of the strong bond between the FQA campaign and the LUCAS survey. Not only the land cover class types are based on the established latter dataset, but also partly the low sample point density. The gaps between the locations may be enough for a Europe-wide project, which covers the whole continent and provides only a rough overview over the situation of the EU member countries. Classifying a small country like Austria with its different terrains, requires some significant adaption of the approach and its design. Most of the reasons why some sample points had to be skipped during the campaign was the inaccessibility due to the terrain and dense vegetation. Also with higher sample point densities the results of the successor projects should be much more precise.

One of the major reasons for the low amount of validated sample points was the lack of motivated participants. In the research area 80 users contributed 592 validations, which equals an average of 7.4 points per person. Comparing it to the author's bachelor project (where the volunteers could contribute their data at home and not in the field), back then 97 unique users contributed nearly 30,000 validations. This meant an arithmetic mean of almost 293 validations per user. This is considerably higher than the FQA statistics. The catch however is that the participants were students and the project was conducted during a remote sensing lecture. Nonetheless, these values show that with the right motivation and desirable incentives, the data volume for crowdsourcing projects can be substantially increased.

As a result of the low density of validated sample points, the data had to be interpolated in order to cover the entire research area. The used method was the Inverse Distance Weighted Interpolation, which predicts possible values for the areas, which were not validated, by measuring existing values and providing near points with a higher weighting. This standard method generated a rather vague mathematical estimation of the gaps between validated sample points. This added another share of uncertainty to the FQA dataset, but there was unfortunately no better way of dealing with this problem at that point. The investigation of the soil imperviousness in the research area however required a dataset, which only showed information of whether an area is sealed or not. After processing the data to a point where they were able to interpolate, the resultant map seemed to be quite plausible. Areas with a high density of validated sample points were much more accurate and detailed (like in Vienna), than areas with a low point density (remote areas like Stockerau or Krems). With a more finely woven sample point distribution, the interpolated map would have been way more reliable and the following analyses would have been much more significant. After choosing a threshold of 0.5, the interpolated data was reclassified into two classes: sealed (1) and unsealed (0). This was meant to be the foundation for further research. It also showed a more or less reasonable distribution of sealed and unsealed areas.

In order to provide a comparable dataset, which was created by using semi-automatic classification plugins with Landsat 8 satellite imagery, the data from supervised classification functioned as a steady companion throughout further analyses. The comparisons with this dataset, which even achieved a higher value at the accuracy assessment, put all the analysed results from the FQA dataset into perspective. One of the major research questions dealt with the comparability of geographic citizen science data with the traditional assessment of satellite imagery via an automated supervised classification. First, all the land cover classes were compared to each other. There were, as expected, some major discrepancies apparent, as the share of artificially sealed areas from the supervised classification was much lower than the FQA classification. It seemed that the classification algorithms have mistaken some of the artificial landscapes with bare land. Other disagreements were found between the different vegetation class types, which had no effect on further analyses, because these areas were reclassified as unsealed either way. The visual interpretation of the sealed areas based on FQA and SVC presented a different picture. The large patches of the interpolated FQA data suddenly seemed to be quite rough and chunky, whereas the pixel-based classification of the SVC method seemed much more refined and detailed, even indicating some of the road network in the research area.

The statistical findings were calculated by comparing all different combinations of agreement and disagreement of sealed and unsealed areas between the FQA data and the SVC. By writing an elaborate formula into the raster calculator, it revealed that most of the area was classified consistently. Nearly 85% of all the classified unique pixels showed either an agreement of sealed areas or an agreement of unsealed areas. Only about 15% were classified inconsistently. This indicates a substantial accordance between the two differently generated maps of sealed surfaces. The significance of these findings however should not be overrated, because both methods have their faults.

After all these analyses, the author focussed on the potential correlations between soil sealing and the extent of the 2013 flood event. All the previously calculated statistics and maps indicated, that the voluntarily generated FQA data was not as precise as the author had hoped. In order to determine the potential correlation, the analyses were split into two parts. The first one investigated any correlations based on the FQA data in comparison with the supervised classification, and the second part focussed on the analysis of traditionally generated geoinformation and open data.

Answering the main research question, required extensive spatial and statistical analyses. The first visual interpretation of the soil sealing maps with the official flood discharge areas already showed, that the majority of sealed surfaces are located outside the inundated areas. Exemplary towns like Stockerau, Korneuburg or Klosterneuburg were very close to the flooded area, but only a comparatively low percentage of sealed surfaces laid inside the discharge area. It seemed to be, if anything, a low negative correlation between soil sealing and the flood extent. The first and rather simple calculation resulted in a very low positive correlation for the FQA data and a very low negative correlation for the supervised classification. These results took the author by surprise, because not only were the values very low and indicated basically no correlation at all, but also the different signs and their meaning for the research question. The geographic citizen science data implied, that there is a nearly not-existent positive correlation, which means that the more sealed surfaces there were, the further the research area distributed. The more detailed supervised classification however indicated a marginal negative correlation, which means that higher numbers of sealed surfaces prevented the flood in 2013 from inundating even more areas.

This interpretation also revealed, that a large amount of sealed surfaces were located far off the Danube and its discharge areas. Insignificant data in remote areas may have influenced the correlation values and as a result of that, all datasets were clipped to a 4500 metre buffer area around the river course. These actions more than tripled the individual values of the correlation, but they were still not high enough to matter in the world of statistics. After processing the data again quite extensively, other statistical measurement methods were applied. The standard correlation and the Pearson’s method both delivered similar values for each data generating method. FQA showed a very low positive correlation again, whereas SVC showed a low negative correlation. The results of the R-squared were both very low, which indicates no linear connection between the two variables. The values for the covariance confirmed the previous correlation methods. This strongly indicated that either there is no real correlation between soil sealing and flood extent in the real world, or that both methods are not accurate enough in order to determine the connections. To find out which of these statements was true, the author compared the calculated results with the findings of traditionally generated information and open data.

This chapter focussed on data from official bodies and agencies like the Federal Ministry for Transport, Innovation and Technology or the Provincial Government of Lower Austria. All of the provided data was collected and marginally processed by trained experts in a GIS environment. Other data sources were established online portals like OpenStreetMap and Earth-Explorer. Some of them are open data platforms of official agencies or companies and some of them are based on crowdsourcing. The first analysis investigated the density distributions of building footprints and road networks from OpenStreetMap. It showed that the building density inside the flooded areas was very low with 2.18% and outside flooded areas quite high with almost 30%. This not only indicated that there is a measurable negative correlation between built-up areas and the 2013 flood extent, but also that the results of the automated supervised classification may have been more accurate than the FQA results. The road density inside flooded areas (~2%) and outside those areas (~25%) confirmed this tendency. After processing and exporting the data to Excel, the standard correlation between sealed surfaces based on OSM data (building footprints plus road networks) and the flood extent, showed a value of -31.14%. This result is way more significant than the previous correlations for the FQA and SVC data. Although implicating a rather moderate negative correlation, this more reliable dataset confirmed the indications from the previous building and road densities. It also suggested, that the SVC method is more useful than the voluntarily generated FQA data. The visual comparison of both building and road densities, inside and outside flooded areas, additionally approved the interpretation of the previous results.

Although the whole FotoQuest Austria campaign collected more than 592 validations of 393 individual sample points in the research area, the quality and density of this data was rather low. This was partly a result of the lack of motivated participants, a lack of validations, a low sample point density in the research area, the missing adaption to the country’s characteristics, the need for interpolating the gaps between the sample points, the setting of a threshold for reclassifying the area into sealed and unsealed regions and the low performance in comparison to the automated supervised classification.

Although there were a number of surprises and disappointments throughout the data analysis, the author thinks that the main research question can be answered. It is possible to determine the rather moderate negative correlations between sealed surfaces alongside the Danube and the extent of the 2013 flood event using traditionally generated information and established crowdsourcing platforms like OSM. The amount of data and the resultant point density from the FQA campaign however is currently not high and precise enough, to determine possible correlations conclusively. There are however numerous ways available of improving future projects.

5.2. Strengths and Weaknesses

As mentioned in the corresponding chapter, crowdsourcing and especially its derivative geographic citizen science are new and fast approaches of generating information about a certain topic with the help of untrained volunteers. With geographic information getting more and more important in today's digitalised society, traditional approaches of collecting data may be too slow and expensive in the near future. This is one field of applicability of crowdsourcing. This type of *participative activity* can contribute large amount of data for universities, organisations or companies in a short period of time. One of the major points of criticism however is the data quality. It is therefore vital to implement safety measures like accuracy assessments and quality management. Throughout the master thesis, numerous comments on the vagueness or incompleteness of the voluntarily generated citizen science data (FQA) were made by the author. Only through the processing and preparation of the FotoQuest Austria data it became apparent, that the research question may be too detailed for such vague data.

The most important weaknesses that were detected while writing this thesis are:

~ the low number of participants that took part. The whole project only attracted 208 individual users (80 in the research area), although due to the Internet, the available potential should be higher.

~ the low amount of total validations and validations per user. In the research area 592 validations of 393 unique sample points mean an average validation of 7.4 (median of 2) per user.

~ the low amount of validations per individual sample. On average, every sample point was validated around 1.5 times (median of 1). These values are not high enough to be statistically significant. One of the reasons could be the questionable incentive, that the more frequent a point has been visited, the less points the user gets awarded (from 100 to 90, 81, 72, 65 etc.).

~ the strong bond between the FQA campaign and the European LUCAS survey. Although being a good basis, there was no real adaption of the survey's structure to specific Austrian characteristics.

~ the rough and widely scattered distribution of individual sample points in the research area. As a result of that, there was no real detailed focus on the Danube River course possible.

~ the inaccessibility of some sample points due to their remoteness or privacy and the temporarily failure of the GPS connectivity.

~ the mediocre result of 71.33% from the accuracy assessment with LUCAS as the ground-truth.

~ some quality issues due to lazy contributors which apparently took the obligatory ground pictures not in the field at the sample point, but in their personal home office.

- the interpolation of not validated sample points. An interpolation is always a rough estimation of possible values and is therefore prone to error and uncertainty.
- the comparison of the citizen science data with an automated supervised classification derived from satellite imagery. The latter dataset showed a higher accuracy of 74.83%. The direct comparison of the interpolated FQA map with the SVC map offered an agreement of nearly 85%.
- The low correlation values of both datasets (FQA and SVC). These low values of the respective calculated correlation (positive and negative), indicate either a non-existent correlation between soil sealing and the flood extent, or that the data is not precise enough to detect any potential connections.

There were however also a number of strong points for the geographic citizen science approach like:

- the fast speed of generating data. In less than a year, the campaign was planned, initiated and completed. This shows that the use of crowdsourcing and its geographic derivatives could accelerate the conduct of such a project in comparison to traditional forms of collecting data.
- the simplicity of conducting a citizen science project. The campaign was initiated by a medium sized institute (IIASA, approx. 200 members), but the same kind of project can be started by big organisations and companies or small non-profit initiatives.
- the rather low costs of conducting the FQA campaign. The few members of staff involved created the application based on the LUCAS survey and were responsible for maintenance and collection of the generated data. A group of scientists processed and analysed the data and published their research results. Other cost factors were the monetary or material prizes, which were used as incentives for the contributing users.
- the possibilities of describing and assessing remote and inaccessible areas due to the interpretation of satellite imagery. In case of the FQA campaign however, it enabled the few members of the project to receive data from all parts of the country without even leaving the office in Laxenburg.
- the high number of potential participants due to the connectivity of modern society. Thanks to the internet accessibility on mobile devices, like tablet computers and smartphones, nearly everyone in the country could function as a human sensor and potentially contribute large amount of data. In case of the FQA campaign the amount of volunteers was quite high with more than 200 users, but the average individual number of validations was surprisingly low.
- the usability of citizen science in the aftermaths of a disaster. Some examples of location-based information, contributed by the affected public via social media services, were used in course of the disaster relief measures after the 2010 earthquake in Haiti or after the events of Fukushima in 2011.

5.3. Applicability in Urban Planning and Disaster Management

One of the main research questions at the beginning of this thesis asked, if there are possible applications for the results of the analyses and how it could help local authorities to manage urban planning strategies in densely populated areas. The next question was quite similar and asked how policy makers and rescue agencies could benefit from the analysed results before, during and after a flood event referring to disaster response? The applicability of volunteered geographic information and geographic citizen science is quite an affair of the heart for the author and the initial reason for choosing the topic and research question for this master thesis. The results of the performed analyses however, paint a dark picture for the usability of the FQA data for urban planning and disaster management.

Throughout the past decade, crowdsourcing became a fast and cheap alternative of collecting vast amount of data, with the help of untrained volunteers. One of its many derivatives, geographic citizen science, has the potential of contributing large amount of information in a short period of time. FotoQuest Austria was such a campaign. It was designed to classify the different land cover and land use types all over the Austrian territory. The focus on the sealed areas should work as the basis for the investigation of vulnerability, risk and the anthropogenic influence on natural disasters.

The first analysis of the generated FQA data already revealed its major problems. The low amount of contributions, unmotivated users and the low density of sample points all over the country resulted in a vague and imprecise dataset, which had to be interpolated in order to cover the whole research area. This mathematical estimation added a large amount of uncertainty and inaccuracy to the dataset. Since that point, detailed investigations were not reliable anymore and too blurry. In case of an emergency, disaster response agencies have to rely on precise and absolute accurate data, otherwise this could lead to the loss of property or human lives. There are several examples, where crowdsourcing solutions for disaster response already worked, but only with a detailed framework and initiated from governmental bodies. The author already mentioned some of them in the course of this thesis. There is however a lot of potential, but just like the results of this thesis' analyses, there is still a lot of uncertainty within the data. These challenges have to be assessed in the near future and the suggestions for improvement have to be incorporated into all coming successor projects.

The analysis of the flood discharge areas along the Danube revealed that spatial planning in Lower Austria is already very good and that planning strategies and flood protection already worked quite well during the 2013 flood event. The low building and road densities inside inundated areas confirm that. The FQA data may not be suitable for improving spatial planning, but it shows the potential of detailed local projects within a community, where citizens are motivated to contribute data. Every local authority could launch their own projects, tailored on their own needs and specifications. If enough citizens take part in such a local campaign, the approach could generate valuable data for communal planning.

5.4. Conclusion

Writing a conclusion is always one of the most challenging bits for the author of this thesis, especially after reading hundreds of scientific papers and abstracts, running after every piece of required data, writing thousands of words in a foreign language and spending 14 hours a day in front of the screen of an old and cheap 17 inch laptop. More than 150 pages of this thesis deal with the thematic introduction to provide the necessary basic understanding, the data preparation and assessment, the analysis of uncooperative datasets and the interpretation of the findings. At this point, it is time for a discussion of possible deductions, potential implications on future research, recommendations, predictions for the future and the author's personal opinion, bringing all the information together and providing the essence of the previous chapters. Recapitulating the findings of this thesis with the research questions in mind, it is obvious that there are still a lot of answers to give and hypotheses to verify

~ One of the main research questions was, whether data from citizen science projects can determine correlations between sealed surfaces and the extent of the flood disaster in 2013. After the previous analyses, the author would answer the question with *not quite yet*. Methods like VGI or geographical citizen science may work for a rough overview about a country's or a continent's land cover, but not for the investigation of a detailed research question in a rather small area. In order to collect sufficient and accurate data for a research like this, the structure of the FQA campaign should be changed and adapted as well as the distribution of the individual sample points has to be much denser.

~ The next questions asks if there are possible linkages between the imperviousness of the soil and the flood extent during the 2013 inundation of the Danube River in Austria. Analysing the results from the FQA campaign and the supervised classification, there was no real final statement derivable. The investigation of open data and traditionally generated information at the end of the chapter about data analysis however indicated, that there is a moderate negative correlation between sealed surfaces and the flood discharge area. So the answer to the question has to be: *yes, but only just*. In terms of statistics, a value of -31.14% only indicates a low or moderate relationship between two variables.

~ Another sub-question addresses the extent, to which crowdsourcing and its derivatives geographic citizen science and volunteered geographic information can complement traditional satellite or ground photo assessments. In order to deal with this issue, an automated supervised classification (SVC) was created, using Landsat 8 satellite imagery. Already the first comparison, the accuracy assessments with the LUCAS dataset as ground-truth, revealed that the SVC had a higher accuracy than the FQA data (74.83% vs. 71.33%). Although every approach has its weaknesses, the calculation of the correlations and building densities from OpenStreetMap data showed, that the traditional assessment and interpretation of satellite imagery is still better and more accurate, than the citizen science approach. However, if the FQA campaign gets more adapted to the specification of the

Austrian territory and the sample point density increases, the results could be much more precise and reliable than before and therefore possibly match the traditional methods.

- The next question, asked more than 150 pages ago, focusses on which additional data is necessary to measure and quantify the potential links between impervious land cover and flood extent. As the last analyses showed, traditionally generated information and open data is still better than the rough geographic citizen science approach used in the FQA project. For a conclusive answer to this question, several datasets had to be added. This included the building footprints and road networks from OpenStreetMap, the flood discharge area and predicted high water areas (HQ30, HQ100, HQ300) from BMVIT, information about hydraulic structures, dams, barriers, locks, ridges and walls from the Electronic Navigational Chart, as well as the Digital Elevation Model based on ASTER data. More on the necessary additional data can be found in the corresponding chapter.
- Based on the previous research question, the next one asks how difficult the integration and collection of additional factors such as elevation and the use of (portable or permanent) flood protection is. As discussed in the corresponding chapter, the collection of traditionally generated information was quite modest. Open data from internet platforms were easily available. Other data from official bodies were either loaded from their online portals or exchanged at a personal appointment. Most of the data from the Federal Ministry for Transport, Innovation and Technology (BMVIT) was provided after several personal meetings with the responsible employee. The implementation of this data however is another story. At the university the author was used to perfectly processed data, which mainly works for one assignment or one course. In the real world, data is mainly unprocessed, chaotic and sometimes unmanageable. The process of preparing all the data, so that they work for the corresponding tasks, costed a lot of time, nerves and working memory.
- Another one of the research questions was the determination of the best ways to assure the quality and the accuracy of the contributed information and the comparison of geographic citizen science data with the data quality of traditional land cover classification methods. As already mentioned before, based on the analysis of open data and traditionally generated information, the supervised classification seems to be a more reliable way of classifying land cover in a small area than the citizen science approach of FQA. The creation of the SVC dataset by the author was one of the main mechanisms to compare and ensure the data quality. The calculation of numerous statistics about the FQA dataset in the chapter about *general statistics* provided a lot of information for the assessment of the data. The accuracy assessment with the LUCAS survey as ground-truth was another way of investigating the reliability and preciseness of the voluntarily generated FQA data. Throughout the course of the master thesis, every analysis of the FotoQuest Austria dataset was compared to the supervised classification in order to guarantee a certain level of data quality. The analysis of traditionally generated data was also a way to determine the quality of the FQA results.
- Almost redundant at this point, the next question asks which method is the best to investigate the potential correlations between soil sealing and flood extent. The two available methods are the

relatively new geographic citizen science approach and the more traditional data collection methods. As indicated a few times before, the traditional approach of generating information about land cover or soil sealing is still more reliable than the crowdsourcing derivative. The geographic citizen science approach offers a lot of potential due to its connectivity with the internet and people's mobile devices. Every person works as a remote sensor and collects data all the time. If only a fraction of this potential could be harvested for such a project as FotoQuest Austria, the possibilities would be nearly endless. The author thinks that if a few hundred people would have taken part in the FQA campaign and contributed just a few validations on their walks or bike rides, the accuracy of the contributed would have easily been enough for significant results. It is apparent that there is a lot of dormant potential, it just has to be used somehow.

~ One of the research questions the author feared throughout the analyses, focusses on the possible applications for the results of the analyses in the field of urban planning and disaster management. The results for the citizen science campaign FQA were not really significant enough, to implement them in any official working procedure. Spatial planning in the research area is actually well established and worked well throughout the 2013 flood event. The information generated during the FQA campaign was not precise enough, because the whole project was probably not intended to investigate small-scale phenomena in the first place. Disaster management and response agencies require absolute accurate and detailed data, which the FQA campaign cannot provide.

~ The last research question asked, how well new technological achievements like geographic citizen science could contribute valuable improvements in the field of disaster response and flood risk management. Just like mentioned before, the accuracy of the FQA data is currently not high enough in order to support important rescue missions. It did however show the weaknesses of the method, which now could be addressed. If the data density could be increased (with more volunteers and higher sample point density), the author is sure that data from geographic citizen science projects could help local authorities to receive a first overview over the situation after a disaster struck.

As usual in a master thesis, there are some hypotheses to check the research questions and its sub-questions on their validity. These statements reflect the most important parts of the research questions.

~ The first hypothesis states that there are significant correlations between the sealing of the ground alongside the Danube River and the resultant extent and damage of the flood in 2013. This statement could be confirmed, but with a small restriction. The analysis of the OSM data revealed, that there is a moderate negative correlation between these two variables. However, a value of -31.14% indicates only a slight relationship between soil sealing and the flood extent, but it cannot be called significant, maybe moderate or noticeable.

~ The next hypothesis says that geographic citizen science and volunteered geographic information can complement the assessment of satellite imagery and ground photos, or in this case, the FQA data and the supervised classification. The accuracy assessments of both these datasets showed, that their values are similar to each other, only 3.5% apart. Every other analysis, like the calculation of the correlation, showed some discrepancies between them. Compared to the analysis of the open data and traditionally generated information, it seems that the SVC data is more accurate and reliable than the FotoQuest Austria data. This hypothesis therefore has to be falsified for now.

~ Another statement discusses the sufficient quality of crowdsourced data and the accuracy in order to produce meaningful and significant results. Generally speaking, the crowdsourcing approach has its reason to exist in the scientific world, although constantly showing a degree of uncertainty because of untrained volunteers. In this case however, the uncertainties of the FQA data were bigger than the benefits from this method. It is apparent that the citizen science approach produces interesting information for a large area, just like countries or continents. If focussing on a smaller problem, the data quality and accuracy is not high enough for producing meaningful and significant results.

~ The combination of remote sensing technologies and volunteered geographic information can facilitate the process of data generation. This was another statement, written down before the author saw the FQA dataset. Apart from the fact, that the FQA campaign is a citizen science project, the combination with remote sensing technologies could improve the results measurably. The last analysis of potential correlations with OpenStreetMap data revealed, that the crowdsourcing approach used in OSM is capable of determining spatial relationships, if the data quality is high enough and a lot of members improve the accuracy constantly. The citizen science approach alone, which has been used in this master thesis, has its limitations and has to be adapted for future research.

~ Geographic Citizen Science and crowdsourcing in general, are a helpful new methodical approach for the acquisition and interpretation of land cover data, which will become more important in the near future. This hypothesis can be confirmed, when the approach is used for larger areas.

~ The last hypothesis states that although citizen science is an interesting approach, traditional methods are still more accurate. Contradicting to some of the previous hypotheses, this statement was definitely confirmed in this master thesis. Citizen science has its limitations when the project design is not adapted properly and when there are not enough volunteers taking part. It is however a promising method for future research, if the level of digitisation continues to increase. Some of the weaknesses of the FQA campaign have already been addressed and adequate solutions have been incorporated into the follow-up project FotoQuest Europe. If these measures work, the resultant information could definitely rival the traditional methods of generating land cover data and would provide a much more accurate foundation for the determination of potential correlations.

6. Summary

The main objective of this master thesis was the investigation of the anthropogenic factors of natural disasters. This was enabled by the investigation of potential correlations between soil sealing alongside the Danube River and the extent of the 2013 flood event. In order to achieve this ambitious aim, different methods of data acquisition and processing were used. This included voluntarily generated geodata from the geographic citizen science project FotoQuest Austria and traditional open data from official bodies or corresponding online portals. The main purpose of the findings should make a small contribution to the constant improvement of official and private disaster management and urban planning strategies.

The author of this thesis always wanted to conduct the kind of research that has a strong connection to real-world problems, not only literature research that has no further use than earn a master's degree. The results of the analyses should give an insight into the possible applications of new data generation methods, based on modern remote sensing technologies. Such a potential field of application is the management of natural disasters. Catastrophic events are as old as planet Earth, but only since the existence of humans and the formation of society, they have become a threat and therefore a disaster. First interpreted as the bad moods of the respective gods, it is known nowadays that these catastrophes are the results of the dynamic system called Nature. From the first established civilisations, to medieval times and the Industrial Revolution, natural disasters have always affected people and their environment.

Every disaster however has its own characteristics and parameters, which makes it very hard to predict. Gaining experience and gathering information is therefore an essential part of understanding complex disastrous events like floods, earthquakes or wildfires, in order to prepare for the next event and mitigate the economic, environmental and social effects. Since the beginning of the 20^{th} century, official and organised disaster management began to develop. Different concepts like risk, hazard, vulnerability or resilience were devised and started to make an impact on spatial planning and emergency response on different political levels. After a considerable amount of time, even international frameworks were installed, to ensure cross-border cooperation in case of a catastrophe.

If for instance a flood disaster strikes, the four phases of the management cycle help to plan and organise all the emergency initiatives. This includes disaster response, recovery, mitigation and preparation. The prevention of or the preparation for a disaster is probably the most appropriate field of application for new methods of data generation, together with the immediate emergency response, because these two phases require the most information in the DMC. As a result of an increasing frequency of flood events, there is room for improvement in the traditional field of data collection and processing. One approach that has been used in this master thesis was geographical citizen science. This concept is a derivative of crowdsourcing and differs from other similar approaches (e.g. VGI) by focussing only on scientific initiatives with a spatial reference. An example is the FotoQuest Austria campaign (FQA), conducted by the International Institute for Applied Systems Analysis (IIASA). Using an application for mobile devices, information about land cover and land use in Austria was collected in 2015. Based on this dataset, an interpolated map was created, which showed the degree of sealed surfaces for the research area. This was necessary, because the main aim of this thesis was the investigation of anthropogenic influences on the severity of natural disasters. By measuring potential connections between the degree of sealed soils and the extent of the Danube River flood in 2013, with the help of voluntarily generated geodata and traditionally acquired open data, statements about the quality and usability of geographic citizen sciences were derivable. Unfortunately the data was not as accurate as the author had hoped.

In the selected research area, there were 592 total validations of 393 individual sample points, conducted by 80 individual users. This already indicated quite a low average validation of one and a half times per point. This was mostly due to the low average validation per user with only two sample points. The maximum validations with 140 points were contributed by one power-user. All these disappointing statistics were the result of a low amount of players and apparently no motivation for contributing more than two validations per player. There were however other problems, for instance inaccessibility of the sample locations or technical problems with the GPS-signal.

Considering the resulting land cover classification, around 28% of the research area was assessed as artificially sealed surfaces, more than 68% were overgrown or uncovered soil and the rest water areas. In order to ensure the quality of the data from the geographic citizen science campaign FQA, the accuracy assessment compared the generated data to a ground-truth dataset, in this case the Land Use and Cover Area Sample survey (LUCAS) from 2015. The aim was to estimate the correctness of crowdsourced data. The overall accuracy was 71.33% and can therefore be seen as a moderate agreement, but probably not good enough for significant results. Another indicator of agreement, Cohen’s kappa, resulted in a value of just 0.6. This means that the accuracy is only 60% better than by chance alone. All these results implied, that the data quality from the citizen science project is not sufficient enough and the number of users and their validations should have been considerably higher to provide a reliable database. Also the number of comparable LUCAS points in the area was too low.

Providing additional data quality included the verification of classifications with the selection of test samples, the re-evaluation of inconsistent classifications, the examination of numerous ground pictures (which were taken by the individual users) and the testing of the individual users’ reliability. As an addition to that, an automated supervised classification based on satellite imagery was also initiated, which offered an even better overall accuracy of almost 75%, with LUCAS as ground-truth again. The results of the supervised classification then were compared to the results of the citizen science project FQA. Before that, both datasets (FQA and SVC) had to be reclassified into two classes: sealed areas and unsealed areas. The preparation of the supervised classification data was relatively simple, because the whole research area was classified into the familiar eight land cover class types on a pixel-level. The data from the citizen science project FQA had to be divided into sealed or unsealed sample points and then interpolated, to estimate the areas which were not classified during the campaign in 2015. This left a very vague map, showing the prognosticated degrees of soil sealing. After setting a reasonable threshold, finally a map with the estimated sealed surfaces in the research area began to emerge.

The comparison of data from a geographic citizen science campaign with the results of the automated supervised classification had the purpose to demonstrate or at least indicate the usability and possibilities of voluntarily generated geodata. Looking at the percentages of the land cover classifications, it became clear, that every one of the two approaches had its issues. Firstly, the supervised classification had its problems with wrongly assessed cloud cover, secondly, both methods may had problems with classifying the different types of vegetation and finally, the number of usable sample points from the FQA campaign in the research area was just too low. After comparing the two soil sealing maps, nearly 85% of all the compared areas showed a total agreement, only about 15% offered disagreement. The majority of the agreement (approx. 81%) was due to the vast unsealed areas, combining forests, farmland, green land and bare soils. Although indicating quite a strong agreement, the significance of this result should not be overrated, because of the uncertainties within the interpolated FQA map.

The next and most important part in this thesis was the determination of potential correlations between soil sealing and the 2013 flood discharge areas, with the help of the estimated soil sealing map from the FQA campaign and the results from the supervised classification. In the course of this thesis it was detected, that the voluntarily generated data from the FotoQuest Austria campaign (FQA) may not be as accurate as possible and required. The final analyses, concerning the investigation of correlations, was split into two parts as a result of that. The first part focussed on the geographic citizen science data and the supervised classification, derived from satellite imagery. The second part offered a short overview over the research area based on traditionally generated information, received from official bodies. The first determination of possible correlations resulted in very low values, ranging from a low negative correlation of -3.15% (SVC) to a positive correlation of 4.14% (FQA). After focussing the investigation on a buffer area of 4500 metres along the river course, the values actually increased quite notably (to -11.44% and 15% respectively). These results are however still too small to reveal a significant correlation, because strong correlations should offer values of approximately +/- 80%. Further refinement only increased the resulting correlations marginally. The use of standard methods of calculating the relationship between the two variables soil sealing and flood extent, resulted in similar low values for the FQA campaign (1.39%) and the supervised classification (-12.12%). Additionally, indicators like R-squared indicate quite strongly that there is no real linear relation between the two variables. The calculation of the covariance as an indicator to which extent the variables change at the same time, confirmed the very low positive relationship for FQA (0.21%) and the very weak negative relationship for SVC (-1.54%).

The investigation of traditionally collected data revealed some factors, which may explain the missing correlation between soil sealing and flood extent as well as the difference between the citizen science campaign FQA and the automated supervised classification (SVC). The analysed data was retrieved from open data platforms like OpenStreetMap and official bodies like the Federal Ministry for Transport, Innovation and Technology or the national Electronic Navigational Chart. The calculation of the building density inside (2.18%) and outside (29.56%) of flooded areas, as well as the road density inside (2.03%) and outside (24.68%) of flooded areas surprisingly indicated a considerable negative correlation between sealed surfaces and the 2013 flood extent. The standard correlation value of both building footprints and road networks was -31.14%, which confirms the moderate negative correlation. This means that the higher the density of sealed surfaces was, the lower were the chances that the area was inundated. This also implied that the results of the supervised classification are more accurate and reliable than the results of the FQA campaign. The sealing densities inside and outside flooded areas for both methods (FQA and SVS) confirmed this statement.

The investigation of official data about flood protection measures and corresponding constructions revealed not only that there is unfortunately no national database about permanent protection structures, but also that not every measure taken actually worked. Some dams and walls worked absolutely fine, while only a short distance away, nearly every protective measure failed and caused material damage. Another finding was that the biggest inundated areas were meadow landscapes and therefore uninhabited and unsealed. The investigation of the terrain elevation revealed that there seems to be no connection between the elevation of the research area and the extent of the flooding.

All the performed analyses showed, that the information generated during the FQA campaign is not precise enough to determine potential correlations between soil sealing and the extent of the 2013 floods conclusively. A number of reasons can be listed of why the geographic citizen science approach was not successful in this case. The lack of participants, the low average validation per user, the low amount of validated sample points, the strong connection to the much more general LUCAS survey, the low density of sample points in the research area, the inaccessibility of some locations, the moderate accuracy compared to other land cover datasets, the rough estimations from the interpolation and the rather insignificant values of the correlation calculation all contributed largely to the unsuitability of this dataset for answering the research question. The analysis of traditionally generated information and open data however determined the mediocre negative correlation between soil sealing and the flood extent. In order to compete with official data, collected by trained experts, the FQA project has to identify its weaknesses and reduce and eradicate its errors in order to initiate a more successful and reliable successor project. The usability of the FQA data for spatial planning or disaster management is unfortunately not given at the moment. The accuracy and reliability are currently not high enough.

7. Literature

7.1. Published Resources

[AAS-13] ALBRECHT, F., AUBRECHT, C., STEINNOCHER, K. (2013): Kapitel 9: Klassifikation. Teaching material for "VU Bildverarbeitung und Fernerkundung", University of Vienna, IfGR, 45.

[ALB-14] ALBRECHT, F., ZUSSNER, M., PERGER, C., DÜRAUER, M., SEE, L., MCCALLUM, I., FRITZ, S., KAINZ, W. (2014): Using student volunteers to crowdsource land cover information. GI_Forum 2014, Geospatial Innovation for Society – Conference Proceedings, 314-317. doi: 10.1553/giscience2014s314

[ALC-02] ALCÁNTARA-AYALA, I. (2002): Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. Geomorphology 47, 107-124.

[ALE-13] ALEXANDER, D. (2013): Resilience and disaster risk reduction: an etymological journey. Natural Hazards and Earth System Sciences 13, 2707-2716.

[ALE-16] ALEXANDER, D. (2016): The game changes: "Disaster Prevention and Management" after a quarter of a century. Disaster Prevention and Management 25, 2-10.

[ALF-15] ALFIERI, L., FEYEN, L., DOTTORI, F., BIANCHI, A. (2015): Ensemble flood risk assessment in Europe under high end climate scenarios. Global Environmental Change 35, 199-212.

[ALI-14] ALI, A., SCHMID, F. (2014): Data Quality Assurance for Volunteered Geographic Information. In: DUCKHAM, M., PEBESMA, E., STEWART, K., FRANK, A. (Eds.): Geographic Information Science. GIScience 2014, Lecture Notes in Computer Science 8728. Springer, Cham, 126-141.

[ALR-08] ALROY, J. (2008): Dynamics of origination and extinction in the marine fossil record. Proceedings of the National Academy of Sciences of the United States of America 105, 11536-11542.

[ART-14] ARTMANN, M. (2014): Institutional efficiency of urban soil sealing management – From raising awareness to better implementation of sustainable development in Germany. Landscape and Urban Planning 131, 83-95.

[ART-15] ARTMANN, M. (2015): Managing urban soil sealing in Munich and Leipzig (Germany) – From a wicked problem to clumsy solutions. Land Use Policy 46, 21-37.

[AYE-08] AYERS, P., MATTHEWS, C., YATES, B. (2008): How Wikipedia works and how you can be a part of it. No Starch Press, San Francisco, 3-58.

[BAL-13] BALICA, S., POPESCU, I., BEEVERS, L., WRIGHT, N. (2013): Parametric and physically based modelling techniques for flood risk and vulnerability assessment: A comparison. Environmental Modelling & Software 41, 84-92.

[BAR-12] BARKAN, S. (2012): A Primer on Social Problems. 1st edition published online under creative commons licence. Chapter 14. Available at: http://2012books.lardbucket.org/pdfs/a-primer-on-social-problems.pdf

[BAY-16] BAYAS, J. C. L., SEE, L., FRITZ, S., STURN, T., PERGER, C., DÜRAUER, M., KARNER, M., MOORTHY, I., SCHEPASCHENKO, D., DOMIAN, D., MCCALLUM, I. (2016): Crowdsourcing In-Situ Data on Land Cover and Land Use Using Gamification and Mobile Technology. Remote Sensing 8 (11), 905.

[BEL-14] BELLO, O., AINA, Y. (2014): Satellite remote sensing as a tool in disaster management and sustainable development: towards a synergistic approach. Procedia – Social and Behavioural Sciences 120, 365-373.

[BEN-16] BENNETT, D., PHILLIPS, B., DAVIS, E. (2016): The future of accessibility in disaster conditions: How wireless technologies will transform the life cycle of emergency management. Futures (in press), http://dx.doi.org/10.1016/j.futures.2016.05.004.

[BET-93] BETTS, J. (1993): John Harrison: inventor of the precision timekeeper. Endeavour 17, new series, 160-167.

[BLL-16] BALLABIO, C., PANAGOS, P., MONATANARELLA, L. (2016): Mapping topsoil physical properties at European scale using the LUCAS database. Geoderma 261, 110-123.

[BLO-13] BLÖSCHL, G., NESTER, T., KOMMA, J., PARAJKA, J., PERIGAO, R. (2013): The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods. Hydrology and Earth System Sciences 17, 5197-5212.

[BRO-13] BROWN, M., SHARPLES, S., HARDING, J., PARKER, C., BEARMAN, N., MAGUIRE, M., FORREST, D., HAKLAY, M., JACKSON, M. (2013): Usability of Geographic Information: Current challenges and future directions. Applied Ergonomics 44, 855-865.

[BRU-08] BRUINS, H., MACGILLIVRAY, J. A., SYNOLAKIS, C., BENJAMINI, C., KELLER, J., KISCH, H., KLÜGEL, A., VAN DER PLICHT, J. (2008): Geoarchaeological tsunami deposits at Palaikastro (Crete) and the Late Minoan IA eruption of Santorini. Journal of Archaeological Science 35, 191-212.

[BUC-16] BUCHECKER, M., OGASA, D., MAIDL, E. (2016): How well do the wider public accept integrated flood risk management? An empirical study in two Swiss Alpine valleys. Environmental Science & Policy 55, 309-317.

[BUR-78] BURTON, I., KATES, R., WHITE, G. (1978): The Environment as Hazard. Oxford University Press, New York.

[CAL-16] CALLAGHAN, C. (2016): Disaster management, crowdsourced R&D and probabilistic innovation theory: Toward real time disaster response capability. International Journal of Disaster Risk Reduction 17, 238-250.

[CAM-14] CAMPONOVO, M., FREUNDSCHUH, S. (2014): Assessing uncertainty in VGI for emergency response. Cartography and Geographic Information Science 41, 440-455.

[CHA-94] CHAPMAN, D. (1994): Natural Hazards. Oxford University Press, Oxford, 174.

[CHO-17] CHOI, M., GLASSMAN, M., CRISTOL, D. (2017): What it means to be a citizen in the internet age: Development of a reliable and valid digital citizenship scale. Computers & Education 107, 100-112.

[COL-15] COLLINS, A., MANYENA, B., JAYAWICKRAMA, J., JONES, S. (2015): Introduction: Hazards, Risks, and Disasters in Society. In: COLLINS, A., JONES, S., MANYENA, B., WALSH, S., SHRODER, J. (Eds.): Hazards, Risks, and Disasters in Society. Elsevier, Amsterdam, 1-15.

[COP-06] COPPOLA, D. (2006): Introduction to International Disaster Management. Butterworth-Heinemann, Burlington, 1-30.

[COR-16] CORNING, E., SADEGHPOUR, A., KETTERINGS, Q., CZYMMEK, K. (2016): The Carbon Cycle and Soil Organic Carbon. Cornell University, Agronomy Fact Sheet 91, 1-2.

[CRA-12] CRABTREE, A. (2012): Climate change and mental health following flood disasters in developing countries, A review of the epidemiological literature: What do we know, what is being recommended? Australasian Journal of Disaster and Trauma Studies 1, 21-30.

[CRG-15] CRAGUN, R. (2015): Science and Religion. International Encyclopedia of the Social & Behavioral Sciences 21, 172-175.

[DOC-12] DOCIU, M., DUNARINTU, A. (2012): The Socio-Economic Impact of Urbanization. International Journal of Academic Research in Accounting, Finance and Management Sciences 2, 47-52.

[DUS-15] DU, S., SHI, P., VAN ROMPAEY, A., WEN, J. (2015): Quantifying the impact of impervious surface location on flood peak discharge in urban areas. Natural Hazards 76, 1457.

[EEA-16] EUROPEAN ENVIRONMENT AGENCY (2016): Rivers and lakes in European cities. Past and future challenges. Luxembourg, EEA Report 26, 11-19.

[ELM-13] ELMQVIST, T., REDMAN, C., BARTHEL, S., COSTANZA, R. (2013): History of Urbanization and the Missing Ecology. In: Urbanization, Biodiversity and Ecosystem Services: Challenges and Opportunities, Springer Netherlands, 13-30.

[ENC-08] INLAND ELECTRONIC NAVIGATIONAL CHART HARMONIZATION GROUP (2008): Inland Electronic Navigational Chart Encoding Guide, Inland ENC Harmonization Group, Version 3.1, 25-160.

[EST-12] ESTELLÉS-AROLAS, E., GONZÁLEZ-LADRÓN-DE-GUEVARA, F. (2012): Towards an integrated crowdsourcing definition. Journal of Information Science 38, 189-200.

[EUC-07] COUNCIL OF THE EUROPEAN UNION (2007): Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks. Official Journal of the European Union L288, 27-34.

[EUC-12] EUROPEAN COMMISSION (2012): Guidelines on best practice to limit, mitigate or compensate soil sealing. Commission staff working document. European Commission, Brussels, 1-65.

[EUS-13] EUROSTAT (2013): LUCAS 2012 – Land Use / Cover Area Frame Survey. Sectoral and Regional Statistics, European Commission – Eurostat, Luxembourg, 4-73.

[EVA-14] EVANS, M. (2014): Water and carbon cycling – New A-Level Subject Content Overview. Royal Geographical Society, Advancing geography and geographical learning, 1-19.

[FAO-85] FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS (1985): Irrigation Water Management: Introduction to Irrigation. Training Manual No. 1, Natural Resources Management and Environment Department, Chapter 2.

[FEI-13] FEICK, R., ROCHE, S. (2013): Chapter 2 – Understanding the Value of VGI. In: SUI, D., ELWOOD, S., GOODCHILD, M. (Eds.): Crowdsourcing Geographic Knowledge. Volunteered Geographic Information (VGI) in Theory and Practice. Springer, Dordrecht, 15-30.

[FEL-16] FELDBACHER, E., PAUN, M., RECKENDORFER, W., SIDOROFF, M., STANICA, A., STRIMBU, B., TUSA, I., VULTURESCU, V., HEIN, T. (2016): Twenty years of research on water management issues in the Danube Macro-region – past developments and future directions. Science of the Total Environment 572, 1297-1306.

[FER-03] FERRIER, N., HAQUE, E. (2003): Hazards Risk Assessment Methodology for Emergency Managers: A Standardized Framework for Application. Natural Hazards 28, 271-290.

[FRA-01] FRANCOIS, R. (2001): The Origin of Water on Earth. Science 293 (5532), 1056-1058.

[GAL-08] GALLEGO, J., BAMPS, C. (2008): Using CORINE land cover and the point survey LUCAS for area estimation. International Journal of Applied Earth Observation and Geoinformation 10, 467-475.

[GAP-10] GANAPATI, S. (2010): Using Geographic Information Systems to Increase Citizen Engagement. IBM Center for The Business of Government, Washington D.C., 13-28.

[GAR-13] GARCÍA, P., PÉREZ, M., GUERRA, A. (2013): Using TM images to detect soil sealing change in Madrid (Spain). Geoderma 214-215, 135-140.

[GIO-12] GIOSAN, L., COOLEN, M., KAPLAN, J., CONSTANTINESCU, S., FILIP, F., FILIPOVA-MARINOVA, M., KETTNER, A., THOM, N. (2012): Early Anthropogenic Transformation of the Danube-Black Sea System. Scientific Reports 2, 582.

[GON-16] GONZÁLEZ-HERNÁNDEZ, Y., CEDDIA, M., ZEPHAROVICH, E., CHRISTOPOULUS, D. (2016): Prescriptive conflict prevention analysis: An application to the 2021 update of the Austrian flood risk management plan. Environmental Science & Policy 66, 299-309.

[GOO-08] GOODCHILD, M. (2008): Commentary: whither VGI? GeoJournal 72, 239-244.

[GOO-10] GOODCHILD, M., GLENNON, J. A. (2010): Crowdsourcing geographic information for disaster response: a research frontier. International Journal of Digital Earth 3, 231-241.

[GOR-15] GONZÁLEZ-RIANCHO, P., GERKENSMEIER, B., RATTER, B., GONZÁLEZ, M., MEDINA, R. (2015): Storm surge risk perception and resilience: A pilot study in the German North Sea coast. Ocean & Coastal Management 112, 44-60.

[GOT-11] GOTO, K., CHAGUE-GOFF, C., FUJINO, S., GOFF, J., JAFFE, B., NISHIMURA, Y., RICHMOND, B., SUGAWARA, D., SZCZUCINSKI, W., TAPPIN, D. R., WITTER, R. C., YULIANTO, E. (2011): New insights of tsunami hazard from the 2011 Tohoku-oki event. Marine Geology 290, 46-50.

[GRI-16] GRILLAKIS, M., KOUTROULIS, A., KOMMA, J., TSANIS, I., WAGNER, W., BLÖSCHL, G. (2016): Initial soil moisture effects on flash flood generation – A comparison between basins of contrasting hydro-climatic conditions. Journal of Hydrology 541, 206-217.

[GRV-01] GRAVLEY, D. (2001): Risk, Hazard, and Disaster. Extract of doctoral thesis, University of Canterbury in New Zealand, 1-7.

[HAK-13] HAKLAY, M. (2013): Citizen Science and Volunteered Geographic Information – overview and typology of participation. In: SUI, D., ELWOOD, S., GOODCHILD, M. (Eds.): Crowdsourcing Geographic Knowledge. Volunteered Geographic Information (VGI) in Theory and Practice. Springer, Dordrecht, 105-122.

[HAN-10] HAND, E. (2010): People power. Networks of human minds are taking citizen science to a new level. Nature 466, London, 685-687.

[HIR-13] HIRTH, M., HOSSFELD, T., TRAN-GIA, P. (2013): Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. Mathematical and Computer Modelling 57, 2918-2932.

[HOW-06] HOWE, J. (2006): The Rise of Crowdsourcing. Wired Magazine 14.06, San Francisco, 1-5. Available under: http://sistemas-humano-computacionais.wdfiles.com/local--files/capitulo%3Aredes-sociais/Howe_The_ Rise_of_Crowdsourcing.pdf (15.02.2017)

[HRD-13] HARDING, J. (2013): Usability of geographic information – Factors identified from qualitative analysis of task-focused interviews. Applied Ergonomics 44, 940-947.

[HUY-14] HUYCK, C., VERRUCCI, E., BEVINGTON, J. (2014): Remote Sensing for Disaster Response: A Rapid, Image-Based Perspective. Earthquake Hazard, Risk and Disasters, 1-24.

[IAS-13] INSTITUTE FOR ADVANCED SUSTAINABILITY STUDIES (2013): Soil Sealing. Global Soil Forum, Potsdam, 1-4.

[JOE-14] JOERIN, J., SHAW, R., KRISHNAMURTHY, R. (2014): The concept of resilience to disasters. In: JOERIN, J., SHAW, R., KRISHNAMURTHY, R. (Eds.): Building Resilient Urban Communities (Community, Environment and Disaster Risk Management 15. Emerald Group Publishing, 35-48.

[JOH-13] JOHNSON, P., SIEBER, R. (2013): Chapter – 5: Situating the Adoption of VGI by Government. In: SUI, D., ELWOOD, S., GOODCHILD, M. (Eds.): Crowdsourcing Geographic Knowledge. Volunteered Geographic Information (VGI) in Theory and Practice. Springer, Dordrecht, 65-82.

[JOY-09] JOYCE, K., WRIGHT, K., SAMSONOV, S., AMBROSIA, V. (2009): Remote sensing and the disaster management cycle. Advances in Geoscience and Remote Sensing, 317-346.

[JUN-14] JUNGWIRTH, M., HAIDVOGL, G., HOHENSINNER, S., WAIDBACHER, H., ZAUNER, G. (2014): Österreichs Donau. Landschaft – Fisch – Geschichte. Universität für Bodenkultur Wien, 170-179.

[KOR-11] KORDZAKHIA, G., SHENGELIA, L., TVAURI, G., TATISHVILI, M., MKURNALIDZE, I. (2011): Remote sensing for early warning of natural meteorological and hydrological disasters and provision of transportation safety over the Black Sea in Georgia. Procedia Social and Behavioural Sciences 19, 532-536.

[LFW-15] BUNDESMINISTERIUM FÜR LAND- UND FORSTWIRTSCHAFT, UMWELT UND WASSERWIRTSCHAFT (2015): Hochwasser Juni 2013 Ereignisdokumentation. Wien, 13-199.

[LIN-15] LIN, W. (2015): Citizen Mapping. International Encyclopedia of the Social & Behavioral Sciences 2, 619-624

[MCL-85] MCLOUGHLIN, D. (1985): A Framework for Integrated Emergency Management. Public Administration Review 45 (Emergency Management: A Challenge for Public Administration), 165-172.

[MIL-14] MILLER, J., GREBBY, S. (2014): Mapping long-term temporal change in imperviousness using topographic maps. International Journal of Applied Earth Observation and Geoinformation 30, 9-20.

[NRC-10] NATIONAL RESEARCH COUNCIL OF THE NATIONAL ACADEMIES (2010): Toward Sustainable Agricultural Systems in the 21st Century. The National Academies Press, Washington D.C., 20-55.

[OLL-16] OLLOBARREN, P., CAPRA, A., GELSOMINO, A., LA SPADA, C. (2016): Effects of ephemeral gully erosion on soil degradation in a cultivated area in Sicily (Italy). Catena 145, 334-345.

[PAN-12] PANAGOS, P., MEUSBURGER, K., ALEWELL, C., MONTANARELLA, L. (2012): Soil erodibility estimation using LUCAS point survey data of Europe. Environmental Modelling & Software 30, 143-145.

[PAN-13] PANAGOS, P., BALLABIO, C., YIGINI, Y., DUNBAR, M. (2013): Estimating the soil organic carbon content for European NUTS2 regions based on LUCAS data collection. Science of the Total Environment 442, 235-246.

[PAP-15] PAPPENBERGER, F., CLOKE, H., PARKER, D., WETTERHALL, F., RICHARDSON, D., THIELEN, J. (2015): The monetary benefit of early flood warnings in Europe. Environmental Science & Policy 51, 278-291.

[PAU-17] PAUL, B. K., RASHID, H. (2017): Tropical Cyclones and Storm Surges. Climatic Hazards in Coastal Bangladesh, 35-81.

[PEN-05] PENDER, G. (2005): Flood risk management. Water and Sewerage Journal 4, 9.

[PIS-15] PISTOCCHI, A., CALZOLARI, C., MALUCELLI, F., UNGARO, F. (2015): Soil sealing and flood risks in the plains of Emilia-Romagna, Italy. Journal of Hydrology: Regional Studies 4, 398-409.

[REV-16a] REVILLA-ROMERO, B. (2016): Improving Global Flood Forecasting using Satellite Detected Flood Extent. Dissertation, Faculty of Geosciences, Utrecht University.

[RIC-16] RICCARDI, M. (2016): The power of crowdsourcing in disaster response operations. International Journal of Disaster Risk Reduction 20, 123-128.

[RIE-14] RIESCH, H., POTTER, C. (2014): Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. Public Understanding of Science 23, 107-120.

[ROY-15] ROY, A. (2015): Crowdsourcing. In: WHERRY, F., SCHOR, J. (Eds.): The SAGE Encyclopedia of Economics and Society. SAGE, Los Angeles, 1-6.

[SAD-17] SADEGHI-POUYA, A., NOURI, J., MANSOURI, N., KIA-LASHAKI, A. (2017): An indexing approach to assess flood vulnerability in the western coastal cities of Mazandaran, Iran. International Journal of Disaster Risk Reduction, article in press, 1-13.

[SCN-88] SCANLON, T. J. (1988): Disaster's little known pioneer: Canada's Samuel Henry Prince. International Journal of Mass Emergencies and Disasters 6, 213-232.

[SEE-16] SEE, L., MOONEY, P., FOODY, G., BASTIN, L., COMBER, A., ESTIMA, J., FRITZ, S., KERLE, N., JIANG, B., LAAKSO, M., LIU, H. Y., MILCINSKI, G., NIKSIC, M., PAINHO, M., PODÖR, A., OLTEANU-RAIMOND, A. M., RUTZINGER, M. (2016): Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. ISPRS International Journal of Geo-Information 5, 55, 1-23.

[SIO-17] SIONE, S. M. J., WILSON, M. G., LADO, M., GONZALEZ, A. P. (2017): Evaluation of soil degradation produced by rice crop systems in a Vertisol, using a soil quality index. Catena 150, 79-86.

[SUI-13] SUI, D., ELWOOD, S., GOODCHILD, M. (2013): Chapter 1 – Volunteered Geographic Information, the Exaflood, and the Growing Digital Divide. In: SUI, D., ELWOOD, S., GOODCHILD, M. (Eds.): Crowdsourcing Geographic Knowledge. Volunteered Geographic Information (VGI) in Theory and Practice. Springer, Dordrecht, 1-14.

[TAN-94] TANTON, J. (1994): End of the Migration Epoch? Time for a new Paradigm. The Social Contract Press 4, 162-173.

[TOR-04] TORREY, B. B. (2004): Urbanization: An Environmental Force to Be Reckoned With. Population Reference Bureau – Inform. Empower. Advance. Available under: (10.01.2017) http://www.prb.org/Publications/Articles/2004/UrbanizationAnEnvironmentalForcetoBeReckonedWith.aspx

[TSE-15] TSENDBAZAR, N., DE BRUIN, S., HEROLD, M. (2015): Assessing global land cover reference datasets for different user communities. ISPRS Journal of Photogrammetry and Remote Sensing 103, 93-114.

[TUC-07] TUCCI, C. (2007): Urban Flood Management: Urban drainage, Flooding, Urbanization. World Meteorological Organization, Geneva. 127-143.

[VAS-08] VASILESCU, L., KHAN, A., KHAN, H. (2008): Disaster Management Cycle – a theoretical approach. Management & Marketing – Craiova 1, 43-50.

[VIT-15] HABERSACK, H., PUCHER, K., SCHOBER, B., SAMEK, R., HACKEL, C. (2015): Hochwasserdokumentation Donau 2013 – Ereignisdokumentation. Bundesministerium für Verkehr, Innovation und Technologie (BMVIT). Wien, 5-239.

[WHI-01] WHITE, G., KATES, R., BURTON, I. (2001): Knowing better and losing even more: the use of knowledge in hazards management. Environmental Hazards 3, 81-92.

[WIL-07] WILSON, S., TEMPLE, B., MILLIRON, M., VAZQUEZ, C., PACKARD, M., RUDY, B. (2007): The Lack of Disaster Preparedness by the Public and it's Affect on Communities. The Internet Journal of Rescue and Disaster Medicine 7, 1-8.

[WIS-01] WISNER, B. (2001): Disasters: what the United Nations and its world can do. Environmental Hazards 3, 125-127.

[WIT-15] WITHERS, C. (2015): The longitude question. Journal of Historical Geography 47, 1-5.

[XUR-17] XU, R., ZHANG, H., WANG, T., LIN, H. (2017): Using pan-sharpened high resolution satellite data to improve impervious surfaces estimation. International Journal of Applied Earth Observation and Geoinformation 57, 177-189.

[YAH-07] YANKO-HOMBACH, V., GILBERT, A., DOLUKHANOV, P. (2007): Controversy over the great flood hypotheses in the Black Sea in light of geological, paleontological, and archaeological evidence. Quaternary International 167-168, 91-113.

[YAL-04] YALCIN, G., AKYUREK, Z. (2004): Analysing Flood Vulnerable Areas with Multicriteria Evaluation. International Society for Photogrammetry and Remote Sensing, Congress Proceedings 35, 1-6.

[YAT-11] YATES, D., PAQUETTE, S. (2011): Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. International Journal of Information Management 31, 6-13.

[ZHA-16] ZHANG, H., LIN, H., LI, Y., ZHANG, Y., FANG, C. (2016): Mapping urban impervious surface with dual-polarimetric SAR data: An improved method. Landscape and Urban Planning 151, 55-63.

7.2. Online Resources

[ADP-17] ASIAN DISASTER PREPAREDNESS CENTER (2017): Capacity Building in Asia using Information Technology Applications – Concepts of Hazards, Disasters and Hazard Assessment. Online: http://www.adpc.net/casita/course-materials/Mod-2-Hazards.pdf, 27.01.2017

[AGA-17a] AUSTRALIAN GOVERNMENT (2017): What causes Floods? Online: http://www.ga.gov.au/scientific-topics/hazards/flood/basics/causes, 01.01.2017

[AGA-17b] AUSTRALIAN GOVERNMENT (2017): Where do Floods occur? Online: http://www.ga.gov.au/scientific-topics/hazards/flood/basics/where, 01.01.2017

[AHE-14] ANCIENT HISTORY ENCYCLOPEDIA (2014): Urbanization. Online: http://www.ancient.eu/urbanization, 09.01.2017

[AUF-17] AUSTRIA FORUM (2017): Donau. Online: http://austria-forum.org/af/AustriaWiki/Donau, 10.01.2017

[AWP-17] ALBERTA WATER PORTAL (2017): What is Snowmelt? Online: http://albertawater.com/what-is-snowmelt, 02.12.2016

[BBC-16] BRITISH BROADCASTING CORPORATION (2016): River flooding and management issues. Online: http://www.bbc.co.uk/schools/gcsebitesize/geography/water_rivers/river_flooding_management_rev1.shtml, 02.12.2016

[BBC-17] BRITISH BROADCASTING CORPORATION (2017): Flood prevention being ignored – MPs. Online: http://www.bbc.com/news/science-environment-38721486, 13.02.2017

[BIO-17] BIOLOGYDISCUSSION (2017): What are the Main Causes of Soil Degradation? Online: http://www.biologydiscussion.com/soil/what-are-the-main-causes-of-soil-degradation/7276, 14.01.2017

[BJV-17] BUNDESMINISTERIUM DER JUSTIZ UND FÜR VERBRAUCHERSCHUTZ (2017): Gesetz zum Schutz vor schädlichen Bodenveränderungen zur Sanierung von Altlasten (Bundes-Bodenschutzgesetz BBodSchG). Online: http://www.gesetze-im-internet.de/bbodschg/__2.html, 06.01.2017

[BMI-17] BUNDESMINISTERIUM FÜR INNERES (2017): Civil Protection in Austria – Disaster Management. Online: http://www.bmi.gv.at/cms/BMI_Zivilschutz_en/management/start.aspx, 03.02.2017

[BRT-16a] ENCYCLOPAEDIA BRITANNICA (2016): Devonian – extinctions. Online: https://www.britannica.com/science/Devonian-extinctions, 28.11.2016

[BRT-16b] ENCYCLOPAEDIA BRITANNICA (2016): End-Triassic extinction. Online: https://www.britannica.com/science/end-Triassic-extinction, 28.11.2016

[BRT-16c] ENCYCLOPAEDIA BRITANNICA (2016): K-T extinction. Online: https://www.britannica.com/science/K-T-extinction, 28.11.2016

[BRT-16d] ENCYCLOPAEDIA BRITANNICA (2016): Ordovician - Silurian extinction. Online: https://www.britannica.com/science/Ordovician-Silurian, 28.11.2016

[BRT-16e] ENCYCLOPAEDIA BRITANNICA (2016): Permian extinction. Online: https://www.britannica.com/science/Permian-extinction, 28.11.2016

[CAM-16a] CAMBRIDGE DICTIONARY (2016): Flood. Online: http://dictionary.cambridge.org/dictionary/-english/flood, 30.11.2016

[CAM-16b] CAMBRIDGE DICTIONARY (2016): Floodplain. Online: http://dictionary.cambridge.org/dictionary/-english/floodplain, 01.12.2016

[CAM-16c] CAMBRIDGE DICTIONARY (2016): Tsunami. Online: http://dictionary.cambridge.org/dictionary/-english/tsunami?fallbackFrom=learner-english, 02.12.2016

[CAM-16d] CAMBRIDGE DICTIONARY (2016): Soil. Online: http://dictionary.cambridge.org/dictionary/learner-english/soil_1, 04.12.2016

[CAM-17a] CAMBRIDGE DICTIONARY (2017): Deforestation. Online: http://dictionary.cambridge.org/de/worter-buch/learner-englisch/deforestation, 14.01.2017

[CAM-17b] CAMBRIDGE DICTIONARY (2017): Urbanization. Online: http://dictionary.cambridge.org/dictionary-/english/urbanization?fallbackFrom=learner-english&q=urbanisation, 08.01.2017

[CAM-17c] CAMBRIDGE DICTIONARY (2017): Hazard. Online: http://dictionary.cambridge.org/dictionary/-learner-english/hazard_1, 24.01.2017

[CCE-15] COMEDY CENTRAL (2015): The Daily Show with Jon Stewart. Online: http://www.cc.com/sitewide/media_player/play.jhtml?itemId=23520, 24.10.2015

[CET-11] CIVIL ENGINEERING TERMS (2011): Definition of Permeability | Pervious Soil | Impervious Soil. Online: http://www.civilengineeringterms.com/soil-mechanics-1/definition-of-permeability-pervious-soil-impervious-soil, 11.01.2017

[CIA-17] CENTRAL INTELLIGENCE AGENCY (2017): The World Factbook. Online: https://www.cia.gov/library/publications/the-world-factbook, 15.01.2017

[CLC-14] CLIMATE CENTRAL (2014): New Analysis Shows Global Exposure to Sea Level Rise. Online: http://www.climatecentral.org/news/new-analysis-global-exposure-to-sea-level-rise-flooding-18066, 12.02.2017

[CLM-17] GLOBAL CLIMATE CHANGE BY NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (2017): Soil – News and Features. Online: https://climate.nasa.gov, 08.01.2017

[CNE-04] CENTRE NATIONAL D'ETUDES SPATIALES (2004): Spot 4 Imagery. Online: http://spot4.cnes.fr/spot4_gb/index.htm, 20.01.2017

[CPC-17] COPERNICUS (2017): Europe's eyes on Earth – Copernicus in brief. Online: http://www.copernicus.eu/main/copernicus-brief, 14.10.2017

[CRH-10] NATIONAL WEATHER FORECAST OFFICE (2010): Hurricane Ike (September 2008) Damage Photos. Online: http://www.crh.noaa.gov/hgx/?n=projects_ike08_bolivar1, 05.01.2017

[CRT-16] UNITED STATES CLIMATE RESILIENCE TOOLKIT (2016): Inland Flooding. Online: https://toolkit.climate.gov/topics/coastal-flood-risk/inland-flooding, 03.12.2016

[CSN-15] CITIZEN SCIENCE NEWS (2015): FotoQuest Austria wurde gestartet. Online: http://www.citizen-science.at/news/item/94-fotoquest-austria-wurde-gestartet, 27.02.2017

[DON-14] DATAONE (2014): Difference between Citizen Science and VGI. Online: https://notebooks.dataone.org/citsci-data/difference-between-citizen-science-and-vgi-week-6, 18.02.2017

[DUE-05] ESA – DATA USER ELEMENT (2005): GlobCover. Online: http://due.esrin.esa.int/-page_globcover.php, 20.01.2017

[EAA-17] ENVIRONMENT AGENCY AUSTRIA (2017): BORIS SOIL Information System in Austria. Online: http://www.umweltbundesamt.at/en/services/services_resources/services_soil_spatial/en_ref_boris 07.01.2017

[EEN-11] EUROPEAN ENVIRONMENT AGENCY (2011): Soil sealing and population density in the capitals of EEA countries and the Western Balkans. Online: https://www.eea.europa.eu/data-and-maps/figures/soil-sealing-in-the-capitals/vienna-eps-file, 24.04.2017

[EEN-17] EUROPEAN ENVIRONMENT AGENCY (2017): Average annual increase of sealing. Online: http://www.eea.europa.eu/data-and-maps/figures/average-annual-increase-of-sealing, 20.01.2017

[EEX-17] EARTHEXPLORER (2017): EarthExplorer by the United States Geological Survey. Online: https://earthexplorer.usgs.gov, 21.05.2017

[EFR-02] DEPARTMENT FOR ENVIRONMENT, FOOD & RURAL AFFAIRS (2002): Flood and Coastal Defence R&D Programme – Temporary and Demountable Flood Protection. Online: http://www.aquadam.net/pdf/130_1_ig.pdf, 13.02.2017

[ENC-17] Electronic Navigational Charts (2017): Donau River Information Services – Inland ENCs. Online: http://www.doris.bmvit.gv.at/en/maps-charts/inland-encs/, 22.06.2017

[EPA-17] ENVIRONMENTAL PROTECTION AGENCY (2017): Corine Land Cover Mapping. Online: http://www.epa.ie/soilandbiodiversity/soils/land/corine, 20.01.2017

[EPB-13] EMERGENCY PLANNING BLOG BY DAVID ALEXANDER (2013): The Bouncy Castle of North London and the Origins of Resilience. Online: http://emergency-planning.blogspot.co.at/2013/02/the-bouncy-castle-of-north-london-and.html, 26.01.2017

[EPB-17] EMERGENCY PLANNING BLOG BY DAVID ALEXANDER (2017): Explosions: A Curious Coincidence. Online: http://emergency-planning.blogspot.co.at/2017/02/explosions-curious-coincidence.html, 27.01.2017

[ESC-10] ESCHOOL TODAY (2010): What causes flooding? Online: http://eschooltoday.com/natural-disasters/floods/what-causes-floods.html, 01.12.2016

[ETY-16a] ONLINE ETYMOLOGY DICTIONARY (2016): Disaster. Online: http://etymonline.com/?term=disaster, 28.11.2016

[ETY-16b] ONLINE ETYMOLOGY DICTIONARY (2016): Flood. Online: http://etymonline.com/index.php?-term=flood&allowed_in_frame=0, 30.11.2016

[EUC-16a] EUROPEAN COMMISSION (2016): The EU Floods Directive – What is flood risk management? Online: http://ec.europa.eu/environment/water/flood_risk/flood_risk.htm, 13.02.2017

[EUC-16b] EUROPEAN COMMISSION (2016): Soil. Online: http://ec.europa.eu/environment/soil/index_en.htm, 04.12.2016

[EUC-16c] EUROPEAN COMMISSION (2016): International Year of Soils 2015. Online: http://ec.europa.eu/environment/soil/iys2015/index_en.htm, 15.01.2017

[EUC-16d] EUROPEAN COMMISSION (2016): Soil. Online: http://ec.europa.eu/environment/soil/-sealing_guidelines.htm, 14.01.2017

[EUC-17] EUROPEAN COMMISSION (2017): The economics of climate change adaption in EU coastal areas – The Netherlands. Online: https://ec.europa.eu/maritimeaffairs/sites/maritimeaffairs/files/-docs/body/netherlands_climate_change_en.pdf, 12.02.2017

[FAO-00] FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS (2016): Land Cover Classification System (LCCS): A Dichotomous, Modular-Hierarchical Approach Classification Concepts. Online: http://www.fao.org/docrep/003/x0596e/X0596e00.htm#P-1_0, 20.01.2017

[FAO-16] FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS (2016): Soil sealing. Online: http://www.fao.org/3/a-i6470e.pdf, 15.01.2017

[FAO-17a] FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS (2017): FAO Soils Portal. Online: http://www.fao.org/soils-portal/en, 01.01.2017

[FAO-17b] FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS (2017): FAO Soils Portal – Biological properties. Online: http://www.fao.org/soils-portal/soil-survey/soil-properties/biological-properties/en, 06.01.2017

[FAO-17c] FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS (2017): FAO Soils Portal – Physical properties. Online: http://www.fao.org/soils-portal/soil-survey/soil-properties/physical-properties/en, 07.01.2017

[FAO-17d] FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS (2017): FAO Soils Portal – Legacy Soil Maps and Soils Databases. Online: http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/en, 07.01.2017

[FAO-17e] FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS (2017): FAO Soils Portal – Soil degradation. Online: http://www.fao.org/soils-portal/soil-degradation-restoration/en, 14.01.2017

[FAO-17f] FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS (2017): World deforestation decreases, but remains alarming in many countries. Online: http://www.fao.org/news/story/en/item/40893/icode, 14.01.2017

[FAO-17g] FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS (2017): Chemical properties. Online: http://www.fao.org/soils-portal/soil-survey/soil-properties/chemical-properties/en, 14.01.2017

[FAZ-17] FRANKFURTER ALLGEMEINE ZEITUNG (2017): Stärkste Ostsee-Sturmflut seit 2006. Online: http://www.faz.net/agenturmeldungen/dpa/staerkste-ostsee-sturmflut-seit-2006-14605349.html, 05.01.2017

[FDF-16] THE FREE DICTIONARY BY FARLEX (2017): Soil. Online: http://www.thefreedictionary.com/soil, 04.12.2016

[FDF-17] THE FREE DICTIONARY BY FARLEX (2017): Flood-prone area. Online: http://financial-dictionary.thefreedictionary.com/flood-prone+area, 12.02.2017

[FEM-16] FEDERAL EMERGENCY MANAGEMENT AGENCY (2016): Mission Areas. Online: https://www.fema.gov/mission-areas, 08.02.2017

[GEO-17] GEOGRAPHICAL ASSOCIATION (2017): Images of Kenya – Flash flood. Online: http://www.geography.org.uk/resources/photosforenquiry/imagesofkenya/flashflood, 05.01.2017

[GET-17] GETTY IMAGES (2017): Coastal Flood – Bilder und Fotos. Online: http://www.gettyimages.at/fotos/coastal-flood, 05.01.2017

[GDR-17] GLOBAL DEVELOPMENT RESEARCH CENTER (2017): The Disaster Management Cycle. Online: https://www.gdrc.org/uem/disasters/1-dm_cycle.html, 08.02.2017

[GWO-17] GEO-WIKI ONLINE (2017): Engaging Citizens in Environmental Monitoring. Online: http://geo-wiki.org/Application/index.php, 27.02.2017

[HAG-17] ÖSTERREICHISCHE HAGELVERSICHERUNG (2017): Schützen wir das Land, von dem wir leben. Online: https://www.hagel.at/site/index.cfm?objectid=22B124FC-5056-A500-6ACDE212863096-55, 10.01.2017

[HOT-17] HUMANITARIAN OPENSTREETMAP TEAM (2017): Humanitarian OpenStreetMap Team. Online: https://www.hotosm.org/about, 18.02.2017

[HYP-17] HYPERSOIL UNIVERSITÄT MÜNSTER (2017): Bodenhorizonte und Bodenprofil. Online: http://hypersoil.uni-muenster.de/0/04/06.htm, 06.01.2017

[IAM-13] IAMYOUASHEISME (2013): Journey to Perplexity – NYC, Storms, and Risk. Online: https://iamyouasheisme.wordpress.com/tag/floods, 05.01.2017

[IFR-16] INTERNATIONAL FEDERATION OF RED CROSS AND RED CRESCENT SOCIETIES (2016): Disaster Management – What is a disaster? Online: http://www.ifrc.org/en/what-we-do/disaster-management/about-disasters/what-is-a-disaster, 28.11.2016

[IFR-17] INTERNATIONAL FEDERATION OF RED CROSS AND RED CRESCENT SOCIETIES (2017): Disaster and crisis management. Online: http://www.ifrc.org/en/what-we-do/disaster-management/, 27.01.2017

[IHE-17] UNESCO-IHE INSTITUTE FOR WATER EDUCATION (2017): Flood Vulnerability Indices (FVI). Online: http://unescoihefvi.free.fr/vulnerability.php, 12.02.2017

[IIA-16] INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS (2016): What is IIASA? Online: http://www.iiasa.ac.at/web/home/about/whatisiiasa/what_is_iiasa.html, 27.02.2017

[JRC-13] JOINT RESEARCH CENTRE (2016): Towards a harmonised soil database for the Danube Basin. Online: http://eusoils.jrc.ec.europa.eu/projects/esoter/Danube/Presentations/Danube%20-%20Jones%20-%20towards%20a%20harmonised%20soil%20DB%20for%20the%20Danube%20basinDLSN-250k-KO.pdf, 23.01.2017

[JRC-16] JOINT RESEARCH CENTRE (2016): European Flood Awareness System (EFAS). Online: https://www.efas.eu, 13.02.2017

[JRC-17] JOINT RESEARCH CENTRE (2016): The Danube Land and Soil Nexus. Online: https://ec.europa.eu/jrc/sites/jrcsh/files/jrc-danube-land-soil-nexus.pdf, 23.01.2017

[LSC-13] LIVESCIENCE (2013): The 20 Cities Most Vulnerable to Flooding. Online: http://www.livescience.com/38956-most-vulnerable-cities-to-flooding.html, 12.02.2017

[LUC-16] LUCAS EUROSTAT (2016): LUCAS – Land use and land cover survey. Eurostat – Statistics Explained. Online: http://ec.europa.eu/eurostat/statistics-explained/index.php/LUCAS_-_Land_use_and_land_cover_survey#Defining_land_use.2C_land_cover_and_landscape, 17.04.2017

[LUF-17] Luftlinie.org (2017): Entfernung Ybbs an der Donau – Wien. Online: https://www.luftlinie.org/Ybbs-an-der-Donau/Wien, 24.01.2017

[MAS-16] MASHABLE (2016): The Halifax explosion – The naval accident that erased an entire city in Canada. Online: http://mashable.com/2016/02/15/halifax-explosion/#Xzpqi20NYuqB, 29.02.2017

[MCK-16] MCKINSEY & COMPANY (2016): Urban world: Meeting the demographic challenge in cities. Online: http://www.mckinsey.com/global-themes/urbanization/urban-world-meeting-the-demographic-challenge-in-cities, 08.01.2017

[MKA-07] MCKAY, J. (2007): Accuracy Assessment, Geography 581, 245-249. Online: http://web.pdx.edu/~jduh/courses/Archive/geog481w07/Students/Mckay_AccuracyAssessment.pdf, 05.05.2017

[MNE-17] M & E STUDIES (2017): Types of Capacities in Disaster Management. Online: http://www.mnestudies.com/disaster-management/capacity-types, 26.01.2017

[MOD-17] MODIS BY NASA (2017): MODIS Characterization Support Team. Online: https://mcst.gsfc.nasa.gov, 20.01.2017

[MOS-17] MOSE VENEZIA (2017): Barriere del Mose in Venezia. Online: https://www.mosevenezia.eu, 13.02.2017

[NAS-13] National Aeronautics and Space Administration (2013): Landsat Science – Landsat 8. Online: https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-bands, 14.05.2017

[NAT-13] NATURE – INTERNATIONAL WEEKLY JOURNAL OF SCIENCE (2013): Common source for Earth and Moon water. Online: http://www.nature.com/news/common-source-for-earth-and-moon-water-1.12963, 30.11.2016

[NBI-16] NEWSBISCUIT (2016): Thames barrier sabotage means city of London technically offshore. Online: http://www.newsbiscuit.com/2016/04/12/thames-barrier-sabotage-means-city-of-london-technically-offshore, 13.02.2017

[NEW-16] NEW SCIENTIST (2016): Floods in Europe will cause five times more damage by 2050. Online: https://www.newscientist.com/article/2075712-floods-in-europe-will-cause-five-times-more-damage-by-2050, 01.12.2016

[NHC-11] NATIONAL HURRICANE CENTER (2011): The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts). Online: http://www.nhc.noaa.gov/pdf/nws-nhc-6.pdf, 08.02.2017

[NOD-14] NOTEBOOKS DATAONE (2014): Citizen Science Gathering. Online: https://notebooks.dataone.org/citsci-data, 19.02.2017

[NOS-16] NATIONAL OCEAN SERVICE (2016): What is a tsunami? Online: http://oceanservice.noaa.gov/facts/tsunami.html, 02.12.2016

[NSS-16] NATIONAL SEVERE STORMS LABORATORY (2016): Severe Weather 101 – Floods. Online: http://www.nssl.noaa.gov/education/svrwx101/floods/types, 03.12.2016

[NTU-14] NANYANG TECHNOLOGICAL UNIVERSITY (2014): Deforestation Throughout Indonesia – Soil Erosion. Online: https://blogs.ntu.edu.sg/hp331-2014-42/?page_id=90, 14.01.2017

[NVZ-15] NUOVA VENEZIA (2015): MOSE nuovo rinvio la fine lavori slitta a giugno del 2018. Online: http://nuovavenezia.gelocal.it/venezia/cronaca/2015/06/12/news/mose-nuovo-rinvio-la-fine-lavori-slitta-a-giugno-del-2018-1.11601780?refresh_ce, 13.02.2017

[NWS-16] NATIONAL WEATHER SERVICE (2016): Introduction to storm surge. Online: http://www.nws.noaa.gov/om/hurricane/resources/surge_intro.pdf, 01.12.2016

[NWS-17] NATIONAL WEATHER SERVICE (2017): Hydrologic Information Center – Flood Loss Data. Online: http://www.nws.noaa.gov/hic, 13.02.2017

[OEC-03] ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (2003): Glossary of Statistical Terms – Urbanization. Online: https://stats.oecd.org/glossary/detail.asp?ID=2819, 08.01.2017

[OEH-14] OFFICE OF ENVIRONMENT & HERITAGE (2014): Wind erosion. Online: http://www.environment.nsw.gov.au/soildegradation/winder.htm, 14.01.2017

[OEH-17] OFFICE OF ENVIRONMENT & HERITAGE (2017): Dustwatch – Wind Erosion. Online: http://www.environment.nsw.gov.au/images/dustwatch/WindErosion_JLeysLG.jpg, 14.01.2017

[OER-12] OEROK ATLAS (2012): Anteil der versiegelten Fläche 2012, 1 km-Raster (gem. Copernicus-Programm). Online: http://www.oerok-atlas.at/#indicator/61, 10.01.2017

[OMA-15] ONTARIO MINISTRY OF AGRICULTURE, FOOD AND RURAL AFFAIRS (2015): Soil Erosion – Causes and Effects. Online: http://www.omafra.gov.on.ca/english/engineer/facts/12-053.htm, 14.01.2017

[PBZ-05] PRESENTATION BLOG ZEN (2005): FEMA chart becomes brunt of joke. Online: http://www.presentationzen.com/presentationzen/2005/10/fema_chart_beco.html, 08.02.2017

[PDX-12] PORTLAND STATE UNIVERSITY (2012): Satellite Image Classification & Change Detection – Accuracy Assessment. Online: http://web.pdx.edu/~nauna/resources/9-accuracyassessment.pdf, 05.05.2017

[PEO-05] PEOPLE CHINA (2005): After 30 years, secrets, lessons of China's worst dams burst accident surface. Online: http://en.people.cn/200510/01/eng20051001_211892.html, 01.12.2016

[PHY-14] PHYS.ORG (2014): How Earth was watered. Online: https://phys.org/news/2014-02-earth.html#jCp, 30.11.2016

[REL-00] RELIEFWEB (2000): Mozambique: Debt payments suspended. Online: http://reliefweb.int/report/mozambique/mozambique-debt-payments-suspended, 09.02.2017

[SCE-16] UNITED NATIONS SCIENTIFIC COMMITTEE ON THE EFFECTS OF ATOMIC RADIATION (2016): The Fukushima-Daiichi nuclear power plant accident. Online: http://www.unscear.org/unscear/en/fukushima.html, 13.02.2017

[SDR-15b] UNITED NATIONS OFFICE FOR DISASTER RISK REDUCTION (2015): Sendai Framework for Disaster Risk Reduction. Online: http://www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf, 10.02.2017

[SDR-17a] UNITED NATIONS OFFICE FOR DISASTER RISK REDUCTION (2017): UNISDR Home Page. Online: https://www.unisdr.org, 08.02.2017

[SDR-17b] UNITED NATIONS OFFICE FOR DISASTER RISK REDUCTION (2017): Terminology on DRR. Online: https://www.unisdr.org/we/inform/terminology, 03.02.2017

[SDR-17c] UNITED NATIONS OFFICE FOR DISASTER RISK REDUCTION (2017): What is Disaster Risk Reduction? Online: http://www.unisdr.org/who-we-are/what-is-drr, 24.01.2017

[SDR-17d] UNITED NATIONS OFFICE FOR DISASTER RISK REDUCTION (2017): Sendai Framework for Disaster Risk Reduction. Online: http://www.unisdr.org/we/coordinate/sendai-framework, 10.02.2017

[SGA-14] GAIKWAD, S., WADHVA, S. (2014): Disaster Management. Online: https://de.slideshare.net/drsagargaikwad/disaster-management-by-dr-sagar-gaikwad, 08.02.2017

[STA-17] STATISTIK AUSTRIA (2017): Bevölkerung zu Quartalsbeginn seit 2002 nach Bundesland. Online: http://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/bevoelkerungsstand_und_veraenderung/bevoelkerung_zu_jahres-_quartalsanfang/023582.html, 23.04.2017

[TEE-14] THE EAST END (2014): The Silvertown Explosion. Online: http://the-east-end.co.uk/the-silvertown-explosion, 29.02.2017

[THB-17] THE THAMES BARRIER (2017): Flooding and coastal change – The Thames Barrier. Online: https://www.gov.uk/guidance/the-thames-barrier, 13.02.2017

[UCL-13] UNIVERSITY COLLEGE LONDON (2013): A Brief History of Resilience. Online: https://www.slideshare.net/dealexander/a-brief-history-of-resilience, 26.01.2017

[UCL-17] UNIVERSITY COLLEGE LONDON (2017): Integrating Cascading Disasters in London's Response and Preparedness Strategies. Online: http://www.ucl.ac.uk/rdr/cascading/01-news-cascades/LRcascading, 10.02.2017

[UFL-12] URBAN FLOOD (2012): Introduction Urban Flood Early Warning System Framework. Online: http://www.urbanflood.eu/Pages/default.aspx, 05.01.2017

[UND-04] UNITED NATIONS DEVELOPMENT PROGRAMME (2004): Reducing Disaster Risk – A Challenge for Development. Online: http://www.planat.ch/fileadmin/PLANAT/planat_pdf/alle_2012/2001-2005/Pelling__Maskrey_et_al_2004_-_Reducing_Disaster_Risk.pdf, 10.02.2017

[UND-14] UNITED NATIONS DEVELOPMENT PROGRAMME (2014): Population Division – World Urbanization Prospects 2014. Online: https://esa.un.org/unpd/wup, 08.01.2017

[USD-01] UNITED STATES DEPARTMENT OF AGRICULTURE (2001): Land Degradation: An overview. Online: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/?cid=nrcs142p2_054028, 15.01.2017

[USG-93] UNITED STATES GEOLOGICAL SURVEY (1993): How much water is on Earth? Online: https://water.usgs.gov/edu/gallery/global-water-volume.html, 30.11.2016

[USG-16] UNITED STATES GEOLOGICAL SURVEY (2016): How much water is there on, in, and above the Earth? Online: https://water.usgs.gov/edu/earthhowmuch.html, 30.11.2016

[USH-17] USHAHIDI (2017): Ushahidi Tracker. Online: http://github.ushahidi.org/, 10.02.2017

[VVB-17] VODOHOSPODÁRSKA VYSTAVBA (2017): Optimalization of the water regime in the Danube river branch system in the stretch Dobrohošt. Online: http://www.vvb.sk/old.gabcikovo.gov.sk/doc/yellow/chapters/preklad.htm, 22.04.2017

[WBG-13] WORLD BANK GROUP (2013): Disaster Risk Management in Guatemala. Online: http://www.worldbank.org/en/results/2013/10/08/promoting-proactive-disaster-risk-management-in-Guatemala, 09.02.2017

[WBG-16] WORLD BANK GROUP (2016): World Bank national accounts data and OECS National Accounts data files. Online: http://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=MZ, 09.02.2017

[WEA-16] WEATHER ABOUT (2016): Urban Flooding. Online: http://weather.about.com/od/u/qt/urban_flooding.htm, 03.12.2016

[WEB-16] MERRIAM-WEBSTER ONLINE DICTIONARY (2016): Hypothesis. Online: https://www.merriam-webster.com/dictionary/hypothesis, 27.11.2016

[WEB-17a] MERRIAM-WEBSTER ONLINE DICTIONARY (2017): Capacity. Online: https://www.merriam-webster.com/dictionary/capacity, 26.01.2017

[WEB-17b] MERRIAM-WEBSTER ONLINE DICTIONARY (2017): Resilience. Online: https://www.merriam-webster.com/dictionary/resilience, 24.01.2017

[WEB-17c] MERRIAM-WEBSTER ONLINE DICTIONARY (2017): Risk. Online: https://www.merriam-webster.com/dictionary/risk, 24.01.2017

[WEB-17d] MERRIAM-WEBSTER ONLINE DICTIONARY (2017): Vulnerability. Online: https://www.merriam-webster.com/dictionary/vulnerability, 25.01.2017

[WHO-17] WORLD HEALTH ORGANIZATION (2017): Flooding and communicable diseases fact sheet – Risk assessment and preventive measures. Online: http://www.who.int/hac/techguidance/ems/-flood_cds/en/, 13.02.2017

[WIK-17] WIKIPEDIA (2017): History of Wikipedia. Online: https://en.wikipedia.org/wiki/-History_of_Wikipedia, 16.02.2017

[WOM-17] WORLDOMETERS (2017): Current World Population. Online: http://www.worldometers.info/world-population, 08.01.2017

[WWF-17] WORLD WILDLIFE FUND FOR NATURE (2017): Deforestation – Overview. Online: http://www.worldwildlife.org/threats/deforestation, 14.01.2017

[YTB-13] YOUTUBE CHANNEL OF UCL MATHEMATICAL AND PHYSICAL SCIENCES (2013): Professor David Alexander – mini inaugural lecture. Online: https://www.youtube.com/watch?v=f7wLS9IOS4s, 27.01.2017

7.3. Data Sources

Federal Ministry for Transport, Innovation and Technology (BMVIT) – Christoph Hackel

~ Flood discharge areas from the 2013 flood event for the Danube River in Austria.
~ Predicted average water depths in case of a 30-, 100- or 300-year flood event.
~ Information about the different types and frequencies of damages along the Danube River, including the locations of sediment shifts alongside the river banks (erosion and accretion).
~ Electronic Navigational Chart (ENC) for the Danube River in Austria 2010 with encoding guide.
~ And many datasets more, which were not used in this master thesis like the Danube Atlas for flood hazard and risk maps 2012, orthophotos, maps from European waterways, statistical information about the floods in 1954, general information about the 2002 flood event and statistical data about water levels from the Austrian Danube in 2010.

International Institute for Applied Systems Analysis (IIASA) – Ian McCallum and Christoph Perger:

~ The statistical output from the 2015 FotoQuest Austria campaign in the form of an Excel spreadsheet, consisting of 19,972 data entries (rows) and 25 data categories (columns).
~ General information about the Central European floods in 2013.
~ Lots of inputs and advices during regular email-exchanges.

European Commission – Eurostat:

~ Land Use and Cover Area Sample (LUCAS) survey, retrieved from http://ec.europa.eu/eurostat/ statistics-explained/index.php/LUCAS_-_Land_use _and_land_cover_ survey (17.04.2017)

Provincial Government of Lower Austria (NOE) – Erik Formann:

~ Predicted flood extent in case of a 30-, 100- or 300-year flood event (HQ30, HQ100, HQ300)

Other Sources:

~ Digital Elevation Model (DEM), retrieved from https://earthexplorer.usgs.gov (21.05.2017)
~ Landsat 8 satellite data via the SCP plugin, retrieved from https://earthexplorer.usgs.gov (no date)
~ Building footprints and roads, retrieved from http://www.openstreetmap.org (11.07.2017)

Curriculum Vitae

- Gelöscht für die anonymisierte Plagiatsprüfung

Affirmation

Ich versichere:

- dass ich die Masterarbeit selbstständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfe bedient habe.
- dass alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Publikationen entnommen sind, als solche kenntlich gemacht sind.
- dass ich dieses Masterarbeitsthema bisher weder im In- noch im Ausland (einer Beurteilerin/ einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe.
- dass diese Arbeit mit der vom Begutachter beurteilten Arbeit übereinstimmt.

Datum Unterschrift