

CatMod: A computational platform for earthquake loss modeling of catastrophic events

Akkar Sinan – Turk Reinsurance Inc., İstanbul, Turkey, sakkar@turkre.com

Yazgan Ufuk – Earthquake Eng. Dpt., Istanbul Tech. University, Istanbul, Turkey, ufukyazgan@itu.edu.tr

Ülkü Onur – Turk Reinsurance Inc., İstanbul, Turkey, oulku@turkre.com

Abstract: We describe two alternative earthquake loss methods being implemented in Turk Reinsurance Inc.'s catastrophic event modeling platform (CatMod) for aggregating insured portfolio losses. The procedures use Convolution Theorem and Monte Carlo sampling to compute insured portfolio loss at target annual exceedance probabilities. We present simple cases studies to discuss the similarities and discrepancies in aggregate portfolio losses by these alternative procedures.

Keywords: Portfolio Loss for Insurance/Reinsurance, Convolution and Event-Based Approaches, Probabilistic Loss Analysis, Monte Carlo Sampling

1. Introduction

In re/insurance industry, decisions on risk transfer against catastrophic events rely on the computation of portfolio losses for target annual exceedance probabilities. Consequently, the modelers establish probability based platforms while computing aggregate losses of insured portfolios subjected to catastrophic events.

Probabilistic portfolio losses can be computed by utilizing either event-based (Bazzurro and Luco, 2007) or convolution (Wesson et al., 2009) approaches. The former uses Monte Carlo sampling and the aleatory variability and epistemic uncertainty in portfolio vulnerability modeling as well as the hazard model triggering the loss are determined by prior distributions. If the catastrophic event of interest is earthquakes, Monte Carlo simulation can (a) mimic the aleatory variability in ground-motion intensity metrics (GMIMs) triggering the portfolio losses, and (b) account for the epistemic uncertainty in source activity, maximum magnitude, seismic source geometry, etc. It can also consider the epistemic uncertainty in portfolio vulnerability such as the uncertainty in damage states, distribution of damage given a GMIM level, etc. The Monte Carlo sampling can easily integrate the spatial correlation in GMIMs, which is of importance when the assets in a given portfolio are closely spaced (within few hundred meters to few kilometers) provided that their geo-coordinates are precisely defined in the portfolio. If spatial correlation is disregarded in Monte Carlo simulations, the resulting motion intensities for all assets at a given site would be either fully correlated or partially correlated depending on the ground-motion sampling strategy implemented in the procedure. Full correlation would be achieved by simulating a single within-event variability component for all assets located at the same region. On the other hand, partial correlation would be obtained if a different within-event component is simulated for each asset within a given region.

Given an earthquake scenario having a magnitude $M = m$, the convolution approach uses the ground-motion probabilities conditioned on predefined fractions of between-event standard deviation, τ , (i.e., $\eta \cdot \tau$, η is standard normal random variable, $\eta \sim N(0,1)$). Consequently, the GMIMs at the sites of interest only differ by the random within-event standard deviation component $\varepsilon \cdot \phi$ ($\varepsilon \sim N(0,1)$) and they can be assumed to be independent. Under the conditionally independent ground motions, the losses to the assets in the insured portfolio can also be considered as independent. The conditionally independent probability densities

of individual losses are convolved through Fast Fourier Transformation (FFT) to obtain the aggregate loss distribution to the insured portfolio. As the within-event standard deviation component ($\varepsilon \cdot \phi$) in the ground-motion model is considered to be an independent random variable, the convolution approach would yield sum of uncorrelated losses in the insured portfolio for each event.

Event-based and convolution approaches have their advantages and disadvantages. The Monte Carlo simulation requires a large sampling space for an accurate portfolio loss distribution. This means powerful computational platforms and management of big data. Conversely, the convolution approach can be faster but requires a careful structuring of cumbersome numerical analyses. Notwithstanding, this approach has been criticized for disregarding the spatial correlation of GMIMs (i.e., independency of ground motions even at closely spaced sites leading to uncorrelated losses) increasing the possibility of biased portfolio losses. The last point may not be significant, if the insured portfolio at a specific region (such as a town), is defined at a single geo-coordinate.

This note is a follow up of Yazgan et al. (2022) and briefs the earthquake loss modeling methodologies in Turk Reinsurance Inc.'s catastrophic event modeling platform (CatMod). The paper starts by explaining the probabilistic convolution framework and its implementation to reinsurance financial structure for computing the aggregate losses at different annual exceedance probabilities. This discussion is followed by explaining the event-based approach that constitutes the alternative to the convolution method while assessing yearly-based reinsurance coverages for treaty and facultative portfolios. The paper concludes by presenting some case studies to show the differences between these two approaches in earthquake loss modeling.

2. Convolution approach for aggregate portfolio losses

The conditionally independent loss probability density function (pdf) of i^{th} asset, L_i , in an insured portfolio conditioned on a specific magnitude $M = m$, source-to-site distance $D = d$, and a fraction of between-event standard deviation $BE = \eta_j \cdot \tau$ ($\eta_j \sim N(0,1)$) is

$$f_{L_i|M,D,BE}(l_i | M = m, D = d, BE = \eta_j \cdot \tau) = \int_0^{\infty} f_{L_i|IM}(l_i | IM) \cdot f_{IM|M,D,BE}(im | M = m, D = d, BE = \eta_j \cdot \tau) di m. \quad (1)$$

$f_{L_i|IM}$ represents the pdf of asset's vulnerability model and $f_{IM|M,D,BE}$ is the ground-motion pdf conditioned on a specific earthquake scenario represented by a magnitude, distance and fraction of between-event standard deviation.

Given a set of $BE = \eta_j \cdot \tau$ covering the entire range of between-earthquake variability for the seismic source of concern and the range of magnitudes having a finite occurrence probability on the seismic source, the recursive use of Eqn. (1) together with Total Probability Theory would yield the loss pdf of the i^{th} asset (f_{L_i}) in the portfolio. Recall that derivation of f_{L_i} relies on conditionally independent ground motions and asset losses in the portfolio.

Eqn. (2) shows the implementation of convolution approach for computing the annual probabilities of aggregate portfolio loss (L_T) exceeding a set of loss levels (i.e., $P[L_T > l]$).

$$P[L_T > l] = \int_0^l f_{L_T}(l) dl \quad \text{where} \quad f_{L_T}(l) = (f_{L_1} * f_{L_2} * \dots * f_{L_{N_e}})(l) \quad (2)$$

where $f_{L_T}(l)$ and $f_{L_i}(l)$ are the pdfs of L_T and L_i , respectively. The operation designated using ‘*’ symbol represents convolution operation. Often closed-form solutions cannot be obtained for $f_{L_T}(l)$ and $f_{L_i}(l)$. Hence, L_T and L_i are typically modelled using discrete random variables. For this purpose, the entire range of considered losses (*i.e.*, $[0, l_{max}]$) are discretized into N intervals. The upper boundary of the j th interval is defined as follows:

$$l_j = (j - 1)\Delta l \quad \text{where} \quad \Delta l = l_{max}/N \quad (3)$$

Once the discretization scheme is defined, an efficient FFT procedure for evaluating the annual probability $P[L_T > l]$ as given in Eqn. (2) is presented in (Wesson et al., 2009). We note that similar to other numerical calculations involving nonlinear functions and convolutions, the discretization scheme has a strong impact on the numerical precision and the computational workload. Therefore, l_{max} and N should be selected with care to achieve reasonable accuracy with efficient use of processing power.

A recursive N and l_{max} strategy, and a sequential procedure for computing f_{L_T} is used in CatMod’s convolution approach in loss assessment. The f_{L_T} is evaluated in multiple convolution steps (Yazgan et al., 2022).

3. Implementation of reinsurance financial structure to convolution approach

3.1 Background

Given an insured asset, the re/insurance structure divides asset’s total value (hereafter sum insured, SI) into three components shown as a financial box structure on the left hand side of Figure 1. The ground up loss distribution (plotted to the left of the financial box) shows the loss probabilities given a specific percentage of SI value. Although this distribution is an important metric showing the likelihoods of loss with regards to SI value, its use is quite limited for re/insurance as it ignores the financial aspects of the agreement signed between the client and the re/insurer. The “limit” l_B designates the fraction of SI and it is determined from asset’s probable maximum loss (PML) value. PML is calculated for the catastrophic earthquake event that is likely to hit the asset. In essence, l_B shows the maximum responsibility level of the re/insurer for loss payment under the signed agreement. Deductible limit (l_A) is the loss limit below which only the client is responsible for loss compensation. Hence, ground up losses over l_B (over limit) and below l_A (below deductible) stay outside the responsibility of re/insurer. Consequently, the hatched areas in ground up loss pdf show the probabilities of ‘no payment’ by the re/insurer. The re/insurer, to decide on next year’s financial coverage and hence the risk transfer, needs a probabilistic model to assess the likelihoods of consolidating specific “gross” loss amounts (L^g) given the catastrophic event occurs.

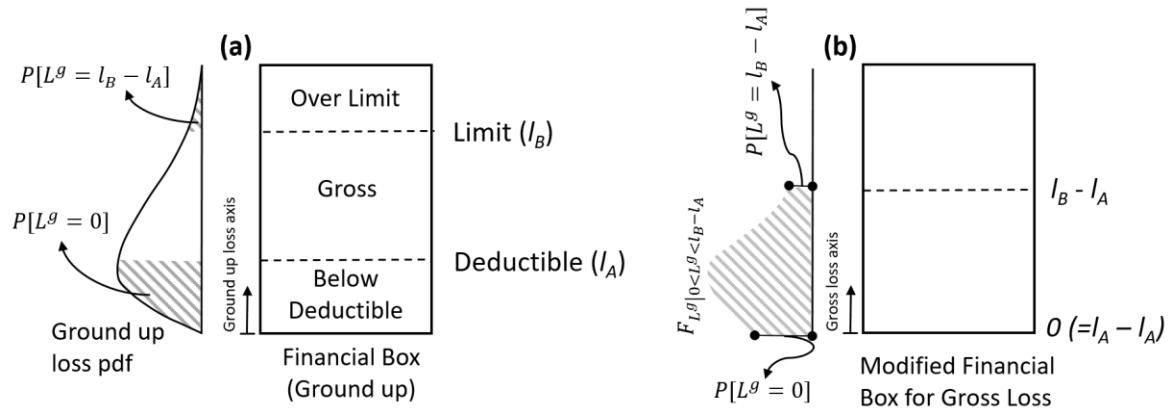


Fig. 1 – Financial re/insurance structure of an asset at risk under a catastrophic event. (a) Ground up loss distribution and financial reinsurance structure depicting “over limit”, “gross loss”, and “below deductible” components, (b) Modified financial box for probabilistic loss modeling of the asset that accounts for gross loss under the responsibility of re/insurer

To establish the probabilistic loss model, we modify the loss probability distribution for gross loss only. This is illustrated by the second financial box structure on the right hand side of Figure 1 (hereafter “gross loss financial box”). The pdf on the right hand side of gross loss financial box shows the conditional loss probabilities in which the re/insurer is not responsible for payments below and above l_A and l_B , respectively, while the loss under re/insurer’s responsibility is somewhere between these limits (i.e., $f_L^g(L^g \leq l | 0 < L^g < l_B - l_A)$). Note that the area under this conditional pdf should be 1 and the probabilities of $P[L^g = 0]$ and $P[L^g = l_B - l_A]$ are computed from the hatched areas under the ground up loss pdf.

The left and right panels of Figure 2 show the cumulative distribution functions (cdfs) of conditional ($F_{L^g|0 < L^g < l_B - l_A}$) and unconditional gross losses ($F_{L^g}(l)$), respectively. $F_{L^g|0 < L^g < l_B - l_A}$ is computed from F_{L^g} as given in Eqn. (4). The cdf on the right hand side of Figure 2 shows all the terms in Eqn. (4).

$$F_{L^g|0 < L^g < l_B - l_A} = \frac{P[L^g \leq l | 0 < L^g < l_B - l_A]}{P[0 < L^g < l_B - l_A]} = \frac{P(L^g \leq l) - P(L^g = 0)}{1 - P(L^g = 0) - P(L^g = l_B - l_A)}. \quad (4)$$

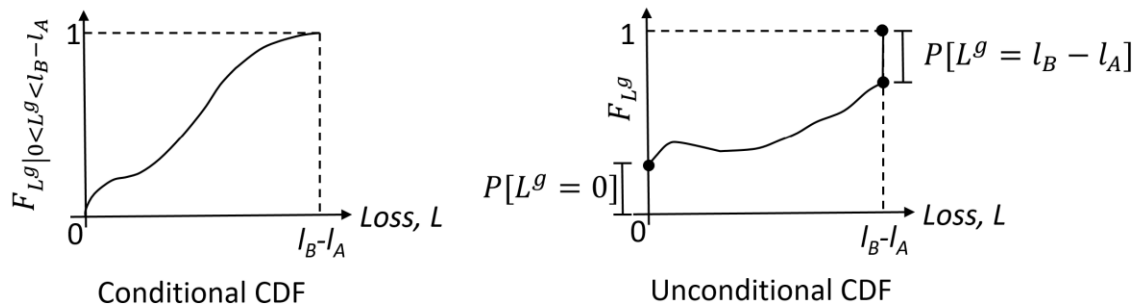


Fig. 2 – Conditional (left panel) and unconditional (right panel) cdfs of gross loss of an insured asset

3.2 Aggregation

Perhaps, the easiest way to explain the implementation of aggregation concept to our probabilistic loss model is to work on a case that aggregates two individual assets underwritten in two different agreements for Client-1 and Client-2. Once the probabilistic case is established for this case aggregating consequent risks is done recursively. Let the

gross losses from the first agreement is designated by L_1^g and let L_2^g designate the gross losses from second agreement. The aggregated loss (i.e., L_T) requires the consideration of multiple (joint) probabilities arising from the reinsurance structures of L_1^g and L_2^g . This issue can be addressed by using Total Probability Theorem that is presented in Figure 3.

$L_2^g \backslash L_1^g$	$L_1^g = 0$	$0 < L_1^g < l_{B_1} - l_{A_1}$	$L_1^g = l_{B_1} - l_{A_1}$
$L_2^g = 0$	I	II	III
$0 < L_2^g < l_{B_2} - l_{A_2}$	IV	V	VI
$L_2^g = l_{B_2} - l_{A_2}$	VII	VIII	IX

Fig. 3 – Conditional probabilities that the probabilistic loss model accounts for when aggregating the gross losses of two individual assets each with different deductible (l_A), limit (l_B) and SI values.

The notation in Figure 3 uses the indices “1” and “2” to designate the re/insurance agreements signed with Client-1 and Client-2, respectively. Each region (shown in Roman numerals) represents a conditional probability for L_T such that $L_T \leq l$ is conditioned on the joint probability describing the specific re/insurance agreement terms set with Client-1 and Client-2. For example, **Region I** represents the conditional probability $P[L_T \leq l | L_1^g = 0, L_2^g = 0]$ describing the annual non-exceedance probability of L_T for a threshold loss value l when the gross losses in both agreements are equal to zero (i.e., $L_1^g = 0$ and $L_2^g = 0$). (Such a condition occurs when the losses in the assets are less than their deductible limits, l_{A_1} and l_{A_2}). The sum of all nine conditional probabilities gives the annual non-exceedance probability of aggregate loss given a threshold loss level, l (i.e., $P[L_T \leq l]$). Eqn. (5) shows the entire set of conditional probabilities to compute $P[L_T \leq l]$.

$$P[L_T \leq l] = P[L_T \leq l | L_1^g = 0, L_2^g = 0] \cdot P[L_1^g = 0, L_2^g = 0] \quad (5.a)$$

$$+ P[L_T \leq l | 0 < L_1^g < l_{B_1} - l_{A_1}, L_2^g = 0] \cdot P[0 < L_1^g < l_{B_1} - l_{A_1}, L_2^g = 0] \quad (5.b)$$

$$+ P[L_T \leq l | L_1^g = l_{B_1} - l_{A_1}, L_2^g = 0] \cdot P[L_1^g = l_{B_1} - l_{A_1}, L_2^g = 0] \quad (5.c)$$

$$+ P[L_T \leq l | L_1^g = 0, 0 < L_2^g < l_{B_2} - l_{A_2}] \cdot P[L_1^g = 0, 0 < L_2^g < l_{B_2} - l_{A_2}] \quad (5.d)$$

$$+ P[L_T \leq l | 0 < L_1^g < l_{B_1} - l_{A_1}, 0 < L_2^g < l_{B_2} - l_{A_2}] \cdot P[0 < L_1^g < l_{B_1} - l_{A_1}, 0 < L_2^g < l_{B_2} - l_{A_2}] \quad (5.e)$$

$$+ P[L_T \leq l | 0 < L_1^g < l_{B_1} - l_{A_1}, 0 < L_2^g < l_{B_2} - l_{A_2}] \cdot P[0 < L_1^g < l_{B_1} - l_{A_1}, 0 < L_2^g < l_{B_2} - l_{A_2}] \quad (5.f)$$

$$+ P[L_T \leq l | L_1^g = 0, L_2^g = l_{B_2} - l_{A_2}] \cdot P[L_1^g = 0, L_2^g = l_{B_2} - l_{A_2}] \quad (5.g)$$

$$+ P[L_T \leq l | 0 < L_1^g < l_{B_1} - l_{A_1}, L_2^g = l_{B_2} - l_{A_2}] \cdot P[0 < L_1^g < l_{B_1} - l_{A_1}, L_2^g = l_{B_2} - l_{A_2}] \quad (5.h)$$

$$+ P[L_T \leq l | L_1^g = l_{B_1} - l_{A_1}, L_2^g = l_{B_2} - l_{A_2}] \cdot P[L_1^g = l_{B_1} - l_{A_1}, L_2^g = l_{B_2} - l_{A_2}] \quad (5.i)$$

The above probability expressions, except **Region V** can be readily computed from the conditional and unconditional L^g cdfs of each agreement. The calculation of conditional probability in Eqn. (5.e) requires additional explanation. As such, we present Figure 4 illustrating a basic representation of the conditional probability in Eqn. (5.e). Given a specific threshold loss, l , the shaded area represents all possible values of L_T 's satisfying the conditions $0 < L_1^g < l_{B_1} - l_{A_1}$ and $0 < L_2^g < l_{B_2} - l_{A_2}$. The computation of aggregate loss

distribution satisfying the given specific condition can be done through the convolution theorem (Section 2).

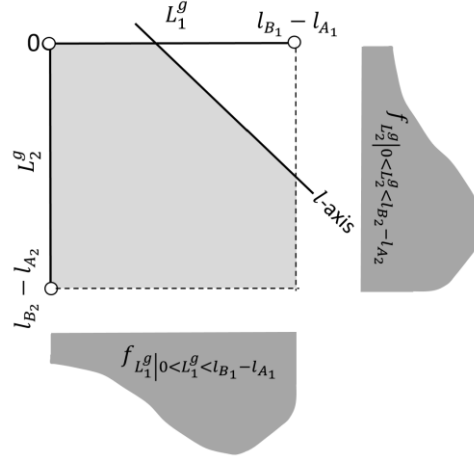


Fig. 4 – Illustration of conditional probability component in Eqn. (5.e). Probable gross losses from Client-1 and Client-2 risks contributing to L_T should fall on the l -axis

The calculation of Eqn. (5) would yield the aggregate loss cdf of two individual assets underwritten in two different agreements. The procedure is implemented sequentially by aggregating subsequent agreement over this cdf, and proceed until the entire insurance agreements in a given portfolio are fully considered. The final result would be the annual loss probability cdf of the entire portfolio and the complementary cdf would yield the annual exceedance probabilities of loss levels that are of interest to re/insurer.

4. Event-based Portfolio Loss Computation via Monte Carlo Simulation

Monte Carlo simulation technique can be used effectively in event-based portfolio loss computation while describing the earthquake hazard and resulting loss to the re/insurance portfolio composed of multiple risks (assets) at multiple sites. Given the seismic sources of interest, Monte Carlo simulation populates the scenario earthquakes by using the source specific magnitude-frequency distributions. The scenario events are distributed randomly along the seismic sources to compute the ground-motion distributions at the insured portfolio sites.

Given a seismic source, the ground-motion distributions at the portfolio sites are computed from Eqn. (6) for events sampled from the source-specific magnitude-frequency distribution.

$$\log(Y_{es}) = \log(\overline{Y_{es}}) + \tau \cdot \eta_e + \phi \cdot \epsilon_{es} \cdot L_{ss} \quad (6)$$

Y_{es} and $\overline{Y_{es}}$ are e by s matrices keeping the spatially correlated and median GMIM distributions, respectively at the portfolio sites. The indices s and e represent number of portfolio sites and number of sampled earthquake scenarios generated by Monte Carlo simulation, respectively. As already explained, τ and ϕ designate, respectively the between-event and within-event standard deviations of the GMIM used in defining the ground-motion distributions. η_e and ϵ_{es} are the randomly sampled between- and within-event residuals that follow normal standard distribution. L_{ss} is the s by s lower triangular matrix obtained from Cholesky decomposition such that $L_{ss} \cdot L_{ss}^T = C$ where C is the positive definite correlation matrix to consider the spatial correlation of ground motion between sites. The correlation matrix C is computed from a relevant spatial correlation model in the literature.

Aggregated loss distribution to insured portfolio (L_e) for the set of sampled events is computed from Eqn. (7). $V(Y_{es}; c_s)$ is the vulnerability model for the set of risks that belong to vulnerability classes c_s at sites s . iv_s is the vector of insured values and Y_{es} is the spatially correlated GMIMs triggering the portfolio loss at the portfolio sites.

$$L_e = V(Y_{es}; c_s) \cdot iv_s \quad (7)$$

Since Monte Carlo simulation technique allows the computation of loss for each individual risk in the portfolio, the re/insurance financial structure imposed by the re/insurance agreements are implemented directly to the ground-up and gross losses without going through the steps discussed Section 3.

5. Case Studies

We present simple cases to delineate the differences and similarities in convolution and event-based portfolio loss calculation methods. As stated previously in the paper, the convolution approach yields uncorrelated losses because the method imposes conditionally independent losses for each risk in the portfolio given a specific seismic event. The Monte Carlo sampling, on the other hand, can determine the loss between full and no correlation depending on the proximity of the sites where the risks are located. This point is illustrated in Figure 5 by displaying the within-event residual sampling and corresponding damage ratios of two similar risks located at two different site. The fully correlated (Figures 5.a) and uncorrelated (Figure 5.b) hazard assumption (represented by within-event residual distributions) lead to fully correlated (Figure 5.c) and uncorrelated (Figure 5.d) losses (represented by damage ratios) of these two assets. The presented marginal loss distributions would eventually affect the total (aggregated) loss of the two risks. Needless to say, the actual (realistic) portfolio losses should lie between the above two marginal ends that depends on the portfolio distribution over the region of interest.

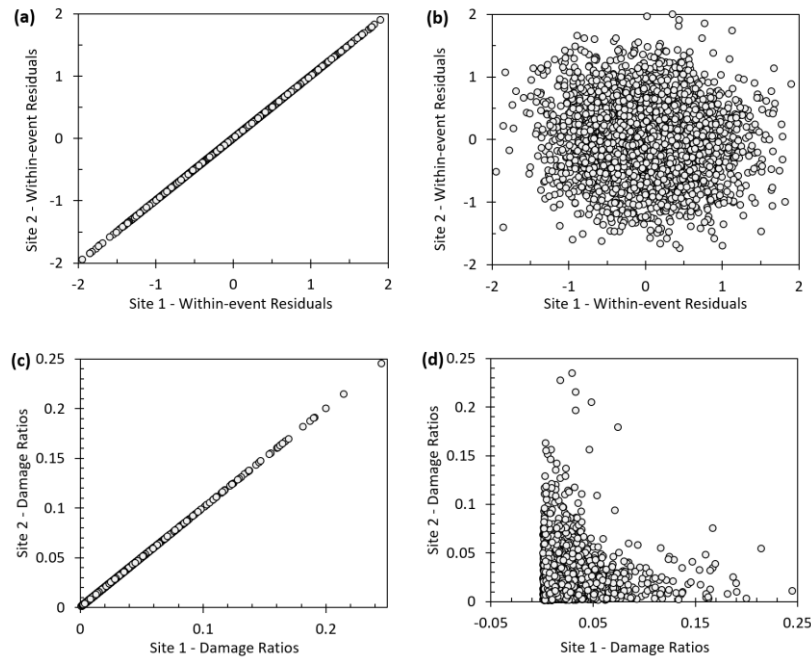


Fig. 5 – Fictitious cases showing (a) fully correlated and (b) uncorrelated residual sampling, and the corresponding (c) fully correlated and (d) uncorrelated loss (damage ratios) of two assets located at two different sites

The case study presented in Figures 6 and 7 discusses how far the correlated and uncorrelated aggregate loss of two risks differ depending on the change in the proximity of two sites. For simplicity, the two risks are assumed to have the same mean vulnerability model with different sum insured (SI) values. Figure 6 defines the fictitious earthquake scenario for this case study. A magnitude **M7.5** earthquake occurs on the fictitious fault and affects the two insured assets. The first asset is located at Site 1 and the second insured asset is assumed to move from Site 2 to Site 5. The aggregate loss of the two risks are computed using convolution and event-based Monte Carlo sampling approaches as the second risk is sequentially moved from Site 2 to Site 5. Note that Site 2 is in the close proximity of Site 1 (0.5 km of difference between Site 1 and Site 2) whereas Site 5 is the farthest point from Site 1 (10 km difference between Site 1 and Site 5). For event-based Monte Carlo sampling, the PGV (GMIM for vulnerability model), spatial correlation model by Esposito and Iervolino (2011) is used. The calculations are done by CatMod.

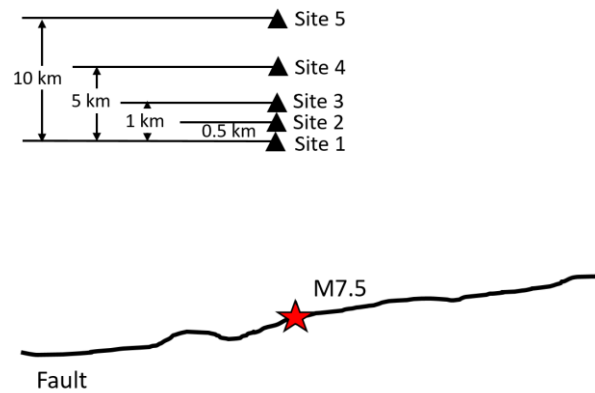


Fig. 6 – Fictitious earthquake scenario, location of the sites and their proximities with respect to Site 1

The left panel of each row in Figure 7 shows the correlated within-event residuals between Site 1 and the other site and the right panel illustrates the aggregated loss of the two risks computed by convolution and event-based Monte Carlo sampling approaches. Recall that the first risk is always located at Site 1 and the location of the second risk is moved each time from Site 2 to Site 5 to compute the aggregate loss. The first row in Figure 7 shows the case for Site 1 and Site 2 (Site 2 is 0.5 km away from Site 1), whereas the last row describes the case for Site 1 and site 5 (Site 5 is 10 km away from Site 1). The sampled within-event residuals in the first row display a very high spatial correlation as the slope of the linear line (dark red color) fitted on the sampled residuals is close to the black diagonal line representative of the full correlation trend. The aggregate loss exceedance probability curves by convolution (that is based on uncorrelated hazard and loss) and event-based Monte Carlo sampling procedures (that accounts for the correlation in hazard and loss) differ from each other for exceedance probabilities smaller than $2 \cdot 10^{-2}$. The spatially correlated hazard between the two sites decrease gradually and becomes almost negligible for Site 1 and Site 5 (i.e., when the distance between the two sites is 10 km) as the dark red trend line fitted on the sampled within-event residuals for this case is almost parallel to the horizontal axis. Consequently, the aggregate loss exceedance probability curves by convolution and MCS approaches follow each other very closely for exceedance probabilities down to $2 \cdot 10^{-4}$. Though not discussed, the observations about aggregate loss exceedance probability curves for Site 1-Site 5 case are also valid for Site 1-Site 4 case where the two sites are 5 km from each other.

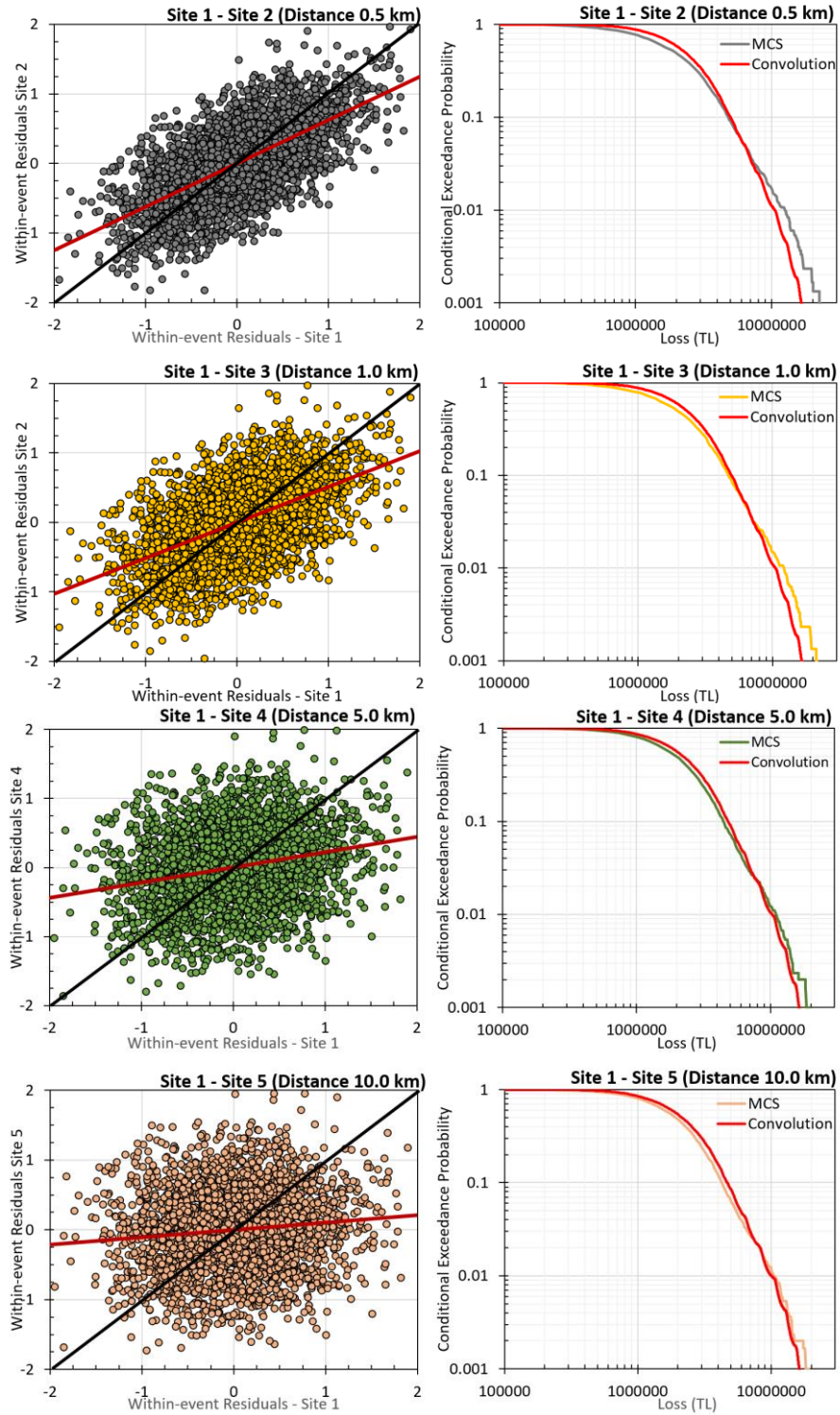


Fig. 7 – Comparisons of convolution and event-based aggregate losses for two insured assets located at two sites separated by 0.5 km (top row), 1 km (second row), 5 km (third row), and 10 km (last row).

6. Summary and Conclusions

Currently, Turk Reinsurance Inc.'s catastrophic event modelling platform for portfolio loss (CatMod) uses convolution and event-based (via Monte Carlo sampling) methods to compute yearly re/insurance portfolio coverages at the desired probability levels. These two

approaches have different limitations for their implementation in practice. The convolution approach is faster and does not require large numbers of sampling but assumes uncorrelated hazard and loss. One can surmount the latter shortcoming by event-based Monte Carlo sampling at the expense of computational burden. The simple case study presented in this paper indicates that the convolution approach can be used with confidence while aggregating the losses when the sites (so the insured assets) are in the proximity of 5+ km with each other.

References

- Esposito S., and Iervolino I. (2011). PGA and PGV Spatial Correlation Models Based on European Multievent Datasets, *Bull. Seism. Soc. Am.*, 101, 2532-2541.
- Bazzurro P., and Luco N. (2007). Effects of Different Sources of Uncertainty and Correlation on Earthquake-Generated Losses, *Australian Journal of Civil Engineering*, 4, 1-14.
- Wesson RL., Perkins DM., Luco N., and Karaca, E. (2009). Direct Calculation of the Probability Distribution for Earthquake Losses to a Portfolio, *Earthq Spectra*, 25, 687-706.
- Yazgan U., Akkar S., and Ulku O. (2022). Estimating Aggregate Earthquake Losses to Insurance Portfolios Using Convolution Approach, 12th US National Conference on Earthquake Engineering, Salt Lake City, Utah 27 June – 1 July, 2022.