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.

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 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)

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 $-\infty$ and $+\infty$. 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 $-\infty$ to $+\infty$)

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

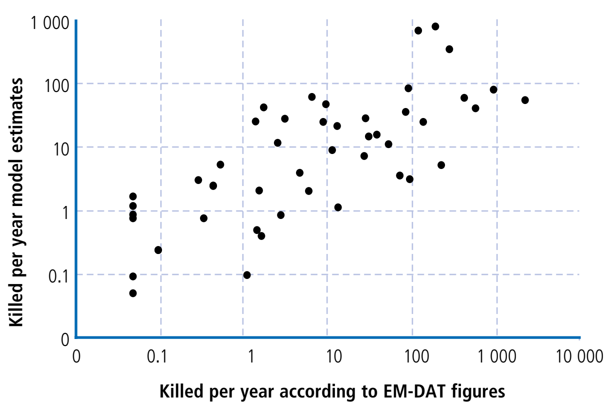
EQ11 $$ln(K) = 1.26ln(PhExp) + 12.27 \cdot U_g - 16.22$$

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

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.

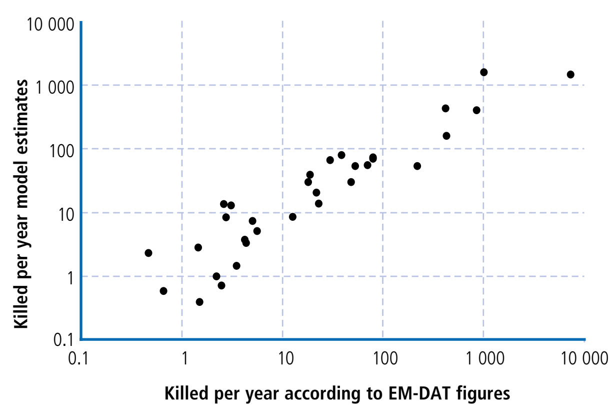
TABLE T.9 EXPONENT AND P-VALUE FOR CYCLONES MULTIPLE REGRESSION

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

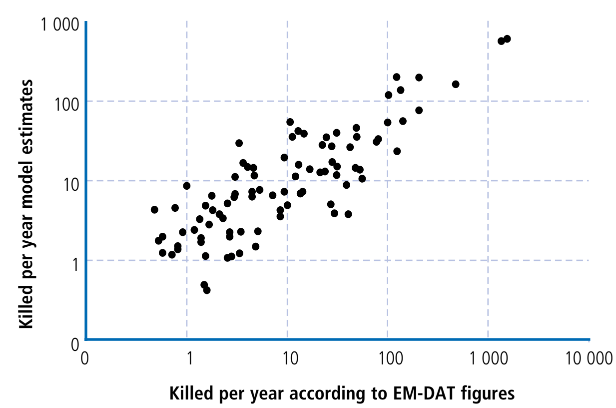
90 countries	B	p-value[i]
Intercept	−5.22	0.00000
ln(PhExp)	0.78	0.00000
$\ln(GDP_{cap})$	−0.45	0.00002
ln(Density)	−0.15	0.00321
$R= 0.84$, $R^2= 0.70$, adjusted $R^2= 0.69$		

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

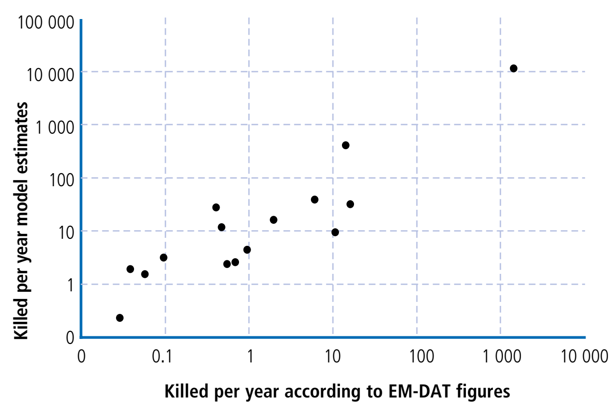
Where
- K is the number of killed from floods
- PhExp is the physical exposure to floods
- GDP_{cap} is the normalised Gross Domestic Product per capita (purchasing power parity)
- D is the local population density (i.e. the population affected divided by the area affected)

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).