TECHNICAL ANNEX

The Technical Annex provides supporting material on methodologies and results to supplement the main body of the Report. In particular, it provides background on the statistical work undertaken in the development of the Disaster Risk Index (DRI).

This is a detailed account of the work that was carried out in the DRI, the challenges that require further attention and the potential that exists for further work.

T.1 Definition of Statistical Terms

In the Glossary, we have included a set of key terms which are referred to throughout the Report. In order to aid comparability, in most cases we stay close to those used in the ISDR Secretariat publication *Living in Risk.* At the same time, the development of the DRI required the adoption of specific working definitions that guided the statistical analysis undertaken.

In this section, we present an extract of terms from the Glossary followed by the specific working definition of the term used in the development of the DRI.

Natural Hazard: Natural processes or phenomena occurring in the biosphere that may constitute a damaging event. Hazardous events vary in magnitude, frequency, duration, area of extent, speed of onset, spatial dispersion and temporal spacing.[1]

In the DRI: Natural hazards refer exclusively to earthquake, tropical cyclone, flood and drought. Only frequencies and area of extent were considered in the model. Magnitude is taken into account indirectly when possible. Secondary hazards triggered by the primary hazards mentioned above (for example, landslides triggered by earthquakes) are subsumed in the primary hazard.

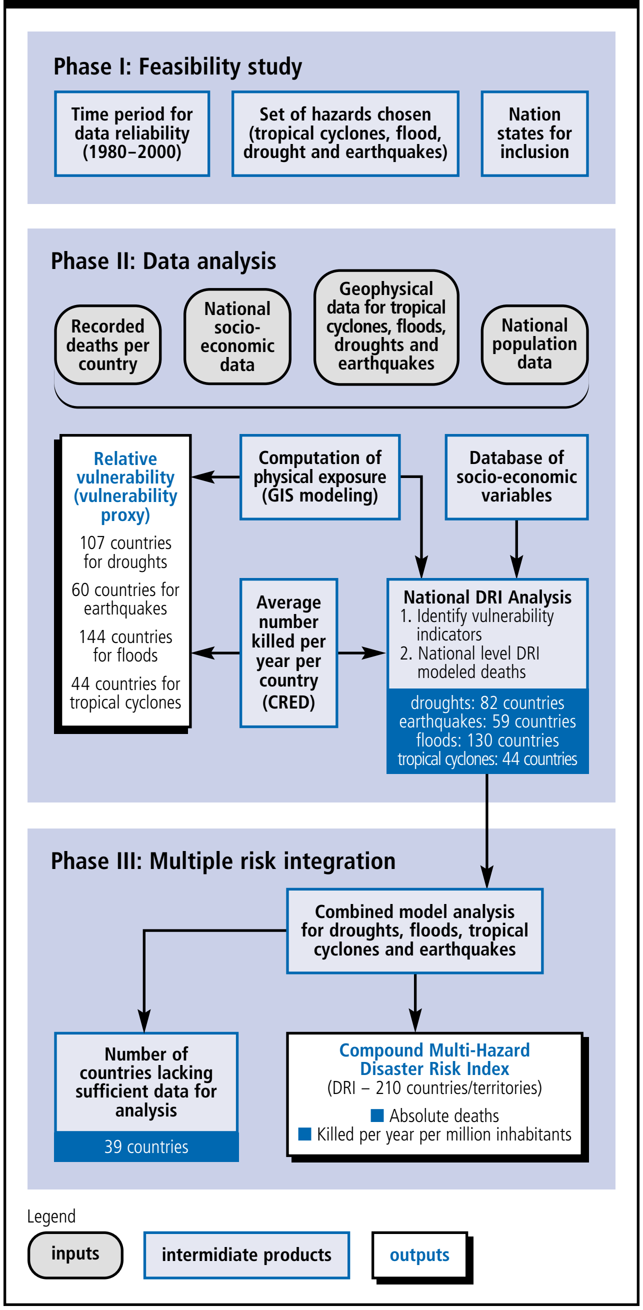
Physical Exposure: Elements at risk, an inventory of those people or artefacts that are exposed to the hazard.[2]

In the DRI: Physical exposure refers to the number of people located in areas where hazardous events occur combined with the frequency of hazard events.

Human Vulnerability: A human condition or process resulting from physical, social, economic and environmental factors, which determine the likelihood and scale of damage from the impact of a given hazard.

In the DRI: Human vulnerability refers to the different variables that make people more or less able to absorb the impact and recover from a hazard event. The way vulnerability is used in the DRI means that it *also* includes anthropogenic variables that may increase the severity, frequency, extension and unpredictability of a hazard.

FIGURE T.1 FLOW CHART OF THE GLOBAL RISK AND VULNERABILITY TREND PER YEAR (GRAVITY) PROJECT



Natural Disaster: A serious disruption triggered by a natural hazard causing human, material, economic or environmental losses, which exceed the ability of those affected to cope.

In the DRI: Disasters are a function of physical exposure and vulnerability.

Risk: The probability of harmful consequences or expected loss (of lives, people injured, property, livelihoods, economic activity disrupted or environment damaged) resulting from interactions between natural or human-induced hazards and vulnerable conditions. Risk is conventionally expressed by the equation Risk = Hazard + Vulnerability.

In the DRI: Risk refers exclusively to loss of life and is considered as a function of physical exposure and vulnerability.

T.2 Sourcing Data

T.2.1 EM-DAT Database

The DRI exercise is calibrated against the mortality data in the EM-DAT global disaster database. It is important to be clear about the data collection and management methods employed by EM-DAT.

The Centre for Research on the Epidemiology of Disasters (CRED) maintains the EM-DAT database at the University of Louvain in Belgium. Events that conform to a consistent definition of a disaster are included in the database. Such events meet at least one of the following criteria: 10 or more people reported killed; 100 people reported affected; a call for international assistance; and/or a declaration of a state of emergency. Information on losses comes from secondary sources (government reports, the International Federation of the Red Cross and Red Crescent Societies (IFRC) and other disaster relief agencies, Reuters, reinsurance company assessments) and is cross-checked where possible. These criteria exclude smaller loss events which are not considered disasters.

One important quality of EM-DAT is its management by an independent academic institution that encourages public access and scrutiny of the dataset. Great care is taken to verify disaster reports and emphasis is placed on the higher confidence that can be placed on the accuracy of deaths over those injured, made homeless or affected by disaster, although information is also made available for these categories.

Two other global disaster databases are maintained by the Munich Re Group and Swiss Reinsurance Company, but are not publicly available. A study by CRED (commissioned by the ProVention Consortium[3]) carried out a comparison of EM-DAT, Swiss Re and Munich Re natural disasters databases for four countries (Honduras, Mozambique, India and Viet Nam) between 1985 and 1999. Although the report stated that all three databases furnish the world community with 'acceptable levels of data on disasters',[4] it discovered significant variations among these datasets in both the events recorded and losses reported.

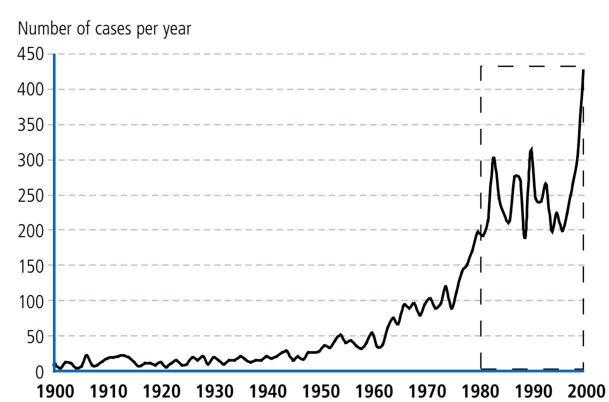
These differences were explained by differences in recording practice: what date each event is given, differences in classificatory methodology for each hazard type (a problem if one hazard triggers another) and the multiple entry of a single disaster event. As a result, the study found considerable differences between the datasets in the number of people affected (66 percent) and to a lesser extent the number of deaths (37 percent) and physical damage (35 percent). This is not surprising, since the definition of people affected varies enormously from disaster to disaster and from reporting source to reporting source. It is the most difficult impact variable to quantify and for this reason has not been used in the DRI work. The report also showed that the differences between the databases reduced significantly with time. This reflects EM-DAT's practice of reviewing its databases to incorporate updated information as it becomes available, even years after an event. A main weakness with global disaster data is the lack of standardised methodologies and definitions. This weakness is being addressed through the development of a unique global identifier for disaster reporting, the GLIDE system discussed in Chapter 2.

As mentioned above, EM-DAT explicitly excludes events where the loss is below defined threshold levels. A study undertaken on behalf of the ISDR Working Group 3 on Risk, Vulnerability and Impact Assessment, compared national disaster databases developed using the DesInventar methodology with the EM-DAT databases in four countries (Colombia, Chile, Panama and Jamaica). In all four countries, small-scale disasters with losses below the EM-DAT threshold represented a variable proportion of total disaster loss. Additionally, the national databases contained data on a number of medium-scale disasters that were above the EM-DAT threshold, but which were not captured by international reporting. It is impossible to arrive at a firm conclusion from a four-country study regarding what percentage of total disaster loss is not captured by international reporting, and in any case this will vary from country to country. Again, the adoption of a unique identifier such as GLIDE in both national and global databases like EM-DAT should progressively improve the consistency of disaster reporting.

Given that the DRI is calibrated against mortality data from EM-DAT, under- or over-reporting of this variable in EM-DAT would affect the DRI results. However, the DRI takes into account the varied reporting for individual disasters by basing its analysis on average losses over a 20-year period (1980-2000). The EM-DAT database provides a very good sample of total disaster loss in this period with a national level of resolution.

This period provides a reasonable length of time to account for fluctuation in the occurrence of most hazard types and also coincides with the most reliable period of data collected in EM-DAT. Figure T.2 shows the total number of disasters recorded by EM-DAT from 1900 to 2000. The upward trend at first suggests an exponential increase in disaster frequency. However, improvement in disaster reporting is a substantial

FIGURE T.2 DISASTERS RECORDED BY EM-DAT



Source: EM-DAT: The OFDA/CRED International Disaster Database contributing factor.[5] While one cannot rule out that the number of hydrometeorological hazard events may have increased, the upward trend in reported disasters is more likely to be tied to improvements in telecommunication technology and the increasingly global coverage of different information networks. This makes the reporting and recording of disaster losses more possible today than in the past.

T.2.2 Choice of hazard types

The decision to limit the DRI to earthquake, tropical cyclone, flood and drought was based on two factors. First, the dominance of these hazard types in being associated with lives lost to disaster in past records (94.43 percent). Secondly, the availability of usable geophysical and hydrometeorological data to model each hazard's comparative extent and potential severity of impact. Data had to be available at the global level but detailed enough to map risk within each country.

During a preliminary investigation, volcanic eruptions were also considered. They were finally excluded because of the complexity of modelling the spatial extent of volcanic hazard events. Other types of hazards that may lead to disasters and influence the process of human development, such as technological and biological hazards, are not covered by the DRI, nor are natural hazards with more prominence at the local scale such as landslides. These could be included in the future when global datasets of events with national resolution come into use.

T.2.3 Choice of country cases

The DRI exercise aims to include all sovereign states in its analysis. This is compromised in two ways. First, there are varying levels of data availability. The decision here was to include all states from the outset, but discount those with inadequate data from detailed analysis. This partly accounts for the uneven number of states entered into the hazard-specific analyses. Secondly, a number of territories are classified as dependent territories or overseas departments. Such dependencies are often small islands or enclaves geographically distant from, but politically and administratively tied to, sovereign states such as France, the United Kingdom, USA or China. Overseas territories and sovereign states often exhibit very different socio-economic and environmental characteristics and hazard profiles. Where possible such territories have been analysed in their own right.

T.2.4 Outline formula and method for estimating risk and vulnerability

The formula used for modelling risk combines its three components. Risk is a function of hazard occurrence probability, the element at risk (population) and vulnerability. The equation below was made for modelling disaster risk.

0 (hazard) x population x vulnerability = 0 (risk)

The three factors used to construct this statistical explanation of risk were multiplied with each other. This meant that if the hazard was null, then the risk was null. The risk was also null if nobody lived in an area exposed to hazard (population = 0). The same situation held if the population was invulnerable (vulnerability = 0, induce a risk = 0).

From this, a simplified equation of risk[a] was constructed:

EQUATION 1 RISK

EQ1 $R = H \cdot Pop \cdot Vul$

Where
R is the risk (number of killed people.
H is the hazard, which depends on the frequency and strength of a given hazard
Pop is the population living in a given exposed area
Vul is the vulnerability and depends on the socio-politial-economical context of this population

Hazard multiplied by the population was used to calculate physical exposure.

EQUATION 2 RISK EVALUATION USING PHYSICAL EXPOSURE

EQ2 $R = PhExp \cdot Vul$

Where
PhExp is the physical exposure, i.e. the frequency and severity multiplied by exposed population

Physical exposure was obtained by modelling the area affected by each recorded event. Event frequency was computed by counting the number of events for the given area, divided by the number of years of observation (in order to achieve an average frequency per year). Using the area affected, the number of people in the exposed population was extracted using a Geographical

a. The model uses a logarithmic regression, the equation is similar but with exponent to each of the parameters.

EQUATION 3 ESTIMATION OF THE TOTAL RISK

EQ3 $$Risk_{Tot} = \sum(Risk_{Flood} + Risk_{Earthquake} + Risk_{Volcano} + Risk_{Cyclone} + ...Risk_n)^{b}$$

Information System (GIS). The population affected multiplied by the frequency of a hazard event for a specified magnitude provided the measure for physical exposure.

Socio-economic variables that could be statistically associated with risk were identified by replacing the risk in the equation with deaths reported in EM-DAT. A statistical analysis was then run to identify links between socio-economic and environmental variables, physical exposure and observed deaths.

The magnitude of events was taken into account by drawing a threshold above which an event is included. In the case of earthquakes, the threshold was placed at 5.5 on the Richter scale. Then the magnitude was partially taken into account by approaching the size of the area affected in relation to the magnitude, for the computation of physical exposure. Estimating event magnitude for use in global assessments is an area where there is great scope for improvement.

Scores for aggregated hazard deaths were calculated at the national level. Expected losses due to natural hazards were equal to the sum of all types of risk faced by a population in a given area. This is summarised in Equation 3 above.

The multi-hazard risk for a country required calculating an estimate of the probability of the occurrence and severity of each hazard, the number of persons affected by it, and the identification of the population's vulnerability and coping capacities. This is very ambitious and not achievable with present data constraints. However the aim is to provide an approach built on existing data that will be refined in subsequent runs of the DRI.

T.3 Choice of Indicators

T.3.1 Spatial and temporal scales

The DRI exercise was performed on a country-by-country basis for the 249 countries defined in the GEO reports.[6]

The socio-economic variables used in the analysis of risk needed to be available to cover the 21-year period under analysis. This period was from 1980 to 2000. The starting date was set at 1980 because access to information (especially on victims) was not considered reliable or comparable before this year. The variables introduced in Equation 2 were aggregate figures (sum or average) of the available data for that period, with the following major exceptions:

- Earthquake frequencies were calculated over a 36-year period, due to the longer return period of this type of disaster. The starting date for the first global coverage on earthquakes measurement is 1964.
- Cyclones frequencies were based on annual probabilities provided by the Carbon Dioxide Information Analysis Center (CDIAC).[7]
- HDI was available for the following years: 1980, 1985, 1990, 1995 and 2000. However, algorithms were applied for computation of every year between 1980 and 2000.
- Population by grid cell (for physical exposure calculations) was available for 1990 and 1995.
- The Corruption Perception Index (CPI) was available for 1995 to 2000.

T.3.2 Risk indicators

Risk can be expressed in different ways (for example by the number of people killed, percentage killed or percentage killed as compared to the exposed population). Each measure has advantages and inconveniences (see Table T.1 on the following page).

The DRI work used two indicators for each hazard type: the number of killed and killed per population. The third indicator is used to indicate relative vulnerability. Exposed populations to different hazards should not be compared as stated in the Report without standardisation.

T.3.3 Vulnerability indicators

Table T.2 (see following page) hows those socio-economic and environmental variables chosen to represent eight separate categories of vulnerability.

b. In the case of countries marginally affected by a hazard type, the risk was replaced by zero if the model could not be computed for this hazard.

TABLE T.1 ADVANTAGES AND DISADVANTAGES OF RESPECTIVE RISK INDICATORS

Indicators for risk	Advantages	Inconveniences
Number of killed	Each human being has the same 'weight.'	10,000 people killed split between 10 small countries does not appear in the same way as 10,000 killed in one country. Smaller countries are disadvantaged.
Killed/Population	Allows for comparisons between countries. Less populated countries have the same weight as more populated countries.	The 'weight' of each human being is not equal, e.g. one person killed in Honduras equals 160 killed in China.
Killed/Population exposed	Regional risk is highlighted, even though the population affected is a smaller portion of the total national population.	This may highlight local problems that are not of national significance and give the wrong priority for a selected country.

TABLE T.2 VULNERABILITY INDICATORS

Categories of Vulnerability	Indicators	Drought	Flood Earthquakes Cyclones	Source[c]
Economic	Gross Domestic Product per inhabitant at purchasing power parity	X	X	WB
	Human Poverty Index (HPI)	X		UNDP
	Total debt service (% of the exports of goods and services)		X	WB
	Inflation, food prices (annual %)		X	WB
	Unemployment, total (% of total labour force)		X	ILO
Type of economic activities	Arable land (in thousand hectares)		X	FAO
	% of arable land and permanent crops		X	FAO
	% of urban population		X	UNPOP
	% of agriculture's dependency for GDP	X		WB
	% of labour force in agricultural sector	X		FAO
Dependency and quality of the environment	Forests and woodland (in % of land area)		X	FAO
	Human-Induced Soil Degradation (GLASOD)	X	X	FAO/UNEP
Demography	Population growth		X	UNDESA
	Urban growth		X	GRID[d]
	Population density		X	GRID[e]
	Age dependency ratio		X	WB
Health and sanitation	% of people with access to improved water supply (total, urban, rural)	XXX		WHO/UNICEF
	Number of physicians (per 1,000 inhabitants)		X	WB
	Number of hospital beds		X	WB
	Life expectancy at birth for both sexes		X	UNDESA
	Under-five-years-old mortality rate	X		UNDESA
Early warning capacity	Number of radios (per 1,000 inhabitants)		X	WB
Education	Illiteracy rate		X	WB
Development	Human Development Index (HDI)	X	X	UNDP

Source: UNDP/UNEP

c. FAOSTAT, the database of the Food and Agriculture Organisation (FAO); GRID, the Global Resource Information Database of UNEP; WB, World Development Indicators of the World Bank; Human Development Report of UNDP; ILO, International Labour Office; UNDESA, the UN Dept. of Economic and Social Affairs/Population Division. Most of the data were reprocessed by the UNEP Global Environment Outlook Team. Figures are available at the GEO Data Portal (UNEP), http://geodata.grid.unep.ch
d. Calculated from UN Dept. of Economic and Social Affairs data.
e. Calculated from UNEP/GRID spatial modelling based on CIESIN population data.

TABLE T.3 DATA SOURCES FOR HAZARDS	
Hazard type	**Data source**
Earthquakes	Council of the National Seismic System (as of 2002), *Earthquake Catalog,* http://quake.geo.berkeley.edu/cnss/
Cyclones	Carbon Dioxide Information Analysis Centre (1991), *A Global Geographic Information System Data Base of Storm Occurrences and Other Climatic Phenomena Affecting Coastal Zones,* http://cdiac.esd.ornl.gov/
Floods	U.S. Geological Survey (1997), *HYDRO1k Elevation Derivative Database,* http://edcdaac.usgs.gov/gtopo30/hydro/
Droughts (physical drought)	IRI/Columbia University, National Centres for Environmental Prediction Climate Prediction Centre (as of 2002), CPC *Merged Analysis of Precipitation (CMAP),* monthly gridded precipitation, http://iridl.ldeo.columbia.edu/

TABLE T.4 DATA SOURCES FOR VICTIMS, POPULATION AND VULNERABILITY VARIABLES	
Theme	**Data source**
Victims (killed)	Université Catholique de Louvain (as of 2002), *EM-DAT: The OFDA/CRED International Disaster Database,* http://www.cred.be/ (for droughts, victims of famines were also included on a case by case basis by UNDP/BCPR)
Population (counts)	CIESIN, IFPRI, WRI (2000), *Gridded Population of the World (GPW), Version 2,* http://sedac.ciesin.org/plue/gpw/; UNEP, CGIAR, NCGIA (1996), *Human Population and Administrative Boundaries Database for Asia,* http://www.grid.unep.ch/data/grid/human.php
Vulnerability factors	
Human Development Index (HDI)	UNDP (2002), *Human Development Indicators,* http://www.undp.org/
Corruption Perceptions Index (CPI)	Transparency International (2001), *Global Corruption Report 2001,* http://www.transparency.org/
Soil degradation (% of area affected)	ISRIC, UNEP (1990), *Global Assessment of Human-Induced Soil Degradation (GLASOD),* http://www.grid.unep.ch/data/grid/gnv18.php
Other socio-economic variables	UNEP/GRID (as of 2002), *GEO-3 Data portal,* http://geodata.grid.unep.ch/ (data compiled from World Bank, World Resources Institute, FAO databases)

The list of factors to be considered for the analysis was set on the basis of the following criteria:

- *Relevance.* Select vulnerability factors (outputs orientated, resulting from the observed status of the population) not based on mitigation factors (inputs, action taken). For example, school enrollment rather than education budget.
- *Data quality and availability.* Data should cover the 1980-2000 period and most of the 249 countries and territories.

Examples of variables that were rejected for these two reasons were the percentage of persons affected by AIDS, the level of corruption and the number of hospital beds per inhabitant.

T.3.4 Data sources

Data sources ranged from universities and national scientific institutions to international data series collected by international organisations. Table T.3 presents the data sources used to obtain data on hazards.

Table T.4 presents the data sources used to obtain data on victims, population and vulnerability variables.

T.4 The Computation of Physical Exposure

T.4.1 General description

Two methods are available for calculating physical exposure. First, by multiplying hazard frequency by the population living in each exposed area. The frequencies of hazards were calculated for different strengths of event, and physical exposure was computed as in Equation 4.

EQUATION 4 COMPUTATION OF PHYSICAL EXPOSURE

EQ $$PhExp_{nat} = \sum F_i \cdot Pop_i$$

Where

$PhExp_{nat}$	is the physical exposure at national level
F_i	is the annual frequency of a specific magnitude event in one spatial unit
Pop_i	is the total population living in the spatial unit

A second method was used when data on the annual frequency of return of a specific magnitude event was not available. In this case (earthquake), physical exposure was computed by dividing the exposed population by the numbers of years when a particular event had taken place as shown in Equation 5.

EQUATION 5 PHYSICAL EXPOSURE CALCULATION WITHOUT FREQUENCY

EQ5 $$PhExp = \sum \frac{Pop_i}{Y_n}$$

Where

Pop_i is the total population living in a particular buffer, the radius of which from the epicentre varies according to the magnitude

Y_n is the length of time in years

PhExp is the total physical exposure of a country, in other words the sum of all physical exposure in this country

EQUATION 6 COMPUTATION OF CURRENT PHYSICAL EXPOSURE

EQ6 $$PhExp_i = \sum \frac{Pop_i}{Pop_{1995}} \bullet PhExp_{1995}$$

Where

$PhExp_i$ is the physical exposure of the current year

Pop_i is the population of the country at the current year

Pop_{1995} is the population of the country in 1995

$PhExp_{1995}$ is the physical exposure computed with population as in 1995

Once the area exposed to a hazard was computed — using UNEP/GRID-Geneva methods for earthquakes, floods and cyclones and using a method for drought from the International Research Institute for Climate Prediction (IRI) — then the exposed population was calculated for each exposed area. This number was then aggregated at the national level to come to a value for the number of exposed people over the last 21 years for each hazard type.

Depending on the type of hazard and the quality of data, different methods were applied to estimate the size of populations exposed to individual hazards. Population data was taken from CIESIN, IFPRI and WRI Gridded Population of the World (GPW, Version 2) at a resolution of 2.5'[f] (equivalent to 5 x 5 km at the equator). This was supplemented by the Human Population and Administrative Boundaries Database for Asia (UNEP) for Taiwan and CIESIN Global Population of the World Version 2 (country level data) for ex-Yugoslavia. These datasets reflect the estimated population distribution for 1995. Since population growth is sometimes very high in the 1980-2000 period, a correction factor using country totals was applied in order to estimate current physical exposures for each year as follows (see Equation 6).

Due to the resolution of the dataset, the population could not be extracted for some small islands. This has meant some small islands had to be left out of parts of the analysis. This is a topic for further research (see recommendations in the Conclusions of the Technical Annex).

The main challenge lay in the evaluation of areas exposed to particular hazard frequency and intensity. At the global scale, data was not complete. Expert opinion was used to review the process of building datasets. Of the four hazards studied, only in the case of floods was it necessary to design a global dataset. This was constructed by linking CRED information with USGS watersheds. Drought maps were provided by IRI. For the other hazards, independent global datasets had already been updated, compiled or modelled by UNEP/GRID-Geneva and were used to extract population. The Mollweide equal-area projection was used when calculations of areas were needed.

T.4.2 The case of earthquake

A choice was made to produce seismic hazard zones using the seismic catalogue of the Council of the National Seismic System. The earthquakes records of the last 21 years (1980-2000) were grouped in five magnitude classes using a buffer with a radius from the epicentre that varied according to the magnitude class (see Table T.5).

The values in Table T.5 show estimated ground-motion duration for specific acceleration and frequency ranges, according to magnitude and distance from the epicentre.[8] Numbers in bold in Table T.5 show the duration for a particular acceleration and frequency range between the first and last acceleration excursions on the record greater than a given amplitude level (for example, 0.05 g).[9]

f. GPW2 was preferred to the ONRL Landscan population dataset despite its five times lower spatial resolution (2.5' against 30") because the original information on administrative boundaries and population counts is almost two times more precise (127,093 administrative units against 69,350 units). Furthermore, the Landscan dataset is the result of a complex model which is not explained thoroughly and which is based, among other variables, on environmental data (land-cover). That makes it difficult to use for further comparison with environmental factors (circularity).

TABLE T.5 LIMITS OF THE RADIUS FOR EARTHQUAKES HAZARD

Distance (km)	Magnitude						
	5.5	6.0	6.5	7.0	7.5	8.0	8.5
10	8	12	19	26	31	34	35
25	4	9	15	24	28	30	32
50	2	3	10	22	26	28	29
75	1	1	5	10	14	16	17
100	0	0	1	4	5	6	7
125	0	0	1	2	2	3	3
150	0	0	0	1	2	2	3
175	0	0	0	0	1	2	2
200	0	0	0	0	0	1	2

Source: [Bolt et al. 1975] Acceleration > 0.05 g = ~ 0,49 m/s2, frequency > 2 Hz

According to these figures, a specific buffer distance was defined for each class of magnitude to limit the area affected by ground motions: 75 km for Magnitude ≤ 6.2, 125 km for M = 6.3 – 6.7, 150 km for M = 6.8 – 7.2, 175 km for M = 7.3 – 7.7, 200 km for M ≥ 7.8. This approach did not take into account local conditions, for instance soil or geo-tectonic characteristics.

Assuming the limitations inherent in a mortality-based conceptual model, there were three key challenges to calculating the earthquake risk index.

The first and most difficult challenge was the necessity to use a restricted time-frame for analysis of risk (1980-2001). Twenty years is a short time-span to analyze the occurrence of geological phenomena such as earthquakes, which are low frequency/high impact events. For this reason, risks are overestimated by the model for some countries and underestimated for others. Armenia provides an example of a high-impact single earthquake in a small-sized country (29,000 square kilometres), with a high population density (117 per square kilometre). The earthquake that affected this former Soviet Republic in 1998 killed 25,000 people, left 514,000 people homeless and prompted the evacuation of almost 200,000 people. The high losses recorded in this event appear to exaggerate Armenia's long-term calculated risk value, in comparison with countries known to be at risk but where no event took place during the time period used to calculate the risk model. An example of this is the Algerian earthquake in 2003, which is later than the period used in the DRI exercise. In order to partly overcome such limitation, frequency was derived using data from 1964-2000 in order to take advantage of the time-span available globally.

Secondly, in the delimitation of areas at risk from individual earthquake zones, it was not possible to consider intervening factors (such as soil types and geology) in the transmission of earthquake energy. In explaining the ground motions of earthquakes and therefore the severity of impact, soil conditions play a major role. Inclusion of this data would have allowed for a more accurate delimitation of areas and thus populations exposed to earthquake risks of various magnitudes and intensities. While values for peak ground acceleration were available from the Global Seismic Hazard Assessment Programme, they did not allow for the calculation of frequencies. Consequently, the analysis was based solely on magnitude values that were taken from the Council of the National Seismic System (CNSS).

FIGURE T.3 POPULATION, INTENSITY AND PHYSICAL EXPOSURE FOR EARTHQUAKES

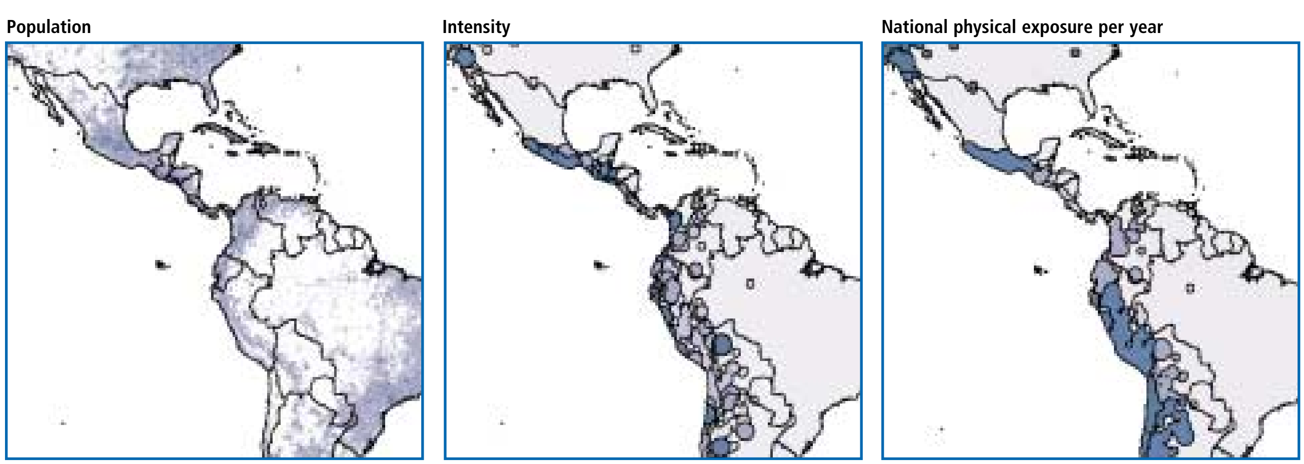


TABLE T.6 WIND SPEEDS AND APPELLATIONS

Wind speeds	Name of the phenomenon
≥ 17 m/s	Tropical storms
≥ 33 m/s	Hurricanes, typhoons, tropical cyclones, severe cyclonic storms (depending on location)
≥ 65 m/s	Super-typhoons

A third and more generic challenge for the risk model was the lack of casualty and death data and a lack of underlying socio-economic and environmental data for some countries. This is particularly problematic for mapping global earthquake risk because some gaps in national level data led to the exclusion of some countries — known to be at particularly high risk from earthquakes — from the calculation of the vulnerability indicators. This was the case for Afghanistan, Sudan, Tajikistan and Guinea. Future improvements in statistical records will enhance the scope of future assessments.

T.4.3 The case of tropical cyclone

The data used to map tropical cyclone hazard areas were produced by the Carbon Dioxide Information Analysis Centre.[10] The spatial unit is a 5 x 5 decimal degrees cell. Return probabilities were based on tropical cyclone activity over a specific record period. Exceptions were made for several estimated values attributed to areas that may present occasional activity, but where no tropical cyclones were observed during the record period.

The Saffir-Simpson tropical cyclones classification is based on the maximum sustained surface wind. Systems with winds of less than 17 m/s are called Tropical Depressions. If the wind reaches speeds of at least 17 m/s, the system is called a Tropical Storm. If the wind speed is equal to or greater than 33 m/s, the system is named, depending on its location:[g] Hurricane, Typhoon, Severe Tropical Cyclone, Severe Cyclonic Storm or Tropical Cyclone. Systems with winds reaching speeds of 65 m/s or more are called Super-typhoons.[11]

The CDIAC provided the probability of occurrence for these three types of events. The average frequency (per year) was computed using Equation 7.

To obtain physical exposure, a frequency per year was derived for each cell. Cells were divided to follow country borders, then population was extracted and multiplied by frequency in order to obtain the average yearly physical exposure for each cell. This physical exposure was then summed by country for the three types of cyclones.

Physical exposure to tropical cyclones of each magnitude was calculated for each country using Equation 5.

There is room for improving the human exposure calculation by more accurate delimitation of exposed population zones for tropical cyclone tracks. Even though accurate zoning was possible for many tropical cyclone-prone countries, data on tracks, central pressure and sustained winds were not available for some heavily populated and high-risk countries, such as India, Bangladesh and Pakistan. While these data exist they were not accessible.

T.4.4 The case of flood

The only global database on floods that was identified was the Dartmouth Flood Observatory, but this database did not cover the time period under study. Due to the lack of information on the duration and severity of floods, only one class of intensity was made. Using the EM-DAT database, a geo-reference of each recorded flood was produced and the watershed related to each flood event was identified. Watersheds affected were mapped for the period 1980-2000. A frequency was derived for each watershed by dividing the total number of events by 21 years. The watersheds were then split to follow country borders. Next, population was extracted and multiplied by the event frequency. The average yearly physical exposure was then summed at a country level using Equation 3.

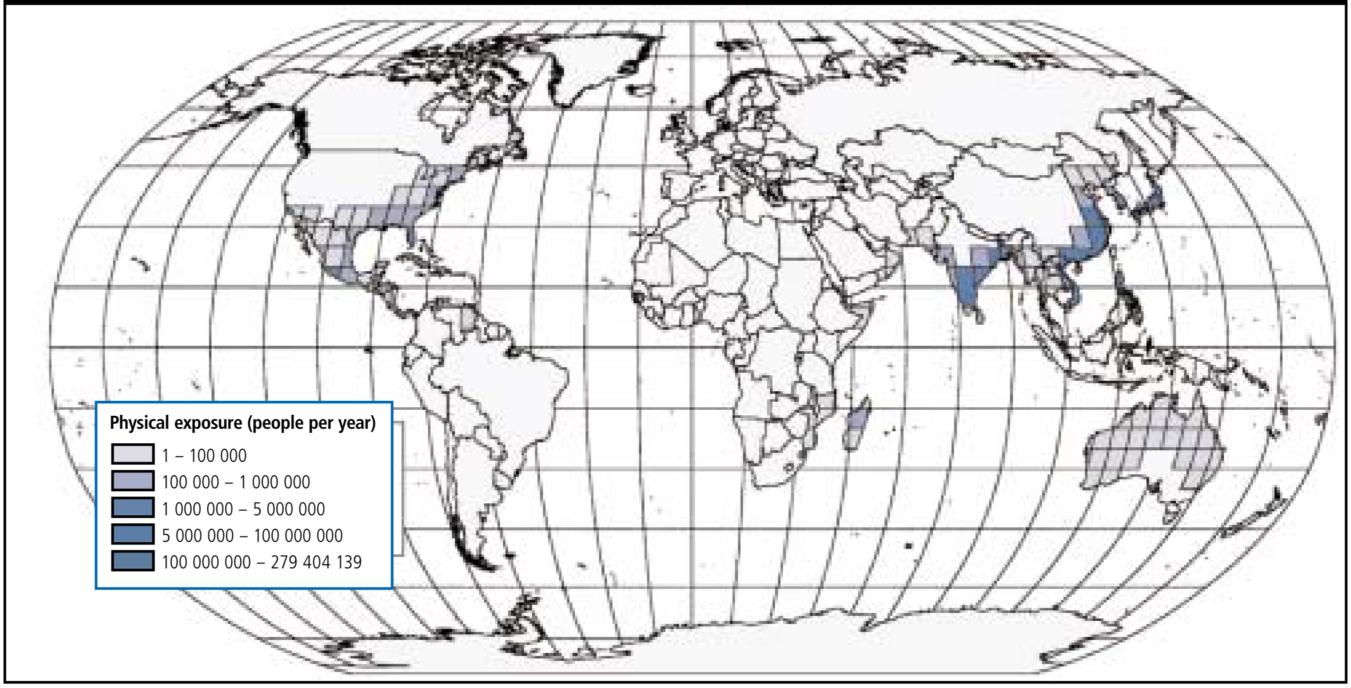
EQUATION 7 FROM PROBABILITY TO ANNUAL FREQUENCY FOR CYCLONES

EQ7 $$E(x) = \lambda = -\ln(1 - P(x \geq 1))$$

Where
E(x) is the statistical expectation, i.e. the average number of events per year = λ
P(x) is the probability of occurrence

g. Hurricane: North Atlantic Ocean, Northeast Pacific Ocean east of the dateline, or the South Pacific Ocean east of 160E; Typhoon: Northwest Pacific Ocean west of the dateline; Severe tropical cyclone: Southwest Pacific Ocean west of 160E and Southeast Indian Ocean east of 90E; Severe cyclonic storm: North Indian Ocean; Tropical cyclone: Southwest Indian Ocean; Source: NOAA/AOML, FAQ: *Hurricanes, Typhoons and Tropical Cyclones,* http://www.aoml.noaa.gov/hrd/tcfaq/tcfaqA.html#A1

FIGURE T.4 AN EXAMPLE OF PHYSICAL EXPOSURE FOR TROPICAL CYCLONES



Source: Carbon Dioxide Information Analysis Centre: A Global Geographic Information System Database of Storm Occurrences and Other Climactic Phenomena Affecting Coastal Zones; CIESIN, IFPRI, WRI: Gridded Population of the World (GPW), Version 2 (population); Compilation and computation by UNEP/GRID-Geneva

Assuming the limitations inherent in a mortality-based conceptual model there were two key challenges to measuring flood risk.

First, there remains a need for refining the calculation of human exposure and vulnerability to floods in the formulation of the DRI. The use of watersheds affected by floods to delimit hazard exaggerates the extent of flood-prone areas, subsequently exaggerating human exposure and diminishing proxies of vulnerability.

Second, in the absence of historical flood event data, annual probabilities of floods should be based on hydrological models rather than being inferred from the flood entries in the EM-DAT database.

T.4.5 The case of drought

Identification of drought

The data used in this analysis consisted of gridded monthly precipitation data for the globe for the period 1979-2001. This dataset was based on a blend of surface

FIGURE T.5 POPULATION, FREQUENCY AND PHYSICAL EXPOSURE FOR FLOODS

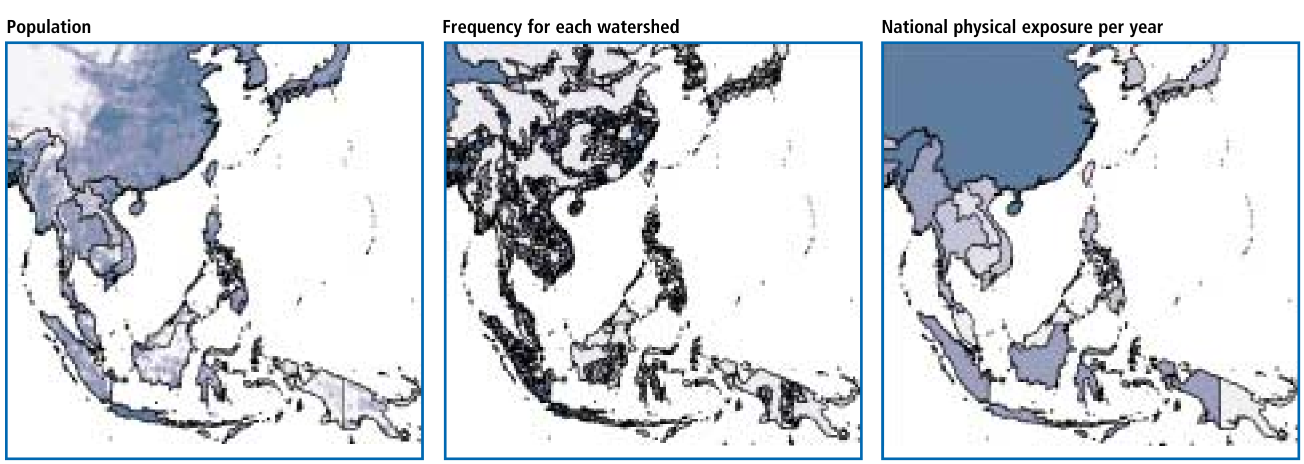


TABLE T.7 DEFINITION OF DROUGHT

Duration	Severity
3 months	90% of median precipitation 1979-2001 (-10%)
3 months	75% of median precipitation 1979-2001 (-25%)
3 months	50% of median precipitation 1979-2001 (-50%)
6 months	90% of median precipitation 1979-2001 (-10%)
6 months	75% of median precipitation 1979-2001 (-25%)
6 months	50% of median precipitation 1979-2001 (-50%)

station observations and precipitation estimates drawn from satellite observations. The first step in assessing exposure to meteorological drought was to compute, for each calendar month, the median precipitation for all grid points between the latitudes of 60S and 70N over the base period 1979-2001 (the 23-year period for which the data was available). Next, for each grid-point, the percent of the long-term median precipitation was computed for every month during the period January 1980 to December 2000. For a given month, grid-points with a long-term median precipitation of less than 0.25 mm/day were excluded from the analysis. Such low median precipitation amounts can occur either during the 'dry season' at a given location or in desert regions. In both cases our definition of drought does not apply.

A meteorological drought event was defined as having occurred when the percent of median precipitation was at or below a given threshold for at least three consecutive months. The different thresholds considered were 50 percent, 75 percent and 90 percent of the long-term median precipitation, with the lowest percentage indicative of the most severe drought according to this method. The total number of events during the period 1980-2000 were thus determined for each grid-point and the results plotted on global maps.

Computation of physical exposure

Using the IRI/Columbia University dataset, physical exposure was estimated by multiplying the frequency of hazard by the population living in an exposed area. The events were identified using different measurements, based on severity and duration as described in Table T.7. For each of the following six definitions, the frequency was then obtained by dividing the number of events by 21 years, thus providing an average frequency of events-per-year.

Physical exposure was computed, as in Equation 5, for each drought definition. The statistical analysis selected the best fit. This was achieved with droughts of three months duration and 50 percent decrease in precipitation.

EQUATION 8 ESTIMATE OF KILLED

EQ8 $$K = C \cdot (PhExp)^{\alpha} \cdot V_1^{\alpha_1} \cdot V_2^{\alpha_2} \dots \cdot V_p^{\alpha_p}$$

Where

K is the number of persons killed by a certain type of hazard

C is the multiplicative constant.

PhExp is the physical exposure: population living in exposed areas multiplied by the frequency of occurrence of the hazard

V_i are the socio-economic parameters

α_i is the exponent of Vi , which can be negative (for ratio)

T.5 Statistical analysis: Methods and results

T.5.1 Defining a multiplicative model

The statistical analysis is based on two major hypotheses. First, that risk can be understood in terms of the number of victims of past hazardous events. Secondly, that the equation of risk follows a multiplicative model as in Equation 8.

Using logarithmic properties, the equation was re-formulated as in Equation 9. This equation creates a linear relationship between logarithmic sets of values. This allows significant socio-economic parameters V_i (with transformations when appropriate) and exponents α_i to be determined using linear regressions.

T.5.2 Detailed process

Data on victims

Numbers of killed were derived from the EM-DAT database and computed as the average number killed per year during the 1980-2000 period.

EQUATION 9 LOGARITHM PROPERTIES

EQ9 $$ln(K) = ln(C) + \alpha(PhExp) + \alpha_1 ln(V_1) + \alpha_2 ln(V_2) + \dots \alpha_p ln(V_p)$$

Filtering the data

The statistical models for each disaster type were based on subsets of countries, from which were excluded:

- Countries with no physical exposure or any victims reported (zero or null values).
- Countries where it was not possible to confirm data on physical exposure (e.g. the case of Kazakhstan for floods) or socio-economic factors.
- Countries with low physical exposure (less than 2 percent of the total population) because socio-economic variables were collected at a national scale. The exposed population needs to be of some significance at a national level to reflect a relationship in the model.
- Countries without all the selected socio-economic variables.
- Eccentric values, when exceptional events or other factors would clearly show abnormal levels of victims, such as Hurricane Mitch in Nicaragua and Honduras or droughts in Sudan and Mozambique.

Transformation of socio-economic variables

For statistical analysis the socio-economic variables being tested had to be converted into 21-year averages and then transformed into a logarithm value. For some of the variables, the logarithm was computed directly. For those expressed as a percentage, a transformation was applied in order that all variables would range between -∞ and +∞. For others, no logarithmic transformation was needed. For instance, 'population growth' already behaves in a cumulative way and could be put directly into the calculation.

EQUATION 10 TRANSFORMATION FOR VARIABLES RANGING BETWEEN 0 AND 1

EQ10 $$V_i' = \frac{V_i}{(1 - V_i)}$$

Where
V_i' is the transformed variable (ranging from −∞ to +∞)
V_i is the socio-economic variable (ranging from 0 to 1)

Choice between variables

One important condition when computing regressions is that the variables included in the model should be independent, i.e. the correlation between two sets of variables is low. This is clearly not the case with HDI and GDPcap purchasing power parity (further referred to as GDPcap), which are highly correlated. GDPcap was used more than HDI because HDI was not available for several countries. In order to keep the sample as complete as possible, a choice between available variables was made choosing variable datasets that were as independent from each other as possible. This choice was performed by the use of both matrix-plot and correlation-matrix (using low correlation, hence low p-value, as the selection criteria).

The stepwise approach

For each type of hazard, numerous stepwise (back and forth steps) linear regressions were performed in order to identify significant socio-economic variables. The validation of each regression result was carried out using R2, variance analysis and detailed residual analysis.

Once the model was derived, the link between modelled estimated-killed and number-of-killed observed from EM-DAT was provided by both graph plots and computation of Pearson correlation coefficients.

If one can intuitively understand that physical exposure is positively related with the number of victims, and that GDPcap is inversely related with the number of victims (the lower the GDP, the higher the number of victims), this is less obvious for other variables such as the percentage of arable land. This method multiple logarithmic regression allows the estimation of the α_i coefficients. Their signs provided information to show if the variables were in a numerator or denominator position and hence the positive or inverse relationship between the variable and modelled deaths.

This model allowed the identification of parameters leading to higher/lower risk, but should not be used as a predictive model. Small differences in the logarithm scale can induce large ones in the modelled number of deaths.

The results following this method were surprisingly high and relevant, especially considering the independence of the data sources and the coarse resolution of the data at the global scale.

T.5.3 Mapping Risk

A judgement was made between the different risk indicators (i.e. killed, killed per million inhabitant, killed per population exposed).

T.5.4 Earthquake

Statistical model

The multiple regression was based on 48 countries. The best-fit regression line followed Equation 11 (see following page).

EQUATION 11 MULTIPLE LOGARITHMIC REGRESSION MODEL FOR EARTHQUAKES

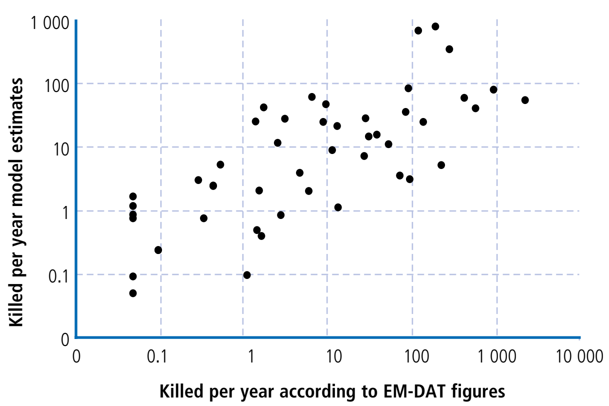
EQ11 $$\ln(K) = 1.26\ln(PhExp) + 12.27 \bullet U_g - 16.22$$

Where
K is the number of killed from earthquakes
PhExp is the physical exposure to earthquakes
U_g is the rate of urban growth (rates do not request transformation as it is already a cumulative value)

TABLE T.8 EXPONENT AND P-VALUE FOR EARTHQUAKE MULTIPLE REGRESSION

48 countries	B	p-value[h]
Intercept	–16.22	0.000000
PhExp	1.26	0.000000
U_g	12.27	0.047686
R= 0.75, R^2= 0.56, adjusted R^2= 0.54		

FIGURE T.6 SCATTER PLOT OF THE OBSERVED NUMBER OF PEOPLE KILLED BY EARTHQUAKES (EM-DAT) AND THE MODEL PREDICTIONS



Source: The EM-DAT OFDA/CRED International Disaster Database and UNEP/GRID-Geneva

The variables retained by the regression include physical exposure and the rate of urban growth. Explained variance is smaller than for flood or cyclones (R^2=0.544), however considering the small length of time taken into account (21 years as compared to the long return period of earthquakes), the analysis delineates a reasonably good relation. Physical exposure is of similar relevance than for previous cases, relevant p-value. Urban growth is also highly negatively correlated with GDP and HDI. Thus, a similar correlation (but slightly inferior) could have been derived using HDI or GDP.

T.5.5 Tropical cyclone

Statistical model

The multiple regression was based on 32 countries and the best-fit regression line followed Equation 12.

The plot delineates a clear linear distribution of the data as seen in Figure T.7.

The parameters highlighted show that physical exposure, HDI and the percentage of arable land were associated with cyclone hazards.

The GDPcap is strongly correlated with the HDI or negatively with the percentage of urban growth. In most of the cases, the variable GDPcap could be replaced by HDI as explained previously. However, these results show with confidence that poor countries and countries with low human development index rank are more vulnerable to cyclones.

With a considerable part of variance explained by the regression (R^2 = 0.863) and a high degree of confidence in the selected variables (very small p-value) over a sample of 32 countries, the model achieved is solid.

In the model, the consequences of Hurricane Mitch could easily be depicted. Indeed, Honduras and Nicaragua were far off the regression line (significantly underestimated). This is explained by the high impact of Mitch compared to other hurricanes. The extreme values given by this event led to Honduras and Nicaragua being rejected from the model.

EQUATION 12 MULTIPLE LOGARITHMIC REGRESSION MODEL FOR TROPICAL CYCLONE

EQ12 $$\ln(K) = 0.63\ln(PhExp) + 0.66\ln(\overline{Pal}) - 2.03\ln(\overline{HDI}) - 15.86$$

Where
K is the number of killed from cyclones
PhExp is the physical exposure to cyclones
$\overline{Pal}$ is the transformed value of percentage of arable land
$\overline{HDI}$ is the transformed value of the Human Development Index

h. In broad terms, a p-value smaller than 0.05 shows the significance of the selected indicator, however this should not be used blindly.

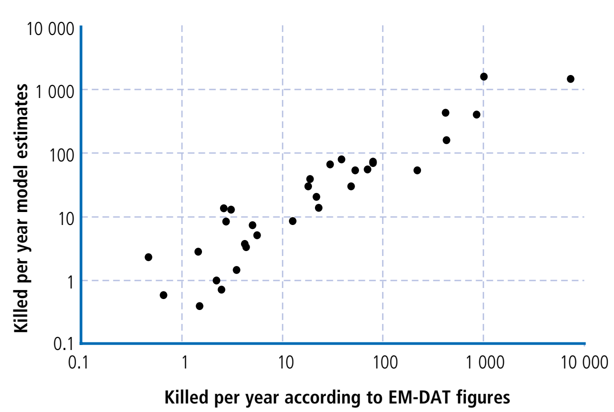
TABLE T.9 EXPONENT AND P-VALUE FOR CYCLONES MULTIPLE REGRESSION

21 countries	B	p-value[i]
Intercept	−15.86	0.00000
ln(PhExp)	0.63	0.00000
$\ln(\overline{Pal})$	0.66	0.00013
$\ln(\overline{HDI})$	−2.03	0.00095
$R = 0.93$, $R^2 = 0.86$, adjusted $R^2 = 0.85$		

TABLE T.10 EXPONENT AND P-VALUE FOR FLOOD INDICATORS

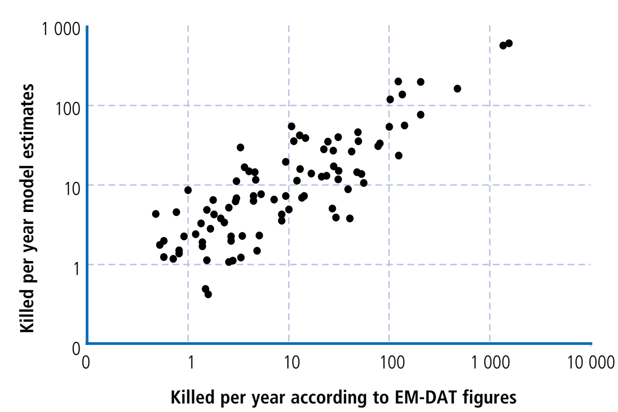
90 countries	B	p-value[i]
Intercept	−5.22	0.00000
ln(PhExp)	0.78	0.00000
$\ln(GDP_{cap})$	−0.45	0.00002
ln(Density)	−0.15	0.00321
$R = 0.84$, $R^2 = 0.70$, adjusted $R^2 = 0.69$		

FIGURE T.7 SCATTER PLOT OF THE OBSERVED NUMBER OF PEOPLE KILLED BY TROPICAL CYCLONE (EM-DAT) AND THE MODEL PREDICTIONS



Source: The EM-DAT OFDA/CRED International Disaster Database and UNEP/GRID-Geneva

FIGURE T.8 SCATTER PLOT OF THE OBSERVED NUMBER OF PEOPLE KILLED BY FLOOD (EM-DAT) AND THE MODEL PREDICTIONS



Source: The EM-DAT OFDA/CRED International Disaster Database and UNEP/GRID-Geneva

T.5.6 Flood

Statistical model

The multiple regression was based on 90 countries. The best-fit regression line followed Equation 13.

Due to space constraints, only a selection of countries was included in the above scatter plot. A comprehensive list of countries affected by floods is provided below:

Albania, Algeria, Angola, Argentina, Australia, Austria, Azerbaijan, Bangladesh, Benin, Bhutan, Bolivia, Botswana, Brazil, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Chad, Chile, China, Colombia, Costa Rica, Côte d'Ivoire, Czech Republic, Dominican Republic, Ecuador, Egypt, El Salvador, Ethiopia, Fiji, France, Gambia, Georgia, Germany, Ghana, Greece, Guatemala, Haiti, Honduras, India, Indonesia, Iran (Islamic Republic of), Israel, Italy, Jamaica, Japan, Jordan, Kenya, Lao People's Democratic Republic, Malawi, Malaysia, Mali, Mexico, Republic of Morocco, Mozambique, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Republic of Korea, Republic of Moldova, Romania, Russian Federation, Rwanda, Saudi Arabia, Sierra

EQUATION 13 MULTIPLE LOGARITHMIC REGRESSION MODEL FOR FLOOD

EQ13 $$\ln(K) = 0.78\ln(PhExp) + 0.45\ln(GDP_{cap}) - 0.15\ln(D) - 5.22$$

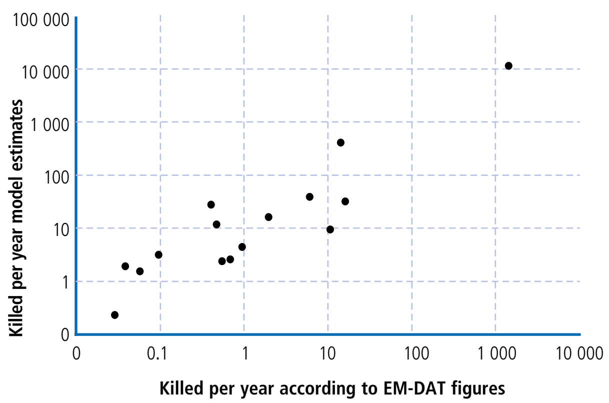
Where
K is the number of killed from floods
PhExp is the physical exposure to floods
GDP_{cap} is the normalised Gross Domestic Product per capita (purchasing power parity)
D is the local population density (i.e. the population affected divided by the area affected)

i. In broad terms, a p-value smaller than 0.05 shows the significance of the selected indicator, however this should not be used blindly.

TABLE T.11 EXPONENT AND P-VALUE FOR DROUGHT MULTIPLE REGRESSION

Predictor	Coef	SE Coef	T	p-value[j]
Constant	14,390	3,411	4.22	0.001
PhExp3_5	1.2622	0.2268	5.57	0.000
$WAT_{TOT}^{(1n)}$	-7,578	1,077	-7.03	0.000
S = 1,345, R-Sq = 0.812, R-Sq(adj) = 0.78				

FIGURE T.9 SCATTER PLOT OF THE OBSERVED NUMBER OF PEOPLE KILLED BY DROUGHT (EM-DAT) AND THE MODEL PREDICTIONS



Source: The EM-DAT OFDA/CRED International Disaster Database and UNEP/GRID-Geneva

Leone, Slovakia, South Africa, Spain, Sri Lanka, Thailand, Tunisia, Turkey, Uganda, Ukraine, United Kingdom of Great Britain and Northern Ireland, United Republic of Tanzania, United States of America, Viet Nam, Yemen and Zimbabwe.

The variables selected by the statistical analysis are physical exposure, GDP_{cap} and local density of population. GDP_{cap} being highly correlated with HDI, this later could have been chosen as well. The GDP_{cap} was chosen due to slightly better correlation between the model and the observed killed, as well as because of lower p–value. Regression analysis supposes the introduction of non-correlated parameters, thus preventing the use of all these variables.

The part of explained variance ($R^2 = 0.70$) associated with significant p–value (between 10^{-23} and $2{\cdot}10^{-3}$) on 90 countries is confirming a solid confidence in the selection of the variables (see Table T. 10 on the previous page).

T.5.7 Drought

Statistical model

The regression analysis was performed using the six different exposure datasets derived from IRI drought maps. In general, the models were based on three-month thresholds to give better results. The dataset based on a drought threshold set at three months, at 50 percent below the median precipitation between 1979-2001, was finally selected as the exposure data.

The multiple regression was based on 15 countries. The best-fit regression line followed Equation 14.

Rejected countries: Swaziland and Somalia (WAT_{TOT} value inexistent), North Korea (reported WAT_{TOT} of 100 percent is highly doubtful), Sudan and Mozambique (eccentric values, suggesting other explanation for deaths).

The small p–values observed suggest a relevant selection of the indicators among the list of available datasets. It is to be noted that the high coefficient for WAT_{TOT} (−7.578) denotes a strong sensitivity to the quality of the data. This implies that even a change of 1 percent in total access to water would induce significant change in the results. This would be especially so for small values where small changes have bigger influence in proportion.

The model could not be used for predictive purposes. Inconsistencies were found in the data that require further verification.

EQUATION 14 MULTIPLE LOGARITHMIC REGRESSION MODEL FOR DROUGHT

EQ14 $$\ln(K) = 1.26\ln(PhExp3\_50) - 7.58\ln(WAT_{TOT}) + 14.4$$

Where
K is the number of killed from droughts
PhExp3_50 is the number of people exposed per year to droughts. A drought is defined as a period of at least three months less or equal to 50 percent of the average precipitation level (IRI, CIESIN/IFPRI/WRI)
WAT_{TOT} is the percentage of population with access to improved water supply (WHO/UNICEF)

j. In broad terms, a p-value smaller than 0.05 shows the significance of the selected indicator, however this should not be used blindly.

The variables associated with disaster risk through statistical analysis were physical exposure and the percentage of population with access to improved water supply. A strong correlation was established (R^2 = 0.81) indicating the solidity of the method as well as the reliability of these datasets for such a scale of analysis.

Figure T.9 shows the distribution (on a logarithmic scale) of expected deaths from drought and as predicted from the model. A clear regression can be drawn. It should be noted that if Ethiopia were to be excluded, the correlation would fall to (R^2 = 0.6). However, the offset and the slope of the regression line do not change significantly, reinforcing the robustness of the model.

As far as 1.26 is close to 1, the number of killed people grows proportionally to physical exposure. Also, the number of killed people decreases as a percentage of population when improved water supply grows. This latter variable should be seen as an indicator of the level of development of the country, as it was correlated to other development variables, such as the under-five mortality rate (Pearson correlation r = –0.64) and Human Development Index (r = 0.65).

Some countries with large physical exposure did not report any deaths to drought (United States of America, Viet Nam, Nigeria, Mexico, Bangladesh, Iran, Iraq, Colombia, Thailand, Sri Lanka, Jordan, Ecuador). This could be for a number of reasons. Either the vulnerability is null or extremely low, e.g. USA and Australia, or the number of reported killed from food insecurity is placed under conflict in EM-DAT, e.g. Iraq and Angola. For other countries, further inquiry might be necessary.

T.6 Multiple Risk Integration

So far, the precision and quality of the data as well as the sensitivity of the model do not allow the ranking of countries for disaster risk.

T.6.1 Methods

How to compare countries and disasters

A multiple-hazard risk model was made by adding expected deaths from each hazard type for every country. In order to reduce the number of countries with no data that would have to be excluded from the model, a value of 'no data' for countries without significant exposure was replaced by zero risk of deaths.

Countries were considered as not affected if the two following conditions were met: a physical exposure smaller than 2 percent of the national population AND an affected population smaller than 1,000 per year.

Some 39 countries were excluded from the analysis. Despite this, it is known that each was exposed to some level of hazard and 37 countries with recorded disaster deaths were in EM-DAT. This list of countries identifies places where improvement in data collection is needed to allow their integration in future work. Reasons that individual countries were excluded were: countries marginally affected by a specific hazard, countries affected but without data; and countries where the distribution of risk could not be explained by the model (for example, for drought in Sudan, where food insecurity and famine is more an outcome of armed conflict than of meteorological drought as defined in the model).

Once the countries to be included in the model were identified, a Boolean process was run to allocate one of five statistically defined categories of multi-hazard risk to each country. Figure T.10 illustrates the different steps taken to incorporate values into a multiple-risk index. Once this process had been completed, three different products were available:

- A table of values for the countries that include the data for relevant hazards or countries without data but marginally affected (210 countries).
- A list of countries with missing data (countries with reported deaths but without appropriate data).
- A list of countries where the model could not be applied (indicators do not capture the situation in these countries, case of countries not explained by the model, or rejected during the analysis because the indicators are not relevant to the situation).

Multiple risk computation

Multiple risk was computed using the succession of formulae as described in Equation 15 (see following page).

Between each addition, the whole process described in Figure T.10 (see following page) needed to be run in order to identify those countries where a value represented by zero needed, either to be replaced by a value calculated from the selected hazard model, or if not, the country was placed in the 'not-relevant' or 'no data' lists (see below).

EQUATION 15 COMPUTATION OF MULTIPLE RISK BY SUMMING CALCULATED DEATHS AS MODELLED FOR RISK FOR CYCLONE, FLOOD, EARTHQUAKE AND DROUGHT

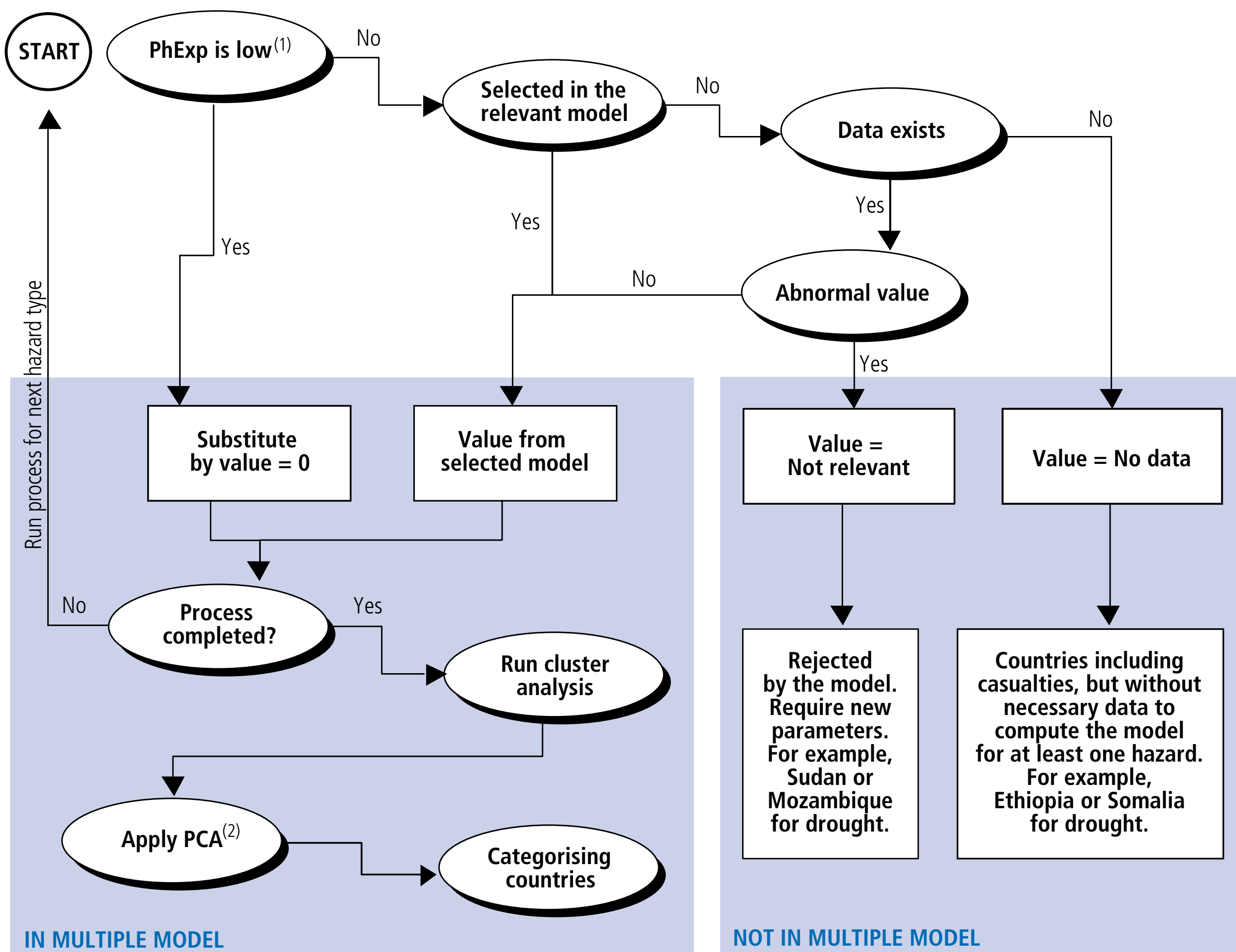
EQ15

$$K_{cyclones}\,(PhExp^{0.63}_{cyclones} \bullet \overline{Pal}^{0.66} \bullet \overline{HDI}^{-2.03} \bullet e^{-15.86}) + K_{floods}\,(PhExp^{0.78}_{floods} \bullet GDP^{-0.45}_{cap} \bullet D^{-0.15} \bullet e^{-5.22}) +$$
$$K_{earthquakes}\,(PhExp^{1.26}_{earthquakes} \bullet U_g^{12.27} \bullet e^{-16.27}) + K_{droughts}\,(PhExp3\_50^{1.26} \bullet WAT^{-7.58}_{TOT} \bullet e^{14.4})$$

Where

e	is the Euler constant (=2.718…)
PhExp	is the physical exposure of selected hazard
HDI	is the Human Development Index
GDP_{cap}	is the Gross Domestic Product per capita at purchasing power parity
D	is the local density (density of population in the flooded area)
U_g	is the Urban growth (computed over three-year period)
WAT_{TOT}	is the access to safe drinking water

FIGURE T.10 MULTIPLE RISK INTEGRATION



(1) Physical exposure is considered as marginal if smaller than 1,000 per year
(2) PCA: Principal Component Analysis, used to combine killed per year and killed per population in one component.

In order to examine the fit between model multi-hazard risk and recorded deaths, data from both sources were categorised into five country risk classes. A cluster analysis minimising the intra-class distance and maximising the inter-classes (K-means clustering method) was performed. This meant that a purely statistical process had been used to identify severities of risk from the model and deaths as recorded by EM-DAT.

In order to take both risk indicators (killed and killed per inhabitant) into account, a Principal Component Analysis (PCA) was performed to combine the two. Then a distinction was made between countries smaller than 30,000 km squared and with population density higher than 100 inhabitants per km squared.

T.6.2 Results

Modelled countries without reported deaths
The multi-hazard DRI was computed for 210 countries. This includes 14 countries where no recorded deaths were reported in the last two decades from EM-DAT: Barbados, Croatia, Eritrea, Gabon, Guyana, Iceland, Luxembourg, Namibia, Slovenia, Sweden, Syrian Arab Republic, The former Yugoslav Republic of Macedonia, Turkmenistan and Zambia.

No data, abnormal values and specific cases
Through the Principal Component Analysis transformation, inferior and superior thresholds were identified. This was performed on both observed and modelled deaths. For 14 countries, a value was calculated in the multi-hazard risk model even though no deaths had been recorded by EM-DAT in the 1980-2000 period. On the other hand, 37 countries where deaths were recorded could not be modelled, either because of a lack of data or because they did not fit with the model assumptions. These countries were: Afghanistan, Azerbaijan, Cuba, Democratic People's Republic of Korea, Democratic Republic of the Congo, Djibouti, Dominica, France, Greece, Liberia, Malaysia, Montserrat, Myanmar, New Caledonia, Portugal, Solomon Islands, Somalia, Spain, Sudan, Swaziland, Taiwan, Tajikistan, Vanuatu, Yugoslavia, Antigua and Barbuda, Armenia, Guadeloupe, Guam, Israel, Martinique, Micronesia (Federated States of), Netherlands Antilles, Puerto Rico, Reunion, Saint Kitts and Nevis, Saint Lucia, United States Virgin Islands.

Countries absent of both EM-DAT and Model
Two countries were absent from both EM-DAT and the model: Anguilla (a dependency of the United Kingdom) and Bosnia-Herzegovina.

EM-DAT-DRI multi-hazard risk comparison results
The results of the comparison of modelled and EM-DAT multi-hazard deaths are presented and discussed in Chapter 2. For more information, including country specific variables, researchers are encouraged to visit the Report website.

T.7 Technical Conclusions and Recommendations

T.7.1 The DRI – A work in progress

The DRI is a statistically robust tool
The results generated by the DRI method were statistically robust with a high level of confidence. This is especially the case considering the independence of the data sources and the coarse resolution of the data available at the global scale. The statistically strong links — both between observed and modeled deaths and between socio-economic variables associated with human vulnerability and levels of risk — that were found in the DRI study are not often found in similar studies that analyse geophysical datasets and socio-economic data. The model has succeeded in opening the great potential for future national level disaster risk assessments. It provides the first, solid statistical base for understanding and comparing countries' disaster risk and human vulnerability.

The DRI is not a predictive model
This is partly a function of a lack of precision in the available data. But it also shows the influence of local context. The risk maps provided in this research allow a comparison of relative risk between countries, but cannot be used to depict actual risk for any one country. Sub-national risk analysis would be required to inform development and land-use planning at the national level.

How to link extreme and everyday risk?
Extraordinary events by their very nature do not follow the normal trend. Hurricane Mitch in 1998, the rains causing landslides in Venezuela in 1999 or the 1988 earthquake in Armenia were off the regression line. This is due to the abnormal intensity of such events. These events are (hopefully) too rare to be usefully included in a two-decade period of study. Incorporating this level of intensity can only be done on an event-per-event approach.

T.7.2 Ways forward

Socio-economic variables
Results showed that global datasets can still be improved both in terms of precision and completeness. However, they already allow the comparison of countries. Other indicators — such as a corruption, armed conflict or

political events — would be interesting to test in the model in the future.

Floods

Geophysical data can be improved. The watersheds used to estimate flood physical exposure were based on a 1 km cell resolution for elevation. A new global dataset on elevation from radar measures taken from a National Aeronautics and Space Administration (NASA) space shuttle is expected in 2004. It consists of a 30m resolution grid for the USA and 90m resolution for global coverage. This dataset will allow the refining of estimated areas exposed to flood risk. This advance will be especially welcome for the central Asian countries, where the quality of globally accessible available data was low.

Earthquakes

If information on soil (i.e. Quaternary rocks) and fault orientations can be generated, it would be possible to compute intensity using a modified Mercali scale, with much higher precision for delineating the affected area. Alternatively, a method for deriving frequency based on the Global Seismic Hazard Map from the GSHAP[13] could be used.

Cyclones

Once data from the North Indian Ocean is available, a vector approach should be applied using the PreView Global Cyclone Asymmetric Windspeed Profile model developed by UNEP/GRID-Geneva. This method computes areas affected, based on central pressure and sustainable winds.

Drought

Other precipitation datasets with higher spatial resolution could be usefully tested. The use of geo-climatic zones might be useful in order to take into account the usual climate of a specific area. Indeed, a drop of 50 percent precipitation might not have the same consequence on a humid climate as on a semi-arid area. The use ofthe Global Humidity Index (from UNEP/GRID UEA/CRU) might help in differentiating these zones. Measuring food insecurity (by using information on conflict and political status) would be also a significant improvement as compared to meteorological drought. Alternatively, drought could be measured in terms of crop failure through use of satellite imagery. This will be closer to drought as it impacts on food security.

The case of small islands and archipelagos

In some cases, small islands and archipelagos were too small to be considered by the GIS-automated algorithms. This was typically the case for population data. The raster information layer for population could not be used to extract the population of small islands. For single island countries, the problem might be overcome by using the population of the country, but for others this was not possible. Indeed, when superimposing cyclone tracks on top of archipelagos, the population is needed for each island. A manual correction is needed, but could not be performed due to the time-frame of the study. The compilation of socio-economic variables was also not complete for the islands. This might be improved by collaborating with agencies such as the South Pacific Applied Geoscience Commission (SOPAC) and Economic Commission for Latin America and the Caribbean (ECLAC) as both agencies are currently working on indicators for island vulnerability.

For all these reasons, the case of small island states and archipelagos would need a separate study.

Death as an indicator for risk

To what extend are deaths proportional to the significance of total losses, including losses of livelihood? In the case of earthquakes, where no early warning exists, this might be a good proxy. But it will depend on whether the earthquake epicentre is located in a rural or urban area. For tropical cyclone and flood, deaths are usually much smaller in relation to losses of houses, infrastructures and crops. In drought, the relationship is even more exaggerated. A much higher number of people are affected through the slow erosion of rural livelihoods and the possible influence of intervening factors, such as armed conflict, economic or political crisis, or epidemic disease such as HIV/AIDS. This makes separating the impact of drought from other factors a big challenge.

The ideal would be to have access to records of livelihood losses in order to calibrate the severity of one hazard type as compared to another (while considering the magnitude of a hazard). Other approaches for obtaining a structured assessment of comparative risk by country could include an assessment on the comparative severity of hazard using local and expert knowledge, or using relief and aid organisation budget data as a proxy for risk severity.

Extending to other hazards

Volcanic eruptions. The variability of volcanic hazards was too complex to be entered into a general model. Volcanic hazard ranges from lahars linked with precipitation level, seismicity, topography and soils characteristics, to tephra falls influenced by the prevailing wind direction and strength, and phreatomagmatic eruption. Despite this complexity, much data is available for volcanic hazard and each active volcano is well described. Data needed for a global assessment of volcanic risk probably exists. But a finer resolution for elevation is needed. It would be necessary to include data on the shape and relief of volcanoes, computing slopes and hazard from lahars. Remote sensing analysis for local assessment of danger and population distribution would also be required.

Tsunamis and landslides. Some countries are not well represented by the model because they are affected by hazards that are not of global significance. This is the case of Papua New Guinea and Ecuador, both affected by tsunamis, respectively 67.8 percent and 14.3 percent of national deaths. Landslides also cause significant losses in Indonesia (13 percent), Peru (33 percent) and Ecuador (10 percent) of recorded disaster-related deaths. As a result, the multi-hazard DRI is under evaluated for these countries.

1. Burton et al. 1993, p.34.
2. Coburn et al. 1991, p. 49.
3. Guha-Sapir, Debatathi and Below, Regina (2002) "Quality and Accuracy of Disaster Data: A Comparative Study of 3 Global Datasets," WHO Centre for Research on the Epidemiology of Disasters, University of Louvain School of Medicine for the Disaster Management Facility of the World Bank, Brussels.
4. Idem, p.14.
5. For a more detailed argument see the CRED-EM-DAT database http://www.cred.be/ and IFRC World Disaster Reports.
6. UNEP, 2002.
7. Birdwell & Daniel, 1991.
8. Bolt et al. 1975.
9. Bolt et al. 1975.
10. Birdwell & Daniel, 1991.
11. Landsea, 2000.
12. Giardini, 1999.