*17th World Conference on Earthquake Engineering, 17WCEE*

*Sendai, Japan - September 13th to 18th 2020*

# SPATIAL CORRELATION OF EARTHQUAKE LOSS AND ITS IMPACT ON LOSS AGGREGATION

Z. Wang[1,2]

[1] *Distinguished Professor, School of Civil and Architectural Engineering, Henan University, zifa@iem.ac.cn*

[2] *Research Professor, Institute of Engineering Mechanics, China Earthquake Administration, zifa@iem.ac.cn*

…

## *Abstract*

Earthquake loss estimation is a complex process where numerous variables contribute to the uncertainty of loss estimation. The variables include seismic source information such as location and magnitude of the earthquakes, the ground motion prediction equations used, the effect of site conditions and the impact of structural vulnerability. In addition, when loss estimation for multiple locations is required, the correlation of loss estimation between different locations is needed to aggregate the loss together. Research on loss correlation in the past has been focused on spatial correlation of ground motion rather than the loss due to the lack of a large amount of detailed loss information from a single event. This approach cannot account for the impact of structural vulnerability, which means that it can only assess the partial impact of spatial correlation of earthquake loss. In the insurance industry, for simplicity the correlation is sometimes assumed to be a constant, independent of distance between the two locations in consideration, which is not true based on theoretical analysis as well as actual loss experience. This paper analyzes the detailed loss information collected after the 2011 Tohoku Earthquake, and using an advanced algorithm called kriging, a distance-dependent formula for spatial correlation of earthquake loss is derived. According to the formula, the correlation decays exponentially over distance between the two locations. The shorter the distance, the higher the correlation. The derived formula is validated against the actual loss information and the validation showed excellent agreement between the derived results and the actual loss information. The distance dependent formula is then implemented in the earthquake loss estimation model for Japan. Based on the loss results for a typical exposure portfolio in Japan, compared against the approach that utilizes a constant for spatial correlation, the distance dependent loss correlation produces a loss estimate that can be 50% or more different for amounts whose return periods are higher than 100 years in Japan, which could have significant impact on capital requirement in the insurance and reinsurance industry..

*Keywords: Spatial correlation; Earthquake loss; Loss aggregation; Earthquake risk modeling; Distance dependency*

## 1. Introduction

There are generally two approaches in estimating earthquake loss. One is deterministic and the other is probabilistic. Although both the mean and the standard deviation of the estimated loss are often provided even in the deterministic approach, loss uncertainty of the estimation is usually associated with the probabilistic approach when stochastic events are used. Uncertainty of the earthquake loss estimation is especially important in quantifying tail risk, when distribution of the loss estimation is used to estimate the tail risk of earthquake event loss.

There are two types of uncertainty. One is epistemic and the other is aleatory [1]. As explained in detail in [1], epistemic uncertainty relates to the incompleteness of inputs, the variation of available data, and the limitation of the method and techniques used. There are a number of methods that can be used to describe this type of uncertainty in earthquake ground motion, such as the logic tree method [2], and the Gaussian copula approach [3,4]. Aleatory uncertainty is used to describe the systemic difference between estimation and actual loss. In other words, it is related to the method itself that does not include all the necessary parameters and components in the loss estimation. This type of uncertainty is often described using the mean and standard deviation of the estimated results, with different variations in the description detail [5]. It should be noted that in practice it is both difficult and unnecessary to clearly separate these two types of uncertainty and they can sometimes be described by the standard deviation of the final estimation.

In a typical framework of earthquake loss estimation [6], uncertainty is inherent in every component in the framework, from event occurrence, ground motion attenuation, site effect, structural vulnerability, financial term application and the aggregation of earthquake loss estimation. In a typical scenario in earthquake loss estimation, exposure can be from many locations when each location could have multiple coverages, such as building, content and business interruption or time element coverages. When uncertainty is included in the loss estimation process, dependency of the loss estimation for each component, each coverage at many locations needs to be considered. Dependency describes the correlation between the loss estimation, and it is very important when the estimated loss needs to be aggregated together to derive the tail risk for the earthquake event impacting a large area. As in the case for uncertainty, dependency needs to be considered for all the components in the framework for earthquake loss estimation, as in ground motion estimation, structural vulnerability and others.

An ideal approach to addressing the dependency is to use a nested approach that encompasses all the correlation effect in the loss estimation framework from earthquake occurrence to final loss aggregation as described in the framework process in [6], but most of the research results have been focused on ground motion correlation [1,7], partly because of the lack of actual loss information and the availability of a large amount of observed ground motion records in recent large earthquakes. The correlation of earthquake ground motion is affected by a few factors such as the earthquake source (magnitude and fault), attenuation (propagation path), the local site effect (soil layering and property), and the distance between locations. Of all the factors affecting the dependency or correlation, distance effect is the most important because distance correlation reduces dramatically as distance increases, therefore most of the correlation studies for earthquake ground motion have been focused on distance effect, that is, the spatial correlation effect of ground motion[1,7].

Since distance correlation impact is the most important factor in quantifying earthquake ground motion uncertainty, it can be easily deduced that it is also one of the most important factors in quantifying earthquake loss estimation uncertainty. As explained in the above, the ideal approach to quantifying the correlation effect is to use a nested approach to account for the impact from all the components in the earthquake loss estimation process, but this approach is infeasible in practice because of a number of reasons, such as the complexity of the quantification process, the unreal demand for computation power, and the lack of supporting data to better describe the distance dependency impact in all factors. Although the study on ground motion distance correlation has been widely conducted and a number of interesting results have been published, it is only addressing its impact from ground motion, while the impact from all other components still remains unknown. To consider the overall impact of distance dependency in earthquake loss estimation considering all the factors, instead of using the nested convolution approach described above, this paper derives the actual distance correlation of earthquake loss based on the large amount of observed loss information from the March 11, 2011 Great East Japan Earthquake, and using this correlation coupled with the impact of structural vulnerability correlation to address the overall impact of distance dependence on earthquake loss estimation.

## 2. Simulating Distance Dependency in Location Level Loss Estimation

In the traditional framework for earthquake loss estimation of stochastic events [6], correlation has to be assumed when aggregating loss from different locations, and more often a constant is used without considering the impact of distance, and this constant of correlation is usually between 0.15-0.25, with bigger constant for smaller footprint of an earthquake. As demonstrated in following sections, this assumption is generally acceptable for large portfolio-level loss estimation when exposure is evenly distributed in space, but there are two major issues in this assumption. One is that this assumption does not work when the exposure has concentrations spatially, which is usually true because of the exposure concentration difference between urban and rural areas. The other issue is that this assumption does not work well when there are only a few locations in consideration where distance impact is more significant. Additionally, there is another major issue in implementing the distance correlation in loss aggregation within the traditional framework. In a traditional framework, loss aggregation has to be repeated multiple times from location coverage, location, account to portfolio level when correlation has to be assumed each time when aggregation happens. Because of the nonlinear behavior of most distribution assumptions, especially in the case of using a beta-function as the distribution function, distribution attributes do not follow the same pattern when aggregated at different levels considering the impact of distance dependency. To address this issue when implementing an improved algorithm to consider the effect of distance dependency, it is important to move the uncertainty simulation from portfolio level in a traditional framework to location level. When loss is simulated at location level considering the effect of distance dependency, aggregation can follow through to higher levels without further assumption of distribution at different levels. Therefore, our major objective is to simulate location level loss using marginal distribution with given location level spatial correlation.

As explained in the previous section, to avoid the complex nested effect of correlation dependency, simulating loss directly based on the spatial correlation of actual loss data at location level is the most straightforward approach. At the location level, two additional layers need to be considered in additional to distance dependency. One is the correlation between structural vulnerability and the other is the correlation between different coverages at the same location, or in the case of location level simulation, within the same location grid, as shown in Fig. 1.

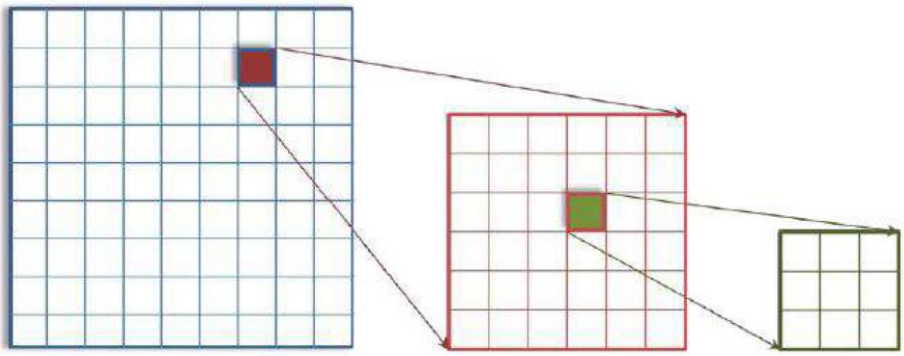


Fig.1 Nested hierarchical correlation matrix at location level



As shown in Fig. 1, this is clearly a multivariate sampling example, and Gaussian copula model can be easily adapted to address this issue. The application of Gaussian copula model in addressing the spatial correlation of earthquake ground motion can be found in [2], and a similar approach can be used in simulating location level loss considering the distance dependency effect. There are two major challenges in applying this approach to simulate earthquake location loss for many locations with nested correlation effect of distance, structural vulnerability and location coverage dependency. They are addressed and explained separately below.

The first challenge is the impact of a large number of locations. When Gaussian copula model is used to sample multivariate random variables that have known marginal distributions (beta-function distribution in our case) with given correlation structures, we need to first sample a multivariate normal random variable, and then transform it into a mathematical way to construct the desired marginal distributions. When the dimension of multivariate variable is low, samples can be drawn directly using this approach. However, when the dimension is high such as in the case of earthquake loss estimation when the number of locations can be easily over a few millions, computation demand increases at the order of $N^3$, where N is the number of locations. It is computationally impossible to directly sample the location level loss at millions of locations using the Gaussian copula model, and in this case, we need to find a way to effectively address this issue. We have proposed to generate samples from a small number of locations and use the location-level loss in these locations to interpolate the losses at much greater number of locations based on the spatial Kriging interpolation method [8]. Kriging belongs to the family of linear least squares estimation algorithms. The aim of kriging is to interpolate/estimate the value of an unknown real-value function f at a location $x^*$ , given the known values of the function at some other locations $x_1, \ldots, x_n$. A kriging estimator is said to be linear because the predicted value is a linear combination that may be written as:

$$f(x^*) = \sum_{i=1}^{n} w_i(x^*) f(x_i) \tag{1}$$

The weights $w_i$s are solutions of a system of linear equations which are obtained by assuming that f is a sample-path of a random process F(x), which is a Gaussian process in our case. More detailed implementation of a kriging approach can be found in [8].

The second challenge is nested effect of both structural vulnerability and location coverage. As shown in Fig. 1, we have other correlations (structural vulnerability and location coverage) nested with spatial correlation. Given that location coverage is always at the same location for the same building, it is acceptable to assume that they are completely dependent. Therefore, in our research we only need to consider vulnerability correlation that reflect the correlation between buildings with different building characteristics. There are typically over 20 basic vulnerability curves for any given earthquake event. This additional nested layer of correlation makes the ultimate correlation matrix (considering both distance and vulnerability) between location losses much bigger. For example, if we only look at a 20× 20 grid with only spatial correlation considered, we will have a 400× 400 spatial correlation matrix for the 400 grid points. However, if add in 20 basic vulnerability curves, we ultimately have an 8000× 8000 huge correlation matrix for all combinations of distance and vulnerability types. To solve this technical issue, we need to introduce the Kronecter product representation of nested correlation matrices. With the Kronecker product representation of the huge ultimate matrix combing both spatial and vulnerability correlation matrix, instead of sampling for the Np × Np huge matrix (N is the number of locations and p is the number of vulnerability curves), we only need to sample for the N×N spatial correlation matrix and the p×p vulnerability matrix separately, and then a Kronecker product of the these two returns the sampling results with decomposition for the huge product matrix. For a short introduction of the Kronecker product representation, please refer to http://en.wikipedia.org/wiki/Kronecker_product for details. Considering the $O(n^3)$ computational complexity of Cholesky decomposition, the speed improvement is huge after the decomposition.

## 3. Distance Dependency based on Actual Loss Data

Due to the lack of a large amount of detailed earthquake loss information from a single earthquake event, little progress has been made on studying distance correlation of earthquake loss. The earthquake catastrophe risk management company RMS has derived an empirical table for earthquake loss correlation over distance as shown in Table 1, which was subsequently applied in estimating earthquake loss for the insurance industry [9].

Table 1 Correlation coefficient of earthquake loss over distance

| D(km) | $m_\rho$ | $\sigma_\rho$ | D(km) | $m_\rho$ | $\sigma_\rho$ |
|---|---|---|---|---|---|
| 0-5 | 0.986 | 0.0119 | 100-110 | 0.129 | 0.1726 |
| 5-10 | 0.934 | 0.0679 | 110-120 | 0.118 | 0.1615 |
| 10-15 | 0.863 | 0.1050 | 120-130 | 0.0870 | 0.1440 |
| 15-20 | 0.774 | 0.1331 | 130-140 | 0.0674 | 0.1171 |
| 20-25 | 0.671 | 0.1679 | 140-150 | 0.0527 | 0.1052 |
| 25-30 | 0.606 | 0.1743 | 150-180 | 0.0454 | 0.1050 |
| 30-35 | 0.520 | 0.2016 | 180-210 | 0.0246 | 0.0777 |
| 35-40 | 0.471 | 0.2226 | 210-240 | 0.0124 | 0.0302 |
| 40-45 | 0.394 | 0.1966 | 240-270 | 0.00850 | 0.0203 |
| 45-50 | 0.392 | 0.2154 | 270-300 | 0.00359 | 0.0104 |
| 50-60 | 0.343 | 0.2228 | 300-330 | 0.00148 | 0.00533 |
| 60-70 | 0.291 | 0.2210 | 330-360 | 0.000579 | 0.00230 |
| 70-80 | 0.229 | 0.1968 | 360-390 | 0.000190 | 0.00135 |
| 80-90 | 0.179 | 0.1978 | 390-420 | 0.000028 | 0.000231 |
| 90-100 | 0.158 | 0.1971 | 420-450 | 0.000008 | 0.000071 |

After the March 11, 2011 Great East Japan Earthquake, a big amount of loss data has been obtained, and this provides us with a unique opportunity to revisit the issue of distance dependency of earthquake loss. For this event, we have collected loss data from more than 580,000 locations, and they are distributed all over the northern part of Japan, as shown in Fig. 2.

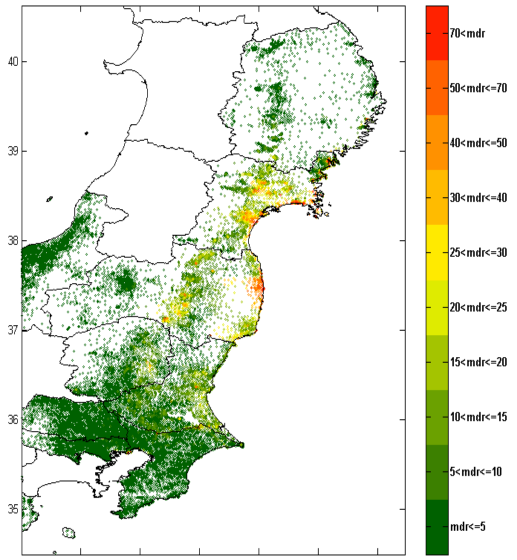



Fig. 2 Earthquake loss information from the March 11, 2011 event in Japan

Using the information from Table 1 and Fig. 2, we can derive a formula that describes the distance dependency as shown in Fig. 3.

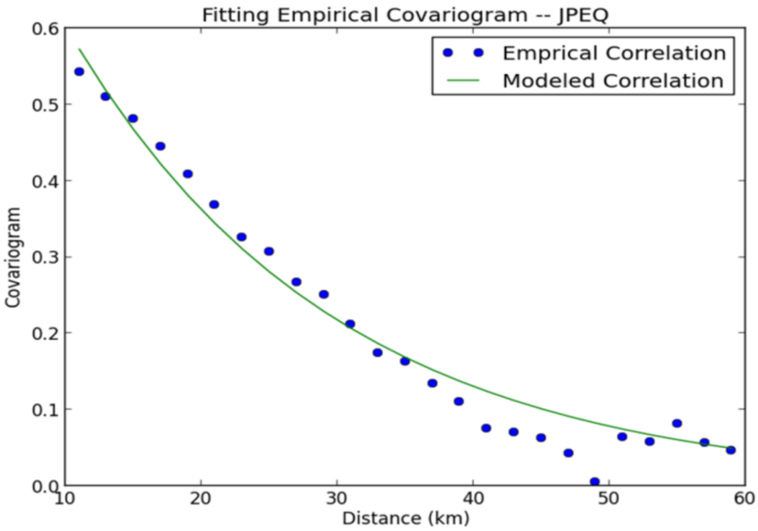



Fig. 3 Distance dependency of earthquake loss

The curve in Fig. 3 can be expressed in the following equation where the coefficient λ is 0.05.

$$\rho(x_1,x_2) = \exp\left(-\lambda \| x_1 - x_2 \|_1\right) \tag{2}$$

where $x_1$ and $x_2$ are two locations under consideration.

Now that we have the distance dependency formula as expressed in Eq. (2) and Fig. 3, we need to verify if kriging approach works well. We first generate a 10,000 grid footprint for an earthquake, and select a 200 points sub-grid from this 10,000 grids. Gaussian copula is applied at the 200 points sub-grid level, and then kriging is applied to all other grid points. Finally we normalize the sampled results so that variance is similar at every grid point. In order to verify if this approach works, we randomly selected 1000 points from the true sampled results without kriging and compare the kriging results against the true sampled results, as shown in Fig. 4.

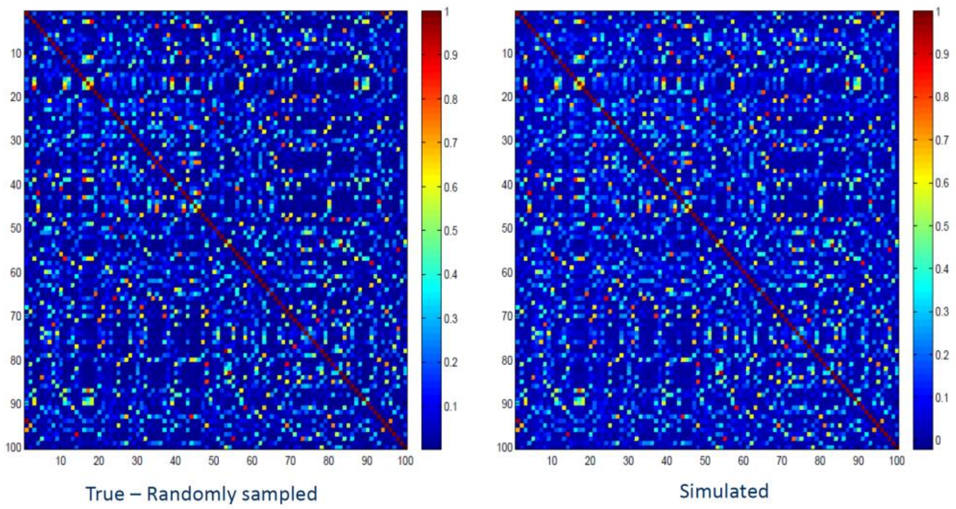



Fig. 4 Comparison of truly sampled results against simulated ones using kriging approach



The error distribution of truly sampled against simulated is shown in Fig. 5.

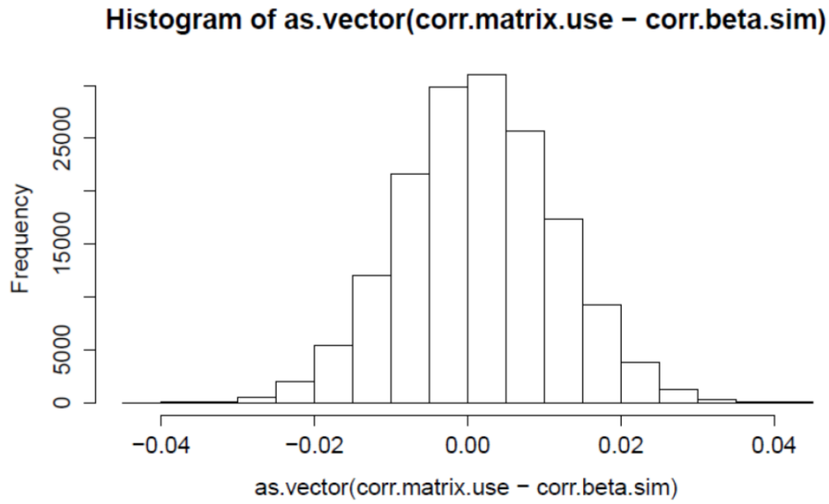



Fig. 5 Histogram of error distribution

As can be seen from both Fig. 4 and Fig. 5, the kriging approach works well in interpolating sampled results at grid points not sampled using the Gaussian copula approach.

## 4. Impact of Distance Dependency on Earthquake Loss Estimation

We plan to use two examples to study the impact of distance dependency in earthquake loss estimation. As explained in previous sections, distance dependency impacts the estimate of variance or standard deviation, thus affecting the estimate of tail risk for an earthquake event. As known in the insurance industry, tail risk is very important in risk quantification, risk management and capital allocation, it is essential to accurately estimate the mean as well as the variance or standard deviation to better quantify the distribution of earthquake loss estimate.

The first example is a hypothetical portfolio with different types of exposure distribution or concentration, as shown schematically in Fig. 6.

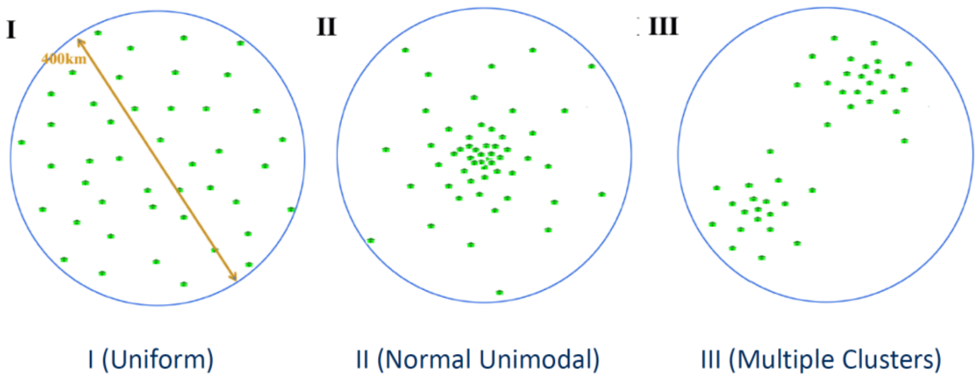



Fig. 6 Three types of exposure distribution where the blue points are exposure locations and the circle is earthquake footprint



In this example, the footprint diameter is 400km, and all exposure locations are within the footprint. There are 20,000 locations which are all within the footprint. We compared the standard deviation between the constant assumption (0.2) and our kriging approach using distance dependent correlation, as shown in Table 2.

Table 2 Comparison of standard deviation for ground-up loss

| | Type I | Type II | Type III |
|---|---|---|---|
| Constant correlation (0.2) | 2.42e-3 | 2.42e-3 | 2.42e-3 |
| Distance dependent | 2.44e-3 | 3.61e-3 | 2.76e-3 |
| Ratio | 1.008 | 1.492 | 1.140 |

As can be seen from Table 2, constant correlation assumption generally reduces the standard deviation, thus resulting a tail risk smaller than that from the distance dependent correlation. It should also be pointed out that for exposure evenly distributed within the footprint, as in Type 1, there is little difference between constant correlation and distance dependent correlation.

Additional study demonstrated that when the footprint is much smaller than the exposure coverage area size, the resulting standard deviation of distance dependent correlation will be much smaller compared against that from the constant correlation assumption. In other words, the constant dependency assumption overestimates the tail risk.

The second example is based on the study of an actual portfolio. This is a residential portfolio in Japan. The first case is the nationwide residential portfolio in Japan, as shown in Fig. 7(a). The Probable Maximum Loss (PML) comparison between distance dependent correlation and constant correlation (different levels of constant assumption) is shown in Fig 7(b).

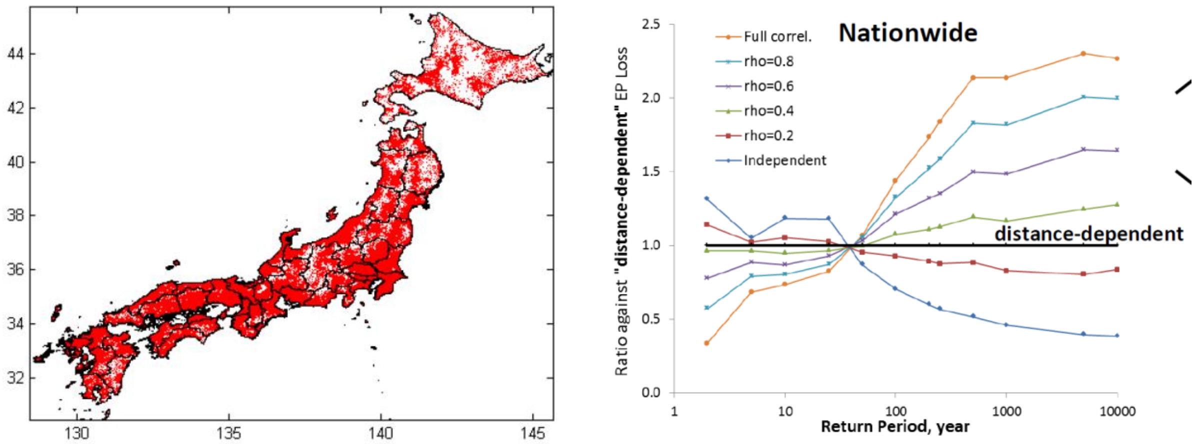


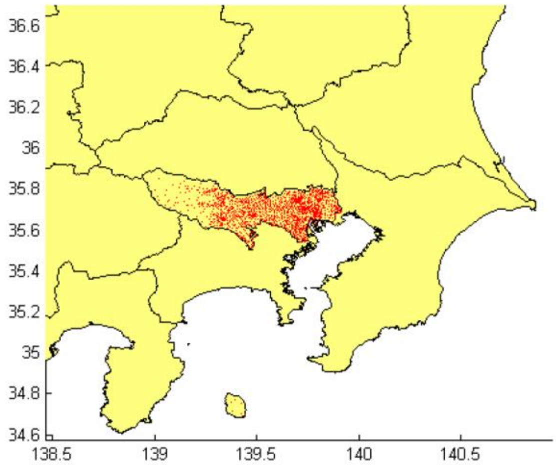

Fig. 7(a) Distribution of nationwide exposure (b) PML comparison ratio against distance dependent

As can be seen in Fig. 7, depending on the return periods, the difference can be very small at around 75-year return period, but a few times different at return periods above 500 years according to different assumption of constant correlation.



The second case is if we only look at a portfolio with exposure in Tokyo, the comparison is shown in Fig. 8.



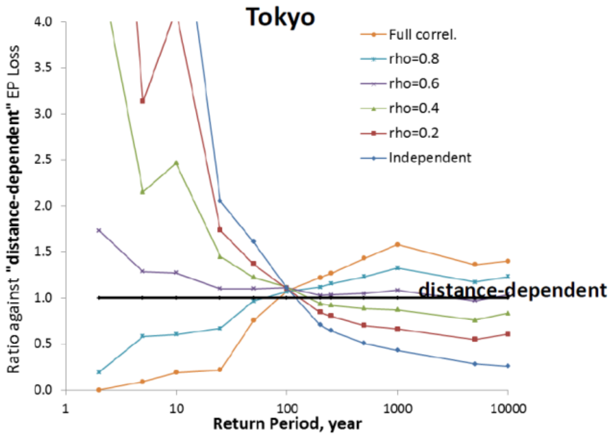




Fig. 8(a) Distribution of exposure in Tokyo (b) PML comparison ratio against distance dependent

As can be seen from Fig. 8, when only exposure in Tokyo is considered, much variation is seen for PMLs at smaller return periods when they are less than 30 years or so. Therefore, the effect of distance dependent correlation is much relied upon the distribution of exposure and its comparable area size against earthquake footprint size. A general trend is that when exposure distribution size is much bigger than typical earthquake footprint, bigger difference is seen at higher return periods, and when exposure area is comparable or smaller than a typical earthquake footprint, bigger variation is observed for PMLs at smaller return periods. The impact can be as big as a few times different to little or small difference at return periods closer to 100-year return periods.

## 5. Concluding Remarks

Distance dependent correlation is very important in studying the uncertainty and tail risk of earthquake loss estimation. Through directly sampling loss at location level based on the distance dependent attribute derived from actual earthquake loss information, this paper proposed an efficient algorithm using kriging technique for a large amount of locations for actual earthquake loss estimation. Simulation errors were used to validate the proposed approach and actual portfolios in Japan were used to study the impact of distance dependency. The results showed that distance dependency can have significant effect on PML estimation at various return periods for different portfolios with variable exposure area sizes. The constant dependency assumption can over- or under-estimate earthquake tail risk depending on the exposure distribution as well as the PML return periods.

## 6. Acknowledgements

The author would like to thank Dr. Kai Cui for his contribution of the theoretical framework of kriging and distance dependent sampling, and Dr. Sangjoon Park for his efforts in helping on the impact study for Japanese residential portfolios. Part of the work was completed when the author was working in Validus Research when both Dr. Kai and Dr. Joon were part of a research team, which developed an advanced platform for Japanese earthquake risk quantification and modelling. Part of the work has been funded by Henan University in Kaifeng, China and Institute of Engineering Mechanics, China Earthquake Administration in Harbin, China,



## 7. References

[1] Sokolov V, Wenzel F (2011): Influence of spatial correlation of strong ground motion on uncertainty in earthquake loss estimation. *Earthquake Engineering & Structural Dynamics*, **40**, 993-1009.

[2] Bommer JJ, Scherbaum F, Bungum H, Cotton F, Sabetta F, Abrahamson NA (2005): On the use of logic trees for ground motion-prediction equations in seismic-hazard analysis. *Bulletin of Seismological Society of America*, **95**,377–389.

[3] Page MT, Carlson JM (2006): Methodologies for earthquake hazard assessment: model uncertainty and the WGCEP-2002 forecast, *Bulletin of Seismological Society of America*, **96**,1624–1633.

[4] Li YH, Shi BP, Zhang J (2008): Copula joint function and its application in probability seismic hazard analysis, *Acta Seismologica Sinica*, **21**(3), 296–305.

[5] Strasser FO, Abrahamson NA, Bommer JJ (2009): Sigma: issues, insights, and challenges, *Seismological Research Letters*, **80**, 40–56.

[6] Shah CH, Dong WM, Stojanovski P, Chen A (2018): Evolution of seismic risk management for insurance over the past 30 years, *Earthquake Engineering and Engineering Vibration*, **17** (1), 11-18.

[7] Jayaram N, Baker JW (2009): Correlation model for spatially distributed ground-motion intensities, *Earthquake Engineering & Structural Dynamics*, **38**, 1687–1708.

[8] Gundogdu KS, Guney I (2007): Spatial analyses of groundwater levels using universal kriging, *Journal of Earth System Science*, **116**, 49–55.

[9] Liu RS, Wang ZF, Zhu M (2006): Study on financial loss and its adjustment in earthquake insurance, *Acta Seismologica Sinica*, **19**, 207–216.

[10] Wang ZF (2008): A preliminary report on the Great Wenchuan Earthquake, *Earthquake Engineering and Engineering Vibration*, **7** , 225-234.

