3 Risk assessment at global scale - the case of earthquakes

In section 2 the theoretical background has been set up for a quantitative risk assessment, considering people at risk from a certain natural hazard. In this following section an approach will be introduced to carry out such a risk assessment worldwide at sub-national scale, specifically for the natural hazard of earthquakes. The underlying datasets do not allow the production of absolute risk values. The result of this work can only identify areas of higher risk than others considering earthquakes.

Problems and limitations that occurred when developing and applying the methodology will be explained according to the three main parameters of the risk equation as introduced in section 2: the hazard, the exposure and the vulnerability. The former two features, the hazard and the exposure part, are based on pre-existing data. In order to understand the limitations of these datasets some basic knowledge about earthquakes and population number measurements, as well as about the methods applied for the generation of the datasets discussed here, is needed. This basic information is provided in chapter 3.2 and chapter 3.3 respectively. The underlying theoretical background for the last parameter, vulnerability, has been elaborated in detail in section 2. Due to the lack of worldwide data regarding the level of people's vulnerability, a composite indicator for the estimation of vulnerability at national scale has been developed and is described in chapter 3.4.

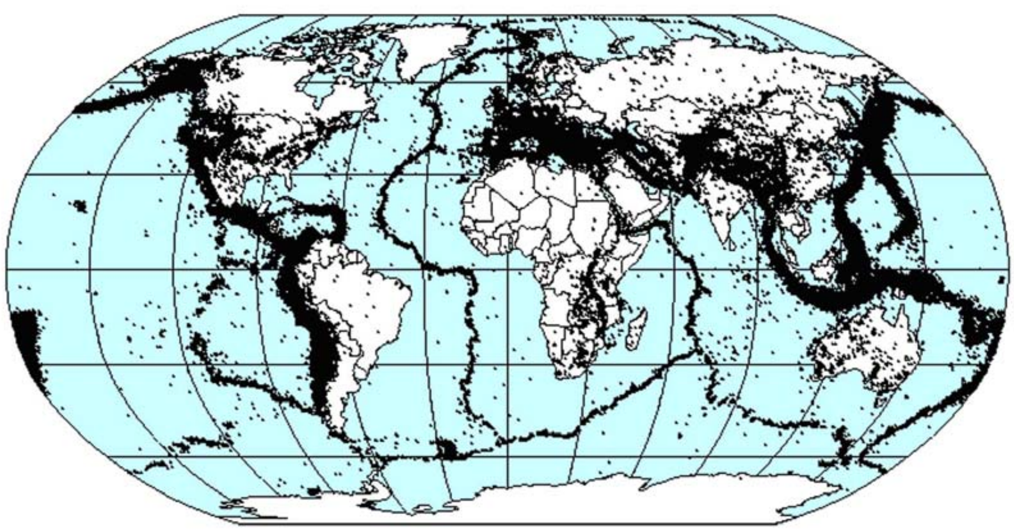

Figure 12: Global distribution of Epicentres during the years 1963-1998 (source: LOWMAN et al. 1999).

3.1 Introduction - earthquake risk

Earthquakes are one of the most devastating natural hazards, causing in average around between 14,000 to 16,000 fatalities per year (COBURN & SPENCE 1992). This is due to the fact that (1) they are the least predictable of all natural disasters and (2) the span of time between the earthquake threat and its impact is the shortest among the major disasters (GUHA-SAPI et al. 2004). Ninety percent of the world's earthquakes occur on, or near, the edges of tectonic plates but they can also occur far from the edges along geological faults, (SHEDLOCK et al. 2000) (see Figure 12).

Usually the destructive impact of an earthquake is limited in its extent, since the energy released by an earthquake diminishes significantly with increasing distance from the epicentre. However, through indirect effects, earthquakes can also have a devastating impact on large areas or areas further away from the epicentre, as the tsunami in SE Asia dramatically proved in December 2004.

The direct effect felt from an earthquake on the earth’s surface is correlated to the seismological characteristics of the event (depth, magnitude etc.) and the geological characteristics of the region (underlying rocks, soil texture etc.). Beside these parameters, the impact on the population is strongly related to the physical parameters of infrastructure (for example the level of earthquake safety applied to the construction of buildings or bridges). It is also dependent on the preparedness of individuals and society, for example, the level of individual knowledge about required precautionary measures to be taken in the case of earthquakes, or the quality of equipment and rapid response of rescue teams.

3.2 Earthquake risk: the hazard component

When an earthquake occurs, it releases energy in the form of waves that radiate in all directions from the earthquake’s source. These seismic waves travel through the Earth at speeds of several kilometres per second causing ground motions. There are different types of energy waves that shake the ground in different ways and travel through the earth at different velocities (see Figure 13). The variety of waves can cause hazardous impacts on people, infrastructure or buildings in a direct or indirect way such as surface faulting, tremors vibration, liquefaction, landslides, aftershocks and/or tsunamis.

There are different parameters used to describe the severity of an earthquake. The magnitude of an earthquake is a purely seismological characteristic calculated from a measurement of either the amplitude or the duration of specific types of recorded seismic waves. The magnitude is only determined from measurements made from seismograms and does not refer to the level of shaking or measures of building damage. More crucial for the potential damaging impact of an earthquake at a particular location is the ground motion or ground shaking and the earthquake’s intensity. The extent of ground motion is a major parameter for the degree of damage or collapsing of buildings and infrastructure, the major cause for earthquake fatalities. The level of ground motion in a specific area is strongly related to the magnitude of an earthquake and the distance of the location to its epicentre, but it is also influenced by the seismic and geologic setting of the site. This includes for example the type of faulting, the depth, kinds and thicknesses of geologic materials exposed at the surface and the subsurface geologic structure (RASMUSSEN et al. 1974; NEWMARK & HALL 1982). The intensity of an earthquake describes the effect that an earthquake has at a particular site on the natural features, the industrial installations and/or the human beings. It should be noted that an earthquake is assigned one magnitude, but it may give rise to reports of ground motion and intensity at many different levels depending on local parameters.

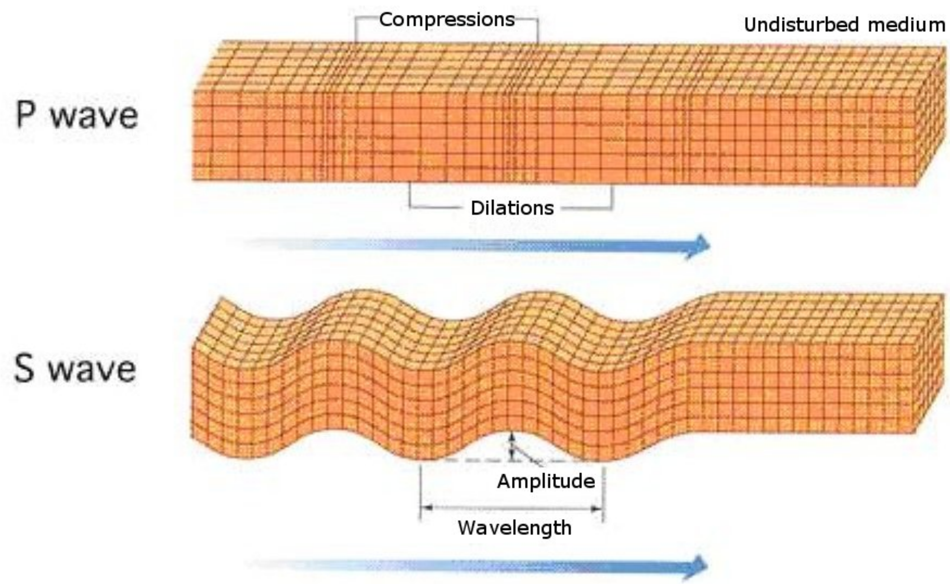

Figure 13: Different types of seismic waves (source: after SKLPC 2006).

It is difficult to predict the extent of damage to infrastructural objects such as lifelines or buildings caused by an earthquake. The resilience of such objects to an earthquake depends on the interaction between structural elements of the object and the direction, frequency, and duration of ground motion. In particular, the frequency of ground motion is an important factor in determining which structures are likely to be affected (NOSON et al. 1988). The frequency of shaking varies with distance. Close to the epicentre, both high (rapid) and low (slow)-frequency motions are present. Further away, low-frequency motions are dominant, a natural consequence of wave attenuation in rock. Tall buildings, bridges, and other large structures are less resistant to low-frequency ground shaking, and small structures are more affected by high-frequency shaking. Additionally individual characteristics such as the shape of the building, the building material, the means of construction, the ties between the foundation and the structure, history of the building and history of previous earthquakes play an important role.

For comparing and ranking earthquake events, measurements of severity are required. The earthquake's magnitude is most often specified as value on the Richter Scale, although nowadays seismologists use a number of different magnitude scales, such as surface wave magnitude, body wave magnitude and moment magnitude, which are extensions of the original Richter Scale. The quantification of ground motion is usually based on measurements of the physical parameters peak ground velocity (pgv), peak ground acceleration (pga) or peak ground displacement. Of these, pga is the parameter most often associated with severity of ground motion, without considering the impulse and frequency of the waves triggered by an the earthquake (ANDERSON & BERTERO 1987). The intensity assessment is based on observations of an earthquake's effects and impact. It is most commonly expressed by using the 'Modified Mercalli Intensity' (MMI) scale, which has also been developed by RICHTER (1958) building upon a work of Mercalli in 1902 (see Annex 1). MMI lower numbers describe the manner in which the earthquake is felt by people. The higher numbers on the scale are based on observed structural damage. A comparable classification of to what level an earthquakes affects a specific place is provided by the European Macroseismic Scale (EMS) which was last updated in 1998 (GRÜNTHAL 1998).

This short introduction regarding the size measurement and damage assessment of earthquakes hints at the complexity of the topic, which accounts for the difficulty in generating a global earthquake hazard dataset. For this study it was deliberated whether to integrate previously observed earthquake events as the hazard part of the risk equation. However, the time period of less than 30 years, for which data is available at a suitable level of accuracy, was evaluated as too short and the region that could have been covered by such a dataset, too small. It was therefore decided to base the study on the final global seismic risk map of GSHAP.

GSHAP was launched in 1992 by the International Lithosphere Program (ILP) with the support of the International Council for Science (ICSU), and was endorsed as a demonstration programme in the framework of the United Nations International Decade for Natural Disaster Reduction. The GSHAP project terminated in 1999. The global seismic hazard map produced within the scope of the project depicts the levels of chosen ground motions that are likely to or will not be exceeded in specified exposure times. The regional hazard assessment programmes commonly specify a 10% chance of exceeding (and 90% chance of not exceeding) pga values for an exposure time of 50 years, corresponding to a return period of 475 years (SHEDLOCK et al. 2000). The expression of pga was introduced earlier as ground motion parameter. Pga is a short period parameter that is proportional to force and it is very commonly mapped because current building codes containing reference to seismic events, specify the horizontal force a building should be able to withstand during an earthquake (GIARDINI 1999). The global GSHAP dataset is a compilation of 7 regional maps. Their production was led by selected regional centres with specific expert knowledge and applying multinational test areas under the coordination of large working groups. The final GSHAP map in the resolution of 6' cell size is shown in Figure 14.

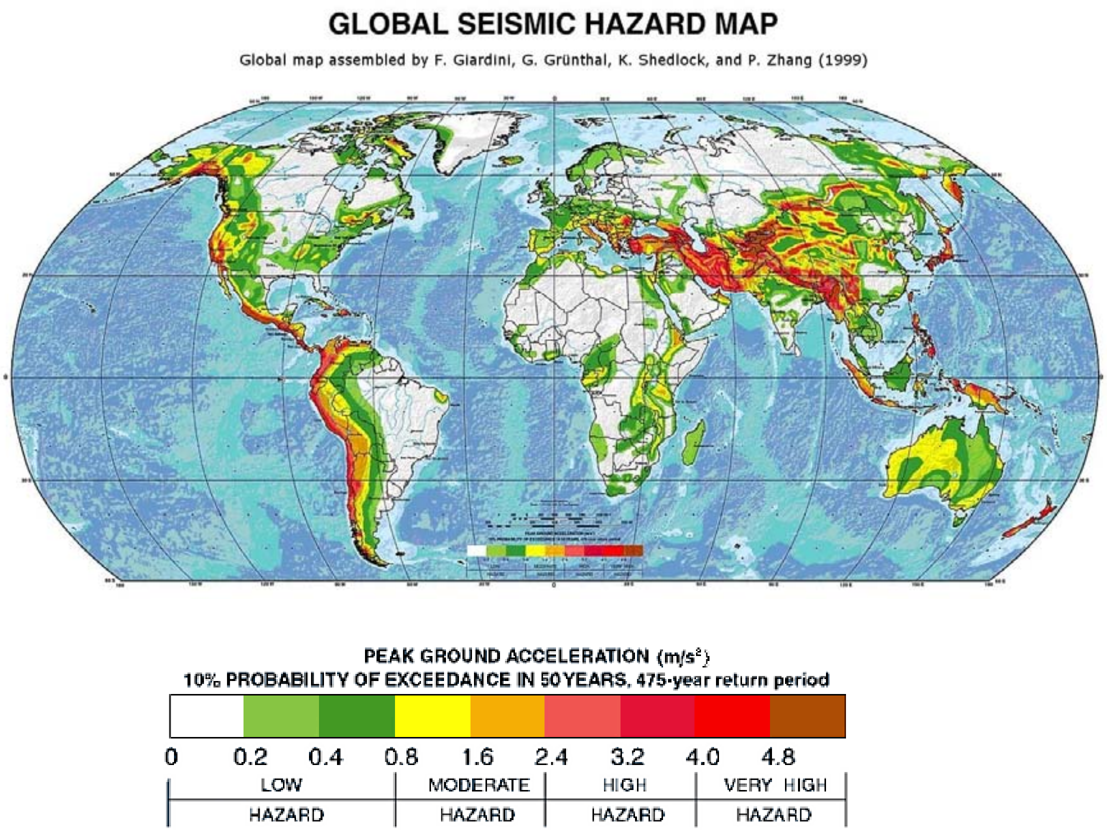

Figure 14: The GSHAP global seismic hazard map (source: SHEDLOCK et al. 2000). The indicated peak accelerations have a 10% probability of being exceeded in 50 years (average recurrence period of 475 years). For example, in the areas mapped in yellow there is a 10 % probability that PGA values between 80-160 cm/s^2 are exceeded within 50 years.

The GSHAP dataset is by far the best estimate of global seismic risk available today. Nevertheless, there are three shortcomings, two of general and one of study-specific type:

(1) The fact that GSHAP results are based on the site classification 'rock' everywhere (except Canada and the United States). Hence any local geological specification which might have a significant influence on the seismic risk is neglected.

(2) The pga value measures short-period ground motions, which are crucial for the estimation of impact on rather small scale buildings (e.g. one-to-two storey buildings). These buildings are the largest class of structures in the world. However, this is not the best indicator for tall buildings with 10+ floors which are more vulnerable to damage from longer-period ground motions.

(3) The GSHAP map shows direct ground motion probabilities due to earthquakes. It does not reflect any indirect impact with high potential risk of harming people, such as soil liquefaction or tsunamis.

3.3 Earthquake risk: the exposure component

The exposure component of the risk equation covers exclusively the amount of people potentially affected by a hazardous event. Ideally, it is the precise number of people who reside within an area affected by a hazardous event at the time when the event strikes or takes place, taking into account hourly, daily and seasonal fluctuations in population density. In reality population density data is only available with limited quality and resolution for large parts of the world. This is even more so for developing countries. Census data is often lacking a frequent update, or is not available to researchers or disaster managers. In nearly all cases population data resulting from national census exists as attribute information allocated to an administrative unit. This data format needs to be changed to be useful for computer supported spatial analysis such as the estimation of risk of disasters described here. Various approaches have been pursued for the process of transforming data from vector to raster format and the generation of population distribution datasets in general is discussed in chapter 4.2. The focus of this chapter is the creation of the Lanscan Global Population Database, produced by the Oak Ridge National Laboratories (ORNL) (DOBSON et al. 2000) for which a critical review is provided.

The number of global population datasets available in gridded format is limited. For this study it was decided to integrate the Landscan dataset (DOBSON et al. 2000). The researchers involved in the Landscan production claim to base their dataset on the “best available census counts (usually at sub-province level) for each country and four primary geospatial input datasets, namely land cover, roads, slope and night time lights” (BHADURI et al. 2002, p 1). The population data was developed by allocating census counts at 30" X 30" cells through a "smart" interpolation based on the relative likelihood of population occurrence. This likelihood is a function of the above mentioned input layers. The Landscan methodology takes into account that the correlation between population numbers and features of the input data layers varies throughout the world by applying different models for different regions of the globe. Unfortunately the Landscan dataset has no qualitative measures attached to it and the model used in disaggregating the data is not documented (SCHNEIDERBAUER & EHRLICH 2005; MIRELLA et al. 2005). Potential errors within Landscan include the integration of the night-time lights dataset. These data tend to underestimate the population density of urban centres and to overestimate the population density of suburban areas (SUTTON et al. 1997; ELVIDGE et al. 2004). Exclusion areas are calculated by digital slope and by the global land cover map of the world; two datasets, for which no accuracy measures are attached. Notwithstanding these deficiencies the Lanscan dataset offers a useful overview of the distribution of population at global scale and is unique regarding its fine spatial resolution.

Landscan is used for this study as the exposure data layer for the risk assessment. The dataset used is the version ‘Landscan 2002’ (ORNL 2002). Considering the time required developing the underlying input datasets and to model and compute the population dataset itself it is realistic to assume that the dataset is representing population numbers not more recent than the year 2000. The Lanscan data is visualised as global map in Figure 15.

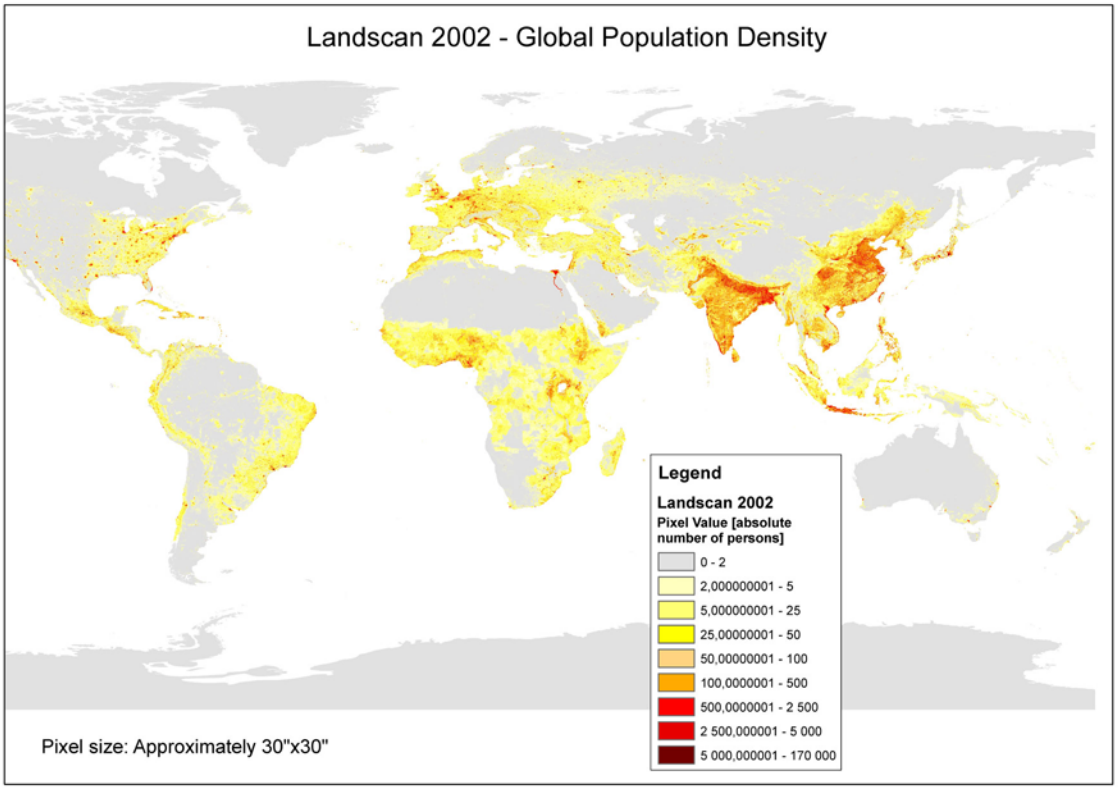

Figure 15: The Landscan Dataset, a worldwide population database compiled on a 30" X 30" latitude/longitude grid (map: author, data source: ORNL 2002).

3.4 Earthquake risk: the vulnerability component

The determination of people’s vulnerability is not straightforward for several reasons already elaborated in chapter 2.3. Hence, a global vulnerability assessment is at best only possible through the use of general proxies, and will always result in relative rather than absolute estimations (Dilley et al. 2005). Nevertheless, a reasonable relative estimation of populations’ vulnerability provides significant support for decision making and to the management of disaster preparedness actions; even if it’s spatial resolution is limited to national scale.

The few studies that have been carried out with the aim to assess vulnerability to natural disasters worldwide base their approaches on data about casualties and / or fatalities of previous hazardous natural events (PEDUZZI et al. 2002, DILLEY et al. 2005). The only data base providing these data worldwide for a significant number of years is the Emergency Events Database (EM-DAT) managed by the Centre for Research on the Epidemiology of Disasters (CRED) which belongs to the Université Catholique de Louvain but is located in Brussels. The EM DAT data base has an invaluable scientific worth. Nevertheless it suffers from a number of constraints and disadvantages most of which are identified by the responsible scientists and some of the problems could be solved during the recent years (GUHA-SAPIR & BELOW 2002, GUHA-SAPIR et al. 2004). To name but a few: The data entering the database are collected from a variety of sources and lack standards. Usually, the data was primarily not recorded for statistical purposes and the data may have been recorded with a specific political intention in mind. The geographical location of disasters is, if at all, often not reported accurately and the extent of the affected area is difficult to determine. In addition, disaster data are often not reported for developing countries and for the majority of years of the dataset the data was entered without distinguishing between the value 0 and 'no data'. The latter two points led to the decision not to use the EM-DAT database for this study but to develop a new layer of an assessed vulnerability of people to natural disaster at national scale.

In order to include the wide range of aspects that contribute to people's vulnerability regarding earthquakes, the development of a composite indicator was pursued. This composite indicator was developed with a focus on the hazard independent part of the vulnerability as stated in 2.4. Information that would have allowed generating a vulnerability assessment specifically for earthquakes would have required global data on construction type and quality of buildings and infrastructure objects, which is not available at moment. Hence, the here created composite indicator is not only applicable to earthquakes but to all types of hazardous events that may have an impact on the population.

In general term a composite indicator summarises a number of underlying sub-indicators or variables (FREUDENBERG 2003). Composite indicators that compare country performance allow simple comparisons of complex and sometimes elusive realities in wide ranging fields (NARDO et al. 2005b). Ideally composite indicators extract and condense information from their sub-indicators and ease interpretation of multi-dimensional issues, in particular for the non-scientific community (decision makers, the public). On the other hand they simplify, may send misleading, non-robust policy messages and are somehow subjective, since their development involves judgments and decisions taken by their creator(s). Advantages and disadvantages of composite indicators are discussed in detail in WALL et al. (1995), SAISANA & TARANTOLA (2002), SAISANA et al. (2005), and NARDO et al. (2005a). The dispute about scientific value of composite indicators is ongoing, though they are increasingly recognised as a useful tool in policy analysis and public communication (NARDO et al. 2005a). Classically, composite indicators of country performance worldwide were developed to measure economic parameters. In the last years they have increasingly been applied in the field of environment, globalisation, society and innovation / technology. See FREUDENBERG (2003) for a list of composite indicators for country performance recently in use.

The development of a composite indicator is one approach to formalise mathematically the rules that guide a complex natural system by establishing a theoretical framework. This model is significantly influenced by the choices made by the scientist on how to observe reality. "When building a model to describe a real-world phenomenon, formal coherence is a necessary property, yet not sufficient. The model in fact should fit objectives and intentions of the user, i.e. it must be the most appropriate tool for expressing the set of objectives that motivated the whole exercise. The choice of which sub-indicators to use, how those are divided into classes, whether a normalization method has to be used (and which one), the choice of the weighting method, and how information is aggregated, all these features stem from a certain perspective on the issue to be modelled" (NARDO et al. 2005a, p 7). Hence, subjectivity is an unavoidable part of composite indicators. FREUDENBERG (2003) claims that all composite indicators at a minimum "should be as transparent as possible and provide detailed information on methodology and data sources. They should always be accompanied by explanations of their components, construction, weaknesses and interpretation” (FREUDENBERG 2003, p 6). In the following a description of the main steps for the creation of the here introduced composite indicator is given.

3.4.1 Purpose of the composite indicator

The overall objective of the vulnerability composite indicator is a relative assessment of the hazard independent part of people’s vulnerability at national scale to natural hazardous events. More specifically, the composite indicator is aimed at reflecting the various identified levels of social scales and dimensional characteristics regarding the vulnerability of people to natural hazards in general, including vulnerability to earthquakes (see the description in chapter 2.3). The composite indicator is restricted to national scale, resulting in only one indicator value for each country. This is due to the fact that data for all countries worldwide, for a variety of factors of importance for populations’ vulnerability, is only available at country level. In fact, some selected indicators included represent characteristics of scale levels at finer ‘resolution’ - such as the individual or the household - but are aggregated to a national scale.

The creation of the composite indictors can be divided into 4 main steps described below:

- pre-selection of sub-indicators,
- data collection and preparation,
- statistical analysis (resulting in the reduction of considered indicators),
- aggregation of sub-indicators.

3.4.2 Pre-selection of sub-indicators

The World Bank database and the Human Development Reports (HDR) of the United Nations Development Program (UNDP) are the most important open sources for relevant datasets with the required spatial coverage and temporal frequency. The datasets of these institutions have been widely analysed by the scientific community, predominantly as single datasets or as a limited number of datasets for a specific topic. For example, the GDP and other monetary indicators for the assessment of economic performance, or indicators regarding malnutrition and death rate for assessing the current health situation. During recent years visualisation tools were introduced allowing a quick comparison of a number of these indicator values for several countries of the world. These tools are mainly aiming at representing the course of development of nations, for example the ‘Google Gapminder’ [http://tools.google.com/gapminder] or the ‘Dashboard for Sustainability’ [http://esl.jrc.it/envind/dashbrds.htm] <both viewed December 2006>. A number of studies developed composite indicators based on structural sub-indicators recorded by the World Bank and the UN. For example, the ‘Global Needs Assessment’ (GNA), initiated by the Directorate General ‘ECHO’ of the European Commission (EUROPEAN COMMISSION 2006), or the Index of Severity for the assessment of a country’s risk potential regarding conflicts, developed by the Country Indicators for Foreign Policy (CIFP) (CARMENT 2001). However, there is a lack of research activities regarding the systematic collection and statistical analysis of structural indicators with worldwide coverage for the purpose of populations’ vulnerability determination.

The selection of sub-indicators for global vulnerability estimation requires a compromise between the availability of datasets and the information provided by them. Therefore, the choice of sub-indicators to build the base for this study was primarily done in order to optimise the coverage of scales and dimensions, which describe the characteristics of vulnerability. Secondly the choice of indicators took into account the frequency of update and the number of countries covered. Based on these criteria 37 indicators were pre-selected, which are summarised with their respective measurement unit and source in Table 5 and described more detailed in Annex 2. Most of the required datasets were available through the World Bank and the UN. Data regarding the level of corruption was taken from the Non Governmental Organisation Transparency International. Data relating to the amount of arable land was gathered from the FAO. Four other sub-indicators were based on existing data but had to be developed particularly for the purpose of this study. These are the percentage of HIV infection and AIDS cases, the situation of ethnic minorities within a country, missing indicator values and the number and severity of armed conflicts. The compilation of values for these indicators is described in more detail below.

HIV/AIDS values [No of sub-indicator: 5]

The values for this indicator are simply the percentage of HIV/AIDS affected people within a country related to the absolute number of population of that nation. The HIV/AIDS infection values were extracted from a large number of country-specific UNAIDS/WHO/UNICEF Epidemiological Fact Sheets, which include the most recent country-specific data on HIV/AIDS available. The population data is sourced from the UNPOP database (http://www.un.org/popin/ <viewed April 2005>).

The situation of ethnic minorities within the country [No of sub-indicator: 17]

The data source for this indicator is the Minorities At Risk Data Generation and Management Program (MARGene) of the University of Maryland (GURR et al. 1993; GURR 2000). The Minorities at Risk (MAR) Project is a university-based independent research project that monitors and analyses the status and conflicts of communal groups in all countries in the world with a current population of at least 500,000. A major work of the project is the collection and analysis of the Minorities at Risk database, a combination of qualitative and quantitative information concerning all communal groups which are seen to be ‘at risk’. Such a minority at risk refers to 'an ethnopolitical group (non-state communal group) that:

- Collectively suffers, or benefits from, systematic discriminatory treatment vis-à-vis other groups in a society; and/or
- Collectively mobilizes in defence or promotion of its self-defined interests”

(MAR 2004, p 5).

The MARGene dataset can freely be downloaded for research purposes together with a software tool running in a standard Microsoft Windows environment, providing an easy to use interface for searching and creating data subsets. It should be mentioned that the MARGene project is aimed at measuring the relative status and condition of the group, not its absolute condition. The level of discrimination is determined in relative terms compared to the situation of other groups within the country. For example, the denial of the right to vote is only relevant if some groups in the country do have the right to vote (MAR 2004).

No.	Indicator Name	Unit of measurement	Source (Year)
1	Agriculture, value added	% of GDP	WB (2002)
2	Employment in agriculture	% of total employment	WB (2002)
3	Agriculture productivity	constant 1995 US$	WB (2002)
4	Aid dependency	% of GNI	WB (2002)
5	HIV/Aids values	% of population	UNAIDS / WHO / UNICEF (2002)
6	Arable Land	% of Land Area	FAO (1999)
7	Corruption	Predefined index scale (0-10)	TI (2003)
8	Exports of goods and services	% of GDP	WB (2002)
9	GDI – Gender – related development index	Predefined index scale (0-1)	UNDP HDR (2003)
10	GDP Per capita PPP	Current international $	WB (2003)
11	HDI	Predefined index scale (0-1)	UNDP HDR 2003 (HDI values: 2001)
12	Imports of goods and services	% of GDP	WB (2002)
13	Improved water source	% of population with access	WB (2002)
14	Life expectancy at birth, total	Years	WB (2002)
15	Literacy rate adults - UNDP	% of illiterate adult population	UNDP HDR (2002)
16	Malnutrition prevalence, weight for age	% of children under 5	WB (2002)
17	MAR - Minorities at risk	Index scale	University of Maryland
18	Mobile phones in 2001	Phones per 1000 people	UNDP HDR (2002)
19	Rural population density	People per sq km	WB (2002)
20	Population growth	Annual %	WB (2002)
21	Population fertility rate (per woman)	Annual %	UNDP HDR (2003)
22	Rural population	% of total population	WB (2002)
23	Trade	% of GDP	WB (2002)
24	Urbanisation	% of total population	WB (2002)
25	Urban growth (95-00) –	% of growth rate	WB (2002)
26	HDI – Life Expectancy	Predefined index scale (0-1)	UNDP HDR (2003)
27	HDI – Education	Predefined index scale (0-1)	UNDP HDR (2003)
28	HDI – GDP	Predefined index scale (0-1)	UNDP HDR (2003)
29	Gross foreign direct investment	% of GDP	WB (2002)
30	Official development assistance and official aid	US $ / capita	WB (2002)
31	Military expenditure	% of GNI	WB (2002)
32	Gross private capital flows	% of GDP	WB (2002)
33	Fertilizer consumption	Grams per hectare of arable land	WB (2002)
34	Arms import	% of total imports	WB (2002)
35	External balance on goods and services	% of GDP	WB (2002)
36	Missing Values for the years 1990 – 2001	Number of missing values	WB (2002)
37	Armed Conflicts after 1980	Index scale	Uppsala University

Table 5: List of pre-selected sub-indicators as base for the factor analysis carried out for the development of a composite indicator for vulnerability estimation at national scale (extract of Table in Annex 2) (source: author).

Currently the database contains information on 285[3] ethnic minorities, which were politically active to an extensive degree at some point between1945 and 2000. Hence these ethnic groups represent all minorities that have been involved in a serious ethnic conflict in this period (FOX 2004). The MARGene database is unique regarding this information and has been widely used for scientific studies concerning ethnopolitical conflicts and discrimination with varying focus. FOX (2004) and FOX & SANDLER (2004) used the data to concentrate on religious aspects as driving force in politics and international relations. CAPRIOLI & TRUMBORE (2004) investigate the correlation between ethnic discrimination within countries and the involvement in international conflict of these countries. SAIDEMAN (2001) elaborates on the reasons for states intervening in other countries' ethnic-based conflicts and the consequences of these interventions. More recently, the MARGene database has also been exploited for the UN Human Development Report in order to assess the countries' 'level of liberty' and to quantify cultural exclusion (UNDP 2004a).

Within the MARGene database ethnic groups are allocated to one or more countries. For each country entry the group is coded separately, i.e. that Kurds in Turkey are assessed differently in their level of discrimination from Kurds in Iraq, Iran or Russia. Each of the 451 variables is coded for each ethnic group entry. The maximum code value that each variable may obtain varies with the variables.

For the creation of the vulnerability composite indicator, six out of the 451 variables were selected to serve as a base for the description of a country's vulnerability depending on potential ethnopolitical conflicts. Of these six variables in Table 6 below the first two variables are composite indictors themselves.

[3] The precise number of groups varies with different project phases: "The Minorities at Risk dataset has developed over four distinct phases. Phase I covered 227 communal groups which met the criteria for classification as a minority at risk for the years 1945-1989. Phase II covered 275 groups from 1990-1995, Phase III covered 275 groups from 1996-1998 and Phase IV covered 285 groups from 1998-2000" (MARGene codebook 2004, p 6).

No.	Name	Period of observation	Description	Coding, (for all: no value = 99)
1	POLRES - Extent of Political Restrictions	1990 - 2000	Political Restrictions Index, based on 9 sub-indicators: Freedom of expression, Freedom of movement, Rights in judicial proceedings, Restrictions on organizing, Restrictions on voting rights, Police/military recruitment, Civil service access, Access to higher office, Other political restriction	The code values 0-2 for each sub-indicator are summed up for the index (maximum code value: 8)
2	CULRES Extent of Cultural Restrictions	1990-2000	Index of Cultural Restrictions, based on 8 sub-indicators: Restrictions on religion, Restrictions on use of language, Restrictions on language instruction, Restrictions on ceremonies, Restrictions on appearance, Restrictions on family life, Restrictions on cultural organizations, Other cultural restrictions.	The code values 0-3 for each sub-indicator are summed up for the index (maximum code value: 7)
3	INTERCON Intercommunal conflict	since 1990	Variable contains information on open hostilities between the minority group and other communal groups. It includes open conflicts with other minorities and the majority or dominant group, but not conflicts with the state, or with dominant groups exercising state power.	Code 0 for no conflict and 1 for conflict
4	INTRACON Presence of intracommunal violence	since 1990	This variable records open hostilities between different fractions within the minority group.	Code 0 for no conflict and 1 for conflict
5	REB Annual Rebellion Index	1985-2000	Only the most serious manifestation of rebellion is coded	Codes: 0 None reported 1 Political banditry 2 Campaigns of terrorism 3 Local rebellion 4 Small-scale guerrilla activity 5 Intermediate guerrilla activity 6 Large-scale guerrilla activity 7 Protracted civil war [4]
6	atrisk1	1998	Is of dichotomous type and determines whether an ethnic group is subject to discrimination *at present* (1998)	Code 0 for no discrimination and 1 for discrimination

Table 6: List of selected variables from MARGene dataset for the creation of an indicator for country's vulnerability depending on potential ethnopolitical conflicts (source: author).

[4] The borderline between the code values in terms of conflict severity are more precisely defined in the MARGene codebook (2004, p 91)

Creating the composite sub-indicator 'Minorities at risk'

Firstly, all those entries of which the minorities were coded as 'not currently at risk' were excluded. Therefore, only those entries were kept, that had been allocated the value 1 to the 6th variable ('atrisk1')[5]. Secondly, for each of the variables No. 1-5 for each ethnic group for each country, the maximum value of the years 1995 – 2000 was extracted from the database. That is, the event with the highest potential to contribute to a conflict situation was selected. Thirdly, the code values for the two index variables 'POLRES' and 'CULRES' and the rebellion variable No. 5 'REB' of the remaining entries were recoded in order to reduce their value weight within the overall composite indicator in comparison to the values of the indicators 'INTERCON' and 'INTRACON' (see Table 7):

Old code value	*New code value*	*Relevance for conflict*
1-2	1	Very low
3-4	2	Low
5-6	3	Medium
7-8	4	High

Table 7: Look-up table for recoding of MARGene values 'POLRES' and 'CULRES' and 'REB' (source: author).

Fourthly, the values of the variables 1-5 were summed up for each minority in each country. Following, the values for all minority groups in each country were added. Finally, the variable was normalised as described later on in this chapter. An example of the computation of the value for Afghanistan is given in Annex 3.

Missing indicator values: [No of sub-indicator: 36]

The statistics compiled by international institutions such as the UN or the World Bank rely to a large extent on data made available within the respective country. The data cannot be collected when administrative systems are not functioning sufficiently or when they are disrupted due to a disaster, a war or any other restrictive conditions. Therefore the lack of indicator values is considered as an indicator itself. For the calculation of this indicator the following 10 datasets were selected, which are usually annually updated by the World Bank and which cover various sectors (health, economics, communication, tourism):

- Life expectancy at birth, total,
- School enrollment, secondary,
- GDP per capita, PPP,
- Unemployment rate,
- International tourism, number of arrivals,
- GNI (Gross National Income),
- Public spending on education,
- Public health expenditure,
- Trade in goods,
- Information and Communication Technology expenditure.

For each country the number of years with missing values, that is years for which there was no entry in the World Bank database, was summed up for the period 1990 – 2001.

[5] The value available in the MARGene database used for this study is referring to 1998

The number and severity of Armed Conflicts: [No of sub-indicator: 37]

The relevant data for this indicator is provided by the Uppsala Conflict Data Program (UCDP) of the Uppsala University[6]. This programme collects continuously and updates yearly data on armed conflicts worldwide (HÖGBLADH 2004). The armed conflict database covers disputes since 1946 and its free download is provided jointly by the Uppsala University and the Center for the Study of Civil War at the International Peace Research Institute, Oslo (PRIO)[7]. Within the frame of PRIO, an armed conflict is defined as "a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in one calendar year" (PRIO 2005, p 3). The conflicts integrated in this database are described by a range of variables. The most relevant for this study is the variable 'intensity' which indicates the severity of each conflict by allocating it to one of the following classes:

1. *Minor*: At least 25 battle-related deaths per year for every year in the period.
2. *Intermediate*: More than 25 battle-related deaths per year and a total conflict history of more than 1000 battle-related deaths, but fewer than 1,000 per year.
3. *War*: At least 1000 battle-related deaths per year.

The data used for this study are based on the 2003 version of the conflict database. From this dataset all conflict data for the period 1989 – 2002 was extracted, that is all events after the end of the Cold War. The data referring to the events concerning conflicts on the territory of the former Soviet Union (SU) between the years 1989-1992 were excluded since it is impossible to disaggregate them and allocate the conflict values to the respective Newly Independent States (NIS) which emerged out of the SU. In order to take appropriate consideration of the severity of the conflicts, the values of the intensity variable were recoded according to the scheme in Table 8. The recoded conflict intensity values for each country for the years 1989-2002 were then summed up.

Conflict intensity - Original value	Conflict intensity – recoded value
1	1
2	3
3	10

Table 8: Recoding of conflict intensity values taken from the PRIO database

3.4.3 Preparation for the statistical analysis of the collected sub-indicators

Missing data

For all indicators the latest value available for each country was taken into account. If there was no data available from 1990 or later the record was considered as 'no value'. A special note was added into the table in Annex 2 when not all the country values for one indicator are based on the same year. Unavoidably there are a number of countries where several indicator values are missing. Countries with too few indicator values available were

[6] URL: http://www.pcr.uu.se/database/index.php <viewed November 2006>

[7] URL: http://www.pcr.uu.se/research/UCDP/our_data1.htm <viewed November 2006>

excluded from the statistical computation when calculating the factor analysis. Those countries with more than 7 missing values of the 10 selected sub-indicators for the final composite indicator were allocated the value ‘no data’ in the final map (see Map 1).

Normalisation

All indicator values were normalised in order to avoid errors due to various measurement units. This was done by the re-scaling method allocating values between 0 and 1 (except the variables GDI [9], HDI [11] and the HDI sub-indicators [26, 27, 28] of which the values were ranked in this way beforehand). Compared to other normalisation methods this method has the advantage of greater increasing the effect on the composite indicator by widening the range of indicators with a small interval (NARDO et al. 2005b). The disadvantage of this method is the potential distortion of the indicators by outliers. Therefore, all values at the minimum or maximum end of the country values with more than 50 % of value difference to the next ranked country value were neglected. For example, for the sub-indicator ‘Agricultural productivity’ Brunei provides an outlier maximum country value with more than double the value of the next ranking country Belgium. Therefore Brunei’s value is neglected for the re-scaling process and the maximum value ‘1’ is manually allocated after the normalisation computation (Table 9). All outliers of the sub-indictors are listed in the table in Annex 2.

Country	Value	Normalised Value
France	55685.07421875	0.9304899498
Netherlands	55992.34765625	0.9356355475
Belgium	59835.921875	1
Brunei	152209.34375	1

Table 9: Example of normalisation of sub-indicator values with outlier (source: author).

After normalisation, where necessary the indicator values were changed in the way that ‘1’ as maximum value signifies ‘high vulnerability’ and the value ‘0’ signifies low vulnerability. For those indicators the values of which were changed, a ‘1-x’ notion was put in the table in Annex 2.

The resulting database comprises latest available normalised and re-scaled values for 37 sub-indicators for 262 countries indicating low vulnerability with a small value (minimum ‘0’) and high vulnerability with a high value (maximum ‘1’).

3.4.4 Implementation and results of statistical analysis

Based on these pre-selected sub-indicators a Principle Component Analysis (PCA) was implemented. PCA is one way of performing factor analyses, the aim of which is to identify one or several measurable variables that best describe complex non-measurable phenomena such as ‘vulnerability’ (BAHRENBERG et al. 2003). Their theoretical principle is that the selected variables can be transformed into linear combinations of hypothesised and unobserved components called ‘factors’. Factors may be associated with one or more of the original variables. The 'Eigenvalues' of these factors indicate the variance of the original set of variables that has been successfully extracted. The relationship between the original variables and the computed factors is specified by the factor loadings. The factors are derived successively resulting in a user-specified number of factors with decreasing correlation between factor and related variables. A detailed explanation of the methodological base of PCA and factor analysis is for example given by JOLLIFFE (1986), MANLY (1994) and LINDEMANN et al. (1980).

PCA and factor analysis are parametric techniques and hence requiring interval-scaled variables. All of the pre-selected sub-indicators possess this scale level with the exception of three variables: No. 7 'Corruption', no. 17 'Minorities at Risk' and no. 37 'Armed Conflicts'[8]. The values of these variables were observed rather than measured and they have an ordinal scale (in the former two cases 'Corruption' and 'Minorities at Risk') or the scale level is arguable (in between ordinal and interval scale in the latter case of 'Armed Conflicts'). It is here assumed that the divisions between categories of these three variables are equidistant and that it is therefore justifiable to integrate them in PCA analyses.

Within the scope of this study a PCA has been carried out in order (1) to determine correlations between the pre-selected sub-indicators and (2) to reduce the number of sub-indicators by focusing on those that explain most of the variance spanned by the set of pre-selected variables[9]. A certain level of correlation between the sub-indicators is the prerequisite for a successful application of a factor analysis aimed at reducing the number of variables (NARDO et al. 2005a). Therefore, in a first step of statistical analysis, a correlation matrix for the sub-indicator values was computed. A great number of variables are correlated with each other with correlation values >.5. Five pairs of those were identified with correlation values > .95. The information of indicator pairs with such a high correlation value can be considered as almost identical. This fact needs to be taken into account when interpreting the results of the statistical analysis. All correlation values for all indicators are shown in the table in Annex 4 where those indicator pairs exceeding the above mentioned thresholds are highlighted in red.

The PCA was carried out using the 'Pairwise Deletion of Missing Data' option thus ensuring a reasonable number of remaining data records. The factors resulting from the PCA were rotated with the 'Varimax normalised' function in order to ease the interpretation of the results. Following all those factors with an Eigenvalue < 1 were dropped according to the 'Kaiser criterion'. This exclusion criterion states that it does not make sense to consider factors that explain less variance than is contained in one sub-indicator (NARDO et al. 2005a). Seven factors were computed with Eigenvalues > 1 accounting for 78 % of the variance of the original dataset (see Table 10).

---

[8] The variables no. 9 'GDI' and no. 11 'HDI' are composite indicators themselves. Their underlying sub-indicators are without exception interval scaled, therefore also the composite indicators are assumed to possess interval scale.

[9] For the purpose of this study from a methodological point of view the application of an Exploratory Factor Analysis (EFA) would have been most appropriate. Unlike a PCA the EFA accounts only for the common variance in the dataset assuming that there is a certain variance which cannot be explained by the extracted factors. The PCA is based on the assumption that the whole variance can be explained by the extracted factors. However, the results of both methods resemble with increasing number of variables and it is accepted to use the PCA instead of the EFA when more than 15 variables are analysed HOLM (1976, p 71-72).

Eigenvalues (phdfinaldataset.sta)				
Extraction: Principal components				
		% total	Cumul.	Cumul.
	Eigenval	***Variance***	***Eigenval***	***%***
1	15,58857	42,131271	15,58857	42,131271
2	3,2818271	8,869803	18,870397	51,001074
3	2,9711858	8,0302319	21,841583	59,031306
4	2,2249311	6,0133274	24,066514	65,044633
5	1,9354368	5,2309102	26,001951	70,275543
6	1,4861707	4,0166777	27,488122	74,292221
7	1,4501619	3,9193564	28,938284	78,211577

Table 10: Results of PCA – the Eigenvalues and the Variance (source: author).

The interpretation of the results is facilitated by the fact that the allocation of the variables to the relating factors is rather unambiguous. That is, the majority of variables have a high value in respect to one factor (> 0.7) and lower values (< 0.3) in respect to the other factors. In general, the variables contributing most to a single factor can easily be grouped within specific topics. The contributions of the introduced new variables 'Minorities at risk', 'Armed conflicts' and 'Missing values' is rather low. This might be caused by the fact that they mainly provide information about 'critical' countries but have no validity for the large group of less vulnerable nations. Below, the results of the factor analysis are interpreted in more detail, focusing on all those variables with factor loadings > 0.5, which are highlighted in orange in Table 11. An exception is factor 1 where only the 16 variables with a loading > 0.7 are taken into account for the detailed analysis.

Factor 1 has by far the strongest link with the 'classical' development indicators, namely HDI, GDI and population fertility. It also embraces the countries' economic situation in the primary sector and the degree of industrialisation. This is proved by the high loading values for the variables describing the added value of agricultural activity to the GDP, the employment rate in the agricultural sector, the rural population and urban growth. The second factor describes the importance of trading for the nations' economy, with high loading values for the variables dealing with trade, import and export. Factor 3 also describes the economic situation but focuses on the external balance of goods and services and the absolute and relative amount of external aid. A considerable contribution in this factor is based on the 'missing values' variable. Factor 4 explains characteristics in the field of conflicts and military expenditure. Accordingly, the variable 'Minorities at risk' has its highest loading value for this factor. Factor 5 deals with the type and quality of the economy with strongest contributions from the variables corruption, mobile phones and the economic indicators 'agricultural productivity', 'gross foreign direct investments' and 'gross private capital flows'. Factor 6 describes another aspect of the level of industrialisation mainly composed of the indicators rural population density and fertiliser consumption. Factor 7 has only one high loading value in the variable 'arable land'.

Factor Loadings (Varimax normalized)								
Extraction: Principal components								
(Marked loadings are > .500000)								
Nr	Name	1	2	3	4	5	6	7
1	AGGR_ADD	0,7260	0,1654	0,0808	0,0040	0,3115	0,0916	-0,1514
2	AGGR_EMP	0,8188	0,0624	0,0478	0,1124	0,1741	0,0736	-0,1902
3	AGGR_PRO	0,3987	-0,0945	0,1122	-0,0070	0,7918	0,0666	0,0088
4	AIDOFGIN	0,4075	0,0180	0,7704	-0,0113	0,0253	0,1047	-0,0114
5	AIDS	0,5173	0,1056	0,0157	-0,2253	-0,1700	0,2125	0,0728
6	ARABLELA	-0,1409	-0,0221	0,0549	-0,0672	0,0548	0,0180	-0,7203
7	CORR	-0,4786	-0,0512	-0,1288	-0,0554	-0,8025	-0,1420	0,0993
8	EXP%GDP	0,2000	0,8983	0,2697	0,0578	0,1489	0,0836	-0,0691
9	GDI	0,9136	0,1262	0,2623	0,0540	0,3431	0,0553	0,1023
10	GDP	0,7968	0,0806	0,1537	0,0398	0,4816	0,0890	-0,0663
11	HDI02	0,9187	0,1095	0,1423	0,0660	0,2875	0,0627	0,0838
12	IMP%GDP	0,0995	0,8900	-0,3133	0,0439	0,1129	0,0814	0,0885
13	IMPWATAC	0,7388	-0,0466	0,0118	0,1934	0,2316	0,1514	0,1439
14	LIFEEXPE	0,8908	0,0739	-0,0289	-0,0113	0,1668	0,2206	0,0730
15	LITRATEA	0,8226	0,2506	0,0578	0,0910	-0,0621	-0,0778	0,1596
16	MAL_5	0,8003	0,0079	-0,0908	0,2149	0,0393	-0,2777	-0,1721
17	MAR	0,0977	0,1508	-0,3895	0,4346	0,2522	0,1436	-0,2566
18	GSM	0,5541	0,0452	0,1749	0,0166	0,6796	0,0486	0,0484
19	POPDENRU	0,1509	-0,0968	0,1865	0,0670	0,2098	-0,7952	0,0870
20	POPGTHAN	0,5948	0,0127	0,0434	0,1924	0,0767	-0,2287	0,4916
21	POPFERT	0,8706	0,0647	0,1483	0,1446	0,1011	-0,0322	0,2830
22	RURPOP	0,7453	0,0105	0,1725	-0,0151	0,3254	-0,1674	-0,3601
23	TRAD%GDP	0,1598	0,9531	-0,0102	0,0559	0,1328	0,0878	0,0054
24	URBANIS	0,7587	-0,0322	0,0707	-0,0164	0,3036	-0,1696	-0,4086
25	URBGTH	0,7694	0,0974	0,0395	0,1359	0,1085	-0,1431	0,2928
26	HDILIFE	0,8699	0,0547	0,0718	0,0061	0,1724	0,2089	0,1336
27	HDIEDU	0,8491	0,1430	0,1076	0,1505	0,1433	-0,0819	0,1493
28	HDIGDP	0,7903	0,0976	0,1930	0,0383	0,4820	0,0597	-0,0508
29	GROSSINV	-0,1196	-0,3866	-0,0253	0,0339	-0,7116	0,1671	-0,0443
30	ODA	-0,0301	-0,1010	0,7724	-0,0308	0,1622	-0,1775	0,0138
31	MILITEXP	0,1052	-0,0819	0,0712	0,8597	0,0293	-0,0322	0,2306
32	GROSSPRI	-0,1232	-0,3992	-0,0805	-0,0302	-0,6672	0,0994	0,1036
33	FERTCONS	-0,3346	-0,1158	-0,0677	-0,0472	-0,2924	-0,7044	-0,0060
34	ARMSIMP	0,0784	0,0773	0,1207	0,7569	-0,0355	-0,0836	0,0923
35	EXTBALAN	-0,1705	-0,1485	-0,8338	-0,0360	-0,1006	-0,0099	0,2040
36	MISSVAL	0,2436	-0,0833	0,5968	0,2396	0,3945	-0,0089	0,3709
37	ARMCONFL	0,2343	0,2051	-0,3196	0,5743	-0,0434	0,1140	-0,2705

Table 11: Results of PCA - the Factor Loadings. All factor loadings >0.5 are highlighted in orange. The grey boxes indicate indicators with inverse values (1-x) (source: author).

Selection of variables for the composite indicator

Based on the fact that the variables with the highest loading values of the seven factors represent well the major aspects of the vulnerability to be determined, it was decided to consider all factors and their main contributing variables in an equalised way. That is, for the selection procedure the spread of variables over various factors received a higher priority than the absolute values of factor loadings or the explanatory contribution of the single factor to the overall variance. In practice, the overall number of variables to be selected was limited to ten and the maximum number of variables chosen per factor was limited to two. The selection of variables for each factor followed specified decision based rules: The first variable selected is the one with the highest factor loading value. The second variable selected is the one with the next highest loading value that additionally has to fulfil the requirement of a correlation < 0.5 with the first variable. If none of the lower loaded variables fulfils these prerequisites only the first variable is selected.

The final composite indicator is based on the 10 first selected sub-indicators as listed in Table 12. The quartiles for the values of each indicator were computed and all values coded according to the respective quartile within the range of ‘0’, indicating low vulnerability, up to ‘3’, representing high vulnerability. The final CI country score is the simple average of the sub-indicator values.

Variable		**Factor**
Nr	**Name**	
11	HDI	1
5	HIV/AIDS	1
23	Trade	2
35	External Balance	3
36	Missing Values	3
31	Military Expenditure	4
37	Armed Conflicts	4
7	Corruption	5
3	Agricultural Productivity	5
19	Rural Population Density	6

Table 12: Finally selected indicators for the creation of the Composite Indicator (source: author).

A frequency table was created for the 262 countries in the database summing up the number of missing sub-indicator values (see Figure 16). Based on the result for the following working steps only those countries with three or more sub-indicator values were taken into account, resulting in 193 country records.

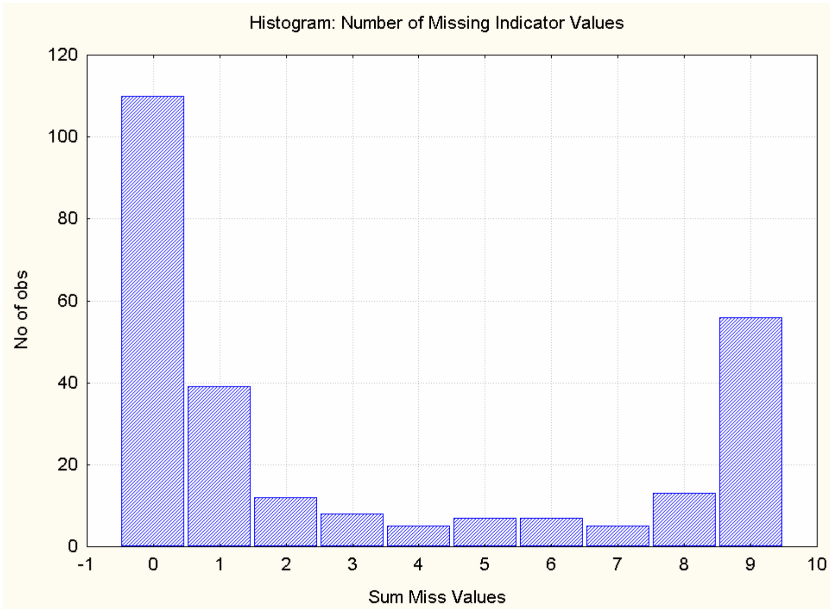

Figure 16: Frequency table of missing sub-indicator values (source: author).

3.4.5 The composite indicator – results and discussion

The final ranking list of countries with sufficient sub-indicator values is presented in Table 13, showing the country rank, the country name, the number of missing sub-indicator values and the computed CI value. The same ranking number was allocated to countries with identical CI values, resulting in 58 ranks for 193 considered nations. The complete list of all countries (including those not considered in the final result due to the lack of indicators values) and sub-indicator values is shown in the table in Annex 5 a) and b). In Annex 5 a) the countries are sorted by CI ranking values, in Annex 5 b) they are sorted by names.

RANK	COUNTRY_NAME	SUM_MISS	RES
1	Burundi	1	2,888889
1	Rwanda	1	2,888889
2	Ethiopia	0	2,700000
2	Sierra Leone	0	2,700000
2	Uganda	0	2,700000
3	Congo, Democratic	1	2,666667
4	Eritrea	2	2,625000
5	Somalia	5	2,600000
6	Chad	1	2,555556
6	Haiti	1	2,555556
7	Mozambique	0	2,500000
8	French Polynesia	5	2,400000
8	New Caledonia	5	2,400000
9	Guinea-Bissau	1	2,333333
9	Yugoslavia	4	2,333333
10	Angola	0	2,300000
11	Djibouti	2	2,250000
12	Burkina Faso	1	2,222222
12	Cambodia	1	2,222222
12	Myanmar	1	2,222222
12	Sudan	1	2,222222
13	Congo, Republic o	0	2,200000
13	Tanzania	0	2,200000
13	Zimbabwe	0	2,200000
14	Guinea	1	2,111111
14	Nepal	1	2,111111
14	Vietnam	1	2,111111
15	Bangladesh	0	2,100000
15	India	0	2,100000
15	Mali	0	2,100000
15	Senegal	0	2,100000
16	Afghanistan	7	2,000000
16	Guam	7	2,000000
16	Iraq	4	2,000000
16	Kenya	0	2,000000
16	Korea, North	6	2,000000
16	Lesotho	1	2,000000
16	Libya	4	2,000000
16	Marshall Islands	6	2,000000
16	Netherlands Antil	7	2,000000
16	Pakistan	0	2,000000
16	Palau	6	2,000000
16	Puerto Rico	7	2,000000
16	Virgin Islands	7	2,000000
16	Yemen	0	2,000000
17	Malawi	0	1,900000
18	Benin	1	1,888889
18	Central African R	1	1,888889
19	Bhutan	2	1,875000

RANK	COUNTRY_NAME	SUM_MISS	RES
22	Bahamas, The	6	1,750000
22	Comoros	2	1,750000
22	Micronesia, Feder	6	1,750000
22	Sao Tome and Prin	2	1,750000
23	Armenia	0	1,700000
23	Cameroon	0	1,700000
23	Ecuador	0	1,700000
23	Iran	0	1,700000
23	Lebanon	0	1,700000
23	Morocco	0	1,700000
23	Peru	0	1,700000
23	Tajikistan	0	1,700000
23	Zambia	0	1,700000
24	Oman	2	1,625000
25	Azerbaijan	0	1,600000
25	Bolivia	0	1,600000
25	Botswana	0	1,600000
25	China	0	1,600000
25	Colombia	0	1,600000
25	Cote d'Ivoire	0	1,600000
25	Ghana	0	1,600000
25	Indonesia	0	1,600000
25	Tonga	5	1,600000
25	Turkey	0	1,600000
25	Uzbekistan	0	1,600000
26	Grenada	3	1,571429
26	Vanuatu	3	1,571429
27	Suriname	1	1,555556
28	Algeria	0	1,500000
28	Dominican Republi	0	1,500000
28	Honduras	0	1,500000
28	Jordan	0	1,500000
28	Macedonia, The Re	0	1,500000
28	Nigeria	0	1,500000
28	Papua New Guinea	0	1,500000
28	South Africa	0	1,500000
28	Western Samoa	4	1,500000
29	Equatorial Guinea	1	1,444444
30	Albania	0	1,400000
30	Kyrgyzstan	0	1,400000
30	Russia	0	1,400000
30	Solomon Islands	5	1,400000
30	Syria	0	1,400000
31	Cape Verde	2	1,375000
31	Seychelles	2	1,375000
32	Cuba	1	1,333333
32	Gabon	1	1,333333
32	Kuwait	1	1,333333
32	Qatar	1	1,333333

RANK	COUNTRY_NAME	SUM_MISS	RES
37	Moldova	0	1,200000
37	Nicaragua	0	1,200000
37	Panama	0	1,200000
37	Paraguay	0	1,200000
37	Thailand	0	1,200000
38	Dominica	4	1,166667
39	Brunei	3	1,142857
40	Israel	1	1,111111
41	Argentina	0	1,100000
41	Croatia	0	1,100000
41	France	0	1,100000
41	Portugal	0	1,100000
41	Ukraine	0	1,100000
42	Bahrain	1	1,000000
42	Brazil	0	1,000000
42	Kazakhstan	0	1,000000
42	Poland	0	1,000000
42	Romania	0	1,000000
42	Saudi Arabia	0	1,000000
42	Trinidad and Toba	0	1,000000
42	Turkmenistan	1	1,000000
42	Venezuela	0	1,000000
43	Costa Rica	0	0,900000
43	Latvia	0	0,900000
43	Mauritius	0	0,900000
43	Mexico	0	0,900000
43	United Kingdom	0	0,900000
44	Barbados	1	0,888889
45	Maldives	2	0,875000
46	Belarus	0	0,800000
46	Lithuania	0	0,800000
46	Malaysia	0	0,800000
46	Spain	0	0,800000
46	Tunisia	0	0,800000
46	Uruguay	0	0,800000
47	Cyprus	1	0,777778
47	Fiji	1	0,777778
48	Malta	2	0,750000
49	Australia	0	0,700000
49	Canada	0	0,700000
49	Chile	0	0,700000
49	Iceland	0	0,700000
49	Japan	0	0,700000
49	Norway	0	0,700000
50	Switzerland	1	0,666667
51	Bulgaria	0	0,600000
51	Czech Republic	0	0,600000
51	Denmark	0	0,600000
51	Germany	0	0,600000

20	Bosnia and Herzeg	0	1,800000
20	Egypt	0	1,800000
20	El Salvador	0	1,800000
20	Gambia, The	0	1,800000
20	Guatemala	0	1,800000
20	Liberia	5	1,800000
20	Madagascar	0	1,800000
20	Namibia	0	1,800000
20	Sri Lanka	0	1,800000
20	West Bank	5	1,800000
21	Laos	1	1,777778
21	Mauritania	1	1,777778
21	Niger	1	1,777778
21	Swaziland	1	1,777778
21	Togo	1	1,777778
33	Greece	0	1,300000
33	Jamaica	0	1,300000
33	Philippines	0	1,300000
34	Antigua and Barbu	3	1,285714
34	Saint Kitts and N	3	1,285714
34	Saint Lucia	3	1,285714
34	Saint Vincent and	3	1,285714
35	Kiribati	6	1,250000
35	United Arab Emira	2	1,250000
36	Guyana	1	1,222222
36	Mongolia	1	1,222222
36	United States	1	1,222222
37	Belize	0	1,200000
37	Georgia	0	1,200000
37	Italy	0	1,200000
51	Korea, South	0	0,600000
52	Estonia	0	0,500000
52	Hungary	0	0,500000
52	Macau	6	0,500000
52	Netherlands	0	0,500000
52	Slovakia	0	0,500000
52	Slovenia	0	0,500000
53	Singapore	1	0,444444
54	Austria	0	0,400000
54	Belgium	0	0,400000
54	New Zealand	0	0,400000
54	Sweden	0	0,400000
55	Luxembourg	2	0,375000
56	Finland	0	0,300000
57	Ireland	1	0,222222
58	Hong Kong	0	0

Table 13: Ranking list of countries according to the people's estimated hazard-independent vulnerability (source: author).

For the visualisation of the final results in the global map 'Hazard Independent Vulnerability of Populations – an estimation at national level' (see Map 1) the resulting dataset of the CI country values was divided into seven classes plus the class 'no value'. The allocation of country values to a specific vulnerability class, that is the definition of threshold values between the vulnerability classes, is geared to those suggested by the Jenks methodology (JENKS 1967) (see Figure 17). The Jenks optimisation method is also known as the Goodness of Variance Fit (GVF). It is used to minimize the squared deviations of the class means. Optimisation is achieved when the quantity GVF is maximized. Originally, class threshold values were computed with the GVF for 8 classes. However, for this study the first two classes were merged into the class 'very low vulnerability'. The remaining class limits suggested by this methodology were adapted slightly in order to achieve a finer distinction between those countries with higher vulnerability values, since these countries are our main concern.

In general the list of countries ranked according to the introduced CI follow expected patterns with lower values for industrialised nations and maximum vulnerability estimation allocated to Africa. In a general tendency the country ranking of vulnerability is similar to a ranking according to the status of development. However, there are some cases where the country values do not coincide with the usual picture of national development progress and which deserve scrutinising. The most obvious discrepancy with respect to the development status is based on the integration of the sub-indicators 'military expenditure' and 'armed conflicts'. Due to high values, particularly regarding military expenditure, and keeping in mind that the values of the CI predominantly reflect the situation of the years 1990-2000, nations such as France, Yugoslavia and Iraq rank relatively high on the list. In contrary, certain countries with low armed conflict values, such as Kazakhstan, Turkmenistan and Mexico, appear relatively low on the vulnerability ranking list compared with their overall development status. Special attention requires the high rank (No. 36) of the United States (US). Firstly, this is the case due to their expenditure of weapons and involvement in conflicts. Additionally, the US receives a high vulnerability value for its low level of trading compared with its overall GDP. Of course, the usefulness of this sub-indicator for very large countries, which are to a certain degree independent from the import of foreign products or services (such as the US, Australia or even Brazil) is debateable. However, the high ranking of the US is also influenced by a high value for AIDS/HIV infection rate which indicates the huge social differences within the country. The fact of social inequality contributes definitively to a higher vulnerability, as demonstrated by the devastating impact on a limited part of the population of hurricane Katrina in New Orleans in summer 2005. Nevertheless, for future development of the indicator it is proposed to cluster the countries prior to the statistical analysis in particular considering the country size. This would overcome some of the drawbacks of the methods applied here, which selects sub-indicators and interprets their value in a generic way.

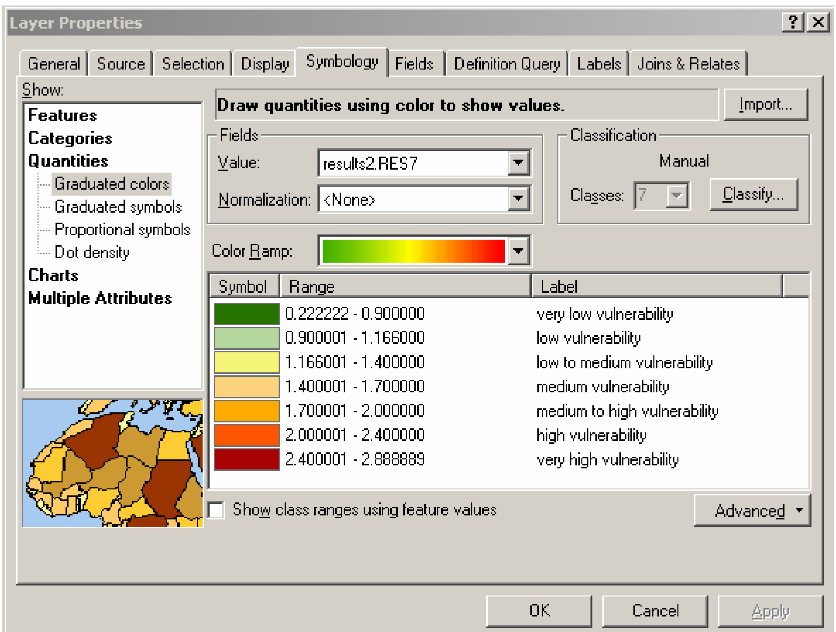

Figure 17: Classification of CI country values according to Jenks' methodology of Goodness of Variance (source: author).

As a conclusion it can be stated, that the vulnerability CI introduced here suffers from the lack of appropriate available data in the same manner as other similar approaches. Its use is limited by its coarse spatial resolution at country scale. Additionally, the creation of the composite indicator required many subjective decisions about selection and values' classification regarding the sub-indicators and their final aggregation. The CI remains a rough overview and the individual sub-indicators might be more or less suitable for a specific type of nation or for particular regions. However, it is felt that the use of sub-indicators aimed at determining administrative efficiency and the level of conflicts at different societal levels is of great importance for vulnerability estimations. Hence, the integration of indicators related to corruption, conflict and lack of available data contributes to a more realistic view of vulnerability than vulnerability indicators that are solely focusing on classical economic and development indicators. Nevertheless It should be emphasised again that the final CI results (1) need to be interpreted with the consideration that severe simplification and generalisation is needed in order to quantify the complex phenomena 'vulnerability' and (2) that all numbers and the selected methodology of data aggregation and ranking is based on the fact that the resulting numbers are at the most only able to reflect vulnerability in relative and not absolute values.

Map 1: Hazard Independent Vulnerability of Populations – an estimation at national level

3.5 The calculation of earthquake risk

In the previous three chapters the underlying components for the risk equation as developed in section 2 have been elaborated. In order to carry out the final calculation, the datasets required the following preparation:

- The GSHAP dataset was taken as reference regarding the dataset extent and the spatial resolution.
- The vulnerability layer, originally existing as vector layer with attribute data, was rasterised resulting in pixels conforming to those of the GSHAP layer and possessing the final vulnerability output values.
- The Landscan 2002 population dataset was processed as follows: First, the natural logarithm (ln) was computed in order to avoid an overvaluation and dominance of the very high pixel numbers of this dataset in relation to the former ones. Second, the spatial resolution was degraded by factor 12 in order to be consistent with the other 2 data layers. As a result, each of the resulting population pixel values represents the mean of the ln of a square of 12x12 original Landscan values.

The final global dataset for the estimation of risk of loss of life due to earthquakes is the result of the simple multiplication of these pre-processed datasets, resulting in values between 0 and 120. 'No value' was allocated to all those pixels with no data in any of the three underlying data layers. For the visualisation of this outcome in Map 2 the final risk values were classified using exponentially increasing class limits and respective colours as shown in Figure 18.

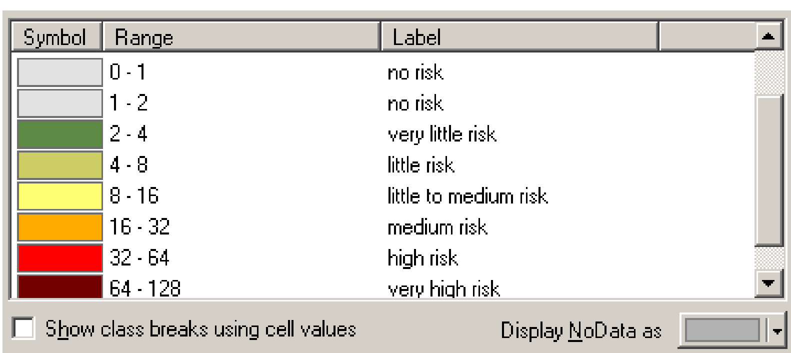

Figure 18: Classification scheme for the map 'Earthquake Risk – risk estimation of the loss of lives due to earthquakes - (source: author).

Map 3 compares the underlying data layers and the resulting overall risk map at a larger scale (1 : 40,000,000) for two regions of interest regarding seismic risk, namely South-East Asia and East Europe / the Middle East.

Map 2: Earthquake Risk – risk estimation of the loss of lives due to earthquakes

Map 3: Risk of Life Due to Earthquakes – example areas at larger scale

3.6 Discussion

The results have to be scrutinised by having in mind:

- the limitations considering accuracy and / or completeness of the three main input layers, namely:
    - Vulnerability
        - People's Vulnerability: as a starting point - due to its complexity – may only be described in an incomplete way.
        - The proposed vulnerability assessment is focusing on the 'harzard independent' component of vulnerability, the hazard specific part could not be considered due to the lack of data.
        - The available data restrict the vulnerability layer to country level, that is all sub-national characteristics and discrepancies within a country are not considered.
        - The vulnerability assessment is based on a model, which can only attempt to represent reality; several decisions during the modelling development had to be made in a rather subjective manner.
        - It is not possible to attach any accuracy information to the vulnerability layer for many reasons, the most important of which is that there is no absolute and correct vulnerability to which an accuracy of the here proposed method could be evaluated.
    - Hazard
        - The GSHAP map simplifies by not considering local geological characteristics.
        - The pga values as base for the seismic risk assessment are crucial for the resilience of a specific type of objects, hence covering the majority of objects but not all of them.
        - Secondary effects such as soil liquefaction or tsunamis are not considered.
    - Exposure
        - There is no detailed information available about the model on which the Landscan data layer is based.
        - There is no accuracy information attached to the Landscan dataset.

Notwithstanding these crucial points of incompleteness and the ignorance about the level of accuracy the final result visualised in Map 2 help to focus on regions of major concern. The hot spots of potential loss of lives due to earthquakes are easily recognisable. Areas of relatively low seismic risk but with a very high population density (example: Jakarta / North West Java) or a very high vulnerability (example: Nepal) are identified as hot spots. However, those areas with very high seismic risk but a low vulnerability value do only appear as medium risk zones (example: parts of Italy and Greece).

As a result the methodology proposed in this study is a valuable alternative for the few existing studies looking at natural disaster hot spots worldwide with the great advantage not to be based on post-event mortality data and not to rely on the limited existing disaster data base. First comparisons with the results of the work from DILLEY et al. (2005) show accordance in the identification of the areas at maximum risk of earthquakes. However, the here presented methodology proves to be more detailed and appropriate for the demarcation of areas assessed to be at low and medium risk of earthquakes.

In principle the method could be relatively easily transferred in order to assess people's risk of other natural hazards. However, in practice this approach is hampered by the – nowadays still - missing data regarding the potential occurrence of those hazards, for example of meteorological type.