universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

„Hospital-based spatial analysis of inpatient characteristics, after the 2008 Wenchuan earthquake“

Verfasser

Thomas Jakubicka

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat.)

Wien, 2012

Studienkennzahl lt. Studienblatt:		A 453
Studienrichtung lt. Studienblatt:		Theoretische und Angewandte Geographie
Betreuerin / Betreuer:		Univ.-Prof. DI Dr. Wolfgang Kainz

ABSTRACT

This thesis presents the spatial analysis of the Deyang CRED dataset (DCD) after the 2008 Wenchuan earthquake. The dataset was collected retrospectively from the hospital records. The dataset contains 1878 individual patient records that are connected to the earthquake.

First of all the dataset had to be geocoded according to the placenames. Then the spatial distribution of the patients and the two selected indicators, the severity of injury, measured with „International ICD-based Injury Severity Score" (ICISS) and the time to reach the hospital, could be analysed. The Methods selected for this analysis were simple descriptive methods, as well as more complex interpolation techniques and statistical tests.

The goal of the thesis is to investigate the spatial patterns of the dataset and characterize possible correlations. Furthermore the dataset is to be described spatially, to illustrate the possibilities of spatial analyses and cartographic visualization for the research of the human impact of disasters.

The analysis shows the very differential and uneven distribution of the patients that were admitted to the hospital. Not only were most patients from Mianzhu Shi county in the Northeast of the study area, but additionally there were a couple of locations, from which the number of patients admitted was above the average. Furthermore the analysis highlights that severely injured patients were admitted to the hospital fast. The thesis also points out the limitations that came to the fore and gives clues for their solution.

ZUSAMMENFASSUNG

In der vorliegenden Arbeit wurde der Deyang CRED Datensatz (DCD) zum Wenchuan Erdbeben, aus dem Jahr 2008, einer räumlichen Analyse unterzogen. Der Datensatz wurde retrospektiv aus den Krankenhaus Registern erstellt. Er beinhaltet 1878 individuelle Patienten Einträge, die alle im Zusammenhang mit dem Erdbeben stehen.

Zunächst musste der Datensatz anhand der Ortsangaben geokodiert werden. Anschließend konnte die räumliche Verteilung der Patienten und der zwei für die Analyse als relevant erachteten Indikatoren, die schwere der Verletzung, gemessen mit dem „International ICD-based Injury Severity Score" (ICISS) und die Zeit bis zur Aufnahme im Krankenhaus, untersucht werden. Zu diesem Zweck wurden sowohl einfache deskriptive Methoden als auch komplexere Interpolationsverfahren und statistische Tests angewandt.

Ziel dieser Arbeit ist es den Datensatz auf räumliche Verteilungsmuster zu untersuchen und mögliche Zusammenhänge zu beschreiben. Außerdem soll der Datensatz räumlich beschrieben werden und so die Anwendungsmöglichkeiten der räumlichen Analyse und der kartographischen Auswertung, für die Erforschung der Auswirkungen von Katastrophen auf den Menschen dargestellt werden.

Die Untersuchungen in dieser Arbeit haben die sehr unterschiedliche und ungleiche Verteilung der im Krankenhaus aufgenommenen Patienten gezeigt. So kommen die meisten Patienten nicht nur aus dem Bezirk Mianzhu Shi im Nord-Osten des Untersuchungsgebietes, es gibt auch noch zusätzlich einige Orte in diesem Bezirk, aus welchen überdurchschnittlich viele Patienten aufgenommen wurden. Die Analysen haben auch gezeigt, dass schwer verletzte Patienten schnell im Krankenhaus aufgenommen wurden. Die Arbeit zeigt auch die Einschränkungen, die aufgetreten sind auf und gibt Anhaltspunkte wie diese gelöst werden könnten.

ACKNOWLEDGEMENTS

I would like to thank José Rodriguez-Llanes, for his guidance and support throughout the process of writing this thesis. Further I want to thank Prof. Wolfgang Kainz for his feedback and input to this thesis, the People's Hospital of Deyang City research team for their assistance and Prof. Debarati Guha-Sapir for enabling this project.
Lastly I want to offer my gratitude and regards to my dear friends and family who supported me all these years, in particular my brother, my sister and my parents.

TABLE OF CONTENTS

List of Abbreviations

CRED	Center for Research on the Epidemiology of Disasters
DCD	Deyang CRED Dataset
EM-DAT	International Disaster Database
FAO	Food and Agriculture Organization
GAUL	Global Administrative Unit Layer dataset
ICD	International Classification of Disease
ICISS	ICD-based Injury Severity Score
IDW	Inverse distance weighted interpolation
MMI	Modified Mercali Intensity
MMS	Moment Magnitude Scale
PGA	Peak Ground Acceleration
USGS	United States Geological Survey

List of Figures, Maps and Tables

Maps:

Tables:

Figures:

PREFACE

Beforehand I would like to explain the formation for this thesis. It started with my internship at the Centre for Research on the Epidemiology of Disasters (CRED), in Brussels. This was the first time I got involved with the matter of the human impact of disasters. Through the internship I got interested in the possibilities of GIS analysis in this research field and also for the first time discussed the possibility of a thesis with Prof. Debarati Guha-Sapir, the director of the CRED. After my internship I stayed in touch with the CRED, through my contribution to the MICRODIS project and thus have been able to expand my interest and experience in this field.

The idea for the following thesis is the result of my discussions with Prof. Debarati Guha-Sapir and Jose Rodriguez LIanes after the matter of writing this thesis materialized. At this time the Deyang Hospital Database, was made available at CRED and was ready to be investigated. Hence the spatial analysis part of the investigation was chosen as topic for this thesis.

The extensive discussions with José Rodriguez LIanes and his guidance during the research process were of essential help, for the realization of this thesis. I furthermore would like to point out that this thesis would not have been possible without the data access and support through the CRED, for which I am deeply thankful.

1. Introduction

The distribution of spatial phenomena is an important part of public health research. One of the most popular examples showing the connection of public health issues and spatial dimensions is that described by Dr. John Snow, who identified the source of a cholera outbreak, in London in 1854, at the water pump in Broad Street, by mapping the cholera deaths on a dot map. The tradition says, that by removal of the handle of the water pump, the outbreak was stopped (Mc Leod, 2000).

This very illustrative example shows how the basic questions of geography: Where? Why there? And therefore what? Are important to public health. Medical and health geography is an own discipline since a long time now and offers a wide scale, of well developed possibilities to asses the connections of the environment with the human condition (Meade & Emch, 2010).

The human impact of disasters is one particular part in the research field of epidemiology. The Centre for Research on the Epidemiology of Disasters (CRED), in Brussels, focuses its research in particular on this human impact. Within the MICRODIS project, funded by the European Commission and carried out by the CRED and its partners, the Deyang CRED Dataset (DCD) was developed. The goal of the MICRODIS project was to assess the health, economic and social impact of disasters at the micro level (selected case studies on floods, earthquakes, volcanic activity and cyclones in Europe and Asia). The DCD was collected in the context of the 2008 Wenchuan earthquake. In this thesis the spatial analyses of the DCD are presented.

1.1. Goals

To understand the geography of a catastrophe, such as the Wenchuan earthquake, is important to be able to direct all the different measures that are part of the “disaster management cycle”. The data used in this thesis provides the relatively rare possibility, of locating the human impact geographically on individual level.

The aim of the thesis is to present possibilities of spatial visualization and analysis of such data, to better understand the human impact and evacuation activities after the 2008 Wenchuan earthquake in particular, and for earthquakes and other disasters in general. Furthermore the intention is to reveal spatial patterns in the data. Thus the main research question for this thesis is:

"Are there spatial patterns in the distribution of the patients and the patient-variables, admitted to the Deyang hospital?"

One reason for the large human impact of earthquakes is their high unpredictability. This is why preparedness is a key issue in earthquake prone areas (CHAN, GAO, & GRIFFITHS, 2010). Hence it is important to understand the causalities and reasons of the human impact of these events. To strengthen this understanding is the overall goal of research, like the one undertaken in this thesis.

1.2. STATE OF THE ART

The research is mainly focused on the physical or economical impacts of earthquakes while the human impact or social or individual vulnerability is considered less frequently. However to be able to asses the general vulnerability of a place, the integration of all these factors is necessary (SCHMIDTLEIN, SHAFER, BERRY, & CUTTER, 2011).

Previous studies have stressed the importance of the initial location during earthquakes, in causing death and injuries, as it determines the environment of the victim that has a great influence on the individual risk of injury and death (ARMENIAN, MELKONIAN, NOJI, & HOVANESIAN, 1997; LIAO ET AL., 2003). Yet there are few studies that use a GIS approach to analyse this fact. On the other hand the use of GIS in epidemiology, to identify the patterns of injuries has been recognized, especially for the direction of services (SCHUURMAN, HAMEED, FIEDLER, BELL, & SIMONS, 2008).

In their study of the "GIS mapping of earthquake-related deaths and hospital admissions from the 1994 Northridge, California, earthquake" (PEEK-ASA, RAMIREZ, SHOAF, SELIGSON, & KRAUS, 2000), the authors show the possibilities that the GIS analysis of individual point data offers for such events. The study is able to highlight several relationships between injury and environmental factors. However, as stated by the authors, the small dataset (171 records) could have been biased through the inability of geocoding 37 of the records. Anyhow the study shows the relevance of such research for the preparedness of earthquake prone regions as well as for the immediate emergency situation:

"Data which can predict the boundaries for severe injuries can facilitate search and rescue efforts and help hospitals prepare for patient treatment. Injury configurations from the Northridge Earthquake indicate that although fatal and severe injuries were concentrated near the epicenter, the impacted area was quite large and did not radiate from the epicenter in concentric patterns. Agencies responding to human needs following disasters must be prepared to provide services to an area much larger than predicted by high earthquake activity […] this study has shown that low median distance, high average Mercalli Index, and high average PGA, and areas with high damage to buildings are related to more severe injury, injuries caused by falling building parts, and injuries involving structural damage. Even with this information, it is difficult to predict with certainty all outcomes and causes of injury solely on these environmental characteristics."(PEEK-ASA ET AL., 2000).

Even though the importance of such research has been recognized (ALEXANDER, 2000), little research since then has been conducted in this direction. However the event discussed in this thesis and the most recent events in Haiti and Japan have triggered some studies that will be discussed in Chapter 8.2.

The most important reason for this lack of research is probably the availability of the relevant data or to collect this sort of data, especially the medical data. It becomes even more difficult when very large areas are affected, like it was the case in the Wenchuan earthquake. Many different institutions collect data, each using their own methods. There are many issues connected to such spatially and individually low or not aggregated data, as the most important data privacy has to be named. Even though such data usually is anonymised, one has to be aware that inferences are still possible and therefore patient confidentiality and privacy is often rated higher than the research (BROWN, MCLAFFERTY, & MOON, 2009). Also the collection of such position accurate data may be difficult, depending on the disaster situation and the environment in which the disaster occurred. Especially during the relief efforts collecting information may not be the priority. Thus studies often have to establish their datasets retrospectively, which as well poses issues especially regarding the location information (ZOLALA, 2010).

The data presented in this thesis is such individual point data. With 1878 records the database offers a large basis and furthermore enables the possibility of refining the data selection criteria within the database, without losing a large percentage of the information. Nonetheless there are also limitations in this dataset, which will be discussed later.

1.3. STRUCTURE OF THE THESIS

The first part of the thesis is meant to give an introduction to the study area, the event of the Wenchuan earthquake and the used data (chapters 2-5). Especially the DCD is presented extensively in an own chapter. The process of geocoding the dataset is as well described here, as I did not include it in the methods chapter, because I regarded it as preliminary work. In chapter 6 the methods that were applied are presented. The methods represent a variety of GIS domains, simple and complex, which can be summarized by the term of “spatial analysis”. Thus descriptive mapping of the data was applied, as well as spatial interpolation or statistical testing. Chapter 7 presents the result maps and their interpretation together with a summary of the results in the last section of the chapter. In chapter 8 the results, the limitations and ways forward are discussed, while in the last chapter a conclusion to the thesis is given. Furthermore a list of all locations is provided in the Annex.

2. GEOGRAPHIC DESCRIPTION OF THE STUDY AREA

The study area is determined by the boundaries of Deyang prefecture, as apart of a few exceptions, this was the origin of the patients in the dataset (see chapter 4). Deyang Prefecture is situated in the Sichuan province in central China (see Map 1). The Province is divided into 18 prefecture level cities and 3 prefectures. Sichuan is also the second largest province in China (without considering the autonomous regions) and the capital Chengdu is situated close to Deyang prefecture.

Map 1 Overview of Sichuan Province and Deyang Prefecture

2.1. GEOLOGY AND GEOMORPHOLOGY

Sichuan is framed by mountains and can be divided into three parts. While the Himalayas with the Tibetan high plain dominate a large part of western Sichuan the Sichuan basin and the mountains in the southwest form the other parts. The Longmenshan fault belt extends in the northwest of the Sichuan basin from southwest to northeast and is regarded as a typical intercontinental orogenic belt. Together with the Minshan fault belt it is an important factor in the South- North Seismic Belt. It is also the locus for several historic earthquakes (WANG & MENG, 2009).

The Deyang prefecture is located in the Sichuan basin. The basin covers an area of 160.000 km² and is one of the largest basin landscapes in China. It is irrigated by several rivers from the surrounding mountains and is draining into the Yangtze River, which crosses the plain in the south (STAIGER, FRIEDRICH, & SCHÜTTE, 2006).

2.2. CLIMATE

The climate in Sichuan is differing very much, reflecting the variable landscape. In the Sichuan basin the climate can be described as subtropical. The average temperature in January is 5,5°C and 25,6°C in July. The yearly rainfall is usually above 750mm often even above 1000mm. The subtropical climate often enables two crops a year (STAIGER ET AL., 2006).

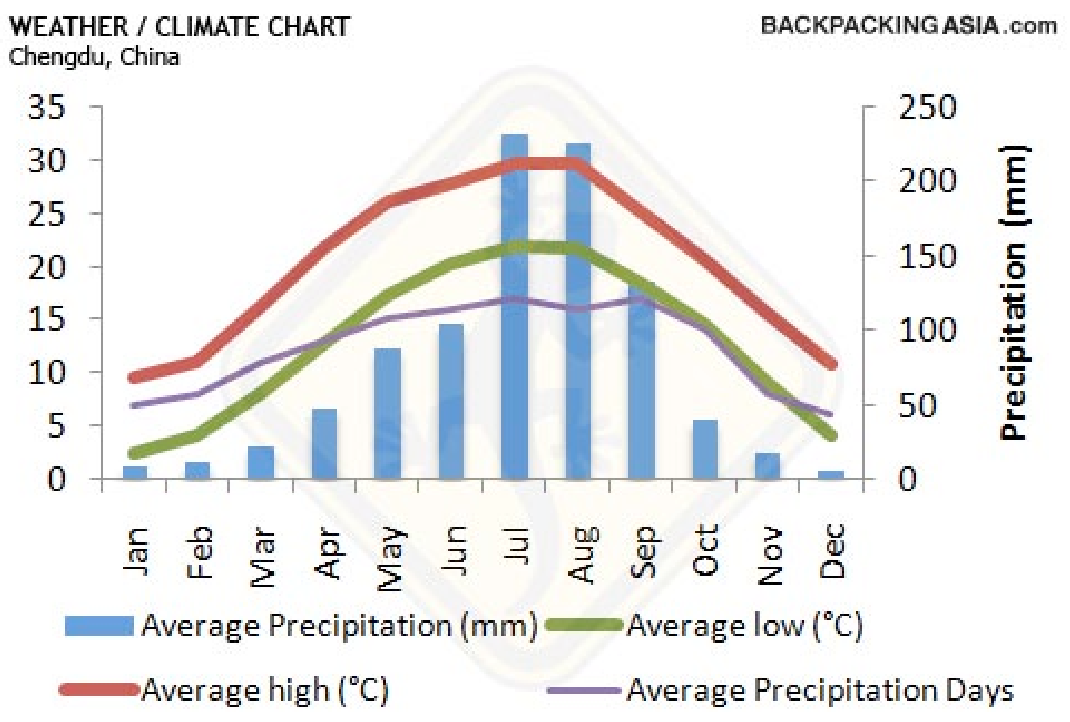


Figure 1 Climate chart of Chengdu; source: www.backpackingasia.com

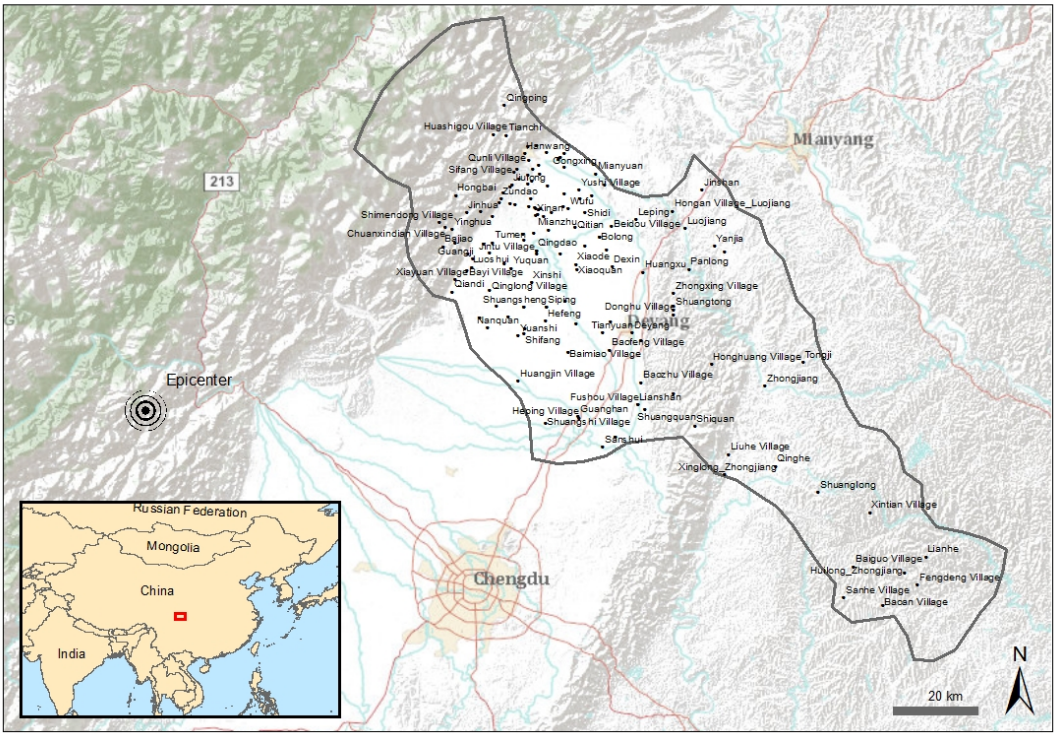
2.3. POPULATION AND ECONOMY

According to the LandScan database Sichuan province has a population of approximately 84 Mio and Deyang Prefecture has a population of 3,8 Mio. There are several Minority groups living in Sichuan, predominately in the mountainous regions. Altogether their number was around 2,5 Mio people at 1995 (STAIGER ET AL., 2006). The major part of the population works in the primary sector (63%). 16% of the employees work in the secondary and 21% in the tertiary sector. The Sichuan basin is often called the breadbasket of the nation, the favourable climate and the good soil conditions being the main reason. Thus especially in this part of Sichuan agriculture is very important. Facilitated by the water availability irrigation farming is the main form of agriculture in the study area. However Sichuan has also been industrialised in the 1960s and 70s, making it one of the most important industry regions in China. With the primary sector remaining strong the gross production value in 1995 equalled 282,6 Billion Yuan.

3. THE 2008 WENCHUAN EARTHQUAKE

3.1. THE WENCHUAN EARTHQUAKE DISASTER

On the 12th May 2008 at 14:28 CST an earthquake with the magnitude of 7,9 on the moment magnitude scale (MMS, Mw) struck in the province of Sichuan. The epicentre was located in Wenchuan county, at 30,986°N and 103,364°E, with a focal depth of 19km (Map 2). Thus the earthquake is known as the Wenchuan earthquake.



Map 2 Location of the epicentre, the Deyang prefecture, and all locations in the DCD

According to The International Disaster Database (EM-DAT) there were 87.476 people killed, 366.596 injured and more than 45 million people affected. Approximately 15 million people were evacuated after the earthquake. 5 million buildings are believed to have collapsed and more than 21 million have been damaged. The economic losses are estimated at US$ 85 billion. This was the deadliest earthquake since the 1976 Tangshan earthquake, where more than 250.000 people were killed and altogether one of the disastrous earthquakes in Chinese history (CRED, 2011; THOMSON REUTERS FOUNDATION, 2011; USGS, 2011).

There were several aftershocks, which caused more casualties and damage, together with other secondary hazards that were triggered by the earthquake, i.e. landslides, rockfalls or barrier lakes. The infrastructure in the region was heavily damaged, especially many roads were blocked or damaged, power supply and telecommunication interrupted. The situation was even worsened by the heavy rainfalls at that time, as many people were homeless, rivers blocked and landslides were a high threat (THOMSON REUTERS FOUNDATION, 2011; USGS, 2011).

Earthquakes are a common threat in China. In EM-DAT it is the country with the most (98) earthquakes reported since 1980. Earthquakes are very unpredictable and they are the disaster type, which causes the most casualties and the highest economic losses (CRED, 2011). China's response to the Earthquake was regarded as very swift and open by western commentators. It was also the first time that China called for foreign help in a disaster situation. The relief operation was very well organised, due to China's experience with earthquakes and large resources (THOMSON REUTERS FOUNDATION, 2011; USGS, 2011).

3.2. INTRODUCTION TO EARTHQUAKES AS A DISASTER

Earthquakes are the deadliest disaster type reported in EM-DAT and the one with the highest economic losses (CRED, 2011). One of the reasons for these facts may be the sudden and very unpredictable onset of this disaster type. Although earthquakes can be expected along the faults, as it is here the Longmenshan fault belt, the exact prediction of time, location and magnitude is not possible can be estimated only very vaguely by the frequency of historic events. This allows no early warning and preparedness is the key issue to reduce the impact of earthquake disasters. Earthquake resistant infrastructure, like for example reinforced buildings, are one part, another is the knowledge of the population what to do in such a situation (CHAN ET AL., 2010; M. RAMIREZ & PEEK-ASA, 2005).

However the literature review of M. RAMIREZ & PEEK-ASA, 2005 describes how difficult it is to generalize these findings. Thus building types that were regarded as earthquake resistant in one region caused a lot of damage in another. And even within the same region and the same earthquake, the same building type can react differently to a quake depending on where it is located and thus on what soil type it is build on. A similar uncertainty can be applied to the behaviour of individuals during an earthquake. Studies of different earthquakes show that in some earthquakes people who left their buildings were more likely to receive injuries while in others the people who did not move and stayed indoors had a higher risk of injury. Thus the vulnerability in an earthquake is very much depending on the location.

There are many more factors to be considered and as they all depend on each other the subject becomes even more complex. These interactions also make it difficult to identify a straightforward way (like the measure of PGA) to asses the impact of such a disaster (M. RAMIREZ & PEEK-ASA, 2005).

Furthermore there are factors like age, gender, ethnic group that also influence the vulnerability. All this leads to striking differences in the effects of the particular event. Therefore the impact of an earthquake depends on the setting where it occurred. This is represented very well by the huge differences of the earthquakes in Haiti and Chile in 2010. They both had a similar magnitude, yet the impact was completely different.

4. The Deyang CRED Dataset

The Database was established in the context of the MICRODIS project, funded by the European Commission. It is part of a hospital-based patient study and a cross sectional survey. The dataset was collected retrospectively with the help of hospital records. This was done by the research teams of the Université catholique the Louvain, the People’s Hospital of Deyang City and The Sichuan Academy of Medical Sciences & Sichuan Provincial People’s Hospital (CRED, 2010).

4.1. The patient database

The study population consist of patients admitted to the hospital between the 12. and 31. May 2008. Only surviving patients were considered in the DCD. Hence outpatients that died are not reflected in this analysis. In total there are 1878 patient records, registered in the database. Almost all patients in the DCD are located in Deyang prefecture. However only patients that were from Deyang prefecture are considered in the analysis. The Database covers 4 main topics:

- Demographic Data
- Admission and Discharge
- Diagnosis and Accordance
- Surgical Treatment and Procedures

For the analysis the following indicators were selected from the database: patient identification (Did), place of living (DD_ha), cause of injury (AD_ci), time of injury (AD_ti), date of injury (AD_di), admission date (AD_ad), admission time (AD_at) and the ICD-based Injury Severity Score (ICISS).

After refining the dataset, excluding records that did not match the criteria or with missing data (as described below in Chapter 4.1.4 and 4.2), the analysis was undertaken on the remaining 1634 records.

4.1.1. Patient ID

The patient identification (Did) is the unique identifier for the records in the dataset.

4.1.2. Place of living

The indicator “Place of living”(DD_ha) is used to identify the location of the patient. DD_ha does not contain address information, but only the placename. Thus there is no exact information on the location of the patients. The other field that contains location information in the dataset is the Field “Place of work”(DD_wp). DD_ha was chosen as the relevant location information, because the field DD_wp is incomplete for many of the records in the dataset. In the case DD_wp contains information it is identical with the information in DD_ha in most of the cases.

It is not clear from the dataset if the patients really were at their home location, or somewhere else. The assumption is that most of the patients were in their home village or town, as these usually are place of living and place of work. For the spatial scale of the analysis in this thesis, the knowledge of the exact location of the patients at their moment of injury is not necessary. Yet the imprecise location information, of the patients during the earthquake and the possibility of error, has to be taken into account in the interpretation of the results, as it could bias them.

4.1.3. Cause of injury

There are two fields for the “cause of injury” (AD_ci1 and AD_ci2). The first one is related to the earthquake and offers two categories: injury in the main shock or in an aftershock. The second field distinguishes between direct and indirect causes. The information in the second field was not relevant for this study, as only 11 patients stated injury due to direct causes.

The 218 patient records that stated an injury in the aftershock were excluded from the analysis, because of missing information for the time of injury in most of the cases. This issue could be solved easily for the cases injured in the main shock (see below), but there was no possibility to gather the time by a different method for the aftershocks.

4.1.4. Time and date of injury and admission

Time and date of injury, for the patients injured during the primary shock, is between 14:28 and 14:32 on the 12th of May 2008. For a few records this information was missing, and the time of 14:30 was assumed for those records. The

Time of admission is recorded by the hospital and states the time when the patient was admitted at the hospital.

4.1.5. *The ICD-based Injury Severity Score*

Injury severity measurement is important for meaningful comparison of outcomes of trauma care, and assessment of burden of injury. In earthquakes, where high numbers of injuries occurred, a proxy of severity of injury would be of interest to investigate accessibility to health care, evaluation of case management and patient rehabilitation for different levels of severity.

The injuries of the DCD were classified using the International Classification of Diseases (ICD), version 10 (WHO, 2007). This classification alone does not give a quantitative estimate of the threat to life, i.e. probability of death. In contrast, the ICISS, International ICD-based Injury Severity Score, provides a reasonable way to estimate severity for databases using ICD-10 or ICD-10-AM (Stephenson, Henley, Harrison, & Langley, 2004). Each ICD-10-AM category is attributed a survival risk ratio (SRR), which is the probability of a patient surviving a single injury, calculated with national Australian and New Zealand datasets (Stephenson et al., 2004).

ICISS is an estimate of the probability of a patient surviving a given set of injuries. Each patient's ICISS score is the product of the SRRs of each of the patient's injuries. Because some patients suffered up to five injuries, this method produces a single value that fully evaluates the threat to life for that patient. This system is also intended to provide an overall score for patients with multiple injuries. Each injury is assigned an Abbreviated Injury Scale (AIS) score and is allocated to one of six body regions (Head, Face, Chest, Abdomen, Extremities, and External). But only the highest AIS score in each body region is used and therefore only the 3 most severely injured body regions are used to produce the ISS score (Baker, O'Neill, Haddon Jr, & Long, 1974).

Stephenson et al., 2004 provides direct equivalencies between ICD-10-AM and SRR. The DCD was compiled using ICD-10. Thus, first ICD-10 was transformed into ICD-10-AM to obtain each patient's SRR. The final ICISS for each patient was obtained through multiplication of all SRRs available. ICISS closer to 0 has a lower survival probability compared to those with values closer to 1.

The ICISS is a continuous variable. For some parts of the analysis a classification was necessary:

Table 1 The ICISS classification

ICISS	Probability of survival
1-0,8	Very high
0,79-0,6	Medium high
0,59-0,4	Medium low
0,39-0	Very low

The cut offs were chosen in respect of the severity of injury. Thus patients classified with “very high” are considered to have received only light injuries and therefore have a high probability of survival. The lowest class was chosen larger than the others, due to the few cases in this class (Table 1).

4.2. Geocoding the patient database

Matching the placenames in the DD_ha field from the dataset to effective locations was an important task before the actual analysis could be carried out. This was done in close collaboration with the research team at the People’s Hospital of Deyang City. In the first place all the different placenames in the dataset had to be identified. The outcome was a list of 261 different placenames. After several revisions, for double entries, misspellings or false placenames, a final list of 156 placenames was established. Especially the different spellings of the Chinese names due to their translation to English posed a major source of error and double entries. There were also 26 records with an empty location field, or a location outside the study area of Deyang prefecture. All these records were as well removed from the analysis.

The coordinates for the locations of the final placename list were then provided as geographic coordinates by the People’s Hospital of Deyang City research team.

5. OTHER DATASETS

In this chapter additional datasets used in the thesis are introduced. They were not the primary variables for the analysis. However, they provided valuable information for a better understanding of the situation as well as an important visual contribution to the cartographic output.

5.1. THE USGS 2008 WENCHUAN EARTHQUAKE DATASET

The United States Geological Survey (USGS) is one of the most important sources for information on earthquakes. The sensors of the USGS register all earthquakes worldwide. The USGS maintains a database with the parameters of the earthquake, such as location or magnitude. For earthquakes that cause a disaster, more detailed information is available. There is a summary of the impact by the earthquake and a dataset with spatial information like peak ground acceleration (PGA) or MMI can be downloaded from the USGS website. The data is provided as ready maps, or can be downloaded in shapefile format.

In this thesis the PGA and MMI were used to show the earthquake intensity in the study area and to determine the level of exposure of the locations to the earthquake. *"PGA is a measure of earthquake acceleration on the ground and an important input parameter for earthquake engineering. Unlike Richter magnitude scale it is not a measure of the total size of the earthquake, but rather how hard the earth shakes in a given geographic area."* (PRASAD, 2011, P. 83). Usually MMI is derived from personal reports and PGA is measured by instruments in *g*, yet both indicators can be related to each other (Table 2). In the USGS data, the MMI map is also called estimated intensity map, as it is not based on real intensity reports: "*The estimated intensity map is derived from ground motions recorded by seismographs and represents Modified Mercalli Intensities (MMI's) that are likely to have been associated with the ground motions. Unlike conventional Modified Mercalli Intensities, the estimated intensities are not based on observations of the earthquake effects on people or structures. Locations within the same intensity area will not necessarily experience the same level of damage since damage depends heavily on the type of structure, the nature of the construction, and the details of the ground motion at that site. For this* *reason more or less damage than described in the MMI scale may occur. Large earthquakes can generate very long-period ground motions that can cause damage at great distances from the epicenter; although the intensity estimated from the ground motions may be small, significant effects to large structures (bridges, tall buildings, storage tanks) may be notable"(USGS, 2011).*

Table 2 Modified Mercalli Intensity and PGA equivalents; source: (FEMA, 2001)

MMI	Acceleration (%g) (PGA)	Perceived Shaking	Potential Damage
I	< 0.17	Not Felt	None
II	0.17-1.4	Weak	None
III	0.17-1.4	Weak	None
IV	1.4-3.9	Light	None
V	3.9-9.2	Moderate	Very Light
VI	9.3-18	Strong	Light
VII	18-34	Very Strong	Moderate
VIII	34-65	Severe	Moderate to Heavy
IX	65-124	Violent	Heavy
X	> 124	Extreme	Very Heavy
XI	> 124	Extreme	Very Heavy
XII	> 124	Extreme	Very Heavy

5.2. THE LANDSCAN POPULATION DATABASE

The LandScan population dataset represents global population data derived by the LandScan algorithm. The spatial resolution of the data is 3 arc seconds (approximately 90m). This is the highest resolution available at present, for such data. The LandScan algorithm uses imagery analysis technologies and multi-variable dasymetric modelling to generate the dataset. To be able to match the differences of individual countries and regions, the LandScan has individual models for each region and country. The database was first established in 1998, particularly to be able to estimate populations at risk (BHADURI, BRIGHT, COLEMAN, & URBAN, 2007; OAK RIDGE NATIONAL LABORATORY, 2011).

Population data is crucial information to address issues in disaster management. This database provides such information in a higher granularity then census data and is not limited by its boundaries (Bhaduri et al., 2007). To compile the dataset the methodology processes different datasets: census information, administrative boundaries, land cover, other spatial data (slope, elevation, roads, etc.), coastlines and imagery. Every year there is an updated version available (OAK RIDGE NATIONAL LABORATORY, 2011). The version used in this thesis is from 2008.

The Oak Ridge National Laboratory develops the LandScan database. The research for the database is funded by the US Department of Defense. The database is usually freely available for non-commercial use.

5.3. THE ESRI BASEMAP DATA

The GIS software company ESRI, is providing an online map service for free accessible basemap imagery. These can be directly integrated into ArcGIS as basemap. There are various themes available, in this thesis the World Topographic Map or the World Shaded Relief were used as background for the maps.

These imagery allow to show the data in the context of the present morphology and infrastructure, such as mountains, rivers or the road network. Therefore this information helps to interpret the data in a meaningful way.

5.4. THE GLOBAL ADMINISTRATIVE UNIT LAYER (GAUL) DATASET

The administrative units used in this thesis were all derived from the Global Administrative Unit Layers (GAUL) dataset. The dataset is provided by the FAO for free use. It is a global dataset that is compiled for all countries using the same methodology. It is updated on a regular basis and has usually a detail down to admin3. For some countries even lower administrative boundary levels are available (GEONETWORK, 2007). The most recent version, which was used here, is from 2009.

6. METHODS

“Spatial analysis” is the concept to approach a geographic problem. There are many methods summarized under the term “spatial analysis”. Therefore the methods that are applicable have to be identified individually for each study. In the following chapter the methods used in this thesis will be introduced. Three different approaches that can be regarded as bottom up were applied in the thesis. In the first stage the dataset was analysed descriptively. In the second stage two indicators were identified for the application of interpolation analysis. At last statistical analysis was performed on the dataset, to check for possible overall correlation in the dataset.

All the analyses performed and maps created in this thesis were done using ArcGIS 10, except the correlation analysis in 6.3.1, which was performed using R version 2.12.2.

6.1. DESCRIPTIVE ANALYSIS

The most basic way to look at geographic data is to map it and interpret the outcome. This is usually also the first approach to get insights into the data. Moreover such maps are understood very easily, as they do not use complicated methods and they stick to the principals of cartographic visualization. In this thesis the spatial distribution of the data as well as the temporal analysis were interpreted in a descriptive way.

6.1.1. SPATIAL DISTRIBUTION

The spatial distribution of the data points is the first impression of such geographic data. The distribution or clustering of the data in geographic space can already give insights into the characteristics of the dataset and draw the attention to hotspots of data point accumulations.

All the data were mapped, according to its geographic location obtained in the geocoding process described in chapter 4.2. As mentioned in chapter 4.1.2 the accuracy of this information has to be considered.

Furthermore the attributing of the locations with information linked to them allows a better characterization of the data. Here the data was aggregated by certain criteria: First, the number of patients per location was obtained. This was then taken further by classifying, by the ICISS classes, defined in chapter 4.1.5.

The data obtained this way was visualized in maps together with different basemap imagery and the LandScan data.

6.1.2. Temporal analysis

Time geography is a useful approach in epidemiology. It is rather constraints orientated then interested in individual choice. Time geography examines the human activity in space over time. It can be characterized by this simple question: *"how does participating in an activity at a given place and time limit a person's abilities to participate in activities at other places and times?"*(Xiong, 2008, p. 1152). However to be applicable for the subject of this thesis the question should be rephrased to: "how does the degree of exposure to the earthquake, personally e.g. injury and environmentally e.g. blocked roads, limit a persons ability to reach the hospital?" Although temporal analysis of geographic problems is present since more than 30 years it has been limited by the ability to store, collect and handle such data. Recently the possibilities to handle and collect such data in temporal GIS databases have very much evolved as well as the GIS analyses tools (Xiong, 2008, p. 1147 ff; 1151 ff). "*The ability to analyze data for both spatial and temporal patterns in a single presentation provides a powerful incentive to develop tools for communication to map users and decision makers."* (Longley, 2005, p. 280).

For the temporal analysis the data was classified by the day of admittance to the hospital and the ICISS class. Although the data is available in minute resolution it was aggregated to days. Next, the data was visualized in a map series for each ICISS class, with one map per day, showing the number of patients admitted to Deyang hospital for each day. To add the context to the whole dataset, the total patients on county level are included in the map.

This analysis can also be visualized as an animation, to provide a more dynamic impression of the trend.

6.2. Interpolation analysis

To be able to interpret the data with a different approach two indicators were identified for interpolation analysis. Interpolation is a method where a raster surface is computed from known points. This is preceded by the assumption, based on the Tobler law, that things that are spatially close have similar characteristics. Based on this, values are predicted for the cells in the raster, from the sample points (ESRI, 2011a). The method can also be described with the following statement: *"Spatial interpolation is the GIS version of intelligent guesswork."* (Longley, 2005, p. 333).

Typically interpolation analysis is used on continuous phenomena, like precipitation or temperature data. In this thesis the selected indicators, for this analysis, were as well assumed to be continuous. Most likely people located in the same area would face similar circumstances and problems. Thus if a road was blocked, all people that were located in this particular area and had to come to the hospital, would take more time. Similarly, but not with the same direct link, it can be assumed that in areas with high ICISS, there is a higher threat of severe injury and that this risk is similar to the other people that live in this area. Therefore "time to reach the hospital" and ICISS were selected for interpolation.

There were two different interpolation methods applied in this thesis, inverse distance weighted and Kriging interpolation. The reason to present two interpolation methods is to show the various possibilities of GIS analysis as well as the comparison of the results received by two different methods. However it is not the subject of this thesis to investigate the differences and the advantages or disadvantages of the two different methods, but to analyse and compare the results derived by these methods.

6.2.1. Inverse distance weighted interpolation (IDW)

IDW is a deterministic interpolation method. This means that the values assigned to a specific location follow a specified mathematic formula. In the case of IDW the influence on the interpolated value is isotropic. Hence, the weight of a point drops by a factor of 4 if the distance to the point doubles. IDW creates a smooth surface (ESRI, 2011b, 2011c; Longley, 2005, p. 335).

If points are coinciding, as it is the case for almost all points in this thesis, the values were averaged. The search radius for the input points is set to variable with the default of 12 points. The power (*k*) value, which controls the significance of the surrounding points, was set to 2.

For the visualization of the IDW interpolation a continuous scale was chosen, as this represents best the smooth surface created by IDW interpolation.

6.2.2. KRIGING

The Kriging interpolation method is well grounded in a theoretical framework. Thus the interpolation process is not a "black box" and the user has several possibilities to influence the result. Kriging is a geostatistical interpolation method by which spatial autocorrelation is also taken into account. This method assumes, that the distance or direction reflects a spatial correlation. Hence the differences in the correlation for different directions and distances can be taken into account. This correlation can be used for the prediction of the output surface. Kriging also has the possibility of validation by a variance raster surface (ESRI, 2011D, 2011E; LONGLEY, 2005, P. 336).

There are many different Kriging methods. In this thesis a basic and often used method called "Ordinary Kriging" was applied. Kriging usually is a two step process: first the dependency rules have to be uncovered, second the predictions are calculated (ESRI, 2011D). ArcGIS offers the possibility to integrate these tasks with the "Geostatistical Wizard". This dynamic and interactive tool was used to create the Kriging surfaces.

On the contrary to the IDW interpolation, a classified scale better visualizes the Kriging surface. The scale was classified by Geometrical Interval. This Method allows to balance between highlighting changes in the middle values and the extreme values.

6.3. STATISTICAL ANALYSIS

The statistical analysis, in particular spatial statistics, is a valuable part of the approach of "spatial analysis". It allows, testing formally for the patterns, observed and described in the process of data visualization (XIONG, 2008, P. 1136). In this thesis the data was tested for bivariate correlation and global spatial autocorrelation.

6.3.1. Bivariate correlation analysis

The overall association between two variables was calculated using the Spearman rank correlation coefficient. This method is preferable because some of the variables of interest are closer to ranks. The Modified Mercalli Intensity scale (MMI), severity of injury measured as per ICISS, and the “time to reach the hospital” (in hours) were correlated in bivariate analyses.

The MMI for each patient record was derived from the USGS MMI data, through a spatial join of the datasets.

6.3.2. Spatial autocorrelation (Global Morans I)

Spatial autocorrelation measures the spatial dependence of a variable. It can be either positive, similar variables are collocated; negative, dissimilar variables are collocated or random, in which case the data is spread independent and identically in space (Xiong, 2008, p. 34).

The Morans I, is one of the oldest and most often used methods to asses spatial autocorrelation. It measures the interdependence of an observed value as a function of the distance. The Morans I tests if the distribution of the data is random or if there is a positive or negative autocorrelation in the data. The Global Morans I is applied to the whole dataset and does not identify where in the dataset this spatial autocorrelation occurs (ESRI, 2011f; Xiong, 2008, p. 754).

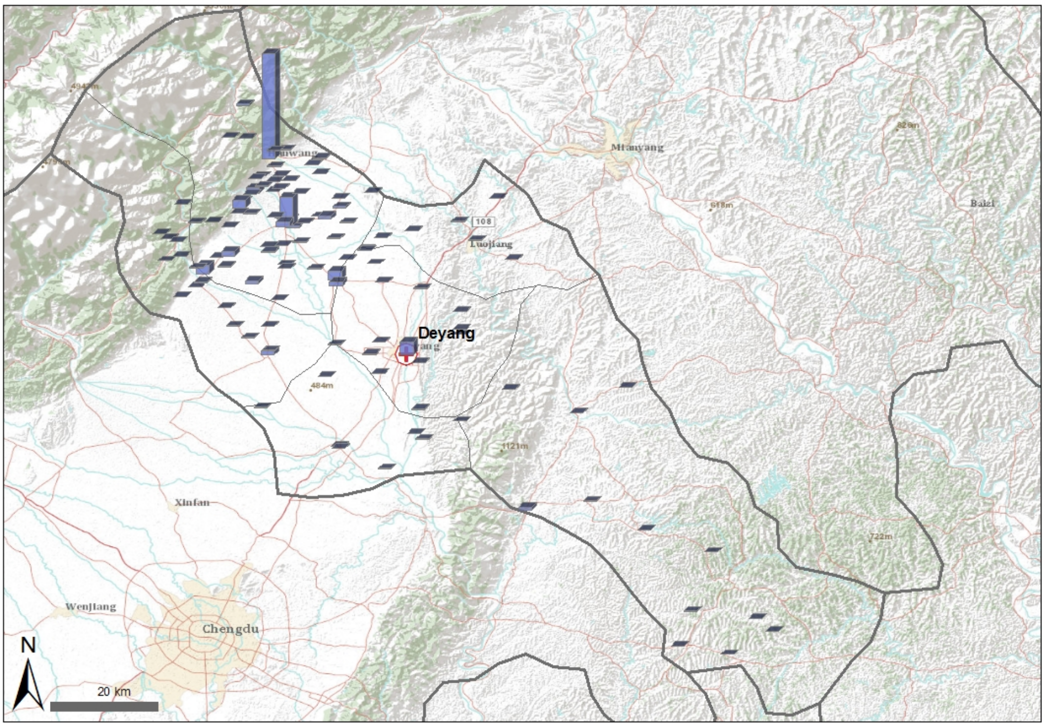
The indicators ICISS and “time to reach the hospital” were tested for spatial autocorrelation. The “Conceptualization of Spatial Relationships” parameter was set to “Fixed_Distance_Band” and the minimum distance was calculated automatically. This ensures that each feature has at least one neighbour. The Euclidean distance parameter was set for the calculation of the distance between the points (ESRI, 2011g).

7. Spatial Analysis of the Deyang CRED Dataset

In this chapter the Results from the spatial analysis will be presented. The order in which the results are presented, illustrates the direction of the analysis, from simple to more complex methods.

7.1. The Spatial Distribution of the Patients Admitted to the People's Hospital of Deyang City

Map 3 shows the amount of patients from every single location in the DCD. The height of the bars represents the number of patients admitted to the hospital from one particular location. One can see that most locations contributed just a few patients to be admitted to the Deyang hospital. The most patients came to the hospital from one particular location, Hanwang, a town in the Northwest of Deyang, at the foot of the Longmen Mountains. There is also a concentration of patient locations towards the mountains in the Northwest. On the other hand very few patients came to Deyang hospital from Zhongijang Xin, the county southeast of Deyang. There were also very few patient locations in this county.



Map 3 Number of patients per location. The tallest bar is Hanwang and equals 704 patients

Most patients came to Deyang from the plain between the Longmen Mountains in the Northwest and the mountains in the Southeast. This first overview already shows the very uneven distribution of the dataset spatially, as well as in terms of patient numbers (compare also with Table 8).

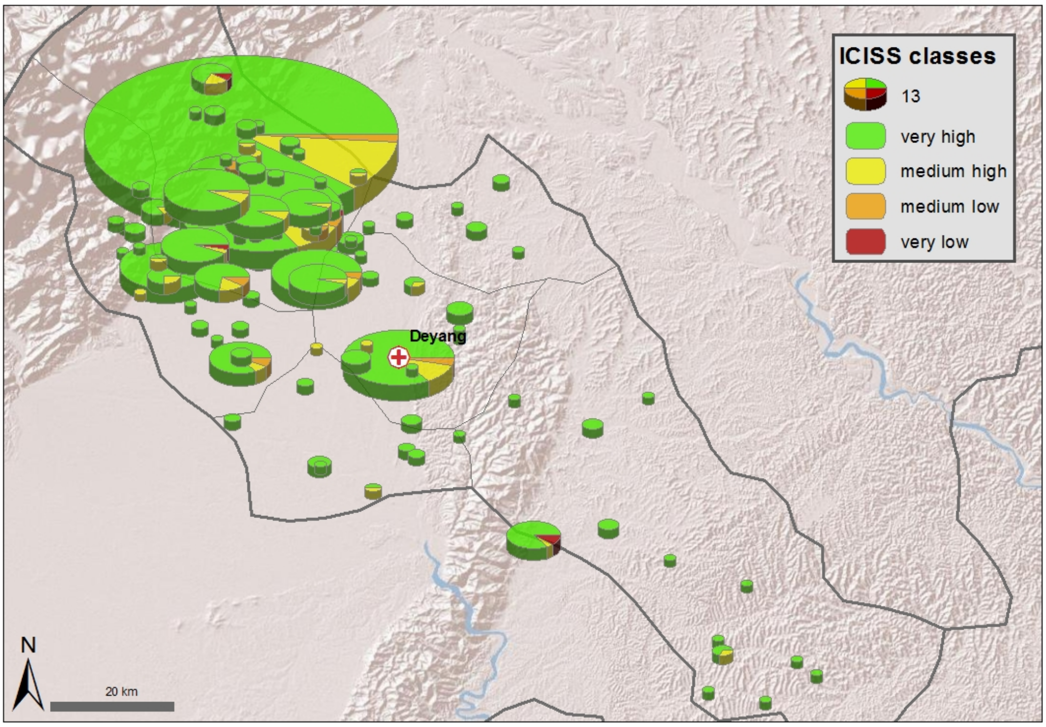
To get a closer look on the data, the DCD was classified using the ICISS. Table 3 shows the number of patients in each class. The vast majority (89%) of the patients in the DCD have an ICISS higher than 0,8, which means a high probability of survival. On the other hand only 15 patients (0,9%) have an ICISS below 0,4. However it has to be taken into account that only patients that did not die were accepted to the Deyang city hospital, so this group could be larger, if the outpatients that did not survive due to their heavy injuries would have been considered. This is of course true for all the other classes (see also chapter 8.1).

Table 3 Total number of patients by ICISS class

ICISS class	Total number of patients
very high	1454
medium high	135
medium low	30
very low	15

The four classes that were established in chapter 4.1.5, were displayed in Map 4 as pie charts.

The size of the chart indicates the number of patients per location. In total the class with a very high probability of survival dominates the dataset. Except for two locations, all records classified below “very high”, are located in Deyang or in the North-West of Deyang. It is important to take into account, that only surviving patients were included in the DCD. Had mortality been included, this probably would have changed the picture. In general one can affirm that further Northwest and closer to the epicentre, there were more severely injured patients. But there were also in general more patients from this region.



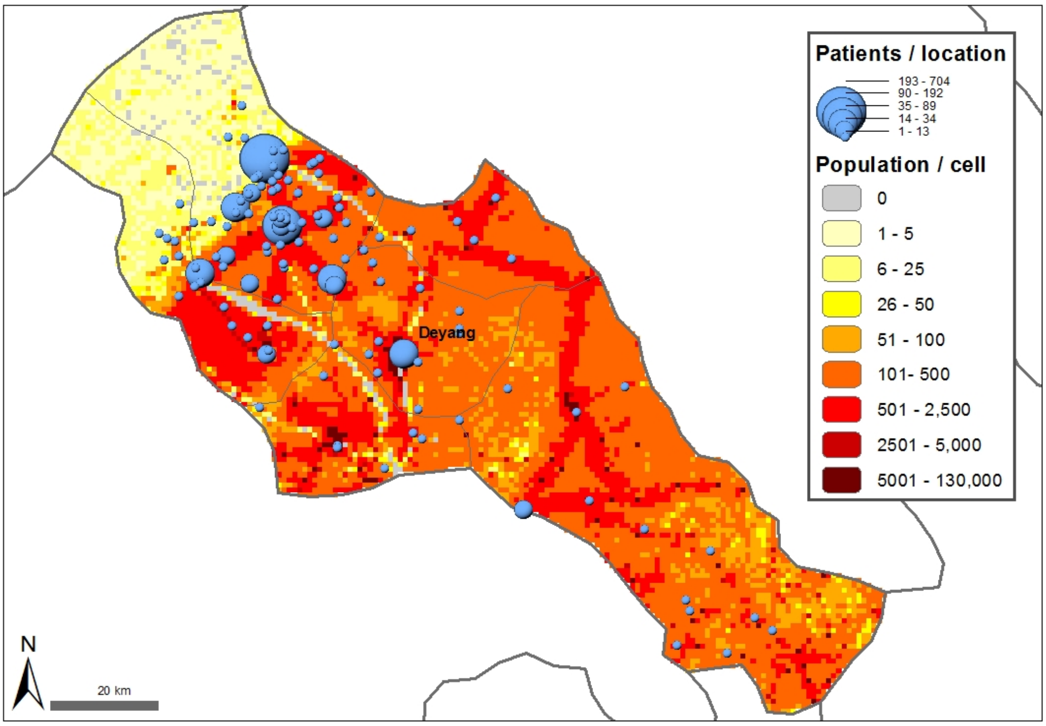
Map 4 ICISS classes by location

In general this is also the most densely populated area in Deyang prefecture. While the population density in the mountainous regions is minor, especially in the Longmen Mountains in the Northwest, population density in the plain that starts at the foot of the Longmen Mountains is high.

When looking at Map 5 one can also see that most population locations coincide, with areas, where more than 2500 people per gridcell live. Table 4 shows the population for the counties in Deyang prefecture. Thus the county with the highest population, Zhongjiang Xian, accounts only for 2,4% of the patients in the DCD. Overall the patients mark only a small portion of the population, the highest patient rate was detected at Mianzhu Shi county, which also accumulated the largest number of patients.

Table 4 Population derived from LandScan and number of patients, on admin3 level

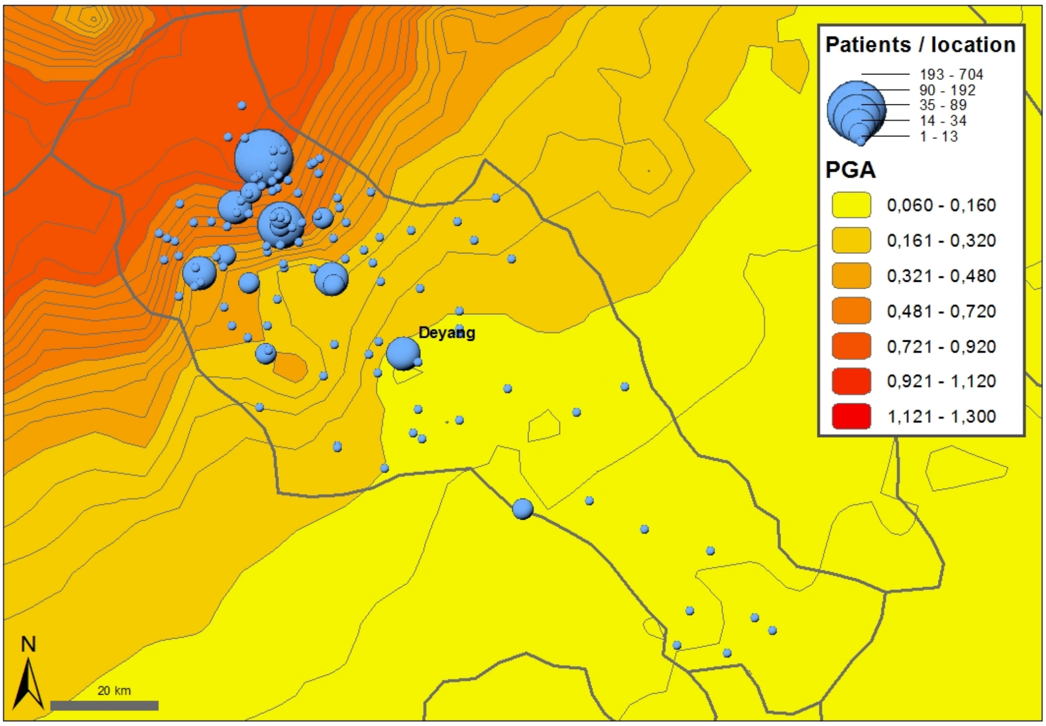
Admin3	Inhabitants	Patients	Rate in ‰
Deyang Shi	553871	120	0,2167
Guanghan Shi	618860	16	0,0259
Luojiang Xian	364616	10	0,0274
Mianzhu Shi	581555	1329	2,2853
Shifang Shi	447326	120	0,2683
Zhongjiang Xian	1470051	40	0,0272



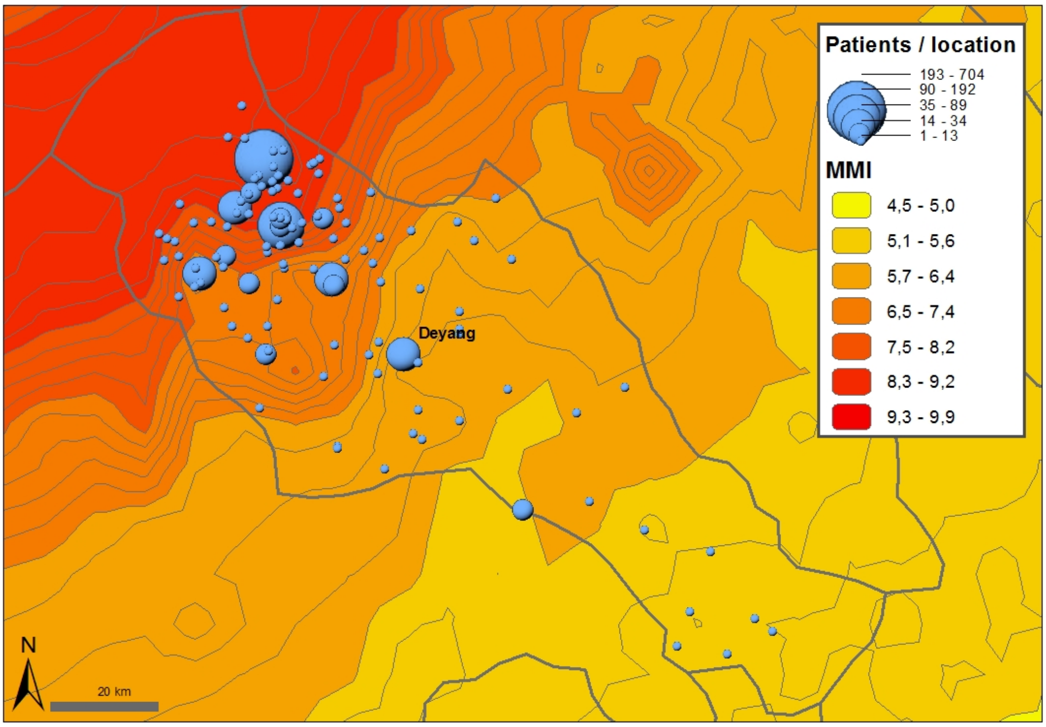
Map 5 Patient locations and population density. Population density source is the LandScan dataset with a cell size of 1 arc sec.

If the earthquake intensity is considered, it is clearly reflected that as one would expect, with decreasing intensity there are also less patients. Map 6 and Map 7 display two common indicators to show the earthquake intensity, PGA and MMI (see also chapter 5.1). Both indicators show a similar trend on different scales: The further from the epicentre and the fault, the lower the intensity.

But when the 2 scales are compared, according to Table 2 a difference between the 2 indicators comes to the fore. While the measured PGA is maximally associated with weak perceived shaking and no damage, the MMI scale varies between V and VI, which would be perceived as moderate or strong shaking and associated with light or medium heavy damage. Thus the MMI in this case seems to represent the actual impact better than the PGA. However none of the 2 indicators were correlated to the ICISS or "time to reach the hospital" (see chapter 7.4). Also the MMI patterns have a slightly different extent than the PGA values and the higher MMI values seem to fit better to the locations with higher patient numbers.



Map 6 Patient locations and peak ground acceleration (PGA)



Map 7 Patient location and MMI

7.2. Temporal analysis of the spatial distribution of the patients, classified by the ICISS score

The data in the DCD offered a good opportunity to analyse the event temporally, as the time of the injury was defined with relative precision by the main shock and the time of admittance to the hospital can be regarded as very reliable. The classification by ICISS gives a better understanding of the behaviour of the 4 patient groups after the quake.

It is important to consider that the indicated time is only stating the time between the injury and the time the patient reached the hospital. Thus it remains unclear how the time in between was spent. So it could be possible that patients did not go directly to the hospital for instance if they had only light injuries, or they received first aid before they were admitted. Yet both would have increased their time to reach the hospital.

Table 5 Number of patients by day of admittance and ICISS class

May the	ICISS class "very high"	ICISS class "medium high"	ICISS class "medium low"	ICISS class "very low"	Total patients
12	455	41	7	5	508
13	445	63	16	0	524
14	287	23	5	9	324
15	83	2	1	1	87
16	53	2	1	0	56
17	16	1	0	0	17
18	50	2	0	0	52
19	18	0	0	0	18
20	2	1	0	0	3
21	1	0	0	0	1
22	18	0	0	0	18
23	6	0	0	0	6
24	1	0	0	0	1
25	4	0	0	0	4
26	6	0	0	0	6
27	5	0	0	0	5
28	4	0	0	0	4

When analysing the trend in the complete dataset it is clearly visible that the first three days are the most important in the disaster. 1356 patients were admitted in these first days. Furthermore severe injuries were received at the hospital only in the first days of the catastrophe (Table 5, Figure 2). This corresponds to the insight in the literature, where the first 3 days of an earthquake disaster are regarded as crucial.

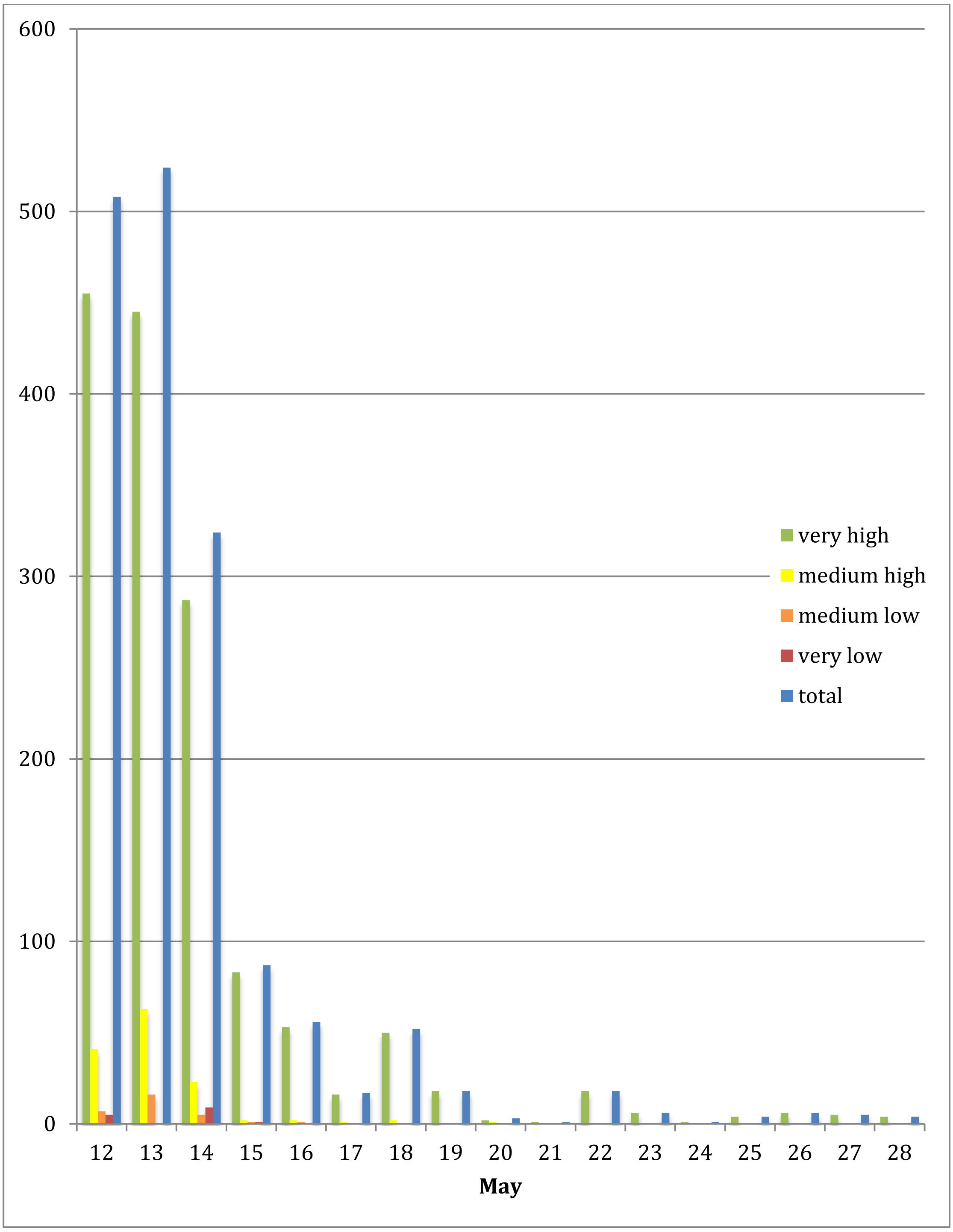
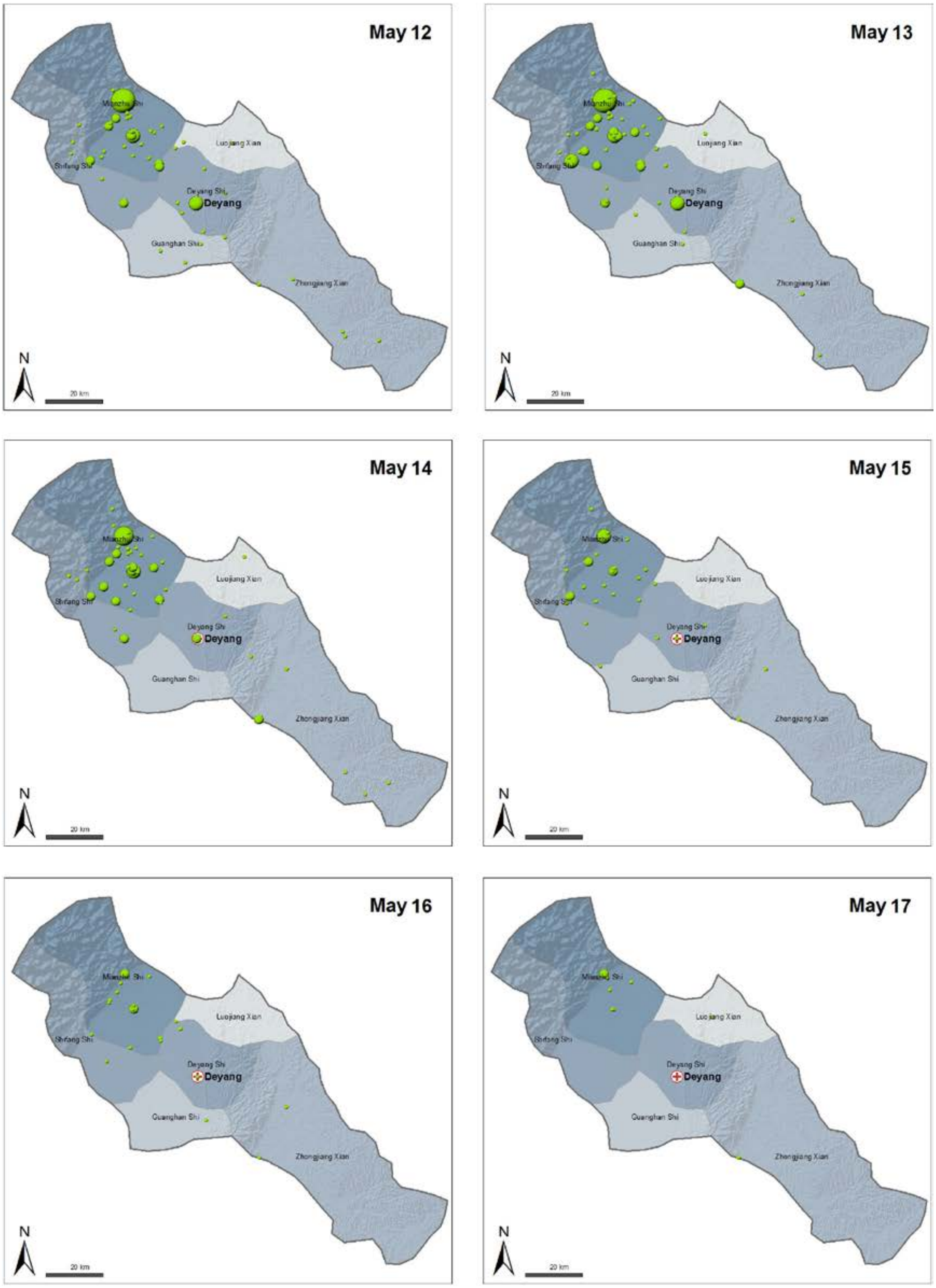
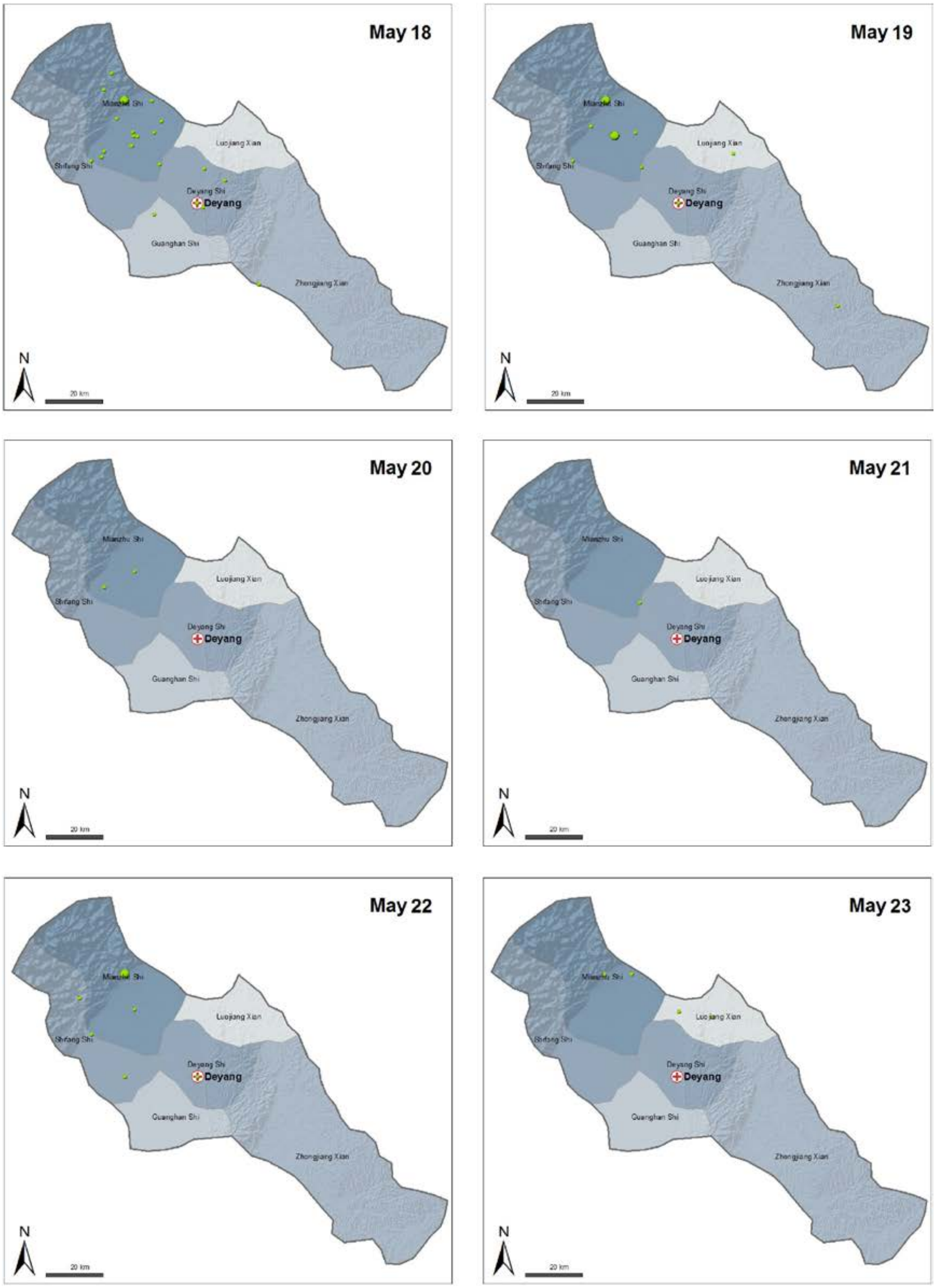


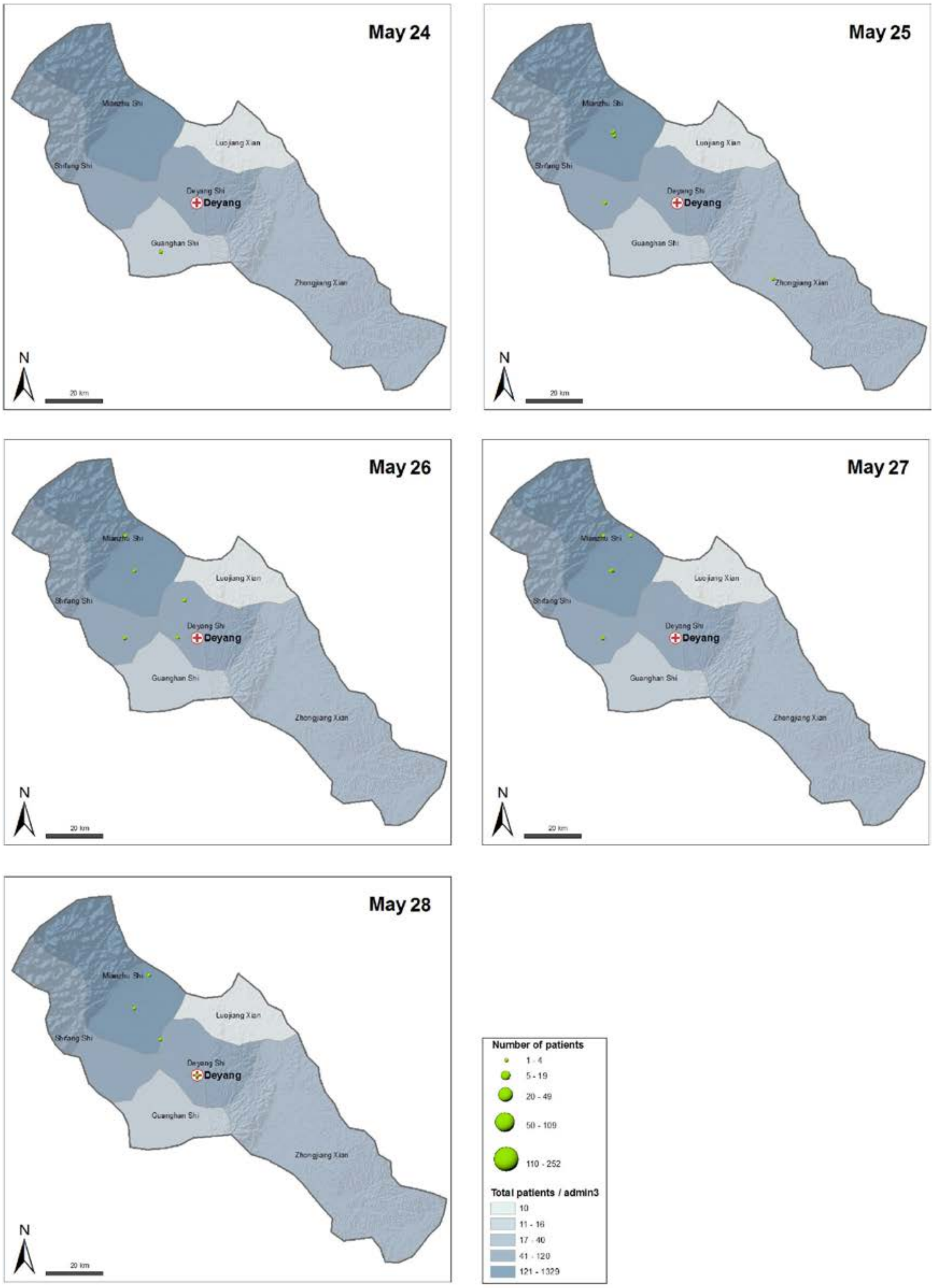
Figure 2 Patients by ICISS class and day

The first map series shows the ICISS class "very high" (Map 8). 1454 patients were classified as very high. That makes it the largest class in the DCD. It is also the only ICISS class in the DCD that has constant admittances, every day of the study period. The peak of patients admitted to the hospital is on the 12th May, the first day of the disaster. On the 20th May the admittances then considerably drop compared to the 8 days before.

From all ICISS classes, the “very high” class is the one, with the most patients from Zhongijang Xin county in the southeast. However most patients in this class are from Mianzhu Shi county, which is similar to the other ICISS classes. Mianzhu Shi county also has the highest number of patients in the DCD. This is furthermore the region from where patients kept coming continuously to the Deyang hospital, also in the later days of the event.


May 12
May 13
May 14
May 15
May 16
May 17
Mianzhu Shi
Luojiang Xian
Shifang Shi
Deyang Shi
Deyang
Guanghan Shi
Zhongjiang Xian
N
20 km

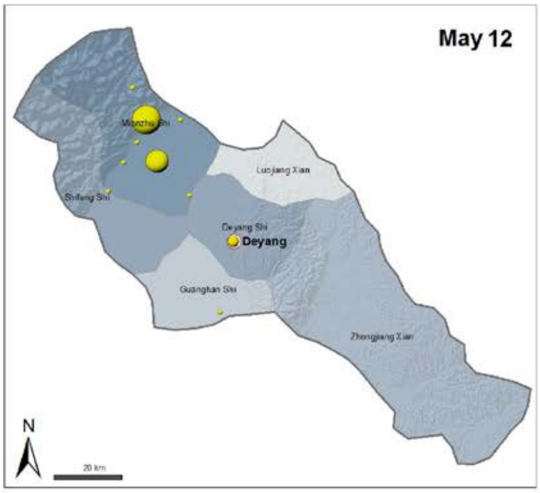

May 18
May 19
May 20
May 21
May 22
May 23
Mianzhu Shi
Luojiang Xian
Shifang Shi
Deyang Shi
Deyang
Guanghan Shi
Zhongjiang Xian
N
20 km

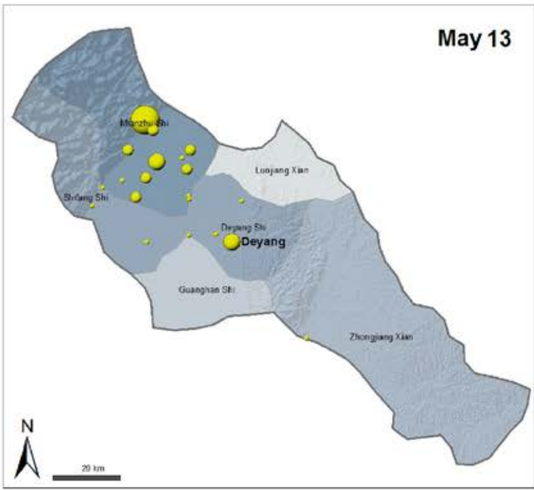


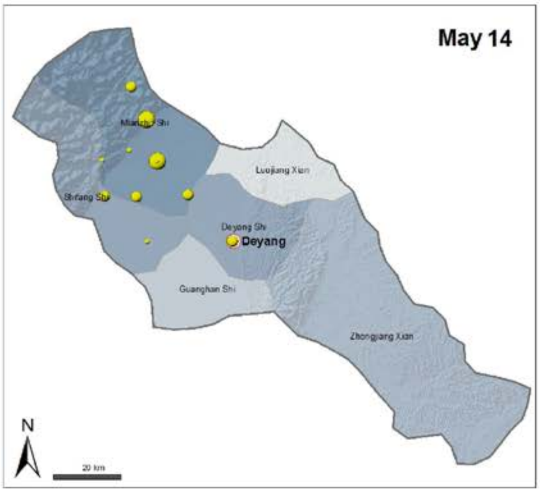
Map 8 Time series of the patient origin for the ICISS class "very high"

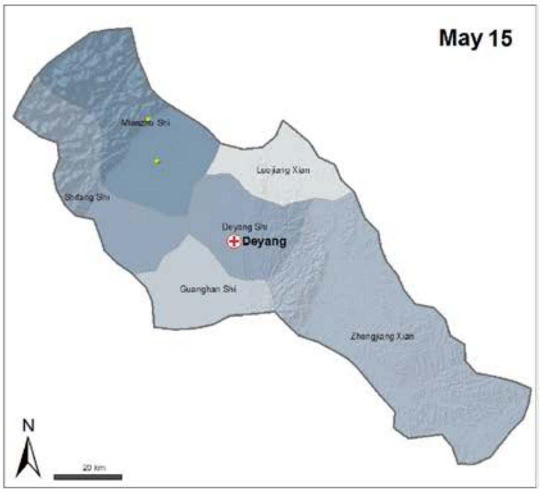
The “medium high” class reflects the second biggest class in the DCD (Map 9). On the second day of the disaster, most patients classified as “medium high” were admitted to Deyang city hospital. Like in the “very high” class most patients in this class came from the county Mianzhu Shi in the North of Deyang prefecture. On the other hand only 2 patients, on the 13^{th} and 18^{th} of May, came from the county Zhongjang Xian.

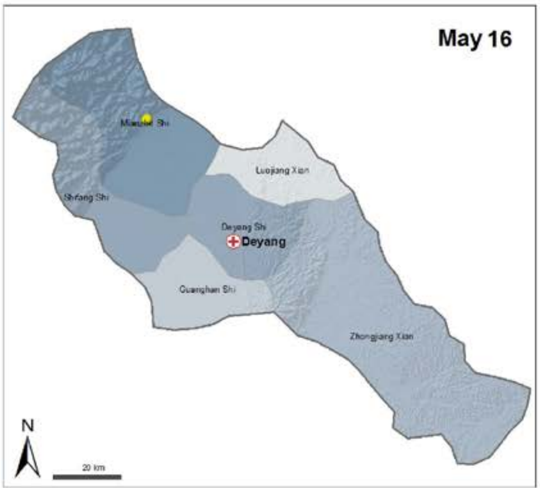
Patients with injuries in this class were coming to the hospital continuously until the 18^{th} of May and then one more patient at the 20^{th} of May. After the 20^{th} of May no more patients in the ICISS class “medium high” were admitted.

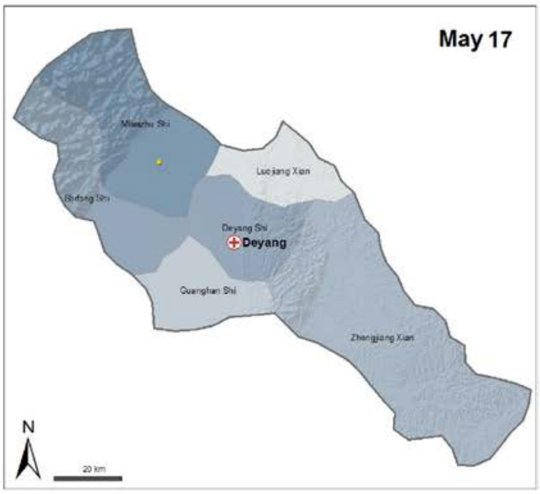

May 12
Mianzhu Shi
Luojiang Xian
Shifang Shi
Deyang Shi
Deyang
Guanghan Shi
Zhongjiang Xian
N
20 km

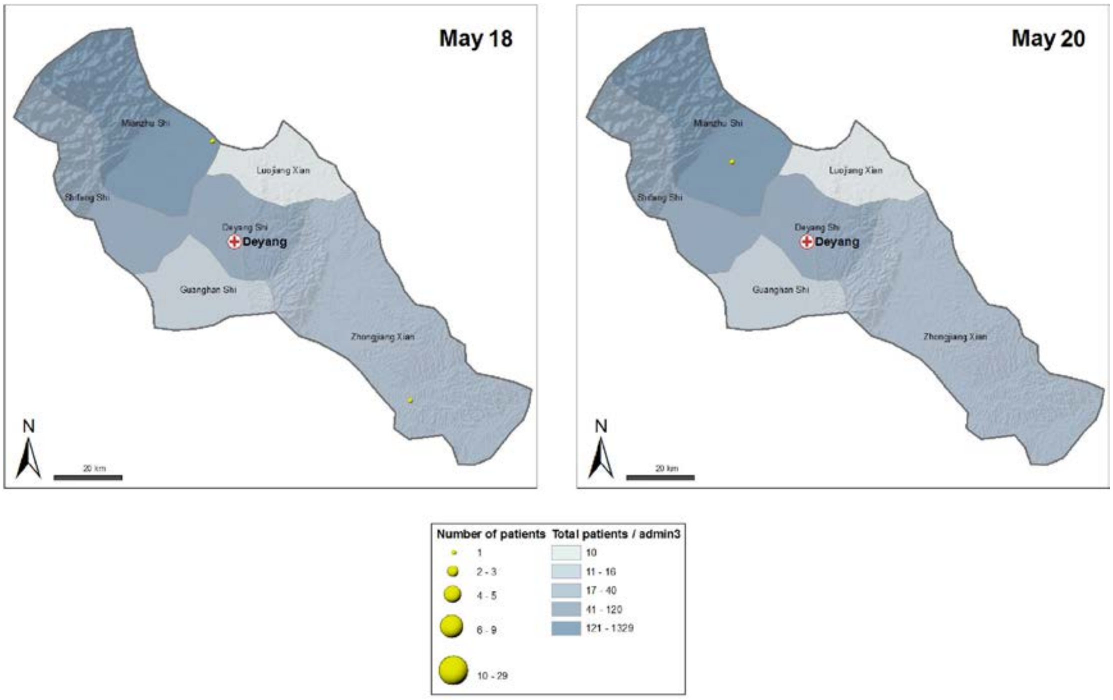

May 13
Mianzhu Shi
Luojiang Xian
Shifang Shi
Deyang Shi
Deyang
Guanghan Shi
Zhongjiang Xian
N
20 km


May 14
Mianzhu Shi
Luojiang Xian
Shifang Shi
Deyang Shi
Deyang
Guanghan Shi
Zhongjiang Xian
N
20 km


May 15
Mianzhu Shi
Luojiang Xian
Shifang Shi
Deyang Shi
Deyang
Guanghan Shi
Zhongjiang Xian
N
20 km

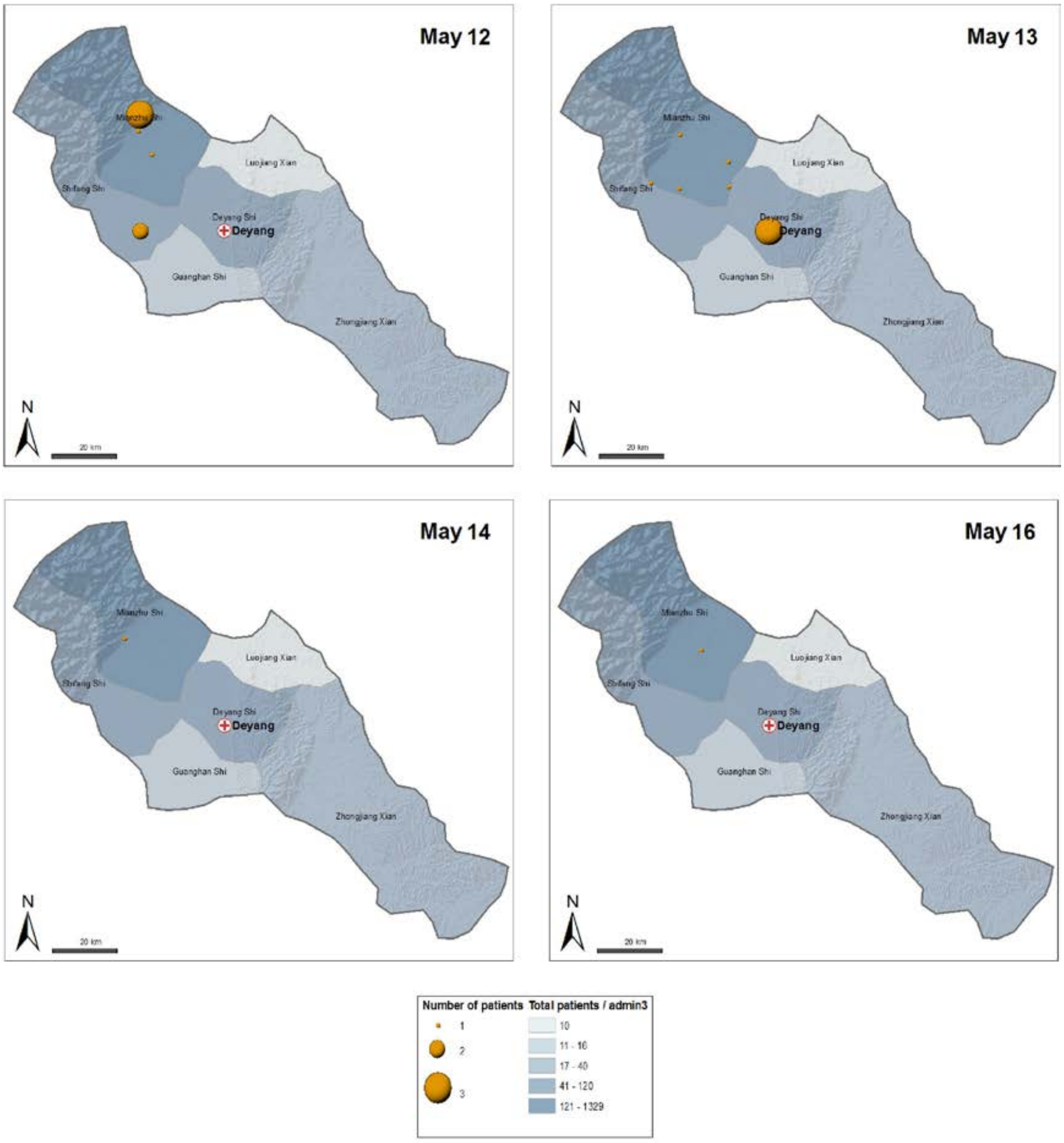

May 16
Mianzhu Shi
Luojiang Xian
Shifang Shi
Deyang Shi
Deyang
Guanghan Shi
Zhongjiang Xian
N
20 km


May 17
Mianzhu Shi
Luojiang Xian
Shifang Shi
Deyang Shi
Deyang
Guanghan Shi
Zhongjiang Xian
N
20 km



Map 9 Time series of the patient origin for the ICISS class "medium high"

In Map 10 the admittance series of patients classified as "medium low" is shown. Similar to the previous two classes the peak of hospital admittances is at the second day. The patients in this class are even more concentrated in the county Mianzhu Shi. Furthermore no patients classified as "medium low" came to the Deyang city hospital from the county Zhongjang Xian in the southeast. Another difference to the other classes is that at no location there were more than 3 patients in this class at the same day.

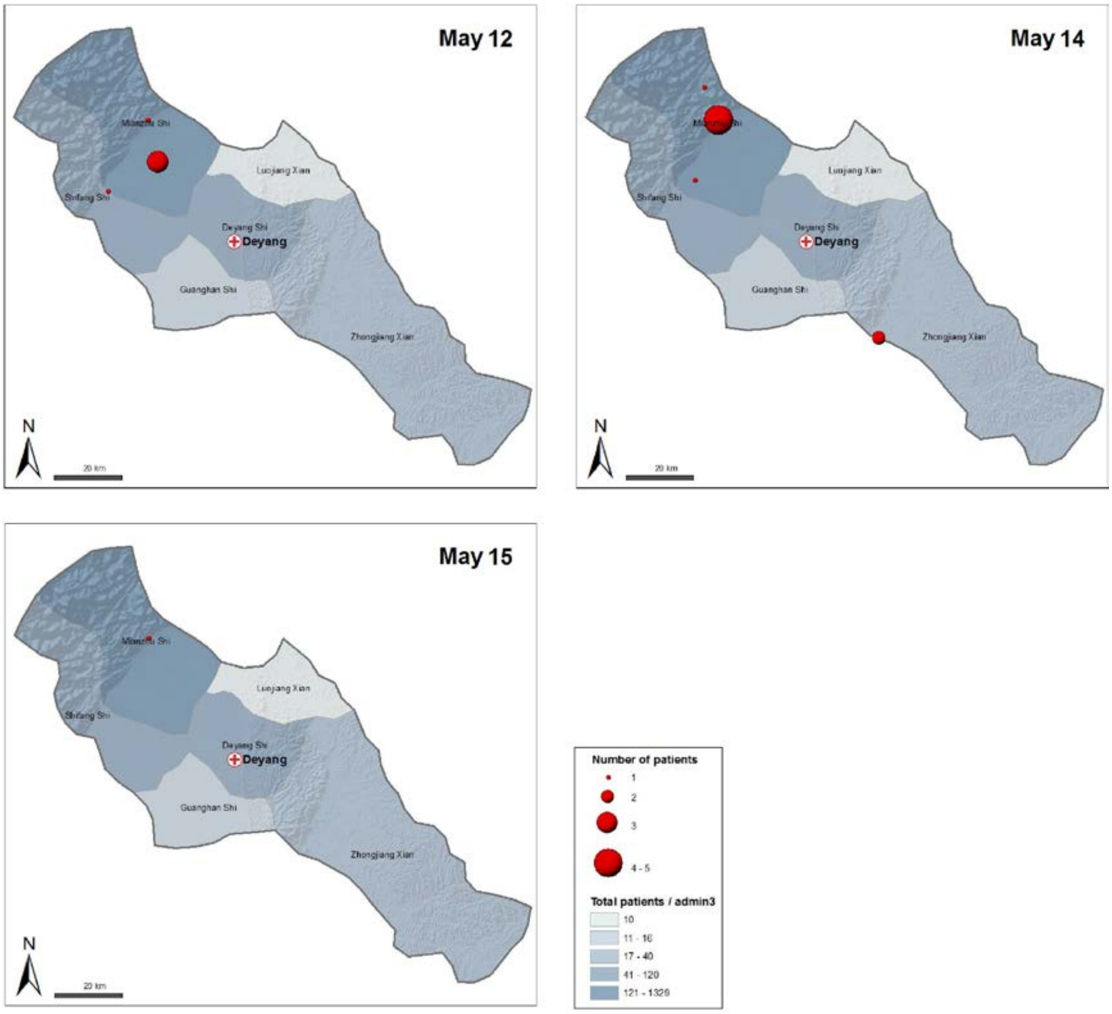


Map 10 Time series of the patient origin for the ICISS class "medium low"

The last map series represents the patients with a low ICISS classification; this means that they were severely injured. First of all it shows that there are only few patients in this class (15 patients). As described previously in chapter 6.1.1 this number has to be considered carefully, especially in this class, due to the missing non-surviving patients.

There were only 3 days, at which patients classified with "very low" came to the Deyang city hospital. On the contrary to the other classes the patient peak in this class was at the third day, while on the second day no patients were classified as “very low”. However, the pattern of patient concentration in the county Mianzhu Shi presents clear analogies to the other injury severity classes.

If considering the total number of patients in this class it has the biggest share of patients from county Zhongiang Xian.



Map 11 Time series of the patient origin for the ICISS class “very low”

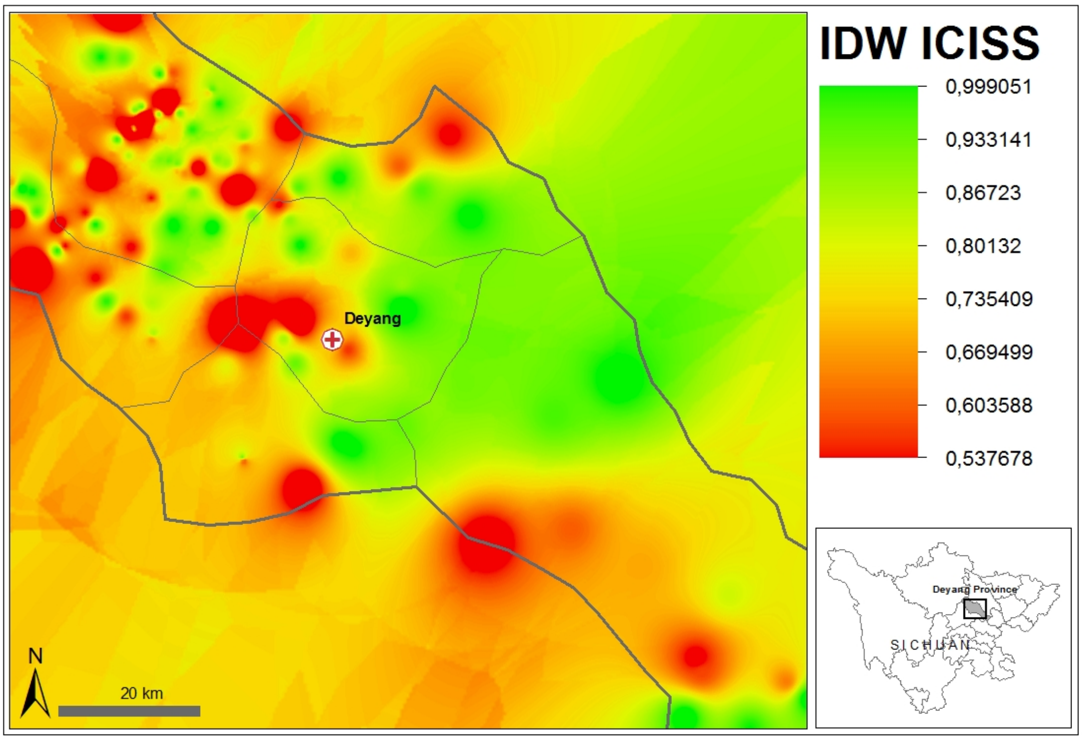
Unsurprisingly the concentration of the patients to the county Mianzhu Shi prevails for all classes. And even for most of the days apart of a few exceptions Mianzhu Shi is the county with the most patients that were admitted at the Deyang city hospital.

7.3. Interpolation of the "time to reach the hospital" and the ICISS

In this chapter the interpolation of the indicators "time to reach the hospital" and ICISS is presented. When looking at the extent of the map, one can see that a smaller area is covered compared to the other maps due to the smaller area of the interpolations. The interpolation area is cut off compared to the study area because it is determined by the outermost data points.

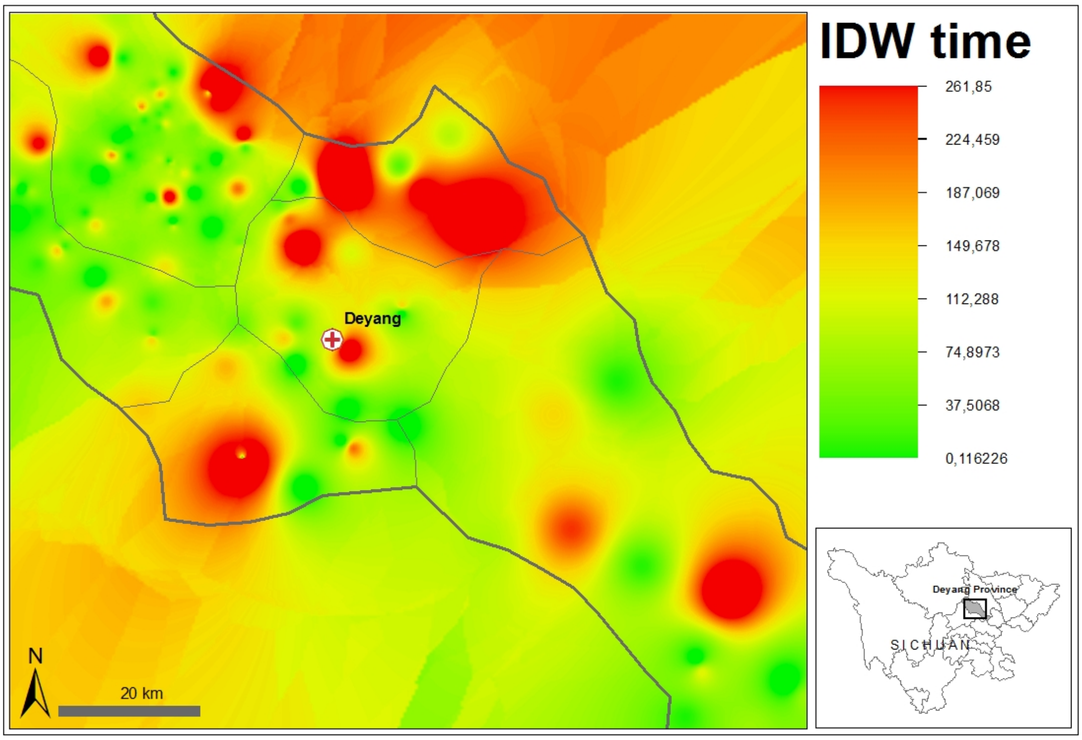
The IDW interpolation of the ICISS and "time to reach the hospital" gives a good impression of the situation represented in the DCD. The ICISS value scale is not showing the extreme values that are in the DCD, because the average value is used in the case of coinciding locations. Of course the interpolation is less reliable towards the edges and in areas with few data points.

In Map 12 the IDW interpolation of the ICISS is shown. One can see immediately that in the area east and also north of Deyang there were few severely injured people whereas in the west and northwest of Deyang there were more people severely injured, as well as in the southwest.



Map 12 IDW interpolation of ICISS

When looking at the time interpolation in Map 13 another pattern is visible. The areas in the east and especially in the north are shown to be the ones with the highest time span to reach the hospital. On the other hand in the areas in the west and northwest, that show a low ICISS (Map 12), it took the patients less time to reach the hospital. One conclusion from this can be that severely injured patients reached the hospital faster than less severely injured patients (see also chapter7.6).



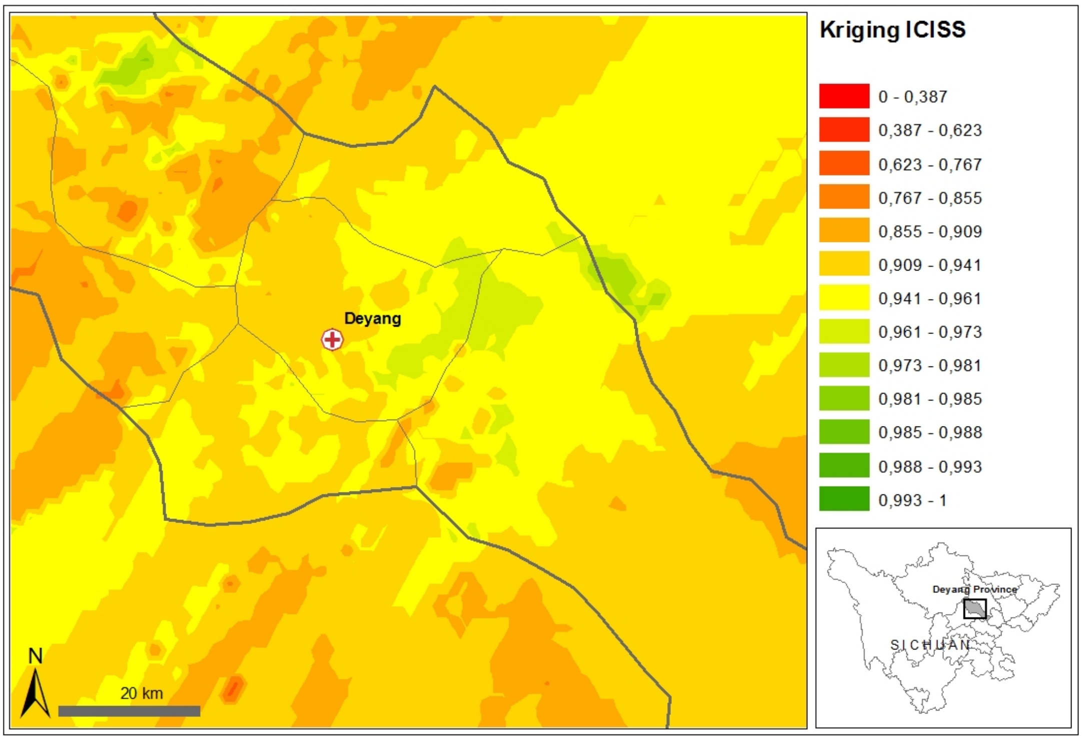
Map 13 IDW interpolation of "time to reach the hospital"

The same pattern is also visible in the maps that were created with the Kriging method. This sophisticated interpolation method gives a more detailed display of the indicators, but the trend in patterns overall stays the same as in the maps, that were created using the IDW method.

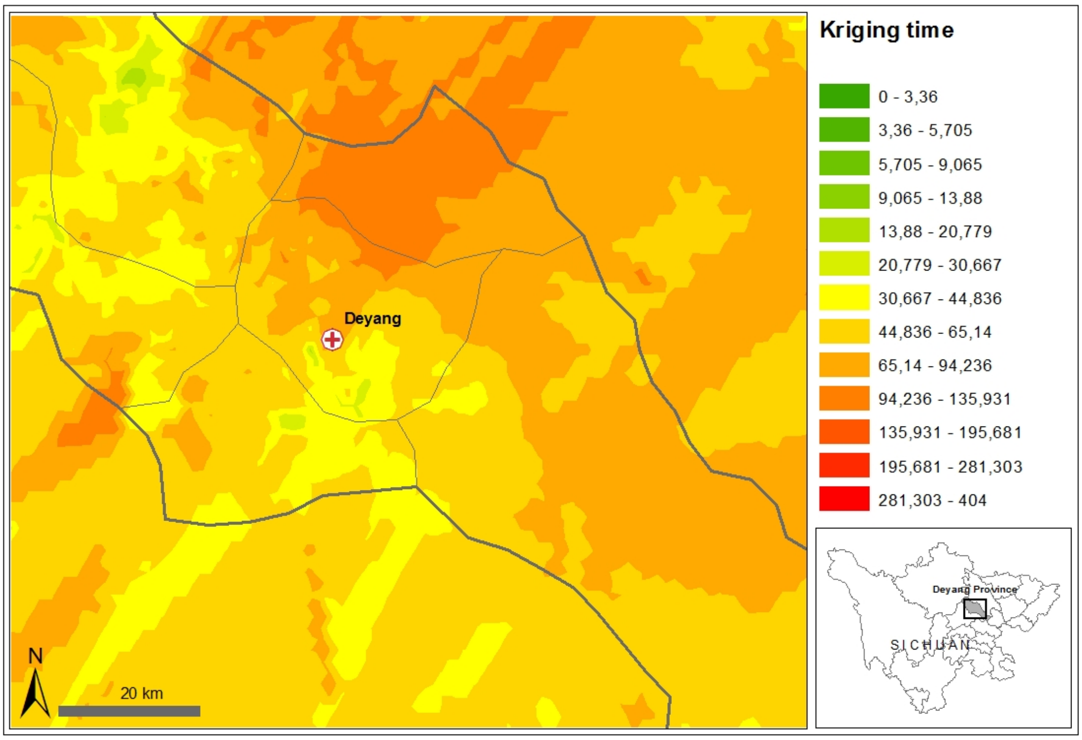
However this method offers a better possibility for an in depth analysis of the situation, especially together with the error prediction, that is available for the Kriging method. So the specific areas are better distinguishable, and can be identified more easily. Also in Kriging the coinciding points do not have to be averaged for interpolation and the information is not lost.

The error prediction is shown in Map 16 for the ICISS indicator and in Map 17 for the "time to reach the hospital" indicator. Both maps show that the error gets bigger towards the edges and that there is an area with an elevated prediction error in the centre of Mianzhu Shi county, despite this being the area with the most datapoints. This is probably due to the different values of the indicators despite their large spatial proximity.

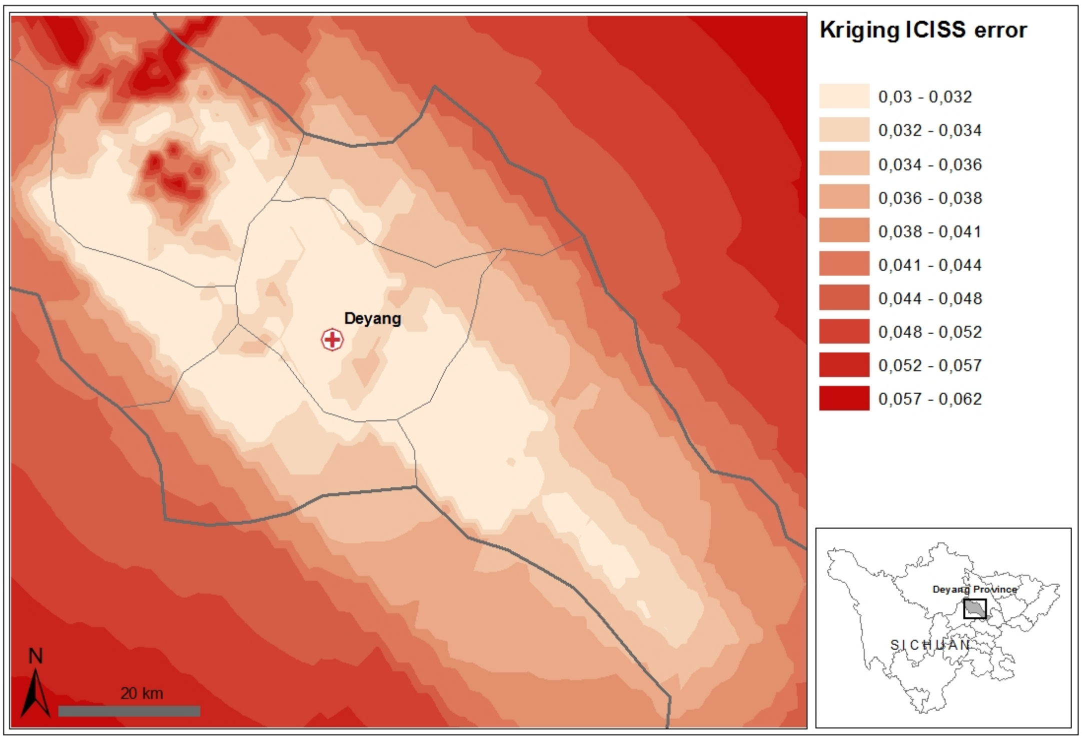
When looking at the graduation of the error, one can also see that the error span is way bigger for the "time to reach the hospital", than for the ICISS.



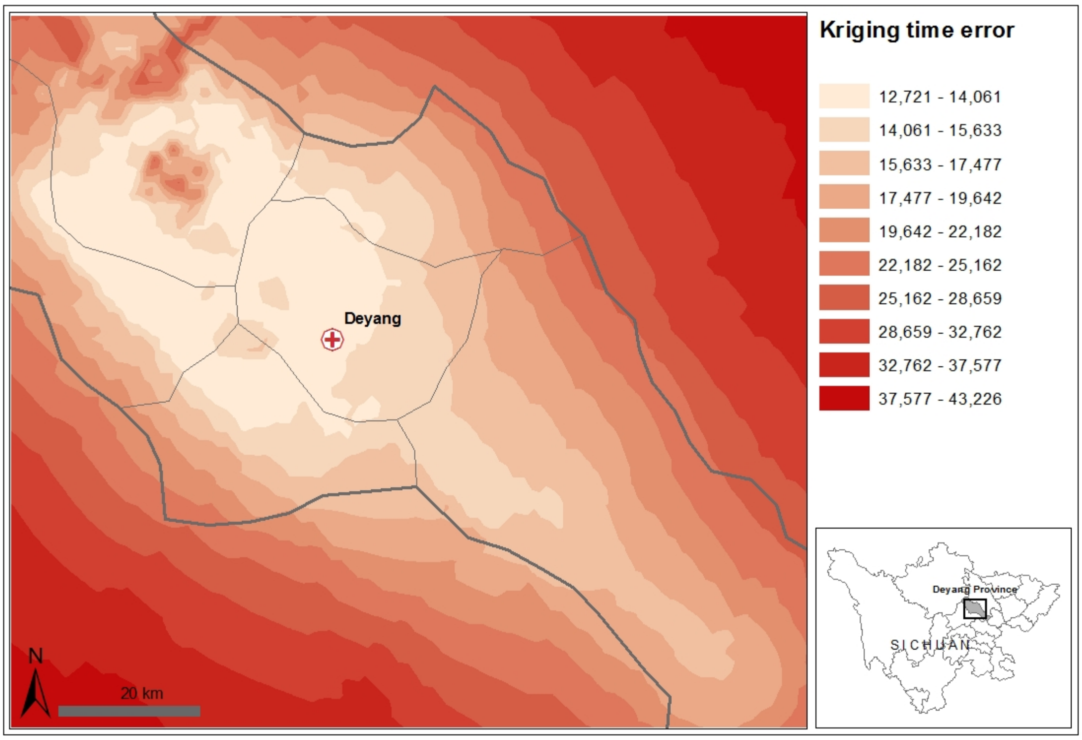
Map 14 Kriging interpolation of ICISS



Map 15 Kriging interpolation of "time to reach the hospital"



Map 16 Kriging prediction error for the ICISS interpolation



Map 17 Kriging prediction error for the "time to reach the hospital" interpolation

Table 6 Summary of interpolation results

	IDW	Kriging
ICISS range	0,538-0,999	0-1
Time to reach the hospital	0,116h-261,85h	0h-404h
ICISS error span	Not available	0,03-0,06
Time to reach the hospital error span	Not available	12,72-43,23

7.4. BIVARIATE CORRELATIOS BETWEEN MMI, "TIME TO REACH THE HOSPITAL" AND THE ICISS SCORE.

The Modified Mercalli Intensity scale (MMI) was not correlated with ICISS (Spearman rank correlation: r_S = -0.01, N = 1635, P = 0.713) but did with "time to reach the hospital" (Spearman rank correlation: r_S = -0.12, N = 1635, $P < 0.001$). The latter indicates an earlier evacuation of injured patients inhabiting areas exposed to higher earthquake intensity. ICISS and "time to reach the hospital" were not correlated (Spearman rank correlation: r_S = -0.01, N = 1635, P = 0.689).

The test also suggests that there could be some biases in the data that have to be identified and should be solved before further statistical analysis (see chapter 8).

7.5. Test for spatial autocorrelation of "time to reach the hospital" and ICISS

To determine if there is a spatial autocorrelation in the DCD the indicators "time to reach the hospital" and ICISS were tested with the Global Morans I. While no spatial autocorrelation was detected for the ICISS (Morans Index = -0,000224 Z score = 0,314, P = 0,753) very high spatial autocorrelation was identified "for the time to reach the hospital" (Morans index = 0,0229, Z score = 19,05, P = 0,0).

The result suggests that there is a very high probability of a cluster of high spatial autocorrelation for this indicator in the DCD.

7.6. Summary of the Results

The spatial analysis shows that the amount of patients from Deyang prefecture that were admitted to the Deyang city hospital is much higher in the north of the prefecture and especially concentrated to Mianzhu Shi county. The number of patients decreases from Northwest to Southeast, which corresponds to the earthquake intensity. Such a trend would also agree with the findings of (Peek-Asa et al., 2000) in their study "GIS mapping of earthquake-related deaths and hospital admissions from the 1994 Northridge, California, earthquake". However the values of the indicators suggest a lower damage (Table 2), than actually occurred. Furthermore the origin of a large part of the patients in the DCD to very few locations, especially Hanwang town with 704 patients, becomes visible.

The more detailed approach to the data through the indicators ICISS and the time of admittance, through which the "time to reach the hospital" was derived, highlights some more trends in the DCD.

In general, most patients were admitted during the second day of the disaster. While the light injured patients admittances peaked at the first day and then continuously declined, the more severe injuries had their peak on the second or third day and then diminished the faster the severer the injury was. One can conclude that the second and third day of the disaster are the key days for severe injuries and speculate that severely injured patients that are admitted in the later stages of the disaster did not survive, as the deaths of patients are not represented in the DCD.

On the other hand for those patients that are represented in the DCD, the time to reach the hospital was no issue if they were severely injured. Thus if the severity of the injuries was high the patient reached the hospital fast, even if the distance to its location was further away. The interpolations of the indicators seem to underline these findings, as in areas with low ICISS the time to reach the hospital is low as well and the other way round.

The high spatial autocorrelation of the time value could be explained through the damages to the roads, which were very high in this disaster. Hence people that were at the same location faced the same difficulties. On the other hand ICISS is probably more related to individual behaviour, age or the building type.

However the findings appear to be difficult to validate with statistical methods. This may be partly due to the limitations and issues of the data, that are also discussed in chapter 8.1. Yet the positive test for spatial autocorrelation gives a hint were to focus and move forward with research on the DCD.

Table 7 Summary of results

Complete dataset	Time to reach the hospital	ICISS
Unequally distributed patients in the study area	Most admittances at the second day	Second and third day of disaster are the most important
Concentration of patients in the Northwest (Mianzhu Shi county)	Severely injured patients are admitted only in the first days	Only lightly injured patients admitted at every day
Concentration of patients to few locations (e.g. Hanwang)	Distance is not an issue for severely injured patients	Severely injured patients decrease rapidly in the later stages and the heavier the injuries are
Number of patients decreases with earthquake intensity	Similar time characteristics within certain areas, indicate that time is influenced by the environmental circumstances	More severely injured patients in the North than in the South

8. Discussion

The following chapter highlights the issues that came to the fore during the analysis as well as the ways forward to solve these problems, which are beyond the scope of this thesis. It also gives a more personal view on some of the aspects of this thesis.

8.1. Limitations

The results presented in the previous chapter show the difficulties in such analysis but also the potential of the data. There are certainly many problems to be addressed of which probably the most important is, the lack of information why and under which circumstances a patient was admitted to the Deyang city hospital. Did the patients receive initial treatment somewhere else? Which would have affected their time to reach the hospital, as well as if they were transferred from an other hospital due to a certain condition. Was there a pre selection of patients, maybe according to the severity of the injuries?

As complementary to the just mentioned questions, the lack of data from other hospitals in this area can be regarded. Such data would allow getting a bigger picture and enabling the identification of more general trends, apart of those discussed in this thesis. The trends identified here are always clearly related to this particular hospital and situation.

Another problem that comes to the fore when dealing with the DCD is the appointment of the patient location. The location was assigned on the basis of certain assumptions (chapter 4.2) that cannot be validated and may be not true for all cases, and might bias the results. However as the locations are aggregated to the place of living and not to the accurate address, a certain tolerance is maintained, that is sufficient for the analysis presented in this thesis. Yet if other analysis should be applied on the data, like for example the correlation of injuries with collapsed buildings, or similar, this would pose a problem.

Another missing point in this thesis is additional impact data or even ordinary spatial data, like infrastructure or socioeconomic data. It would be interesting to layer such data together with the present dataset and investigate if this could explain any of the observed patterns. So for example if the “Manhattan distance” and not the “Euclidean distance” could be calculated and if furthermore road damage or roadblocks through landslides or similar could be brought into the analysis, maybe the differences in the time to reach the hospital, could be explained more satisfactory.

Besides the issues mentioned above, it would be interesting to have some baseline data, so that changes in the patterns to the normal situation could be identified. Socioeconomic data as well would be interesting, to see if the patterns could be related to any socioeconomic groups.

A completely different topic that has to be brought to attention is the privacy issues that always come a long with such data. As already mentioned in the introduction these issues are often hindering the collection of such sensitive data. The lower the level of aggregation the easier it becomes to make inferences, especially with geographic data. On the other hand, the more the data is aggregated the more is lost from the primary information and may even exclude some of the analysis that is presented here. This is an on going discussion and there are different approaches how to target these challenges on the different levels (social, ethical, legal and technological) (Brown et al., 2009; Longley, 2005; Xiong, 2008). The DCD is anonymised and aggregated only to the place of living. Curtis, Mills, & Leitner, 2006 suggest that guidelines should be developed and applied for the research with such point data. Therefore it has to be considered carefully how the data is handled and presented. Thus this question always has to be targeted when undertaking research with such data.

8.2. Ways forward

Starting to address the issues discussed above would be a good next step to continue the research on the DCD. Especially the implementation of the criteria and circumstances of admittance could give valuable insights. But even when looking at the topic of research considering the human impact in disasters from a broader perspective, several possibilities arise that could be investigated with similar methods or are associated to the subject.

One study that has to be mentioned here is the "Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti" by BENGTSSON, LU, THORSON, GARFIELD, & VON SCHREEB, 2011. The authors assessed the population movement after the Haiti earthquake. The geographic locations were derived from the SIM card data, provided by the mobile network operator. This study shows an interesting new approach to retrieve the locations of the affected population in a disaster. It would be interesting to investigate the use of such data for the assessment of the human impact of disasters.

But geographic coordinates can also be derived with other methods, like GPS devices at the moment of data collection or by geocoding addresses. In general it is important to put the effort on the assessment of the geographic conception of a study or dataset during the study design. This may be a matter of course for geographers, but not always for participants in the disaster research that come from other disciplines, who are quiet frequent, as disaster research is very multidisciplinary.

Databases, like the DCD could also help to create deductive and predictive models, although each region and disaster is very different (ALEXANDER, 2000). One only has to compare the Haiti and Wenchuan or even the most recent Fukushima earthquake, to recognize the large differences of each event. HU, WANG, ZHU, & REN, 2011 created a model in their study of "Mapping under-five mortality in the Wenchuan earthquake using hierarchical Bayesian modelling", using the under-five mortality data from the Wenchuan earthquake. The data was very unsymmetrical and suggested a larger variability than actually existed, similar to the DCD and the model was used to overcome the issue and smooth the data. This model was then applied in an adjacent study by HU, WANG, LI, REN, & ZHU, 2011 for the risk assessment of under-five mortality, in the Wenchuan earthquake. The risk assessment then could be provided to decision makers. These two studies show how such research can be applied to create results that can support decision-making.

9. CONCLUSION

The here presented work can be regarded as the first step of the spatial analysis of the DCD. The thesis explores several approaches, towards the investigation of the DCD and describes the spatial characteristics of the dataset. Thus the uneven distribution of the patient origins that were admitted at Deyang city hospital is shown. Moreover the temporal patterns of the admittances at the hospital in relation to their origin and the severity of injury could be displayed and described. During the examination of the data for relations, spatial autocorrelation of the patient time to reach the hospital was identified.

However the thesis also highlighted some of the issues in the dataset, like the missing of information of patient mortality outside the hospital or the lack of data from other hospitals, to allow a broader assessment of the situation.

Yet the thesis offers several starting points for further research. This could be modelling, or in depth statistical analysis of specific aspects, or the cross section analysis with other datasets.

Furthermore the question of individual confidentiality and privacy has been discussed and can be regarded as a key matter in the presented research because it is crucial for the availability of such data. One could argue that disaster situations are a very exceptional state of emergency that requires exceptional measures. On the other hand such argumentation has already often been misused and the right for privacy is unquestionable. In the end the sensitivity of the subject has to be recognised and there is no general answer. Probably this subject has to be solved for each research case individually.

Though this dataset can be regarded as a rare possibility for an in depth analysis, especially spatially, of the human impact of this earthquake disaster on a high resolution. It will be certainly a challenge to get some of the additional information or data mentioned in this thesis, due to the restrictive data policy on high resolution data in China.

It is probable that data of the kind and therefore research similar to the one presented here will increase. Geographic data is becoming more and more important and the awareness and demand for it in the scientific community as well as in the general public is rising. Such a development will hopefully help to prevent, mitigate or at least reduce the impact of catastrophes, like the one investigated here.

BIBLIOGRAPHY

Alexander, D. (2000). On the spatial pattern of casualties in earthquakes. *Annals of Epidemiology*, *10*(1), 1-4.

Armenian, H. K., Melkonian, A., Noji, E. K., & Hovanesian, A. P. (1997). Deaths and injuries due to the earthquake in Armenia: a cohort approach. *International Journal of Epidemiology*, *26*(4), 806 -813. doi:10.1093/ije/26.4.806

Baker, S. P., O'Neill, B., Haddon Jr, W., & Long, W. B. (1974). The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *Journal of Trauma*, *14*(3), 187-196.

Bengtsson, L., Lu, X., Thorson, A., Garfield, R., & von Schreeb, J. (2011). Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in haiti. *PLoS Medicine*, *8*(8).

Bhaduri, B., Bright, E., Coleman, P., & Urban, M. L. (2007). LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal*, *69*, 103-117. doi:10.1007/s10708-007-9105-9

Brown, T., McLafferty, S., & Moon, G. (2009). *A Companion to Health and Medical Geography*. John Wiley and Sons.

Chan, E. Y. Y., Gao, Y., & Griffiths, S. M. (2010). Literature review of health impact post-earthquakes in China 1906–2007. *Journal of Public Health*, *32*(1), 52 -61. doi:10.1093/pubmed/fdp078

CRED. (2010). *Wenchuan earthquake study report*. Center for the Research on the Epidemiology of Disasters.

CRED. (2011). Database | EM-DAT. Retrieved October 27, 2011, from http://www.emdat.be/database

Curtis, A. J., Mills, J. W., & Leitner, M. (2006). Spatial confidentiality and GIS: Re-engineering mortality locations from published maps about Hurricane Katrina. *International Journal of Health Geographics*, *5*. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-33750469045&partnerID=40&md5=f41b8175e9553496ccea30cbbbc77642

ESRI. (2011a). Desktop Help 10.0 - Understanding interpolation analysis. Retrieved October 18, 2011, from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Understanding_interpolation_analysis/009z0000006w000000/

ESRI. (2011b). Desktop Help 10.0 - How IDW works. Retrieved October 17, 2011, from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/How_IDW_works/009z00000075000000/

ESRI. (2011c). Desktop Help 10.0 - IDW (Spatial Analyst). Retrieved October 18, 2011, from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009z0000006m000000.htm

ESRI. (2011d). Desktop Help 10.0 - How Kriging works. Retrieved November 2, 2011, from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/How_Kriging_works/009z00000076000000/

ESRI. (2011e). Desktop Help 10.0 - Kriging (Spatial Analyst). Retrieved October 18, 2011, from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Kriging/009z0000006n000000/

ESRI. (2011f). Desktop Help 10.0 - How Spatial Autocorrelation (Global Moran's I) works. Retrieved November 2, 2011, from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/How_Spatial_Autocorrelation_Global_Moran_s_I_works/005p0000000t000000/

ESRI. (2011g). Desktop Help 10.0 - Spatial Autocorrelation (Global Moran's I) (Spatial Statistics). Retrieved November 2, 2011, from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Spatial_Autocorrelation_Global_Moran_s_I/005p0000000n000000/

FEMA. (2001). *Understanding Your Risks: Identifying Hazards and Estimating Losses* (No. FEMA 386-2). FEMA.

GeoNetwork. (2007, July 2). GeoNetwork opensource portal to spatial data and information. InteractiveResource. Retrieved December 2, 2011, from http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691

Hu, Y., Wang, J., Li, X., Ren, D., & Zhu, J. (2011). Geographical detector-based risk assessment of the under-five mortality in the 2008 Wenchuan earthquake, China. *PLoS ONE*, *6*(6). Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-79959585033&partnerID=40&md5=ee302903ce34efa5f4f424f383b4fa14

Hu, Y., Wang, J., Zhu, J., & Ren, D. (2011). Mapping under-five mortality in the Wenchuan earthquake using hierarchical Bayesian modeling. *International Journal of Environmental Health Research*, *21*(5), 364-371.

Liao, Y.-F., Hwang, L.-C., Chang, C.-C., Hong, Y.-J., Lee, I.-N., Huang, J.-H., Lin, S.-F., et al. (2003). Building collapse and human deaths resulting from the Chi-Chi earthquake in Taiwan, September 1999. *Archives of Environmental Health*, *58*(9), 572-578.

Longley, P. (2005). *Geographic information systems and science*. John Wiley and Sons.

Mc Leod, K. S. (2000). Our sense of Snow: the myth of John Snow in medical geography. *Social Science & Medicine*, *50*(7-8), 923-935. doi:10.1016/S0277-9536(99)00345-7

Meade, M. S., & Emch, M. (2010). *Medical Geography*. Guilford Press.

Oak Ridge National Laboratory. (2011). LandScan - Documentation. Retrieved October 28, 2011, from http://www.ornl.gov/sci/landscan/landscan_documentation.shtml

Peek-Asa, C., Ramirez, M. R., Shoaf, K., Seligson, H., & Kraus, J. F. (2000). GIS mapping of earthquake-related deaths and hospital admissions from the 1994 Northridge, California, earthquake. *Annals of Epidemiology*, *10*(1), 5-13.

Prasad, B. B. (2011). *Advanced Soil Dynamics and Earthquake Engineering*. New Delhi: PHI Learning Pvt. Ltd.

Ramirez, M., & Peek-Asa, C. (2005). Epidemiology of traumatic injuries from earthquakes. *Epidemiologic Reviews*, *27*, 47-55.

Schmidtlein, M. C., Shafer, J. M., Berry, M., & Cutter, S. L. (2011). Modeled earthquake losses and social vulnerability in Charleston, South Carolina. *Applied Geography*, *31*, 269-281. doi:10.1016/j.apgeog.2010.06.001

Schuurman, N., Hameed, S. M., Fiedler, R., Bell, N., & Simons, R. K. (2008). The spatial epidemiology of trauma: the potential of geographic information science to organize data and reveal patterns of injury and services. *Canadian Journal of Surgery*, *51*(5), 389-395.

Staiger, B., Friedrich, S., & Schütte, H.-W. (2006). *China: Lexikon zu Geographie und Wirtschaft* (1., Aufl.). Primus Verlag.

Stephenson, S., Henley, G., Harrison, J. E., & Langley, J. D. (2004). Diagnosis based injury severity scaling: Investigation of a method using Australian and New Zealand hospitalisations. *Injury Prevention*, *10*(6), 379-383.

Thomson Reuters Foundation. (2011). China earthquake 2008 - AlertNet. Retrieved October 27, 2011, from http://www.trust.org/alertnet/crisis-centre/crisis/china-earthquake-2008

USGS. (2011). Magnitude 7.9 - EASTERN SICHUAN, CHINA. Retrieved October 27, 2011, from http://earthquake.usgs.gov/earthquakes/eqinthenews/2008/us2008ryan/#summary

Wang, E. C., & Meng, Q. R. (2009). Mesozoic and cenozoic tectonic evolution of the Longmenshan fault belt. *Science in China, Series D: Earth Sciences*, *52*(5), 579-592.

WHO. (2007). *International Classification of Diseases (ICD) version 10.* World Health Organization. Retrieved from http://www.who.int/classifications/icd/en/

Xiong, H. (2008). *Encyclopedia of GIS*. Springer.

Zolala, F. (2010). Data collection after massive natural disasters (focusing on Bam earthquake, Iran). *Disaster Prevention and Management*, *19*(5), 541-547.

ANNEX

Table 8 List of all patients and locations in the DCD

Location names (Dd_Ha)	Patient count
Baiguo Village	4
Baimiao Village	2
Bajiao	1
Banqiao	10
Baoan Village	1
Baofeng Village	1
Baozhu Village	3
Bayi Village	1
Beidou Village	1
Boguo Village	1
Bolong	5
Changzhao Village	2
Chuanxindian Village	1
Dexin	2
Deyang	89
Dongbei	22
Donghu Village	1
Dumao Village	1
Fangquan Village	1
Fengdeng Village	1
Fushou Village	2
Fuxin	8
Gaoan Village	1
Gaobo Village	2
Gaolong Village	1
Gongxing	7
Guanghan	4
Guangji	61
Guangming Village	1
Guangping Village	4
Guilan Village	1
Guwang Village	2
Hanwang	704
Heping Village	1
Hongan Village_Luojiang	1
Hongbai	2
Honghuang Village	1
Huangjin Village	2

Huangxu	3
Huashigou Village	1
Huilong_Zhongjiang	1
Jiannan	13
Jinhua	7
Jinshan	2
Jintu Village	2
Jiuling Village	3
Jiulong	34
Jixian Village	1
Leping	2
Lianhe Village	1
Linfa Village	2
Longjing Village	1
Luojiang	3
Luoshui	8
Magui Village	1
Mawei	4
Mazu Town	1
Mianzhu	192
Niubizi Village	2
Penghua Village	1
Qianjin Village	1
Qinghe	3
Qinglong Village	1
Qingping	12
Qingquan Village	1
Qitian	6
Qunli Village	2
Renhe Village	2
Ronghua Village	2
Sanhe Village	1
Sanyuan Village	1
Shidi	3
Shifang	28
Shihe Village	1
Shimendong Village	2
Shuanglong	1
Shuangquan	2
Shuangtong	1
Shuangwa Village	1
Sifang Village	1
Songbo Village	1
Tanmu Village	2

Tianchi	3
Tianping Village	2
Tianyuan	6
Tongji	1
Tumen	33
Tuqiang Village	1
Wandeng Village	3
Woyun Village	2
Wudu Town	3
Wudu Village	5
Wufu	20
Wuxing Village	1
Xiaode	59
Xiaoquan	24
Xiayuan Village	1
Xieyue Village	1
Xinan	29
Xinglong_Mianzhu	2
Xinglong_Zhongjiang	21
Xinkai Village	3
Xinquan Village	2
Xinshi	2
Xintian Village	1
Xinzheng Village	1
Xuemen Village	1
Yanhe Village	1
Yanjia	1
Yinghua	3
Yongning Village	2
Yunhua Town	3
Yuquan	22
Yushi Village	1
Zhongjiang	3
Zhongxing Village	5
Zundao	52

AFFIRMATION

Unless otherwise indicated in the text or references, or acknowledged above, this thesis is entirely the product of my own scholarly work. This thesis has not been submitted either in whole or part, for a degree at this or any other university or institution.

Vienna, February 2012 Thomas Jakubicka

Curriculum Vitae

Personal Information

Name: Thomas Jakubicka

Date of birth: 18. April 1982

Place of birth: Munich

Nationality: German

Education

Geography studies at the Department of Geography and Regional Research, University of Vienna, 2002-2012

Secondary school St. Anna Gymnasium, Munich, 1992-2001

Primary school Klenzeschule, Munich, 1988-1992

Publications

D. Guha-Sapir, J.M. Rodriguez-LIanes, and T. Jakubicka, "Using disaster footprints, population databases and GIS to overcome persistent problems for human impact assessment in flood events", 2011

T. Jakubicka, F. Vos, R. Phalke, M. Marx, and D. Guha-Sapir, *Health impacts of floods in Europe, Data gaps and information needs from a spatial perspective*. Brussels, 2010