
Mapping Mine Hazard

Estimating the Global Threat by Explosive Remnants of War

Jonas Herzog



Mapping Mine Hazards
Estimating the Global Threat by
Explosive Remnants of War

by Jonas Herzog

Department of Geography
University of Zurich

Master's Thesis

Under supervision of Prof. Robert Weibel

In collaboration with the
Geneva International Centre for Humanitarian Demining (GICHD)

and

University of Geneva

October 2010

Jonas Herzog

Agnesstrasse 47

CH-8004 Zürich

jherzog@geo.uzh.ch

“Landmines are among the most barbaric weapons of war, because they continue to kill and maim innocent people long after the war itself has ended. Also, fear of them keeps people off the land, and thus prevents them from growing food.”

Kofi Annan, Former UN Secretary-General

Zusammenfassung

Als Folge vieler Kriege und Schlachten in den letzten Jahrzehnten wurden zahlreiche Regionen auf allen Kontinenten mit Landminen und nicht gezündeten Sprengkörpern verseucht. Deren verheerenden Auswirkungen beschränken sich nicht nur auf getötete oder verletzte Frauen, Männer und Kinder, sie stellen zudem eine ernst zu nehmende Bedrohung für Gesellschaft und Wirtschaft betroffener Länder dar. Die bei den betroffenen Ländern vorhandenen Daten beschreiben die Situation in den einzelnen Ländern im Detail, es fehlt jedoch ein globaler Überblick über die Bedrohungslage. Diese Daten sollen nun so verarbeitet werden, dass die Verbreitung auf globaler, nationaler und regionaler Ebene erkennbar wird und betroffene Länder miteinander verglichen werden können. Dabei muss darauf geachtet werden, dass die exakte Lage mancher Gefahrenherde nicht veröffentlicht werden darf, da die Datenlieferanten ein berechtigtes strategisches Interesse daran haben, gewisse Informationen nicht preiszugeben. Das Geneva International Centre for Humanitarian Demining (GICHD) verfügt über einige Datensätze, mit welchen Tests durchgeführt werden können und welche in dieser Arbeit verwendet wurden. Mittels Methoden der geographischen Informationswissenschaften konnten zunächst Flächen-daten von vier untersuchten Ländern (Afghanistan, Laos, Somalia und Zypern) in gewichtete Punktdaten umgewandelt werden und diese wiederum in ein Bedrohungsraster überführt werden. Je nach Gebiet unterscheiden sich die Daten bezüglich Qualität und Quantität stark. Nach der Aufbereitung und räumlicher statistischer Analyse dieser Datensätze wurde eine Kernel Density Estimation (KDE) durchgeführt, wobei die generierten Punkte je nach Relevanz gewichtet wurden. Die Einflussfaktoren der KDE wie die Grösse der Kernels und die Rasterweiten wurden variiert und zahlreiche Klassierungsmethoden getestet. Anschliessend wurden verschiedene Experimente für die Darstellung vorgenommen, die Resultate evaluiert und daraus konkrete Lösungsansätze erarbeitet. Mit geeigneten Klassierungsmethoden können nun Länder mit unterschiedlichsten Gefahrenpotentialen auf derselben Karte dargestellt und miteinander verglichen werden. Ausserdem wurden Wege beschrieben, wie vertrauliche Daten angemessen verschleiert werden können, ohne signifikante Informationsverluste in Kauf nehmen zu müssen. Die KDE stellte sich als eine adäquate Methode heraus, die Bedrohungslage in einer geeigneten Form zu veranschaulichen, gleichzeitig leistet sie einen zusätzlichen Beitrag zum Datenschutz. Die generierten Daten können in Zukunft als Basis für weitere Berechnungen im Kontext der Landminenbekämpfung verwendet werden, zum Beispiel können Schätzungen über negative gesellschaftliche und ökonomische Auswirkungen in den betroffenen Ländern verlässlicher und standardisierter erfolgen. Durch den richtigen Einsatz von Methoden der geographischen Informationswissenschaft und der verfügbaren räumlichen Daten konnte zum ersten Mal eine globale Visualisierung erstellt werden, welche die von Landminen und nicht detonierten Sprengkörpern ausgehende Bedrohung veranschaulicht. Diese wird ab nächstem Jahr mittels einer Web-Applikation für alle interessierten Personen zugänglich gemacht und sie soll so mögliche Spender, Behörden und Entscheidungsträger auf das Thema aufmerksam machen.

Abstract

As a result of many wars and battles in the last decades, numerous regions on all continents have been contaminated with explosive remnants of war, such as landmines and unexploded ordnances (UXO). They do not only leave behind killed and injured women, men and children, but also have a serious impact on the society and economy of affected countries. The data available at mine action authorities of the affected countries describe the situation in the individual countries in detail, but a global overview on the hazard situation is missing. These data will now be processed to make the distribution visible on a global, national, or regional level. Also, it will be possible to compare affected countries. When working with the data, in some cases one must be very cautious in order not to publish the exact locations of hazards, since data providers have a legitimate strategic interest in not revealing certain information. The Geneva International Centre for Humanitarian Demining (GICHD) has some datasets available for test purposes that have also been used in this thesis. Working with GIScience, it was possible to turn polygon data of the four studied countries (Afghanistan, Cyprus, Laos, and Somalia) into point data, which then could be converted into a hazard raster. The quality and quantity of the data can be very different depending on the conditions under which they have been recorded. After preparation and a spatial statistical analysis of the datasets a kernel density estimation (KDE) has been performed, where the points have been weighted according to their relevance. The parameters of the KDE such as the kernel size and the cell size have been diversified and numerous classification methods have been examined. Subsequently, miscellaneous experiments to find an adequate form of visualization has been conducted, the results have been evaluated and concrete solutions have been elaborated. With a suitable classification method, countries with very different hazard rates could be presented on the same map and compared to each other. Furthermore, ways have been described as to how confidential data can be obfuscated appropriately without loosing any significant information. The KDE turned out to be an adequate method to visualize the hazard situation in a useful form, while it also contributes to the non-disclosure issue. In the future, the generated data can be used as a base for other evaluations in the context within mine action. For instance, estimations of socioeconomic impacts on affected countries can be conducted more reliably and in a more adequate way. With the proper application of GIScience and the available spatial data, for the first time a global visualization that shows the global threat of landmines and UXO could be presented. Next year, this map will be made accessible to all interested persons and shall make donors, officials, and decision makers aware of this topic.

Table of contents

Zusammenfassung	V
Abstract	VII
Acknowledgements	xiii
1 Introduction	1
1.1 Motivation.....	1
1.2 Problem Definition.....	2
1.3 Constraints	3
1.3.1 Confidentiality	3
1.3.2 Software and hardware related constraints	4
1.4 Limitations	4
1.5 Definitions	4
1.6 Structure of the thesis	5
2 Background and State of the Art.....	7
2.1 Threats and Hazards.....	7
2.1.1 The landmine threat	7
2.1.2 The Threat of Explosive Remnants of War	9
2.1.3 The cluster munitions threat	10
2.2 Mine action	10
2.3 Information Management in Mine Action.....	11
2.4 State of the Art.....	13
2.4.1 Visualization of spatial phenomena	13
2.4.2 Spatial Interpolation	14
2.4.3 Spatial analysis and Mine Action	16
3 The Dataset.....	17
3.1 Countries	17
3.1.1 Afghanistan	17
3.1.2 Somaliland	18
3.1.3 Lao PDR	19
3.1.4 Cyprus.....	20
3.2 Format	21
3.3 Kinds of Hazards.....	23
3.4 Point Pattern Analysis	23
4 Methodology	27
4.1 Premises.....	27
4.2 From polygons to points	28
4.2.1 Comparing of different kinds of ERW	28

4.2.2 Overlapping polygons	30
4.3 Possible methods	30
4.4 Kernel Density Estimation (KDE)	31
4.4.1 How KDE works.....	31
4.4.2 The kernel method for multivariate data	34
4.4.3 Data processing.....	36
4.4.4 Kernel Size.....	37
4.4.5 Cell Size	38
4.4.6 Values.....	39
4.5 Classification.....	39
4.5.1 Classification methods	39
4.5.2 Number of classes	42
4.6 Projection	42
4.7 Neighboring countries	43
4.8 Obfuscation	43
4.9 Workflow	46
5 Results	47
5.1 Presentation of the generated maps.....	47
5.2 Examples.....	49
5.3 Modifying parameters	56
5.3.1 Varying kernel size.....	56
5.3.2 Classification methods	59
5.4 Histograms of the Density Maps.....	62
5.5 Combination with road data	65
6 Discussion	67
7 Conclusion	71
7.1 Achievements	71
7.2 Reflections / Critique.....	72
7.3 Future Projects	72
8 Bibliography	75
9 Appendix.....	81

List of figures

fig. 1: Blast mines.....	7
fig. 2: Fragmentation mines.....	7
fig. 3: Bounding and directional mines.....	8
fig. 4: Anti-tank mines (GICHD 2003, p.13)	8
fig. 5: The remaining of an unexploded cluster bomb in Lebanon. (www.mineaction.org 2010) .	10
fig. 6: Workflow of IMSMA (CSS 2005).....	12
fig. 7: Some examples for interpolation methods for a DEM from a set of points: (a) given data and Voronoi polygons, (b) TIN-based linear interpolation, (c) IDW, (d) Kriging. (Mitas & Mitasova 1999, p.483)	15
fig. 8: The distribution of the hazard point pattern in Afghanistan.....	18
fig. 9: The distribution of the hazard point pattern Somaliland	19
fig. 10: The distribution of the hazard point pattern in Lao PDR.....	20
fig. 11: The hazard point pattern of Cyprus	21
fig. 12: Example of the map pane in IMSMA ^{NG} showing hazardous areas, hazard reductions and accidents.....	22
fig. 13: A draft of the options data providers will have when generating rasters (courtesy of GICHD).....	29
fig. 14: Kernel estimate showing individual kernels. Kernel size 0.4 (Silverman 1998, p.13)	33
fig. 15: Kernel size: 0.2 (Silverman 1998, p.14)	33
fig. 16: Kernel size: 0.6 (Silverman 1998, p.14)	34
fig. 17: The Epanechnikov kernel (Silverman 1998, p. 42)	35
fig. 18. Kernel density estimation of a point pattern (Gatrell A.C. et al. 1996, p.260)	35
fig. 19: The Kernel Density dialog in ArcMap	36
fig. 20: The six classification methods provided by ArcGIS	41
fig. 21: Perturbation of a point pattern with a growing ε . The number indicates the size of ε in decimal degrees.....	44
fig. 22: Kernel density estimations with an exponentially growing amount of perturbation. The number indicates the size of ε in decimal degrees and the kernel size is 0.05 decimal degrees.	45
fig. 23: Workflow of the SERWIS project.....	46

fig. 24: A global overview of how the created layers could be presented (P. Lacroix & H. Santiago 2010).....	48
fig. 25: The interface of the web application, with example data.....	49
fig. 26: Afghanistan, regional level (not showing the current situation).....	52
fig. 27: Somalia, regional level (not showing the current situation)	52
fig. 28: Laos, regional level (not showing the current situation)	53
fig. 29: Cyprus, regional level (not showing the current situation)	53
fig. 30: Afghanistan, global level (not showing the current situation)	54
fig. 31: Somalia, continental level (not showing the current situation)	54
fig. 32: Laos, national level (not showing the current situation)	55
fig. 33: Linearly growing kernel size (in decimal degrees) for Lao PDR.....	56
fig. 34: Linearly growing kernel size (in decimal degrees) for Afghanistan	57
fig. 35: : Linearly growing kernel size (in decimal degrees) for Somaliland	58
fig. 36: 3D-plot of the Laos hazards, with an linearly growing kernel size (in decimal degrees) ...	59
fig. 37: Classification with k-Means clustering method with different number of classes	60
fig. 38: Growing number of classes with the equal interval method	61
fig. 39: Geometric Interval, various geometric series of class intervals	62
fig. 40: A histogram of the results from the four countries combined	63
fig. 41: A logarithmic histogram of all values combined (kernel size: 200 km)	64
fig. 42: Scatter plot on hazard vs. road distance	66

List of Tables

tab. 1: Results of KS-Tests and test for Poisson distribution.....	24
tab. 2: Statistical values of the generated raster maps	64
tab. 3: R ² values for the four countries.....	66

Acknowledgements

The completion of this thesis would not have been possible without valuable inputs, support and information provided by other persons.

I am deeply grateful to Robert Weibel of the University of Zurich for the uncomplicated mentoring and the extensive guidance and support.

I would like to thank Daniel Eriksson, Anne-Li Nauclér and Inna Radchenko of the Geneva International Centre for Humanitarian Demining (GICHD) and Pierre Lacroix and Helder Santiago of the University of Geneva, for the providing of data, software and valuable information on mine action, information management and GIS.

I'm very grateful to my father Adrian Herzog for his help with statistical and other problems that emerged while elaborating this thesis.

Also, I would like to thank Rada Leu, Rahel Nüssli and Peter Leibacher for the reviewing of the thesis and their important inputs.

Lastly, I would like to express my gratitude to all of those who supported me in any respect during the completion of this master thesis.

1 Introduction

1.1 Motivation

The many wars that have been waged on this planet inevitably left their marks on our world and on the people living in it. The impact of an ongoing war on economies, societies and people's daily life is tremendous. But even if the acts of war come to an end, there are still consequences that have to be dealt with. One of them is the problem of remaining landmines and other explosive remnants of war (ERW) that still pose a threat to persons living in war torn areas.

One can imagine that the problem of landmines and unexploded ordnances does not only affect potential direct victims, but also many other fields. Harris (2002) describes how clearance of explosive remnants of wars could improve the economy of an affected region: production can increase on cleared land, saved lives and fewer disabilities imply reduced costs for medical treatments in the short and long term. Travelling time is saved and transport costs are reduced, while costs of supporting refugees are minimized, etc. Many benefits are difficult to estimate, e.g. saved psychological costs, stemming from fear by those suffering of the threats. Furthermore, hazardous regions can have mutual effects on not affected areas, since they can be linked to each other or used to have important network functions.

Thanks to the development of applications based on Geographic Information Systems (GIS), database technology, and space-based navigation satellite systems (such as GPS), it became possible to collect accurate data of ERW hazards and store them in databases. These data can be used to make mine action much more effective, as long as it is processed and maintained in a reasonable way. Having only a large amount of information does not allow seeing the problem as it is manifested in reality, as long as the data are not handled correctly. An aim of this thesis is to process the available data to make the phenomenon visible. Although the ERW problem is quite well known, no attempts have been made to make these threats visible on a global scale. Also, a way has to be found to compare different regions and / or nations, whilst at the same time making sure to remain neutral, thus not emphasizing certain countries while neglecting others.

So far, there is no possibility to make these threats visible on a global scale that rests on a solid base respectively no solutions have been presented yet. Thanks to the collection of data in a standardized way, however, the option now exists to make the problem visible and comparable. This would allow decision makers to see where efforts have to be made. Donors would be able to see where resources are required and affected countries are encouraged to deliver data, which would allow a more efficient process of mine action.

On the other hand, non-disclosure has to be guaranteed, since in many cases data on hazards and hazard reductions are considered state secrets. Thus, the obfuscation¹ of delicate information cannot be disregarded.

The use of GIS in mine action is nothing new. Today, GPS, spatial analysis, and many other techniques of GIScience play an important role in mine action. Various programs and applications have been designed to make mine action more effective and to support decision-making processes. GIS can be employed at different levels to support mine action, for instance in coordination, strategic analyses, data management, but also to make estimates of impacts on an economic or sociological level.

It will be necessary to convert the available data into a map that represents the hazard contamination of the countries in the world. The conversion of data into other representations is an aspect as old as spatial analysis itself and its methods will play an important role in this thesis. Interpolation, density estimations, etc. are some of the basic topics in GIScience. Similar methods have been applied in various other fields of science.

There have been successful attempts to estimate probabilities of landmine contamination using spatial information. Knowing only about mined and mine-free land, it was possible to make a probabilistic forecast of landmine contamination in unknown areas. Riese (2001) tried to use data from Bosnia to make predictions which places are likely to be hazardous. Thus, forecasts can be made about the presence of landmines. Although other kinds of hazards are not included, the result is very impressive. Delhay et al. (2005) also made successful attempts to use GIS tools to support humanitarian demining. Their system helps demining campaign planners to take rational decisions and works with satellite imagery instead of a map. Another very innovative tool that has been enhanced over the past years is the Information Management System for Mine Action (IMSMA) (Eriksson 2008).

1.2 Problem Definition

The problem this thesis aims to solve is how to process available data into a format that makes a widespread global problem visible to any interested person. The presentation of the data has to be balanced in order not to disadvantage any affected country and the processing steps used to achieve the results should be transparent and comprehensible.

Creating a visualization that represents the distribution of hazards on a global scale poses some challenges:

1. The available data have to be processed so that different countries with different ERW related problems become comparable.

¹ "Obfuscation concerns the practice of deliberately degrading the quality of information in some way, so as to protect the privacy of the individual to whom that information refers." (Duckham & Kulik 2005, p.152)

2. An adequate method has to be found to convert this data into a representation that makes the problem visible. To achieve this, appropriate parameters have to be chosen in order to obtain an optimal solution.
3. All affected countries have to be visible, while still being comparable to each other.
4. Certain constraints regarding non-disclosure, software and hardware environment have to be respected.
5. The results have to be tested and statistical values have to be calculated, so that they can be compared quantitatively.

In a few words, the goal of this thesis is to use GIS methods to create a global map that shows ERW contamination. Claims for non-disclosure have to be respected, and the maps have to be adequate for presentation to a larger public.

In this master thesis knowledge about mine action is combined with GIS methods to obtain optimal results. First-hand information about the terminology, practice and background information of mine action are crucial in order to successfully accomplish this thesis. Hence, the thesis has been developed in collaboration with the Geneva International Centre for Humanitarian Demining (GICHD)², which assisted with data, know-how, and other support.

1.3 Constraints

1.3.1 Confidentiality

A very important issue of mine action is the confidentiality of the data. The locations of mine-fields can be a crucial strategic issue for affected countries. Especially, information on border regions is not likely to be published, in order to maintain security for the country. For this reason the generated maps will not show any vector data. Respecting delicate information represents a very central part of this thesis. A way to obfuscate certain information has thus to be found.

The created maps should encourage affected countries and their mine action authority to provide data that can be used for further studies. To encourage them to do so, it must be guaranteed that no confidential data are being published. Of course data providers are being motivated to provide as much data as possible, but they have to be sure to maintain total control over how spatially explicit data will leave their hands. The choice of the degree of detail has to be set easily and the data providers should be completely aware of which information they are actually delivering.

² “The GICHD is an international expert organisation, legally based in Switzerland as a non-profit foundation. It was established by Switzerland and several other countries in April 1998. In March 2003, the GICHD concluded a status agreement with the Swiss Government guaranteeing its independence and freedom of action. The Centre has over 50 staff members and is financially supported by more than 12 countries and international organisations.” (GICHD 2010b)

1.3.2 Software and hardware related constraints³

The calculations for the resulting raster maps are processed on the client side, hence this step has to be carried out by authorities or mine action authorities in the affected countries. Consequently, circumstances on-site must be respected. Furthermore, this means that data providers with no or little GIS knowledge must be able to process calculations. This requires an understandable, user-friendly interface. On the other hand, it must be taken into account that most users are working with standard hardware and might not have large capacity computers, as they are available in universities, international organizations, or enterprises. In addition, data must be processed in a fixed time frame, because in some areas blackouts may occur at any time.

Moreover, analysis will be done with ESRI ArcGIS Engine Runtime with Spatial extension since these are the only commercial products available for non-governmental organizations, meaning that only their functionalities can be used. The entire processing workflow has to be automated such that operating staff need only execute the program and the data are processed to create the maps that are sent to the GICHD. While working on this project, other products also have been used or at least tried out.

1.4 Limitations

It must be stressed that there is no intention to develop maps that can be used in reality to show which places are safe and which ones should be avoided. No map that is created in the process of this thesis can be used for navigational purposes at the local scale, or should be used on site to guide and manage the mine action process. The focus of this thesis lies on developing methods to create a basis for ERW hazard at the regional to global scale, and analyze the results of mine action critically.

1.5 Definitions

The following terms and expressions will be used in this thesis. Some of them are mine action related, others stem from geographical information science.

- ERW: Explosive Remnants of War (ERW) means all kinds of explosive ordnances that are left behind in the post-conflict environment. The term includes all forms of conventional explosive weaponry – anti-personnel mines, anti-tank mines, mortar bombs, rockets, grenades, artillery shells and air-dropped bombs.
- Mine action: Mine action stands for any kind of activities related to removing or minimizing the threat emanating from explosive remnants of war.

³ Based on personal communication with Daniel Eriksson and Anne-Li Nauclér (April 26, 2010 at GICHD)

- Hazard: The negative effect emanating from any kind of explosive remnants of war on humans, societies or economies.
- Hazard value: This virtual value represents a level of threat emanating from ERW calculated with density estimation (see section 4.4).
- SERWIS project: SERWIS stands for Service Explosive Remnants of War Information System that represents the project in the framework of which this thesis has been developed, launched by the GICHD.
- Density estimation: Density estimation represents a fundamental concept used in statistic to approximate unknown values of a given function (see section 4.4).
- GIScience (or geographic information science): This academic theory stands behind the development, use, and application of geographic information systems. The key issues are collection and measurement, statistics, modeling and theory, algorithms, processes, display, analysis of spatial data (M. F Goodchild 1992).
- ESRI ArcMap: ArcMap is the main component of Esri's ArcGIS suite that can be used primarily to view, edit, create, and analyze geospatial data. ArcMap allows the user to explore data within a data set, symbolize features accordingly, and create maps.

1.6 Structure of the thesis

The remainder of this thesis is structured as follows:

Chapter 2 gives background information about mine action and an introduction into the different kinds of threats. Furthermore, an overview of the state of the art is given. In chapter 3 the dataset used in this study and the studied areas are presented. In chapter 4 the methods used are illustrated and explained. The results achieved by applying the methods are presented in chapter 5 and discussed in chapter 6. Conclusions are drawn and future research possibilities are described in chapter 7.

2 Background and State of the Art

2.1 Threats and Hazards⁴

In this chapter, I would like to introduce the problem I want to elaborate in this thesis. Landmines still are a cheap and efficient mean of war and represent a long-lasting problem to many countries in the world. Furthermore, explosive remnants of war (abandoned stockpiles of munitions and unexploded ordnances) pose a huge problem for post-conflict societies around the globe.

Three main threats can be distinguished, upon which I am focusing this thesis: the landmine threat, the explosive remnants of war threat, and the cluster munitions threat.

2.1.1 The landmine threat

Usually, landmines are explosive traps that are victim-activated regardless of whether the target is a person or a vehicle. Experts commonly divide anti-personnel mines into four categories: blast, fragmentation, bounding, and directional fragmentation. Some examples are shown in fig. 1 – 4.



fig. 1: Blast mines



fig. 2: Fragmentation mines

⁴ The content of this section is based mainly on the *Guide of Mine Action*, published by GICHD (2003), if not otherwise indicated.



fig. 3: Bounding and directional mines



fig. 4: Anti-tank mines (GICHD 2003, p.13)

Many landmines are not primarily built to kill soldiers, but at least to hurt them. An injured, suffering soldier can demoralize the troops more than a casualty and their cure can be much more expensive. Anti-tank mines (fig. 4) are designed to destroy tanks and armored vehicles, thus they are definitely lethal for humans.

Landmines have been used in dozens of wars in this and the last century and before. They were adopted in the American Civil War and were already then condemned as improper and barbaric. With the development of tanks in World War I, first anti-tank mines were introduced. In World War II landmines played an important role, as hundreds of millions were produced and many new types have been developed and enhanced.

After World War II anti-personnel mines were widely used in the wars in Korea and Vietnam and were responsible for a large number of casualties. In Vietnam, they have often been delivered aerially, which was very cheap and made it easy to deliver mines deep into enemy territories (Miguel & Roland 2006).

Although technology has advanced rapidly (e.g. self-destroying mines), the use of cheap low-tech anti-personnel and anti-tank mines is widespread in internal armed conflicts by both government forces and opposition groups. In Afghanistan, Angola, Bosnia-Herzegovina, Cambodia, Colombia, Ethiopia, Iraq, Mozambique, Nicaragua, Somalia, Sudan and many other war-torn nations, anti-personnel mines were widely used as part of a deliberate military strategy or simply to terrorize civilians or control their movements.

It is impossible to know the exact number of landmines in the world. Therefore, estimations are very vague. Numbers go up to hundreds of millions (ICBL 2009b), but this can only be guessed and some estimates are much higher. The numbers of casualties present a similar problem. The Landmine Report of the International Campaign to Ban Landmines (ICBL), a network of more than 1,400 NGOs, reported the numbers of casualties in the last decade. Between 1999 and 2009 73,576 casualties in 119 countries / areas were counted. The total number of survivors is not

known but estimated to be in the hundreds of thousands. In 2008, 5,197 deaths have been recorded, which is slightly less than the year before (5,473 in 2007) (ICBL 2009b).

But the landmine threat goes far beyond killing innocent people in mine accidents. Although after the success of the International Campaign to Ban Landmines, the laying of new landmines has been reduced, the presence of remaining landmines still has a large impact on economies and societies of affected countries. Especially children suffer from the consequences of wars and landmines. Surviving children face severe medical problems and the treatment can cost much more than an average person can earn in a developing country (Machel et al. 1996).

The Ottawa Convention

In December 1997, over 120 countries have signed a treaty that prohibits the use, stockpiling and transfer of anti-personnel mines and regulates their destruction. This convention led to a reduction of production and trade while the pressure on non-signatories has increased. Finally, it is estimated that already in 2001 thanks to this convention among other factors, at least less landmines are planted than are being taken out every day (Spearin 2001). By the year 2010, more than three quarters of the world's countries had become parties of this convention and an international norm against anti-personnel mines has been established.

2.1.2 The Threat of Explosive Remnants of War

The term 'explosive remnants of wars' refers to unexploded ordnances (UXO)⁵ and abandoned explosive ordnances. Explosive remnants of war (ERW) are munitions that have been abandoned or that have been used but which have failed to explode. Mines, booby-traps⁶ or other devices are not included. It is not possible to give an estimation of the number of explosive remnants of war. Nevertheless, their number is estimated much higher than the number of landmines. There are still ERW of the World Wars being discovered – for example in Belarus even from the Napoleonic Wars. One can realize that their threat deserves as much attention as the one of landmines.

The threat by ERW, which sometimes is subordinated to concern about the humanitarian impact of landmines, must not be underestimated. In some countries civilians collect items of ordnances for their value as scrap metal or even the explosive, and children may be killed or injured while playing with ERW. The consequences of not disposing ERW correctly often can be fatal (Vosburgh 2006).

⁵ Unexploded Ordnance (UXO) refers to munitions that have been used, but which have failed to detonate as intended. Failure rate can vary between 1 to 40 per cent, depending on many factors. (GICHD 2003)

⁶ A booby trap is an apparently harmless object containing a concealed explosive device designed to kill or injure anyone who touches it.

2.1.3 The cluster munitions threat

Cluster munitions are conventional weapons that are designed to release multiple submunitions, called ‘bomblets’ or ‘bombies’. The submunitions are dispersed over an area that may cover several hundred square meters. There is a heavy threat to civilian population emanating from cluster munitions during and after deployment. With the wide-area effect, the likelihood of civilian victims or collateral damage increases. The other problem is the high failure rate of many submunitions that leaves many small, unexploded but still lethal devices of submunition remains (fig. 5), which represent a great danger for civilians. The sensitive fusing system leads to explosions caused by minimal disturbances. Thus, submunitions have killed and injured many civilians in war-torn countries all over the world. Further, presence of clustermunition affects agricultural areas and endangers the safe return of displaced people.



fig. 5: The remaining of an unexploded cluster bomb in Lebanon. (www.mineaction.org 2010)

The Convention on Cluster Munitions of 2008 bans weapons that cause civilian casualties both during and after attacks (Docherty 2009, p.2). Until August 2010, this convention has been signed by 108 states and ratified by 38 of them. Nevertheless, important countries that produce or stockpile large amounts of cluster munitions, including Brazil, China, India, Israel, Pakistan, Russia, and the United States have not yet signed the convention (Burns 2008).

2.2 Mine action

Mine action is not only clearing land from landmines. It consists of the following five pillars:

- Removing and destroying landmines and explosive remnants of war and marking or fencing off areas contaminated with them.
- Mine-risk education to help people understand the risks they face, identify mines and explosive remnants of war and learn how to stay out of harm's way.
- Medical assistance and rehabilitation services to victims, including job skills training and employment opportunities.
- Advocating for a world free from the threat of landmines and encouraging countries to participate in international treaties and conventions designed to end the production, trade, shipment or use of mines and to uphold the rights of persons with disabilities.
- Helping countries destroy their stockpiles of mines as required by international agreements, such as the 1999 anti-personnel mine-ban treaty.

(UNMAS 2008)

Among other objectives, this thesis aims to support mine action by creating new visualizations of the problem that can be presented on a web application. The thesis should help to achieve these goals and therefore make mine action more efficient. However, the results of this study will not support all of the above pillars and cannot be expected to solve the problem completely, but hopefully they will provide valuable inputs and bring mine action one step further.

2.3 Information Management in Mine Action

Like in many other fields, mine action generates plenty of information. It is crucial to find a way to manage the knowledge gained about processes, locations, transformations, etc.

The data I was working with was collected with the information management tool called IMSMA (Information Management System for Mine Action). It was developed to make mine action safer, faster, more effective, and more efficient. Requested by the mine action community, this tool was developed with the Centre for Security Studies (CSS) of ETH Zurich in collaboration with the GICHD, where it is now being enhanced (GICHD 2009).

IMSMA is a self-contained information system developed specifically for humanitarian demining centers in ERW affected countries. It combines a powerful database with a geographical information system, which is used for many tasks: planning, managing reporting and mapping the results of survey and field data collection; reporting and mapping of mine, UXO and other ERW threats; recording and reporting of clearance activities (CSS 2005). Moreover, it helps managers to track the progress of their work and to analyze and support decisions for prioritization. Currently, IMSMA is in use in more than 80 % of mine action programs around the world and is the United Nations preferred information management system for mine action (GICHD 2009).

Figure 5 shows the IMSMA workflow, developed by CSS in cooperation with GICHD.



fig. 6: Workflow of IMSMA⁷ (CSS 2005)

IMSMA supports operations by providing ways to record, display and analyze the results of field operations (Bollinger n.d.). Users are allowed to enter any ERW related kind of data by using mobile devices. Hazardous areas can be recorded as points or polygons.

There are different types hazard reductions, methods or activities that identify or reduce the effect or impact of a hazard:

- Impact survey, an assessment of the socio-economic impact caused by the actual or perceived presence of ERW, in order to assist the planning and prioritization of mine action programs
- Non-technical survey, survey activity that involves collecting and analyzing new and/or existing information about a suspected hazardous area. A non-technical survey does not normally involve the use of clearance or verification assets.
- Technical survey, a detailed intervention with clearance or verification assets. It should confirm the presence of ERW leading to the definition of one or more defined hazardous area and may indicate the absence of ERW, which could allow land to be released when combined with other evidence.
- Post clearance, surveys to assess the effectiveness and efficiency of mine action planning, priority setting, and implementation processes, aiming to enhance the productivity and effectiveness of mine action, monitor post-clearance land use, ensure priority-setting processes are clear, transparent and carried out correctly, and help identify problems faced by communities in transforming the outputs of mine action (e.g. cleared land) into sustainable developmental outcomes.

⁷ This figure contains the outdated symbology, which has been updated in 2006.

The value of this system was first verified 1999 in Kosovo: until 2001 it supported successfully the complete clearance of this country. The program manager confirmed:

"Information is a vital component of mine action. During the successful clearance operation in Kosovo, IMSMA enabled us to rapidly collate and analyze an enormous amount of data. This in turn helped us to plan and prioritize clearance effort, and assisted with the integration of other activities such as mine awareness education. Throughout the entire mine action program in Kosovo, IMSMA was constantly used to manage the ongoing operational activities, and without it, our task would have been much more difficult."

John Flanagan, Program Manager UN Mine Action Coordination Center, Kosovo (CSS 2005)

With this system, costs of demining operations are reduced and standardized data can be collected more easily. The system's users maintain control over the data at the country level and IMSMA allows to publish the collected data.

The data are not stored in a worldwide repository. The individual mine action programs retain the control over the data and decide if they want to provide them or not. For this thesis it was possible to obtain data of some affected countries. The available data consist of polygons and points of minefields, battle areas, etc. These data can be exported from the database. On this base, methods have been developed to create maps.

In 2006 an updated version of IMSMA was published, called IMSMA^{ng} (Next Generation), which provides many helpful updates. This information management system is regularly updated and user support is provided (GICHD 2010a).

2.4 State of the Art

2.4.1 Visualization of spatial phenomena

Since ERW are usually not viewable or detectable with remote sensing or similar imaging methods (unlike land cover, elevation, clouds, etc.), a representation has to be created that makes them visible. Working with a large set of spatial data, where information on x- and y-coordinates, as well as a z-value is available, a way to represent these data has to be found. The difficult part is to find the mapping function (Keim et al. 2004). Slocum et al. (2008) suggest a wide spectrum of visualization methods, including interpolation, choropleth mapping⁸, symbolization, etc. For example, it would also be possible to calculate hazard values for regions and generate a choropleth map, which would show their distribution by areal units (e.g. administrative districts).

There are almost infinite ways to represent spatial phenomena numerically and graphically; the hard part is to find a suitable solution that satisfies the requirements stated in Chapter 1.

⁸ A choropleth map is a thematic map that shows a generalized depiction of quantitative area distributions (Peterson 1979).

2.4.2 Spatial Interpolation

Using GIS methods to estimate unknown values based on events or a set of points is nothing new. Spatial analysis has always been a very important part of GIScience and it provides various possibilities for the examination of data (Michael Goodchild et al. 1992). In the last decades, several methods have been developed to provide solutions for this problem, such as interpolation (Flowerdew & Green 1992; Journel & Huijbregts 1978; Mitas & Mitasova 1999). Mitas & Mitasova (1999) discuss the most common approaches (fig. 7). There is the local neighborhood approach that supposes that any point influences the result up to a certain distance. The inverse distance weighted interpolation (IDW), the natural neighbor method, or the interpolation based on a triangular irregular network (TIN) are some examples with this background. Another related method is the Voronoi approach, which is one of the most fundamental data structure in computational geometry. A plane is divided by Voronoi diagrams according to the nearest neighbor rule, which associates each point to the region of the plane closest to it. These planes can than be used to interpolate between the points (Aurenhammer 1991). On the other hand there is the geostatistical approach, which works with the concepts of Kriging, as described by Journel & Huijbregts (1978). A typical example where these methods can be applied is the interpolation of elevation models. Here, some known, arbitrarily distributed points of altitude are taken to estimate a digital elevation model, which provides dense estimates of elevations, for instance in grid arrangements.

Another important approach is the kernel density estimation (KDE). Usually this method is applied to smoothen histograms, while working with one-dimensional data. However, it can also be applied to spatial, two-dimensional data (Silverman 1998) (see chapter 4.4). The important difference between the interpolation methods discussed above is that the former are intended to work with continuous data (i.e. data measured at an interval/ratio scale) or possibly ordinal data, while KDE is restricted to the interpolation of counts data. Still, KDE is a method that is appropriate for individual point locations, which will be processed within this thesis.

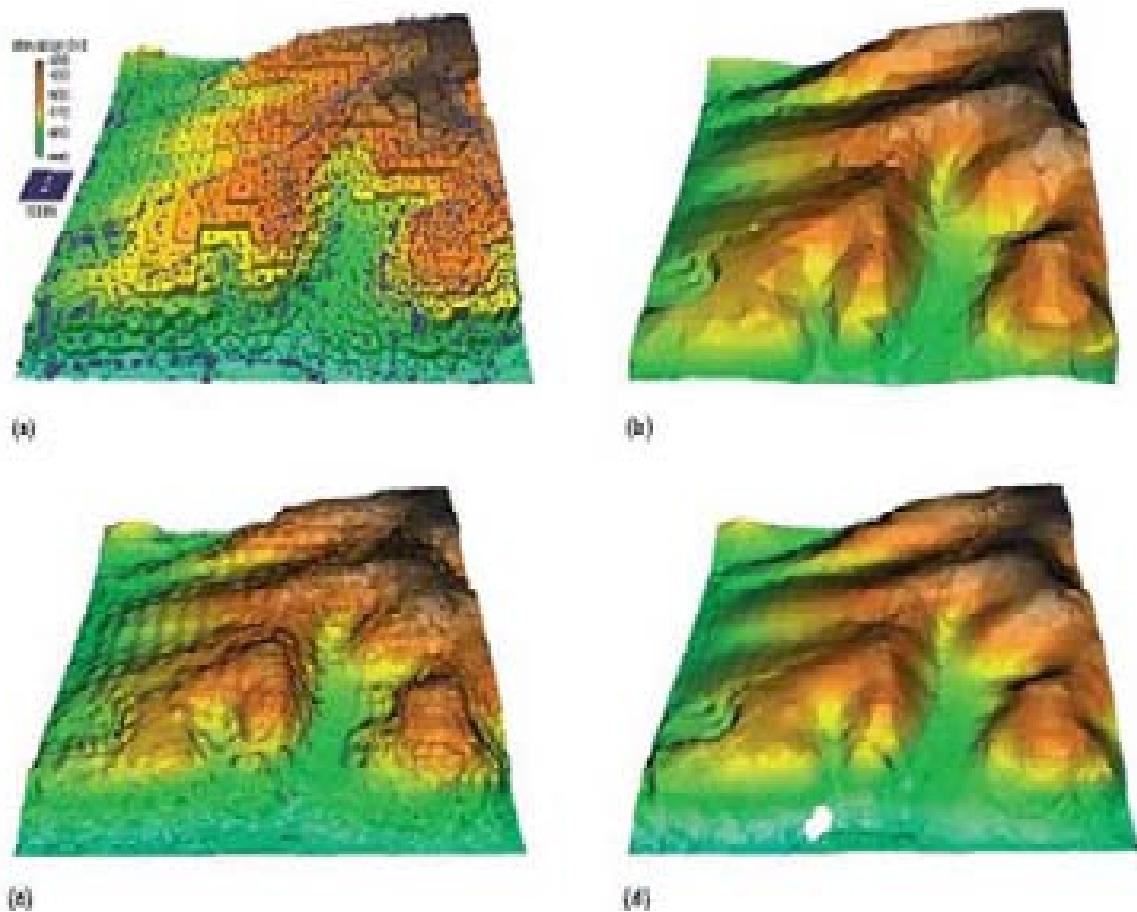


fig. 7: Some examples for interpolation methods for a DEM from a set of points: (a) given data and Voronoi polygons, (b) TIN-based linear interpolation, (c) IDW, (d) Kriging. (Mitas & Mitasova 1999, p.483)

The methods mentioned above have been applied in various fields. Especially the kernel density estimation, which plays an important role in this thesis, has been used in a large spectrum of fields of research. To mention a few of them: crime mapping (Boba 2005; Chainey & Ratcliffe 2005; Manzoni 2008; Levine 2006; Eck et al. 2005); public health (Rushton 2003; Cromley & McLafferty 2002) traffic accidents (Xie & Yan 2008), home range analysis (Seaman & Powell 1996; Worton 1989); marketing research (Donthu & Rust 1988; Donthu & Rust 1989). This list could be extended with many other topics.

In crime mapping for instance, reliable packages have been developed that allow analyzing crime patterns, mapping clusters of crime occurrence, measuring geographic distributions, etc. (ESRI n.d.). Crime mapping has much in common with the mapping of ERW contamination. The aim of both applications is to find concentrations of a hazard and to make it visible. The geographic distribution of the point pattern has an immense impact on the output and connections can be unveiled that are not visible at first glance. Nevertheless, a main goal of crime mapping is to find certain hot spots, while in this thesis the attention is turned to the overall concentration as well as the comparison between different countries or regions. Moreover, in both applications the presentation of the results can have a political impact.

An example – Costumer distribution in markets

Donthu & Rust (1989) used KDE already in the late 1980ies, when spatial analysis of this kind was still in its infancy, to evaluate the customer distribution in markets. The aim was to find out where to locate new stores, where high customer concentrations could be found, etc. They used data on hospitals and population density to test their methods. Using KDE allowed finding the best spots to plan additional hospitals, as the density map generated by KDE showed where population density was high, with no hospitals close by. They concluded that:

“Kernel density estimation provides an attractive method for estimating flexibly the geographic distribution of customers. Whatever the shape of the underlying density surface, kernel is capable of fitting that shape, given a large enough sample size.” (Donthu & Rust 1989, p.201)

This gives an indication of the strength of this method and manifests its many different application fields. But it has to be remembered that for any application the right method has to be found and for any method the parameters have to be chosen appropriately in order to obtain the result as desired.

Although some of the above topics are quite different to mine action, they deal with a very similar problem. Single events and observations are used to make estimates for unknown areas and visualize a spatial problem. Thus, some results and ideas of this thesis could also be applied to other fields of research, while efforts already made by other researchers can be used to achieve this thesis' goals. The achievements of this thesis could provide interesting new inputs for further projects.

2.4.3 Spatial analysis and Mine Action

As mentioned before, attempts have already been made previously to estimate ERW threats in affected countries spatially. With GIS and statistical techniques, it became possible to analyze hazards numerically and spatially. Riese (2001) tried to create a forecast of the probability of landmine presence with real data of Bosnia, collected by the U.S. Army. Based on information of contaminated areas, examples of areas free of mines, feature data as well as an initial estimate of the overall mine contamination of the area a probabilistic model was created. This model was able to make sufficiently accurate forecasts of the presence of landmines. Unfortunately, no more attempts have gone into a similar direction. Also, it is not very common to have data available at this fine level of detail for many other countries, which would make a comparison very difficult.

Since so little has been done in this direction besides few previous studies, this MSc-project is legitimate and represents a large potential to gain new insights into GIS-supported mine action. Because it is one of the first of its kind, there are many obstacles that had to be overcome and I had to make a lot of effort in order to get acquainted with the topic.

3 The Dataset

In order to get better acquainted with the data used, I will give an overview about the investigated countries and their attributes. To better understand the condition the spatial distribution has been analyzed with statistical tests as well.

3.1 Countries

The results of this thesis are based on data provided by the GICHD, mainly on four different countries: Afghanistan, Somalia, Cyprus and Laos PDR. These countries have been chosen because each of them represents a particular example of how a country can be affected. Also, the data of these states are not as confidential as for others, and can therefore be displayed and presented up to a certain level of detail. Now follow portraits of the four affected countries and the motivation as to why they have been selected.

3.1.1 Afghanistan

The Islamic Republic of Afghanistan is a landlocked country in south-central Asia with an estimated population of almost 30 millions. It lies along important trade routes from southern and eastern Asia to Europe. Its economy and society have been affected heavily by the many wars fought on its territory in the last decades (Rubin 2000).

Afghanistan is one of the countries most affected by injuries caused by landmines and UXO in the world. During 2000 and 2001 in no other country the number of reported casualties were higher (Bilukha et al. 2003, p.516). The most of the mines were laid during the soviet occupation between 1978 and 1989 and the subsequent civil war. The air strikes of the coalition troops in 2001 provided a new problem: the cluster munitions. Among the victims are a lot of children, because a lot of them mistook the yellow bright not exploded submunitions for the humanitarian daily rations, dropped earlier during the conflict. 81 % of the injured were civilians and 45.9 % of them were 16 years or younger (Bilukha et al. 2003).

The impact of ERW on this country is huge and quantitative estimations have shown how landmines affect daily activities in Afghanistan and in other countries. (Andersson et al. 1995)

Afghanistan represents a large landlocked country with landmines and UXO, where data are widely available. The vegetation and the land cover are very diversified and affected areas are around roads, settlements, borders, former battle areas, etc. The hazards originate from a very broad period of time, namely from the 1970ies until today and still there are combat operations going on in some parts of the country down to the present day.

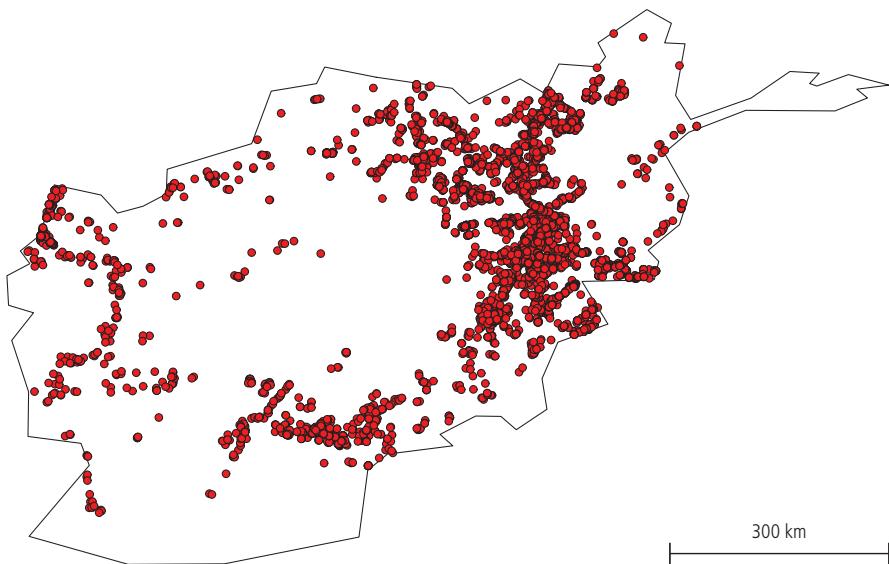


fig. 8: The distribution of the hazard point pattern in Afghanistan

In fig. 8, the pattern of the known hazards can be seen (for security reasons it is not possible to collect data from some parts of the country). Every point represents a center point of a minefield or another threat of any size. Thus, it does not allow draw inferences on the actual contamination, but the main concentrations can be spotted. Eventually, Afghanistan was chosen because of its large size and of the detailed available data.

3.1.2 Somaliland

The data used in this thesis is limited to the extent of Somaliland, the self-declared republic in the north of Somalia. In 1991, from the traditional tribal mode of governance emerged a representative democracy and in 1997 the first provisional constitution was established. (Farley 2010)

From the 1960s to today, this country located at the Horn of Africa, has been war-torn for many years. The main part of the mines have been laid during the Ogaden wars⁹ in 1964 and 1977, the 1988 – 1991 civil war, and during an internal conflict in 1994/1995 (Lardner 2008).

Somaliland was chosen because many affected areas are in border regions, which can lead to inconsistencies with bordering countries. The most affected areas are close the neighbor countries Eritrea and Ethiopia.

⁹ This war between Ethiopia and Somalia was one of the biggest conflicts in contemporary times in Africa. It is one of the many proxy wars during the Cold War. Among others, the Soviet Union changed from supporting Somalia to Ethiopia, which prompted the United States to do the contrary. The war ended with a withdrawal of the Somali army and a truce was declared. (Tareke 2000)

Hazard Distribution in Somaliland

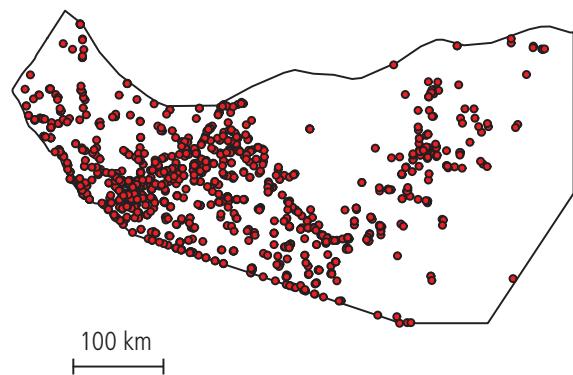


fig. 9: The distribution of the hazard point pattern Somaliland

In fig. 9, one can see that the hazards are distributed differently than in Afghanistan. Although a concentration in the western part of the country can be seen, the hazards seem to be much more irregularly distributed. Furthermore, a large concentration along the southern border cannot be ignored.

3.1.3 Lao PDR

The Lao People's Democratic Republic is a landlocked country in South-East Asia surrounded by Thailand, Vietnam, Cambodia, Myanmar and China. It has the distinction of being per capita the most heavily bombed nation in the world (Sisavath 2006). During the Indochina conflicts in the 1960s and 1970s, Laos has been affected by heavy bombardment and ground battles. From 1964 until 1973, over half a million bombing missions were carried over the country and more than 2 million tons of bombs were dropped, mainly by US Air Force operating from secret bases within the country (Hamilton-Merritt 1999). The mine action NGO UXO Lao estimates that up to 30 percent of all ordnance failed to explode on impact. These not detonated objects still remain a threat for over 30 years (UXO Lao 2009).

This large contamination resulted in over 13'000 ERW-related casualties since 1975 and wide areas of the country have become unusable for agriculture. Like in other countries, there is a huge socio-economic impact that affects at least 25 % of the total villages. Due to UXO many people die or get injured decades after the war (Sisavath 2006).

Lao PDR is a good example for a very imbalanced ratio of UXO and landmine contamination. The data used in this thesis covers not the complete dataset, but retains the spatial distribution. It represents a medium sized country, mainly affected by UXO. In Laos, the most of them are

submunitions from cluster bombs that spray shrapnel up to 300 m and have a lethal radius of 150 m (Morikawa et al. 1998).

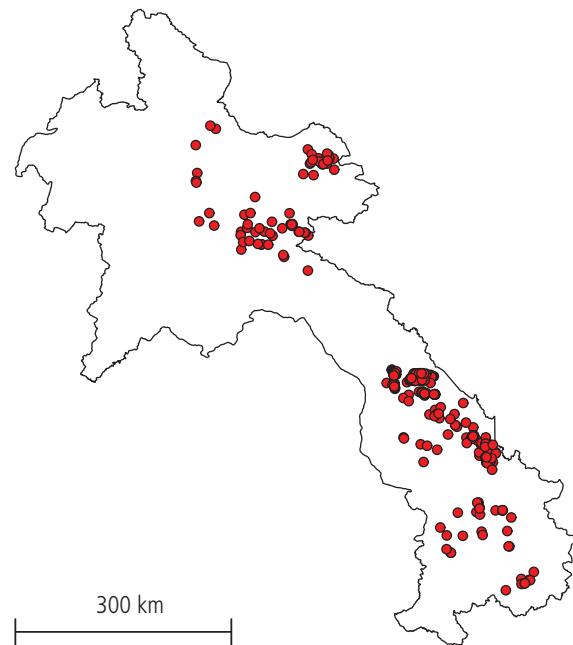


fig. 10: The distribution of the hazard point pattern in Lao PDR

As shown in fig. 10, in Laos the distribution is not as clustered as in other countries, but it is concentration is higher in the western and northern part of the country. The most of them are concentrated in the northern and eastern provinces of the country, where the Ho-Chi-Minh-Trail¹⁰ and other important positions of the North Vietnamese Army were situated.

The Lao government estimates the contaminated area as more than 50 % of the whole country area (Legacies of War 2010). Other sources estimate the percentage much higher: Up to 80 percent of agricultural land is affected mainly by UXO, arisen from the bombardments of the U.S. Air Force according to Kyung (2009). Since the collection of data is not complete, the threat in this country is highly underestimated. In spite of this known fact, only available and reliable data can be considered for this use.

3.1.4 Cyprus

Affected countries are not only to be found in the developing world. Countries in Europe are also affected, some still by landmines or ERW of the World Wars, like France or Belgium.

The Republic of Cyprus is an island in the Eastern Mediterranean that has been divided into two parts by a buffer zone in 1974, as a result of the ongoing conflict between Greek and Turkish Cypriots. This zone is controlled by the UN Forces in Cyprus (UNFICYP) and occupies three

¹⁰ “A covert system of trails along Vietnam's western frontier, a major supply route for North Vietnamese forces during the Vietnam War.” (New Oxford American Dictionary 2001)

percent of the island, including valuable agricultural land. During the conflict, both parties laid defensive minefields within and outside of the buffer zone (ICBL 2009a). Therefore, the landmines are concentrated around this buffer zone and the contaminated areas are more or less known.

Generally, the data are more reliable than those of other affected countries, but they have to be processed likewise. The hazards are mostly minefields, the vast majority concentrated in and around the buffer zone, which can be seen in fig. 11. It was selected mainly because of its size, since very small countries can be a challenge for representation on a global map.



fig. 11: The hazard point pattern of Cyprus

3.2 Format¹¹

The IMSMA database consists of more information than hazardous areas that could be of interest when mapping the ERW hazards in a country:

- Hazard reductions done indicates where there have been a hazard present (clearance and different types of surveys (see 2.3))
- Accidents also give an indication where hazards are. To generate the rasters in this thesis, they have not been taken into account.

Hazard reduction will be used for another layer on field activities, while accidents and the other information in the IMSMA database could not be considered in this thesis, but can be used for other purposes not discussed in this thesis.

¹¹ Based on personal communication with Anne-Li Nauclér (15. Sept. 2010)



fig. 12: Example of the map pane in IMSMA^{NG} showing hazardous areas, hazard reductions and accidents.

In fig. 12 a situation of recorded ERW data are illustrated, how it is displayed in IMSMA^{NG}.

The collected data have geographical references that are points, lines and/or polygons. The size of the polygons varies from few square meters up to some hundreds of hectares. The emphasis of this thesis lies on hazards emanated from mines and UXO; therefore the hazard records are used for calculations. Hazards are stored mostly as polygons.

When retrieving the data from the database with several SQL queries (Gennick 2008) the GICHD has made sure that each hazard is only represented once in the calculations. If the hazard had a polygon then

- the centroid of the polygon is used, else
- the benchmark¹² is used, else
- the landmark or the reference point¹³ is used.

This information is stored in local databases, from which the mine action authorities in affected countries can query the required records. To generate the rasters, only the hazard records are taken into account.

¹² Benchmark: a fixed point of reference used to locate a marked and recorded hazard or hazardous area. It should normally be located a short distance outside the hazardous area.

¹³ Landmark and Reference point: A fixed point of references, some distance outside the hazardous area. It should be an easily recognizable feature (such as a cross-roads or a bridge), which can be used to assist in navigating to one or more benchmarks.

3.3 Kinds of Hazards

There are many different kinds of landmine or ERW related hazards:

- Minefields
- Battle areas
- Dangerous areas
- Stockpiles
- Abandoned areas
- And many others

Each one of them represents a different threat, but since only the degree of hazard and not the kind of hazard is of general interest, other ways had to be found to compare them. Since the aim of this thesis is to map the general ERW contamination, their qualitative difference does not have a crucial impact on the results.

3.4 Point Pattern Analysis

In order to better understand the distribution of the points that will be used for calculations, a point pattern analysis has been carried out. The analysis of spatial point patterns has been applied in many fields, for example in plant ecology (Dale 2000), available and exploited resources (Lancaster & J. Downes 2004), zoology (Diggle 1986), etc. One of the first applications of this kind was the classic work of John Snow in Victorian London that linked the clustering of cholera deaths around a water pump to the source of infection (Snow et al. 1965; Gatrell A.C. et al. 1996).

A first approach is to test the distribution of the points for spatial randomness. Performing a Kolmogorov-Smirnov-Test¹⁴ a value is displayed that tells how randomly the points are distributed. KS-Tests have been conducted for the point patterns of all four countries in focus. Using the statistical software environment R and the package Spatstat to analyze spatial point patterns, the data could be examined in detail (Baddeley & Turner 2005).

A Kolmogorov-Smirnov-Test of Complete Spatial Randomness (CSR) has been performed and the likeliness of being Poisson-distributed (PPM) has been examined. The latter has been described by Berman (1986). The results are presented in tab. 1.

¹⁴ The Kolmogorov-Smirnov statistic tests whether a set of observations is from some completely specified continuous distribution. An alternative is provided by the chi-square test. (Lilliefors 1967)

	Spatial Kolmogorov-Smirnov test of CSR	PPM	Number of available records	Estimated overall hazard rate¹⁵
	D-Value	D-Value		
Afghanistan	0.4678	0.2481	6644	741.783
Somaliland	0.3885	0.1198	1183	770.938
Laos	0.4537	0.2170	285	0.292
Cyprus	0.1817	0.2236	94	1.392

tab. 1: Results of KS-Tests and test for Poisson distribution

The D-value of the Spatial Kolmogorov-Smirnov-Test of CSR shows how randomly the points are arranged. A value of 1 indicates complete spatial randomness, while low values down to zero suggest a very clustered distribution. If the value is higher than 1, a regularity of the point pattern can be assumed. It can be seen that the randomness of the points in Afghanistan and Laos is significantly higher than those of Cyprus, while Somaliland is somewhere in between. In Cyprus the hazards are highly concentrated around the border, while the other countries (especially Laos) have been affected also by air bombing, which tend to a less regular point distribution. Of course, this can only be guessed and the sample is quite small, but it becomes very obvious that the point patterns can vary heavily. Further, in the third row of tab. 1, it can be seen that the number of records is much larger in Afghanistan with more than 6000 points than in Cyprus with less than 100 and in the fourth row an overall hazard rate has been calculated.

The D-values of the test of Poisson distribution (PPM) are low for all countries, which suggests that they are not distributed regularly and not Poisson-distributed. Hence, this allows to draw inferences on how mines and UXO have been deployed in space, or also on how they have been recorded. This is because hazard information is usually not collected at random locations all over the country, but in regions that are accessible and more or less secure.

The values found with these analyses could help finding an optimal bandwidth that depends on the size of the clusters that are aggregated. Since the values are different in many countries, it is not possible to find a universal bandwidth that fits every point pattern. Hence, a compromise had to be found that meets the requirements.

Another fact that should be considered is the influence of terrain and accessibility on the pattern of hazards. Usually minefields are laid in strategic areas, while remote areas are not affected. There are numerous factors that have an influence on these patterns, which can be different in every country. Some key factors could be attributes like proximity to cities and other strategic points, the presence of country borders, the vegetation, the proximity to war zones. To find all these factors and estimate their influence is almost impossible and they are varying in every country. (Benini et al. 2002)

¹⁵ This theoretical value has been calculated by dividing the total weighted values by the area of the country. It illustrates the crucial differences in the available datasets, but it says nothing about the actual threats in reality.

The analysis of the point patterns shows that minefields and other ERW, as anticipated, are not randomly deployed. In some cases, the points are heavily clustered or ordered in some way, mostly depending on the strategic situation and the terrain. By analyzing point patterns inferences can be drawn on how the hazards have been laid and what the influencing factors might be.

4 Methodology

In this chapter I would like to give some information on the developed methods and on how the maps representing hazard contamination have been created. The data used for this thesis was very heterogeneous, as described in the previous chapter. There are some countries where the positions of hazards are quite well known, while in some of them large areas are considered as contaminated with no reliable information. Thus, the available data had to be processed into comparable units. Available data exist not only in the form of points, but also mainly in polygons depending on how data have been collected.

Since targeted readers of this thesis are mainly not GIS-experts, basic information on classification techniques, data handling, and other GIS-related topics are also explained in this section.

4.1 Premises

The aim of the method is to generate maps that describe the ERW threats spatially and adequately. Under certain circumstances this method can be used for other data, such as mine action activity or others. Considering this and also the fact that the records are very heterogeneous, rasters are generated that can be compared and are also a good way to face the problem of obfuscation.

Since the data will remain at the mine action authorities of the participating states, responsible persons in affected countries will execute all the calculations by themselves. They will have the possibility to choose how much information will be published. It is guaranteed that data providers have full control on how much and which data will be delivered. Giving them this possibility, it is more likely to receive data on more countries than to ask them to hand over all available datasets. Of course data providers also have the possibility to transmit all their data in vector format, in which case processing will take place at the GICHD.

In any case, a method had to be developed that is executed under the affected countries authority and whose functionality is transparent and understandable to the data providers. The developed method has to be working locally and under any given circumstances.

4.2 From polygons to points

The data used to create the maps consist basically of polygons that each represent a recorded hazardous area. To estimate densities, there are no common methods that work with polygon data. Only point or line features can be processed with the methods provided. In order to apply density estimation methods on the available data, they first have to be transformed from polygons to points, without losing too much important information.

Given that the final resolution of the representation cannot be too high (i.e. up to 200 m) due to confidentiality restrictions, and as the recorded polygons are relatively small, it was decided to assign to every polygon its centroid. This point is then used for calculations and represents the polygon. The sizes of the polygons are varying, which means that just using these centroids would overestimate small polygons while underestimating large ones. Thus, the area of the polygon was assigned to every point as well. Since the generated maps will not be used to find the exact locations of threats (e.g. to find a safe way) and confidentiality has to be guaranteed, working with the centroids is.

4.2.1 Comparing of different kinds of ERW

In order to include different ERW in the model they had to be made comparable. Although the data from different countries were collected with the same tool (IMSMA^{NG}), there are differences and local customizations done in each country. Therefore, the parameters may not be static and the GICHD is in the process of creating a highly flexible client application that can be seen in fig. 13).

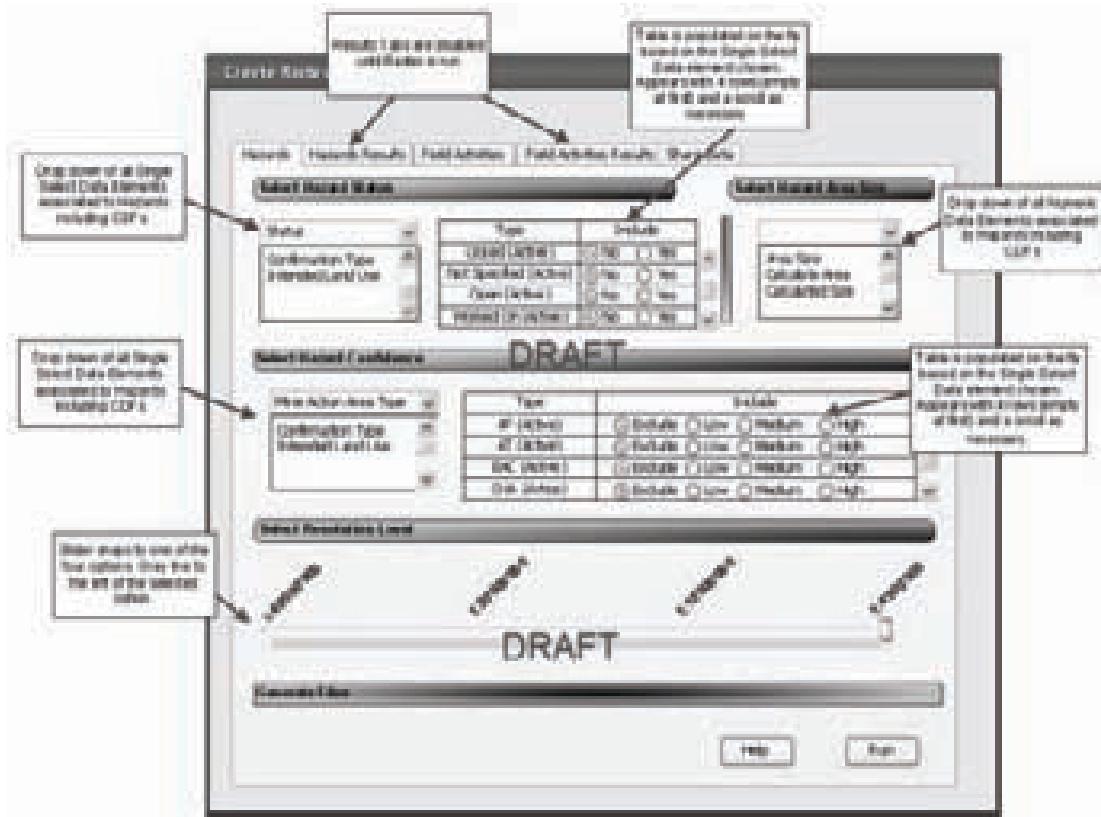


fig. 13: A draft of the options data providers will have when generating rasters (courtesy of GICHD).

The data provider will choose which fields in the database he wants to use as:

1. A status parameter and choice of which values and categories are representing active hazards.
 2. A size parameter.
 3. A confidence parameter and the level of confidence for each value / category.

After selecting these parameters the user indicates the finest resolution, which can be published on the SERWIS server.

An example: If in Somaliland the data regarding size on CHA (Confirmed Hazardous Area) is collected more precisely than information on BAC (Battle Area Clearance), a high confidence for CHA is then being selected and a lower for BAC.

The choice of these confidence factors has an effect on the raster calculation. A hazard rated with “high confidence” will be weighted five times stronger than a “low” value and three times stronger than a “medium” value. Since some countries’ categories are defined in another way than in others, the choice of the weighting is ceded to the data providers. They know best how the recording has been done and therefore they can better assess the values.

Metadata will be transmitted to the SERWIS server to ensure full transparency about values that have been used for the calculations.

4.2.2 Overlapping polygons

Working with real data, one has to deal with the conditions that influenced their recording. When a minefield is detected on-site, in order to ensure not to leave out a hazardous part, data collectors usually begin every new sector inside the previously recorded polygon and thus record a small corridor mostly twice. In certain instances this may mean that regions could appear more affected than they are since some parts are recorded twice.

This effect has been investigated: intersecting polygons, called sliver polygons¹⁶, have been separated from the original polygons and have been clipped from them. Thus, every affected part has been counted only once. The area of the sliver polygons and their centroid have been calculated, hence independent polygons respectively points have been created. The number of points increases drastically by carrying out these steps, because on every polygon, several of these sliver polygons are generated.

Hence, respecting this issue the amount of data that has to be processed increases greatly. However, although there are numerous sliver polygons, the effect of generating rasters by including them is tiny. In order not to overload the hardware capacities, it was decided to ignore these effects since their impact is literally impossible to distinguish¹⁷. Also other problems would have to be considered, such as how to deal with intersecting polygons of different kinds of hazards. This leads to uprising questions such as if an area can be a minefield, while being a cluster bomb area. Answering these questions, however, is not a part of this thesis. Since the benefit of doing much more complex calculations is very small and the maximum resolution of the output is relatively low, there is no need to waste time and resources for a minimal improvement as time and resources are limited.

4.3 Possible methods

Density estimation has been developed for univariate datasets but many methods can be applied on multivariate data as well. Silverman (1998) provides a survey of the existing methods to estimate densities: The spectrum of methods ranges from the simple histogram over nearest neighbor algorithms to very specific kinds of density estimations.

To calculate a density raster from a given set of points, various tools are provided in ESRI ArcMap, but also in other software packages, such as Intergraph Geomedia or R. Besides kernel density estimation, ArcMap for instance also provides a tool called point density that might be applied on mine action data. Using this tool, around every point a user-defined shape (circle, rectangle, ring) is defined and if these shapes intersect, their values are added. These val-

¹⁶ Sliver polygon: “A small, narrow, polygon feature that appears along the borders of polygons following the overlay of two or more geographic datasets. Sliver polygons may indicate topology problems with the source polygon features, or they may be a legitimate result of the overlay.” (wiki.gis.com n.d.)

¹⁷ See Appendix F for results.

ues then are assigned to a raster and a representation of the density distribution is created. Thus, the values will get highest where the shapes around the points intersect, because they would be added to each other at the intersection. At the point's locations itself, however, the values are not necessarily very high. This method could be applicable for example to find areas, which have a certain kind of entities nearby (e.g. a neighborhood with many restaurants within a certain distance), but for the data concerning landmine, UXO and other ERW contamination, it is not suitable.

In the case of ERW contamination the values are assumed be high, where a record is confirmed and lower, where ERW are nearby. The further away the confirmed hazard, the lower the probability of a hazard. In the z-dimension, the hazard around a confirmed ERW would have a shape of a mountain with the peak at the point's location. Hence, in the case of ERW, where not only the proximity to multiple hazards is important, but the distance to one or more hazard, an appropriate approach is required.

Possible interpolation methods could be Kriging, Spline¹⁸, trend surfaces, or local regression models. The estimation of hazards emanating by landmines, UXO, etc. between known points cannot be interpolated like for other topics. The named methods are adequate to estimate unknown values of continuous surface between known data points. Thus, the points are expected to have values on an interval or ratio scale. The case of ERW data is different. Although the data are a weighted point distribution, mainly the presence of points is important and less their values. Thus, the interest lies in the density of points and less in the unknown values between. A salient method to estimate this density is the kernel density estimation (KDE) that has been described in detail by Silverman (1998) and which fulfills the requirements of this project.

4.4 Kernel Density Estimation (KDE)

As mentioned before, it is presumed that the hazard emanating of a minefield or any other ERW is getting lower by increasing the distance to the source. This follows Tobler's first law of geography: "Everything is related to everything else, but near things are more related than distant things." (Tobler 1970, p. 236) The concept of spatial kernel density estimation (KDE), in the row with interpolation methods, is widely based on this perception.

4.4.1 How KDE works

Based on the reasoning described above, it was decided to use a kernel density algorithm to generate the maps representing ERW contamination. A good solution is to choose an algorithm that calculates the density of features in a search radius around other features (ESRI 2007).

¹⁸ "The Spline tool uses an interpolation method that estimates values using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input points." (ESRI 2007)

To get an impression of how KDE for spatial data is working, it is helpful to know how the kernel method for univariate data is designed. This method is applied to samples of independent, identically distributed observations from which the probability density function has to be estimated (Silverman 1998). The kernel estimator improves the normal histogram function that is:

$$f(x) = \frac{1}{nh}$$

where n is the number of observations and h the width of the bins. The histogram can be generalized or specified, if the bin widths are allowed to vary. This can be made with a naïve estimator¹⁹ or by implementing a kernel function.

Instead of just generating a histogram of the distributed record, an attempt is made to smooth out the histogram to achieve a more natural probability density function. Therefore, a kernel function is defined that describes the weight function between the bins in the histogram. The kernel function looks like this:

$$\int_{-\infty}^{\infty} K(x) dx = 1$$

The kernel function is usually a symmetric density function, for instance the normal density. The kernel estimator is defined by

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

where h is the kernel size (also called bandwidth, window width or smoothing parameter). I will use the term ‘kernel size’ from now on. While at the histogram boxes are being put on each other, to show higher densities, the kernel estimator sums up “bumps” that are placed on the observations. The functions vary depending on the kernel size, which can be seen in the following figures (the observations are always the same, only the kernel size changes).

¹⁹ By choosing a naïve estimator, the values between the bins are estimated to make the histogram more precise. Every point is the centre of a sampling interval, which allows freeing the histogram from a particular choice of the bin positions. The data are smoothed strongly and a more realistic pattern of the distribution is produced. (Silverman 1998)

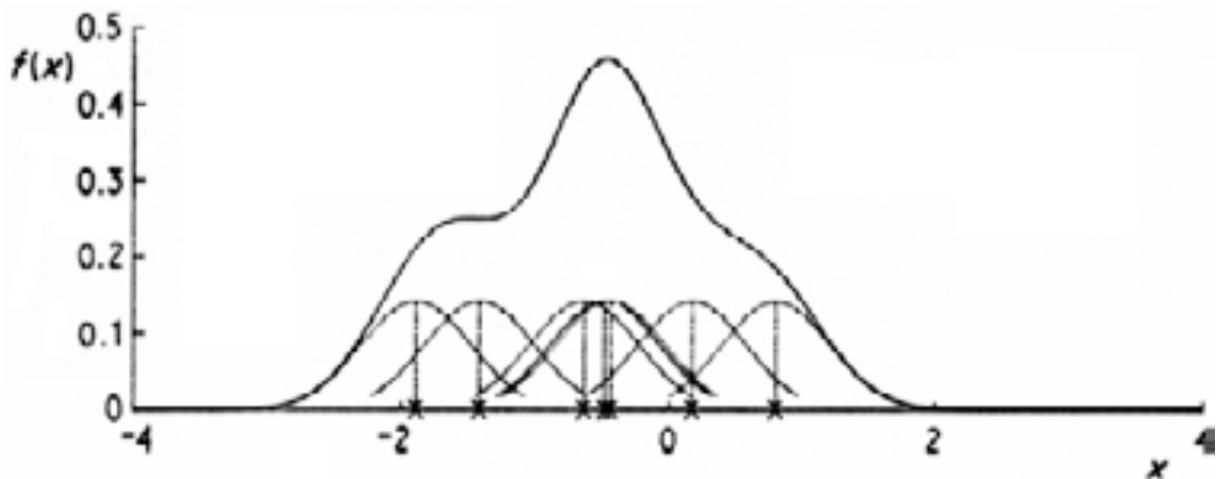


fig. 14: Kernel estimate showing individual kernels. Kernel size 0.4 (Silverman 1998, p.13)

In fig. 14 the results of a medium sized kernel size can be seen. The “bumps” sum up to the overall estimate, which has its maximum where the concentration of records is highest. In fig. 15 the kernel size is smaller, hence the differences are higher as well as the maximum values. The estimated curve is rougher than the one before, the autocorrelation is smaller.

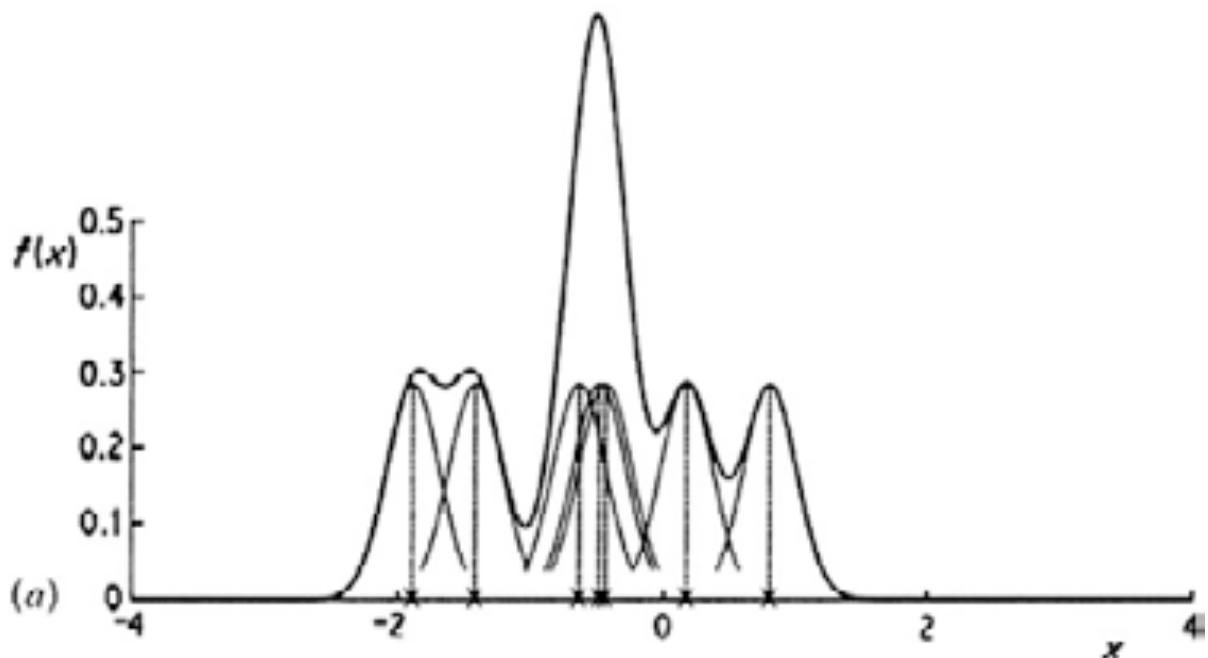


fig. 15: Kernel size: 0.2 (Silverman 1998, p.14)

In fig. 16 the kernel size is three times larger than in fig. 15, hence the estimated function is getting much smoother, namely it has become only one curve that resembles strongly a normal distribution.

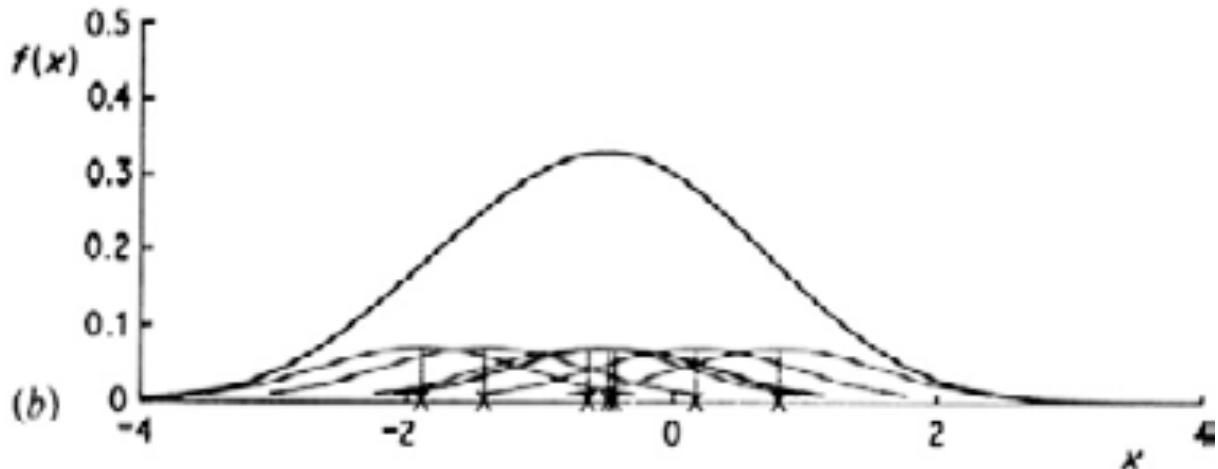


fig. 16: Kernel size: 0.6 (Silverman 1998, p.14)

Hence, the choice of the kernel size affects the results crucially. Therefore, it has to be chosen with care. The method of the kernel estimator is one of the most commonly used methods for density estimations and has been studied most (Silverman 1998). Furthermore, the smoothing property of KDE can also be utilized to make information fuzzier, so that the conditions in terms of obfuscation are easier to fulfill.

4.4.2 The kernel method for multivariate data

In the case of this thesis the kernel density estimation has been used for real data that are distributed in two dimensional space. Thus, multivariate or bivariate density estimation had to be applied. Bivariate kernel density estimation is a feature of Spatial Analyst in ArcToolbox. It works similarly to the described method above, but is applied on bivariate – i.e. spatial data.

Silverman (1998, p. 88) describes how this method can be applied on multivariate data. It can be used for d dimensions. In the case of two-dimensional spatial data d is 2. Here, “bumps” are created that are centered at the observations and if they intersect they are added in analogy to the method for univariate data. Usually the kernel function is a radially symmetric, unimodal probability density function, for example the standard multivariate normal density function:

$$K(x) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2} x^T x\right)$$

Another kernel function is the multivariate Epanechnikov kernel that is used in the method provided by ArcGIS:

$$K_e(x) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2)(1-x^T x)^{-1} & \text{if } x^T x < 1 \\ 0 & \text{otherwise} \end{cases}$$

(Silverman 1998, p.76)

c_d is the volume of the unit d -dimensional sphere, in c_2 it is π . The curve of this kernel can be seen in fig. 17. The advantages of the Epanechnikov kernel are higher differentiability properties for kernels and their resulting estimates as well as shorter calculation times. That is why this shape has been implemented in ArcGIS. For other common shapes, see Appendix E.



fig. 17: The Epanechnikov kernel (Silverman 1998, p. 42)

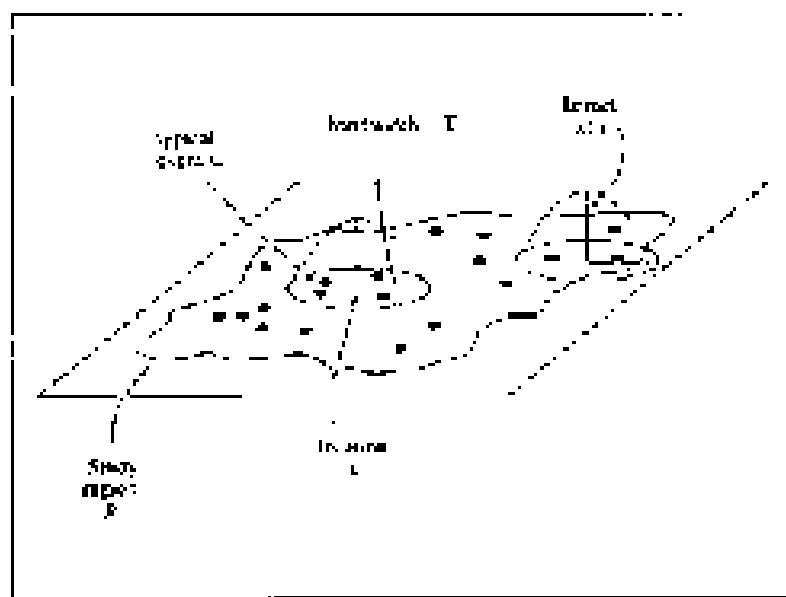


fig. 18. Kernel density estimation of a point pattern (Gatrell A.C. et al. 1996, p.260)

In fig. 18 is illustrated how a KDE of a point pattern is assembled schematically and which values have an influence on the output. Applied on ERW contamination estimation the study area R is an affected countries and the locations s are recorded ERW hazards. Additionally they have been weighted, which influences the kernel $k()$ for each record. The choice of the kernel size τ is explained in section 4.4.4.

4.4.3 Data processing

For the implementation of the method, ESRI ArcToolbox with Spatial Analyst offers a large variety of utilities adaptable on spatial data. Hence, the kernel density tool has been applied on the available datasets.

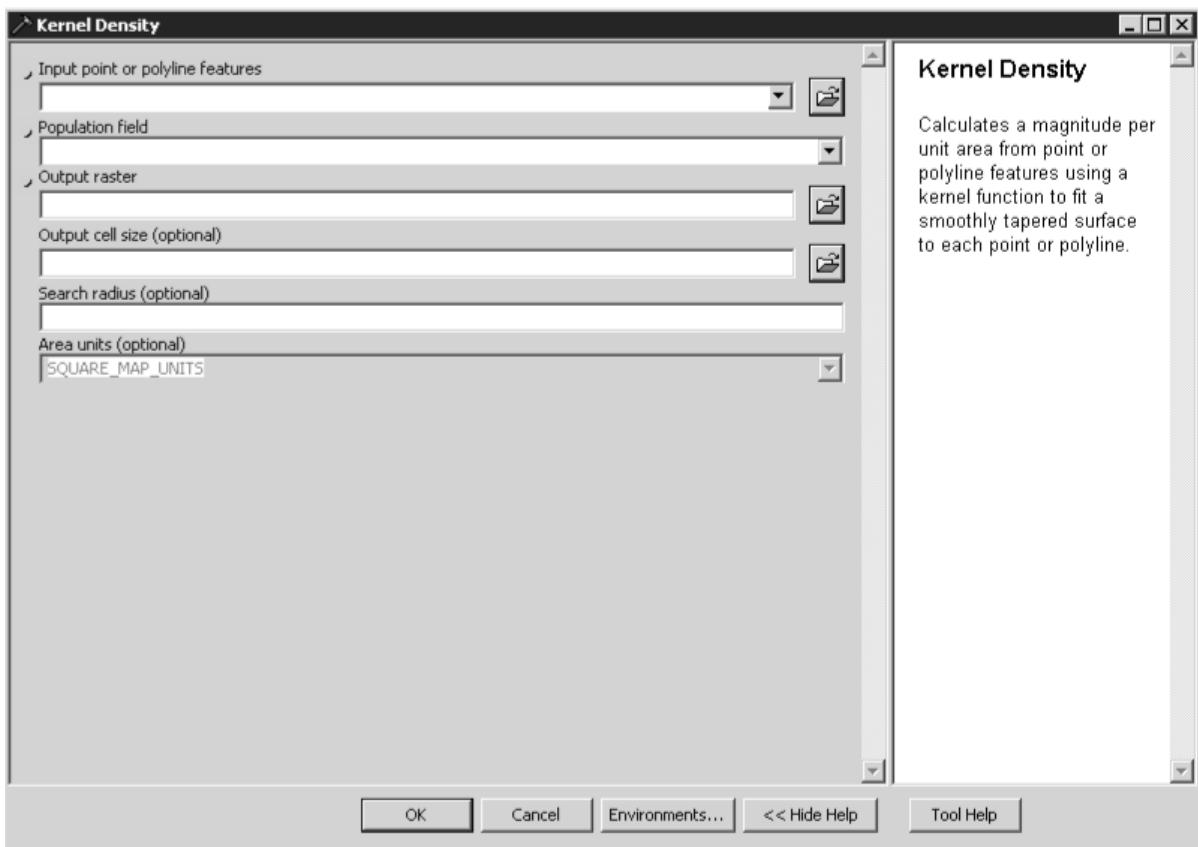


fig. 19: The Kernel Density dialog in ArcMap

fig. 19 shows how the KDE dialog in ArcMap looks like and which values can be set. If the user chooses a valid input (a point or a polyline feature class), besides the “population field” (this is the weighting value), optionally he can assign the “cell size” and the “search radius” (kernel size or kernel size). Of course in KDE there would be some other parameters that could be adjusted; not only the kernel size, but also the shape of the kernel can have an influence. Unfortunately, in ArcGIS this cannot be changed.

4.4.4 Kernel Size

Choosing the kernel size is a very important and difficult part of the process. There are methods to find a good kernel size, depending on the point pattern. But there is no right or wrong choice, as long as the requirements of the target application are met. According to De Smith et al. (2007, p.178):

“Bandwidth selection is often more of an art than a science, but it may be subject to formal analysis and estimation, for example by applying kernel density estimation (KDE) procedures to sets of data where actual densities are known.”

There are methods to estimate the kernel size value mathematically (Silverman 1998) and it was initially tried to find the best solution by this way. However, since the point patterns are unknown and vary strongly a way had to be found to have a kernel that works with every distribution for every country. An adaptive kernel is not a sensible choice, since it must be possible to compare the results of different countries. Moreover, changing the kernel sizes produces different estimated values, thus if the kernel was adapted in each country they would not be comparable anymore.

It was decided to have the same kernel sizes for all datasets, in order to keep the countries comparable, which is one of the main goals of the SERWIS project. Having countries as little as Cyprus ($9'248 \text{ km}^2$) and as large as Sudan ($2'505'813 \text{ km}^2$), the use of the same kernel sizes is a challenge that must not be underestimated. Comparing countries worldwide is a challenge for any kind of data, since the values are often very heterogeneous and country sizes differ heavily.

The purpose of the generated maps within this thesis will be to show the threat of ERW on different scales. Interested persons get the possibility to pan and zoom to see the distribution of hazards worldwide. Heavily affected areas should become visible on the global scale; moreover it has to be made possible to see the hazards on more detailed scales, i.e. to see regional differences. Therefore, different maps with different properties have been generated. On a global scale, the human eye cannot perceive objects of a few kilometers neither can they be shown. Therefore, for different scales different kernel sizes have been chosen. On the global scale the largest kernels have been selected to draw a very generalized picture of the ERW threats in the world. Here, the radius of the kernel is 200 km. On the other hand on the more detailed scales, the picture gets much more circumstantial, which allows the viewer to recognize patterns inside the countries. This does not go as far as choosing a different kernel size for any possible zoom level; there are only four levels of detail. Having these levels, allows to give a diversified representation of the phenomenon without overloading the provider's computation capacities.

The following kernel sizes have been selected in order to meet the projects requirements:

- Global scale (~ 1 : 40'000'000): 2'000 km kernel radius
- Continental scale (~ 1 : 20'000'000): 100 km kernel radius
- National scale (~ 1 : 10'000'000): 50 km kernel radius

- Regional scale ($\sim 1 : 4'000'000$): 20 km kernel radius

In this case generalization is one of the central concepts, like in any map design. Generalization means to remove the ‘specific’, ‘random’ and ‘unimportant’, while emphasizing the ‘important’ (Brassel & Weibel 1988). In this case, on a global level, the focus is on presenting the centers of hazard concentration world wide while on the regional level local differences can be made visible.

Furthermore, at coarse scales the results can be very doubtful. Performing a KDE in a small country like Cyprus (which means less than 100 km across) with a kernel that is several times larger than the country itself, estimated density values do not make a lot of sense. Nevertheless, on a global scale the interest lies in showing where the affected countries are, but not their detailed situation. With a good classification and color ramp this problem can be handled as well. (See section 4.5)

Eventually, in order to not present very unrealistic data, it was decided to place markers on affected countries that are too small to be visible on the map at global scale, and color them according to their severity of ERW hazard threat²⁰. This makes the maps more readable and by zooming in, the user can still see the situation in the regional level of detail.

Nevertheless the smallest kernel size will still be 20 km, in order that single records cannot be identified. The risk of presenting confidential information is therefore minimized. (See also section 4.8)

4.4.5 Cell Size

The cell size does not affect the result strongly. By choosing a higher resolution, the density surface becomes smoother, but also the calculation time increases. Thus, a value has been chosen, which does not overcharge the provider’s computers, while still generating high-resolution rasters. Of course the maps for different scales have a different resolution. It was decided to connect the cell size with the kernel size, in order to have a visually similar impact on the results. Thus it was decided to select a 200 times smaller value than the one of the kernel diameter.

Eventually, four levels of detail have been chosen – from a global to a regional scale. Their resolutions have been chosen as the following:

- Global scale: 2'000 m cell size
- Continental scale: 1'000 m cell size
- National scale: 500 m cell size
- Regional scale: 200 m cell size

²⁰ To get the value of this color, the average threat level of the country overall is estimated, and the marker’s color is chosen according to the corresponding class in the color ramp.

By choosing these values it becomes possible to display maps for different extents, with reasonable dimensions. In no level of detail cells will be visible at display scale and the computation times for the highest resolution remain in the limits of the hardware capacities. If a country did not submit at the finest resolution, only the coarser levels of detail will be presented and it is indicated that no data are available for that specific level.

4.4.6 Values

The estimated values obviously depend strongly on the input values. Since there is no value that measures the contamination of landmine, UXO, or other kinds of ERW hazards, a virtual value has been calculated. The area of the polygon and a confidence factor determine this value. The choice of this factor is explained in section 4.2.1.

$$\text{hazard value} = \frac{\text{hazard area}}{\text{confidence factor}}$$

This hazard value is assigned to any hazard point and determines its weight for the KDE that is performed subsequently. The integral of the kernel function remains constant, independent of the kernel size. Thus, with a small kernel size the values get much higher than for the large ones. Therefore, the resulting values can be as high as hundreds of thousands (since the area is calculated in square meters), suggesting that around these high peaks the contamination is very high. Eventually, the calculated values represent a virtual hazard value that will not be displayed numerically but in the shade of the color. (See also chapter 4.5)

4.5 Classification

Having calculated the hazard values with KDE while choosing the appropriate settings, the rasters need to be represented adequately. A big challenge of this project was to find the right classification for the final representation. As a matter of fact the amount of hazards can vary a lot. Thus, to find a classification method that respects the asymmetric distribution is very important. A method had to be found, which does not let low-affected countries disappear while not having countries that seem over-affected or that seem as if they are completely contaminated with ERW. There are several methods that calculate suitable classifications; the hard part was to find the appropriate one.

4.5.1 Classification methods²¹

ArcGIS supports the following classification methods: equal intervals, defined intervals, natural breaks, quantiles, geometric intervals and the standard deviation method. If desired, the user can

²¹ The descriptions base mainly on the ArcGIS Desktop Help 9.2 by (ESRI 2007) if not indicated otherwise.

define class intervals manually. This allows developing own classification, which can represent the data more adequately.

Equal interval

This classification scheme divides the range of values into subranges of same size. The user can specify the number of intervals, while the program determines where the breaks should be.

Defined Interval

Different to the equal intervals scheme, here the user chooses an interval, while the program finds the corresponding number of intervals.

Quantile

This method places the same number of observations in every interval. A quantile classification is recommended for linearly distributed data. The result of this map can be misleading, because values in the same class can be very different, while in some classes they span only a very small range. With a large number of classes, this distortion can be minimized.

This method seems to be quite reasonable, but caution is advised here. To calculate the quantiles of the values of one method is very easy and the result is visually rather appealing. It is also possible to combine the calculated maps of some countries and to calculate the new breaks. In practice, this method cannot be recommended unconditionally. It is expected and hoped that in the course of time more data will be available and more and more countries are going to participate in this project. Thus, the classification has to be recalculated every time a new map is added, which is not very convenient.

Natural breaks

Here, the classes are based on natural groupings inherent in the data. The break points are picked to best group similar values and maximize the difference between the classes (Jenks & Caspall 1971).

In the data related to ERW, natural breaks occur very rarely or never. Therefore, this method is not suitable to classify the data.

Geometric Intervals

In this scheme the class breaks are based on class intervals that follow a geometrical series. In order to optimize the class ranges, the geometric coefficient changes after the first intervals to its inverse, which allows a better differentiation for low values. In ArcGIS intervals are chosen in order to minimize the sum of elements per class, while the user chooses the number of classes. The result is visually appealing and cartographically comprehensive.

Standard Deviation

Here the classification scheme is based on the values' variation from the mean. The class breaks are created using mean values and the standard deviation. This classification method is not recommended for skewed distributions like the ERW data.

In fig. 20, the mentioned classification methods are applied to the same data. The underlying dataset depicts the results of a KDE for a part of Afghanistan at a relatively high level of detail. Only the class breaks have been changed, the parameters for the KDE constant (20 km kernel size, 200 m cell size). The color range goes from yellow to red, while zero values have no color.

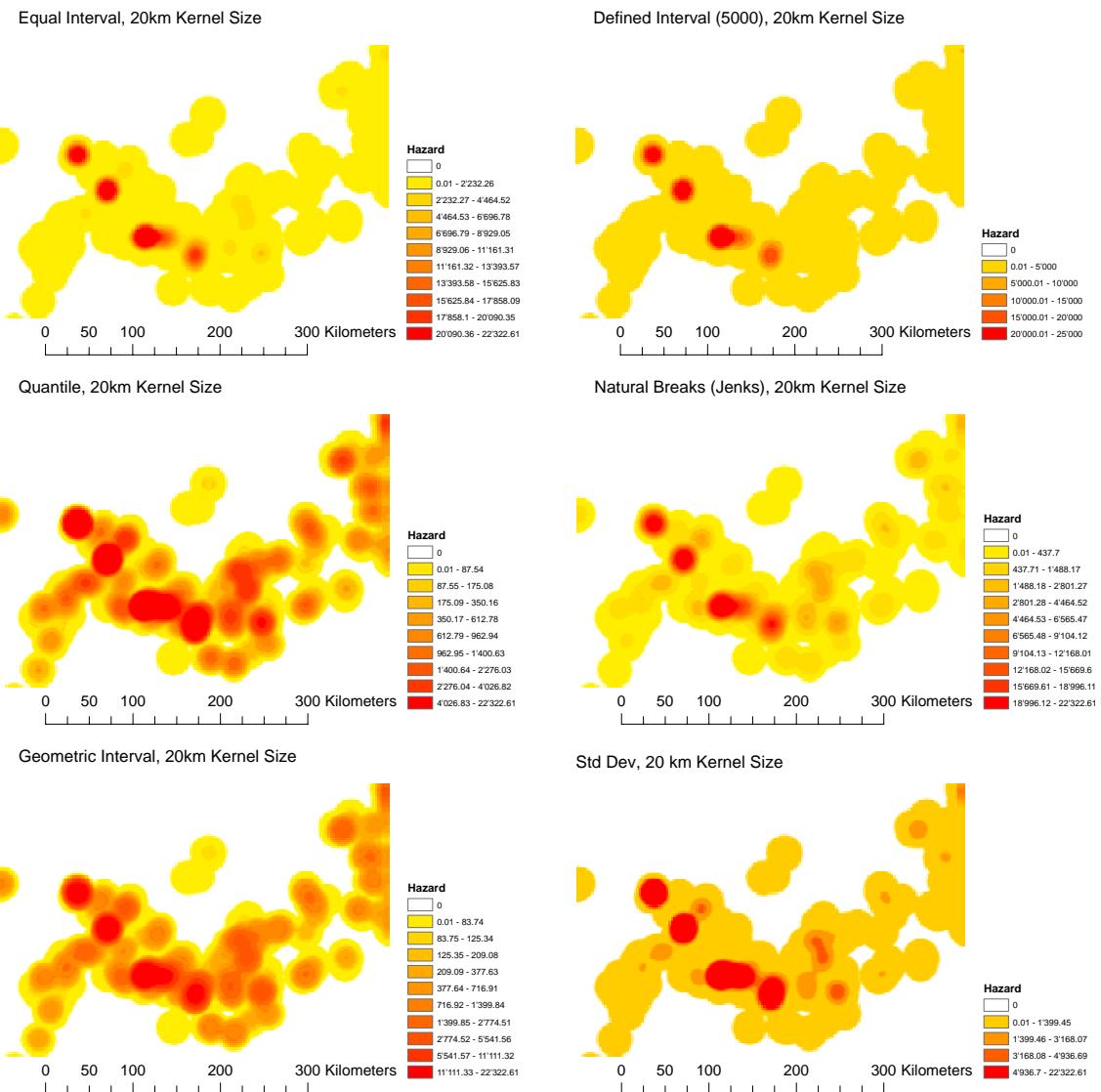


fig. 20: The six classification methods provided by ArcGIS

This example illustrates that by changing only the classification method the results can look very different. Looking at these six options one can see that it is difficult to identify one solution as representing the data better than any other. Further, the color ramp has a strong influence on the results. Adding more shades of colors or to start with a white color at low levels can generate a

very plausible solution, but it also the message can change. Finally, after discussing with experts of GICHD it was decided to choose the shown color ramp (yellow to red).

4.5.2 Number of classes

Another question that arises concerns the number of classes that should be used to represent the data. Even a continuous progression – without any class breaks, is conceivable. Therefore, either a number of class limits is chosen, so that the intervals are defined; or there are no classes, i.e. the values have no steps, the color changes with the progression of the hazard values. Both methods have their advantages and disadvantages.

Since the hazard rate does not have any natural steps and is a continuous phenomenon, few classes would not make a lot of sense. Working with few classes also neglects the very details of the distribution, and so it seems to make more sense to choose a continuous classification or many classes (30 or more). But displaying many classes can make the map more complicated and more difficult to read. If the user only has to distinguish between few shades of colors, he or she would not get overstrained. Nevertheless, a continuous classification comes much closer to reality and the human eye is used to fuzzy transitions of colors. That is why this solution has been selected.

4.6 Projection

Working with global data, the spherical shape of our planet has to be considered. The data derived from IMSMA are stored in WGS84²² format. This implies that x- and y-coordinates are stored as decimal degrees and by calculating distances and areas the latitudes have to be taken into account. To perform calculations that are depending on distances (like KDE), it is necessary to project the data into another coordinate system. Having them converted into the UTM Grid System²³ Euclidian (rather than spherical) distances can be calculated. Thus, for any country the UTM zone has to be identified and finally when putting the pieces together in a global map, they have to be turned back into a global format. This can be done with simple ArcGIS functions, but ignoring this issue would generate wrong distances.

²² The World Geodetic System 1984 (WGS84) is a complete standard for georeferencing. It was established by the US Department of Defense. It compromises a standard coordinate frame for the Earth (Boucher & Altamimi 2001).

²³ The Universal Transverse Mercator (UTM) system is one of the most widely used map projections. It is divided into 60 zones each 6 degrees wide in longitude. The zones are narrow enough to keep distortion and scale variation within the zone to an acceptable level (Langley 1998).

4.7 Neighboring countries

A very delicate aspect that affects the results of the process is the problem of countries that share a border. Many mines have been laid in conflicts concerning sovereignty over border regions and many borders of mine-affected countries are still disputed. Therefore, a way had to be found how to present the situation in these areas. Since the kernel size can be as large as 200 kilometers, estimated values can be as far away as the kernels size from the actual borders. Hence, the conflation of the hazard rasters of bordering countries that overlay each other has to be handled with care.

There are approaches of how to deal with this problem. The easiest way would be to clip the calculated raster with the borderline of the country. But this apparently simple solution causes some other crucial problems: In cases where we have bordering countries and therefore overlapping rasters, often there is no border that is accepted by both countries. Thus, to guarantee neutrality, it is not possible to use any available border to clip with, because to choose a border would mean to choose a party. The policy of the GICHD is to prevent showing any border as long as possible, thus an alternative solution had to be found.

A reasonable solution is to calculate average values of the overlaying cells and displaying them. All values are estimated and therefore they will not be the same and since the rasters of each country are calculated locally and not on a central server, it has to be dealt with this problem with already processed data. Furthermore, this can become a very complicate issue: new data are coming piece by piece and therefore every time an additional raster is added the border area has to be recalculated. This also means that values can change only because data are added.

A very straightforward solution that was chosen for the first version will be to display the map that is closest to the center point the viewer is looking at. Bordering countries remain in the back until they come to the center of attention, when they are moved to the top level.

4.8 Obfuscation

As described by Duckham & Kulik (2005) obfuscation concerns the practice of degrading information somehow to protect the privacy of the individual to whom the information refers. In the case of this project this concerns the ERW hazard data that must not be revealed to the public. Many data providers have a legitimate interest in hiding information, as described in section 1.3.1.

Although the calculated rasters are going to be presented at a relatively low resolution and only the centroids of the minefield polygons will be taken into account, it has to be assured that the privacy agreement is not violated. KDE already contributes a lot to an obfuscated representation, nevertheless, looking at the results one could argue that a hazard can be expected in the center of the resulting circular patterns, which applies mainly to isolated observations in remote areas. Thus, with the smallest kernel sizes, potentially inferences can be drawn to locations of ERW.

In research, obfuscation techniques have been used mainly to ensure an individual's privacy for location-based services, e.g. by spatial and temporal cloaking (Gruteser & Grunwald 2003) or by elaborating sophisticated privacy concepts (Snekkenes 2001). In this project, the conditions are less complicated. Unlike in this domain of location-based services, there is no temporal dimension involved in ERW data that has an influence on confidentiality. Thus, besides the KDE, a solution is formulated that only deals with the records' locations.

The maps are not designated to help people navigate, hence in some cases the exact locations of hazards are less important than their general distribution over space. Thus, their location can be changed slightly, which does not alter the pattern too much. This can be achieved by performing a perturbation of the point pattern. In other words, the points' locations had to be changed by adding a random distance ϵ . It could be possible to allow data providers to select the size of ϵ . With small values of a few hundred meters, the original location will be shifted, while the characteristics of the point pattern remain intact. If the parameter ϵ is too large, the differences can be too severe, thus I recommend not choosing a value that is too high.

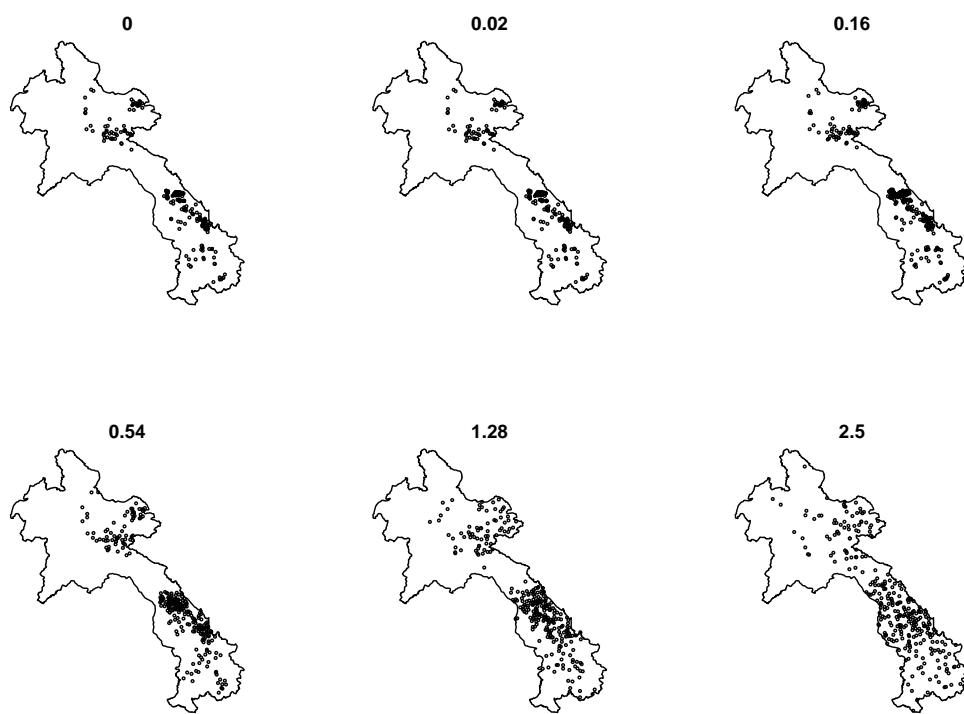


fig. 21: Perturbation of a point pattern with a growing ϵ . The number indicates the size of ϵ in decimal degrees.

The effects of a growing ϵ on the point pattern of records in Laos are illustrated in fig. 21. If the patterns were changed strongly like in the lower plots, computing a KDE with a small kernel size would not make sense anymore. The pattern has nothing to do with the original and the results would be misleading. Nevertheless performing a KDE with a large kernel would be considerably

less impaired. The perturbation of the points does not have a large influence if the kernel size is considerably larger than ε .

Hence, it is not recommended to perform a high-resolution KDE with data that have been randomized strongly, but what about a smaller perturbation? If we now calculate the KDE for the “jittered” new point pattern the result seem to be quite similar, although the data have been manipulated decisively (see fig. 22).

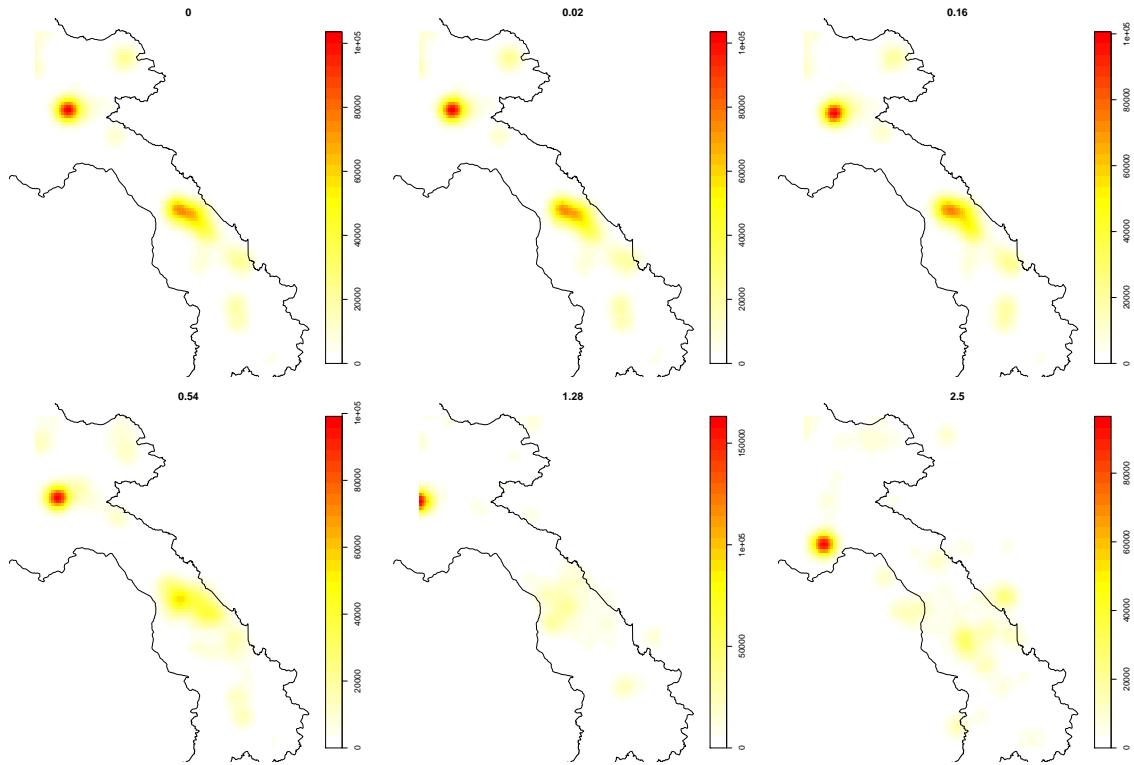


fig. 22: Kernel density estimations with an exponentially growing amount of perturbation.
The number indicates the size of ε in decimal degrees and the kernel size is 0.05 decimal degrees.

In fig. 22 the results of the KDE with the perturbation of its points are presented. Here ε grows exponentially while all other parameters remain constant. While the first results do not seem to differ a lot from the original data (upper left) the latter do not have a lot in common with the actual situation anymore.

It can also be seen that one point in the pattern has one very high value (it generates the red point in the upper left). Compared to the others its weight is considerably higher due to the large area of the hazard. The hazardous area in the center, on the other hand, blurs out almost completely with a high ε .

This solution seems to work very well, although some additional issues have to be regarded as well. For example, after perturbation points may end up in unlikely positions, such as lakes, the ocean or across national borders. To prevent this, points moved across borders or to impossible

locations should either be ignored, i.e. not taken into account when calculating the KDE, or these points could be perturbated again until they end up in at a valid location.

Thus, for the data providers an option could be added that allows obfuscating the point pattern by this method. The user could be prompted to choose an error factor from zero to a defined value. In order not to encourage users to deliver data that is heavily modified from its original status and in order to keep the interface simple, it was decided not to implement this function. Nevertheless, it would be possible to add this obfuscation method, if data providers hesitate to furnish information, or if they require more obfuscation features.

4.9 Workflow

The steps mentioned in the sections above are carried out either locally at affected countries or at the GICHD. All users of IMSMA (mine action authorities or organizations filling this role for a specific regions, or other organizations in support of operations in ERW-affected countries) can submit their information. To get an idea on how the different actions are distributed among them, fig. 23 illustrates a simplified workflow of the SERWIS project.

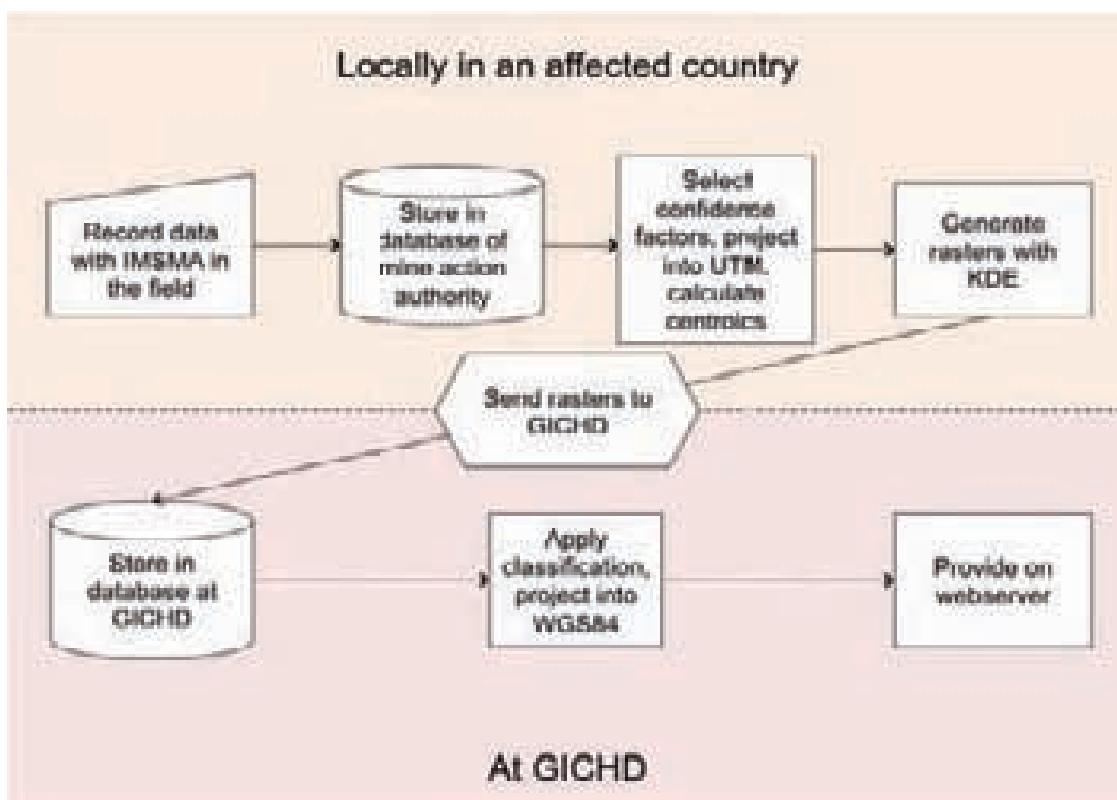


fig. 23: Workflow of the SERWIS project

5 Results

In this chapter I would like to present some possible solutions that could be used to visualize the generated data for the public. Furthermore, an overview is given on how results vary by changing the various parameters.

For statistical analyses and generating most of the figures, the statistical environment and programming language R²⁴ (www.r-project.org) has been used, explained extensively by Ripley (2001). Bivand et al. (2008) offer good explanations of R's possibilities for spatial statistics. In this thesis, it was also applied to perform KDE that use a slightly different algorithm than ArcMap does. This and other algorithms have been described by Baddeley & Turner (2005). Other figures have been generated using ESRI ArcMap and edited with Adobe Illustrator.

5.1 Presentation of the generated maps

The generated maps will be made available on the Internet, viewable for donors, decision makers, and any kind of interested people. There will not be any restrictions, i.e. any interested person with Internet access will be able to see the maps. To present the maps in a web application is the best possibility to reach a large public, since online maps have found their way into many people's daily lives. GIS knowledge cannot be expected from the target users, thus the application has to be developed keeping that in mind. Hence, only basic functions will be available, such as zooming and panning. When the SERWIS project goes on-line in spring 2011, around 10 countries are expected to have provided their hazard rasters. The development of the web interface is still in progress (status in October 2010).

An example of how the maps could be presented is shown in fig. 24. The five zoomed-in areas show how the representations look at a higher level of detail. For all representations the same classification and color ramp has been applied.

²⁴ Code fragments used to generate the figures and calculating statistical values are presented in Appendix G.

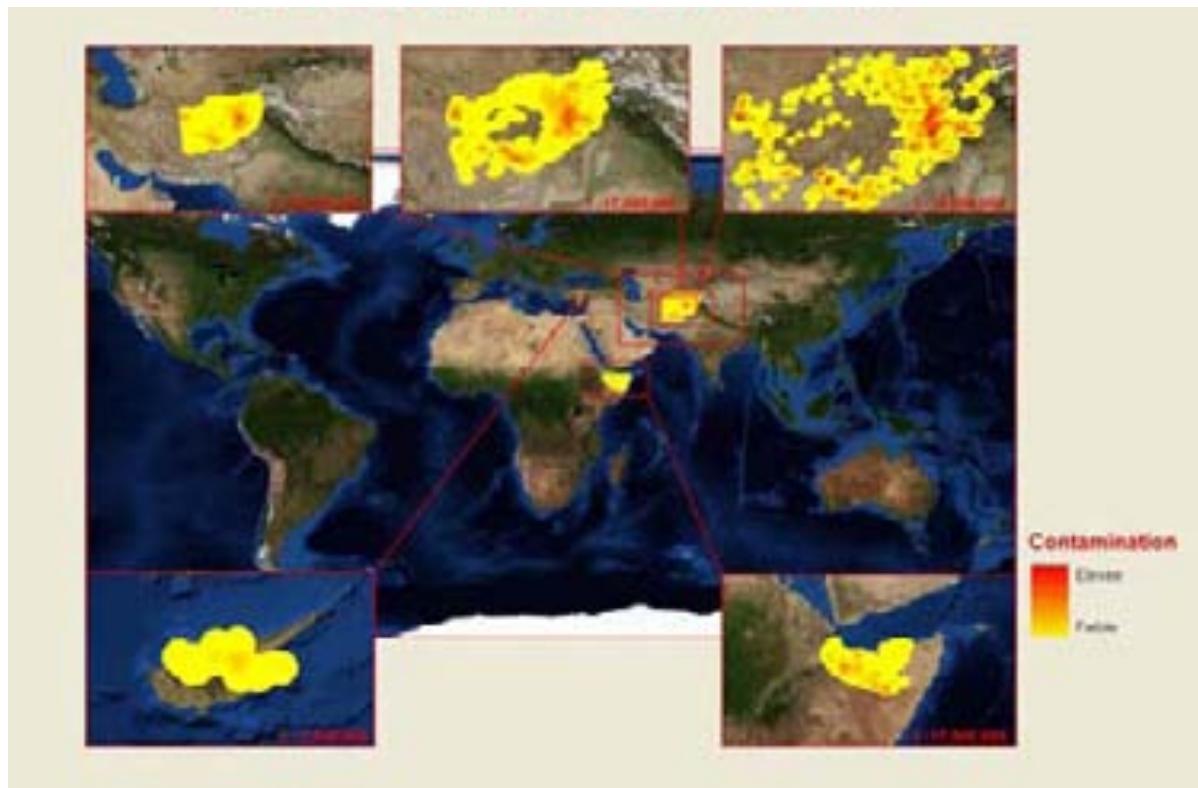


fig. 24: A global overview of how the created layers could be presented (P. Lacroix & H. Santiago 2010).

Another example on the presentation can be seen in fig. 25, where a draft layout of the web application's interface is shown. In the frame on the left different layers will be available, which can be selected by the user.

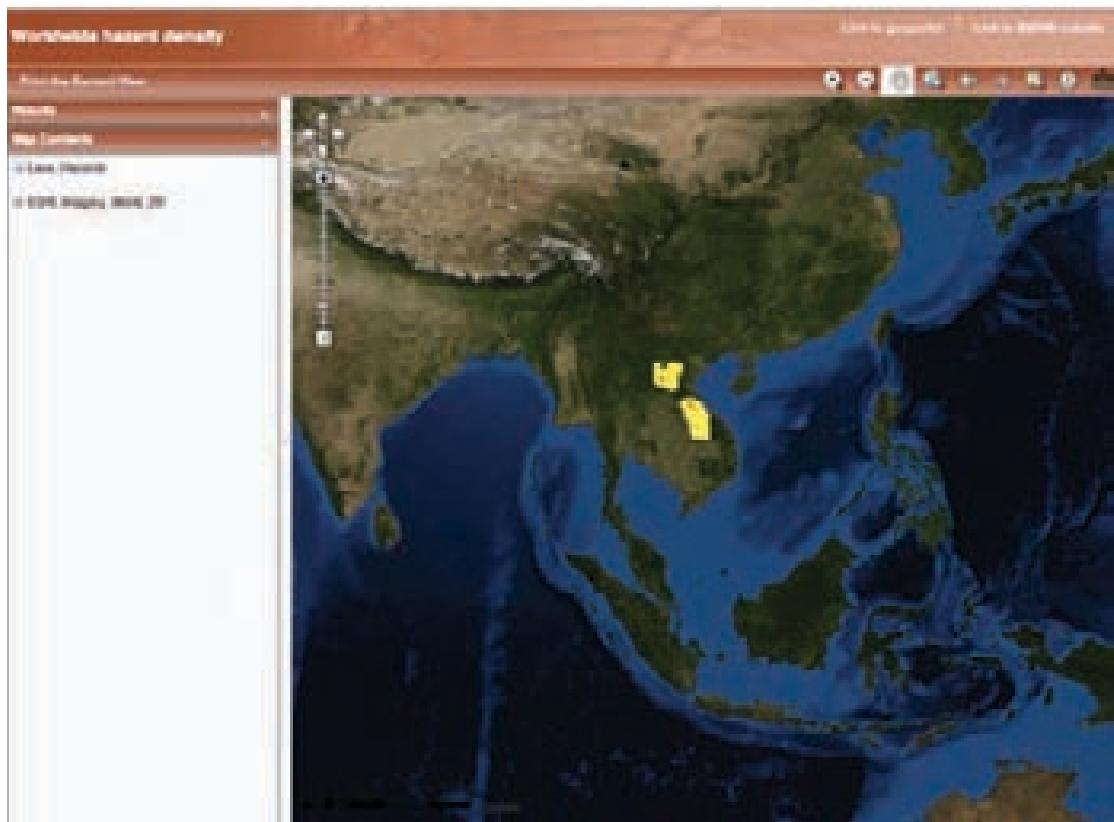


fig. 25: The interface of the web application, with example data.

Besides the information on global ERW contamination, the GICHD has plans to develop more layers. For the moment, three other layers²⁵ are being developed as a first phase of the SERWIS project. They have not been developed in line with this thesis, but will strongly rely on the hazard rasters generated in this thesis.

5.2 Examples

Working with a well-prepared point pattern, the right choice of weights and reasonable parameters for the density estimation, it is possible to create convincing maps that show the ERW hazard distribution. The following examples have been created with ArcMap and represent examples of how the final product could be presented. To make them more readable, settlements, main roads²⁶, country borders and a digital elevation model have been added, only for the purposes of this thesis. For every country, four maps that visualize hazard have been generated, thus sixteen

²⁵ A description of these layers can be found in Appendix B.

²⁶ The road data are taken from OpenStreetMap (2010) and their quality varies heavily between countries and only main roads have been selected. That is why some gaps and inconsistencies occur in some of the maps.

final maps have been created. There are some examples printed in this thesis, of the four countries and with the selected kernel sizes²⁷:

- Each country on the regional scale, i.e. with a kernel size of 20 km (fig. 26 - fig. 29).
- A map for Afghanistan with a large kernel size of 200 km that corresponds to the global scale (fig. 30).
- One for Somalia with a 100 km kernel size that will be presented as a continental scale. (fig. 31).
- For Laos a representation with 50 km kernel size, which would be at a rather detailed level (national scale) see fig. 32.

For all representations, the same classification method has been applied and the kernel and cell sizes are chosen as described in sections 4.4.4 and 4.4.5. The representations do not show the current situation, since the underlying data is not up-to-date.

²⁷ All of the following generated maps have been generated using ArcMap. Moreover, following base data has been used: Digital elevation model by GAEZ (2008), road data by OSM (2010), settlements by Geofabrik GmbH (2010).

ERW Hazard in Afghanistan: 20 km Kernel Size

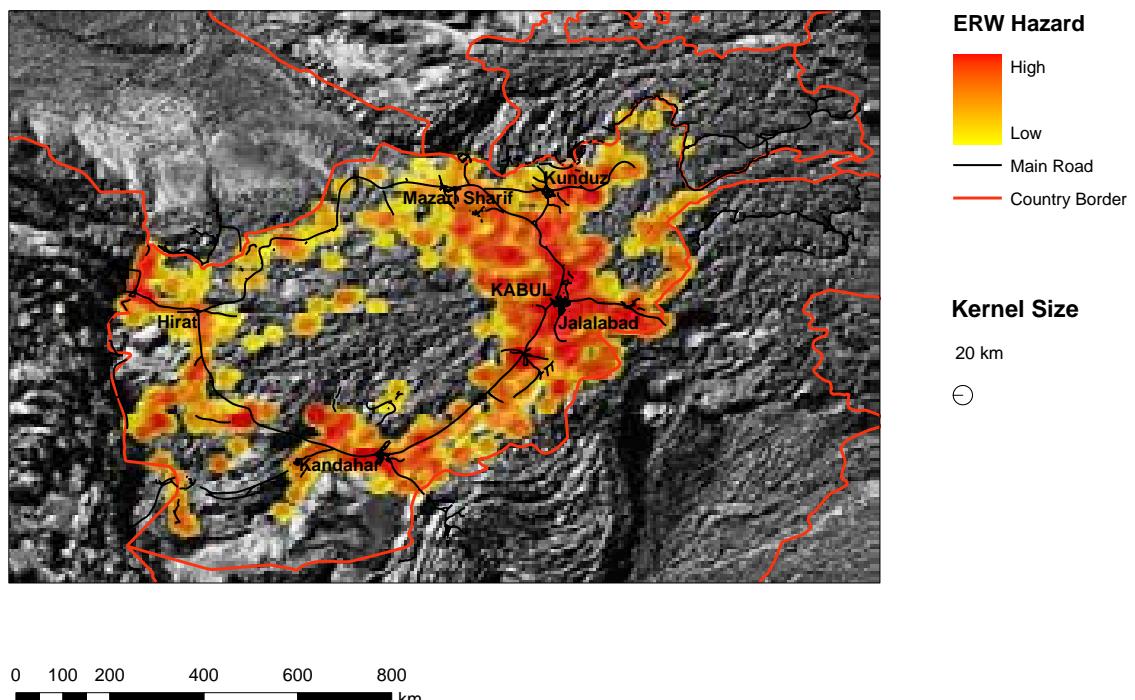


fig. 26: Afghanistan, regional level (not showing the current situation)

ERW Hazard in Somalia: 20 km Kernel Size

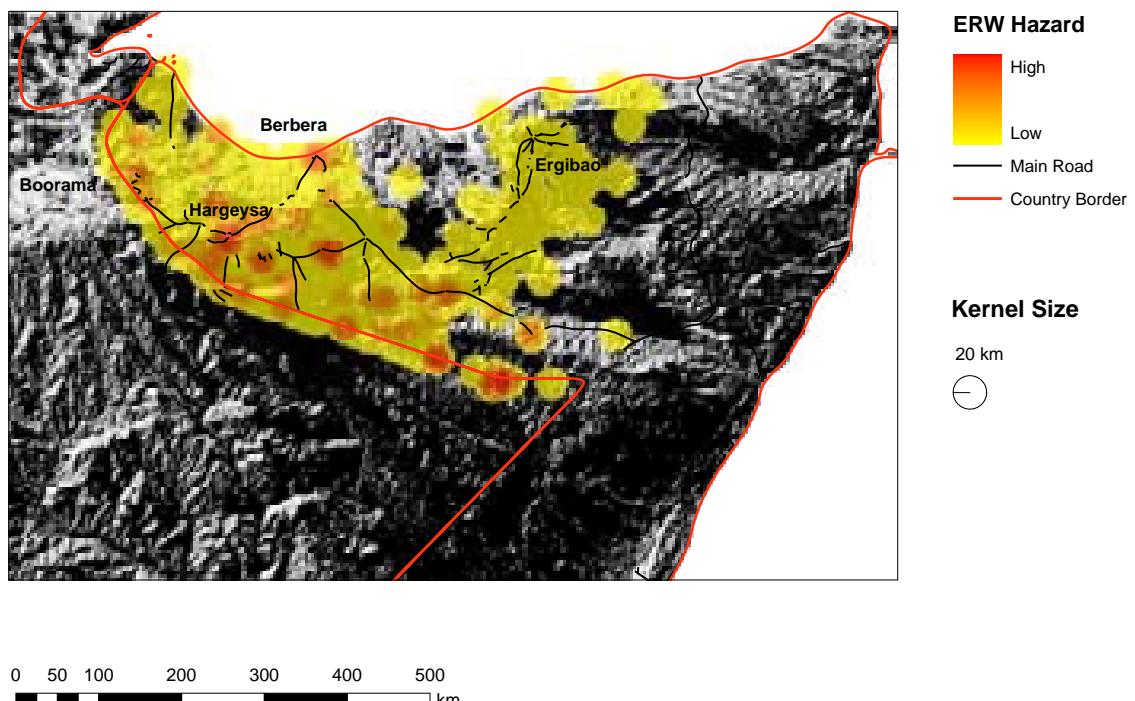


fig. 27: Somalia, regional level (not showing the current situation)

ERW Hazard in Laos: 20 km Kernel Size

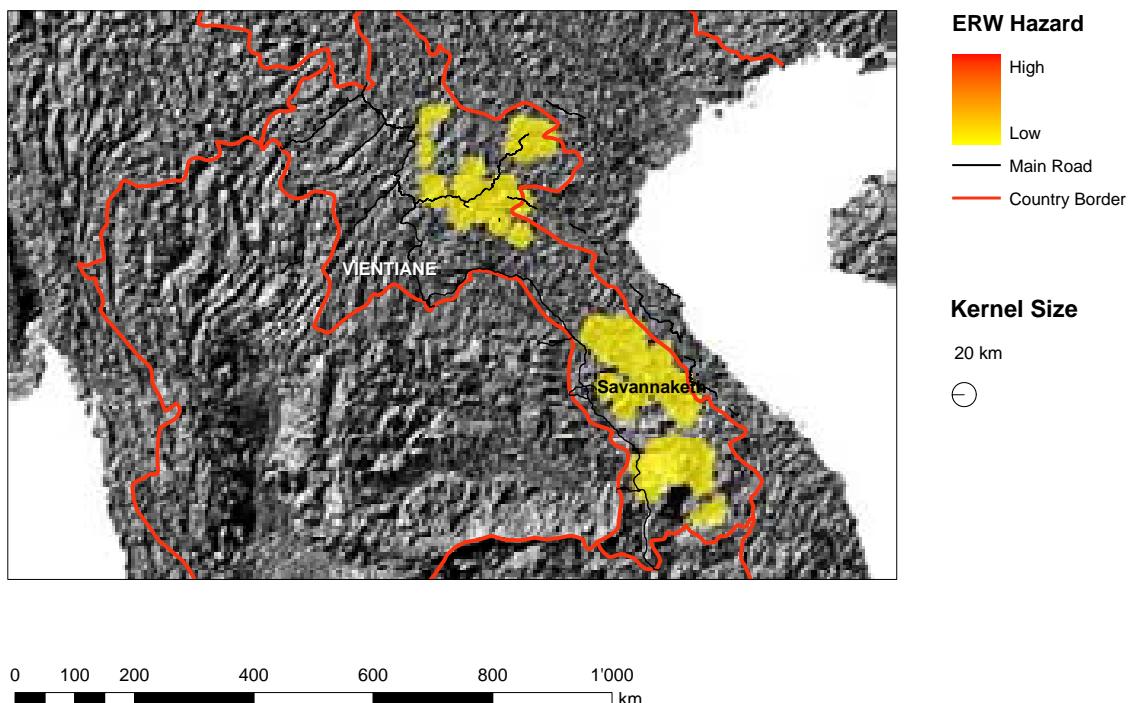


fig. 28: Laos, regional level (not showing the current situation)

ERW Hazard in Cyprus: 20 km Kernel Size

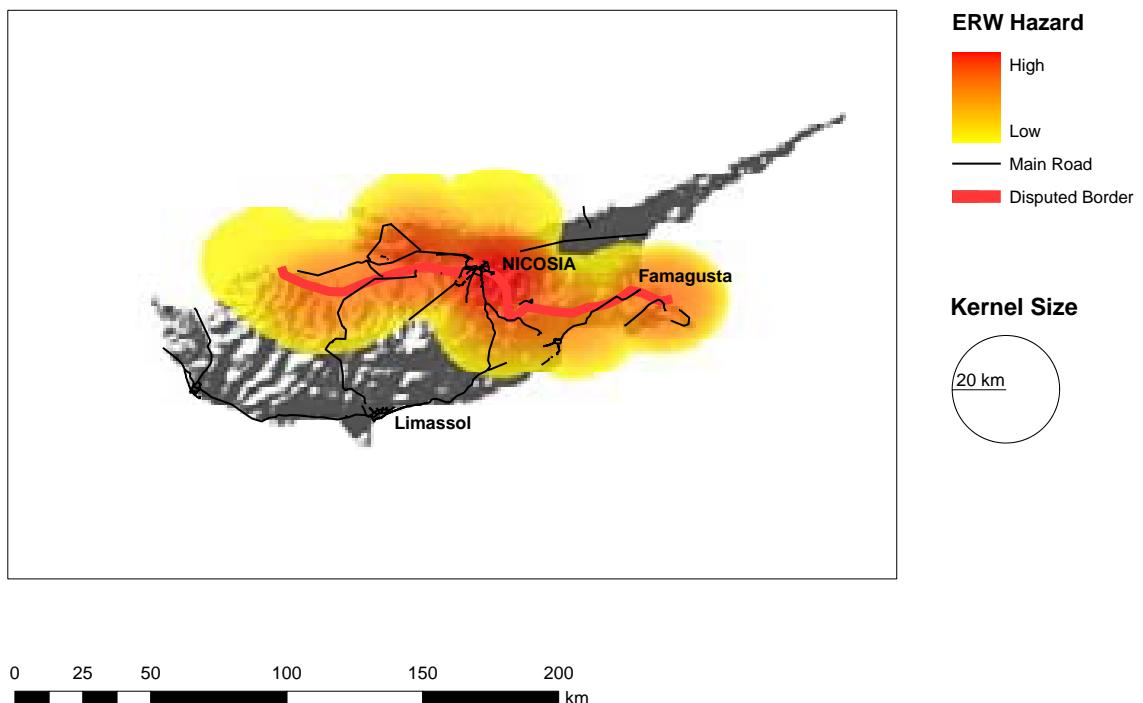


fig. 29: Cyprus, regional level (not showing the current situation)

ERW Hazard in Afghanistan: 200 km Kernel Size

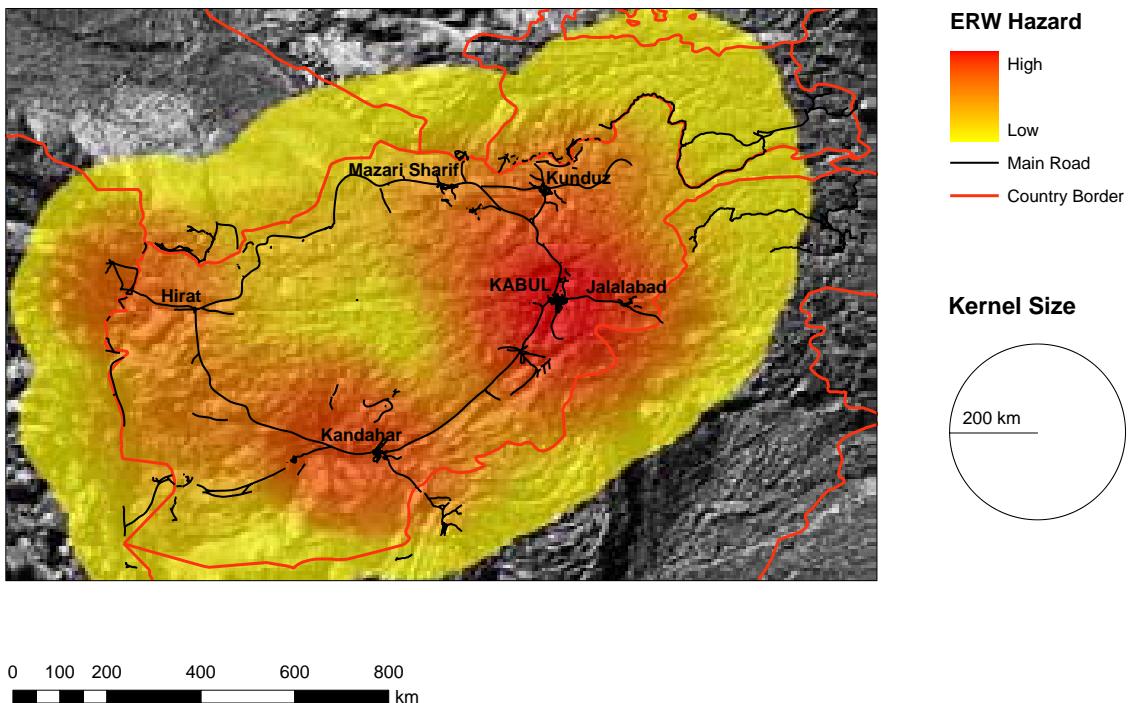


fig. 30: Afghanistan, global level (not showing the current situation)

ERW Hazard in Somalia: 100 km Kernel Size

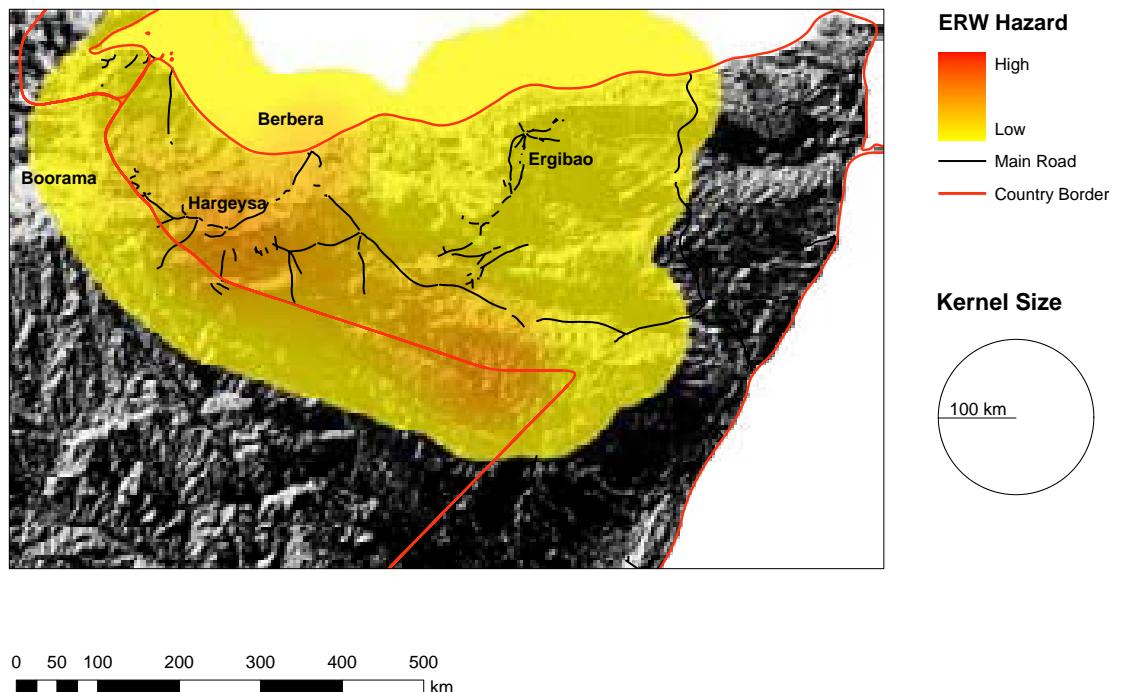


fig. 31: Somalia, continental level (not showing the current situation)

ERW Hazard in Laos: 50 km Kernel Size

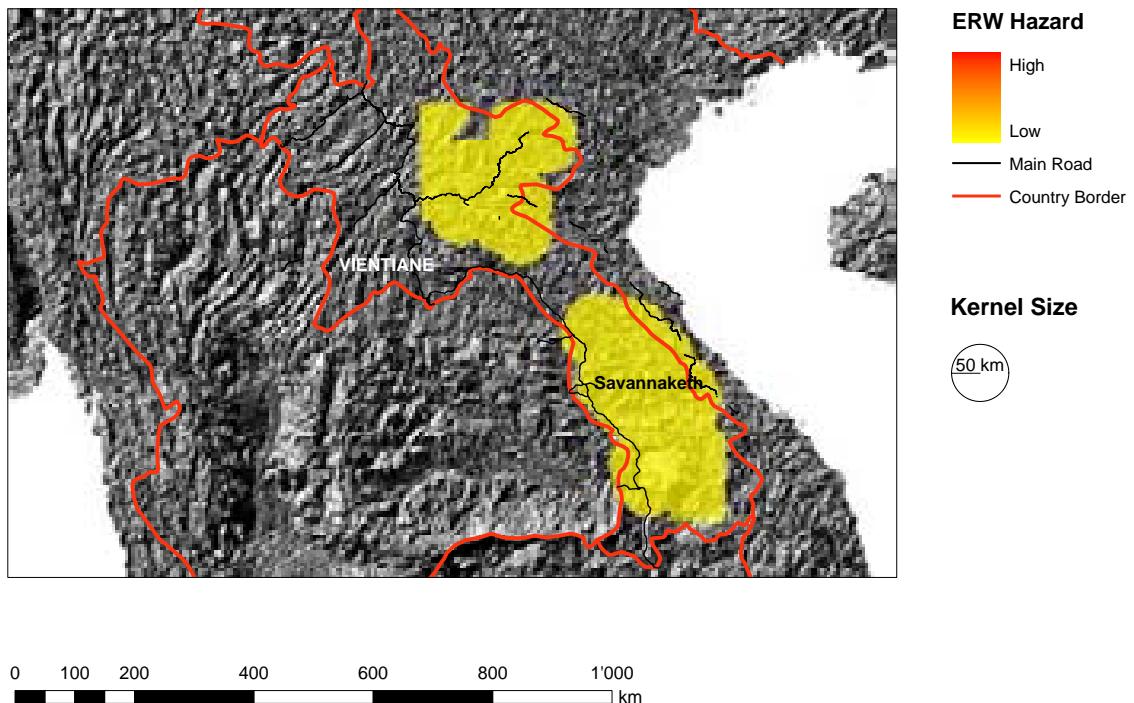


fig. 32: Laos, national level (not showing the current situation)

Looking at these representations, one may wonder where the many hazards of Laos have ended up or why Cyprus with relatively few hazards seems to be affected quite heavily. To explain this one has to be aware that for all representations the same color ramp with the same classification intervals has been applied, namely one based on geometric intervals. Reality is that although umpteen hazards are present in Laos, only a subset of the data was available. Therefore, compared to other countries its estimated overall hazard level is crucially smaller. Especially at the larger extents the range of values for this country lies at the lowest end of the color ramp. Although the low hazard concentrations now may be overrepresented, the selected classification and color ramp avoided to let low-affected countries disappear completely. Hence, displaying any value higher than zero prevents the exclusion of lightly affected countries.

5.3 Modifying parameters

In this section the effects of modifying the most important parameters are presented. By changing the kernel size or by applying different classification methods the results may vary heavily.

5.3.1 Varying kernel size

As mentioned in section 4.4.4, the choice of the kernel size has an important impact on the results of the maps. Thus, several different kernel sizes have been tested. The following figures feature results of KDE, by changing its kernel size.

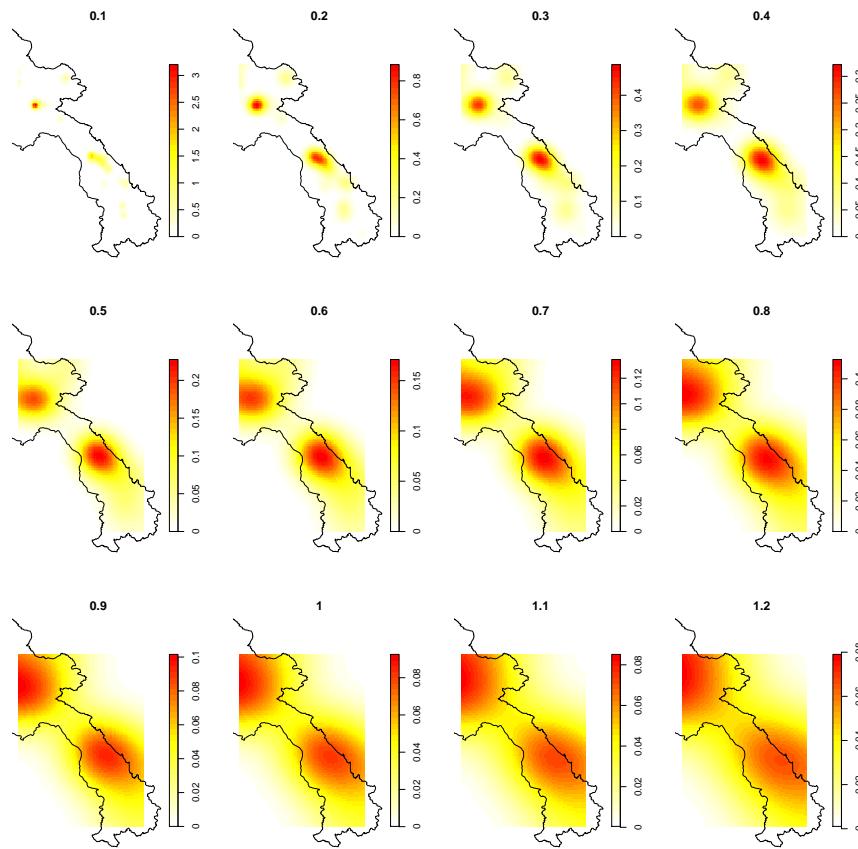


fig. 33: Linearly growing kernel size (in decimal degrees) for Lao PDR

In fig. 33 the effects can be seen for the Lao PDR dataset. A very detailed pattern (with a small kernel size) changes into a representation of only two very generalized main hazard areas. Since only few records are available, with a very small kernel size, it is almost possible to recognize individual records.

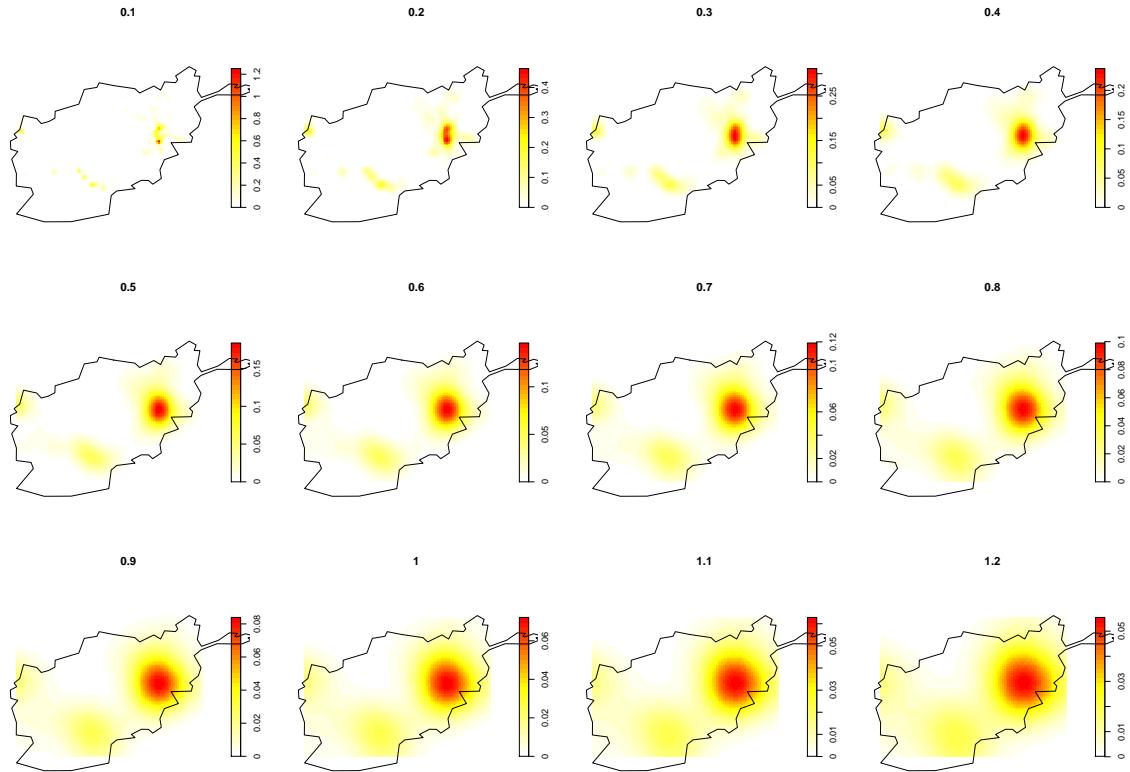


fig. 34: Linearly growing kernel size (in decimal degrees) for Afghanistan

In fig. 34 the same effects are demonstrated for Afghanistan. While in the plots on the top the distribution is still visible quite well, on the lower plots there is only one main focus in the east of the country. The details get fuzzier and the message becomes different. The same parameters for KDE have been chosen as for Laos.

In fig. 35 the effects of a growing kernel size for Somaliland are presented. Having a concentration at the border, values in other countries' territories are estimated like it is the case in Somaliland. This is not completely wrong, since the probability for more hazards is very high on the other side of the borderline. Nevertheless, it can be possible that data of the neighbor country are available, which are incompatible. (See also section 4.7)

In a very small country like Cyprus large kernel sizes are not very plausible, thus this country will be presented only as a marker for the smaller scales.

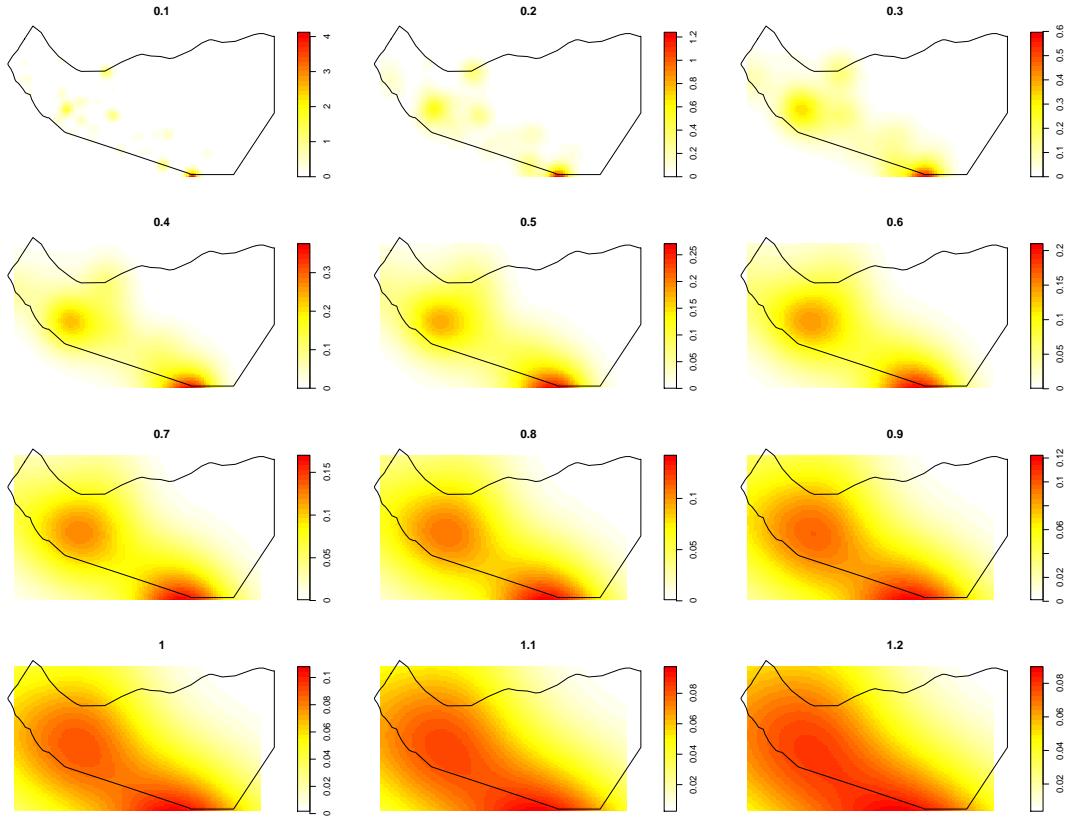


fig. 35: : Linearly growing kernel size (in decimal degrees) for Somaliland

Of course, on a detailed scale it does not make a lot of sense to work with very large kernel sizes. Nevertheless, if we are looking at a global scale, shapes with very small kernel sizes will not be visible anymore, thus a large kernel size is recommendable to make worldwide ERW contamination visible.

An important detail one should not miss is the classification. In the plots above for every representation the class intervals have been adjusted automatically, in order to make the effects of the growing kernel sizes visible. The classifications are always continuous, and depending on the maximum values intervals are chosen correspondingly. In section 5.3.2, this and other classification methods and their effect on the representations are illustrated.

Another way to represent the calculated KDE is to present them as three-dimensional plots. By having a look at fig. 36, it is possible to see how ERW contamination is represented as the height of the density surface generated by KDE.

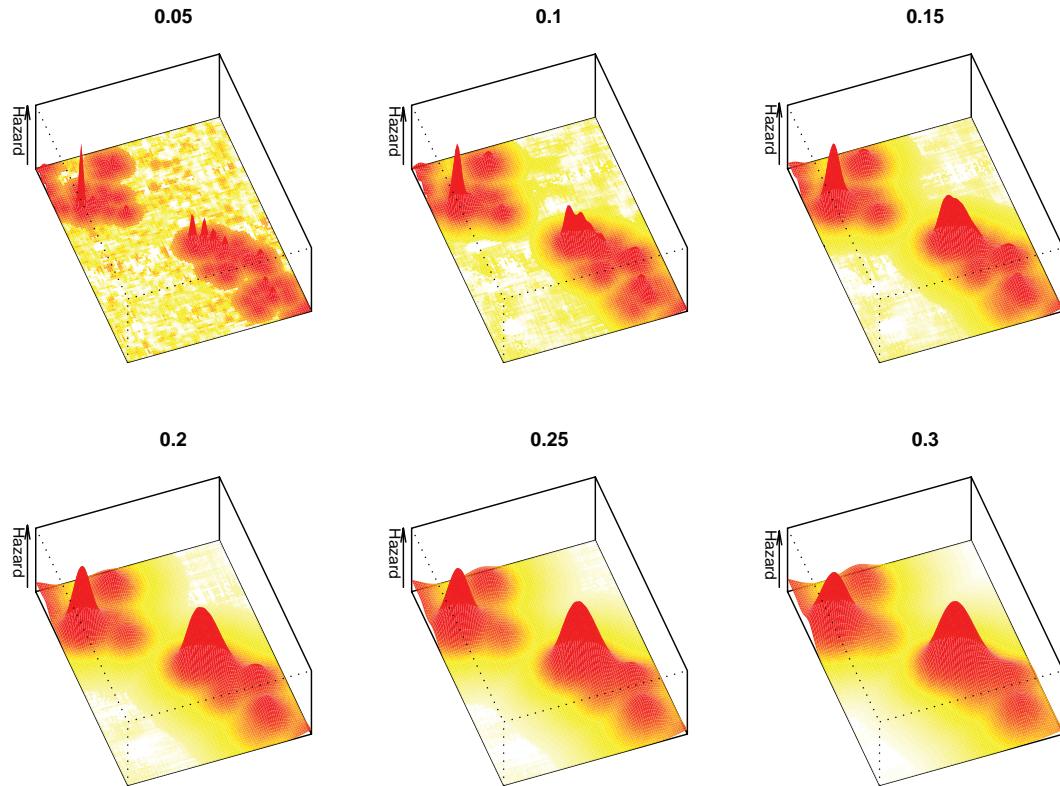


fig. 36: 3D-plot of the Laos hazards, with an linearly growing kernel size (in decimal degrees)

This figure illustrates the effect of an increasing kernel size: the surface gets smoother and the information is blurred. The functionality of the KDE can be comprehended quite well, as one can see how the “bumps” are getting flatter by reducing the kernel size. For more examples of this kind, see Appendix E.

5.3.2 Classification methods

As described in section 4.5, there are numerous methods to classify the values computed with KDE. In the following illustrations, differences between few or many classes and existing classification methods are presented. The class intervals have been calculated with R, but results that have been computed with ArcGIS are similar to identical.

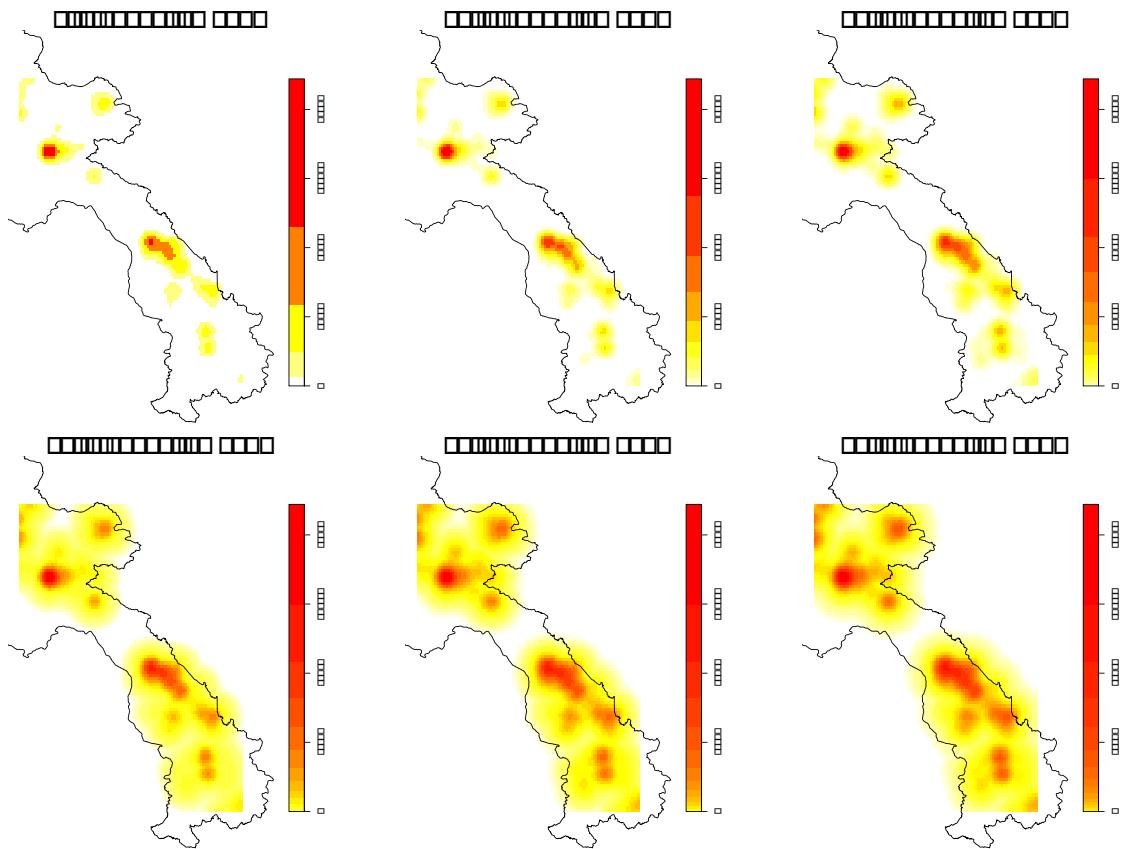


fig. 37: Classification with k-Means clustering method with different number of classes

In fig. 37, the hazard data of Laos is classified using the k-Means clustering method. In this case the values are divided into k clusters, in which the values are assigned to the cluster with the nearest mean (MacKay 2003). By adding classes the representation changes significantly. Not only the color ramp gets smoother, also the general level of hazard contamination seems to become stronger. Less values fall into the first class that has a white color. Thus, the influence of the color ramp and the number of classes can be tremendous.

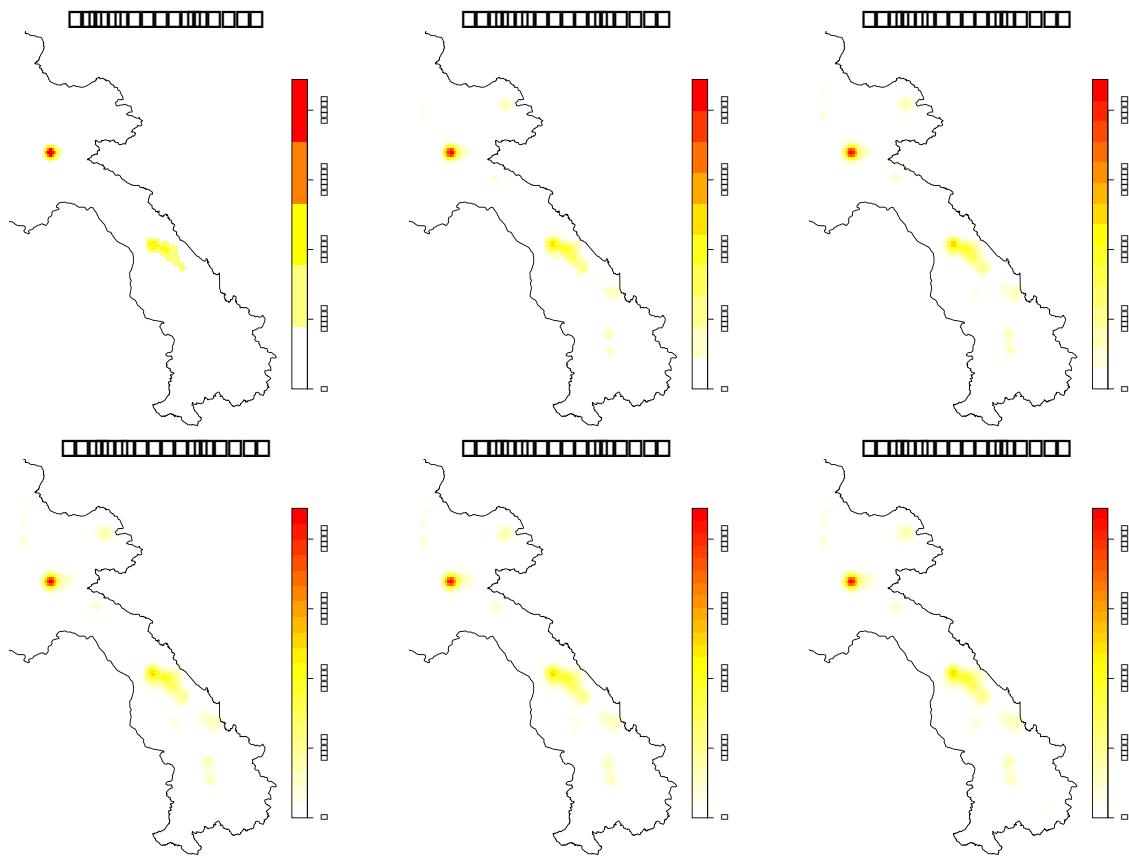


fig. 38: Growing number of classes with the equal interval method

In fig. 38, the effect of adding classes can be seen. In the case of an equal interval method (described in section 4.5.1) it is very small. By only adding more intervals the level of ERW contamination remains constant, but the level of detail increases. Since the values are not at all distributed equally, displaying the generated maps with this classification method is not recommendable. A classification has to be found that respects the values' distribution.

Eventually, for the representation of the hazard rasters a variant of the geometric interval method has been chosen. A geometric series has been generated which the class intervals follow, similar to the one proposed by ArcGIS. To improve the representation, the values of the coefficient and the number of classes before the coefficient changes have been adjusted.²⁸ Choosing the size of the first interval has a strong influence on the output and also determines the smallest values that are going to be displayed. The enormous impact on the results is presented in fig. 39.

²⁸ The source code for the algorithm is added in Appendix G.

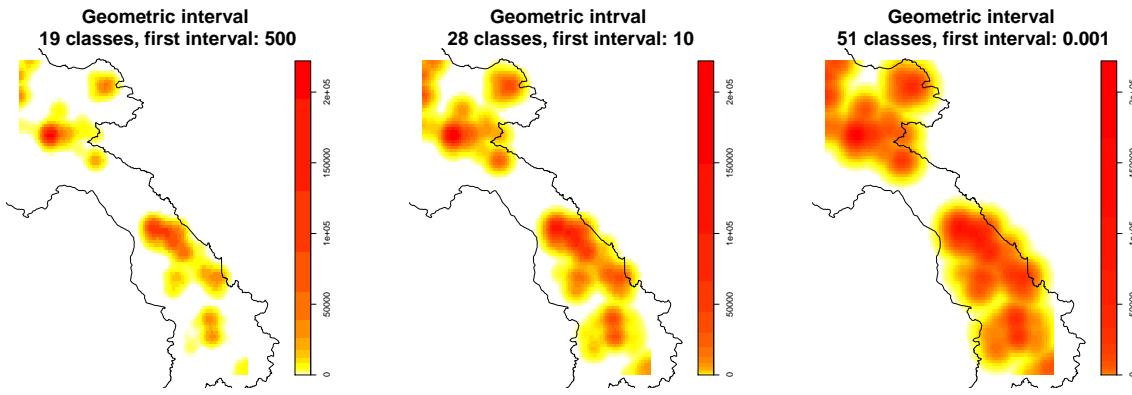


fig. 39: Geometric Interval, various geometric series of class intervals

By changing the parameters of the geometric interval classification, the representation can change crucially. For the three plots in fig. 39, the same data have been used and the same KDE with the same point data have been performed. Only the first interval of the geometric interval method has been modified. One can see that by only altering one parameter, the image can differ dramatically and its message can change crucially. Hence, choosing the method and its parameters is a very decisive step and has an immense impact on the result, but by adjusting some settings a lot can change again. These consequences were discussed intensively with mine experts and it was decided to use create a series of intervals that fits all tested countries. The essential difference of the geometric intervals is the fact that by choosing a very small first interval small values are visible, but by adding exponentially growing intervals, large values can be represented as well.

In spite of the countries' very different levels of contamination, using this classification method allows applying the same class intervals to all of them, since it respects the quadratic characteristics of the KDE.

5.4 Histograms of the Density Maps

The generated rasters provide a good overview of the hazards' geographical distribution. To go deeper into the data, histograms of the density maps and statistical measures have been plotted. The frequencies of hazard values have been computed and the results plotted as a histogram. These histograms have been generated for all countries and all kernel sizes and they show that the hazard values are distributed in a very skewed way (see fig. 40). Most cells have relatively low hazard values, only little higher than zero (zero values are excluded); only few of them have significantly high values. The histograms also make the variation of the kernel size visible. The larger the kernel size, the more skewed is the histogram. Thus, in histograms of a smaller kernel size, the distribution is barely recognizable. Hence, it is recommendable to plot a logarithmic histo-

gram to make the data better visible (Sigworth & Sine 1987). Logarithmic histograms have been plotted and one combining all hazards is presented in fig. 41.²⁹

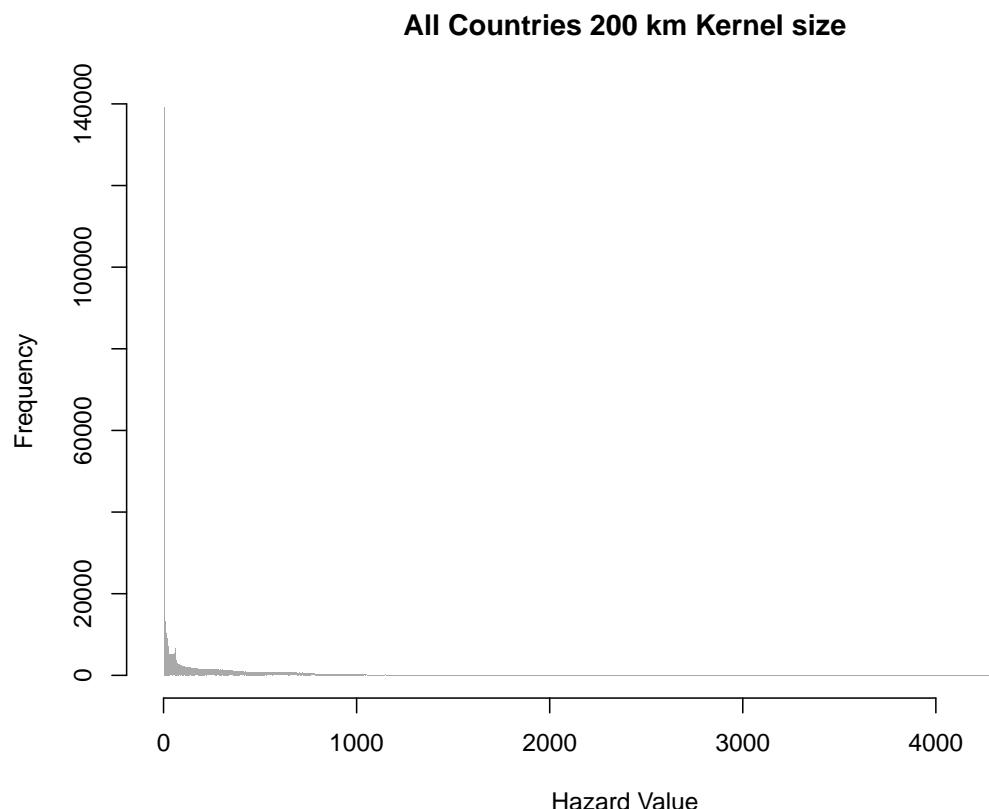


fig. 40: A histogram of the results from the four countries combined

²⁹ For more histograms see Appendix C

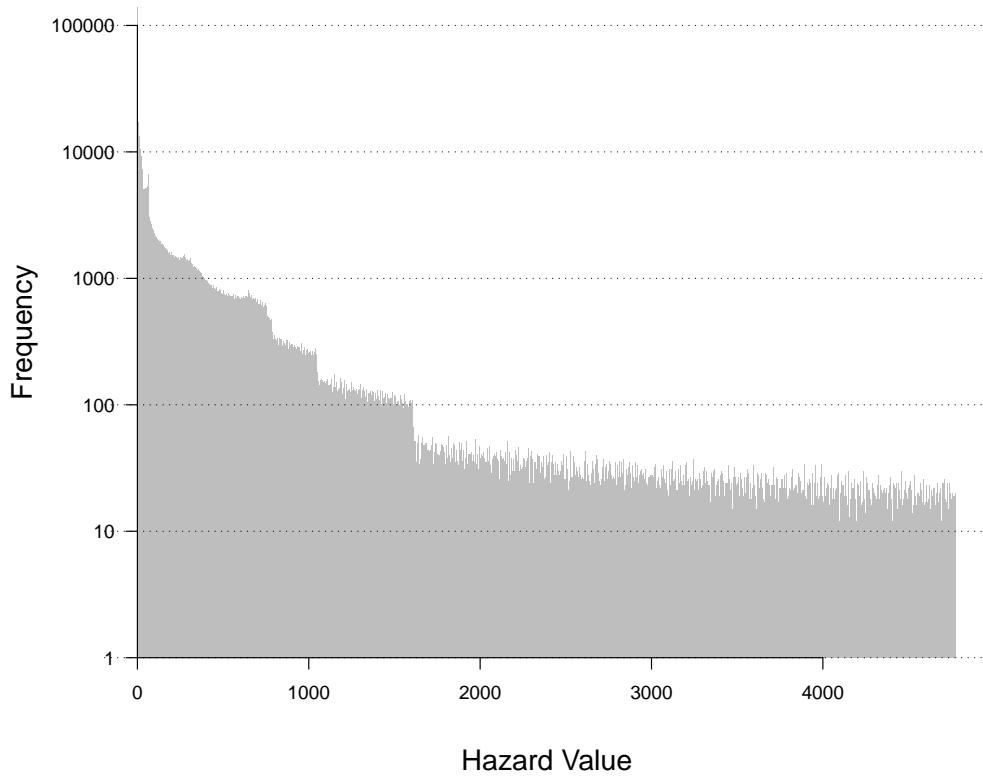


fig. 41: A logarithmic histogram of all values combined (kernel size: 200 km)

Representing a logarithmic histogram of the distribution allows a much more differentiated view of the hazard values' distribution. There are only few values that show up very often, which are generally low. This is influenced mainly by the characteristics of the KDE. Also, one can see that there are some steps or jumps, although the frequency is dropping more or less constantly. The jumps that are visible are due to the fact that this histogram is a combination of four others. Only one country has values of more than 2000, thus fewer values are in this range. In the lower range the values are superposing each other.

Furthermore, main statistical values have been calculated and are presented in tab. 2.

Kernel size	20 km	50 km	100 km	200 km
Mean	1'149.80	624.33	446.56	319.60
Max	88'299.91	22'322.61	11'663.20	4'775.49
Median	82.45	54.52447	53.56	55.99
Std Dev	3'860.64	1'765.157	1'109.05	648.94

tab. 2: Statistical values of the generated raster maps

There is regularity behind these values. Due to the three-dimensional shape of the kernel function, the resulting statistical values can be estimated. They are inversely proportional to the size of the kernel, i.e. the greater the kernel, the smaller the values. The fitted power function is estimated as follows, with an R^2 of 0.99:

$$\text{average hazard} = 5779 \cdot \text{kernel size}^{-0.55}$$

This function is based on the values presented in the table above, which have been derived from the combined results of the analysis of generated rasters. A curve was fitted the frequencies of the hazard values, which has the parameters mentioned above. Having this power function, the quadratic characteristics of the KDE become clearly visible. By increasing the kernel size, the “bumps” flatten and get smoother and consequently the average values decrease. With small kernel sizes, the “bumps” have a smaller area, but they are much steeper. Hence, the relation between kernel and values is known and can be integrated into further calculations of ERW datasets.

5.5 Combination with road data

Of course the calculated hazard values provide numerous ways to make further analyses, limited almost only by the availability of data. Naturally, one has to be careful not to jump to conclusions, by interpreting the results in the wrong way. As an example, a simple analysis of the proximity to roads has been made and its results have been examined.

By counting the hazards inside a buffer around main roads, the relation between hazard concentrations and the road network could be observed. The nearness to roads can also be a rough estimate for accessibility. For that reason data from Open Street Map³⁰ have been combined with the hazard data. Unfortunately, the available road data is inconsistent and the denominations vary for every country. Nevertheless, to calculate this example, for any country its main road types have been chosen and with their spatial attributes the hazard values have been calculated.

The results have been compared with random points that have been scattered inside the country border³¹.

³⁰ “OpenStreetMap creates and provides free geographic data such as street maps to anyone who wants them. The project was started because most maps [...] have legal or technical restrictions on their use, holding back people from using them in creative, productive, or unexpected ways.” (OpenStreetMap 2010)

³¹ For this purpose, with ArcMap random points have been generated inside the borders of Afghanistan and the same calculations as for real data has been made.

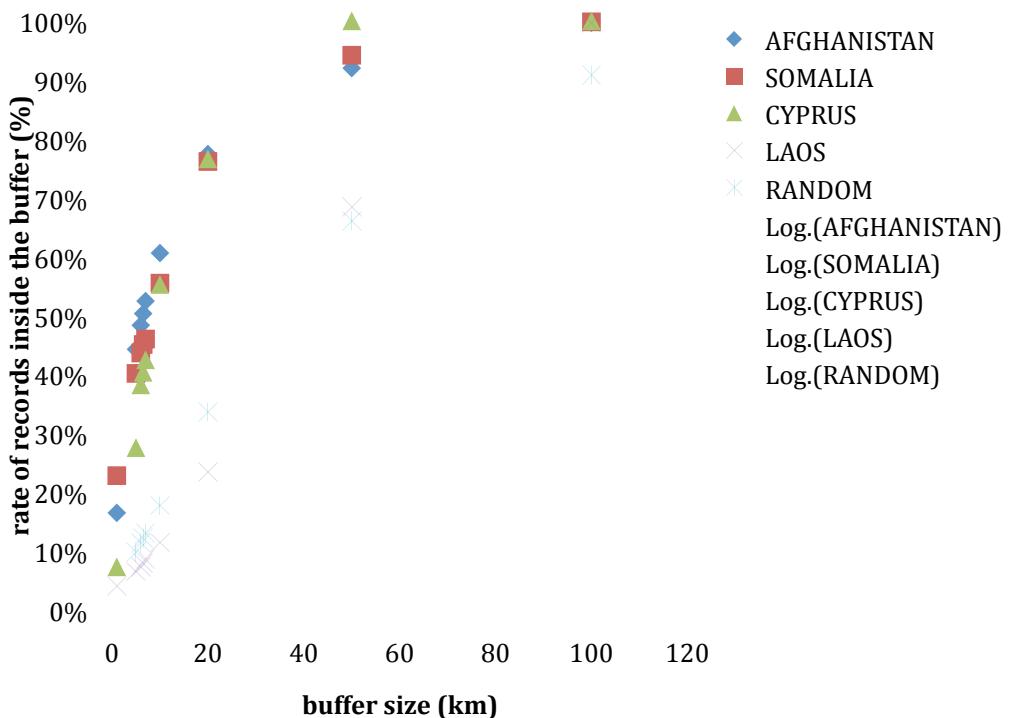


fig. 42: Scatter plot on hazard vs. road distance³²

	Afghanistan	Somalia	Cyprus	Laos	Random
R ²	0.99101	0.95797	0.97529	0.76608	0.97529

tab. 3: R² values for the four countries

In a scatter plot (fig. 42), hazard records' counts and their proximity to roads (from OSM (2010)) are presented. A logarithmical function can be fitted to the calculated values, whose R²-values are presented in tab. 3. The functions are very similar for Afghanistan, Somalia and Cyprus, while Laos seems to be more similar to random point distribution.

³² Map data by OpenStreetMap contributors, CC-BY-SA (OSM 2010)

6 Discussion

Applying GIScience methods on landmine, UXO and other ERW data has been shown to be a adequate way to create maps that show ERW contaminations on a regional to global scale. Since the geographic locations of hazards have a large impact on affected societies, analyzing and visualizing their distribution helps to deal with them and to make the public aware of the problem. Only with the adequate application of GIScience techniques was it possible to achieve these goals. Eventually, it became possible to make a problem visible that before could only be described by numbers of casualties and hectares or square kilometers of hazardous areas, which can now be better realized. Maps are great tools to communicate geographic phenomena, and knowledge can be disseminated to the person who reads them, as described by MacEachren (2004). Finally, the maps of the generated hazard rasters allow the public to see where in the world there are high hazard concentrations to be expected. By looking at more detailed scales regional differences can be observed. This global problem, which was very difficult to grasp before, now can be presented in a convincing way and thus raise the understanding and awareness of it.

It has been shown that there are many different possibilities of how to estimate hazard densities, how to set parameters, and how to represent the results. There are no right or wrong solutions; the important thing is to be aware of the consequences of the choices. By changing the kernel sizes not only the characteristics change, but also the estimated values and their average. The effects are different on small and large countries. While a KDE with a large kernel size can give an interesting overview of hazardous areas in large-area countries, there are no reasonable insights for small-area countries. On the other hand, to represent a global distribution with smaller kernel sizes makes the output impossible to read. Representing the data at different levels of detail, with corresponding kernel sizes allows different perspectives. A similar problem is manifested in the huge differences of levels of ERW contamination between the countries. Only a very sophisticated color ramp can make low hazard concentrations visible, without rendering hazard estimates of heavy affected countries illegible. Having a color ramp that suits one country well does not mean that it will work with all other countries; either some countries could disappear completely or otherwise be represented as entirely hazardous. Hence, a compromise had to be found that works well with most countries, while keeping the global aspect of this topic always in mind. The customized geometric interval method provided an appropriate solution to this problem, since it respects the skewed distribution of the hazard values.

The requirements concerning non-disclosure were approached firstly with the KDE which has a blurring effect and secondly by choosing not too small kernel sizes and resolutions, since larger kernel sizes will not reveal any secret information. Using further obfuscation methods, such as perturbation of point data, guarantees a high degree of confidentiality. Despite these restrictions

are present, it is possible to generate detailed high-resolution maps. Instead of generating only one map, but different maps for four different levels of details, it is possible to show a clear image of the situation up to the regional level, without breaching security constraints.

A cell size of at least 200 m for the most detailed representation turned out to be a good solution for representing the hazard data distribution. The cell sizes for the other levels of detail follow the size of the kernel and do not have large impact on the results. Thus, the requirements for computing time and the demands for a readable map can be fulfilled with the chosen resolutions.

The representation in a raster format also makes comparing hazard concentration with other data possible, in order to find other correlations and dependencies. It is already assured that the results of this thesis will be used for a series of evaluations at GICHD. Furthermore, making the final product viewable on the Internet will allow spreading reliable information to a broad spectrum of people and their feedbacks can be included for further enhancements.

In order to develop an appropriate visualization and to better understand how the data have been generated, the histograms of the calculated hazard values have been examined. The strongly skewed distribution presented a problem that could be solved by choosing a suitable classification method and an adequate color ramp.

Most data classification methods shaped turned out not to be convincing for a good representation. Either they do not respect the classification at all (equal interval, defined interval, standard deviation) or they cannot be applied on the dataset easily, since there are likely to be frequent data updates (quantile, k-means). However, a convincing solution could be presented. Instead of taking one of the predefined solutions of the software system used, it was more adequate to develop a method that respects not only the distribution of the values but also the situation in different countries. By customizing a geometric interval classification it became possible to present countries with very low values and very highly affected countries on the same map, thus making them comparable. Furthermore, a classification that follows growing intervals takes into account the nature of the KDE that implies a quadratic relationship, since points are aggregated over an area to produce a density surface.

The distributions of the point patterns have been analyzed and it could be confirmed statistically that hazards are distributed neither randomly nor following a Poisson distribution. Thus, their locations are related to other geographic objects that can be examined. Nevertheless, the statistical measures change form country to country, for example the Lao data tend more to a Poisson distribution than others. This can be traced back to different origins of the ERW deployment, such as bombing instead of laying minefields.

A relation to an influencing factor has been shown in section 5.5, where a connection between hazards and other data was found. The comparison of ERW data and road data showed some interesting connections. The presented results suggest that in some countries most hazards are concentrated around important traffic connections, while remote areas seem to be less affected. For random values the connection is different, unsurprisingly their correlation is much smaller. Of course there are more hazards inside larger buffers, since the area covered by buffers gets

larger, nevertheless, the coefficient of determination is higher with a fitted logarithmical function for the real data. Surprisingly, the Lao data have similar values to the randomized dataset. Of course, there can be various conclusions to this, which mostly can be neither confirmed nor denied. Obviously, mines are more likely to be laid close to roads, because of the better accessibility and most battles are fought in areas where the channels of supply are short. But it must not be ignored that most hazards in Laos arose during the Vietnam War, mostly by dropped cluster bombs, which left over a huge amount of UXO. Landmines have been delivered from airplanes and many carpet bombings were carried out (Barrett 2000). This could suggest that they have not been dispersed mainly close to roads, but on wide areas of the country, while minefields in other countries tend to be placed in accessible, strategically important areas. The different relationship could also depend on the road network that has another form or is less developed, or on other influences. Of course remote areas are less likely to have complete data than areas close to important transport links. Indisputably, there are many other spatial entities somehow connected to the distribution of the hazard points, e.g. country borders, settlements, resources, etc. With the right data available, these relations could be examined.

Finally, the generated maps have been presented in the way they will eventually look in the target application. The figures in section 5.2 show possible examples with the color ramp (geometric interval) and the different kernel sizes calculated for real data in the four investigated countries. A color ramp from yellow to red was chosen, which gives a good impression of ERW contamination. Any value higher than zero (affected areas) will be colored yellow, and if more heavily affected, the color changes to orange and red. Thus, any hazardous area will be visible and heavily affected zones can be distinguished by the hue. Contaminated areas in weakly affected countries will be shown with mainly yellow colors, especially at representations calculated with large kernel sizes.

The same method that has been applied to the four countries (Afghanistan, Cyprus, Laos and Somaliland) can be used to generate rasters for any ERW affected country that gathered spatial data using IMSMA. The color ramp and the kernel function have been tested on other countries and the results have been very satisfying, since the worldwide ERW hazard threat was easily recognizable and local differences could be examined on the regional level. It is expected that after receiving feedbacks from mine action authorities, demining experts, etc. the method will be re-adjusted and fine-tuned. When published, this map will be the first of its kind to visualize the ERW problem worldwide.

However, the full experience of the results will only become available when the target map application will be ready and online. It will be possible to examine various countries on different continents and the user can navigate inside the application, thus he or she will be able to gain new perspectives of the topic. The way in which the maps are represented will be very similar to those printed in this thesis. Furthermore, the user will be able to view other layers in combination with the generated hazards maps. A detailed description of the topic will be presented and interested persons will be enabled to make their own interpretations.

7 Conclusion

Needless to say, generating maps of landmine and unexploded ordnances hazards and estimating their emanating threats cannot solve the problem of explosive remnants of wars neither can it be done by calculating statistical values or collecting large amounts of related data. Organizations and mine action authorities in affected countries have to deal with the problem on-site and to operate on the local level with machines, animals and manpower.

Nevertheless, this highly valuable work of mine action has to be made as efficient as possible and this is where this thesis tries to make a contribution. By turning the available data into a meaningful visualization, people will hopefully not only become aware of the problem, but resources can be better deployed and priorities can be set with a better basis of decision-making. Donors, mine action authorities and any kind of interested persons can see where there is need for action. Moreover, people get aware of this problem, which affects the lives of millions of humans and poses a threat to entire societies and economies.

By generating reliable rasters and by guaranteeing confidentiality, data providers in affected countries are encouraged to furnish usable ERW data, which then can be used to improve the model and its representation. Consequently, with information of more affected regions other research can be done, e.g. to estimate socioeconomic impacts.

Time will tell if the efforts of this thesis will have a sustainable influence on mine action and on how this problem is perceived by the public. Subsequent studies and projects will be crucial for a lasting collaboration of GIScience and mine action that could provide benefit for affected countries and their societies.

7.1 Achievements

By working on this thesis various achievements have been made, which can be useful for further developments in other projects or to respond to future challenges:

- The kernel density estimation could be applied successfully to generate hazard rasters. Its functionality has been studied, its parameterization has been adjusted to the given requirements by systematic experiments and the results interpreted.
- The preparation and processing of ERW data turned out to be not at all trivial and many lessons had to be learnt on how to deal with real datasets that have a global extent.
- A set of classification methods have been presented and evaluated, and one of them has been customized to better fit this project's requirements.

- The distribution of the hazard's point patterns has been analyzed and differences in the spatial randomness between different countries could be discovered. Possible reasons for these differences have been formulated.
- Solutions for the adequate handling, display, and processing of sensitive data have been proposed. Guaranteeing a high level of confidentiality is expected to encourage countries to furnish more of their data.
- Examples of how the results will be eventually displayed in a productive application that is currently under development have been presented and the way in which generated maps will be made available has been described.

7.2 Reflections / Critique

Of course there are still some difficulties and problems that this thesis that could not solve. To be sure to make the right choice of methods and their parameters, it would have been necessary to test the generated results and their representation. Ideally, real data could be compared with the estimated hazard values, which would allow validating the model or estimating its suitability. Unfortunately, nothing about this aspect can be said by now.

Further, it would have been desirable to present a worldwide representation of the ERW problem, but unfortunately the data are not yet available. Moreover, the definite representation method has not been selected yet. Thus, only possible solutions could be presented and these still have to be evaluated among experts in mine action.

The available data had to be simplified, thus possibly helpful available information had to be disregarded. It was not possible to include other information (e.g. accidents or victims) and the weighting of the different kinds of hazard still is not solved completely. The effects of changing them should definitely be explored in future.

7.3 Future Projects

Using knowledge and data gained with this project, a series of further studies is planned at GICHD. The four layers listed in Appendix B will be examined carefully. The methods developed and evaluated in this thesis will be used as a base for these projects. In the near future the model will be applied on at least 10 more countries, which will allow to evaluate the methodology in order to get closer to reality. This thesis can be regarded as the design and the evaluation of the basic data, which will be used for further projects, in which I will participate as an intern at GICHD.

It would also be interesting to create time series in order to show where hazard reduction has been accomplished or where new hazard have been deployed or recorded. This will be possible as

soon as the project has run for a certain time. Moreover, if datasets of previous years and decades could be obtained, it would be possible to make a historical analysis.

This thesis is just one part of the much larger SERWIS project. The layers explained in Appendix B will have to be developed further and evaluated, assuming that the required spatial data will be available. Having these layers will give interested persons deeper insight into this problem that still has a large impact on societies all over the world.

8 Bibliography

- Andersson, N., Palha da Sousa, C. & Paredes, S., 1995. Social cost of land mines in four countries: Afghanistan, Bosnia, Cambodia, and Mozambique. *BMJ*, 311(7007), 718 -721.
- Aurenhammer, F., 1991. Voronoi diagrams - a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23(3), 345-405.
- Baddeley, A. & Turner, R., 2005. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6), 1-42.
- Barrett, N., 2000. Holding Individual Leaders Responsible for Violations of Customary International Law: The US Bombardment of Cambodia and Laos. *Columbia Human Rights Law Review*, 32, 429.
- Benini, A.A., Moulton, L.H. & Conley, C.E., 2002. Landmines and Local Community Adaptation. *Journal of Contingencies and Crisis Management*, 10(2), 82-194.
- Berman, M., 1986. Testing for Spatial Association Between a Point Process and Another Stochastic Process. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 35(1), 54-62.
- Bilukha, O.O., Brennan, M. & Woodruff, B.A., 2003. Death and injury from landmines and unexploded ordnance in Afghanistan. *JAMA : The journal of the American Medical Association*, 290(5), 650-3.
- Bivand, R.S., Pebesma, E.J. & Gómez-Rubio, V., 2008. *Applied Spatial Data Analysis with R*, New York, NY: Springer New York. Available at:
<http://www.springerlink.com/content/uw07v1/#section=147788&page=1> [Accessed October 1, 2010].
- Boba, R., 2005. *Crime analysis and crime mapping*, Thousand Oaks, CA: Sage.
- Bollinger, T., *IMSMA for Operations: An Overview*, GICHD.
- Boucher, C. & Altamimi, Z., 2001. ITRS, PZ-90 and WGS 84: current realizations and the related transformation parameters. *Journal of Geodesy*, 75(11), 613-619.
- Brassel, K.E. & Weibel, R., 1988. A review and conceptual framework of automated map generalization. *International Journal of Geographical Information Systems*, 2(3), 229.
- Burns, J.F., 2008. Britain Joins a Draft Treaty on Cluster Munitions. *The New York Times*. Available at:
http://www.nytimes.com/2008/05/29/world/europe/29cluster.html?_r=2&em&ex=1212206400&en=0e28528fd0c65072&ei=5087%0A [Accessed August 25, 2010].

- Chainey, S. & Ratcliffe, J., 2005. *GIS and crime mapping*, London: John Wiley and Sons. Available at: <http://eprints.ucl.ac.uk/11293/> [Accessed August 5, 2010].
- Cromley, E.K. & McLafferty, S., 2002. *GIS and public health*, New Yourk: Guilford Press.
- CSS, 2005. Information Management System for Mine Action (IMSMA). Available at: http://www.css.ethz.ch/services/imsma_EN [Accessed March 22, 2010].
- Dale, M.R.T., 2000. *Spatial pattern analysis in plant ecology*, Cambridge University Press.
- De Smith, M.J., Goodchild, M.F. & Longley, P., 2007. *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*, Troubador Publishing.
- Delhay, S., Idrissa, M. & Lacroix, V., 2005. PARADIS: GIS Tools for Humanitarian Demining. In *The 2nd International ISCRAM Conference*. pp. 213–219.
- Diggle, P.J., 1986. Displaced amacrine cells in the retina of a rabbit: analysis of a bivariate spatial point pattern. *Journal of Neuroscience Methods*, 18(1-2), 115-125.
- Docherty, B., 2009. Breaking New Ground: The Convention on Cluster Munitions and the Evolution of International Humanitarian Law. *Human Rights Quarterly*, 31(4), 934–963.
- Donthu, N. & Rust, R.T., 1988. A Programming and Positioning Strategy for Cable Television Networks. *Journal of Advertising*, 17(4), 6-13.
- Donthu, N. & Rust, R.T., 1989. Estimating Geographic Customer Densities Using Kernel Density Estimation. *Marketing Science*, 8(2), 191-203.
- Duckham, M. & Kulik, L., 2005. A formal model of obfuscation and negotiation for location privacy. In *Pervasive*. pp. 152–170.
- Eck, J.E. et al., 2005. *Mapping crime: understanding hot spots*, Washington, DC: National Institute of Justice. Available at: <http://eprints.ucl.ac.uk/11291/> [Accessed February 26, 2010].
- Eriksson, D., 2008. Total Quality Management in Mine Action. *Journal of Mine Action*, 12(1).
- ESRI, 2007. ArcGIS Desktop Help 9.2. Available at: http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Density_calculation_s [Accessed August 4, 2010].
- ESRI, Powerful Spatial Statistics Tools in ArcGIS 9. Available at: http://www.esri.com/news/arcuser/1104/spatial_statistics.html [Accessed September 8, 2010].
- Farley, B.R., 2010. Calling a State a State: Somaliland and International Recognition. *Emory International Law Review*, 24(2). Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1676428#%23 [Accessed October 5, 2010].
- Fischer, G. et al., 2008. *Global Terrain Slope and Aspect Data GAEZ*, ed., IIASA, Laxenburg, Austria and FAO Rome, Italy: Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008).

- Flowerdew, R. & Green, M., 1992. Developments in areal interpolation methods and GIS. *The Annals of Regional Science*, 26(1), 67–78.
- Gatrell A.C. et al., 1996. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers*, 21, 256-274.
- Gennick, J., 2008. *SQL Pocket Guide* 2nd ed., O'Reilly Media.
- Geofabrik GmbH, 2010. GEOFABRIK. Available at: <http://www.geofabrik.de/> [Accessed October 14, 2010].
- GICHD, 2010a. *Computerised Information Management designed for Mine Action*, Available at: <http://www.gichd.org/operational-assistance-research/information-management-imsma/imsma/overview/>.
- GICHD, 2010b. GICHD Website. Available at: <http://gichd.org/> [Accessed September 16, 2010].
- GICHD, 2009. *IMSMA User Guide*, Geneva: Geneva International Centre for Humanitarian Demining.
- GICHD, A., 2010. *Guide to Mine Action* 4th ed., Geneva: GICHD.
- Goodchild, M.F., 1992. Geographical information science. *International Journal of Geographical Information Science*, 6(1), 31–45.
- Goodchild, M., Haining, R. & Wise, S., 1992. Integrating GIS - and spatial data analysis: problems and possibilities. *International Journal of Geographical Information Systems*, 6(5), 407.
- Gruteser, M. & Grunwald, D., 2003. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *Proceedings of the 1st international conference on Mobile systems, applications and services*. San Francisco, California: ACM, pp. 31-42. Available at: <http://portal.acm.org/citation.cfm?id=1189037> [Accessed October 19, 2010].
- Hamilton-Merritt, J., 1999. *Tragic Mountains: The Hmong, the Americans, and the Secret Wars for Laos, 1942-1992*, Indiana University Press.
- Harris, G., 2002. The Economics of Landmine Clearance in Afghanistan. *Disasters*, 26(1), 49–54.
- ICBL, 2009a. *Cyprus Landmine Report 2009*, Available at: <http://www.mineaction.org/country.asp?c=62>.
- ICBL, 2009b. *Landmine Monitor Report 2009*, Mines Action Canada. Available at: <http://www.the-monitor.org/index.php/publications/display?url=lm/2009/>.
- Jenks, G.F. & Caspall, F.C., 1971. Error on Choroplethic Maps: Definition, Measurement, Reduction. *Annals of the Association of American Geographers*, 61(2), 217-244.
- Journel, A.G. & Huijbregts, C.J., 1978. *Mining geostatistics*, London: Academic Press.
- Keim, D. et al., 2004. Visual data mining in large geospatial point sets. *Computer Graphics and Applications, IEEE*, 24(5), 36-44.

- Kyung, L.Y., 2009. Noch tausend Jahre Bomben. *WOZ - Die Wochenzeitung*. Available at: <http://www.woz.ch/artikel/archiv/17561.html> [Accessed June 8, 2009].
- Lacroix, P. & Santiago, H., 2010. Contribution des SIG à la lutte contre les mine antipersonnel et / ou autres restes explosifs de guerre. *Géomatique Expert*, 7, 68-76.
- Lancaster, J. & J. Downes, B., 2004. Spatial point pattern analysis of available and exploited resources. *Ecography*, 27(1), 94-102.
- Langley, R.B., 1998. The utm grid system. *GPS world*, 2, 46–50.
- Lardner, T., 2008. *GICHD : Country*, Geneva: GICHD. Available at: <http://www.gichd.org/links-information-database/research-and-evaluation-reports/country/linksdb/evaluations/external-assessment-of-ddgs-humanitarian-mine-action-programme-in-somaliland/>.
- Legacies of War, 2010. Cluster Bomb Fact Sheet. Available at: <http://legaciesofwar.org/resources/cluster-bomb-fact-sheet/> [Accessed September 29, 2010].
- Levine, N., 2006. Crime mapping and the Crimestat program. *Geographical Analysis*, 38(1), 41–56.
- Lilliefors, H.W., 1967. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association*, 62(318), 399-402.
- MacEachren, A.M., 2004. *How maps work: representation, visualization, and design*, The Guilford Press.
- Machel, G., United Nations. General Assambly and UNICEF & United Nations Centre for Human Rights, 1996. *Impact of Armed Conflict on Children*, New York, NY: United Nations.
- MacKay, D.J.C., 2003. *Information theory, inference, and learning algorithms*, Cambridge University Press.
- Manzoni, P., 2008. *Crime Mapping als Instrument der Einbruchsprävention*, Zürich: Schulthess.
- Miguel, E. & Roland, G., 2006. The Long Run Impact of Bombing Vietnam. *SSRN eLibrary*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=877459 [Accessed September 2, 2010].
- Mitas, L. & Mitasova, H., 1999. Spatial interpolation. *Geographical information systems: principles, techniques, management and applications*, 1, 481–492.
- Morikawa, M., Taylor, S. & Persons, M., 1998. Deaths and injuries due to unexploded ordnance (UXO) in northern Lao PDR (Laos). *Injury*, 29(4), 301-304.
- New Oxford American Dictionary, 2001. *The New Oxford American Dictionary* Book only., Oxford University Press, USA.
- OpenStreetMap, 2010. OpenStreetMap Wiki. Available at: http://wiki.openstreetmap.org/wiki/Main_Page [Accessed October 2, 2010].

- OSM, 2010. Openstreetmap: Map feature Tags. Available at:
http://wiki.openstreetmap.org/Map_Features [Accessed September 2, 2010].
- Peterson, M.P., 1979. An evaluation of unclassed crossed-line choropleth mapping. *Cartography and Geographic Information Science*, 6(1), 21–37.
- Riese, S.R., 2001. *Estimating the probability of landmine contamination in non-combat environment*, Charlottesville: University of Virginia.
- Ripley, B.D., 2001. The R project in statistical computing. *MSOR Connections. The newsletter of the LTSN Maths, Stats & OR Network*, 1(1), 23–25.
- Rubin, B.R., 2000. The Political Economy of War and Peace in Afghanistan. *World Development*, 28(10), 1789-1803.
- Rushton, G., 2003. Public Health, GIS and Spatial Analytic Tools. *Annual Review of Public Health*, 24(1), 43–56.
- Seaman, D.E. & Powell, R.A., 1996. An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology*, 77(7), 2075–2085.
- Sigworth, F. & Sine, S., 1987. Data transformations for improved display and fitting of single-channel dwell time histograms. *Biophysical Journal*, 52(6), 1047-1054.
- Silverman, B.W., 1998. *Density estimation for statistics and data analysis*, Chapman & Hall/CRC.
- Sisavath, B., 2006. UXO Lao's Fight against Unexploded Ordnance. *Journal of Mine Action*, 28–29.
- Slocum, T.A. et al., 2008. *Thematic Cartography and Geovisualization* 3rd ed., Prentice Hall.
- Snekkenes, E., 2001. Concepts for personal location privacy policies. In *Proceedings of the 3rd ACM conference on Electronic Commerce*. Tampa, Florida, USA: ACM, pp. 48-57. Available at: <http://portal.acm.org/citation.cfm?id=501164> [Accessed October 19, 2010].
- Snow, J., Frost, W.H. & Richardson, S.B., 1965. *Snow on cholera*, Hafner New York.
- Spearin, C., 2001. *Ends and Means: Assessing the Humanitarian Impact of Commercialised Security on the Ottawa Convention Banning Anti-Personnel Mines*. YCISS.
- Tareke, G., 2000. The Ethiopia-Somalia War of 1977 Revisited. *The International Journal of African Historical Studies*, 33(3), 635-667.
- Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234-240.
- UNMAS, 2008. United Nations Mine Action Service Brochure. Available at:
<http://www.mineaction.org/doc.asp?d=1032> [Accessed August 4, 2010].
- UXO Lao, 2009. UXO LAO - Annual report 2008. Available at:
<http://www.uxolao.org/downloads.html> [Accessed June 10, 2009].
- Vosburgh, A., 2006. Mine Risk Education and the Amateur Scrap Metal Hunter. *Journal of Mine Action*, 10(2).

wiki.gis.com, GIS Glossary. Available at: http://wiki.gis.com/wiki/index.php/GIS_Glossary/ [Accessed September 21, 2010].

Worton, B.J., 1989. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*, 70(1), 164–168.

www.mineaction.org, 2010. Cluster Munitions and Mine Action. Available at: http://www.mineaction.org/overview.asp?o=1417&status_flag=L&rand=0.5413935 [Accessed September 2, 2010].

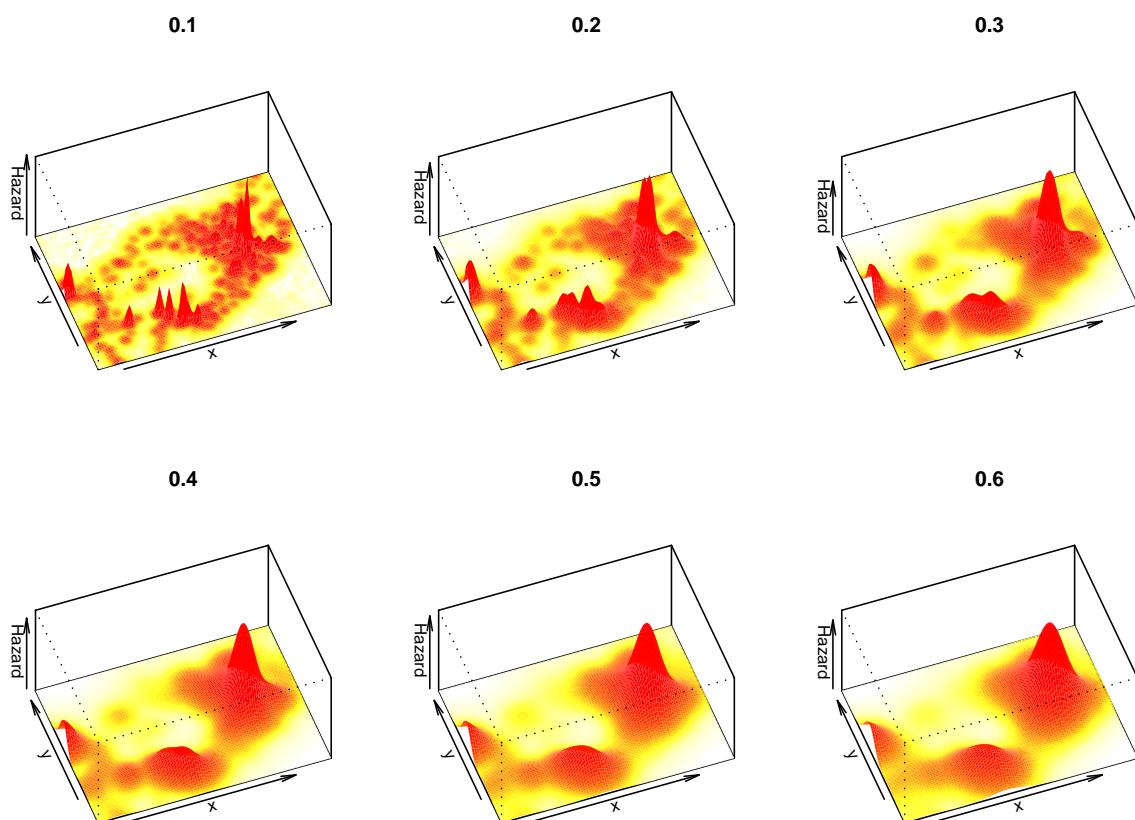
Xie, Z. & Yan, J., 2008. Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, 32(5), 396-406.

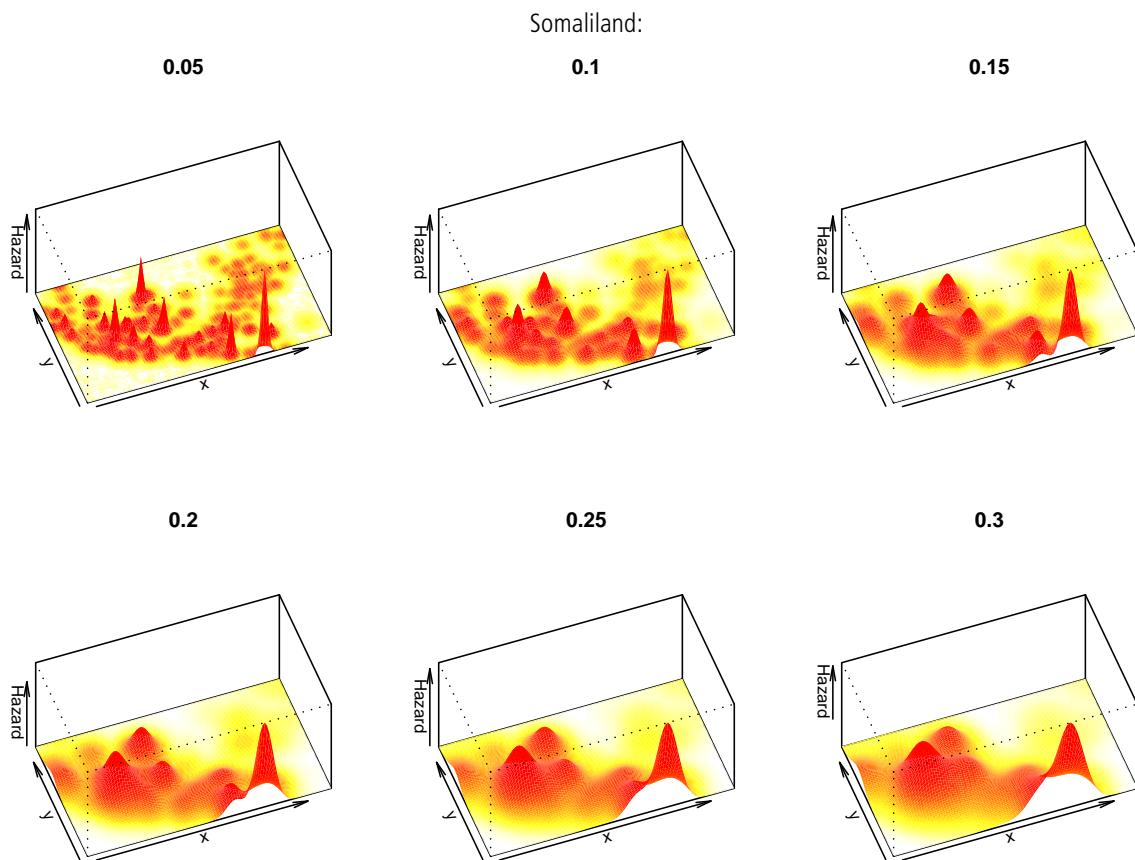
9 Appendix

Appendix A 3D-Plots

KDE with an increasing kernel size for Afghanistan and Somaliland in a three-dimensional plot

Afghanistan:





Appendix B The Layers

In the first phase of the SERWIS project the GICHD is planning to generate four layers.

Layer 1 – Hazards

As described before, the goal of this layer is to make the level of landmine and UXO hazard visible. In other words, the possibility of a threat for every cell will be shown. Only data taken from IMSMA about minefields, battle areas, etc. will be accounted for. This layer can be used later as a base for other analyses.

Layer 2 – Field activity³³

In this layer the field activities conducted the last year will be shown. This information is derived as well from IMSMA. I believe that in this case to carry out a KDE is very doubtful. Nevertheless, a KDE can make visible in which country mine reduction has been done the last year. Also it contributes to an obfuscation that can also be important in this layer.

³³ Field activity includes all efforts that are made to reduce hazards of all ERW related objects.

Layer 3 – Population

To calculate this layer many other datasets are being combined with ERW data. Here, social aspects are included within the representation. Population density, relevant places, points of interest, traffic routes, development areas, etc. might be integrated in order to make influences for the population visible. The contents of this layer not only stress hazards in highly populated areas, as a result decision makers may see where mine action resources may be applied where they make to strongest effect.

Layer 4 – Operational Difficulty

This Layer also is based upon the maps derived from the first layer. But additionally spatial data about operational difficulties is added. The goal of the layer is to find out how the hazards could be cleared most efficiently or where it is least difficult to clear minefields. For this layer land cover, slope, accessibility, climate data, etc. will have to be take into account.

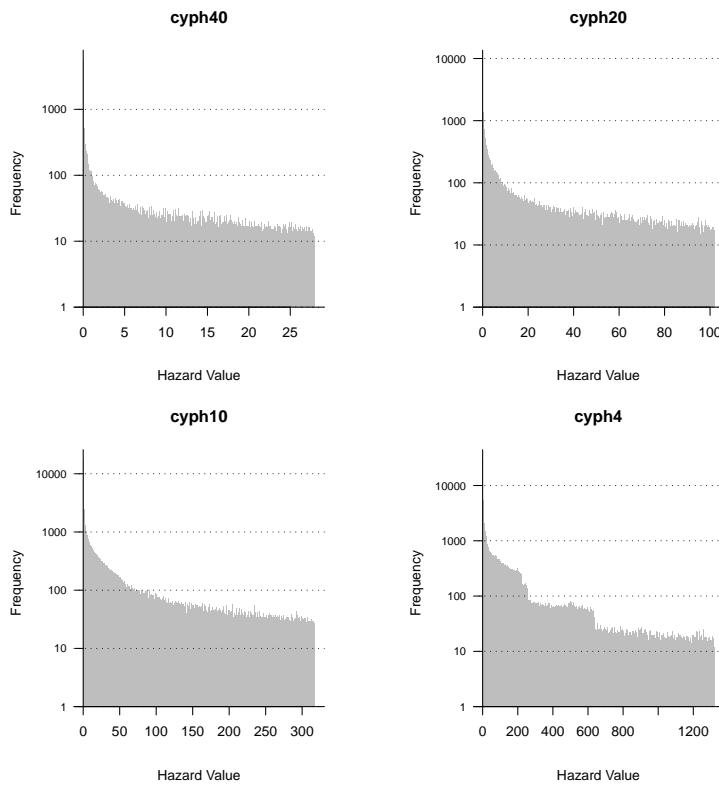
Since there are basically three main ways of demining (with machines, animals, or humans), for any of this kind it can be calculated, where to apply best. Considering many different factors, it can be estimated where hazards can be reduced easier than in other areas.

The goal of the layer is to show if an area is suitable for mechanical demining or animal detection. For this layer land cover, slope, accessibility, climate data, etc. will have to be take into account.

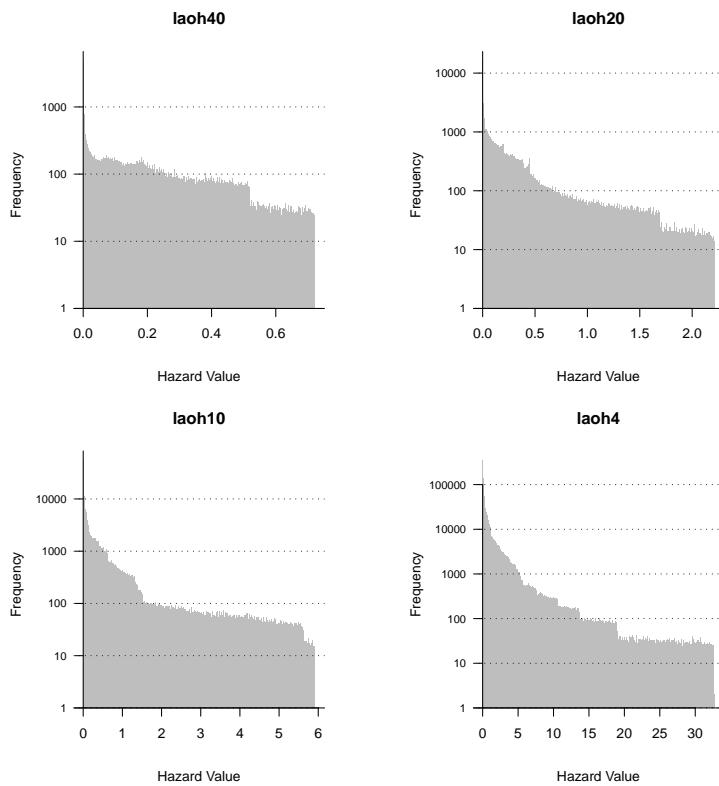
In SERWIS project the focus lies generally on layer 1 and 2, since the methods and technical solutions are going to be used for the others as the base. Hence, to create the first layer it is very important to find the right values and parameters, because important other parts rely on this.

Appendix C Histograms

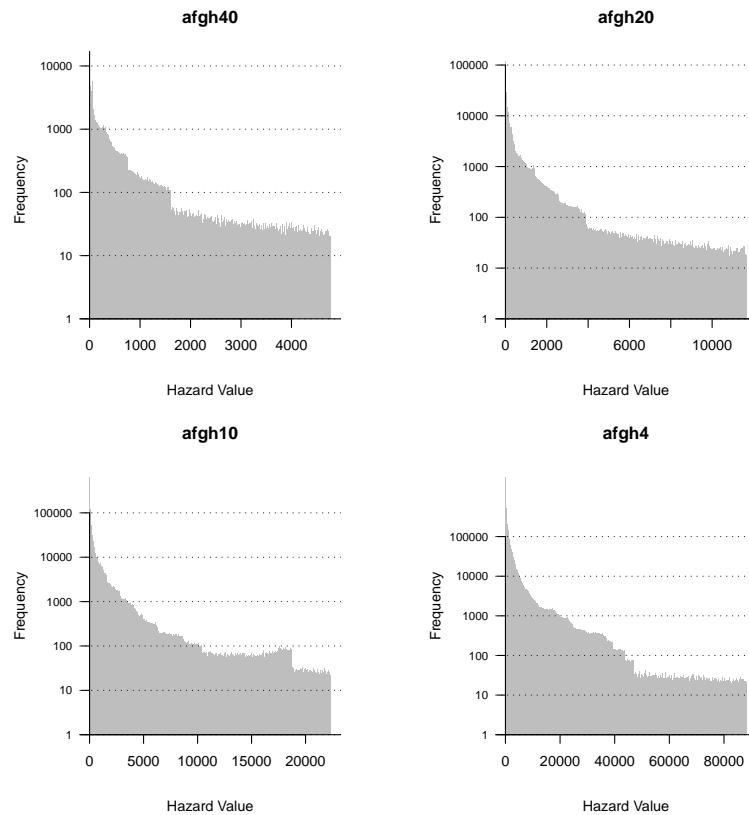
Logarithmic histograms of hazards in Cyprus:



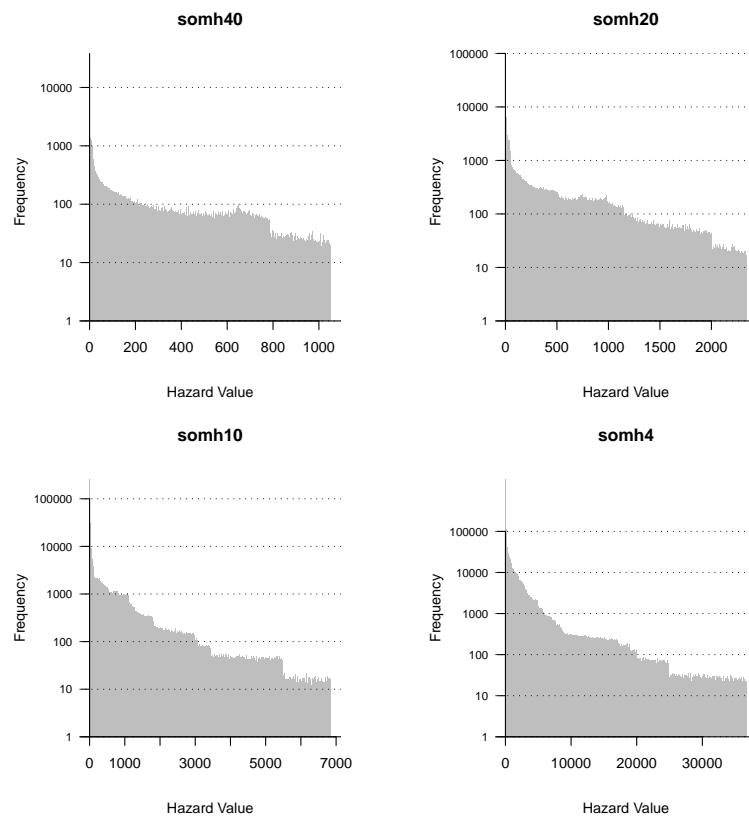
Logarithmic histograms of hazards in Laos:



Logarithmic histograms of hazards in Afghanistan:

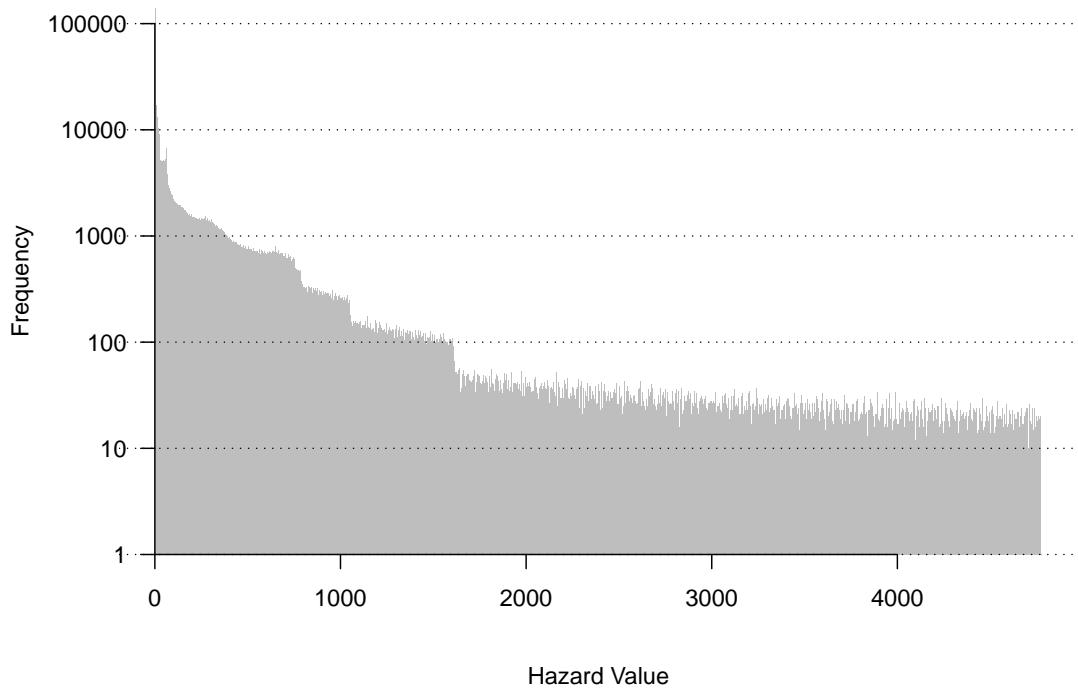


Logarithmic histograms of hazards in Somaliland:

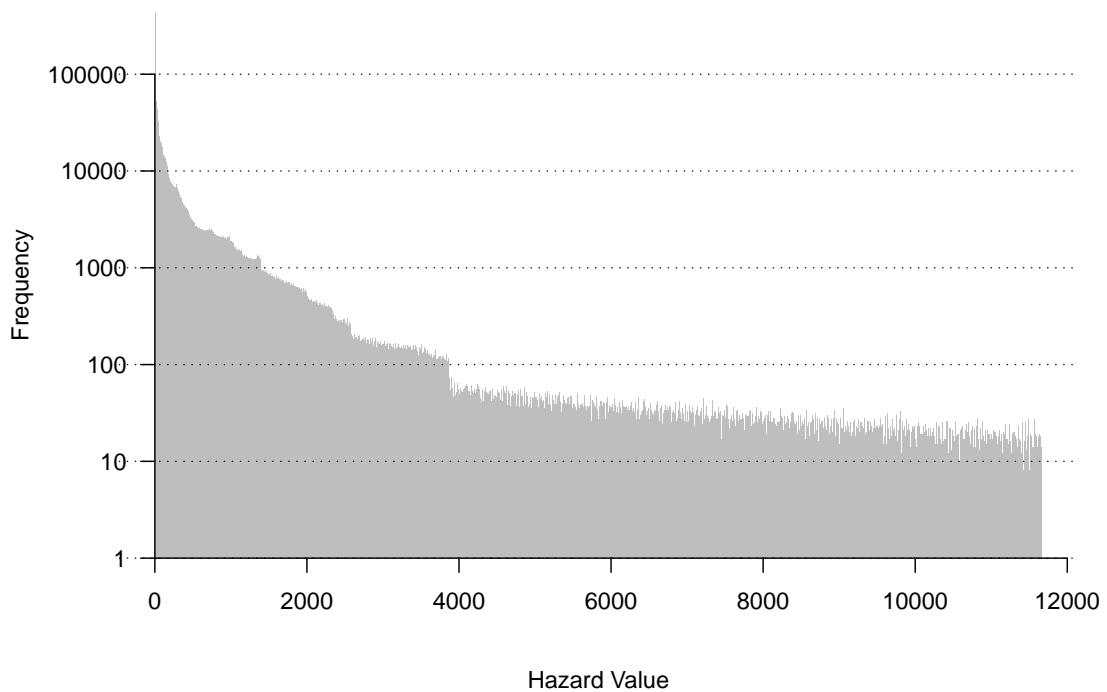


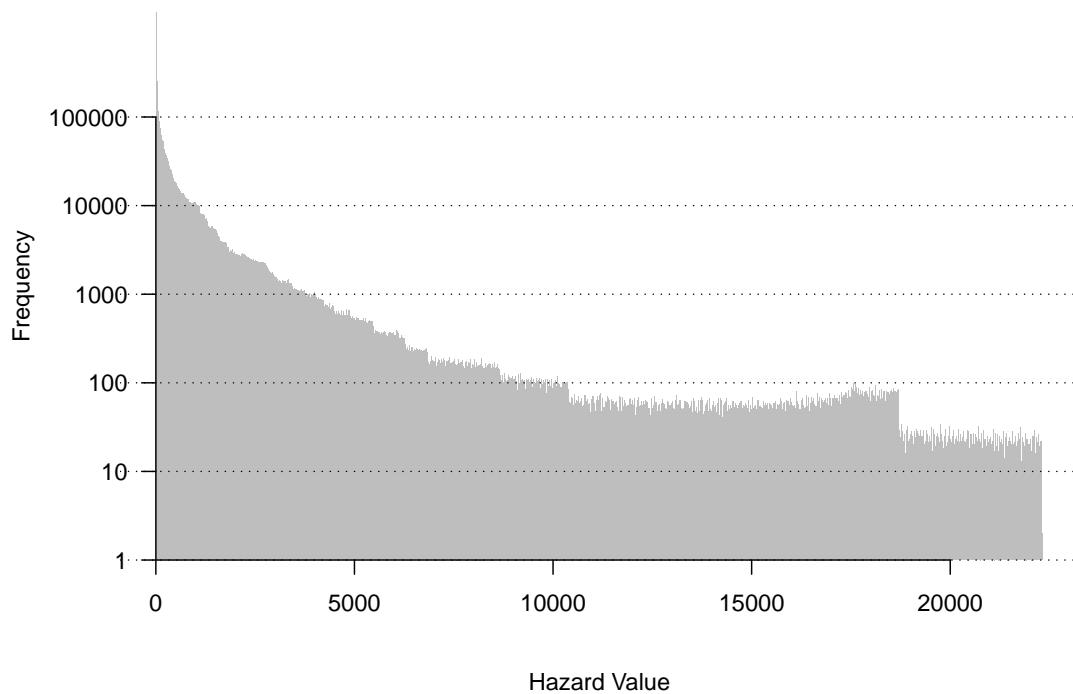
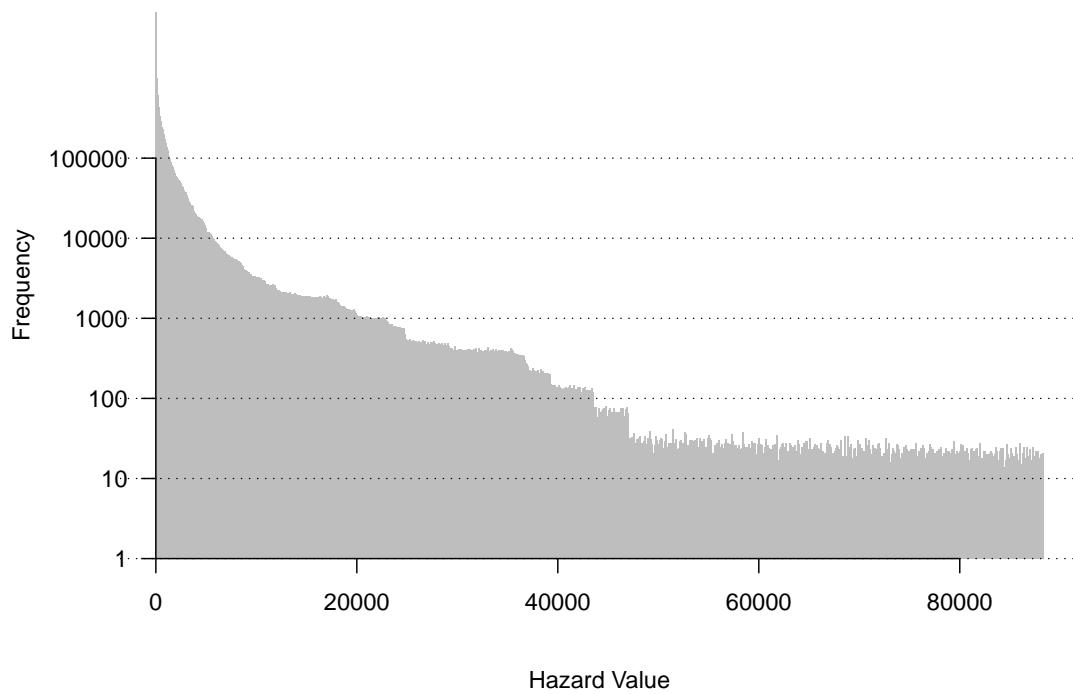
Logarithmic histograms of hazards of the four countries combined

All Countries 200 km Kernel size



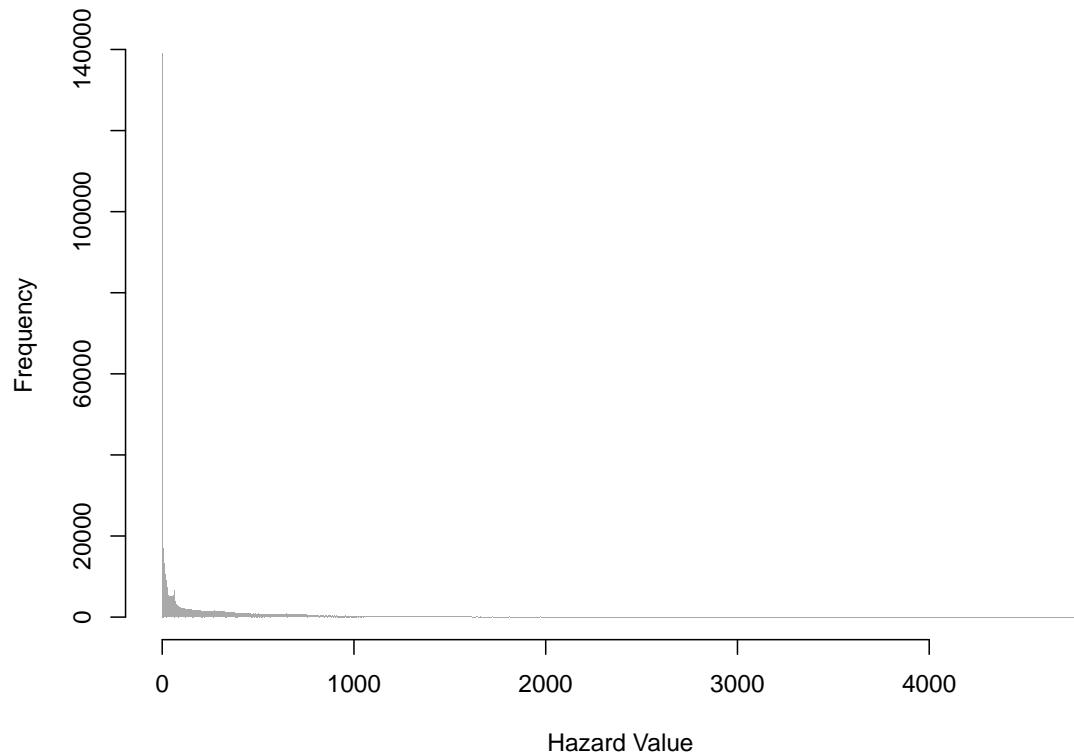
All Countries 100 km Kernel size



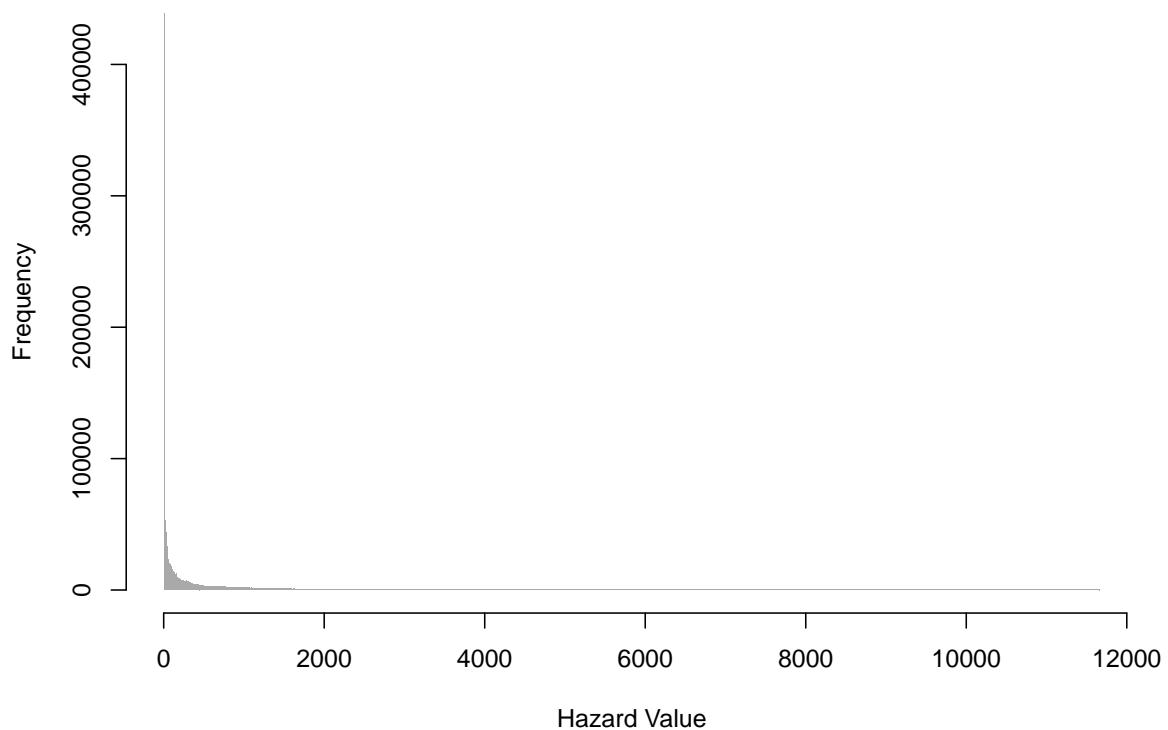
All Countries 50 km Kernel size**All Countries 20 km Kernel size**

Histograms of hazards of the four countries combined:

All Countries 200 km Kernel size

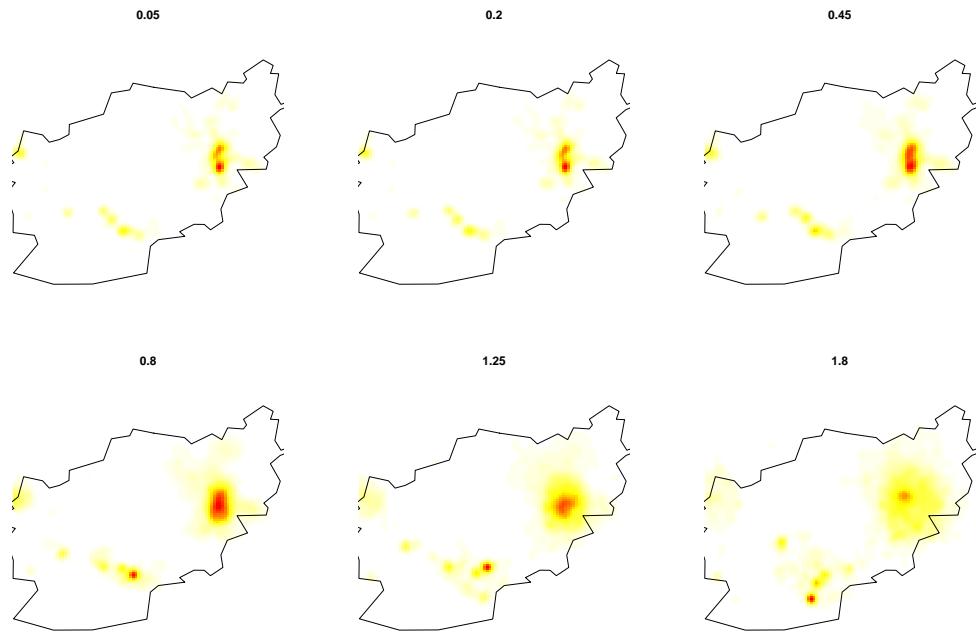


All Countries 100 km Kernel size

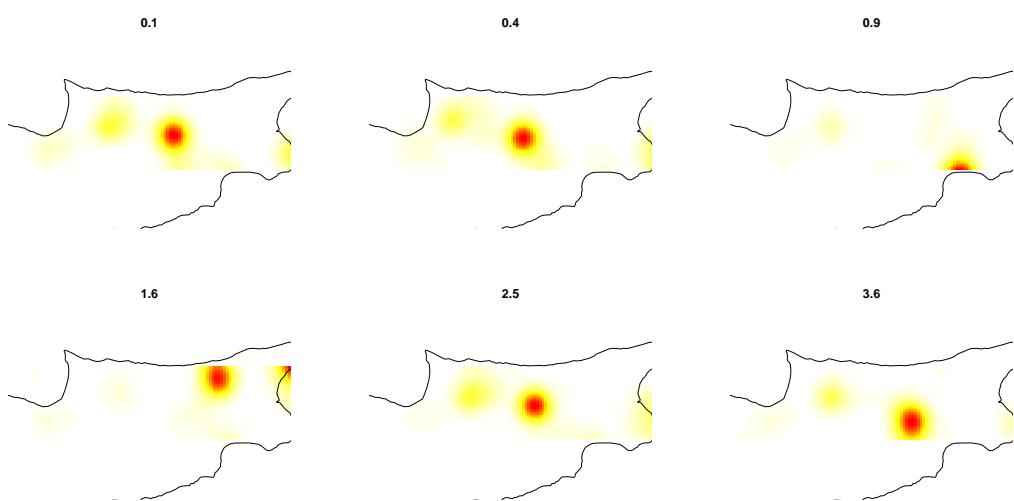


Appendix D Perturbation

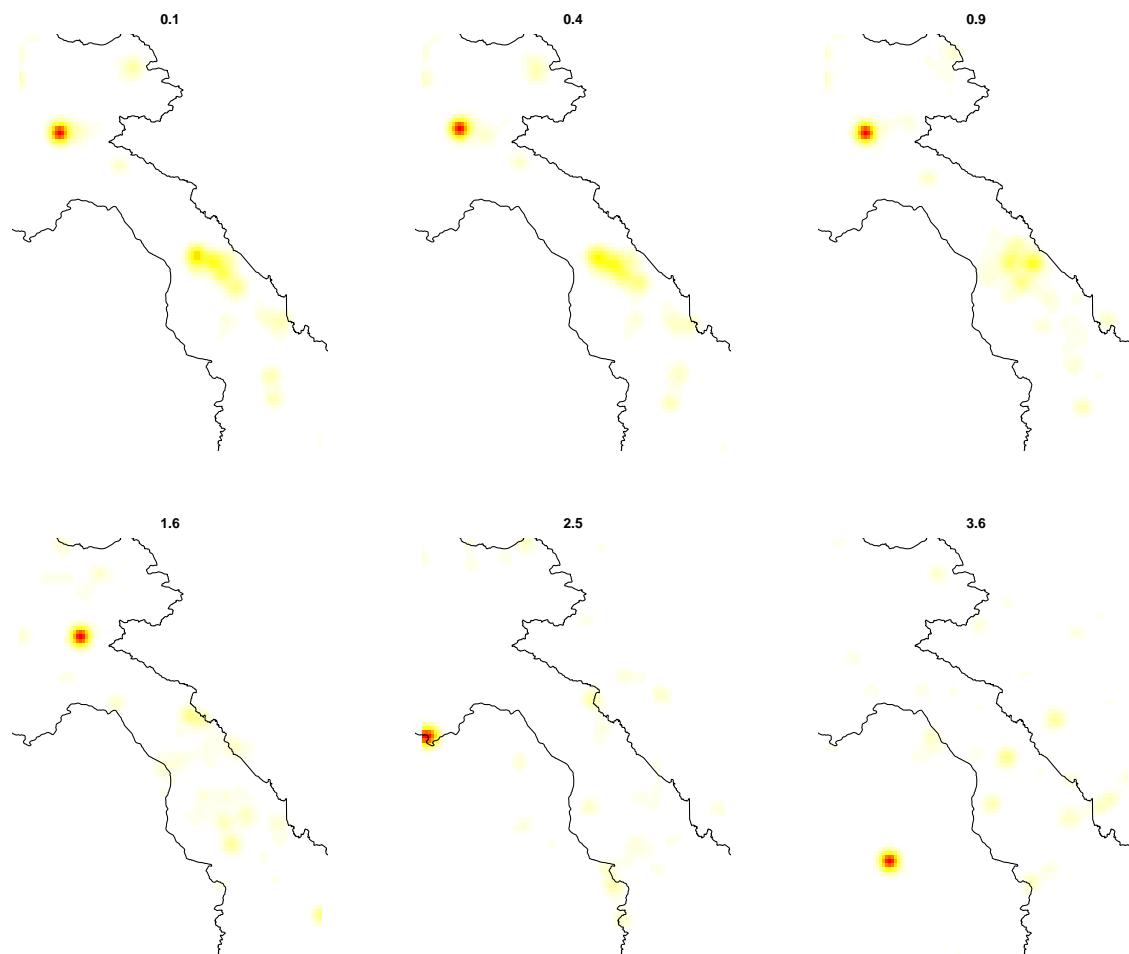
Perturbation with an exponentially growing ε in Afghanistan (ε is in decimal degree):



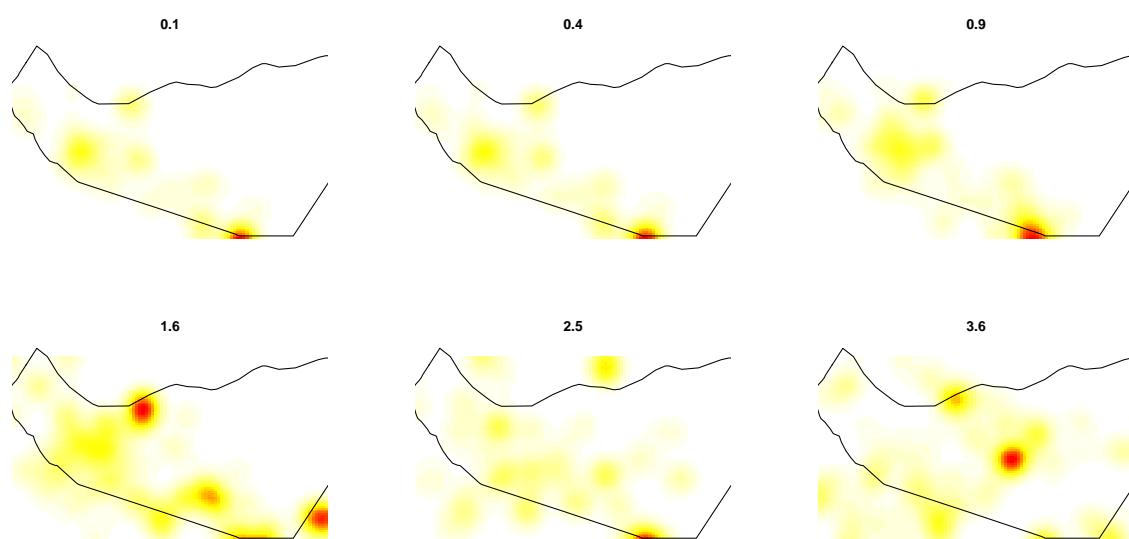
Perturbation with an exponentially growing ε in Cyprus (ε is in decimal degree)::



Perturbation with an exponentially growing ε in Laos (ε is in decimal degree)::



Perturbation with an exponentially growing ε in Somalia (ε is in decimal degree)::



Appendix E Kernels

Biweight:

$$\begin{cases} \frac{15}{16}(1 - t^2)^2 & \text{for } |t| < \sqrt{5} \\ 0 & \text{otherwise} \end{cases}$$

Triangular:

$$\begin{cases} 1 - |t| & \text{for } |t| < 1 \\ 0 & \text{otherwise} \end{cases}$$

Gaussian:

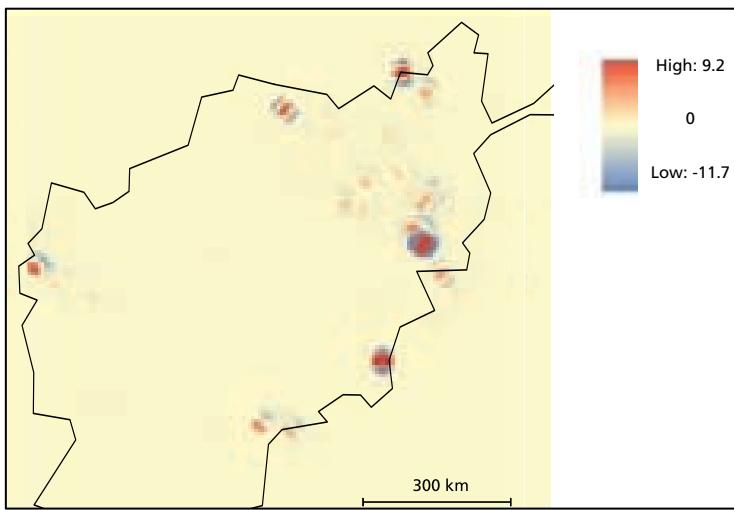
$$\frac{1}{\sqrt{2\pi}} e^{-(\frac{1}{2})t^2}$$

Rectangular:

$$\frac{1}{2} \text{ for } |t|, 0 \text{ otherwise}$$

Appendix F Overlapping Polygons

In the following figure, the differences between the calculations with respecting or ignoring the issue of overlapping polygons are shown, by subtracting their values from each other. Certainly there are differences, but they variation is tiny.



Appendix G Code samples

Reading in shapefiles and calculating their kernel density estimations (in R).

```
library(rgdal)
```

```

library(spatstat)
library(shapefiles)
library(maptools)
ow = as(unuu, "owin")
jj = read.shapefile(filename)
names(xx) = c("GUID", "Xcoord", "Ycoord", "PondSize", "UTM", "header")
hist(xx$PondSize, breaks=100, col="gray")
rgb.palette <- colorRampPalette(c("yellow", "red"), space = "rgb")
jp = ppp(xx$Xcoord, xx$Ycoord, c(min(xx$Xcoord), max(xx$Xcoord)),
c(min(xx$Ycoord), max(xx$Ycoord)))
sigma = 1
par(mar=c(2, 2, 2, 2)+0.1, mfrow=c(4, 2))
for (i in 1:8)
{
  sigma = i/50
  jd = density(jp, sigma, weights=xx$PondSize, edge=T)
  jd$v = jd$v / sum(xx$PondSize)
  image(jd, main=sigma, col = rgb.palette(30))
  plot(ow, add = TRUE, cex = 0.5)
}
dev.off()

```

Code fragment to generate 6 3D-Plots for KDE with a growing kernel size:

```

for(i in 1:6)
{
  sigma = i/50
  jd = (density(jp, sigma, weights = xx$PondSize))
  persp(jd, colmap = rgb30, theta = -20, phi = 50,
  zlab="Hazard", main = sigma)
}

```

For a jitter-effect (in R):

```

for (i in 1:21)
{
  jr = rjitter(jp, 0.02)
  plot(jr)
  plot(jp)
}

```

```
}
```

Function for generating histograms (in R):

```
hist <- function(title,bins)
{
  hhh=readIN(title)
  x=hhh$z

  x = x[!is.na(x)]
  x=x[x>1]
  d <- density(sqrt(x), bw= 100)
  plot(d)

  lot(histogram(log(x), xlab = "Hazard", type = "density",col =
  "black", nint = bins, panel = function(x, ...))
  panel.histogram(x, ...)
}

)
}

plot(hist("somh40", 1000))
```

Code fragment to perform a Kolmogorov-Smirnov-Test (in R):

```
xcoord <- function(x,y) { x }
kstest(jp, xcoord)
model <- ppm(jp, ~x + y)
kstest(model, xcoord)
```

Java method to calculate the geometric intervals:

```
public class Classifier {

  public static void main(String[] args) {
    double FIRST = 10000;
    double c = 1.5;
    int decr = 3;
    double max = 222083;
    double interval;
    System.out.print(FIRST + ",");
    interval = FIRST;
```

```
double klasse = FIRST;
int counter = 1;
for (int i = decr; i > 0; i--) {
    counter++;

    klasse = klasse + interval * (1/c);
    System.out.print(klasse + ", ");
    interval = interval *(1/c);

}

while (klasse < (max)) {
    counter++;
    interval = interval * c;
    klasse += interval;
    System.out.print(klasse + ", ");
}

System.out.println(counter);
}
```

Personal declaration: I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Jonas Herzog