An empirical relationship between fatalities and instrumental MMI

Keith A Porter, David Wald, Trevor Allen, and Kishor Jaiswal

An empirical model is created to estimate the probability distribution of earthquake fatalities in a single event, given a ShakeMap and the number of people exposed to various levels of shaking intensity, anywhere in the world, for possible use in the USGS's Prompt Assessment of Global Earthquakes for Response (PAGER) program. No information is used about structure types or their vulnerability; rather, the vulnerability of buildings in the affected region is accounted for with a scalar index assigned to the country where the earthquake occurs. The effect of time of day is accounted for using a sinusoid. Four model parameters are fit to a historical database of 42 worldwide earthquakes between 1970 and 2005. The model hindcasts fatalities in those events with a residual logarithmic standard deviation of 2.0 and uncertainty that fits a lognormal distribution. Thus the ±1 standard deviation bounds for a given event reflect an uncertainty factor of 7.4x in the number of deaths, or just under 1 order of magnitude. This model is preliminary and subject to further development. A model that uses a larger database of earthquakes and alternative country index is in development.

INTRODUCTION

The USGS' Prompt Assessment of Global Earthquakes for Response (PAGER) project employs earthquake location and magnitude determinations from the National Earthquake Information Center (NEIC), applies seismic attenuation relationships described in Wald et al. (2006) and topographically-derived site soil conditions described in Wald and Allen (2007), to estimate ground-shaking intensity maps. These maps show estimated instrumental intensity (MMI) for current, historic, and hypothetical earthquake events, on a 1-km grid basis. PAGER then draws on two global population databases to overlay population on the Shakemap and estimate the number of people exposed to various levels of MMI, on the same 1-km grid. The population databases are Landscan2006 by the Oak Ridge National Laboratory (http://www.ornl.gov/sci/landscan/) and GRUMP by Columbia University's Center for International Earth Science Information Network (http://sedac.ciesin.columbia.edu/gpw/).

The combination of data—estimated population by instrumental MMI level for historical, worldwide earthquakes with known total fatalities—offers a new loss-estimation opportunity: the estimation of worldwide, earthquake-induced fatalities using intensity and population distribution as explicit model parameters in an completely open risk model: open, peer-reviewed methodology, source data, and software code. The Russian program Extremum (Shakhramanian et al. 2000) includes such a model, but the data and code are not open.

To create such a global open fatality model requires (in addition to the source information, attenuation relationships, and population database) a relation between shaking intensity and fatality rate. Fatality rate is defined here as the number of people killed through earthquake-induced injuries, divided by the number of people exposed to earthquake shaking. Fatality rate could conceivably be affected by shaking intensity, shaking duration, kind and quality of construction, time of day, adjacency of buildings, social aspects such as public emergency preparedness training, secondary perils such as tsunami, fire, and ground failure, human response, and other variables. It would be a very simple relationship that ignored all parameters other than shaking intensity. A somewhat more sophisticated one would also employ a few, readily-available variables such as time of day and some general measure of regional vulnerability. This memo addresses the development of such fatality-rate models.

OBJECTIVE

This manuscript tests the hypothesis that fatality rate can be estimated as a simple, worldwide function of MMI, without requiring knowledge of the distribution of structure types or fragility functions. It is limited to fatalities induced by collapse of buildings due to shaking, i.e., excluding the effects of tsunami, landslides, fire, disease, etc. We test this hypothesis by attempting to fit various parametric functions to reports of total number of fatalities in a small number of past earthquakes, for which shaking intensities and population by 1-km grid can also be estimated. If a parametric form of a fatality-rate function can be found that produces estimated total fatalities for these events within, on average, ±½ to 1 order of magnitude of the reported total number of fatalities, the hypothesis will be deemed to have been supported. (This error term is defined more precisely later.)

LITERATURE REVIEW

A number of researchers have created earthquake fatality models. Engineers have tended to create models that explicitly model earthquake magnitude and location, distribution of shaking intensity, distribution of building types, vulnerability of the various building types, and deaths and injuries to occupants given various building damage states. Examples include Wiggins et al. (1976), ATC-13 (1985), and HAZUS-MH (NIBS and FEMA 2003). A notable challenge to implementing these models is the creation of a building inventory—the geographic and categorical distribution of buildings and other facilities in the affected regions.

Others have modeled earthquake fatalities using a direct relationship between earthquake magnitude and the expected number of victims. Samardjieva and Badal (2002) for example performed regression analyses of 20th Century earthquake deaths in 478 shallow-focus events that produced deaths and occurred in densely population locations, producing linear relationships between magnitude and the base-10 logarithm of fatalities. In Badal et al. (2005), the authors add the ability to disaggregate casualty estimates by intensity ranges. (Note that this is not the same thing as calculating population by intensity range, applying casualty rates at each range, and then summing—the method disaggregates, rather than aggregates.)

Several researchers have examined factors affecting earthquake casualties. Lomnitz (1970) found that the occurrence of foreshocks appeared to reduce casualties in large Chilean earthquakes in the preceding 400 years, and that time of day strongly affected earthquake casualties, with more casualties occurring at night (see Figure 1). Scawthorn et al. (1978) quantified the latter effect, examining worldwide 20th Century earthquakes and finding that time of day accounts for an average ±62% fluctuation in fatality rate (see Figure 2).

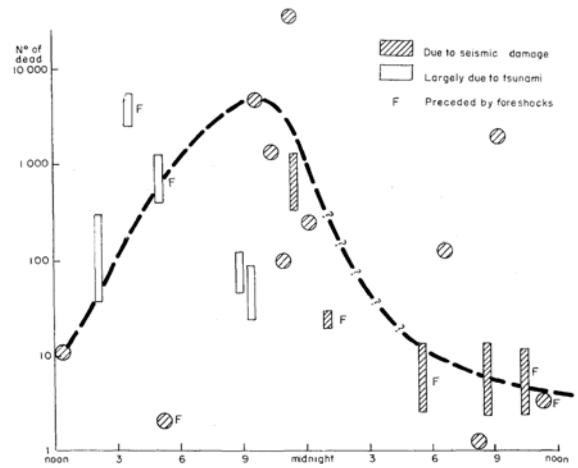

Figure 1. Lomnitz's (1970) observations of the effect of foreshocks and time of day on Peruvian earthquake fatalities in the period 1570-1970

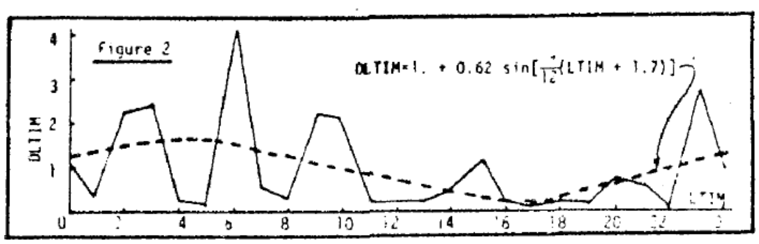

Figure 2. Scawthorn et al. (1978) effect of time of day on fatality rate

Davidson and Shah (1997) offer a scalar index of a city's seismic vulnerability based on six indicators: its seismic code, wealth, age, population density, speed of development, and fraction of its population under 4 or over 65. They exercise their methodology on 10 cities worldwide, producing the vulnerability index values shown in Table 1. In the table, a high index indicates greater vulnerability, i.e., greater loss when subjected to the same seismic excitation.

Table 1. Davidson and Shah (1997) vulnerability index

City	Boston	Istanbul	Jakarta	Lima	Manila	Mexico City	San Francisco	Santiago	St Louis	Tokyo
Vulnerability	1.9	2.0	2.4	2.3	2.6	2.1	1.3	2.3	1.9	1.2

The Global Disaster Alert and Coordination System (GDACS) also includes a vulnerability index (European Commission 2006). GDACS issues alerts to the international community regarding major sudden-onset disasters, including earthquakes among others, to facilitate international relief. GDACS does not estimate losses, but its authors have developed a 3-category vulnerability index that "identifies countries likely to suffer more than others from a humanitarian perspective in the event of a disaster." The index is calculated from a weighted average of various measures of life expectancy, education, wealth, health and health care, and refugees. The GDACS vulnerability index includes no direct measures of the built environment: no indicator of construction materials, building code requirements, building code enforcement, etc.

METHODOLOGY

In the present approach, a limited database of events is compiled, containing at least 5 earthquakes of magnitude 5 and above from each of 5 vulnerability regions (discussed later). Each event is characterized by its date, time, location, depth, reported fatalities, and estimated population exposed to MMI 5, 6, … 10. These last are estimated by applying the ShakeMap methodology to calculate the expected value of instrumental MMI on a 1-km-grid basis. The population in each gridcell is drawn from Oak Ridge National Laboratory's Landscan2003 population database. The total population exposed to shaking intensity $x - 0.5 \leq \text{MMI} < x + 0.5$ is then summed and referred to here as the population exposed to MMI x, where $x \in \{5, 6, \ldots 10\}$.

Three parametric functions are tested: normal cumulative distribution function (CDF), lognormal CDF, and beta CDF. Let the model be denoted by θ (meaning the selection of which CDF to use, the assumptions employed in ShakeMap and LandScan2003, etc.), the evidence be denoted by X (meaning the particular earthquakes employed in creating and testing the model, the estimated number of people exposed to shaking of various intensities in each earthquake, the total number of fatalities recorded for each earthquake, etc.). Let i index the earthquakes in X, and let N denote the number of earthquakes in X. Let j index the shaking intensity levels produced by the ShakeMap, and s_j denote a discrete shaking intensity (e.g., a particular value of MMI). Let $y(s_j)$ denote the fatality-rate function evaluated at s_j, i.e., the normal, lognormal, or beta CDF evaluated at s_j, with parameters fit to the data. Let $P_i(s_j)$ denote the estimated population exposed to shaking intensity s_j in event i, and O_i denote the recorded number of fatalities in event i. Let the objective function be denoted by $g(X,\theta)$, and given by

$$g\left(X,\theta\right) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\ln\left(\frac{E_i + 0.5}{O_i + 0.5}\right)\right)^2} \tag{1}$$

where E_i denotes the estimated fatalities in event i, calculated by

$$E_i = \sum_{j} y\left(s_j\right) P_i\left(s_j\right) \tag{2}$$

$y(s_j)$ is given by one of the following:

Normal
$$y\left(s_j\right) = \Phi\left(\frac{s_j - m}{s}\right) \tag{3}$$

Lognormal
$$y\left(s_j\right) = \Phi\left(\frac{\ln\left(s_j / x_m\right)}{\beta}\right) \tag{4}$$

Beta
$$y\left(s_j\right) = I_x\left(a,b\right) \qquad x = \frac{s_j - s_0}{s_1 - s_0} \tag{5}$$

and the addition of 0.5 deaths to the estimated and observed fatalities avoids numerical problems with the events with zero deaths. In Equations (3) through (5), Φ denotes the cumulative standard normal (Gaussian) distribution, I_x denotes the regularized incomplete beta function, and all parameters other than s_j are fit to the data.

APPLICATION

The evidence X employed here is summarized in Appendix 1; a total of 3,929 events are included in the catalog. Earthquake locations and magnitudes are summarized in Figure 3. The parameters of y are fit to the evidence by minimizing g using Microsoft Excel's solver routine. As shown in Equation (3), the parameters fit to the normal distribution form of $y(s)$ are the mean, denoted here by m, and the standard deviation, denoted here by s, which was constrained to be greater than 0.1. The parameters of the lognormal fit are the median, x_m, and logarithmic standard deviation, denoted here by β, as shown in Equation (4), both constrained to be greater than 0.1. The parameters of the beta distribution are a and b, lower bound s_0, and upper bound s_1. In the beta distribution, the lower bound was constrained to be less than 4.5, and the upper bound was constrained to be greater than 10.5.

Figure 3. Locations of 3,929 earthquakes used to develop model parameters.

Resulting parameter values are summarized in Table 2. All three models produce approximately the same value of g, suggested equal accuracy. Since the beta distribution requires two additional parameters and yields no additional accuracy, it is deleted from subsequent consideration. Since the lognormal distribution offers the constraint that it has zero probability at MMI ≤ 0, whereas the normal distribution does not, only the lognormal distribution is considered henceforth. As shown in Table 2, after minimizing $g(X,\theta)$, all the models have a residual logarithmic standard deviation g of 2.9, indicating that the 90^{th} percentile of fatalities for any given earthquake would be approximately 40 times the median estimate, and the 10^{th} percentile would have $1/40^{th}$ the median estimate.

Table 2. Summary of simple models examined

Normal		Lognormal		Beta	
				a	5.8
				b	5.0
m	12.6	x_m	14.7	s_0	4.5
s	1.6	β	0.22	s_1	27
g	2.91	g	2.90	g	2.90

* derived from other parameters, for comparison purposes.

ACCOUNTING FOR REGION, TIME OF DAY, AND POPULATION GROWTH

Why does the basic model offer such poor accuracy? Several reasonable hypotheses offer themselves:

1. ***Regional vulnerability***. The simple model does not account for regional differences in construction. Table 3 and Figure 4 reflect the authors' judgmental assignment of countries to groups whose building stock might perform approximately equivalently because of structure types in use, quality of construction, development of building codes, etc.

   If we denote by R_i the region in which event i occurred, one can make a preliminary estimate of whether these assignments might make a difference by calculating the correlation between $\log_{10}(O_i/E_i)$ and R_i. The correlation coefficient indicates the degree to which the model's underprediction of fatalities is correlated with the region's vulnerability level. As shown in Figure 5a, there is significant correlation between vulnerability region and modeling error. It is noted that region numbers are ordinal, rather than scalar, values, meaning that countries in region 2 are simply deemed more vulnerable than those in region 1, not necessarily twice as vulnerable. However, as an preliminary adjustment factor to account for region, one could factor fatality estimate as follows:

   $$\Delta_R = 10^{c \ln R + d} \tag{6}$$

   where R denotes the vulnerability region of the gridcell with the highest estimated MMI and c and d are parameters of the function, which Figure 5 indicates as c = 1.92 and d = -2.25, respectively, but which may change over time as the model is trained with more data.

Table 3. Assignment of geographic regions to vulnerability levels

Level and geographic regions	
1	California, New Zealand
2	U.S. (not California), Japan*, Taiwan, Canada, Northwest Europe, Australia
3	South Europe, Eastern Europe, Turkey, Eastern Russia, Mexico, Belize, Costa Rica, Philippines, South Africa
4	Indonesia, Korea, Africa, Southeast Asia, South America, Central America (not Mexico, Belize, or Costa Rica), southern Former Soviet Union
5	Iran, Afghanistan, Pakistan, Mongolia, India, China, Mongolia

* Assigned 2 because of traditional woodframe

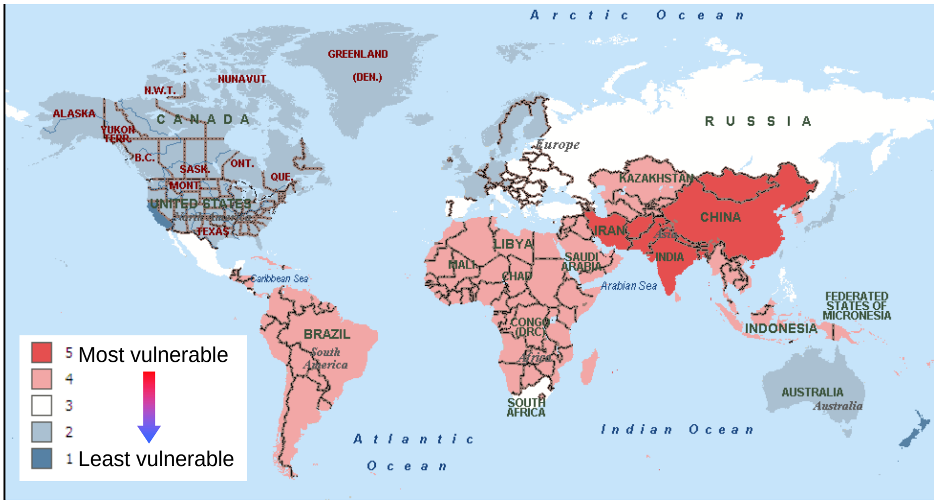

Figure 4. Vulnerability regions

2. ***Time of day.*** No account is made of time of day, which as noted above may account for ±60% fluctuation in fatality rate. Equation (7), taken from Scawthorn et al. (1978), is adopted here to account for time of day, where t denotes local time of day in hours:

$$\Delta_t = 1 + 0.62\sin\left(\frac{\pi}{12}(t + 1.7)\right) \tag{7}$$

3. ***Population growth.*** Recorded deaths and estimated population are based on different years, with events in the dataset reflecting population of almost 50 years ago. To refine the model, it is valuable to estimate population as it existed at the time of the earthquake, rather than 2003. Using current-year population growth (e.g., from the CIA factbook at https://cia.gov/cia//publications/factbook/fields/2002.html), and denoting population growth rate and number of years ago λ and T, respectively, one could apply a years-ago adjustment factor to the fatality estimate of

$$\Delta_T = (1 + \lambda)^{-T} \tag{8}$$

where

T = (Landscan baseline year) – (year of earthquake)
λ = population growth rate (varies by country; use the country of the gridcell with the highest estimated MMI on land)

To employ these adjustments, we substitute the following expression for the median estimate of total fatalities in event i, denoted by $E_{0.5i}$, in place of Equation (2):

$$E_{0.5i} = \Delta_R \Delta_t \Delta_T \sum_j y(s_j) P_i(s_j) \tag{9}$$

where Δ_R, Δ_t, and Δ_T are as defined in Equations (6), (7), and (8), respectively, $y(s_j)$ is as defined in Equation (4), and $P_i(s_j)$ denotes the estimated population exposed to shaking intensity s_j in event i. Using instrumental MMI as the intensity measure, $s_j \in \{5, 6, \ldots 10\}$, and the population in any gridcell is included in $P_i(s_j)$ if the gridcell is estimated to experience shaking of intensity

MMI such that $s_j - 0.5 \le \text{MMI} < s_j + 0.5$. That is, denoting by k a gridcell index, and denoting by MMI_k the estimated MMI in gridcell k,

$$\begin{aligned} V\left(s_j\right) &= \sum_k P_k H_{jk} \\ H_{jk} &= 1 \ \textit{if} \ \ s_j - 0.5 \le MMI_k < s_j + 0.5 \\ &= 0 \ \textit{otherwise} \end{aligned} \tag{10}$$

In Equation (10) the event index i is omitted for clarity. Minimizing Equation (1) using the lognormal function $y(s_j)$ yields the following model parameters: $x_m = 16.0$, $\beta = 0.25$, $c = 1.92$, $d = -2.25$, and $g = 2.0$. Figure 5a shows the effect of region, while Figure 5b shows the estimated versus observed fatalities in the 42 events used to create the model; the median fatality estimates E_i are shown in Appendix 1.

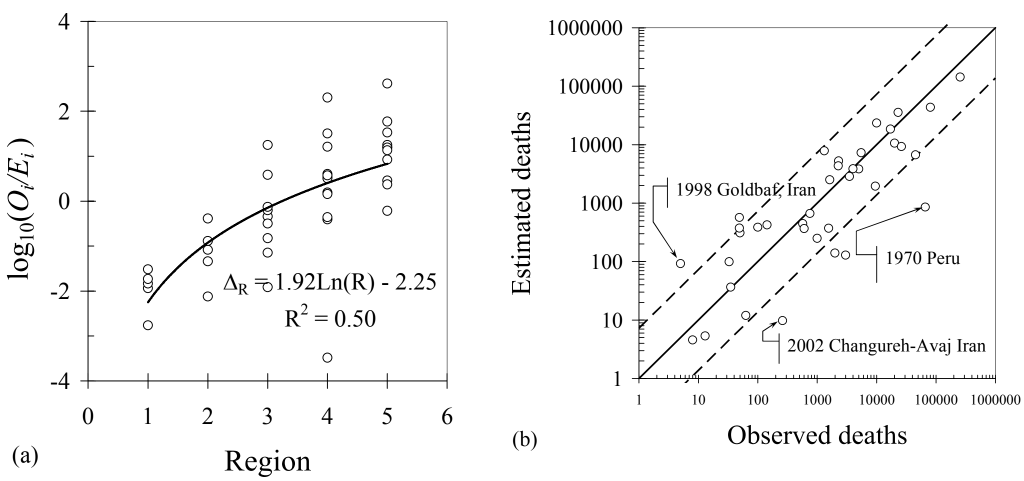

Figure 5. Enhanced model, after accounting for region, years ago, and time of day: (a) effect of region, and (b) accuracy of the model, with some major outliers noted. The solid diagonal line in (b) indicates perfect agreement, and the dashed lines indicate the ± 1 logarithmic standard deviation bounds (a factor of 7.4).
Table 4. Event-by-event fatality estimates for the enhanced model that accounts for region, time of day, and population growth . "Observed" means total recorded fatalities, "Estimated" means number of fatalities estimated by the enhanced model.

RESIDUAL ERROR

The samples of the residual error term $\varepsilon = \ln(E_i/O_i)$ were compared with a normal distribution using the Lilliefors (1967) goodness-of-fit test and the 5% significance level. Lilliefors is a stricter version of the Kolmogorov-Smirnov goodness of fit test applicable when the parameters of the normal distribution are estimated from the sample. The data pass the Lilliefors test: $N = 42$; Max| $F^*(X) - S_N(X)$ | = 0.11, as shown in Figure 6; which is less than the critical value $D_{0.05} = 0.886N^{-0.5} = 0.14$. This indicates that it is reasonable to model the fatalities as lognormally distributed about the median estimate with a logarithmic standard equal to the value of g calculated in Equation (1). A similar test of the absolute residual error $E_i - O_i$ fails, indicating that it is not reasonable to model fatalities as normally distributed about the estimate.

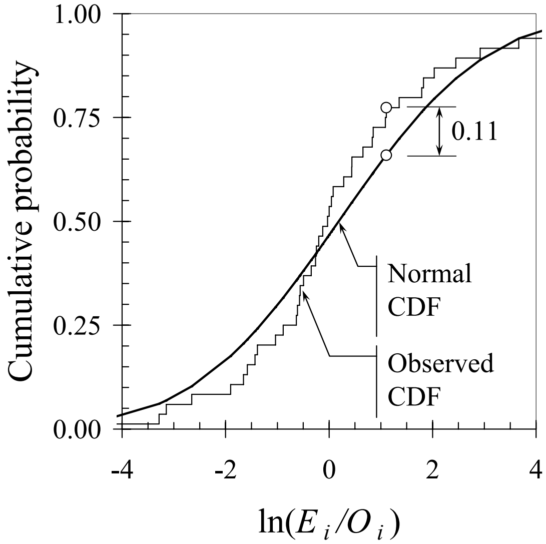

Figure 6. Lilliefors test of lognormal distribution for residual uncertainty

With the knowledge that the residual error is approximately lognormally distributed together with the value of its logarithmic standard deviation (denoted herein by *g*), one can estimate the probability distribution of fatalities given the estimated shaking intensity and population exposed. One can estimate the number of fatalities associated with probability *p* of not being exceeded, denoted here by E_p, as follows:

$$E_p = E_{0.5} \exp\left(g \cdot \Phi^{-1}\left(p\right)\right) \tag{11}$$

where $E_{0.5}$ denotes the median fatality estimate from Equation (9) and $\Phi^{-1}(p)$ denotes the inverse of the standard normal cumulative distribution function evaluated at *p*. Stated another way, with approximate probability *p*, the fatality estimate is within a factor *f* of the median estimate, or *m* orders of magnitude, where:

$$\begin{aligned} f &= \exp\left(g \cdot \Phi^{-1}\left(0.5 + \frac{p}{2}\right)\right) \\ m &= \log_{10}\left(f\right) \\ &= 0.434 \cdot g \cdot \Phi^{-1}\left(0.5 + \frac{p}{2}\right) \end{aligned} \tag{12}$$

Table 5 shows some values of *f* and *m* for *g* = 2. For example, it shows that approximately half the time, the model's median estimate of fatalities would be within a factor of 4 (just over ½ an order of magnitude) of the reported shaking-related fatalities in the earthquake. It shows the 75% of the time, the median fatality estimate would be would be within 1 order of magnitude of the true result. In approximately 2% of cases, the median would be off by more than 2 orders of magnitude.

Table 5. Maximum ratio of observed to estimated fatalities with probability *p*

p	*f*	*m*
50%	4	0.6
68%	7	0.9
75%	10	1.0
90%	27	1.4
98%	100	2.0

The residual error results from a number of sources, among which are:

- Differences between actual shaking intensities and ShakeMap estimates;
- Differences between the actual population distributions and LandScan estimates;
- Differences between reported and actual fatalities;
- Geographic and temporal variability in building damageability absent from the model;
- Variability in the processes that lead from building damage to human deaths.

COMPARISONS

Vulnerability region versus other vulnerability indices. The assignment of vulnerability region, R, to each country was initially intended only as an ordinal ranking of countries, i.e., a country with region r is assumed to experience fewer fatalities, on average, than one with region $r+1$, given the same population in each shaking intensity level. However, as shown in Figure 5a, region assignments can usefully be used as scalar values. The assignment of R values was made independently of the Davidson and Shah (1997) or GDACS (European Commission 2006) vulnerability indices, but it is noteworthy that the correlation coefficient is 0.9 between the Davidson and Shah 1997 vulnerability index for each city in Table 1 and the regional vulnerability level from Table 3. There is little correlation between R and the GDACS vulnerability index, ($\rho = 0.1$), possibly because the latter focuses on the vulnerability of the people (in terms of health, education, and welfare), as opposed to the vulnerability of their buildings.

The World Housing Encyclopedia (www.world-housing.net) currently offers reports detailing the design and seismic performance of more than 100 housing types worldwide, generally written by architects or engineers familiar with local construction. Each report includes the authors' estimate of the seismic vulnerability rating a 6-level (A-F, *worst to best*) for each type, and in some cases indicates the fraction of housing in the country represented by that type. The data are fairly sparse and not entirely consistent, but a qualitative comparison with the vulnerability regions defined here is possible. Let i denote an index to housing types within a country for which the WHE contains a report. Let V_i denote a value assigned to the housing type's seismic vulnerability rating, as shown in Table 6. Let p_i denote the fraction of all housing in the country represented by that type, according to the WHE report's authors. Finally, let V denote the weighted average of the V_i for a given country:

$$V = \left(\sum_{i=1}^{N} p_i V_i \right) \Big/ \left(\sum_{i=1}^{N} p_i \right) \qquad (13)$$

where N is the number of housing types in the country for which both V_i and p_i are given in the WHE. V must range between 1 to 6 (best to worst), while the values of R shown in Table 3 range from 1 to 5 (best to worst). Sixteen countries have adequate data in the WHE to evaluate

Equation (13). A regression analysis of V versus R produces a correlation coefficient of 0.42—not strong enough correlation to conclude that a trend exists (or more precisely to reject the null hypothesis that no trend exists, at the 5% significance level). Does this mean anything? Perhaps not. Figure 7(a) shows V calculated using Equation (13) and the few reports with adequate detail in WHE. Notice that the value of V for the United States is the same as that of Serbia and Montenegro; Canada's V is the same as that of Kazakhstan. No V can be calculated for China, Japan, Indonesia, or a variety of countries with high seismicity. Figure 7(b) shows that, for those countries whose V can be calculated, many are based on reports representing only a few percent of the total housing stock.

Table 6. Numeric values assigned to World Housing Encyclopedia seismic vulnerability ratings

Rating	A	B	C	D	E	F
V_i	6	5	4	3	2	1

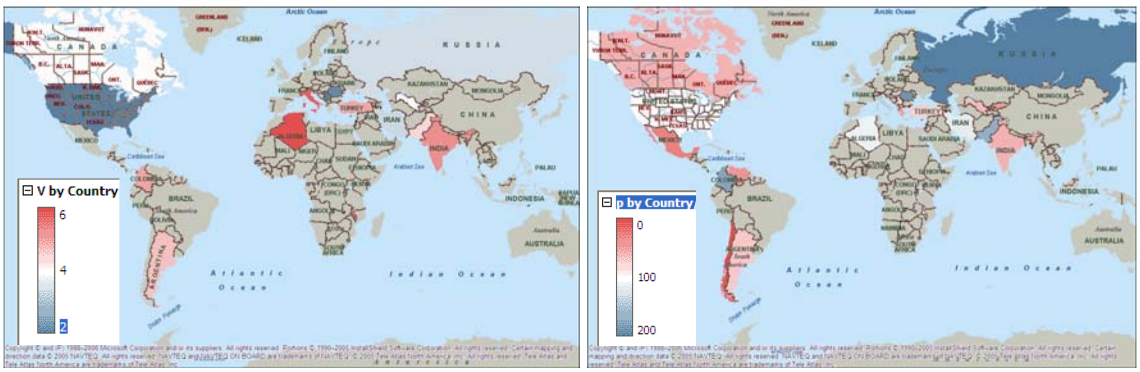

Figure 7. (a) V from WHE (b) WHE coverage

CONCLUSIONS

The USGS' PAGER product has been used to calculate population by MMI level for 42 historic earthquakes from around the world. The events were sampled to reflect earthquakes with large numbers of people exposed to a wide range of shaking intensity from each of 5 general vulnerability regions, including events occurring between 1970 and 2005. A simple fatality-rate model was developed using these data to relate shaking intensity (MMI) to fatality rate, using a parametric cumulative distribution function (CDF) for a basic relationship between MMI and fatality rate, and three factors to account for regional vulnerability, time of day, and population change.

Model parameters were fit to the data, with the objective of minimizing the squared error (in the logarithmic domain) in the estimated total event fatalities. Three competing CDFs were considered: normal, lognormal, and beta distributions to estimate basic fatality rate. The lognormal fit the data best. Prescribed functions (i.e., with parameter values not assigned here) were used to account for time of day and population growth. An additional, 2-parameter function was fit to account for geographic region, reflecting the notion that the kinds and quality of buildings in one part of the world may be more or less deadly than in another part of the world. Thus, to create the final model involved fitting 4 parameters to the data: two for the median and logarithmic standard deviation of the lognormal CDF, and two for the regional effect.

The hypothesis being tested was that such a simple model can estimate total fatalities with approximately 90% confidence that the model would not *underestimate* fatalities by more than 1 order of magnitude, and similar confidence that it would not *overestimate* fatalities by more than 1 order of magnitude. The hypothesis is just supported by the data: the model produces a residual logarithmic standard deviation of 2, equivalent to the two confidence levels just stated. A more sophisticated model may be required for greater accuracy. For convenience, the model is recapped below.

$$y(s_j) = \Phi\left(\frac{\ln(s_j/x_m)}{\beta}\right), \quad x_m = 16.0, \beta = 0.25 \tag{4}$$

$$\Delta_R = 10^{c\ln R + d}, \quad c = 1.92, d = -2.25 \tag{6}$$

$$\Delta_t = 1 + 0.6\sin\left(\frac{\pi}{12}(t+2)\right) \tag{7}$$

$$\Delta_T = (1+\lambda)^{-T} \tag{8}$$

$$E_i = \Delta_R \Delta_t \Delta_T \sum_j y(s_j) P(s_j) \tag{9}$$

where s_j is the shaking intensity in terms of instrumental MMI, $P(s_j)$ is the year-2003 population exposed to shaking s_j, R is the vulnerability region of the event per Table 3 and Figure 4, t is the local time of day in hours (24-hr clock), λ is the annual population growth rate, and T = 2003 – the year of the earthquake. Appendix 2 contains a table of R and λ by country.

The foregoing model has at least two important limitations that we are working to overcome. First, the database of earthquakes is of modest size, and may be biased toward fatal earthquakes. We have developed a larger database of nearly 4,000 events to further test or improve the methodology. Second, the assignment of countries to regions is based solely on judgment and experience. We are working with the World Housing Encyclopedia and considering other methods to assign regions, possibly with a probabilistic aspect. The findings presented here should therefore be seen as preliminary, indicative of PAGER's goal of fatality estimation and one promising method possibly to achieve it. Further development is ongoing.

ACKNOWLEDGMENTS

This research was supported by the U.S. Geological Survey. Thanks also to Shingo Nishida for data extracted from the World Housing Encyclopedia and to Charlie Scawthorn and Dominic Dowling for advice regarding regionalization.

REFERENCES CITED

(ATC) Applied Technology Council, 1985. *ATC-13, Earthquake Damage Evaluation Data for California,* Redwood City, CA, 492 pp.

Badal, J., M. Vazquez-Prada, and A. Gonzalez, 2005. Preliminary quantitative assessment of earthquake casualties and damages. *Natural Hazards* 34:353-474.

Davidson, R., and H. Shah, 1997. *An Urban Earthquake Disaster Risk Index*. Report 121. John A. Blume Earthquake Engineering Center, Stanford, CA., 269 pp.

European Commission Directorate-General for Humanitarian Aid – ECHO, 2006. *Technical Note, Assessment of Humanitarian Needs and Identification of Forgotten Crises*. Brussels. 24 pp.

Lilliefors, H., 1967. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, *Journal of the American Statistical Association*, 62 (318). (June 1967), 399-402

Lomnitz, C., 1970. Casualties and behavior of populations during earthquakes. *Bull. Seism. Soc. Am*. 60 (4), 1309–1313

(NIBS and FEMA) National Institute of Building Sciences and Federal Emergency Management Agency, 2003. *HAZUS-MH MR1 Advanced Engineering Building Module Technical and User's Manual,* Federal Emergency Management Agency, Washington, DC, 119 pp.

Samardjieva, E. and J. Badal, 2002. Estimation of the expected number of casualties caused by strong earthquakes. *Bulletin of the Seismological Society of America*, 92 (6), pp. 2310-2322, Aug 2002.

Scawthorn, C., H. Iemura, H., and Y. Yamada, 1978. World large destructive earthquakes since 1900. *Proc. 33rd JSCE Annual Meeting, Sendai, Japan, September 1978*, pp. 318-319.

Shakhramanian, M.A., V.I. Larionov, G.M. Nigmetov, and S.P. Sutschev, 2000. *Assessment of the Seismic Risk and Forecasting Consequences of Earthquakes While Solving Problems on Population Rescue (Theory and Practice)*. Ministry of Russian Civil Defense and Disaster Management Research Institute, Moscow, Russia, 180 pp.

Wald, D.J. and T.I. Allen, 2008 (expected). Topographic slope as a proxy for seismic site conditions and amplification. Accepted April 2007 for publication in the *Bulletin of the Seismological Society of America,* http://earthquake.usgs.gov/research/hazmaps/interactive/vs30/inc/wald_allen_2007_toposlope.pdf [viewed 5 Nov 2007]

Wald, D.J., B.C. Worden, V. Quitoriano, and K.L. Pankow, 2006. ShakeMap Manual ver. 1.0. US Geological Survey, Golden, CO., 156 pp. http://pubs.usgs.gov/tm/2005/12A01/pdf/508TM12-A1.pdf [viewed 31 Jul 2006]

Wiggins, J.H., J. Slosson, and J. Krohn, 1976. *Natural Hazards: Earthquake, Landslide, Expansive Soil Loss Models*. J.H. Wiggins Company, Redondo Beach, CA.

Appendix 1. Evidence used to create fatality-rate models

Event	Date	TimeUTM	LocalTime	Location	LatN	LonE	Depthkm	M	Pop5Est	Pop6Est	Pop7Est	Pop8Est	Pop9Est	Pop10Est	Deaths O	Estimated E	Note
19700531	5/31/1970	20:23:00	1523	Peru	-9.4	-78.9	64	7.8	13000000	1430000	90000	387000	12000	0	66,000	858	
19721223	12/23/1972	6:29:00	0029	Managua, Nicaragua	12.15	-86.27	6.2	6.0	2730000	747000	473000	671000	0	0	5,000	3,838	
19750204	2/4/1975	11:36:07	1936	Haicheng, China	40.67	122.65	8	7.0	19800000	5680000	1420000	1330000	210000	0	10,000	23,456	
19760204	2/4/1976	9:01:43	0301	Guatemala	15.32	-89.1	5	7.5	12600000	6060000	3180000	2720000	717000	1000	23,000	35,744	
19760517	5/17/1976	2:58:41	0858	Gazli, USSR	40.38	63.47	10	6.8	2710000	35000	3000	3000	0	0	0	8	
19760727	7/27/1976	19:42:54	0342	Tangshan, China	39.38	118.11	23	7.8	49600000	12000000	10200000	2910000	1570000	427000	255,000	143,895	
19780620	6/20/1978	20:03:21	2303	Thessaloniki, Greece	40.74	23.23	22	6.6	2280000	539000	747000	16000	0	0	50	310	
19801010	10/10/1980	12:25:23	1325	El Asnam, Algeria	36.2	1.35	10	7.7	13300000	3260000	974000	548000	66000	0	3,500	2,857	
19801123	11/23/1980	18:34:00	1934	Iripinia, Italy	40.77	15.3		6.9	2480000	474000	111000	78000	2000	0	3,000	129	
19850919	9/19/1985	13:17:47	0717	Michoacan, Mexico	18.19	-102.53	27	8.0	12400000	3520000	1410000	617000	0	0	9,500	1,955	
19861010	10/10/1986	17:49:24	1149	San Salvador, El Salvador	13.79	-89.17	10.9	5.7	3030000	2130000	191000	0	0	0	1,000	250	
19870302	3/2/1987	1:42:34	1442	Edgecumbe, New Zealand	-37.92	176.76	15	6.5	425000	157000	34000	3000	0	0	0	0	
19871001	10/1/1987	14:42:00	0742	Whittier Narrows, CA	34.06	-118.14	14.4	5.9	6110000	6190000	1020000	0	0	0	8	5	
19891018	10/18/1989	0:04:15	1704	Loma Prieta, CA	37.11	-121.76	16	6.9	6090000	4020000	2240000	1580000	49000	0	63	12	
19891227	12/27/1989	23:26:57	1024	Newcasatle, Australia	-32.95	151.61	9.7	5.6	446000	274000	95000	0	0	0	13	5	
19900620	6/20/1990	21:00:09	0030	Manjil, Iran	36.96	49.41	25	7.4	13400000	3280000	1080000	86000	0	0	45000	6,700	
19900716	7/16/1990	7:26:34	1526	Luzon, Philppines	15.68	121.17	18	7.7	16800000	20200000	4170000	1270000	385000	28000	1,621	2,514	
19910628	6/28/1991	14:43:54	0743	Sierra Madre, CA	34.26	-118	11	5.6	14600000	2150000	0	0	0	0	1	1	
19920413	4/13/1992	1:20:00	0220	Heinsberg, Germany	51.18	5.84	18	5.3	7540000	272000	1000	0	0	0	0	4	
19940117	1/17/1994	12:30:55	0430	Northridge, CA	34.16	-118.56	19	6.6	5510000	6230000	3450000	1510000	493000	45000	33	100	
19950116	1/16/1995	20:46:52	0546	Kobe, Japan	34.58	135.02	21	6.9	22300000	6810000	12000000	3570000	1960000	141000	5,502	7,288	
19950527	5/27/1995	13:03:52	0104	Sakhalin Is., Russia	52.63	142.83	11	7.1	17000	34000	3000	1000	3100	0	1,989	140	1
19951009	10/9/1995	15:35:53	0935	Colima-Jalisco, Mexico	19.05	-104.21	26.4	8.0	7090000	877000	752000	159000	0	0	49	568	
19970204	2/4/1997	10:37:47	1407	Garmkhan, Iran	37.66	57.29	10	6.5	1540000	108000	272000	1000	0	0	100	388	
19970510	5/10/1997	7:57:29	1227	Ardakul, Iran	33.82	59.74	10	7.3	3390000	227000	49000	20000	0	0	1,560	373	
19980314	3/14/1998	19:40:27	2310	Golbaf, Iran	30.15	57.6	9	6.6	1270000	157000	4000	1000	0	0	5	93	
19980530	5/30/1998	6:22:28	1052	Afghanistan	37.11	70.11	33	6.6	50000	918000	930000	100000	0	0	4,000	3,880	
19990817	8/17/1999	0:01:39	0301	Izmit, Turkey	40.75	29.86	17	7.5	6640000	6210000	6340000	1900000	783000	33000	17,118	18,517	
19990907	9/7/1999	11:56:49	1456	Athens, Greece	38.12	23.61	10	6.0	552000	617000	3040000	2000	0	0	143	425	
19990920	9/20/1999	5:47:20	0147	Chichi, Taiwan	23.85	120.82	8	7.6	22700000	7440000	2940000	2660000	1680000	164000	2,297	5,280	
19991112	11/12/1999	16:56:19	1856	Duzce, Turkey	40.76	31.16	10	7.1	11700000	935000	162000	223000	109000	0	755	668	
20010126	1/26/2001	3:16:40	0846	Bhuj, India	23.42	70.23	16	7.7	32200000	8150000	854000	130000	21000	0	20,023	10,626	
20010623	6/23/2001	20:33:14	1533	Peru	-16.26	-73.64	33	8.4	903000	1380000	253000	77000	0	0	49	374	
20020622	6/22/2002	2:58:20	0728	Changureh-Avaj, Iran	35.63	48.93	10	6.5	219000	19000	0	0	0	0	261	10	
20030521	5/21/2003	18:44:20	1944	Boumerdes, Algeria	36.96	3.63	12	6.8	6610000	3610000	1990000	502000	37000	0	2,266	4,356	
20031226	12/26/2003	19:15:56	0526	Bam, Iran	29	58.34	10	6.6	3230000	359000	253000	49000	59000	0	26,200	9,310	
20040224	2/24/2004	2:27:46	0227	Tetouan, Morroco	35.23	-3.96	1.7	6.4	2930000	191000	67000	27000	0	0	571	446	
20041023	10/23/2004	8:56:00	1756	Niigata, Japan	37.23	38.75	16	6.6	15600000	1320000	398000	166000	0	0	35	37	
20041115	11/15/2004	9:06:55	0406	Buenaventura, Columbia	4.61	-77.54	10	7.2	6940000	347000	37000	12000	11000	0	0	764	
20050222	2/22/2005	2:25:22	0555	Kerman, Iran	30.73	56.9	14	6.4	581000	101000	67000	0	0	0	602	368	
20050328	3/28/2005	16:09:36	2309	Nias, Sumatra	2.08	97.01	30	8.6	8870000	3690000	128000	854000	10000	0	1,313	7,872	
20051008	10/8/2005	3:50:40	0850	Kashmir, Pakistan	34.49	73.63	26	7.6	36500000	19100000	2060000	668000	147000	0	80,361	43,575	

1: Pop9Est reflects pre-earthquake population, not Landscan2003

Appendix 2: Vulnerability region *R* and population growth by country

ID	Country	R	λ
1	Afghanistan	5	2.7%
3	Albania	3	0.5%
4	Algeria	4	1.2%
5	American Samoa	3	-0.2%
6	Andorra	3	0.9%
7	Angola	4	2.5%
8	Anguilla	3	1.6%
9	Antigua and Barbuda	4	0.6%
10	Argentina	4	1.0%
11	Armenia	3	-0.2%
12	Aruba	2	0.4%
14	Australia	2	0.9%
15	Austria	2	0.1%
16	Azerbaijan	3	0.7%
17	Bahamas, The	3	0.6%
18	Bahrain	4	1.5%
19	Baker Island	3	1.5%
20	Bangladesh	5	2.1%
21	Barbados	4	0.4%
23	Belarus	3	-0.1%
24	Belgium	2	0.1%
25	Belize	2	2.3%
26	Benin	4	2.7%
27	Bermuda	2	0.6%
28	Bhutan	5	2.1%
29	Bolivia	4	1.5%
30	Bosnia and Herzegovina	3	1.4%
31	Botswana	4	0.0%
32	Bouvet Island	4	0.0%
33	Brazil	4	1.0%
34	British Indian Ocean Territory	5	1.0%
35	British Virgin Islands	2	2.0%
36	Brunei	4	1.9%
37	Bulgaria	3	-0.9%
38	Burkina Faso	4	3.0%
39	Burma	4	0.8%
40	Burundi	4	3.7%
41	California, USA	1	0.9%
42	Cambodia	4	1.8%
43	Cameroon	4	2.0%
44	Canada	2	0.9%
45	Cape Verde	4	0.6%
46	Cayman Islands	2	2.6%
47	Central African Republic	4	1.5%
48	Chad	4	2.9%
49	Chile	4	0.9%
50	China	5	0.6%
51	Christmas Island	5	0.0%
53	Cocos (Keeling) Islands	5	0.0%
54	Colombia	4	1.5%
55	Comoros	5	2.9%
56	Congo, Democratic Republic of the	4	3.1%
57	Congo, Republic of the	4	2.6%
60	Cook Islands	4	-1.2%
62	Costa Rica	3	1.5%
63	Cote d'Ivoire	4	2.0%
64	Croatia	3	0.0%
65	Cuba	4	0.3%
66	Cyprus	3	0.5%
67	Czech Republic	3	-0.1%
68	Denmark	2	0.3%
70	Djibouti	4	2.0%
71	Dominica	4	-0.1%
72	Dominican Republic	4	1.5%
73	East Timor	4	2.1%
74	Ecuador	4	1.5%
75	Egypt	4	1.8%
76	El Salvador	4	1.7%
77	Equatorial Guinea	4	2.1%
78	Eritrea	4	2.5%
79	Estonia	3	-0.6%
80	Ethiopia	4	2.3%
83	Falkland Islands (Islas Malvinas)	4	2.4%
84	Faroe Islands	2	0.6%
85	Fiji	4	1.4%
86	Finland	2	0.1%
87	France	2	0.4%
88	French Guiana	4	2.0%
89	French Polynesia	4	1.5%
90	French Southern and Antarctic Lands	4	1.5%
91	FYRO Macedonia	3	1.5%
92	Gabon	4	2.1%
93	Gambia, The	4	2.8%
94	Georgia	3	-0.3%
95	Germany	2	0.0%
96	Ghana	4	2.1%
97	Gibraltar	3	0.1%
99	Greece	3	0.2%
100	Greenland	2	0.0%
101	Grenada	4	0.3%
102	Guadeloupe	4	0.9%
103	Guam	3	1.4%
104	Guatemala	4	2.3%
105	Guernsey	2	0.3%
106	Guinea	4	2.6%
107	Guinea-Bissau	4	2.1%
108	Guyana	4	0.3%
109	Haiti	4	2.3%
110	Heard Island and McDonald Islands	5	2.3%
111	Holy See (Vatican City)	3	0.0%
112	Honduras	4	2.2%
113	Hong Kong	3	0.6%

ID	Country	R	λ
114	Howland Island	2	0.6%
115	Hungary	3	-0.3%
116	Iceland	2	0.9%
118	India	5	1.4%
119	Indonesia	4	1.4%
120	Iran	5	1.1%
121	Iraq	4	2.7%
122	Ireland	2	1.2%
123	Isle of Man	2	0.5%
124	Israel	2	1.2%
125	Italy	3	0.0%
126	Jamaica	4	0.8%
127	Jan Mayen	2	0.8%
128	Japan	2	0.0%
129	Jarvis Island	2	0.0%
130	Jersey	3	0.3%
131	Johnston Atoll	2	0.3%
132	Jordan	4	2.5%
134	Kazakhstan	4	0.3%
135	Kenya	4	2.6%
136	Kingman Reef	2	2.6%
137	Kiribati	4	2.2%
142	Kuwait	4	3.5%
143	Kyrgyzstan	4	1.3%
144	Laos	4	2.4%
145	Latvia	3	-0.7%
146	Lebanon	3	1.2%
147	Lesotho	4	-0.5%
148	Liberia	4	4.9%
149	Libya	4	2.3%
150	Liechtenstein	3	0.8%
151	Lithuania	3	-0.3%
152	Luxembourg	2	1.2%
153	Macau	4	0.9%
154	Macedonia	3	0.3%
155	Madagascar	4	3.0%
156	Malawi	4	2.4%
157	Malaysia	4	1.8%
158	Maldives	4	2.8%
159	Mali	4	2.6%
160	Malta	4	0.4%
161	Marshall Islands	2	2.3%
162	Martinique	4	0.7%
163	Mauritania	4	2.9%
164	Mauritius	5	0.8%
165	Mayotte	4	3.8%
166	Mexico	3	1.2%
167	Micronesia, Federated States of	4	-0.1%
168	Midway Islands	3	-0.1%
169	Moldova	3	0.3%
170	Monaco	3	0.4%
171	Mongolia	5	1.5%
172	Montenegro	3	3.5%

ID	Country	R	λ
173	Montserrat	4	1.1%
174	Morocco	4	1.6%
175	Mozambique	4	1.4%
176	Namibia	4	0.6%
177	Nauru	4	1.8%
178	Navassa Island	2	1.8%
179	Nepal	5	2.2%
180	Netherlands	2	0.5%
181	Netherlands Antilles	4	0.8%
182	New Caledonia	4	1.2%
183	New Zealand	1	1.0%
184	Nicaragua	4	1.9%
185	Niger	4	2.9%
186	Nigeria	4	2.4%
187	Niue	4	0.0%
188	Norfolk Island	4	0.0%
140	North Korea	4	0.8%
189	Northern Mariana Islands	4	2.5%
190	Norway	2	0.4%
191	Oman	4	3.3%
192	Pakistan	5	2.1%
193	Palau	3	1.3%
194	Palmyra Atoll	2	1.3%
195	Panama	4	1.6%
196	Papua New Guinea	4	2.2%
197	Paraguay	4	2.5%
198	Peru	4	1.3%
199	Philippines	3	1.8%
200	Pitcairn Islands	4	0.0%
201	Poland	3	-0.1%
202	Portugal	3	0.4%
203	Puerto Rico	3	0.4%
204	Qatar	4	2.5%
205	Reunion	5	1.3%
206	Romania	3	-0.1%
207	Russia	3	-0.4%
208	Rwanda	4	2.4%
209	Saint Helena	3	0.6%
210	Saint Kitts and Nevis	4	0.5%
211	Saint Lucia	4	1.3%
212	Saint Pierre and Miquelon	4	0.2%
213	Saint Vincent and the Grenadines	4	0.3%
214	Samoa	4	-0.2%
215	San Marino	3	1.3%
216	Sao Tome and Principe	4	3.2%
217	Saudi Arabia	4	2.2%
218	Senegal	4	2.3%
219	Serbia and Montenegro	3	2.3%
221	Seychelles	5	0.4%
222	Sierra Leone	4	2.3%
223	Singapore	3	1.4%
224	Slovakia	3	0.2%
225	Slovenia	3	-0.1%

ID	Country	R	λ
226	Solomon Islands	4	2.6%
227	Somalia	4	2.9%
228	South Africa	3	-0.4%
229	South Sandwich Islands	4	-0.4%
141	South Korea	4	0.4%
230	Spain	3	0.1%
231	Sri Lanka	4	0.8%
232	Sudan	4	2.6%
233	Suriname	4	0.2%
234	Svalbard	2	0.0%
235	Swaziland	4	-0.2%
236	Sweden	2	0.2%
237	Switzerland	2	0.4%
238	Syria	4	2.3%
239	Taiwan	2	0.6%
240	Tajikistan	4	2.2%
241	Tanzania	4	1.8%
242	Thailand	4	0.7%
243	Togo	4	2.7%
244	Tokelau	4	0.0%
245	Tonga	4	2.0%
246	Trinidad and Tobago	4	-0.9%

ID	Country	R	λ
248	Tunisia	4	1.0%
249	Turkey	3	1.1%
250	Turkmenistan	4	1.8%
251	Turks and Caicos Islands	4	2.8%
252	Tuvalu	4	1.5%
253	Uganda	4	3.4%
254	Ukraine	3	-0.6%
255	United Arab Emirates	4	1.5%
256	United Kingdom	2	0.3%
257	United States	2	0.9%
259	Uruguay	4	0.5%
260	Uzbekistan	4	1.7%
261	Vanuatu	4	1.5%
262	Venezuela	4	1.4%
263	Vietnam	4	1.0%
264	Virgin Islands	3	-0.1%
265	Wake Island	2	-0.1%
266	Wallis and Futuna	4	0.0%
267	Yemen	4	3.5%
268	Zambia	4	2.1%
269	Zimbabwe	4	0.6%