NOVA IMS
Information Management School

MEGI

Mestrado em Estatística e Gestão de Informação
Master Program in Statistics and Information Management

Exploring vulnerability and impact of floods in Malawi
A first step towards impact-based forecasting

Maud Broeken

Dissertation presented as partial requirement for obtaining the Master’s degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

2015

Title: Exploring vulnerability and impact of floods in Malawi

Subtitle: A first step towards impact-based forecasting

Student

full name: Maud Broeken

MEGI

2015

Title: Exploring vulnerability and impact of floods in Malawi

Subtitle: A first step towards impact-based forecasting

Student

full name: Maud Broeken

MGI

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

EXPLORING VULNERABILITY AND IMPACT OF FLOODS IN MALAWI

A FIRST STEP TOWARDS IMPACT-BASED FORECASTING

by

Maud Broeken

Dissertation presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Management and Analysis.

Advisor 1: M. J. C. van den Homberg

Advisor 2*:* Ana Cristina Costa

June 2018

ACKNOWLEDGEMENTS

I would like to take the opportunity to thank some people that had a great influence on me while writing my master thesis. First, I would like to thank my two supervisors that played a big role during the whole process of my thesis. Marc has helped me from the beginning to the end and shared a lot of his knowledge and networks with me. I'm very grateful for all the effort he took to help me with my thesis, the valuable feedback he provided me and the opportunity he gave me to be present on a congress in Vienna. Cristina Costa, our meeting in Lisbon was great for moving forward with the thesis. You really helped me with the methodology and data analysis and you always gave a fast response to clarify my questions. Thank you very much! I learned a lot from both of you!

Furthermore, thank you Maarten and Stefania and the rest of the 510 team for giving me much opportunities, new ideas, knowledge and made me suffer during the planking sessions. In particular, I want to thank Aki that helped me to go in the rigth direction at moments that I was lost.

Finally, I am very happy with the continuous support of my family and friends!

NOVA IMS Information Management School

510 AN INITIATIVE OF THE NETHERLANDS RED CROSS

ABSTRACT

Forecast-based financing is a methodology to trigger early actions when a forecast exceeds a danger level in a vulnerable intervention area. The first three implementation steps aim at impact-based forecasting: (1) Understanding risk scenarios, (2) Identifying available hazard forecasts and (3) Identifying danger levels. Impact-based forecasting requires timely, complete, reliable and accurate data at a subnational level, which is however – especially in developing countries with a high data poverty – challenging. 510, The Netherlands Red Cross data team, has developed a Community Risk Assessment dashboard, that visualizes data on the INFORM risk index with three dimensions: Hazard & Exposure, Vulnerability and Lack of Coping Capacity. However, the number of available indicators decreases sharply when one goes from national down to district or even to the community level. The aim of this study was to downscale the vulnerability index to subnational level and to examine the relation between vulnerability and impact in Malawi. A literature review was conducted to understand the existing frameworks of vulnerability and the concept of impact. Thereafter, global and national open data sources were accessed to collate data of vulnerability and impact on subnational level. To determine the vulnerability level on subnational level, the gap in vulnerability data is characterized both vertically, in terms of data missing at lower administrative levels on indicators already used, as well as horizontally by adding new indicators. Thereafter, factor analysis was performed to reduce dimensionality of the dataset (in which there are a large number of uncorrelated variables) and to determine the vulnerability level. Reducing the dimensionality of the dataset makes it easier to visualize and understand the differences in vulnerability level across areas and to examine the relation with impact of floods. Five factors were identified and subsequently the five factors and the total vulnerability were successfully mapped to visualize the vulnerability level on Traditional Authority level in Malawi. The mapping revealed large differences between TAs and made it clear that data on a subnational level is essential in order to have a proper understanding of the reality on the ground. Finally, relations between these factors and impact data were examined. Impact data consisted of Internally Displaced Persons, Food Deficit and People Affected after being exposed to a flood. In conclusion, three relations were found between the vulnerability factors and impact which is a first essential step towards Impact-based Forecasting.

KEYWORDS

Forecast-based financing, Vulnerability, Floods, impact-based forecasting, Malawi

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

CRA	Community Risk Assessment
DHS	Demographic and Health Survey
DRR	Disaster Risk Reduction
FbF	Forecast-based Financing
FA	Factor analysis
GVH	Group Village Headman
HH	Household
IbF	Impact-based Forecasting
IFRC	International Federation of the Red Cross and Red Crescent
IFPRI	International Food Policy Research Institute
INFORM	Index For Risk Management
IOM UN	International Organization for Migration United States
GFDDR	Global Facility on Disaster Reduction and Recovery
IDP	Internally Displaced People
IPCC	Intergovernmental Panel on Climate Change
MOVE	Methods for the Improvement of Vulnerability Assessment in Europe
NSO	National Statistics Office
PCA	Principal Component Analysis
RCMRD	Regional Centre for Mapping of Resources for Development
TA	Traditional Authority
UNISDR	United Nations Office for Disaster Risk Reduction
USAID	U.S. Agency for International Development
VIF	Variance Inflation Factor

INTRODUCTION

1.1 Context

Future global climate change challenges humanitarian organizations even more to reduce risk from natural disasters (Birkmann et al., 2013). An effective method is Forecast-based financing (FbF) where funds are made available prior to the hazard to undertake action before the disaster causes enormous damage in the exposed areas (de Perez et al., 2016). With the movement from post-disaster to pre-disaster strategies of humanitarian aid organizations, millions of lives can be saved and a considerable part of exposed houses can be prevented from significant damage (de Perez et al., 2016; Galindo & Batta, 2013; Harriman, 2014; Rogers & Tsirkunov, 2013).

New technology- and information systems improve the accuracy and reliability of weather forecasts, which causes many opportunities for humanitarian organizations (de Perez et al., 2016). After the forecast of a hazard there is precious time that gives the opportunity to respond in advance, such as transporting relief items before villages are unreachable due to a flood (de Perez et al., 2015). One of the fundamental principles of the Red Cross states that people should be provided with aid according to their needs, where the most urgent cases of distress have first priority (IFRC, n.d.). Currently, identification of the most vulnerable and affected communities is a challenge in most developing countries, given the data poverty. This gap causes humanitarian responders to face difficulties in their decision-making process. Enabling identification of these vulnerable areas will make it easier to take rapid decisions for early actions (Suarez & Tall, 2010; Wilkinson et al., 2018).

Furthermore, weather forecasts are highly relevant for humanitarian organizations however information about the impact that the natural disaster will cause is still lacking. This information is useful in the decision-making process before taking actions in exposed areas. The methodology of predicting the consequences, such as loss and damage, that a natural disaster would cause is called impact-based forecasting. Impact-based forecasting demonstrates that impact is related with exposure to a hazard, vulnerability and coping capacity. The objective is to provide organizations with detailed information in order to make more goal-oriented decisions and to improve the effectivity of the provided aid.

1.2 Problem definition

The Netherlands Red Cross data team has developed a Community Risk Assessment (CRA) dashboard (see: https://dashboard.510.global/#!/) for 14 countries that visualizes data on the INFORM (Index For Risk Management) risk index with three dimensions Hazard & Exposure, Vulnerability and Lack of Coping Capacity. However, data on the indicators determining vulnerability decrease sharply when downscaling to subnational level. The goal of relief workers is to provide aid to the most vulnerable people first however the vulnerability level on subnational level is not yet determined. The lack of having a vulnerability map on a more detailed level causes an obstruction for relief workers to provide aid in the most vulnerable places first. Therefore a vulnerability map is needed.

Moreover, predicting the impact of a natural hazard is of major concern to manage pre-determined actions related with the expected damage and loss. It is of major importance to understand the impact of floods on subnational level in order to implement impact-based forecasting. However, complete and reliable datasets on subnational level are required which is a challenge in developing countries.

1.3 Scope

Scope within the Red Cross:

This study is conducted together with 510. 510 is an initiative of the Netherlands Red Cross and aims to improve humanitarian aid by the smart use of (big) data. The team of 510 runs projects where data is analyzed and converted into understandable knowledge to increase the efficiency of humanitarian aid in disaster prone countries. 510 is NLRCs fast-growing data initiative with over 40 volunteers, MSc students and core staff. The 510 team brings together a variety of skills in data science, geo-graphical information management and applications, data visualization, responsible use of data and disaster risk management. 510 is a member of the Missing Maps Project and partner of the INFORM risk index consortium (led by EU JRC and UN OCHA), in relation to which 510 has currently an assignment with UNDP to develop the INFORM Sub-national dashboard.

Scope geographically:

510 is active in Malawi since the 2015 floods. NLRC and 510 digitally supported Vulnerability and Capacity Assessments (VCAs) and a pilot for flood FbF in early 2018. 510 is implementing the Missing Maps project in Malawi since 2016, coordinating OpenStreetMapping efforts with the World Bank, USAID, MSF, National Statistics Office of Malawi Government and others. 510 is also part of Data4SDGs, funded by the UN and World Bank (Global Partnership for Sustainable Development Data). This is done in partnership with the National Statistics Office. Data collected during ongoing ECHO supported Actions in Malawi ranges from government data at different administrative levels, open data from online data platforms, statistics offices, NGOs and private sector on the INFORM risk index dimensions, i.e. Hazard& Exposure, Lack of Coping Capacity and Vulnerability. This data is visualized through 510's Community Risk Assessment dashboard.

Scope hazard type:

When this study started, the Malawian, Belgian, Netherlands and Danish Red Cross Societies were planning a pilot Forecast-based financing early 2018 in two districts in the Lower Shire River basin in Malawi for floods. Malawi is one of the poorest countries in Africa that is regularly exposed to floods. In 2015 it suffered from the biggest floods in years where 638.000 people were affected according to the International Disaster Database EM-DAT (www.emdat.be). Therefore, this study was conducted to determine the vulnerability level and find a relation with impact of floods in Malawi.

1.4 Research objectives

The objective of this study is to improve the current Community Risk Assessment tool into a vulnerability composite index on subnational level. Open data sources via geospatial data sharing platforms, governmental information systems that are publicly accessible are approached to collect vulnerability data. Furthermore, the objective is to analyze the impact of historical floods in Malawi. Finally, the relation between vulnerability and impact of floods is examined which is a first essential step toward impact-based forecasting.

1.5 Research questions

The following research question is formulated to reach the objectives of this study: *How can vulnerability be used to predict the impact of floods?*

To answer this research question, the following research sub-questions are formulated:

1. *Which frameworks and associated indicators are currently used to characterize vulnerability?*
2. *Which data sources are available on the vulnerability in Malawi on subnational level?*
3. *Which exploratory (unobserved) factors can be extracted from the (observed) vulnerability indicators?*
4. *Which data sources are available on the impact of historical floods in Malawi on subnational level?*
5. *To what extent is vulnerability related with impact of a flood in Malawi?*

1.6 Societal and scientific relevance

This study is a first essential step towards impact-based forecasting with the objective to improve the efficiency of humanitarian aid in vulnerable areas. Forecasting the impact of floods in Malawi enables the Red Cross society to provide more tailored help. In particular, efficient decisions can be taken within the (usually short) lead time; it saves lives and reduces costs (de Perez et al., 2016). Moreover, this study enables to underpin the framework of impact-based forecasting by scientific results.

1.7 Structure of the dissertation

Chapter 2 is dedicated to the literature review, which includes the topics of Forecast-based Financing and Impact-based Forecasting. This chapter reviews issues related to vulnerability and impact and the relation between them.

Chapter 3 describes the methodology that is used to examine the research questions. Section 3.2 describes the characteristics of the study area more in detail. Thereafter, each step for examining the research questions is explained starting with data collation and data preparedness. Section 3.4 explains the method of Factor analysis and eventually the method to relate vulnerability with impact is discussed.

Chapter 4 shows the results of the data analysis. The data collation and preparation is presented first. Thereafter, descriptive statistics are displayed to gather a general impression of the data. Deeper examinations on the data are present in section 4.3. Finally, maps of vulnerability are displayed and the relation with impact is examined.

Chapter 5 and 6 bring respectively a discussion and conclusion and include limitations and future recommendations of this study. Thereafter the references will follow together with appendix and extra annexes.

LITERATURE REVIEW

2.1 Forecast-based Financing

Forecast-based financing is aimed at allocating financial resources prior to a natural disaster in order to support disaster preparedness (Suarez & Mendler de Suarez, n.d.). Currently, hazards can be forecasted, and humanitarian organizations are informed about the location and severity of the extreme weather event. The objective of FbF is to strengthen the preparedness capacities of humanitarian organizations by making funds available after the pre-determined threshold of a hazard forecast is exceeded. Therefore, this methodology triggers early actions when a forecast exceeds a danger level in a vulnerable intervention area. These actions are pre-defined for each event in the country in order to reduce disaster risk and increase efficiency of humanitarian aid.

In particular, countries that are prone to natural disaster highly rely on climate and weather information because of the changing climate and its devastating consequences (de Perez et al., 2016). With innovative information technologies, forecasters improve models to enhance accuracy and reliability which is subsequently relevant for organizations to take decisions on early actions (Rudari, Beckers, De Angeli, Rossi, & Trasforini, 2016). In particular, forecasters share knowledge in terms of weather conditions, such as when and where the disaster is likely to happen, which supports humanitarian organizations in their preparations towards the potential hazard (Suarez & Mendler de Suarez, n.d.). However, the risk of acting in vain (taking action prior to an extreme event while the action is not followed by an actual disaster) is always present which prevents stakeholders from financial investments (de Perez et al., 2015, 2016; Wilkinson et al., 2018). Although, studies prove that taking action prior to the disasters increase the cost-efficiency even with the potential to act in vain (Mechler, 2005). Therefore, the improvement of forecasting models and its corresponding response is important for the success of FbF.

In the past decade, early warning systems were effective in disaster-prone areas. The information that livelihoods received in advance is highly relevant for the management in Disaster Risk Reduction (DRR). However, the full potential of early warning systems was not realized since actions become valuable when they are directly followed after the warning. Implications of previous pilot studies show that there is a high need of actions tailored to the forecast of the disaster (de Perez et al., 2015). To determine which actions have significant effect to alleviate human suffering and limit the amount of damage, the impact should be forecasted prior to the actual hazard. With the knowledge of what the impact will be, decision makers can make precise decisions when taking early actions. This process is defined as impact-based forecasting where the consequences of the forecast are emphasized.

2.2 Impact-based Forecasting

Impact-based forecasting translates the meteorological and hydrological forecast into potential impact on people and their livelihoods (GFDRR, 2016). The forecast of the impact is specified for each location where different impact sectors are examined. This enables organizations to take more situation- and location specific actions in areas that are exposed to the extreme event.

Figure 2.1 depicts the steps that are necessary before implementing impact-based forecasting. The risk analysis is important to identify the areas that are (1) prone to natural hazards, (2) susceptible when exposed to natural disasters and (3) lack the capacity to cope during and after the hazard. These areas have the priority for the Red Cross to provide aid since the suffering is expected to be the highest.

Thereafter, the disaster forecasts analyze when a threshold is exceeded and if early actions should be executed because natural hazards do not always turn into a disaster. Namely, a disaster disturbs the normal functioning of a social system due to natural or man-made changes (Birkmann et al., 2013; Quarantelli, 1998; United Nations General Assembly, 2016). Moreover, a disaster causes high levels of damage and loss in a society due to significant physical environmental changes. Furthermore, exposure plays a central role in this framework since it is the extent to which humans and other systems (e.g. social or economic) are located in the area that is prone to the hazard. With other words, if people are not exposed to a disaster, the direct impact of that disaster will be nihil. This study focuses only on one type of natural hazards, namely floods.

The third step implies the development of a composite index which is based on historical impact information where expert knowledge is indispensable. In addition, historical impact data can be related with the intervention map with different modelling methods in order to reach the final step; impact-based forecasting. When the exposure, vulnerability and forecast map are developed they can be merged together into one intervention map. With this intervention map, organizations can undertake more targeted actions and anticipate to the situation which makes early actions more effective (Wilkinson et al., 2018).

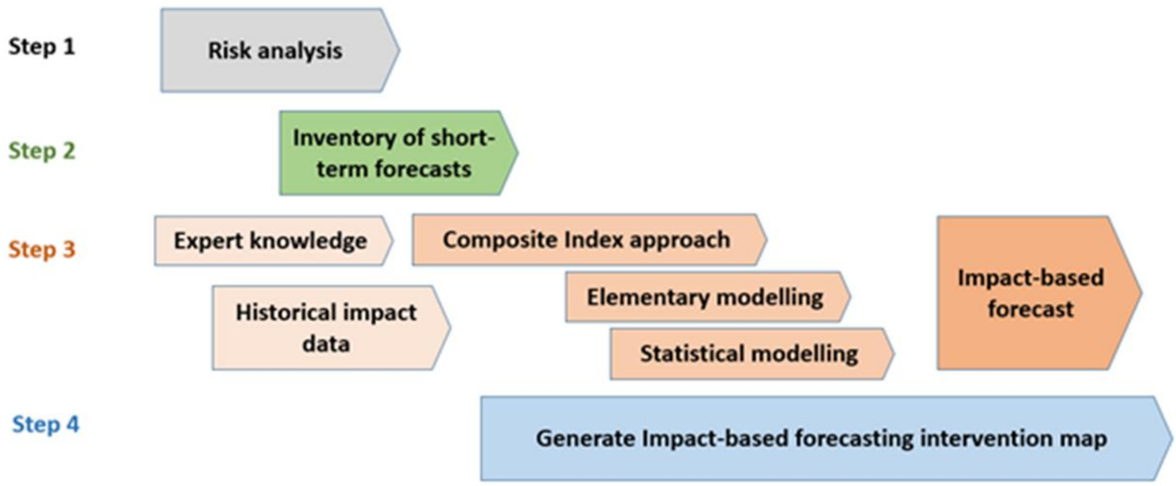

Figure 2.1 – Model for Impact-based Forecasting (van den Homberg & Visser, 2018)

2.3 Understanding Vulnerability

Vulnerability is examined in different fields and has therefore multiple explanations, concepts and methods in the scientific world (e.g. Birkmann et al., 2013; IPCC, 2007; Schneiderbauer & Ehrlich, 2004; Wisner, Blackie, Cannon, & Davis, 2004). The Intergovernmental Panel on Climate Change (IPCC, 2007) wrote a report on climate change and the understanding of impacts, adaptation and vulnerability. In their research, IPCC define vulnerability as "*the degree to which a system is susceptible to, and unable to cope with, adverse effects of climate change, including climate variability and extremes. Vulnerability is a function of the character, magnitude, and rate of climate change and variation to which a system is exposed, its sensitivity, and its adaptive Capacity*. In the field of Disaster Risk Reduction (DRR), the (UNISDR, 2009) define vulnerability as: *The conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards.* This definition is very applicable to this study since it is considered as a resource and circumstance that influences effective risk management which is in line with the objective of this study. Moreover, Pelling (1997) explained vulnerability with regard to floods as the manner in which households can deploy resources in case a flood event occurs to decrease the likelihood of a living space being flooded and reduce the negative impact. Therefore, in this study vulnerability is explained as the resources and assets households can deploy to strengthen the resilience and response to a flood, in order to reduce the negative consequences when being exposed to the extreme event. It includes several factors (such as economic, political and social) that can be destabilized during an extreme event. The following features should be taken into account in this definition: (1) Vulnerability is considered to vary across physical space and among and within social groups, (2) Vulnerability is scale dependent since different vulnerability factors (i.e. units of analysis) are determined to measure this latent variable (Birkmann, 2006; Vogel & O'Brien, 2004).

The explanation above indicates that there are different factors of vulnerability which, however, are used differently over time. The factors that are predominantly mentioned in other studies, and are relevant for this study, are explained here. First, the social factor in vulnerability plays a central role when determining components of vulnerability (Cannon, Twigg, & Rowell, 2003; Cutter, Boruff, & Shirley, 2003). Social vulnerability is the extent to which people can protect themselves and how the social system is affected after a hazard. This factor covers a broad dimension, such as mental damage, educational disruption and discrimination. Social vulnerability is examined by estimating individual and demographic characteristics (Cutter et al., 2003). Second, economic vulnerability is a frequently examined factor and is usually the most easy dimension to measure (Van der Geest & Schindler, 2017). Economic vulnerability determines the damage to tangible assets that can be replaced after a flood (Birkmann, 2006; Birkmann et al., 2013). An example of an economic indicator is the capacity to build stable houses and protection towards danger (Cannon et al., 2003). The INFORM index merges the above mentioned factors into one dimension and calls it a socio-economic vulnerability category where the physical conditions of a livelihood are measured along with well-being (INFORM, 2016). Another dimension of vulnerability where the attention raised since the adverse impacts of global climate change is biophysical vulnerability. Biophysical vulnerability is the concept that focusses on systems that are vulnerable to the environmental change due to climate movements and to what extent they can adapt to the impacts of this change (Birkmann, 2006; WBGU, 2005); Lastly, physical vulnerability refers to the probability that physical assets will be damaged (e.g. water sources, hospitals, or infrastructure) that disable households to utilize or supplement resources (Birkmann et al., 2013).

2.4 IDENTIFYING VULNERABILITY INDICATORS FOR ANALYZING FACTORS OF THE INDEX

Hence, vulnerability is a variable which consists of different factors and is measured with a relative value. 510 calculates vulnerability in the Community Risk Assessment Tool for each administrative level. At the admin level just below the national level there are usually -dependent on the specific country and the amount of time that has been invested already – between five and ten indicators for which data is available; at the lowest admin level this goes down to usually only one or two. Poverty data is usually available at this lowest level; however poverty on its own is insufficient to represent the whole concept of vulnerability. Therefore, variables should be added to the index to calculate a more reliable and valid vulnerability level in Malawi and to cover all the factors. Justification for a composite indicator does not have a universally accepted scientific guideline but instead mathematical models are used to compute whether the indicator adds value to the index (Rosen, 1991; Sharpe, 2004).

In order to tailor the identification of variables to the measurement of vulnerability, indicators were identified. Literature focusing on natural disasters or developing countries were accessed. Even though this study is focused on floods, the vulnerability parameters should be hazard independent since the hazard component is independently incorporated in the calculation of risk and impact. Therefore, the objective is to identify holistic vulnerability factors that are applicable for multiple disasters. From this literature research, three relevant and applicable studies were selected to identify vulnerability indicators.

The first methodology that is highly relevant for this research is the INFORM index. The INFORM index is a collaboration of the Inter-Agency Standing Committee Task Team for Preparedness and Resilience and the European Commission (www.inform-index.org). The INFORM index gives a risk index between 1-10 and which is computed with 3 dimensions; Hazard & Exposure, Vulnerability and Lack of Coping Capacity. Each dimension contains categories and its related components. The INFORM index is displayed in Table 9.1 in the Annex. This model is useful for calculating the risk score for each country on a national level. Even though the scores give a clear overview between countries, vulnerability patterns within countries are still lacking. Therefore, the methodology is valuable for this study to calculate vulnerability scores on TA level but a more extensive research among vulnerability indicators is required.

The second research that is accessed is the study conducted by Birkmann et al. (2013). They designed a framework to assess vulnerability, risk and adaptation. The MOVE framework (Methods for the Improvement of Vulnerability Assessment in Europe) is applicable for this study since they examine how vulnerability should be assessed in the context of natural hazards and climate change. Their framework is in line with the INFORM index, where Exposure, Vulnerability and Lack of Resilience (in INFORM index called Lack of Coping Capacity) are components to calculate risk. However, the vulnerability key factors are differently organized. Table 9.2 in the Annex shows their approach.

Wannewitz, Hagenlocher and Garschagen (2016) developed a multi-hazard risk index for the Philippines (Wannewitz et al., 2016). This study was conducted because a risk index on municipal level was still lacking in this disaster-prone country. The model diverges from the latter two studies since the composite index consists of two indicators; Exposure and Vulnerability. Coping and adaptive capacities are comprised in the vulnerability indicator. This study demonstrates how to downscale a risk index and is therefore relevant for research. However, their approach is based on the World Risk Index (WRI) whereas this study follows the methodology of INFORM. In Table 9.3 in the Annex, vulnerability indicators used for the study of Wannewitz et al. (2016) are depicted.

2.5 IMPACT

Impact are the effects of natural disasters that cause significant changes in the normal functioning of livelihoods and natural systems (IPCC, 2012). Although, impact can be avoided by human actors, sometimes impact is unavoidable since there will always be damage and/or loss. Loss and damage are the severe effects of the hazardous event, despite mitigation and adaptation (Van der Geest & Schindler, 2017). Impact can be examined by several perspectives, for instance material versus non-material losses, economic versus non-economic losses, human and environmental loss and damage (Morrissey & Oliver-Smith, 2013; UNISDR, 2009). Floods have a major impact on peoples live in terms of damaged houses, people injured due to the flood, loss of crops and even the loss of lives (UNISDR, 2009). In the longer term, other consequences usually arise, for instance susceptibility to Malaria and outbreaks of diseases due to the lack of hygiene.

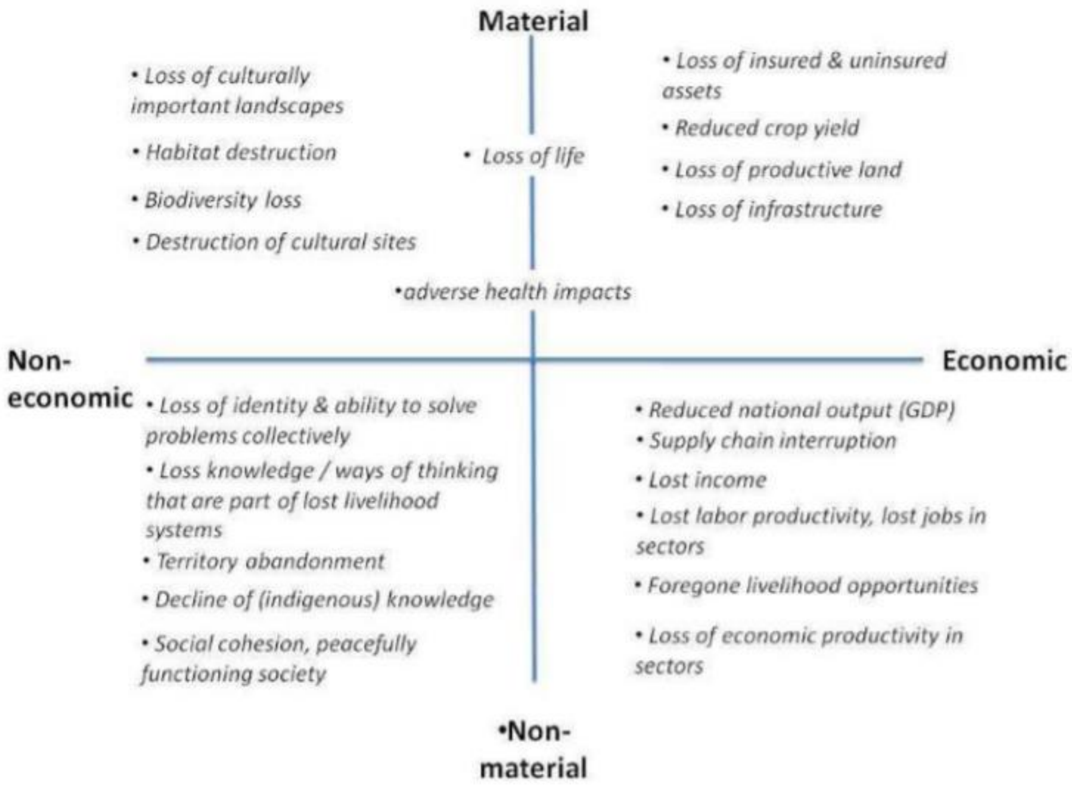

Figure 2.2 - Economic and non-economic losses (Morrissey & Oliver-Smith, 2013)

As indicated above, various dimensions should be considered when examining impact. The measurement of impact is differs between countries and within situations (Van der Geest & Schindler, 2017). For example, the impact of a flood is different than the impact of a drought. Identifying the scale, severity and timing of different impacts is important when responding to the forecast of a hazard (Wilkinson et al., 2018). Furthermore, Wilkinson et al. (2018) state that performing analysis on loss and damage in developed countries contains difficulties due to the limited amount of available data. Usually developed countries lack the documentation of relevant information to determine the loss and damage beforehand, such as amount of houses in a village, labor hours or amount of land the people own. As mentioned above, the assessment tool should be adapted for each situation and area therefore mixed tools can be used to assess damage and loss after a natural hazard. The study about the natural hazard and the household survey are central tools to understand the context of the field. Furthermore, interviews and discussion groups are other measurement tools next to surveys (Morrissey & Oliver-Smith, 2013; Van der Geest & Schindler, 2017). These tools provide the assessor with indigenous knowledge which yields relevant information when identifying impact in a country with a scarce amount of data (Mercer, Kelman, Suchet-Pearson, & Lloyd, 2009). Mercer et al. (2009) state that interaction between indigenous and scientific knowledge is important to eventually reduce community vulnerability exposed to hazards. With their knowledge, a better understanding in the study area will be enlightened.

The Risk Matrix (Figure 2.3) is a useful tool to examine impact (Van der Geest & Schindler, 2017). The Matrix relates the expected impact of a natural disaster to the likelihood of the occurrence of the disaster in that area (Neal, Boyle, Grahame, Mylne, & Sharpe, 2014). When the likelihood and the severity of a flood is high (the red area) people should take actions to prevent themselves of severe impact, for instance by avoiding certain places or situations (Van der Geest & Schindler, 2017). On the other hand, when the likelihood and the impact are expected to be low (the green area), people should be aware of the (low) risk however they do not need to take adverse actions (Van der Geest & Schindler, 2017).

Figure 2.3 – Risk Matrix (Van der Geest & Schindler, 2017)

2.6 Vulnerability and the relation with impact

Actions prior to the hazard are effective when loss and damage is prevented from the situation (Wilkinson et al., 2018). Therefore impact should be forecasted by multiplying vulnerability, coping capacity and hazard exposure (INFORM, 2016). These three explanatory variables determine the risk level which will eventually give a forecast of the impact. The framework in Figure 2.4 shows that risk is climate and socio-economic state dependent and therefore differs among countries and hazards (IPCC, 2014). The framework depicts the three components (i.e. Vulnerability, Exposure and Hazards) determining the risk which is related with impact. The right column shows that the socio-economic processes, such as adaptation and mitigation towards a risk, have a direct influence on the risk.

Moreover, different factors of vulnerability are related to different types of impacts and responses (Turner et al., 2003). The human and environmental conditions together with the coping mechanisms influence the extent to which an event has an impact in the exposed area. In particular, the conditions effect how people can respond to a flood and how their assets endure the flood which subsequently determines the impact (Turner et al., 2003). This is in line with Bogardi and Birkmann (2004) who relates vulnerability with impact using the Union Framework. The framework relates vulnerability with different impacts (in this case economic and social losses). They state that economic losses are solved easier, however, when going beyond the tangible loss the flood event causes negatively interferes with the functioning of the social system (Birkmann, 2006).

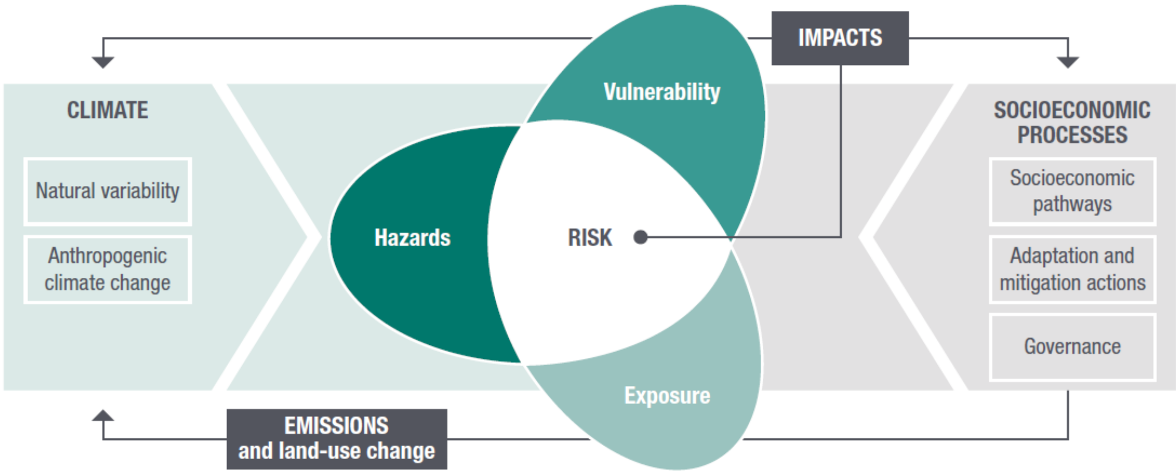

Figure 2.4 – The Interaction of climate-related hazards, vulnerability and exposure (IPCC, 2014)

3. METHODOLOGY

3.1 RESEARCH DESIGN

The objective of this study was to examine the vulnerability level and impact after floods. Furthermore, the goal was to examine the relation between vulnerability and impact. Therefore, the following research question was formulated: *How can vulnerability be used to predict the impact of floods?*

The sub-research questions were examined first to reach the initial objective of this study. Figure 3.1 summarizes the research process. First, an extensive literature research was conducted to understand the concept of vulnerability in the context of DRR. Indicators that measure vulnerability were identified by examining different frameworks. In this study, an indicator is a small aspect that measures vulnerability where one indicator can be measured with different variables. Thereafter, data sources were accessed to collate data for the vulnerability composite index on subnational level in Malawi. After merging the different datasets a data matrix was created that presented the vulnerability variables on Traditional Authority (TA) level. Subsequently, the data matrix was prepared for further analyses.

A Factor analysis was performed to reduce the dimensionality of the dataset. This reduction was achieved by transforming variables into factors where each factor represents a number of interrelated variables. The vulnerability level was calculated for each TA and different maps were produced to visualize the factor levels for each TA. Thereafter, open sources were approached to collect impact data of historical floods in Malawi. Data of historical floods are useful for estimating impact in different areas (Lindell & Prater, 2003). To this end, different databases as well as indigenous and expert knowledge about previous floods had to be used to examine the impact of floods in different livelihoods of a country. However, damage and needs assessment tools are usually not consistently defined over events and within organizations. For example, the IFRC (International Federation of the Red Cross and Red Crescent) reports the amount of people missing, injured, killed and affected in standard so-called Disaster Relief Emergency Fund (DREF) reports. However, this report is filled out by different people and is not specified by country which causes different documentations among the historical flood events. But there are several initiatives to reach more harmonization such as Desinventar and the World Bank's DaLa (Damage, Loss, and Needs Assessment) and PDNA (Post Disaster Needs Assessments) approach (GFDRR, 2010, 2017). Furthermore, databases such as EM-DAT or databases from insurance companies such as - http://natcatservice.munichre.com are usually at a higher aggregate level and often the poor and vulnerable are not included in the reported damage and losses as in economic terms their assets are negligible and uninsured.

Finally, the relation between the vulnerability factors and impact was examined with an Ordered Logit model.

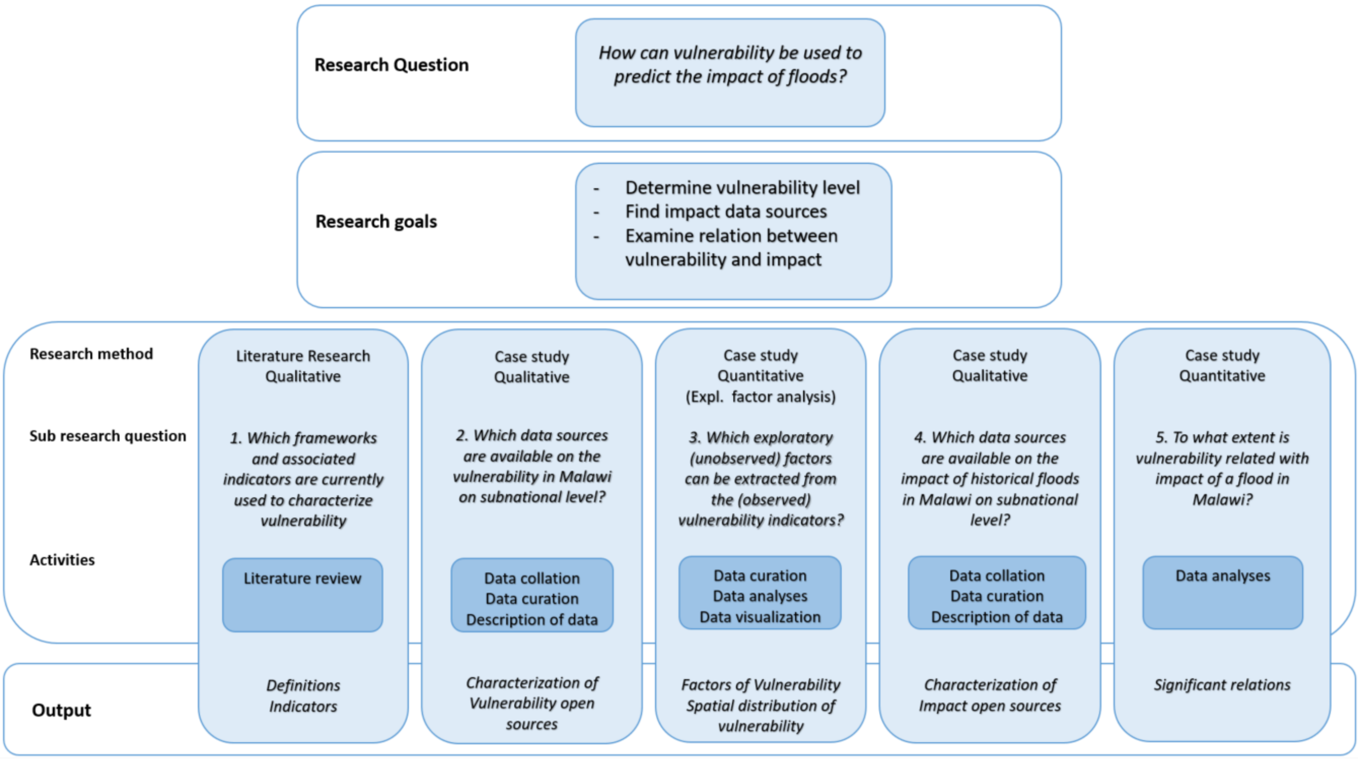

Figure 3.1 – Research Model

3.2 Case Study

Malawi is a republic located in the east of Africa and bordered by Zambia, Tanzania and Mozambique. In total, Malawi has a surface area of 118.000 km^2 where approximately one third of the surface consists of Lake Malawi (see Figure 3.2). Malawi is regionally divided into three regions (Administrative level 1) which consists of the North, Central and South Malawi. Those regions are divided into 28 districts which are considered as Administrative 2 level. Thereafter, there is an Administrative 3 level which consists of 350 smaller areas, called Traditional Authorities (TAs). The TAs consists of several GVHs, Group Village Headman, which is called the Administrative 4 level. The aim of this study is to determine the vulnerability on Administrative 3 level, which is in Traditional Authorities. In Table 3.1 the country characteristics are displayed.

Table 3.1 – Country Characteristics of Malawi

Characteristic		**Source**
Population rate	18.09 Million	The World Bank, 2016
Population living in urban	54.3% of total	The World Bank, 2016
Fertility rate	5.3	UNstats, 2010-2015
Life expectancy at birth	62.5 years	The World Bank, 2010
Under-five mortality	85.329 per 1000 live births	UNstats, 2010-2015
Maternal mortality	634 per 100.000 live births	UNstats, 2015
GDP	5.433 Billion dollar	The World Bank, 2016
Poverty headcount ratio at national poverty lines	50.7% of population	The World Bank, 2010
Illiterate population (>15 years)	3.548.455	UNESCO, 2015**
HIV	1.000.000	UNAIDS, 2016

Floods and droughts are the main natural hazards causing considerable problems in Malawi. Since Malawi is highly dependent on agriculture, these extreme climate events adversely affect the inhabitants. The losses that the country endures is significant; on average, Malawi loses 1.7 percent of its gross domestic product due to the effects of floods and droughts (Pauw, Thurlow, & van Seventer, 2010).

Since 1970, flooding has occurred once in every 2 to 5 years and they have affected more millions of people (Atkins, 2011). Malawi suffers from two different types of flood. On the one hand it experiences flash floods that arise due to excessive rainfall in a short time period. On the other hand, Malawi suffers from riverine floods which happens when the river exceeds its capacity due to heavy rainfall over an extended period of time. In particular, the districts in the southern part of Malawi are the prone to floods (see Figure 3.2) because of the flowing water coming from the hillsides in the Lower Shire River basin that are unable to absorb the water. The consequences are enormous, not only farmers suffer from crop loss, the food prices increase significantly making it impossible for lower income households to feed their family (Pauw et al., 2010).

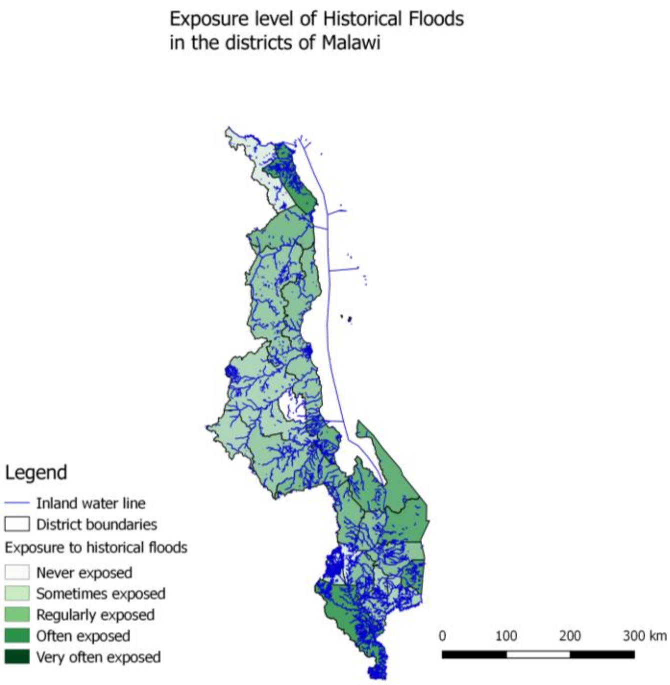

Figure 3.2 – Historical flood exposure in Malawi

Flood 2015 in Malawi

176 People were killed and more than 200.000 were people homeless (EM-DAT, 2017).

Figure 3.3 – Picture of the impact of the flood in 2015.

3.3 Data collation

Indicators for the vulnerability composite index were identified after the literature study. Thereafter, data sources were accessed that contained data on these indicators in Malawi. In this study, the term collation refers to exploring for open source data and data collection refers to gathering primary data through filling out questionnaires, for instance in a case study in Malawi. Data collation will occur by approaching global and national providers of socio-economic and environmental data. Data collation plays a major role in this study, among others because of the data-scarcity in Malawi. To collate vital data, for some sources access is requested from open data providers that contain sensitive information about households in Malawi. Figure 3.4 depicts the framework for data collation.

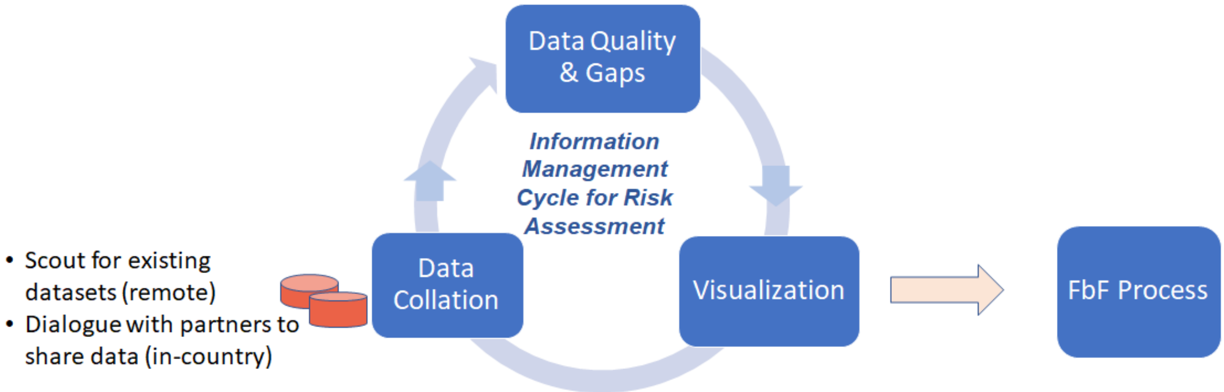

Figure 3.4 – Framework for data curation in Malawi (Van den Homberg & Plaatsman, 2018)

3.4 Data Preparedness

Data was collated from different open sources and extracted to Rstudio for further calculations (Rstudio Team, 2016). The problem of missing values arose when the datasets were merged together into one data matrix. The data matrix cannot contain missing values when conducting a Factor analysis therefore the Amelia II method was applied to create a complete dataset as described below. Furthermore, high correlations between variables cause extreme large variances of regression coefficients which would lead to incorrect estimates of the regression equation (Jolliffe, 2002). This is called multicollinearity and is avoided in the dataset as detailed below. Nevertheless, it is important to note that there are methods to rotate the factors after extraction (e.g., *Varimax*) that guarantee that the factors are orthogonal (i.e., uncorrelated), which eliminates problems of multicollinearity in regression analysis. Variables causing multicollinearity were removed from the dataset so that, in the future, it will be more efficient to only collect the data of the most relevant variables to replicate this study.

3.3.1 Missing values

An option to deal with missing values is to delete the TAs that show many missing values in the dataset. However, this limits the amount of variables that can be added to the data matrix since the ratio of TAs (individuals) to variables (columns) should be at least 5:1 (Bryant & Yarnold, 1995; Nunally, 1978). Therefore, Amelia II was applied to fill in the missing values to continue the analysis with a complete dataset (Honaker, King, & Blackwell, 2011; Schafer & Olsen, 2010). Amelia II is a method to fill in missing values which is less biased compared with other methods, such as mean imputation or deleting rows.

Amelia II has developed an algorithm that performs a multiple imputation on the dataset with missing values. These *m* imputations calculate point estimates for the missing values based on the average of two different variances; the variance within each completed dataset and the variance of point estimates between the datasets. The point estimates are subsequently multiplied by the factor that corrects for the bias.

The estimated variance of q from dataset j is (squared standard error):

$$SE(q_j)^2$$

The sample variance between the *m* point estimates:

$$S_q^2 = \sum_{j=1}^{m} \frac{\left(q_j - \bar{q}\right)^2}{(m-1)}$$

The variance of the multiple imputation point estimate is:

$$SE(q)^2 = \frac{1}{m} \sum_{j=1}^{m} SE\left(q_j\right)^2 + S_q^2\left(1 + \frac{1}{m}\right)$$

After Amelia II has run *m* times (by default m=5) the average of the *m* separated datasets is calculated (King, Tomz, & Wittenberg, 2000). The formula below describes how the datasets can be merged together to remain one estimated parameter $\bar{q}$, the average of the m separate estimates.

$$\bar{q} = \frac{1}{m} \sum_{j=1}^{m} q_j$$

3.3.2 Collinearity

The complete data matrix represented was measured with different variable units and the inertia (total variance) was distributed disproportionally over the variables. Therefore, normalization was required to create values between 0 and 1. The following formula was applied:

$$z_i = \frac{x_i - \mu}{\sigma}$$

Where:

z_i is the z-score for individual i

x_i is the real value for individual i

μ is the average of the sample

σ is the standard deviation of the sample

The correlation matrix was developed to identify underlying relations between the variables. The Variance Inflation Factor (VIF) was calculated to remove variables that showed high correlations. The VIF is calculated with the squared multiple correlation (R^2) between the variable i and the $(i-1)$ other predictors. When VIF reaches the threshold value of 5, the collinearity will be reduced by eliminating one or more variables form the analysis until no variable exceed a VIF-score higher than 5 (Jolliffe, 2002). VIF-score of 5 means a correlation (R^2) of 0.8 which is considered as a high correlation between variables. The formula to calculate the VIF score is written below;

$$VIF_i = \frac{1}{1 - R_i^2}$$

Where:

R_i^2 is the correlation for variable i

And VIF_i is the VIF score of variable i

3.3.3 Bartlett's test of sphericity and Kaiser-Meyer-Olkin-Criterium

The data matrix was complete and without inter-correlations above 0.8 between variables. Additional tests were required to examine whether Factor analysis could be performed on this matrix. Bartlett's test of sphericity and the Kaiser-Meyer-Olkin-Criterium (KMO) are tools to examine whether structures in the data matrix can be detected with a Factor analysis. Bartlett's test of sphericity tends to examine whether the correlation matrix is an identity matrix (Bartlett, 1951). When variables are uncorrelated, a Factor analysis cannot be performed because there are no underlying relations and therefore no structures to be detected. The following hypothesis is formulated:

H_0: Variables in the data matrix are uncorrelated

H_1: Variables in the data matrix are correlated

If the p-value is smaller than significance level ($\alpha = 0.05$) a Factor analysis can be performed.

The KMO test measures the proportion of variance of the variables that could be caused by underlying components (Kaiser, 1970). The output is a Measure of Sampling Adequacy (MSA) and the rule of thumb is that MSA variables with a value smaller than 0.5 are unacceptable for Factor analysis.

3.4 Factor Analysis

Descriptive statistics were calculated to describe the sample population and to display the characteristics of each variable. Thereafter, a factor analysis was performed to expose patterns among the inter-relations of the variables. The central idea of factor analysis is to reduce the dimensionality of a dataset in which there are a large number of uncorrelated variables, while retaining as much as possible of the variation present in the dataset. The objective is to develop a model that explains the variance between the variables by a set of fewer observed factors and their weightings (Osborne & Costello, 2009). Besides, Factor analysis aims to describe the variance explained by the factor where the unique variance (i.e. the variance explained by an external factor) is excluded (see Figure 3.5, the unique variance is represented as u_1.).

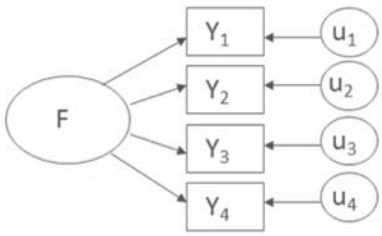

Figure 3.5 – Model of Factor Analysis (Jolliffe, 2002)

Factor analysis can be performed on latent variables; variables that are not directly observable or measurable. Hence, the vulnerability level can be gauged based on its features. The communality of the factor analysis is the proportion of variance explained by a set of factors which are common to the other observed variables. The formula is displayed below.

$$Y_1 = a_{11}F_1 + a_{12}F_2 + \ \dots + a_{1m}F_m + u_1$$
$$Y_2 = a_{21}F_1 + a_{22}F_2 + \ \dots + a_{2m}F_m + u_2$$
$$Y_Q = a_{Q1}F_1 + a_{Q2}F_2 + \ \dots + a_{Qm}F_m + u_Q$$

Where

Y_Q is the name of the variable with Q = 1,2,3...Q

F_m are the underlying factors with with m = 1,2,3...m

a_{Qm} is the weight of variable Q on factor m

u_Q is the unique variance of variable Q

First, eigenvalues are examined to estimate the amount of factors with the Kaiser Criterium. Thereafter, the factor loadings are calculated for each variable to analyze how variables are grouped together and measure the same factor. To optimize the factor solution elimination of variables is necessary for variables that are crossloading or have high correlation with more factors. Moreover, variables that have low loadings on all factors (below 0.5) were removed. Deletion of variables was executed sequentially since it changes the coefficients for all the other variables and other deletions may not be necessary or additional deletions are needed. Deletions were made until a simple structure was obtained, this means that every variable clearly falls under a single factor.

After determining to which factor each variable is assigned to, the variables in each factor were analyzed and examined whether they have something in common or if they are all assessing a distinct dimension.

3.5 Impact

After the factors were identified, the relation with impact was examined. Different data sources are available about impact of a flood however there is a lack of structured impact information about historical floods in Malawi. Currently, impact of floods in Malawi is measured by different organizations with different interests, such as flood type, number of people evacuated, injuries, and crop loss. Even though their data consists of all kinds of data related to floods, the assessments are conducted with different tools which forms an obstruction when merging the data. Therefore the determination of Cost and Data Quality of data sources was a key issue and will be done as follows (Van Den Homberg, Visser, & Van Der Veen, 2017):

The cost of a dataset determines the effort that has to be taken to extract information from a dataset. Structured files consider low costs since information can be easily extracted from the data file (e.g. Excel file). On the other hand, an unstructured data file indicates high costs since it is more difficult to extract information or data, such as a PDF report.

The quality of a dataset consists of four components:

- *Recency* determines when the dataset was updated for the last time and how long a dataset is representative of the reality. Recency will be relevant for historical impact data but in a slightly different way. For example, damage and needs assessments that are executed right after a disaster hits will be different from DNAs that are done two weeks later.
- *Source reliability* is the extent to which the data source is trustworthy, authentic and competent.
- *Content accuracy* is the extent to which the information is consistent with or confirmed by other independent sources.
- *Granularity*: the deeper the available granularity level, the higher the quality of the dataset is considered.

Different data sources, (such as reports, information sources, databases and journals) that examine the impact of a flood are assessed. Thereafter, the Cost and Data quality are examined. For some data sources, access was requested because of the protection of privacy sensitive information of the residents.

After collating the variables for impact, the relation with vulnerability factor was examined with the ordered logit model (McCullagh, 1980). The Ordered Logit Model is a method of ordinal regression where the independent variable, the impact data, is categorized.

4. RESULTS

4.1 Data matrix

The literature study yielded three relevant studies that examined indicators for the vulnerability level (i.e. Birkmann, 2006; INFORM, 2016; Wannewitz et al., 2016). These studies explained which indicators measure vulnerability which subsequently determined which variables should be selected. Different datasets were approached with the goal to cover the majority of the indicators. To downscale the index to subnational level, only datasets that include variables on TA level were accessed. Two large Household Survey datasets were approached which both were conducted by the National Statistics Office (NSO), a government department that conducts surveys in Malawi. The first was an Integrated Household Survey (IHS) that is conducted every five year, this fourth edition is conducted in 2016/2017 in 282 TAs. The second dataset is the Demographic and Health Survey (DHS) whose objective is to estimate basic demographic and health indicators. This dataset was conducted in 2015/2016 in 259 TAs. Next to the household surveys, the Malawi Hazards and Vulnerability Modeling Tool of the RCMRD (Regional Centre for Mapping of Resources for Development) was accessed to collect more variables on TA level (see http://tools.rcmrd.org/vulnerabilitytool/). Their data was collected in 2015. Finally, a previous study of 510 conducted by Wilbrink (2017) about remoteness indicators contained valuable data for this study.

In Table 4.1 the indicators that are covered in the dataset and that were not included in the dataset are displayed. Indicators that were not included in the dataset was due to the missing of the data or the unavailability of the variable on TA level. Some indicators were deleted from the dataset due to low data quality (e.g. recency, reliability or granularity) (Van Den Homberg et al., 2017). All the variables that are used in the dataset are displayed and further explained in the Appendix (Table 8.1).

Table 4.1 – Overview of indicators included in the dataset

Indicators suggested from literature review (INFORM[1], MOVE[2] framework, or Wannewitz[3])	**Variables included in the dataset**	**Indicators suggested from literature review but not included in data set (INFORM[1], MOVE[2] framework, or Wannewitz[3])**	**Cause***
Income[1]	Inc_ade	Life expectancy[1]	1
Health[1]	Chron_ill	Deprivations[1]	1
Population without access to electricity[3]	Dwell_elec	Income inequality[1]	1
Population living in informal settlements[3]	Dwell_trad	Education inequality[1]	1
Households without a mobile phone[3]	Mobile	Health inequality[1]	1
Living standards[1]	Food_wor Food_def Mark_acc Dwell_age	Differences in distribution of achievements[1]	1
population without access to safe drinking water[3]	Water_wo TT_WP	Number of refugees[1]	1
Population without a bike/boat[3]	Bike	Returned refugees[1]	1
Poverty as measured by the Wealth Index[3]	WI	Internally Displaced Persons[1]	1
Households without access to media (radio/TV)[3]	Radio TV	Tuberculosis prevalence[1]	1
Average number of poor people[1]	Pov_I	Cultural Vulnerability[2]	1
Malaria Mortality Rate[1]	Mal_sus	Institutional Vulnerability[2]	1
Children mortality[1]	Inf_mort	Good governance index[3]	1
Physical Vulnerability[2]	Road_dens	Stunting in children under 5 (per km2)[3]	3

Indicators suggested from literature review (INFORM[1], MOVE[2] framework, or Wannewitz[3])	Variables included in the dataset	Indicators suggested from literature review but not included in data set (INFORM[1], MOVE[2] framework, or Wannewitz[3])	Cause*
Population density[3]	Pop_dens House_dens	Protected areas[3]	1
Illiteracy[3]	Literacy_I	Forest cover change[3]	1
Environmental Vulnerability[2]	TT_TC	Density of aid projects[3]	1
Shelter density: primary/secondary schools[3]	TT_PS TT_SS	Dependency ratio[3]	1
Density of emergency services[3]	TT_H TT_C	Dependency on agriculture[3]	1
Education[1]	Edu_mother School_none School_prim School_sec	Conflict density km2[3]	1
Population living in poorly constructed houses[3]	Floor_nat Roof_nat	Prevalence of HIV-AIDS above [1]	2
Social Vulnerability[2]	Health_ade House_ade Father_wo Mother_wo Sens_sc Fe_head Child_5	Unemployment (%)[3]	2
Economic Vulnerability	Mosq_net Food_ade Cloth_ade Dwell_age Bank_acc	GDP per capita ($)[3]	2
		Children under weight[1]	3
		Relative number of affected population by natural disasters in the last three years[1]	3
		Population without a car/motorcycle[3]	3
		Households without access to the internet[3]	3
		Population without access to sanitation[3]	3

Cause: 1 = Not found, 2 = not available on TA level, 3 = excluded due to low data quality*

4.2 Data Preparedness

The original dataset contained 362 rows that represented the Traditional Authorities and 44 columns that represented the variables. First, 59 TAs were deleted since 27 variables were not measured in these TAs which was the majority of the missing values. Deleting the remained TAs with missing values would decrease the dataset with 74 TAs which would diminish the power of the dataset and eliminate a high amount of information, therefore the Amelia II method was applied.

The figure below shows the missing values for each TA. The Amelia II function was performed to fill in the missing values which resulted in a complete dataset. Thereafter, the dataset was standardized to create a dataset measure with the same unit of analysis.

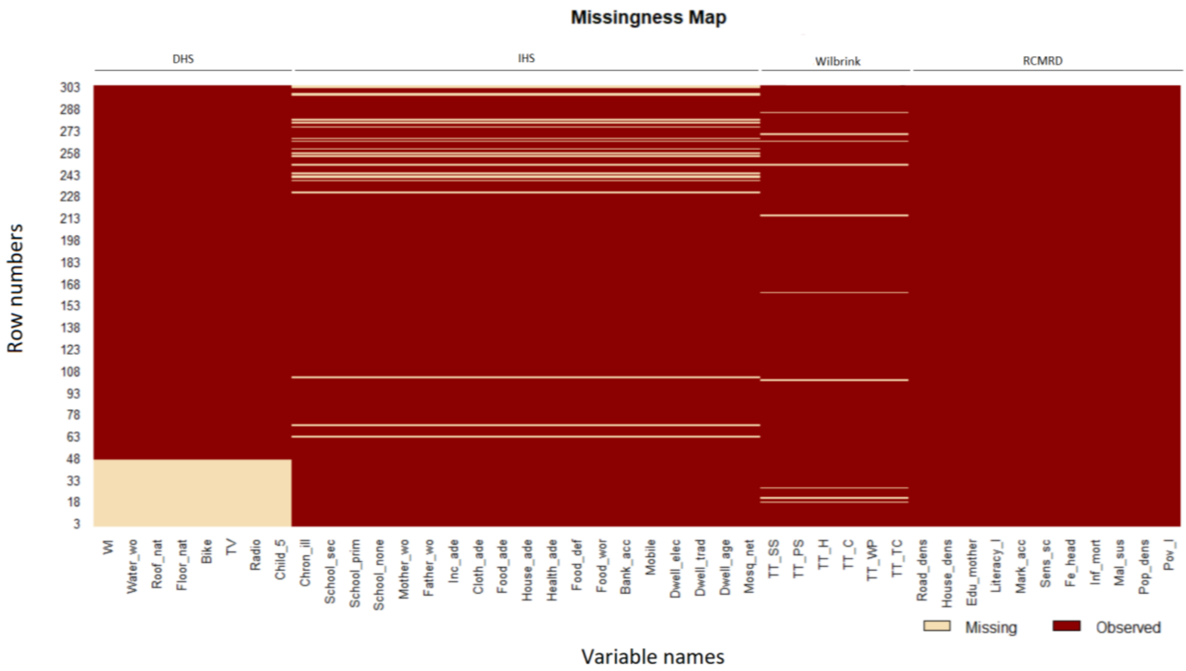

Figure 4.1 – Map of missing values

Thereafter, correlations between variables were calculated (see Figure 4.2). The blue circles represent positive correlations whereas the red circles represent negative correlations. The size of the circle represents the correlation level, where a big circle means a high correlation. The matrix shows that there are correlations among different variables. For future replications of this research, it is important to examine the correlations to understand the context better and to detect which variables need more attention in future data collection. For instance, WI is highly positive correlated with seven variables (e.g. Mobile) and highly negative correlated with twelve other variables (e.g. TV and Floor_nat). Furthermore, Mobile and Roof_nat show many positive and negative correlations with several variables. To delete variables gradually (because the correlations change after one variable is deleted) the VIF factor is used (see Figure 4.3). All variables with a VIF score higher than five were removed. In total, the following 11 variables were removed from the data matrix; TV, Floor_nat, Roof_nat, Dwell_elec, Mobile, School_none, School_prim, School_sec, WI, Pov_l and Sens_sc, due to multicollinearity. The VIF score before and after removing the variables is displayed in Table 8.2 in the appendix for each variable.

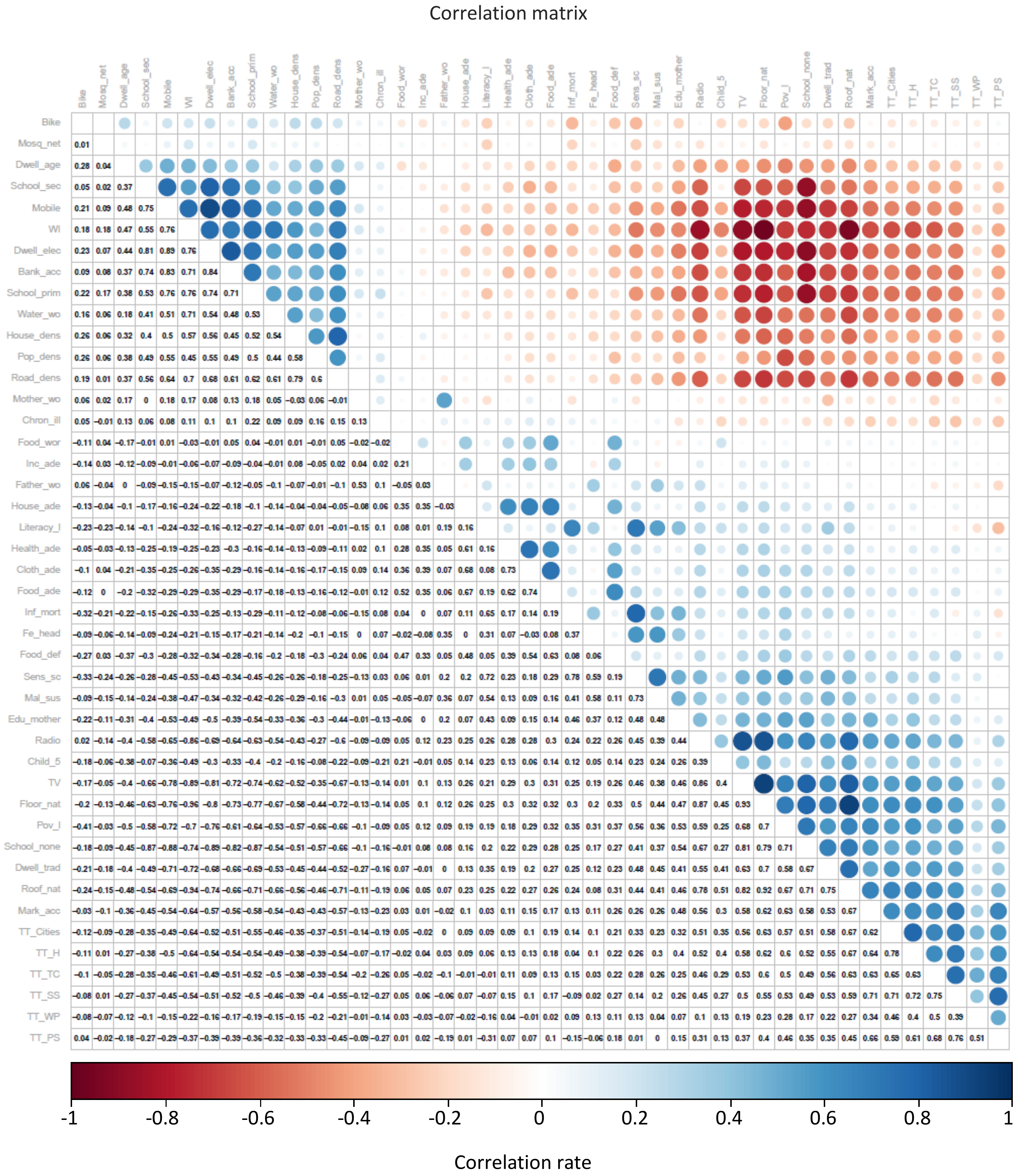

Figure 4.2 – Correlation matrix between all variables

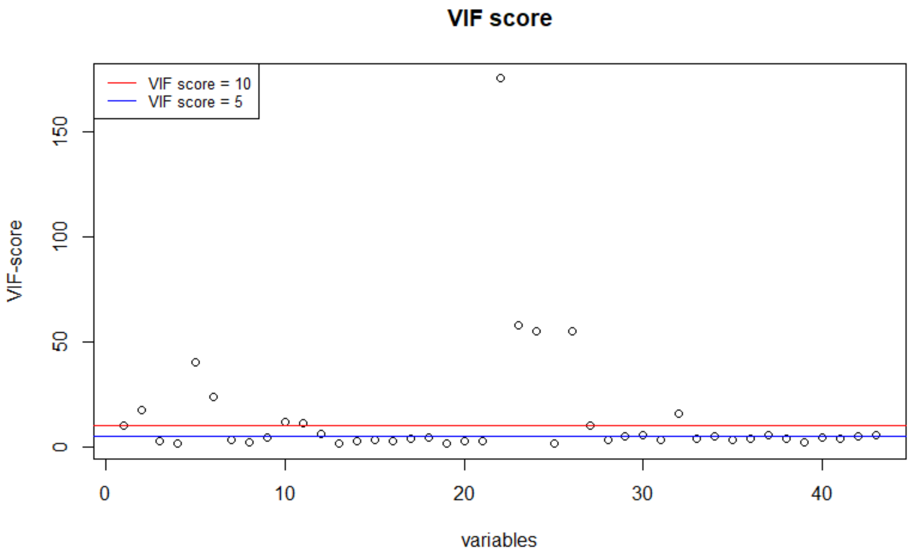

Figure 4.3 – VIF score for each variable

Thereafter, The Bartlett's Test of Sphericity and the KMO-test were performed. The Bartlett's Test of Sphericity is significant which rejects the null hypothesis (see Table 4.2). Moreover, Table 4.3 shows that all values of the KMO-test are above 0.5. Therefore, a factor analysis can be performed to find underlying relations in the data matrix.

Table 4.2 – Bartlett's test of Sphericity

X2	*= 2251.409*
Df	*= 528*
p-value	*< 2.22e-16*

Table 4.3 – KMO test

Variable name	MSA	Variable name	MSA	Variable name	MSA
Child_5	0.79	House_ade	0.84	Mark_acc	0.92
Radio	0.89	Food_ade	0.88	Literacy_l	0.75
Bike	0.64	Cloth_ade	0.80	Edu_mother	0.88
Mosq_net	0.60	Inc_ade	0.84	House_dens	0.85
Water_wo	0.88	Father_wo	0.61	Road_dens	0.88
Dwell_age	0.84	Mother_wo	0.52	TT_Cities	0.91
Dwell_trad	0.90	Chron_ill	0.80	TT_WP	0.82
Bank_acc	0.91	Pop_dens	0.90	TT_TC	0.88
Food_wor	0.75	Mal_sus	0.77	TT_H	0.91
Food_def	0.87	Inf_mort	0.67	TT_PS	0.90
Health_ade	0.83	Fe_head	0.61	TT_SS	0.90

4.2 Descriptive Statistics

In Table 4.4 the descriptive statistics of eight variables are displayed after the missing values are filled in using Amelia. The complete table is present in the Appendix. The table shows the mean and standard deviation (SD) as well as the skewness (see in the appendix Table 8.3 for the descriptive statistics of all variables). For the variable Food_wor there were 282 TAs examined which meant there were 21 missing values. On average, 62 per cent was worried about having enough food with a Standard Deviation of 0.11. The median was 0.63 and the minimum and maximum score were respectively 0.25 and 0.94. Looking at the skewness, the distribution is approximately symmetric and the kurtosis is 0.41. Another example is the TT_Cities which was collected from the study of Wilbrink (2017). In 293 TAs the Travel Time to Cities was calculated. The average Travel Time was 141.97 minutes with a standard deviation of 109.17 minutes and a median of 113 minutes. The minimum is 0 which are the people that already live in a city and the maximum was 500. The distribution is highly positive skewed.

Table 4.4 – Descriptive statistics for 8 variables in the dataset

Variable name	N	Missing values	Mean	SD	Median	Min	ax	Skew	Kurtosis	SE
Food_wor	282	21	0.62	0.11	0.63	0.25	0.94	-0.40	0.41	0.01
Cloth_ade	282	21	0.67	0.12	0.69	0.19	0.94	-0.70	1.00	0.01
Mal_sus	303	0	53.81	20.23	52.55	0.00	93.71	-0.03	-0.56	1.16
Inf_mort	303	0	80.34	12.01	84.11	0.00	100.00	-2.18	11.30	0.69
Literacy_l	303	0	23.28	10.47	24.39	0.00	50.43	-0.18	-0.45	0.60
House_dens	303	0	135.38	334.17	1.88	0.00	1951.04	3.28	11.15	19.20
TT_Cities	293	10	141.97	109.72	113.00	0.00	500.00	1.02	0.52	6.41
Pov_l	303	0	75.04	20.07	82.60	0.00	97.27	-1.60	1.86	1.15

4.3 Explorative Factor analysis

The screeplot in Figure 4.4 depicts the eigenvalues per factor. According to the Kaiser criterion, the amount of eigenvalues that are higher than 1 should be the total factors however this is not a rule of thumb (Jolliffe, 2002). Variables with a loading higher than 0.5 are assigned to a factor. The figure depicts that the estimated amount factors for Factor analysis is three and for Principal Component analysis five. After examining different factors, the total amount of five factors fits the model the best. All variables were assigned to a factor, except for inc_ade since this variable did not have a loading higher than 0.5 with any factor. See Figure 4.5 for an overview of all variable loadings per factor. Furthermore, Table 4.5 displays the variance explained by the factors. The first two factors explain most of the variance with respectively 17% and 16% however the proportion of variance explained is approximately equally divided. The total variance explained by the five factors is 59%.

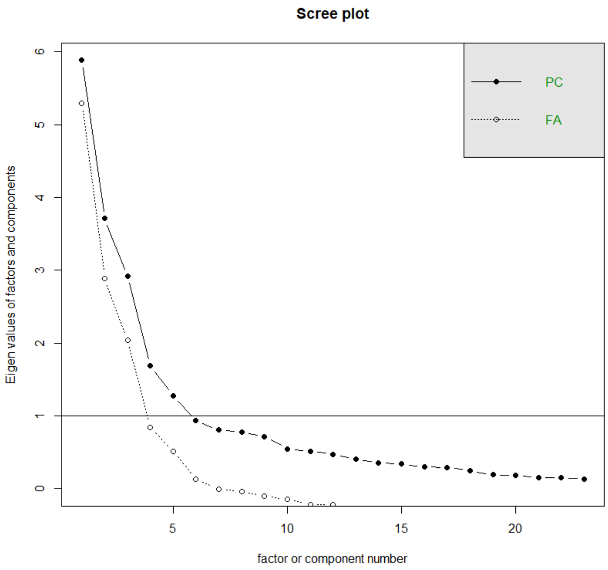

Figure 4.4– Screeplot eigenvalues

Figure 4.5 - Variable loading for each factor

Table 4.5 – Explained variance

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*
SS loadings	3.99	3.57	2.65	2.10	1.33
Proportion Var	0.17	0.16	0.12	0.09	0.06
Cumulative Var	0.17	0.33	0.44	0.54	0.59

Figure 4.6 shows factors and the corresponding variables with a loading higher than 0.5. The factor names were chosen to describe what the several variables under a factor represented together. The variables in the first factor are related to the Average Travel Time to Important Locations, variables in the second factor refer to the Lack of Household Needs, variables in the third factor refer to Densities, variables in the fourth factor are related with Education level and Vulnerable Groups and the last factor refers to wether parents are part of the household.

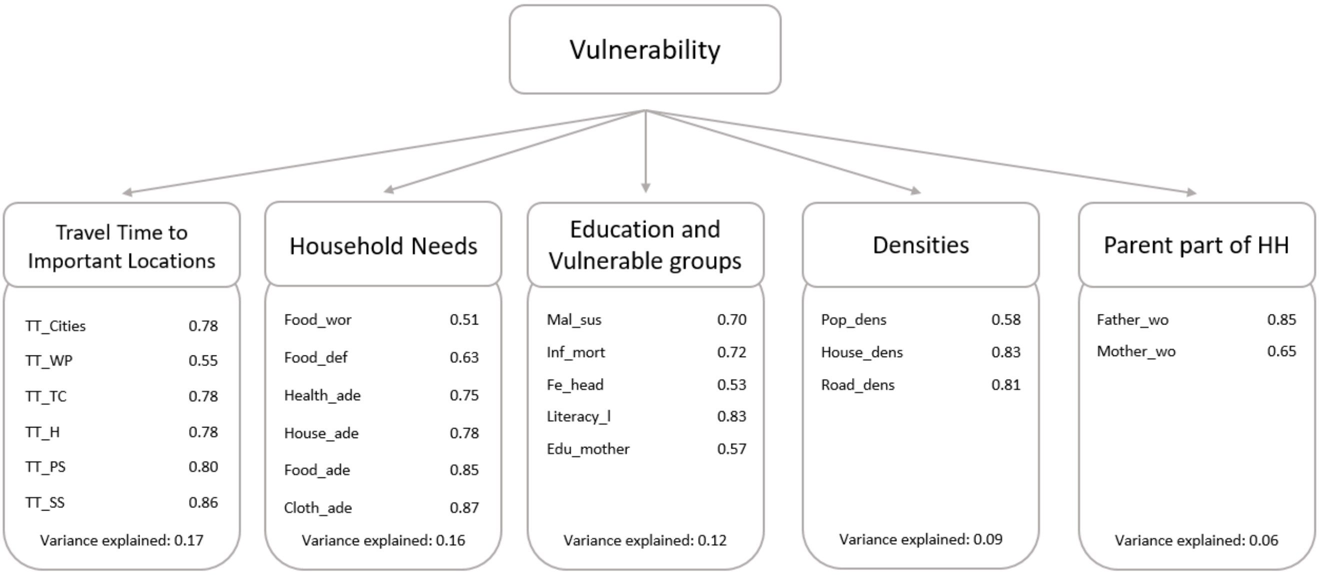

Figure 4.6 – Variables per factor

4.4 Mapping vulnerability on TA level

The calculation of the vulnerability level resulted in the development of six maps (made with QGIS 2.18) visualizing the Malawi's vulnerability level with five different factors and the overall vulnerability level (see Figure 4.7). The maps were created with analyzing quantiles, a method that classifies data into a certain number of categories with an equal number of units in each category. Therefore, it can be easily observed that the vulnerability level differs considerably with respect to their spatial distribution throughout Malawi. These differences between the TAs emphasize the importance of collecting data on subnational level.

Looking at the distribution of the overall vulnerability level, there are differences, meaning that the TAs in this region have to face different levels of vulnerability. In order to understand where the vulnerability heterogeneity originates from, the factors can be analyzed with the help of the other maps (Factor 1 until Factor 5).

Further validation of these results could be done with a panel of experts since indigenous knowledge is highly relevant. Contextual and situational knowledge of the country is still important to validate the results.

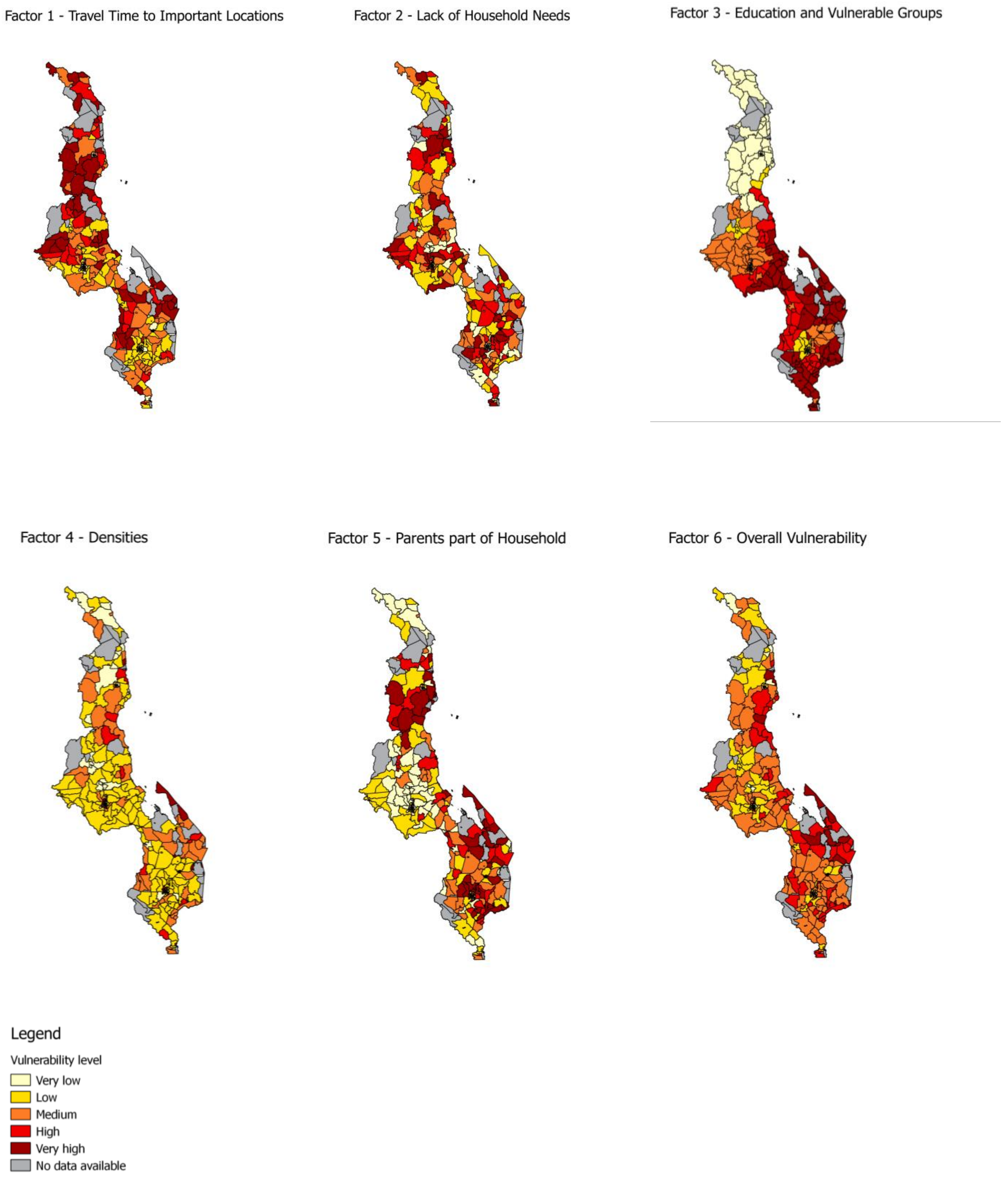

Figure 4.7 – The downscaled vulnerability index for Malawi on TA level

4.5 Impact data of Malawi

The table below is an overview of different data sources that were available on impact of a flood in Malawi. The data sources show many differences in conducting the surveys and reporting results therefore the reliability is difficult to estimate. Besides, some data was not available on TA level and therefore not useful for this study. The table was filled in with a Likert-scale ranging from 1 *(poor)* to 5 *(very good)*.

Based on the analysis in Table 4.6 the following two data sets were selected:

The first dataset that was used to examine impact of historical floods was the International Organization for Migration United States (IOM UN). They conducted surveys in refugee camps after 2015. In the survey they detected from which TA most of the refugees originally fled from. They are called Internally Displaced People (IDP).

The Integrated Household Survey included a theme that covered impact data of natural disasters. Two relevant questions were selected for this study. The first one was: *In the last 12 months, have you been faced with a situation when you did not have enough food to feed the household?* Afterwards they asked: *What was the cause of this situation?* Where the cases that mentioned flood were selected. The second question was: *What share of the community was affected?* Where the multiple choice consisted of *25%, 50%, 75%* and *almost all*.

Table 4.6 – Datasources examining impact of floods in Malawi

Description		**Data Quality**				**Costs of data**	
Data set	**Source of data**	**Date of source/disasters covered**	**Source reliability**	**Content accuracy**	**Granularity**	**Type**	**Structure level**
Flood report	IFRC relief workers	Historical floods	3	4	District	Report	unstructured
Disaster Database	EM-DAT	1967-2017	4	4	National	Database	structured
Internally displaced persons	OCHA	2010-2016	4	3	District	Dataset	structured
Damage and loss assessment	Government	2015	3	4	District	Report	Semi-structured
Survey in camps of IDPs	*IOM UN migration*	*January and February 2015*	*4*	*4*	*GPS - pixel*	*Dataset*	*structured*
Household survey	*NSO*	*2012-2017*	*3*	*4*	*TA level*	*Dataset*	*Structured*

4.6 Regression analysis on Vulnerability and Impact

To identify underlying relations with vulnerability, a Ordered Logit regression analysis was performed with two data sources selected form Table 4.6. In total, people fled from 27 different TAs due to the flood in 2015. Two relations were identified with the ordered logit model. First, Education level and Vulnerable Groups had a significant relation with IDP level (*p-value < 0.05, coeff = 7.55)*. This indicates that when the vulnerability level of *education level and vulnerable groups* increases in a TA, this area more likely is to experience a higher rate of IDP after a flood. Second, Parents part of Household has a significant result however a negative coefficient (*p-value < 0.05, coeff = -2.08).* This indicates that when the vulnerability level of *Parents part of Household* increases, this area more likely is to experience a lower rate of IDP after a flood which is contrary to the expectations. The other factors and the total vulnerability does not seem to have a relation with this impact variable (see Table 4.7).

In some TAs, people suffered from a food deficit while in other TAs this consequence was not mentioned. After a Ordered Logit Regression, two factors of vulnerability showed a relation (see Table 4.8). Areas where the Travel time to Important Locations is higher are more likely to suffer from food deficit than other areas (*p-value < 0.1, coeff = 0.94)*. Furthermore, Parents part of Household seems to have a negative relation with impact which is again in contrast to expectations. When the vulnerability of this factor increases, people are less likely to experience a food deficit after a flood (*p-value < 0.05, coeff = -0.82).* The other factors and the total vulnerability does not seem to have a relation with this impact variable.

Moreover, the integrated household survey of NSO included another variable that was relevant for this study and measured on TA level. The dataset contain the share of community affected after previous floods (time period: 2012 until 2016). There were 4 categories identified; 25%, 50%, 75% and 100%. After performing an Ordered Logit regression, the vulnerability factor Densities showed a relation with this impact variable (see Table. 4.9). When the Densities vulnerability factor increases, people are more likely to get affected by a flood than others (*p-value < .05, coeff = 1.10).* The other factors and the total vulnerability does not seem to have a relation with this impact variable.

Table 4.7 – Ordered Logit Model between Vulnerability and IDP

Dependent variable: IDP			
Y is a categorical variable of amount of IDPs per TA after the flood of 2015, with: 1: < 100 IDPs 2: 100 < 1000 IDPs 3 > 1000 IDPs.			
Independent variable:	*Coefficient*	*S.E.*	*Pr (> | Z |)*
Travel Time to Important Locations	- 3.1842	2.1188	0.1329
Lack of Household Needs	- 1.5690	1.0888	0.1496
Densities	- 0.8686	1.0826	0.4224
Education level and Vulnerable Groups	7.5493	3.7240	0.0426*
Parents part of Household	- 2.0827	0.8491	0.0142*
Total Vulnerability	- 0.1106	0.3617	0.7597

*Notes: Significance level: *p <.05.*

Table 4.8 – Ordered Logit Model between Vulnerability and Food deficit

Dependent variable: Food deficit			
Y is a categorical variable indicating whether people suffered from food deficit after previous floods, with: *0: did not suffer from a food deficit after a flood.* *1: Suffered from food deficit after a flood*			
Independent variable:	*Coefficient*	*S.E.*	*Pr (> | Z |)*
Travel Time to Important Locations	0.9411	0.5280	0.0747**
Lack of Household Needs	0.2943	0.3238	0.3635
Densities	- 0.1771	0.4738	0.7085
Education level and Vulnerable Groups	- 0.7908	0.5483	0.1493
Parents part of Household	- 0.8215	0.3582	0.0218*
Total Vulnerability	0.0459	0.1467	0.7543

*Notes: Significance level: *p <.05, **p< .1*

Table 4.9 – Ordered Logit Model between Vulnerability and Share affected

Dependent variable: Share affected			
Y is a categorical variable indicating the share of TA affected after past floods, with: *0.25: 25% affected* *0.50: 50% affected* *0.75: 75% affected* *1: 100% affected*			
Independent variable:	*Coefficient*	*S.E.*	*Pr (> | Z |)*
Travel Time to Important Locations	-0.1410	0.4757	0.7670
Lack of Household Needs	-0.2932	0.3067	0.3391
Densities	1.0989	0.4348	0.0115*
Education level and Vulnerable Groups	-0.5282	0.4449	0.2351
Parents part of Household	-0.1551	0.3126	0.6199
Total Vulnerability	0.0680	0.1468	0.6433

*Notes: Significance level: *p <.05*

5. DISCUSSION

The objective of this study was to examine the relation between vulnerability and impact after floods in Malawi. The findings of each research question will be discussed below.

1. *Which frameworks and associated indicators are currently used to characterize vulnerability?*

Three studies were examined to understand the concept of vulnerability and to identify indicators that measure this latent variable. The first study that characterize vulnerability was the INFORM index which is currently used by 510. The INFORM index uses a framework that is used to calculate the risk for each country based on three dimensions; Hazard & Exposure, Vulnerability and Lack of Coping Capacity. Each dimension is divided into categories and components which gave a clear understanding of the indicators that measure vulnerability. The second study is the MOVE Framework of Birkmann et al. (2013). They explain vulnerability in the context of natural hazards and divide it into different dimensions, for instance, Social Vulnerability, Economic Vulnerability and Physical Vulnerability. The third study is conducted by Wannewitz et al. (2016). This study developed a multi-hazard risk index on sub-national level in the Philippines and identified indicators that determine vulnerability. All frameworks were related to Disaster Risk Reduction, risk of natural hazards or downscaling indices and therefore highly useful for this study. After examining the frameworks, it was clear what indicators determine vulnerability in this context.

2. *Which data sources are available on the vulnerability in Malawi on subnational level?*

Different global open sources were approached to collate data and develop a data matrix of vulnerability in Malawi on Traditional Authority level. The Integrated Household Survey and the Demographic and Health Survey contained variables about demographics. Wilbrink (2017) conducted a study that calculated the travel times to different locations that have an impact on social vulnerability, such as travel times to water points, markets, cities or schools. Furthermore, the RCMRD contained many variables that were not included in the previous open sources, such as literacy level, malaria susceptibility and infant mortality. These four sources contained structured data on TA level and was collected from 2015 onwards. The reliability and accuracy was checked by comparing the different sources and was considered as sufficient. However, when data quality was insufficient, the variables were not included in the dataset.

3. *Which exploratory (unobserved) factors can be extracted from the (observed) vulnerability indicators?*

Data preparedness played a central role after collating data from different open sources and before analyzing it. Among others, missing values had to be filled in with Amelia II and the VIF score was calculated to avoid multicollinearity and reduce the number of variables of the data matrix. The Bartlett's test of Sphericity and KMO test were performed to detect whether there were underlying relations between the remained variables. Since both tests were significant, a Factor analysis could be performed to reduce dimensionality and describe variability among the observed variables in terms of lower number of unobserved factors. In total, five factors were identified that measure vulnerability: Average Travel Time to Important Locations, Lack of Household Needs, Densities, Education level and Vulnerable Groups, and Parent part of Household. Subsequently, factor scores for each TA were calculated and vulnerability maps for all factors and the total vulnerability were successfully created to visualize the results.

Following from an in-depth analysis for Malawi's vulnerability pattern, it can be easily observed that the overall vulnerability as well as its corresponding factors vary in different Traditional Authorities. While on a global scale vulnerability is calculated for each country where a homogeneous distribution of the concept is implied throughout the country, the maps show that this is a strong generalization (INFORM, 2016). These heterogeneous patterns emphasize the importance of subnational assessments and examinations. Thus national vulnerability indices cannot contribute significantly to national Disaster Risk Reduction and even hold the potential to draw wrong or misleading conclusions for the management of DRR. The comparison of the six maps illustrates how vulnerability is composed differently with respect to their spatial distribution within Malawi. The maps show that the Northern part and the South-East of Malawi have to deal with longer Travel Times to Important Locations and Parents are mainly not Part of the Household which makes the TA more vulnerable. The South is more vulnerable in terms of Education Level and Vulnerable Groups, for instance people that are more susceptible for malaria. Furthermore, some TAs show high Density rates which make the areas more vulnerable when a flood strikes since an increasing amount of people and assets are exposed to the disaster. Moreover, the vulnerability factor about Household Needs is divided throughout the country. All the five factors merged together results in the Total Vulnerability level where the most vulnerable places are in the South-East and North-East. The vulnerability maps enable humanitarian organizations, such as The Red Cross, to detect which areas should be provided with aid first and the factors could be included in the CRA dashboard.

4. *Which data sources are available on the impact of historical floods in Malawi on subnational level?*

Different open data sources were approached to examine the impact of historical floods in Malawi. However, the different organizations examined the impact in different manners. For example, the EM-DAT database reports amount of people affected, missing and injured while the governmental Damage and loss assessment calculates the impact per district in US dollars. Therefore, the Data quality was evaluated before performing an analysis on impact and vulnerability. Only two datasets were useful for this research since they were reliable and measured on TA level. The IOM UN migration reported the amount of Internally Displaced People per TA. The integrated Household Survey measured whether people suffered from Food Deficit and the Share of People Affected after a flood.

5. *To what extent is vulnerability related with impact of a flood in Malawi?*

Three relations between impact and vulnerability factors were found after analyzing the data. They are explained below.

Relation between vulnerability and Internally Displaced People.

- *The lower the education level and the more vulnerable groups in a TA, the more likely the people in this TA will flee when a flood strikes.*
- *The more households where one or two of the parents are not part of the household, the less likely the people in this area will flee when a flood strikes.*

The first relation shows that vulnerable groups and people with a low education level have more difficulties when being exposed by a flood. This can be explained by the expectation that these groups have less resources, such as knowledge, assets or more concrete houses, to stay at the same place when a flood strikes (Brouwer, Akter, Brander, & Haque, 2007). Therefore, these people are forced to move to another place for a certain time period to endure the hazard. The second relation is against expectations. An explanation could be that households with only a female adult prefer to stay at home instead of fleeing given gender based violence (Bhadra, 2017). Nevertheless, the Parents part of Household factor only accounts for 6% of the variability of the variables characterizing the Vulnerability, thus it is a less relevant factor.

Relation between vulnerability and food deficit.

- *The longer the travel distance to important places, the higher the chance that people will suffer from a food deficit after a flood.*
- *The more households where one or two of the parents are not part of the household, the smaller the chance that people will suffer from a food deficit after a flood.*

The first relation did not have a strong significance, however it shows the importance of infrastructure in an undeveloped country. When people have to travel longer to important locations, such as hospitals, schools, markets and city centers, they will experience even more difficulties in case a flood strikes. The infrastructure is usually scarce and will be damaged or even destroyed after a flood which makes it more difficult to travel to places to gain food. The second relation is again against expectations. An explanation can be that families move in to other families due to the impact of the flood and therefore might not experience food deficit since the families share needs.

Relation between vulnerability and shared affected.

- *The higher the density level in a TA, the higher share will be affected after a flood.*

Even though this relation might be straightforward, it is important for humanitarian organizations to realize this outcome. When a flood hits a TA with a high population rate, more people will be affected. Furthermore, the infrastructure density might indicate higher levels of coping capacity and resilience however it also indicates that more roads will be destroyed when exposed to a flood and thus the impact level increases.

Main research question: How can vulnerability be used to predict the impact of floods?

This study shows different relations between vulnerability and the impact of a flood. An important note is that different factors of vulnerability should be examined to find relations with the impact. Furthermore, the vulnerability level should be observed on sub-national level because the distribution of vulnerability in this study was heterogeneous among the different Traditional Authorities. On the other hand, impact should be examined in different types as well, since it shows different relations with the vulnerability factors. Finally, the conclusion of this study can be drawn that the impact of a flood can be estimated with vulnerability. Therefore, the identification of vulnerability factors and the determination of vulnerability levels within a country can be a great tool for predicting the impact of a flood.

6. CONCLUSION

In this study, prior flood-related impact and vulnerability were examined to explore the factors that influence how households experience flooding. It is highlighted that it is of major importance to determine the vulnerability level on subnational level since the levels move across space. The heterogeneous distribution depicted from the vulnerability maps confirm this importance. Furthermore, it is important to understand the negative consequences of a flood which should be collected by (1) gaining indigenous knowledge and (2) collect structured impact data with, for instance, post-disaster needs assessments. A deep understanding of the impact of floods is essential for future forecasts. In this study, the data availability and the scale of application caused a critical dilemma on the examination. However, the data showed interesting relations between vulnerability and impact which is an essential first step towards impact-based forecasting.

7. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

The biggest limitation of this research was the lack of available (valuable) data. Therefore, gaps were substituted with assumptions, past field experience of Red Cross members and statistical methods. The reliability of the conclusions drawn from the analysis is threatened due to data scarcity, especially when zooming in from national to sub-national level. However, this study was performed in order to conduct the first steps of impact-based forecasting. The limitations experienced in this research are fundamental for future works. Therefore, this chapter contains several recommendations for future research in this topic.

First, the study indicates that there is a relation between vulnerability and impact therefore it is of great importance to study the vulnerability level more in depth. The spatial distribution should be deeper examined to increase reliability of the factors and the general vulnerability. Artificial Intelligence techniques, such as Machine Learning techniques, could be implemented to gain more data. With Machine Learning techniques, different vulnerability indicators can be detected with geospatial information, such as infrastructure, agriculture and even the size of the villages. This subsequently means that the vulnerability level can be examined on even a smaller scale.

Second, this study demonstrates that impact related to flood events present high variability and diversity in different places due to space variability of people's vulnerability. The lack of impact data of historical floods caused limitations in these results. Therefore, the lack of reliability of these results should be taken into consideration. However, the results show that there is a relation with vulnerability, which is step forward in validating the method of impact-based forecasting. Future research needs more and reliable impact data. A structured way of reporting impact is of great importance for different stakeholders with different interests and highly necessary. Field assessments should be executed where the cooperation between different organizations is strongly advised.

The third limitation is the data collation where different datasets are merged into one data matrix. It caused missing values in the dataset which were eventually solved with the Amelia II algorithm. However, remarkably different organizations conduct intensive studies in data collection, for example household surveys, which are overlapping. It would be more efficient for organizations, as well as more convienient for inhabitants, to cooperate in conducting field assessments. It would lead to a higher sample where both needs of the organizations are met.

However, the data that was collected is still valuable despite its limitations. It is recommended to validate the results with a panel of experts that have more knowledge about the context in Malawi. This validation will give more information on the reliability of the sources that are used and where gaps are still existing.

Despite the shortcomings and limitations detailed above it can be concluded that a vulnerability index is a highly valuable tool for humanitarian aid organizations. The method is applicable informing decision-makers in the field of DRR on subnational scale. For 510 it is useful to include factors in their CRA dashboard to increase the reliability and validity of the current toolbox. Furthermore, the method ensures a great anticipation on forecasting impact and should therefore be more examined.

REFERENCES

Atkins. (2011). *Integrated Flood Risk Management Plan for the Shire Basin project Inception Report*, (December), 107 pages.

Bartlett, M. S. (1951). The Effect of Standardization on a χ2 Approximation in Factor Analysis. *Biometrika*, *38*(3/4), 337–344. https://doi.org/10.2307/2332580

Bhadra, S. (2017). Women in Disasters and Conflicts in India: Interventions in View of the Millennium Development Goals. *International Journal of Disaster Risk Science*, *8*(2), 196–207. https://doi.org/10.1007/s13753-017-0124-y

Birkmann, J. (2006). *Measuring vulnerability to promote disaster-resilient societies: Conceptual frameworks and definitions*. *United Nations University Press*. https://doi.org/10.1111/j.1539-6975.2010.01389.x

Birkmann, J., Cardona, O. D., Carreño, M. L., Barbat, A. H., Pelling, M., Schneiderbauer, S., … Welle, T. (2013). Framing vulnerability, risk and societal responses: The MOVE framework. *Natural Hazards*, *67*(2), 193–211. https://doi.org/10.1007/s11069-013-0558-5

Bogardi, J., & Birkmann, J. (2004). Vulnerability assessment: the first step towards sustainable risk reduction. In: Malzahn D, Plapp T (eds) Disasters and society—from hazard assessment to risk reduction. Logos, Berlin.

Broeken, M., Homberg, M. van den, Visser, J., Bolton, E., Jong, N. de, Jaramillo, R., … Larsen, J. L. (2018). Data Preparedness for Forecast-based Financing: a pilot study on vulnerability and impact data for Malawi. In *Geophysical Research Abstracts*. EGU General Assembly 2018.

Brouwer, R., Akter, S., Brander, L., & Haque, E. (2007). Socioeconomic Vulnerability and Adaptation to Environmental Risk: A Case Study of Climate Change and Flooding in Bangladesh, *27*(2). https://doi.org/10.1111/j.1539-6924.2007.00884.x

Bryant, F. B., & Yarnold, P. R. (1995). Principal components analysis and exploratory and confirmatory factor analysis. American Psychological Association Books.

Cannon, T., Twigg, J., & Rowell, J. (2003). *Social Vulnerability. Sustainable Livelihoods and Disasters*.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social Vulnerability to Environmental Hazards. *Social Science Quarterly*, *84*(2), 242–261. https://doi.org/10.1111/1540-6237.8402002

de Perez, E. C., Van den Hurk, B., Van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., … Zsoter, E. (2016). Action-based flood forecasting for triggering humanitarian action. *Hydrology and Earth System Sciences*, *20*, 3549–3560. https://doi.org/10.5194/hess-20-3549-2016

de Perez, E. C., Van Den Hurk, B., Van Aalst, M. K., Jongman, B., Klose, T., & Suarez, P. (2015). Forecast-based financing: An approach for catalyzing humanitarian action based on extreme weather and climate forecasts. *Natural Hazards and Earth System Sciences*. https://doi.org/10.5194/nhess-15-895-2015

Dube, T., Moyo, P., Ncube, M., & Nyathi, D. (2016). The Impact of Climate Change on Agro-Ecological Based Livelihoods in Africa: A Review. *Journal of Sustainable Development*, *9*(1), 256–267. https://doi.org/10.5539/jsd.v9n1p256

Galindo, G., & Batta, R. (2013). Socio-Economic Planning Sciences Prepositioning of supplies in preparation for a hurricane under potential destruction of prepositioned supplies. *Socio-*

Economic Planning Sciences, *47*(1), 20–37. https://doi.org/10.1016/j.seps.2012.11.002

GFDRR. (2010). Damage, Loss, and Needs Assessment, Guidance Notes. Design and Execution of a Damage, Loss, and Needs Assessment, *1*, 1–86. Retrieved from https://openknowledge.worldbank.org/handle/10986/19047

GFDRR. (2016). Implementing Multi-Hazard Impact-based Forecast and Warning Services. *Workshop Organized by China Meteorological Administration – Shanghai Meteorological Service and the Global Facility for Disaster Reduction and Recovery*, (December), 76 pages.

GFDRR. (2017). PDNAs. Retrieved from https://www.gfdrr.org/en/post-disaster-needs-assessments

Hammond, M. J., Chen, A. S., S. Djordjević, Butler, D., & Mark, O. (2014). Urban flood impact assessment: A state-of-the-art review. *Flood Impact Assessment Literature Review.*, University of Exeter (UK), CORFU project, Report N. https://doi.org/10.1080/1573062X.2013.857421

Harriman, L. (2014). Cyclone Phailin in India: Early warning and timely actions saved lives. *Environmental Development*, *9*, 93–100. https://doi.org/10.1016/j.envdev.2013.12.001

Honaker, J., King, G., & Blackwell, M. (2011). AMELIA II : A Program for Missing Data. *Journal Of Statistical Software*, *45*(7), 1–54. https://doi.org/10.1.1.149.9611

IFRC. (n.d.). The Seven Fundamental Principles. Retrieved from http://www.ifrc.org/en/who-we-are/vision-and-mission/the-seven-fundamental-principles/

INFORM. (2016). *Concept and Methodology*. Retrieved from http://www.inform-index.org/Portals/0/InfoRM/2016/INFORM Concept and Methodology Version 2016 updated cover.pdf?ver=2015-11-26-123614-937

IPCC. (2007). *Climate Change 2007: Impacts, Adaptation and Vulnerability*.

IPCC. (2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. https://doi.org/10.1017/CBO9781139177245

IPCC. (2014). Climate Change, Adaptation, and Vulnerability. *Organization & Environment*, *24*(March), 1–44. https://doi.org/http://ipcc-wg2.gov/AR5/images/uploads/IPCC_WG2AR5_SPM_Approved.pdf

Jolliffe, I. T. (2002). Principal Component Analysis, Second Edition. *Encyclopedia of Statistics in Behavioral Science*, *30*(3), 487. https://doi.org/10.2307/1270093

Joshua, M. K., Ngongondo, C., Monjerezi, M., Chipungu, F., Liwenga, E., Majule, A. E., … Lamboll, R. (2016). Climate change in semi-arid Malawi : Perceptions , adaptation strategies and water governance. *Jàmbá: Journal of Disaster Risk Studies*, *8*(3), 1–10. https://doi.org/10.4102/jamba.v8i3.255

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401–415. https://doi.org/10.1007/BF02291817

King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, *44*(2), 347–361. https://doi.org/10.2307/2669316

Lindell, M. K., & Prater, C. S. (2003). Assessing Community Impacts of Natural Disasters. *Natural Hazards Review*. https://doi.org/10.1061/(ASCE)1527-6988(2003)4:4(176)

McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *42*(2), 109–142.

Mechler, R. (2005). *Cost-benefit Analysis of Natural Disaster Risk Management in Developing Countries*. Retrieved from https://pdfs.semanticscholar.org/1870/e6d76e9515eed7d53c2904b48425943b832b.pdf

Mercer, J., Kelman, I., Suchet-Pearson, S., & Lloyd, K. (2009). Knowledge bases for disaster risk reduction in Papua new guinea. *Geografiska Annaler: Series B, Human Geography*, *91*(2), 157–183. https://doi.org/10.1111/j.1468-0467.2009.00312.x

Morrissey, J., & Oliver-Smith, A. (2013). *Perspectives on Non-economic Loss and Damage – Understanding values at risk from climate change*. Loss and Damage Series, United Nations University Institute for Environment and Human Security Bon. Retrieved from http://loss-and-damage.net/download/7213.pdf+

Musyoki, A., Thifhulufhelwi, R., Murungweni, F. M., Sciences, G., Africa, S., Africa, S., … Musyoki, A. (2016). The impact of and responses to flooding in Thulamela Municipality, Limpopo Province, South Africa. *Jàmbá: Journal of Disaster Risk Studies*, *8*(2), 1–10. https://doi.org/10.4102/jamba.v8i2.166

Neal, R. A., Boyle, P., Grahame, N., Mylne, K., & Sharpe, M. (2014). Ensemble based first guess support towards a risk-based severe weather warning service. *Meteorological Applications*. https://doi.org/10.1002/met.1377

Nunally, J. (1978). *Psychometric theory*. New York: McGraw-Hill.

Osborne, J. W., & Costello, A. B. (2009). Best practices in exploratory Factor Analysis: four recommendations for getting the most from your analysis.

Pauw, K., Thurlow, J., & van Seventer, D. (2010). Droughts and Floods in Malawi: Assessing the Economywide Effects. *IFPRI Discussion Paper*, *962*(April). Retrieved from http://www.ifpri.org/publication/droughts-and-floods-malawi?print

Pelling, M. (1997). What Determines Vulnerability To Floods; A Case Study In Georgetown, Guyana. *Environment & Urbanization*, *9*(1), 203–226. https://doi.org/10.1177/095624789700900116

Quarantelli, E. L. (1998). *What is a disaster?* Routledge, New York.

Rogers, D. P., & Tsirkunov, V. V. (2013). *Weather and Climate Resilience: Effective Preparedness through National Meteorological and Hyrological Services,*. Washington, D.C.

Rosen, R. (1991). *Life Itself: A Comprehensive Inquiry into Nature, Origin, and Fabrication of Life*. Columbia University Press.

Rstudio Team. (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. Retrieved from http://www.rstudio.com/

Rudari, R., Beckers, J., De Angeli, S., Rossi, L., & Trasforini, E. (2016). Impact of modelling scale on probabilistic flood risk assessment: the Malawi case. *E3S Web of Conferences*. https://doi.org/10.1051/e3sconf/20160704015

Schafer, J. L., & Olsen, M. K. (2010). Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behavioral Research*, *33*(4), 545–571. https://doi.org/10.1207/s15327906mbr3304_5

Schneiderbauer, S., & Ehrlich, D. (2004). *Risk, Hazard and People's Vulnerability to Natural Hazards: A Review of Definitions, Concepts and Data, Brussels: European Commission–Joint Research Centre*. Brussels: European Commission–Joint Research Centre.

Sharpe, A. (2004). *Literature Review of Frameworks for Macro-indicators,*.

Suarez, P., & Mendler de Suarez, J. (n.d.). Linking machine learning and rapid financing for flood preparedness: Red Cross innovations to manage changing climate risks in Togo, 1–36. https://doi.org/10.1016/B978-0-12-802121-7.01601-0

Suarez, P., & Tall, A. (2010). Towards forecast-based humanitarian decisions: Climate science to get from early warning to early action. *Humanitarian Futures Programme*, 10. Retrieved from http://www.alnap.org/resource/12320

Turner, B. L., Kasperson, R. E., Matson, P. A., Mccarthy, J. J., Corell, R. W., Christensen, L., … Schiller, A. (2003). A framework for vulnerability analysis in sustainability science. *Elements*, *100*(14).

UNISDR. (2009). 2009 UNISDR Terminology on Disaster Risk Reduction. *International Stratergy for Disaster Reduction (ISDR)*, 1–30. https://doi.org/978-600-6937-11-3

United Nations General Assembly. (2016). Report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction. A/71/644, *21184*(December), 1–41.

van den Homberg, M., Monné, R., & Spruit, M. (2017). Bridging the information gap of disaster responders by optimizing data selection using cost and quality, accepted for publication, Computers & Geosciences, Special Issue on Big data and natural disasters: New approaches for spatial and temporal massive da. *Computers & Geosciences*. https://doi.org/10.1016/j.cageo.2018.06.002

Van Den Homberg, M., & Plaatsman, T. (2018). 510 - Framework for data curation.

Van Den Homberg, M., & Visser, J. (2018). *510 – Impact-based forecasting methodology presentation*.

Van Den Homberg, M., Visser, J., & Van Der Veen, M. (2017). Unpacking Data Preparedness from a humanitarian decision making perspective Unpacking Data Preparedness from a humanitarian decision making perspective : toward an assessment framework at subnational level.

Van der Geest, K., & Schindler, M. (2017). *Handbook for assessing loss and damage in vulnerable communities*.

Vogel, C., & O'Brien, K. (2004). *Vulnerability and Global Environmental Change: Rhetoric and Reality*. *Information Bulletin on Global Environmental Change and Human Security*.

Wannewitz, S., Hagenlocher, M., & Garschagen, M. (2016). Development and Validation of a Sub-national Multi-hazard Risk Index for the Philippines. *Journal for Geographic Information Science*, *1*, 133–140. https://doi.org/10.1553/giscience2016_01_s133

WBGU. (2005). *World in Transition: Fighting Poverty through Environmental Policy*. London: Earthscan.

Wilbrink, J. G. (2017). *Remoteness as a proxy for social vulnerability in Malawian Traditional Authorities*. TU Delft.

Wilkinson, E., Weingärtner, L., Choularton, R., Bailey, M., Todd, M., & Kniveton, D. (2018). Forecasting hazards , averting disasters early action at scale, (March).

Wisner, B., Blackie, P., Cannon, T., & Davis, I. (2004). *At Risk: Natural Hazards, People's Vulnerability and Disasters*. Routledge, London and New York.

Zulu, L. (2017). *Existing Research and Knowledge on Impacts of Climate Variability and Change on Agriculture and Communities in Malawi*. *GCFSI Publication series*. Retrieved from file:///C:/Users/Jan/Documents/510/Bronnen/Chirstina/een.pdf

8. APPENDIX

Table 8.1 – Data indicators and sources accessed for calculating vulnerability on TA level

Variable name	Survey	Explanation variable / characteristics	N (ind/HH)
Child_5	DHS	Percentage of children at the age of 5 and under per household	120492 Ind
Radio	DHS	Percentage of the HH in a TA that lives without a radio	26361 HH
TV	DHS	Percentage of the HH in a TA that lives without a tv	26361 HH
Bike	DHS	Percentage of the HH in a TA that lives without a bicycle	26361 HH
Floor_nat	DHS	Percentage of the HH in a TA that lives in a house with a natural floor (i.e. made of sand/earth/dung)	26361 HH
Roof_nat	DHS	Percentage of the HH in a TA that lives in a house with a natural roof (i.e. no roof/thatch/palmleaf/sod)	26361 HH
Water_wo	DHS	Percentage of the HH in a TA where in the past two weeks, water was not available for at least one full day.	26361 HH
Dwell_age	IHS	Average age of the dwellings	12447 HH
Dwell_trad	IHS	Percentage of traditional dwellings in the TA	12447 HH
Dwell_elec	IHS	Percentage of dwellings with electricity in the TA	12447 HH
Mobile	IHS	Average amount of phones per HH in the TA	12447 HH
Mosq_net	IHS	Percentage of HHs that sleep under a Mosquito net	12447 HH
Bank_acc	IHS	Percentage of the HH in a TA that has a bank account	12447 HH
Food_wor	IHS	Percentage of the HH in a TA that is worried that the HH would not have enough food	12447 HH
Food_def	IHS	Percentage of the HH in a TA that was faced with a situation when they didn't have enough food to feed the HH in the past 12 months	12447 HH
Health_ade	IHS	Percentage of the HH in a TA that considered the standard of health care they received for HH members less than adequate for household needs	12447 HH
House_ade	IHS	Percentage of the HH in a TA that considered housing less than adequate for households needs	12447 HH
Food_ade	IHS	Percentage of the HH in a TA that considered HH's food consumption (over the past months) less than adequate for household needs	12447 HH
Cloth_ade	IHS	Percentage of the HH in a TA that considered household's clothing less than adequate for households needs	12447 HH
Inc_ade	IHS	Percentage of the HH in a TA that considered household's income less than adequate for households needs	12447 HH
Father_wo	IHS	Percentage of children (until 16 years old) where father is not part of the Household	26737 Ind
Mother_wo	IHS	Percentage of children (until 16 years old) where mother is not part of the Household	26737 Ind
School_none	IHS	Percentage of people (above 15 years old) that have no school degree	28585 Ind
School_prim	IHS	Percentage of people (above 15 years old) that only has a primary school degree	28585 Ind
School_sec	IHS	Percentage of people (above 15 years old) that only has a secundary school degree	28585 Ind
Chron_ill	IHS	Percentage of people with a chronic Illness	53873 Ind
WI	DHS	Wealth Index	26361 HH
Pov_l	RCMRD	Poverty level in each TA where a high level is likely to be associated with high sensitivity to climate hazards	368 TAs
Pop_dens	RCMRD	Population density per TA.	368 TAs
Mal_sus	RCMRD	Malaria susceptibility	368 TAs
Inf_mort	RCMRD	Infant Mortality	368 TAs
Fe_head	RCMRD	Female headed household;	368 TAs
Sens_sc	RCMRD	Sensitivity score	368 TAs
Mark_acc	RCMRD	Market accessibility	368 TAs
Literacy_l	RCMRD	Literacy levels	368 TAs
Edu_mother	RCMRD	Education level Mother	368 TAs
House_dens	RCMRD	House Density	368 TAs
Road_dens	RCMRD	Road density	368 TAs
TT_PS	Wilbrink	Average Travel time in minutes to primary school	9149 places
TT_SS	Wilbrink	Average Travel time in minutes to secondary school	9149 places
TT_TC	Wilbrink	Average Travel time in minutes to Trading Center	9149 places
TT_C	Wilbrink	Average Travel time in minutes to City	9149 places

TT_WP	Wilbrink	Average Travel time in minutes to water points	9149 places
TT_H	Wilbrink	Average Travel time in minutes to hospital	9149 places

Table 8.2 – VIF score for each variable

Variable Name	VIF before removing Var	VIF after removing Var	Variable Name	VIF before removing Var	VIF after removing Var
Child_5	2.4	1.8	School_none	176.7	removed
Radio	10.3	3.4	School_prim	58.0	removed
TV	17.2	removed	School_sec	55.7	removed
Bike	2.8	1.7	Chron_ill	1.6	1.3
Mosq_net	1.4	1.2	WI	56.7	removed
Floor_nat	40.2	removed	Pov_l	10.2	removed
Roof_nat	23.8	removed	Pop_dens	3.6	2.2
Water_wo	3.3	2.4	Mal_sus	5.5	3.0
Dwell_age	2.3	1.9	Inf_mort	5.7	3.0
Dwell_trad	4.6	3.5	Fe_head	3.1	2.2
Dwell_elec	11.7	removed	Sens_sc	16.8	removed
Mobile	11.9	removed	Mark_acc	4.1	3.5
Bank_acc	6.2	3.1	Literacy_l	4.7	3.2
Food_wor	1.9	1.8	Edu_mother	3.0	2.3
Food_def	3.0	2.4	House_dens	4.1	3.5
Health_ade	3.1	2.7	Road_dens	5.5	5.1
House_ade	3.0	2.7	TT_Cities	4.0	3.7
Food_ade	3.8	3.7	TT_WP	1.9	1.8
Cloth_ade	4.5	4.2	TT_TC	4.6	4.2
Inc_ade	1.5	1.4	TT_H	4.2	3.9

Table 8.3 – Descriptive statistics

Variable name	N	Missing values	MEAN	SD	Median	Min	Max	Skew	Kurtosis	SE
Child_5	257	46	0.18	0.03	0.18	0.10	0.28	0.22	0.01	0.00
Radio	257	46	0.56	0.16	0.59	0.08	0.86	-0.88	0.14	0.01
TV	257	46	0.84	0.19	0.93	0.03	1.00	-1.53	1.89	0.01
Bike	257	46	0.61	0.15	0.60	0.17	0.97	0.23	-0.18	0.01
Mosq_net	282	21	0.86	0.10	0.88	0.44	1.00	-1.21	1.92	0.01
Floor_nat	257	46	0.67	0.27	0.79	0.00	1.00	-0.89	-0.52	0.02
Roof_nat	257	46	0.48	0.27	0.53	0.00	0.94	-0.40	-1.05	0.02
Water_wo	257	46	0.23	0.21	0.17	0.00	0.90	1.27	0.98	0.01
Dwell_age	282	21	8.46	4.11	7.51	0.00	40.00	3.02	15.54	0.24
Dwell_trad	282	21	0.31	0.23	0.30	0.00	0.94	0.34	-0.86	0.01
Dwell_elec	282	21	0.18	0.27	0.03	0.00	1.00	1.52	1.15	0.02
Mobile	282	21	0.95	0.70	0.69	0.02	4.25	1.43	2.02	0.04
Bank_acc	282	21	0.28	0.19	0.23	0.00	1.00	1.08	0.76	0.01
Food_wor	282	21	0.62	0.11	0.63	0.25	0.94	-0.40	0.41	0.01
Food_def	282	21	0.70	0.12	0.70	0.31	1.00	-0.43	0.28	0.01
Health_ade	282	21	0.49	0.12	0.50	0.19	0.81	-0.02	0.20	0.01
House_ade	282	21	0.54	0.12	0.56	0.19	0.81	-0.23	0.22	0.01
Food_ade	282	21	0.62	0.11	0.63	0.25	1.00	-0.31	0.59	0.01
Cloth_ade	282	21	0.67	0.12	0.69	0.19	0.94	-0.70	1.00	0.01
Inc_ade	282	21	0.46	0.12	0.46	0.00	0.81	0.17	1.30	0.01
Father_wo	282	21	0.37	0.13	0.37	0.00	0.75	0.13	-0.01	0.01
Mother_wo	282	21	0.17	0.09	0.16	0.00	0.53	1.18	2.11	0.01
School_none	282	21	0.88	0.14	0.94	0.28	1.00	-1.80	3.00	0.01
School_prim	282	21	0.08	0.08	0.05	0.00	0.45	1.43	1.89	0.00
School_sec	282	21	0.03	0.07	0.00	0.00	0.47	3.68	17.11	0.00
Chron_ill	282	21	0.07	0.03	0.07	0.00	0.19	0.48	0.39	0.00
WI	257	46	3.19	0.91	2.85	1.84	5.00	0.62	-1.00	0.06
Pov_l	303	0	75.04	20.07	82.60	0.00	97.27	-1.60	1.86	1.15
Pop_dens	303	0	2.05	4.18	0.52	0.00	31.83	4.08	20.60	0.24
Mal_sus	303	0	53.81	20.23	52.55	0.00	93.71	-0.03	-0.56	1.16
Inf_mort	303	0	80.34	12.01	84.11	0.00	100.00	-2.18	11.30	0.69
Fe_head	303	0	60.29	14.63	59.06	0.00	100.00	0.13	1.38	0.84
Sens_sc	303	0	71.01	11.20	72.85	0.00	88.75	-1.83	8.76	0.64
Mark_acc	303	0	23.45	19.38	22.57	0.00	95.60	0.88	0.81	1.11
Literacy_l	303	0	23.28	10.47	24.39	0.00	50.43	-0.18	-0.45	0.60
Edu_mother	303	0	58.13	35.52	71.66	0.00	100.00	-0.42	-1.44	2.04
House_dens	303	0	135.38	334.17	1.88	0.00	1951.04	3.28	11.15	19.20
Road_dens	303	0	2684.62	3607.79	941.29	0.00	21618.28	1.99	4.05	207.26
TT_Cities	293	10	141.97	109.72	113.00	0.00	500.00	1.02	0.52	6.41
TT_WP	293	10	100.17	75.70	79.49	0.00	399.00	1.80	3.32	4.42
TT_TC	293	10	74.66	45.63	70.54	0.00	314.86	1.38	4.05	2.67
TT_H	293	10	145.65	109.67	128.69	0.00	498.67	0.93	0.33	6.41
TT_PS	293	10	66.32	60.10	49.70	0.00	377.00	2.33	6.71	3.51

TT_SS	293	10	77.86	64.01	68.21	0.00	391.00	1.70	4.33	3.74

9. ANNEXES

Table 9.1 – Vulnerability indicators according to INFORM index

Definition vulnerability: the intrinsic predispositions of an exposed population to be affected, or to be susceptible to the damaging effects of a hazard, even though the assessment is made through hazard independent indicators		
Category 1		
Socio-economic vulnerability: (in)ability of individuals or households to afford safe and resilient livelihood conditions and well-being.		
Components	Definition	Variable
Development & Deprivation	how a population is doing on average	Life expectancy, educational attainment, income, living standards, health, and education, average number of poor people and deprivations with which poor households contend.
Inequality	the dispersion of conditions within population	Income-, education- and health inequality, differences in the distribution of achievements between men and women
Aid Dependency	the countries that lack sustainability in development growth due to economic instability and humanitarian crisis	Total ODA (official Development Assistance) in the last two years per capita OECD, Global Humanitarian Funding per capita, Net ODA Received in percentage of GD
Category 2		
Vulnerable groups: the population within a country that has specific characteristics that make it at a higher risk of needing humanitarian assistance than others or being excluded from financial and social services		
Components	Definition	Variable
Uprooted people	People that are not a part of the society as well as the social system, only partially supported by the community	Number of refugees, returned refugees, Internally Displaced Persons
Other Vulnerable Groups / Health Conditions	people in weak health conditions	Prevalence of HIV-AIDS above , Tuberculosis prevalence, Malaria Mortality Rate
Other Vulnerable Groups / Children under-5	the health condition of children	Children under weight, children mortality
Other Vulnerable Groups / Recent Shocks	people affected by natural disasters in the past 3 years	Relative number of affected population by natural disasters in the last three years
Other Vulnerable Groups / Food Security	"A situation that exists when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life".	Prevalence of Undernourishment, Average Dietary Energy Supply Adequacy, Domestic Food Price Level Index, Domestic Food Price Volatility Index

Table 9.2 – Vulnerability indicators according to the MOVE Framework of Birkmann et al. (2013)

Dimension	Definition
Vulnerability	the propensity of exposed elements such as physical or capital assets, as well as human beings and their livelihoods, to experience harm and suffer damage and loss when impacted by single or compound hazard events
Social Vulnerability	propensity for human well-being to be damaged by disruption to individual (mental and physical health) and collective (health, education services, etc.) social systems and their characteristics (e.g. gender, marginalization of social groups).
Economic Vulnerability	propensity for loss of economic value from damage to physical assets and/or disruption of productive capacity.
Physical Vulnerability	potential for damage to physical assets including built-up areas, infrastructure and open spaces
Cultural Vulnerability	potential for damage to intangible values including meanings placed on artefacts, customs, habitual practices and natural or urban landscapes.
Environmental Vulnerability	potential for damage to all ecological and bio-physical systems and their different functions. This includes particular ecosystem functions and environmental services (see, e.g., Renaud 2006) but excludes cultural values that might be attributed.
Institutional Vulnerability	potential for damage to governance systems, organizational form and function as well as guiding formal/legal and informal/customary rules—any of which may be forced to change the following weaknesses exposed by disaster and response.

* The vulnerability indicators of Birkmann et al. (2013) are based on specific hazards. However, this vulnerability composite index will be based on hazard independent indicators. The type of hazard will be measured separately in the Hazard & Exposure explanatory variable. The majority of assets and systems exposed to hazard will exhibit more than one dimension of vulnerability.

Table 9.3 – Vulnerability indicators according to the study of Wannewitz et al. (2016)

Vulnerability (Susceptability)	Vulnerability (Lack of Coping Capacity)	Vulnerability (Lack of adaptive capacity)
Population density (per km2)	Good governance index (%)	Illiteracy (%)
population without access to sanitation	Road density: primary, secondary, tertiary roads (per km2)	Protected areas
population without access to safe drinking water	Population without a car/motorcycle (%)	Forest cover change (%)
population without access to electicity	Population without a boat (%)	Density of aid projects (per km2)
population living in informal settlements	Shelter density: primary/secondary schools (per km2)	
population living in poorly constructed houses	Density of emergency services (per km2)	
stunting in children under 5 (per km2)	Households without a mobile phone (%)	
dependenccy ratio	Households without access to the internet (%)	
dependency on agriculture	Households without access to media (radio/TV) (%)	
Unemployment (%)		
poverty as measured by the Wealth Index %		
GDP per capita ($)		
conflict density km2		