



METHODS FOR EVALUATION AND TREATMENT OF EPISTEMIC UNCERTAINTY IN PORTFOLIO LOSSES DUE TO EARTHQUAKES

M. Markhvida⁽¹⁾, G. Cremen⁽²⁾, O. Grujic⁽³⁾, L. Ceferino⁽⁴⁾, J. Baker⁽⁵⁾

⁽¹⁾ PhD, Stanford University, markhvid@stanford.edu

⁽²⁾ Research Fellow, University College London, g.cremen@ucl.ac.uk

⁽³⁾ PhD, Stanford University, ogvg@alumni.stanford.edu

⁽⁴⁾ Postdoctoral Fellow, Princeton University, ceferino@princeton.edu

⁽⁵⁾ Professor, Stanford University, bakerjw@stanford.edu

Abstract

Assessment of seismic losses in a portfolio of buildings can be a challenging task, since there can be large epistemic uncertainties associated with the different steps of the probabilistic seismic risk analysis: hazard estimation, exposure modeling, fragility functions, and damage-to-loss estimation. Refining models and gathering more data to reduce the epistemic uncertainties can require substantial time investment and incur significant costs; therefore, to make this process more efficient, variables that drive the epistemic uncertainty must be identified. This paper explores the use of two sensitivity analysis methods to evaluate the effect of uncertain variables on the epistemic uncertainty of portfolio losses from earthquakes: 1) a well established variance-based sensitivity analysis technique and 2) a novel method that leverages regression tree ensemble methods with functional outputs. The two methods are examined using a fictional portfolio of 20 buildings in the San Francisco Bay Area. The results from the methods are compared, and advantages and disadvantages of the regression tree ensemble method are highlighted. Also discussed are recommendations for treatment of uncertain input variables based on insights about epistemic uncertainty in the losses.

Keywords: portfolio losses; epistemic uncertainty; sensitivity analysis; seismic risk.



1. Introduction

Probabilistic seismic risk analysis of building portfolios can be challenging, since there are often large epistemic uncertainties in each of the steps involved, i.e. in hazard estimation (e.g. [1]), exposure modeling (e.g. [2]), fragility functions (e.g. [3]) and damage-to-loss estimation [4]. Gathering more data to reduce the epistemic uncertainties can require substantial time investment and incur significant costs [5], so it is beneficial to make the process more efficient by identifying the variables that drive the epistemic uncertainty in the losses.

This paper explores the use of two sensitivity analysis methods to quantify the effect of uncertain variables on the epistemic uncertainty of portfolio losses due to earthquakes: (1) a well established variance-based sensitivity analysis technique and (2) a novel method that leverages regression tree ensemble methods with functional outputs.

Variance-based sensitivity analysis measures sensitivity in terms of the effect of a model input on the variance of a model output. The use of variance-based methods in sensitivity analysis has developed from the work of Sobol [6], Homma and Seltelli [7], and Jansen et al. [8] in particular. The technique has been applied in seismic risk analyses of individual buildings [9] as well as in a wide variety of other fields, including biological modeling [10].

Regression trees are a non-linear regression method that was introduced in the 1980's [11] and further popularized with the introduction of ensemble methods such as bagging, random forest, and boosting [12], [13]. Benefits of regression trees include the ability to handle missing data, and it is a powerful technique to assess variable importance since it is inherently used for variable selection. Tree ensemble methods have been widely used for variable selection and importance studies in bioinformatics [14] and more recently in flood risk modeling [15]. In this paper, we use an extension of tree ensemble methods, regression trees with functional output [16], to analyze variable importance in epistemic uncertainty of building portfolio losses.

The two methods are used to examine the sensitivity of average annual loss (AAL) to eight input variables with epistemic uncertainty, in a fictional San Francisco Bay Area portfolio of 20 buildings. The results from the two methods are compared, and advantages and disadvantages of the regression tree ensemble method are highlighted. Finally, based on the results, modeling recommendations related to reduction of epistemic uncertainty in the losses are provided.

2. Sensitivity methods

2.1 Variance-based sensitivity analysis

Given a model of the form $Y = g(\mathbf{X})$, variance-based sensitivity analysis measures the sensitivity of the output Y to the inputs \mathbf{X} in terms of a reduction in the variance of Y . Let $V_{\mathbf{X}}[Y]$ denote the variance of Y across the whole input space. $V_{\mathbf{X}}[Y]$ can be decomposed as follows:

$$V_{\mathbf{X}}[Y] = \sum_{i=1}^p V_i + \sum_{1 \leq i < j \leq p} V_{ij} + \dots + V_{1\dots p} \quad (1)$$

where p is the number of input variables. V_i measures the main effect of the input X_i on Y , and is defined as [17]:

$$V_i = V_{X_i}[E_{\mathbf{X}_{\sim i}}(Y|X_i)] \quad (2)$$

where $V[\cdot]$ denotes variance, $E(\cdot)$ denotes expectation, and $\mathbf{X}_{\sim i}$ includes all inputs but X_i . It is the expected reduction in variance that would be obtained if X_i could be fixed. The inner expectation operator takes the mean of Y over all possible values of $\mathbf{X}_{\sim i}$ for a fixed value of X_i . The outer variance is then taken over all possible values of X_i . The sensitivity measure associated with V_i is the first order sensitivity coefficient, defined as:



$$S_i = \frac{V_{X_i}[E_{\mathbf{X}_{-i}}(Y|X_i)]}{V_{\mathbf{X}}[Y]} \quad (3)$$

We use this sensitivity coefficient as the metric of sensitivity in this study.

S_i is estimated as follows. We generate N_s Monte Carlo samples of each of the inputs, according to the corresponding probability distributions. The Monte Carlo samples are used to construct matrix \mathbf{A} , termed the 'sampling matrix':

$$\mathbf{A} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_p^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_p^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N_s)} & x_2^{(N_s)} & \dots & x_p^{(N_s)} \end{bmatrix} \quad (4)$$

We generate a further N_s Monte Carlo samples of each of the input variables, independent of matrix \mathbf{A} . These Monte Carlo samples are used to construct matrix \mathbf{B} , termed the 're-sampling matrix':

$$\mathbf{B} = \begin{bmatrix} x_1^{(N_s+1)} & x_2^{(N_s+1)} & \dots & x_p^{(N_s+1)} \\ x_1^{(N_s+2)} & x_2^{(N_s+2)} & \dots & x_p^{(N_s+2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(2 \times N_s)} & x_2^{(2 \times N_s)} & \dots & x_p^{(2 \times N_s)} \end{bmatrix} \quad (5)$$

We construct a third sampling matrix \mathbf{C}_i , which is matrix, \mathbf{B} with the i th column substituted for the i th column of matrix \mathbf{A} .

The output variables obtained using each input matrix (Y_A, Y_B, Y_{C_i}) are then used to estimate S_i as [18]:

$$S_i = \frac{\left(\frac{1}{N}\right) \left(\sum_{j=1}^N Y_A^{(j)} Y_{C_i}^{(j)}\right) - \hat{f}_0^2}{\left(\frac{1}{N}\right) \sum_{j=1}^N (Y_A^{(j)})^2 - \hat{f}_0^2} \quad (6)$$

where $\hat{f}_0^2 = \frac{1}{N} \sum_{j=1}^N Y_A^{(j)} Y_B^{(j)}$ from [19]. These estimates can be derived from the well-known identity $V[Z] = E(Z^2) - E^2(Z)$.

Note that when inputs are correlated, they are grouped together as a multidimensional variable \mathbf{X}_r [20]. The first order sensitivity coefficient becomes:

$$S_r = \frac{V_{\mathbf{X}_r}[E_{\mathbf{X}_{-r}}(Y|\mathbf{X}_r)]}{V_{\mathbf{X}}[Y]} \quad (7)$$

The corresponding \mathbf{C}_r matrix is obtained by substituting the columns of the relevant correlated inputs in matrix \mathbf{A} for those of matrix \mathbf{B} .

2.2 Regression tree ensemble methods

The tree-based methodology used to conduct the sensitivity (variable importance) analysis is adapted from [16], where both scalar and vector outputs can be considered. First, a training sample, $\mathcal{T} = \{\mathbf{x}^{(i)}, y^{(i)}\}_1^{N_s}$, is considered, where $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_p^{(i)})$ is a vector of input variables in R^p space, and $y^{(i)}$ is a scalar output. Regression tree methods are greedy algorithms that recursively partition the R^p input space into M disjoint sub-regions (R_m), by splitting the space on one input variable at a time. A split s into sub-regions R_{ml} and R_{mr} is made along some input, p , such that it achieves the highest split quality, $Q_{s,p}^m$:



$$Q_{s,p}^m = G(R_m) - [G(R_{ml}) + G(R_{mr})] \quad (8)$$

where $G(R_m)$ is the cost function. In the case of scalar outputs, y , the sum of squared errors cost function, $G_{sse}(R_m) = \sum_{y^{(i)} \in R_m} (y^{(i)} - \mu_m)^2$, can be used, where the true output is approximated by a local mean, μ_m . In the case of functional outputs, $y(t)$, a commonly used cost function shown in Eq. (9) can be used, where $\mu_m(t)$ is the mean function of sub-region m .

$$G(R_m) = \sum_{y^{(i)}(t) \in R_m} \int_T (y^{(i)}(t) - \mu_m(t))^2 dt \quad (9)$$

Building just one regression tree can lead to over-fitting, therefore, tree ensemble methods such as bagging and random forest are often used. In this case study we use bagging to grow multiple trees, where each tree is built using data sampled with replacement from the original data. The predicted output is taken as the mean of all of the outputs produced by the multiple trees.

In order to determine variable importance, a variable importance index, S_p , can be calculated for each of the variables, p , in accordance with Eq. (10) where N_m is the number of training samples in a region m .

$$S_p = \frac{1}{M} \sum_{m=1}^M \frac{1}{N_m} \max\{Q_{s,p}^m, \forall s\} \quad (10)$$

When using ensemble methods with multiple trees, such as bagging or random forest, the importance index is taken as the average S_p across all trees.

3. Epistemic uncertainty in portfolio losses

3.1 Case study description

The fictional portfolio used in this study consists of 20 buildings in the San Francisco-Bay Area (Fig. 1). The properties of the buildings are obtained using the HAZUS methodology [21] and vary with different simulations, as explained in Section 3.2.2.

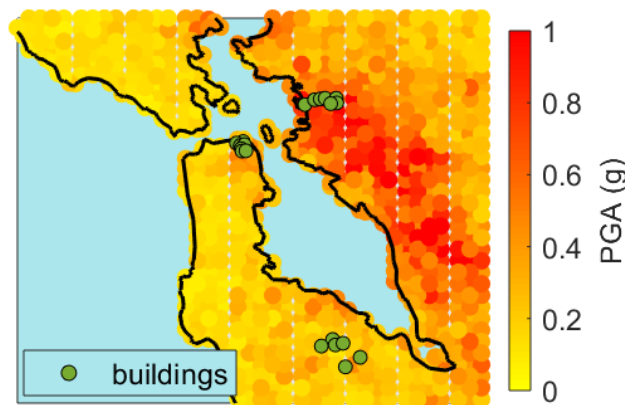


Fig. 1 – The fictional San Francisco Bay Area portfolio of 20 buildings examined in this study. Also included is a sample of peak ground accelerations (PGA's) for one of the earthquake rupture scenarios

The epistemic uncertainty is captured by randomly drawing 1000 samples of each of the uncertain variables (i.e., $N_s = 1000$ in Eqs. (4) and (5); variable distributions are described in Section 3.2). For each of the 1000 samples, a probabilistic risk analysis for the portfolio is performed using the UCERF2 Earthquake Rupture



Forecast scenarios [22], where for each of the scenarios an additional 300 correlated ground motion fields and damages are simulated to take into account aleatory uncertainty.

3.2 Sources of epistemic uncertainty

We examine epistemic uncertainties in eight variables associated with each step in the probabilistic seismic risk analysis of a portfolio of buildings: hazard estimation, exposure modeling, fragility functions, and damage-to-loss estimation.

3.2.1 Hazard estimation

Ground motion prediction equation (GMPE): GMPE's are used to predict the ground shaking for a particular earthquake scenario and site. The USGS hazard maps [23] use a logic tree to capture epistemic uncertainty in various GMPE's. In this analysis we examine four of the five GMPE's used in these maps – ASK, BSSA, CB, CY [24] – by sampling one model for each of the 1000 samples using equally weighted categorical distribution. The fifth model (IM) was excluded from the analysis since the between-event standard deviation values for this model are not provided by the UCERF2 Earthquake Rupture Forecast tool.

Median ground motion (GM): It is necessary to account for epistemic uncertainty in the median ground motion intensity obtained from GMPE's, due to within-model uncertainties as well as model-to-model differences. We quantify this epistemic uncertainty with the model proposed in [25], which uses a three-point discrete approximation to a normal distribution to represent epistemic uncertainty in the median prediction of a given GMPE. This approach involves development of three separate models for each GMPE. One of the models is equal to the original GMPE median and has weight 0.63. The other two models are equal to the median $\pm 1.645\sigma_{\mu \ln(gm)}$ and have weight 0.185. Note that for strike-slip faults:

$$\sigma_{\mu \ln(gm)} = \begin{cases} 0.083, & M < 7 \\ 0.056 \times (M - 7) + 0.083, & M \geq 7 \end{cases} \quad (11)$$

where M is earthquake magnitude.

Spatial correlation: Spatial correlation accounts for the joint occurrence of ground motion intensities at different sites within the portfolio, during a given earthquake. We use a spatial correlation model with the same functional form as that of [26]: $\rho(h) = \exp\left(\frac{-3 \times h}{b}\right)$, where $\rho(h)$ is the correlation between normalized intra-event ground motion residuals located h km apart, and b is the range of the correlation distance for the residuals. Epistemic uncertainty in the model is introduced via the b variable, as this depends on local-site conditions that are often unknown. b is assumed to vary according to the following lognormal distribution:

$$\ln(b) \sim \mathcal{N}(\ln(20\text{km}), 0.5) \quad (12)$$

3.2.2 Exposure modeling

To capture epistemic uncertainty due to a lack of knowledge about building characteristics (i.e., *design level*, *structural type*, and *occupancy type*) in the portfolio, we assume that the proportion of buildings in the portfolio associated with each of the building characteristics is unknown. The proportions for a given building characteristic (variable) are then drawn from a Dirichlet distribution, where the sum of the samples equals one. The distribution is parameterized by a vector α as follows:

$$[P_1, \dots, P_i, \dots, P_N] \sim \text{Dir}(\alpha = [\mathbf{1}, \dots, \mathbf{1}, \dots, \mathbf{1}]) \quad (13)$$

where P_i is the proportion of buildings associated with the i th possible value of the variable, and N is the number of possible values. For variance-based sensitivity analysis, the proportions can be viewed as random correlated sub-variables, so we use the coefficient defined in Eq. (7) to estimate the sensitivity for each variable, where \mathbf{X}_r is the matrix of all associated P_i values. In order to use vector $[P_1, \dots, P_N]$ as an input into the regression tree ensemble method, the samples are ordered using hierarchical clustering and the Bar-Joseph leaf reordering algorithm (see [16]).



Design level: This variable indicates the level of seismic design in a building, which affects the vulnerability of both structural and non-structural components. We investigate all HAZUS design levels in this study: (1) Pre-Code, (2) Low-Code, (3) Moderate-Code, and (4) High-Code. The proportions of each design level are obtained using Eq. (12) for $N = 4$.

Structural type: This variable indicates the type of lateral system in a building, which affects the vulnerability of structural components. We consider the following HAZUS structural types: (1) W2 - Wood, commercial and industrial (>5000 square feet), (2) S1L - Steel moment frame, (3) S2L - Steel braced frame, (4) S4L - Steel frame with cast-in-place concrete shear walls, (5) C1L - concrete moment frame, and (6) C2L - concrete shear walls. The proportions of each structural type are found using Eq. (12) for $N = 6$.

Occupancy type: This variable indicates how the building is used, which affects the losses that occur for a given level of damage. We examine the following HAZUS occupancy types: (1) RES3 - Multi-family dwelling, (2) RES4 - Temporary Lodging, (3) COM1 - Retail Trade, (4) COM4 - Professional/Technical/Business Services, (5) COM6 - Hospital, and (6) EDU1 - Schools/Libraries. The proportions of each structural type are found using Eq. (12) for $N = 6$.

3.2.3 Fragility functions

Building level fragility functions are used to represent the vulnerability of buildings in a portfolio, and are a function of both the design level and the structural type. We assume that epistemic uncertainty in the fragility functions is associated with the type of analytical model used to derive the functions and we quantify this uncertainty using the double-lognormal model, in which the median (θ) of a given function is considered a random variable, i.e.,

$$\ln(\theta) \sim \mathcal{N}(\ln(y_\theta), \beta_E) \quad (14)$$

where y_θ is the median of the HAZUS equivalent peak ground acceleration (PGA) fragility function for the specific design level, structural type, and damage state (as reported in Section 5.4.4 of [21]), and β_E is the epistemic uncertainty. The final fragility function is expressed as:

$$P(DS \geq ds_k | PGA = pga_i) = \Phi\left(\frac{\ln\left(\frac{pga_i}{\theta}\right)}{\beta_a}\right) \quad (15)$$

where $P(DS \geq ds_k | PGA = pga_i)$ is the probability of being in or exceeding damage state ds_k when $PGA = pga_i$, $\Phi(\cdot)$ is the standard normal cumulative distribution function, and β_A is the aleatoric uncertainty. It is assumed that $\beta_E = 0.1$, and $\beta_T = \sqrt{\beta_E^2 + \beta_A^2}$ is the log standard deviation of the HAZUS equivalent PGA fragility function.

We assume that there is perfect correlation among all fragility functions in the analysis (i.e., they were all derived using the same model), such that the same standard score is used to calculate each θ value in a given simulation. We analyze fragility function sensitivity specifically in terms of these scores.

3.2.4 Damage-to-loss modeling

Damage-to-loss models translate the damage predicted by the fragility functions to repair cost ratios, and are a function of the occupancy type. We quantify epistemic uncertainty in a given HAZUS mean repair cost ratio (μ_{RCR}) according to the following normal distribution:

$$RCR \sim \mathcal{N}(\mu_{RCR}, CV \times \mu_{RCR}) \quad (16)$$

where RCR is the simulated mean repair cost ratio and CV is the coefficient of variation for the given damage state, as reported in Table VI of [4]. Note that $RCR < 0$ is set to 0 and $RCR > 1$ is set to 1 (though such values are rare).



We assume that there is perfect correlation in all mean repair cost ratios that occur for a given occupancy in a given simulation, such that the same standard score is used to calculate each value. The sensitivity analyses for damage-to-loss focus on these scores specifically.

4. Results

Results of both sensitivity analysis methods for AAL are very similar (Fig. 2). The ranking of the five most important variables are identical for both methods (i.e., design level, fragility functions, median ground motion, ground motion prediction equations, and structural type, in order of importance) and the magnitude of the sensitivity parameters are also similar. While the rankings of the three other variables (i.e., repair cost ratio, occupancy type, and correlation distance) slightly differ, these variables can be considered negligible since their sensitivity coefficient in the variance-based analysis is ~ 0 , and the sensitivity index in the regression tree ensemble method is comparable to that of an independent random variable sampled outside of the analysis.

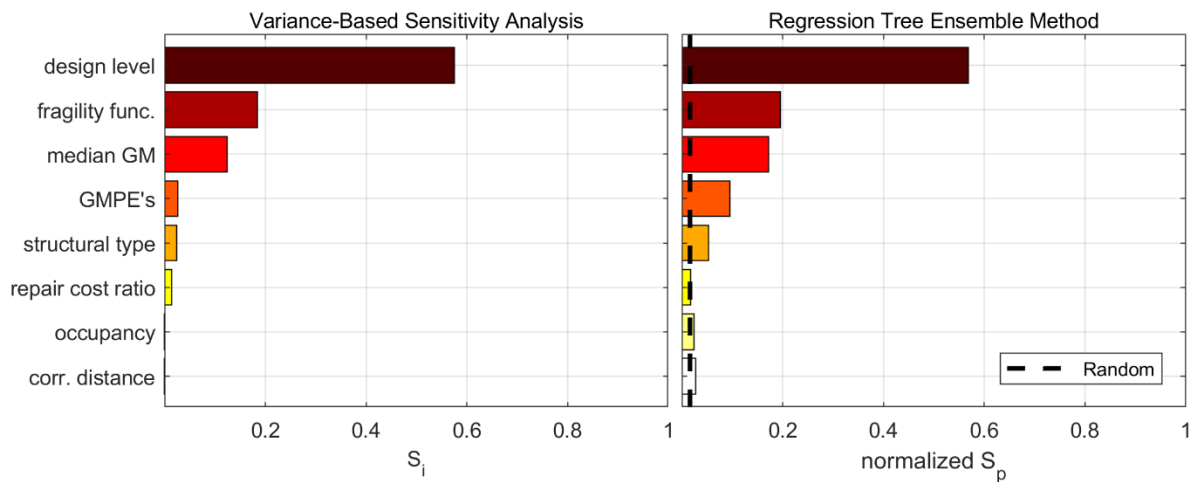


Fig. 2 – AAL sensitivity results for both analysis methods. In the case of regression tree ensemble, the sensitivity indices are normalized by the sensitivity index of using AAL as an input, i.e. the perfect predictor.

There are two main advantages to using regression tree ensemble over variance-based sensitivity analysis: computational efficiency and ability to analyze functional outputs. While both methods give a similar ranking, regression tree ensemble is $2+n$ times more efficient, where n is the number of input variables. In our study, variance-based sensitivity analysis took 10 times longer (23 hours) than the regression tree ensemble method (2.3 hours).

Since the regression tree ensemble method can handle functional outputs, it can also be used to analyze the sensitivity of the portfolio's annual loss exceedance probability (EP) to the input variables. In this case study, the ranking and magnitude of variable importance for loss EP curves is the same as that for AAL. Variance-based sensitivity analysis cannot be used in this context, since it is designed for univariate outputs.

One disadvantage of regression trees is that they become less reliable when using multi-modal functions or input vectors with high dimensionality, since optimal ordering of these vectors becomes less effective.

4.1 Recommendations on the treatment of uncertain variables

Based on the results of the sensitivity analyses, we provide two recommendations for treating epistemic uncertainty in portfolio losses. First, the uncertainty on the variables with the largest sensitivity should be reduced either through refining the model or obtaining more data. In our case study, as a first step we would acquire more data on the distribution of design levels within the portfolio. If we assume that the distribution



of design levels (DL) is known, we can see in Fig. 3 that the epistemic uncertainty in both AAL and loss EP is reduced.

The second recommendation is to ‘prune’ variables, or assume values for the unimportant variables (e.g. mean value). In the case of regression tree ensemble, a variable can be considered unimportant if its sensitivity index is comparable to the index of a randomly drawn input. For variance-based sensitivity analysis, a variable can be considered unimportant if its first-order sensitivity coefficient is close to zero. In our case study, the unimportant variables are the uncertainty on the repair cost ratio, occupancy type, and correlation distance. The effects of the two recommendations on AAL and loss EP uncertainty are shown in Fig. 3, where it can be seen that obtaining more data for important variables (i.e. design level) helps reduce the epistemic uncertainty, while assuming any value for the unimportant variables (i.e. repair cost ratio, correlation distance and occupancy type) preserves the distribution.

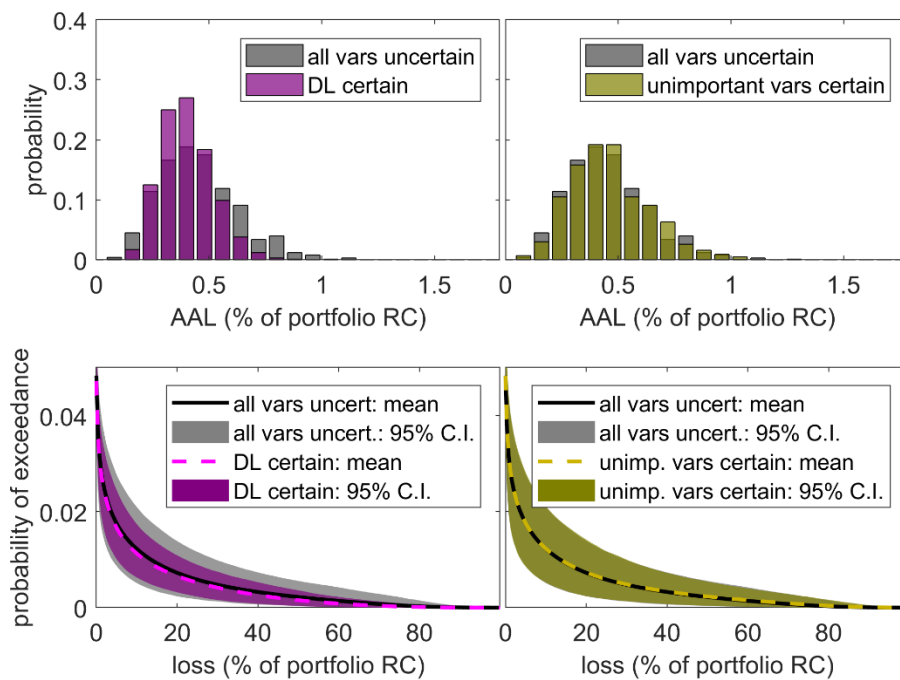


Fig. 3 – The effect of reducing epistemic uncertainty of important variables (design level, DL) and unimportant variables (repair cost ratio, correlation distance, and occupancy type) on the distribution of AAL (top sub-figures) and loss EP curves (bottom sub-figures). Portfolio losses are expressed as a percentage of the total replacement cost (RC). Changes in the mean and the 95% confidence interval of EP curve are also shown.

5. Conclusions

This paper explored the use of two sensitivity analysis methods – 1) a well established variance-based sensitivity analysis method and 2) a novel regression tree ensemble method with functional outputs -- to quantify the effect of eight uncertain variables on the epistemic uncertainty of average annual losses of a fictional portfolio of 20 buildings in the San Francisco Bay Area. We found that the two methods produce very similar results in terms of ranking and magnitude of sensitivities; the five most important variables according to both methods are building design level, fragility functions, median ground motion, ground motion prediction equations, and building's structural type (in order of importance).

The regression tree ensemble method was found to be significantly more efficient than variance-based sensitivity analysis. Unlike variance-based sensitivity analysis, the regression tree ensemble method can also



handle functional outputs, such as a portfolio's annual loss exceedance probability curve. However, it should be noted that regression trees become less reliable when using inputs that are multi-modal functions or vectors with high dimensionality.

To deal with epistemic uncertainty, we recommend reducing uncertainty in variables with the largest sensitivities (such as building design level in this study). We also showed that assuming values (e.g. the mean value) for unimportant variables does not change the epistemic uncertainty in the output of interest.

6. Acknowledgements

We would like to thank Prof. Jef Caers for his insight during this project.

7. References

- [1] Michel G (2017): *Risk modeling for hazards and disasters*. Elsevier.
- [2] Bommer JJ, Spence R, Erdik M, Tabuchi S, Aydinoglu N, Booth E, Del Re D, Peterken O (2002): Development of an earthquake loss model for Turkish catastrophe insurance. *Journal of Seismology*, **6**(3), 431–446.
- [3] Silva V, Crowley H, Varum H, Pinho R, Sousa R (2014): Evaluation of analytical methodologies used to derive vulnerability functions. *Earthquake Engineering & Structural Dynamics* **43**(2), 181-204.
- [4] Martins L, Silva V, Marques M, Crowley H, and Delgado R (2016): Development and assessment of damage-to-loss models for moment-frame reinforced concrete buildings. *Earthquake Engineering & Structural Dynamics* **45** (5), 797-817.
- [5] Crowley H, Bommer JJ, Pinho R, Bird J (2005): The impact of epistemic uncertainty on an earthquake loss model. *Earthquake Engineering & Structural Dynamics*, **34** (14), 1653–1685.
- [6] Sobol IM (1993): Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, **1** (4), 407–414.
- [7] Homma H, Saltelli A (1996): Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, **52** (1), 1–17.
- [8] Jansen MJW, Rossing WAH, Daamen R (1994): Monte Carlo estimation of uncertainty contributions from several independent multivariate sources. *Predictability and Nonlinear Modelling in Natural Sciences and Economics*, 334–343.
- [9] Cremen G and Baker JW (2020): Variance-based sensitivity analysis and uncertainty quantification for FEMA P-58 consequence predictions. *Earthquake Engineering & Structural Dynamics*, (in review).
- [10] Kamp F, Brüningk S, Cabal G, Mairani A, Parodi K, Wilkens JJ (2014): Variance-based sensitivity analysis of biological uncertainties in carbon ion therapy. *Physica Medica*, **30**(5), 583–587.
- [11] Breiman L, Friedman J, Olshen R, Stone C (1984): *Classification and regression trees*. Wadsworth International Group.
- [12] Breiman L (2001): Random forests. *Machine Learning*, **45** (1), 5–32.
- [13] Freund Y (1995): Boosting a weak learning algorithm by majority. *Information and Computation*, **121**, (2), 256–285.
- [14] Díaz-Uriarte R De Andres SA (2006): Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, **7** (1), 3.
- [15] Wang Z, Lai C, Chen X, Yang B, Zhao S, Bai X (2015): Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, **527**, 1130–1141.
- [16] Grujic O (2017): Subsurface modeling with functional data. *PhD thesis*, Stanford University, USA.
- [17] Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S (2010): Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, **181** (2), 259–270.



- [18] Saltelli A (2002): Making best use of model evaluations to compute sensitivity indices,” *Computer Physics Communications*, **145** (2), 280–297.
- [19] Yun W, Lu Z, Zhang K, Jiang X (2017): An efficient sampling method for variance-based sensitivity analysis., *Structural Safety*, **65**, 74–83.
- [20] Jacques J, Lavergne C, Devictor N (2006): Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering & System. Safety*, **91**, (10–11), 1126–1134.
- [21] FEMA (2015): *HAZUS MH-2.1 Earthquake Model Technical Manual*. Federal Emergency Management Agency Washington, DC.
- [22] Field EH, Jordan TH, Cornell CA (2003): OpenSHA: A developing community-modeling environment for seismic hazard analysis. *Seismological Research Letters*, **74** (4), 406–419.
- [23] Petersen MD, Moschetti MP, Powers PM, Mueller CS, Haller KM, Frankel AD, Zeng Y, Rezaeian S, Harmsen SC, Boyd OS, Field EH, Chen R, Rukstales KS, Luco N, Wheeler RL, Olsen AH (2014): Documentation for the 2014 update of the United States national seismic hazard maps. *Report No. 2014-1091*, United States Geological Survey, USA
- [24] Gregor N, Abrahamson NA, Atkinson GM, Boore DM, Bozorgnia Y, Campbell KW, Chiou BSJ, Idriss IM, Kamai R, Seyhan E, Silva W (2014): Comparison of NGA-West2 GMPEs. *Earthquake Spectra*, **30** (3), 1179–1197.
- [25] Al Atik L, Youngs RR (2014): Epistemic uncertainty for NGA-West2 models. *Earthquake Spectra*, **30** (3), 1301-1318.
- [26] Jayaram N, Baker JW (2009): Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering & Structural Dynamics*, **38** (15), 1687-1708.