

# Risk Analytics for Estimating and Validating Magnitude of Earthquake Losses

A. Astoul, C. Filliter, A. Rau-Chaplin, K. Shridhar, B. Varghese<sup>1</sup> and N. Varshney  
*Risk Analytics Lab, Faculty of Computer Science  
Dalhousie University, Halifax, Canada*

**Abstract**—The support for data ingestion and data integration from multiple data sources for rapid loss estimation and visualisation is highly desirable in post-event catastrophe modelling systems. In this paper, the design and development of a distributed system for real-time estimation and visualisation of insured losses incurred due to earthquakes is presented. The system incorporates a model for estimating losses due to earthquakes in near real-time and a geo-browser for visualisation of hazard, exposure and loss data. Preliminary validation of the system is performed using as test cases six global earthquakes and their associated industry loss data. The key result is that the system can generate reasonable order of magnitude estimates with limited data.

**Keywords**—rapid data processing; rapid loss estimation; catastrophe losses; validation technique

## I. INTRODUCTION

Risk analytics is now playing an important role in making key decisions for managing disasters and financial instruments. Traditionally, such decisions were based on estimating losses for catastrophes which could occur in the future with respect to an assigned probability of occurrence. However, quick and imminent decision making were not possible as losses were estimated even before the catastrophe occurred.

Currently, sensor fusion technologies are making it possible to obtain up-to-date, timely and periodic catastrophe related data. For example, Shakemaps [1], one representation of earthquake hazard data in real-time is available. Real-time data can be employed to estimate losses in the minutes, hours, days and weeks immediately following a catastrophe. Examples of systems that perform risk analytics in real-time are Extremum [2], INLET - Internet-based Loss Estimation Tool [3] and PAGER - Prompt Assessment of Global Earthquakes for Response<sup>2</sup>.

Risk analytics in real-time is different from traditional methods since there is a focus on a single earthquake event that has occurred rather than a large catalogue of possible future events. Moreover, there is an evolving view of the event as it unfolds, and therefore data related to the event changes from time to time. For example, PAGER/ShakeMap

released fifteen alerts for the 2011 Tohoku earthquake that struck off the Pacific coast of Japan at 05:46 UTC on Friday, 11 March 2011 in time periods ranging from within an hour to six months after the earthquake. The first alert issued twenty three minutes after the event reported a magnitude 7.9 earthquake, and over the course of the first day alone four additional alerts were issued each updating the data available. In addition, in real-time risk analytics there is a need for rapid estimation of losses. Earthquake data alone is not always sufficient to produce loss estimates (for the Tohoku earthquake between 06:15 UTC and 07:52 UTC a tsunami struck the coastal towns). Additional data from multiple disparate sources are required for complete loss estimation.

This paper, presents a distributed server-client system that performs real-time risk analytics for estimating and validating losses in hours, days and weeks following an earthquake. The server system comprises a loss estimation module, a visualiser module and a database module, and the client system comprises the Google geo-browser on a light-weight visualiser. The loss estimation is based on PAGER and Shakemap for accessing real-time global earthquake data. The feasibility of the distributed system is demonstrated using the evolving view of a test event. The loss model employed in the distributed system is validated using a standard methodology for six global earthquakes. The key observation is that the distributed system produces reasonable order of magnitude losses.

The remainder of this paper is organised as follows. Section II proposes a distributed server-client architecture for earthquake loss estimation and visualisation. Section III presents the feasibility of the distributed system using the recent Tohoku earthquake data and a validation study of the system is pursued for six global earthquakes. Section IV concludes the paper.

## II. LOSS ESTIMATION AND VISUALISATION

The distributed client-server system (refer Figure 1) that facilitates loss estimation and visualisation allows decision makers to determine expected losses due to the occurrence of an earthquake and graphically display these losses. The estimated losses can be visualised at four geographic levels, namely country, state, county and city, on a geo-browser.

<sup>1</sup>Corresponding author. Email: varghese@cs.dal.ca. Web: <http://www.risk-analytics-lab.ca>

<sup>2</sup><http://earthquake.usgs.gov/earthquakes/pager/>

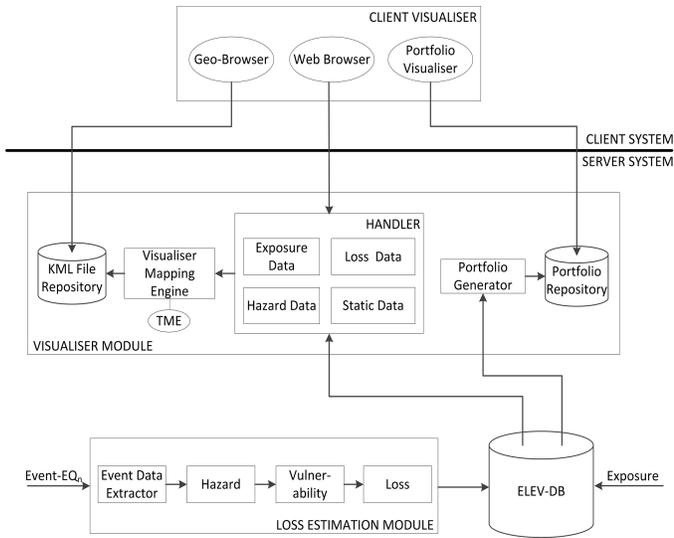


Figure 1: The distributed for estimating and visualising losses due to earthquakes

The country, state and county levels are sometimes referred to as regions, while the city level is referred to as both point and population centre. The server is composed of three primary modules, namely the Earthquake Loss Estimator and Visualizer Database (ELEV-DB), the Loss Estimation Module and the Visualiser Module. The ELEV-DB module is a collection of tables related to an event and geographic data.

ELEV-DB contains tables required by the loss estimation and visualisation modules. For example, industrial data for exposure, earthquake data, geographic information, and data generated by a Mean Damage Ratio (MDR) model [4] along with data generated by the modules are stored in ELEV-DB [5].

The Loss Estimation Module comprises sub-modules, namely the Event Data Extractor, Hazard, Vulnerability and Loss modules. ShakeCast Lite [6] is used as the alerting system to notify the Event Data Extractor of an earthquake event,  $EQ_n$  that has just occurred or is unfolding. The real-time data of the earthquake is obtained from the Prompt Assessment of Global Earthquakes for Response (PAGER). The Hazard module produces the measure of severity of an earthquake, otherwise referred to as the Modified Mercalli Intensity (MMI), in a city and region. The MMI values are then used by the Vulnerability module to produce MDR values. These values are used by the Loss module along with the exposure to generate losses. The numerical computation used in the model can be obtained from the preliminary research presented in [5].

The Visualiser Module comprises the Data Handler, the Mapping Engine, the Portfolio Generator and two repositories. The Data Handler is as an interface between client requests and the data available for visualisation that is stored

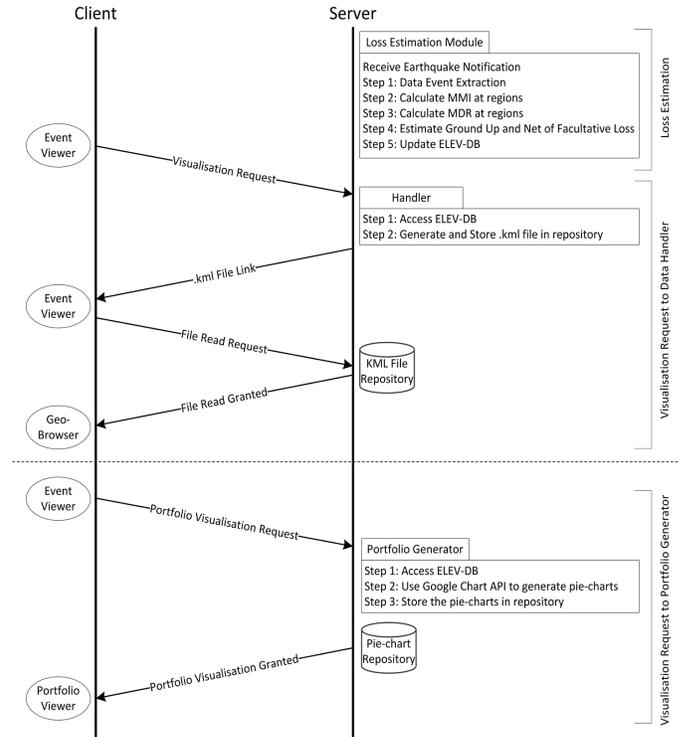


Figure 2: Interaction between client and server system

in the database. Four handlers are available, namely the Exposure Data Handler, the Hazard Data Handler, the Loss Data Handler and the Static Data handler. The Exposure Data Handler retrieves the exposure for different geographic levels. The Loss Data Handler retrieves the losses for different geographic levels. The Hazard Data Handler retrieves MMI and MDR for different geographic levels. Static Data Handler retrieves geography-specific indicators.

The Mapping Engine receives data from the handlers and facilitates the visualisation of data on the client system. It is built on the Thematic Mapping Engine (TME) [7] and generates .kml files. The KML File Repository stores the .kml files generated by the mapping engine. The Portfolio Generator is built on the Google Chart API and presents a comparison of losses and exposures as pie-charts.

The client system is a client visualiser that consists a geo-browser (Google Earth is employed in this research), an Event Viewer and a Portfolio Viewer.

#### A. Communication Sequence

Figure 2 is the illustration of interactions between the client and server modules. The Loss Estimation Module executes Step 1 to Step 5 after it receives an earthquake notification, and finally stores loss values in the database.

The client system can raise two type of visualisation requests, those to the Data Handler and to the Portfolio Generator. A visualisation request to the Data Handler is

made by the Event Viewer. Based on the type of data that needs to be visualised, the Exposure, Loss, Hazard or Static Data Handlers are invoked. The handler retrieves data from ELEV-DB and a .kml file is generated in the KML File Repository. The Event Viewer after receiving a .kml file link requests to read the file and is accessed by the geo-browser on the client system.

A visualisation request to the Portfolio Generator again retrieves loss and exposure data from ELEV-DB. The Google chart API is used to generate pie-charts in a repository. The Portfolio Viewer can then access the pie-charts on the client system.

### III. EXPERIMENTAL STUDIES

The feasibility and validity of the distributed system is considered in this section. The feasibility is confirmed using a test event with an ID USC0001XGP in PAGER. The validation study considers six global earthquakes and the estimated losses are compared against normalised historic loss data.

The earthquake data was available on the PAGER archive<sup>1</sup> and ShakeMap archive<sup>2</sup>. Geometry data for the geographic levels was obtained from the Global Administrative Areas Database<sup>3</sup>, as shapefiles. The boundary specifications were simplified using the MapShaper tool [8].

#### A. Feasibility Study

The test event is of magnitude 9.0, which occurred in Tohoku, Japan on 11 March 2011 that struck off the Pacific coast of Japan at 05:46 UTC on Friday, 11 March 2011. This recent earthquake was a major catastrophe and affected 28 prefectures.

Figure 3 shows screenshots of different alert versions,  $A_1 - A_{15}$  of the earthquake. The figure shows the evolving view of the earthquake and how losses can be rapidly estimated. The MMI of the affected prefectures are shown using choropleth visualisation technique and the height of the prisms are indicative of the Ground Up losses.

#### B. Validation Study

A study that compares the predicted losses of six global earthquakes against historic loss data was pursued in order to validate the loss estimation system. Table I shows the list of earthquakes selected for this study, their date of occurrence (dd-mm-yyyy), magnitude, latitude and longitude, historic losses in millions of USD in the year of occurrence of the earthquake, adjustment multipliers to normalise the historic losses to 2012 USD, and predicted losses in millions of USD. The earthquakes were selected such that (a) they are from different regions, (b) their magnitude were over 5.5, (c) and had occurred in the last 30 years.

<sup>1</sup><http://earthquake.usgs.gov/earthquakes/pager/archives.php>

<sup>2</sup><http://earthquake.usgs.gov/earthquakes/shakemap/list.php?n=sc&y=2010>

<sup>3</sup><http://www.gadm.org/>

The historic data related to all the earthquakes were collected from multiple sources, namely the National Geophysical Data Centre (NSDC)<sup>4</sup>, United States Geological Survey (USGS)<sup>5</sup>, PAGER<sup>1</sup>, ShakeMap<sup>2</sup>, EM-DAT<sup>6</sup> and CAT-DAT [9]. The information collected includes, event data, exposure data, hazard data and loss data. The collected loss data is denoted as  $D_y$  which are in USD of year  $y$  in which the earthquake occurred.

The historic loss data is normalised to 2012 USD, denoted as  $D_{2012}$  using the normalisation method described in [10] and [11]. Three adjustment multipliers are used for the normalisation. Firstly, the Inflation Multiplier, denoted as  $IPD_y$ , which uses the Implicit Price Deflator (IPD) for Gross Domestic Product, and is available from Economic Research of the Federal Reserve Bank of St. Louis<sup>7</sup> and the US Bureau of Economic Analysis<sup>8</sup>. Secondly, the Population Multiplier, denoted as  $\Delta P_{2012-y}$ , which is the ratio of the population in 2012 and the year of occurrence of the earthquake. Thirdly, the Wealth Multiplier, denoted as  $W_y$ . Inflation-corrected wealth adjustment, denoted as  $ICW_y$  for year,  $y$  is performed ( $ICW_y = \frac{\text{Ratio of wealth of 2012 to } y}{IPD_y}$ ) before  $W_y$  is computed as  $W_y = \frac{ICW_y}{\Delta P_{2012-y}}$ . The normalisation equation is

$$D_{2012} = D_y \times IPD_y \times W_y \times \Delta P_{2012-y}$$

PAGER data (MMI at city level, affected cities due to an earthquake) for global earthquakes are only available after 2007. Therefore, for earthquakes prior to 2008 data was extracted from (i) a list of cities whose population is greater than one thousand<sup>9</sup>, and (ii) the ShakeMap file which is a representation of the affected grid on a map due to an earthquake and comprises a large set of point data (latitude, longitude and the MMI at that point). The list of cities that are affected within the grid and their MMIs are extracted, and the cities are mapped onto their respective regions using the latitude and longitude information. The exposure data for the geographic levels are collected from publicly available sources. The losses are then computed using the numerical computation in [5].

A column chart (refer Figure 4) was generated based on increasing historic losses. In the chart, the predicted and historic losses are shown in USD on the logarithmic scale for all events shown in Table I.

The loss predicted by the distributed system is a mean value for an earthquake. To study the probability of a loss threshold  $(a, b)$  the  $\phi$  distribution which is the standard normal cumulative distribution function is employed as

<sup>4</sup><http://www.ngdc.noaa.gov/>

<sup>5</sup><http://www.usgs.gov/>

<sup>6</sup><http://www.emdat.be/>

<sup>7</sup><http://research.stlouisfed.org/>

<sup>8</sup><http://www.bea.gov/>

<sup>9</sup><http://www.geonames.org/>

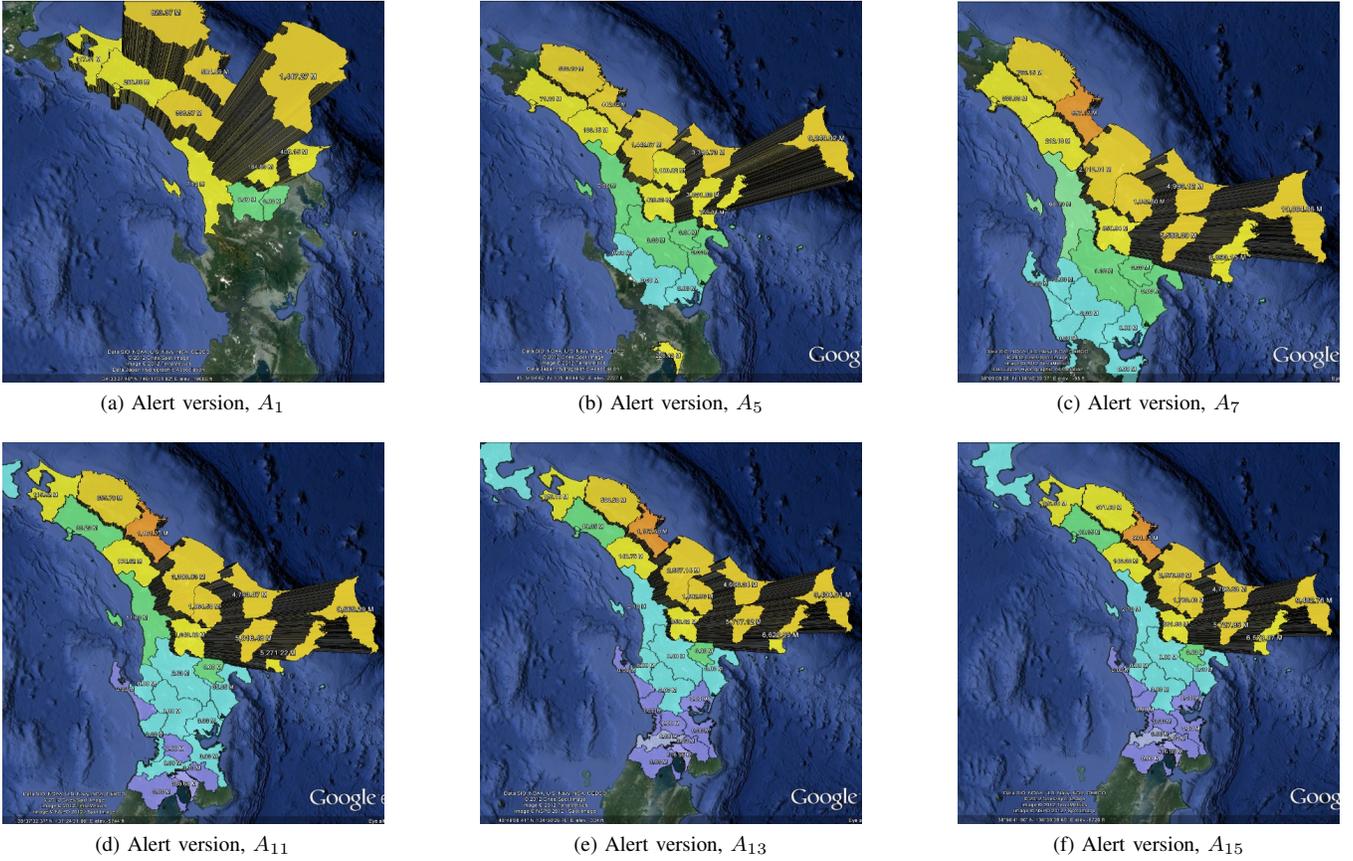


Figure 3: Screenshots of alert versions  $A_1, A_5, A_7, A_{11}, A_{13}$  and  $A_{15}$  of magnitude 9.0, Tohoku, Japan, 11 March 2011 earthquake

follows [4]:

$$P(a < L \leq b) = \phi \left[ \frac{\ln(b) - \mu_{\ln(L)}}{\zeta} \right] - \phi \left[ \frac{\ln(a) - \mu_{\ln(L)}}{\zeta} \right]$$

where  $\mu_{\ln(L)}$  is the predicted value of the logarithm of loss obtained from the system and is assumed to be a lognormal random variable, and  $\zeta$  is the normalised standard deviation of the logarithm of loss obtained from [12].

Figure 5 shows the estimate of probability of different loss thresholds ( $0 < 11$ ,  $1 < 10$ ,  $10 < 100$ ,  $100 < 1,000$ ,  $1,000 < 10,000$ ,  $10,000 < 100,000$ ,  $100,000 < 1,000,000$ ) represented in millions of USD for the earthquakes of Table I.

In this section we have evaluated the performance of the distributed loss estimation system in terms of how well its data acquisition and visualization facilities are able to capture the evolving history of earthquake alerts and the performance of its simplistic loss model. The Tohoku earthquake used in evaluating the feasibility demonstrates how data can be rapidly ingested from multiple sources to visualise earthquake alerts as the data related to the event evolves over hours, days and months after its occurrence.

Evaluation of loss systems is tricky at best due to the inherent difficulty in collecting consistent exposure and loss data for historic events. In the case of the distributed model the goal is to rapidly produce a crude loss estimate of a global event as it evolves, often based on very limited information. In this context, the distribution of expected losses is much more important than the point estimates. Our validation demonstrated that the methodology pioneered in PAGER [4] for economic loss can be usefully applied in the context of portfolio losses.

In 66% of our evaluation events the observed historical losses and the predicted losses fall into the same loss threshold. In 100% of our test events the observed historical losses and the predicted losses fall into the two highest loss thresholds. Given the limited data, the loss system gives reasonable order of magnitude estimates, but it is important that users be aware of the inherent limitations of the underlying approach.

#### IV. CONCLUSION

Often in the timeline of an earthquake data provided by sources such as PAGER/ShakeMap evolves over time.

Country	Date	Mag	Historic Losses in millions of USD for year $y$ , $D_y$	Adjustment Multipliers				Normalised Historic losses (NHL) in millions of 2012 USD, $D_{2012}$	Predicted Losses (PL) in millions of 2012 USD
				Inflation Multiplier, $IPD_y$	Inflation-corrected Wealth Multiplier, $ICW_y$	Wealth Multiplier, $W_y$	Population Multiplier, $\Delta P_{2012-y}$		
USA	09/01/2010	6.5	25.00	1.0352	0.9850	0.9683	1.0172	25.49	16.86
USA	18/10/1989	6.9	2,510.00	1.6103	1.2119	0.9525	1.2724	4,898.69	7,316.61
New Zealand	13/06/2011	6.0	2,816.45	0.9909	1.0165	1.0099	1.0066	2,836.79	3,132.12
New Zealand	21/02/2011	6.1	13,000.00	1.0025	1.0047	0.9976	1.0070	13,093.86	17,660.64
USA	17/01/1994	6.7	22,920.00	1.4381	1.1106	0.9204	1.2066	36,606.39	4,787.64
Japan	11/03/2011	9.0	37,200.00	0.9935	0.9978	0.9873	1.0106	36,877.45	4,611.44

Table I: Earthquakes used as test cases in the validation study

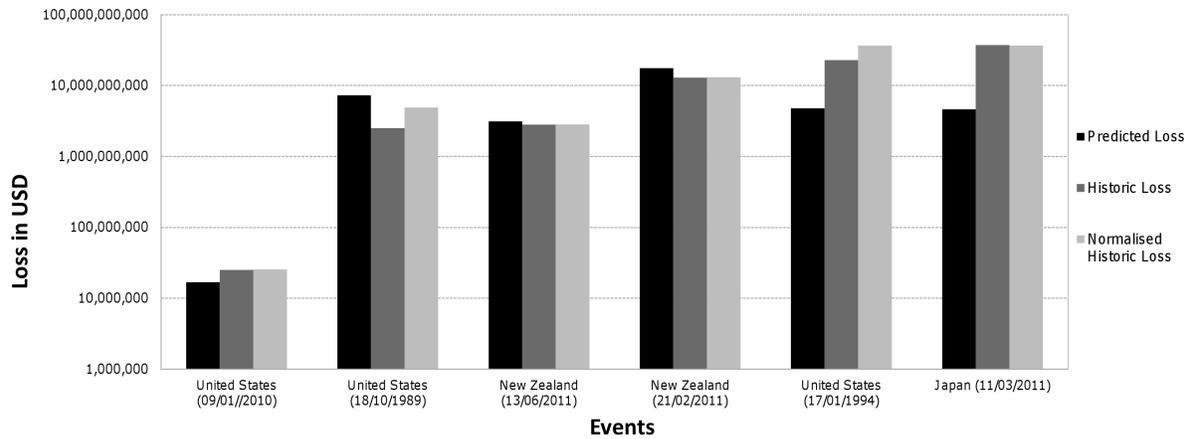


Figure 4: Column charts for predicted, historic and normalised historic losses for earthquakes shown in Table I.

For example, data for the Tohoku earthquake was updated fifteen times ranging from within an hour to six months after the earthquake. The data available initially is not sufficient to produce reliable loss estimates. Therefore, post-event analysis is challenging and important to generate reliable loss estimates.

For a post-event earthquake system to be useful in days and weeks after the event, it needs to support rapid data ingestion, loss estimation, visualisation and integration of hazard, exposure, and loss data from multiple sources. This paper has presented the design and development of such a distributed system, comprising server and client modules. The server comprises a loss estimator, a visualiser and a database, and the client is a geo-browser for visualisation. The loss estimation model incorporated in the system relies on multiple data sources for accessing real-time earthquake data. The recent Tohoku earthquake is used as a test event to demonstrate the feasibility of the distributed system and how an evolving view of the event is generated using the

system. The system is validated using a set of six global earthquakes. The results indicate that the system generates reasonable order of magnitude loss estimates.

Future work will aim to refine the system by calibrating the PAGER vulnerability curves (for economic losses) for a more accurate use in portfolio insured loss models. Efforts will also be made towards augmenting the loss model results with any available historical data points.

## REFERENCES

- [1] D. J. Wald, B. C. Worden, V. Quitoriano and K. L. Pankow, "ShakeMap Manual: Technical Manual, User's Guide, and Software Guide," U.S. Geological Survey, Version 1, 2006.
- [2] N. Frolova, V. Larionov and J. Bonnin, "Earthquake Casualties Estimation in Emergency Mode, Human Casualties in Earthquakes," Advances in Natural and Technological Hazards Research 29, R. Spence, R. and C. Scawthorn, (Editors), Springer, 2011, pp. 107-124.

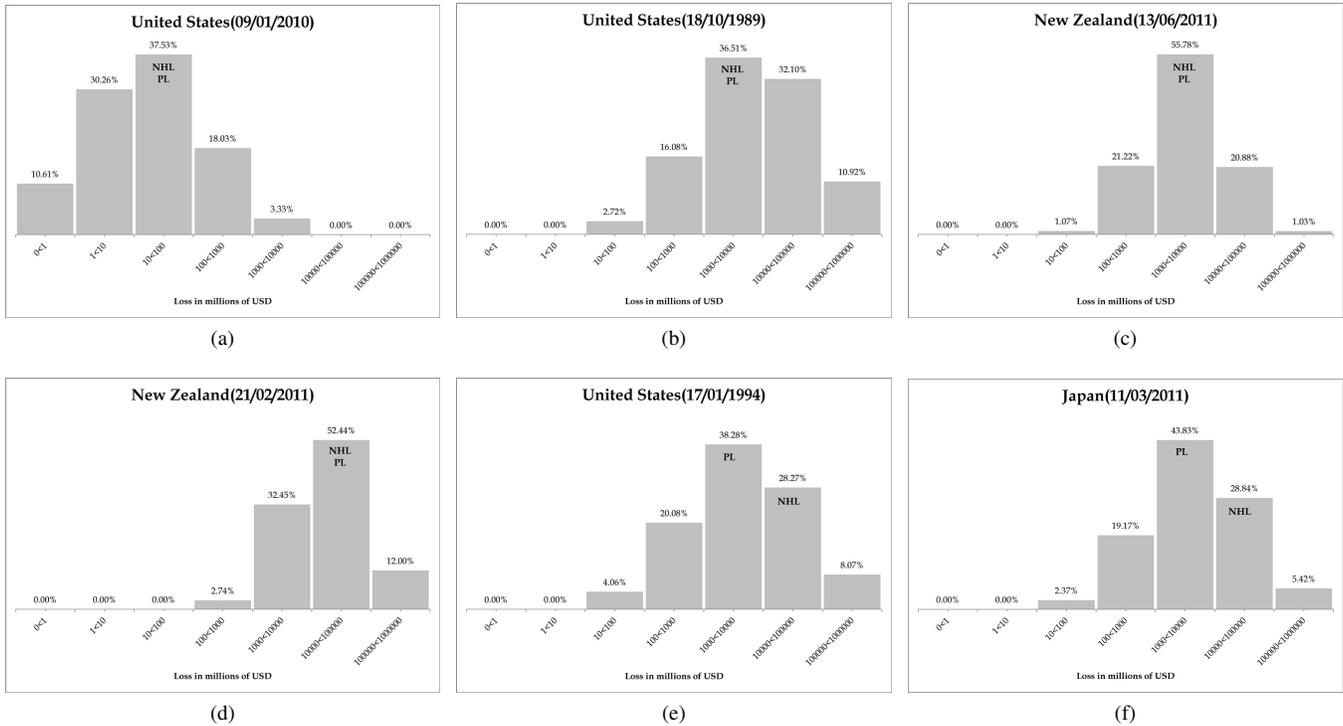


Figure 5: Probability of loss thresholds for earthquakes of Table I. The threshold between which the Normalised Historic Loss (NHL) and Predicted Loss (PL) falls is indicated for each earthquake.

[3] C. K. Huyck, H. -C. Chung, S. Cho, M. Z. Mio, S. Ghosh, R. T. Eguchi and S. Mehrotra, "Centralized Web-based Loss Estimation Tool: INLET for Disaster Response," Proceeding of SPIE 6178, 61780B, 2006.

[4] K. Jaiswal and D. J. Wald, "Rapid Estimation of the Economic Consequences of Global Earthquakes," U.S. Geological Survey Open-File Report 2011-1116, 2011.

[5] E. Mason, K. Shridhar, A. Rau-Chaplin, B. Varghese and N. Varshney, Rapid Post-Event Catastrophe Modelling and Visualization, Proceedings of the Workshop of the 23rd International Conference on Database and Expert Systems Application (DEXA), Vienna, Austria, 2012.

[6] D. Wald, K. -W. Lin, K. Porter, and L. Turner, "Shake-Cast: Automating and Improving the Use of ShakeMap for Post-Earthquake Decision-Making and Response," Earthquake Spectra, Volume 24, No. 2, 2008, pp. 533-553.

[7] B. Sandvik, "Thematic Mapping Engine, Part 2: Supporting Document," MSc dissertation in Geographical Information Science, Institute of Geography, School of Geosciences, University of Edinburgh, 2008.

[8] M. Harrower and M. Bloch, "MapShaper.org: A Map Generalization Web Service," IEEE Computer Graphics and Applications, Volume 26, Issue 4, 2006, pp. 22-27.

[9] J. E. Daniell, B. Khazai, F. Wenzel and A. Vervaeck, "The CATDAT Damaging Earthquakes Database," Natural Hazards Earth System Science, Vol. 11, 2011, pp. 2235-2251.

[10] R. Pielke Jr., J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders and R. Musulin, "Normalized Hurricane Damage in the United States: 1900-2005," Natural Hazards Review, 2008, pp. 29-42.

[11] K. Vranes and R. Pielke Jr., "Normalized Earthquake Damage and Fatalities in the United States: 1900-2005," Natural Hazards Review, 2009, pp.84-101.

[12] Appendix-I of Rapid Estimation of the Economic Consequences of Global Earthquakes, U.S. Geological Survey Open-File Report 2011-1116: <http://pubs.usgs.gov/of/2011/1116/> [Last checked: 29 March, 2013]