Classification algorithms for collapse prediction of tall buildings and regional risk estimation utilizing SCEC CyberShake simulations

Nenad Bijelić
Postdoctoral Scholar, Unit of Applied Mechanics, University of Innsbruck, Innsbruck, Austria

Ting Lin
Assistant Professor, Dept. of Civil, Environmental, and Construction Engineering, Texas Tech University, Lubbock, TX, USA

Greg Deierlein
Professor, Dept. of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA

ABSTRACT: Quantification of collapse risk of buildings in seismically active regions is one of the key elements for informed decision making for building design and establishment of public policies to promote seismic safety and resilience. This paper focuses on development, testing and application of efficient and reliable collapse classification algorithms using machine learning tools. To this end, a large database of structural responses is developed by performing around two million nonlinear time history analyses of an archetype 20-story tall building. Unscaled seismograms simulated for the Los Angeles region as part of the Southern California Earthquake Center (SCEC) CyberShake project are used as inputs for the analysis. Feature selection is performed using regularized logistic regression to identify intensity measures with strong predictive power for classification of collapse. Results of regularization generally confirm the understanding of important predictors as gained from scaling of recorded motions as well as highlight additional important features. Logistic regression and support vector machine (SVM) binary classifiers are then trained on the data to develop collapse prediction models. The resulting collapse assessment models achieve high values of precision and recall and show good performance when tested using benchmark collapse responses. Finally, trained collapse classifiers are utilized to perform regional estimation of collapse risk. Collapse predictions are made using CyberShake data from 336 sites across Southern California where there are around 500,000 simulated seismograms at each site. Regional estimation of mean annual frequency of collapse is performed to generate maps of collapse risk. Higher values of risk correlate well with geologic features such as presence of sedimentary basins and the surface trace of the San Andreas fault.

1. INTRODUCTION

The emergence and maturity of physics-based simulations offer a unique opportunity for application of machine learning tools for efficient evaluation of building response. Collapse response is singled out in this paper as a limit state of practical significance that is at the same time numerically demanding to estimate, particularly for regional estimations. Additionally, conventional approaches of demand estimation based on recorded seismograms usually resort to scaling of seismograms in order to deal with lack of data on extreme motions that can induce collapse. This in turn implicitly requires assumptions on intensity measures that control the response.

In contrast to approaches based on recorded seismograms, using extensive sets of site-specific unscaled motions requires no assumptions about how properties of motions change with scaling and also allows for statistically consistent estimation of regional response. For instance, the Southern California Earthquake Center's (SCEC) CyberShake project [Graves et al. (2011)] performs probabilistic seismic hazard analysis (PSHA) for Southern California by completely relying on numerical simulation of earthquake wave propagation explicitly capturing 3D effects such as seismic focusing and basin amplifications that otherwise could not be assessed by other methods. In this process millions of site-specific ground motions are simulated, including very extreme ground motions, which offers a wealth of data to explore the potential of utilizing earthquake simulations for engineering applications.

Previous work in engineering utilization of ground motion simulations mostly focused on validation [e.g. Jayaram and Shome (2012); Galasso et al. (2013); Burks et al. (2014); Bijelić et al. (2018b)]. For example, it was shown [Bijelić et al. (2018b)] that simulated motions can provide reliable estimates of seismic performance when used in the way in which recorded motions are conventionally used (i.e. ground motions scaled and selected to match hazard targets based on empirically calibrated ground motion prediction models). Further, unscaled physic-based simulations from the SCEC CyberShake project were also used for performance assessment of tall buildings [Bijelić et al. (2018a)] yielding new insights into reliable assessment of collapse risk in the presence of basin effects. This paper takes a step further by applying machine learning techniques to examine the utility of different intensity measures (IMs) for collapse assessment and performing efficient regional collapse risk estimations.

Although the idea of using machine learning for structural response estimations is not new [e.g. Koutsourelakis (2010); Yazdi et al. (2016)] and the efficiency of intensity measures has received a lot of research attention [e.g. Luco and Cornell (2007); Eads et al. (2016)], this is the first time that machine learning tools are applied to a large database of unscaled, site-specific ground motion simulations with specific goals to: 1) identify seismogram features that control collapse response of buildings, 2) develop efficient and reliable collapse classification algorithms, and 3) perform regional collapse risk estimations. To this end, we utilize a structural response database obtained by performing around two million nonlinear time history analyses of an archetype 20-story tall building using input ground motions from the SCEC CyberShake simulations for sites in Southern California. Regularized regression is employed as a primary tool for feature selection considering spectral accelerations and significant durations. Following feature selection, model selection is preformed using logistic regression and support vector machines to highlight the utility of different IMs for collapse prediction in a manner not possible with the limited database of recorded motions. An efficient and reliable collapse classification algorithm is developed and tested using benchmark results. Finally, the developed classifier is utilized to perform regional estimation of collapse risk for Southern California.

2. BUILDING MODEL AND STRUCTURAL RESPONSE DATABASE

2.1. Building model description

An archetype 20-story reinforced concrete special moment is used in this study. The 20-story building was designed as part of a previous benchmark study [Haselton and Deierlein (2007)], according to the governing provisions of the 2003 IBC, ASCE7-02 and ACI 318-02. The frame is idealized as a 2D analysis model using OpenSees [McKenna et al. (2006)], where the first three modal periods are 2.63s, 0.85s and 0.45s. The nonlinearities are captured in concentrated plasticity models in panel zones and plastic hinges at the ends of columns and beams. For additional details regarding the design and modelling assumptions, see [Haselton and Deierlein (2007)].

2.2. Ground motion database

Ground motions used in this study are unscaled, site-specific physics-based hybrid broadband simulated seismograms generated for sites in Southern California (Figure 1) as part of the SCEC CyberShake project [Graves et al. (2011)]. In particular, we utilize ground motions simulated for four CyberShake sites as given in Table 1. The LADT (Los Angeles downtown) is a site of societal importance due to its proximity to a large inventory of tall buildings. The STNI site is situated at one of the deepest basin depths in the regions, where the effects of basin structure on the resulting ground motions are very pronounced [Graves et al. (2011)]. The WNGC and SBSM sites are interesting from the perspective that they exhibit coupling of basin and directivity effects in the ground motions [Graves et al. (2011)].

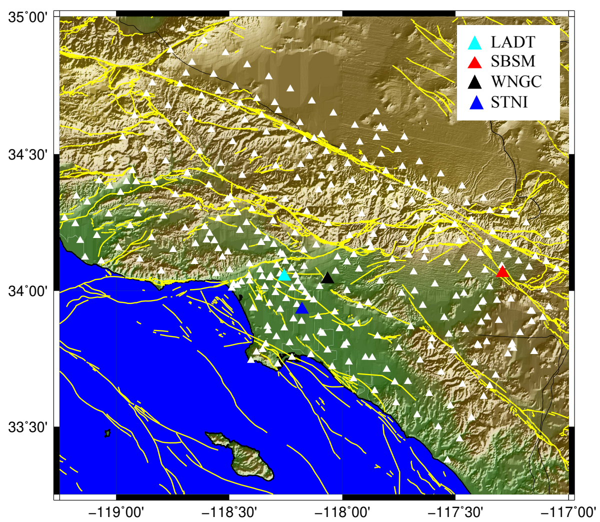

Figure 1: Case study region in Southern California. CyberShake sites indicated with white triangles, faults indicated with yellow lines.

Table 1: CyberShake sites for training data and benchmark results

Site	Latitude	Longitude	Z2.5 [km]	# of seismograms	# of collapses
LADT	34.05	-118.26	2.08	417,954	2,348
STNI	33.93	-118.18	5.57	475,910	17,378
WNGC	34.04	-118.06	2.44	478,210	9,135
SBSM	34.06	-117.29	1.77	483,672	27,873

2.3. *Structural response database*

To obtain the data for application of machine learning techniques, we develop a structural response database by performing nonlinear time history analyses of the archetype 20-story model for all of the seismograms simulated at the considered CyberShake sites (a total of 1,855,746 analyses performed using the Sherlock HPC cluster at Stanford University). No modifications of seismograms are made for the analysis, i.e. as-simulated, site-specific CyberShake seismograms are used as inputs to nonlinear time-history analyses. The database contains information on peak transient story drifts (SDR) at all stories, the timing of the SDRs in the waveform, residual SDRs for each story, roof drift ratio, peak transient floor accelerations (PFA_{rel}), and an indication whether a ground motion induced a collapse. In this paper we limit our attention to collapse responses although other EDPs are also available in our database and will be used in future studies.

3. Feature selection for collapse prediction

To examine the predictive power of commonly used IMs – in particular, Sa values at periods between $0.2T_1$ and $4T_1$, and 5-75% and 5-95% significant durations – feature selection was performed using the L1-regularized logistic regression [Ng (2004)]. The regression, as implemented in LIBLINEAR library [Rong-En et al. (2008)] for scikit-learn (a Python module for machine learning, Pedregosa et al. (2011)), solves the following unconstrained optimization problem:

$$\min_{\theta} \quad ||\theta||_1 + C\sum_{i=1}^{l} \log(1+e^{-y_i\theta^T \mathbf{x}_i}), \tag{1}$$

where θ are coefficients associated with features $\mathbf{x}$ and $||\theta||_1$ is the L1-norm. Parameter C is a cost parameter that penalizes the absolute values of coefficients, i.e. 'encourages' the sum of the absolute values of the model parameters to be small [e.g. Ng (2004)]. Note that this formulation involves minimization and applies the cost parameter to the inverse of the logistic function. Therefore, models with smaller values of C penalize the coefficients more heavily effectively causing more coefficients to shrink towards zero (those models are hence termed 'less complex' models).

Shrinkage of coefficients with cost parameter C along with the corresponding measure of model performance is shown in Figure 2. Models with different values of parameter C were examined by performing 5-fold cross validation on 70% of the data; class membership ratios were preserved when splitting the data into folds. Since different IMs have different units and span different numerical ranges, features were standardized (zero mean, unit variance) to give them equal importance in regularization [Hastie et al. (2009)]. Algorithm performance was measured using precision-recall for following reasons: (a) use of precision (fraction of examples classified as positive that truly are positive) rather than the false positive rate (ratio of false positive classifications over total number of negative samples) captures the effect of class imbalance on algorithm performance [Davis and Goadrich (2006)]; (b) use of area under the precision-recall curve (AUC-PR) rather than a single-threshold value (e.g. precision) is informative from the viewpoint of algorithms ability to minimize both false positive as well as false negative errors which is important from application perspective. In order to get reliable estimates of the AUC-PR, integrations and threshold averaging were performed as recommended in [Davis and Goadrich (2006); Fawcett (2004)].

The values of coefficients associated with different intensity measures that are preserved in different models (i.e. for different values of cost parameter C) are shown in Figure 3. The figure indicates the results of the regularization and a number of observations can be made. First, as shown in Figure 3a, spectral accelerations for periods past the fundamental structural mode (T_1), in particular $\mathrm{Sa}(1.5T_1 < T < 3T_1)$, have primary importance for collapse as they appear in the least complex models (see Figure 2). This is consistent with previous observations based on recorded motions [e.g. NIST (2011)]. Further, the Sa values at very long periods ($\sim$3-4T_1) are also included in the best performing models (Figure 3b) – an observation that is generally not reported in the literature and may stem from the capability of physics-based simulated earthquakes to better characterize longer period energy content. The best performing model additionally includes significant durations as predictors (Figure 3c), conforming with recent findings from a study of earthquake ground motion duration effects on structural response [Chandramohan et al. (2015)]. Finally, including Sa values at higher structural modes ($0.2T_1$<T<$0.5T_1$) as predictors generally does not improve classifier performance. Further increasing model complexity results in added weight to Sa values at small periods without improvement in model performance (suggesting overfitting problems, Figure 3d).

4. MODEL SELECTION FOR COLLAPSE CLASSIFICATION

Using the results of feature selection, we next perform model selections with the objectives to: 1) examine the utility of different IMs for collapse classification, and 2) explore machine learning techniques for development of reliable and efficient collapse prediction algorithms for an archetype tall building. Based on the performed feature selection, the following features are kept in consideration when performing model selection: Sa at periods between $0.2T_1$ and $4T_1$, significant duration (Da5-75%), and $\mathrm{SaAverage}(0.2T_1 - 3.5T_1)$. When comparing different models, we use precision-recall curves as a measure of algorithm performance.

To interpret the utility of different IMs for collapse classification we use relative performance of linear logistic regression classifiers. For instance, using $\mathrm{SaAverage}(0.2T_1 - 3.5T_1)$ as a single predictor yields significantly better performance than $\mathrm{Sa}(T_1)$ or $\mathrm{Sa}(2T_1)$, as seen from precision-recall curves shown in Figure 4, where classifier performance is related to the area under the curve (AUC-PR). Further, including significant duration in addition to SaAverage improves model performance by bringing in the information on duration of ground motions. The best performing models are obtained by including Sa spectral acceleration values at a range of different periods. While not shown in the figure, it is noted that in the case of using $\mathrm{Sa}(T > 0.2T_1)$ as predictors, the addition of significant duration did not improve the model performance. This suggests that the earthquake magnitude, which is strongly correlated to duration, is already reflected in the Sa values of the unscaled motions. Note that this observation would not apply when ground mo-

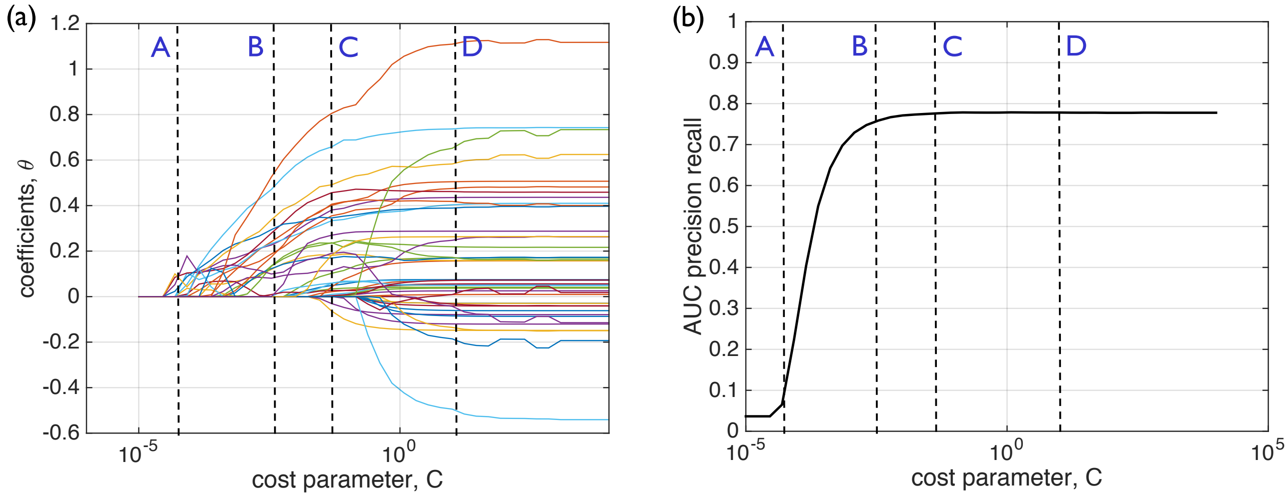

Figure 2: Regularized logistic regression: (a) standardized coefficients vs. cost parameter – shrinkage of the coefficients; (b) algorithm performance, measured as area under the precision recall curve (AUC-PR) vs. cost parameter. Vertical lines indicated with A-D represent snapshots of the shrinkage process; associated features that are preserved in the model are indicated in Figure 3.

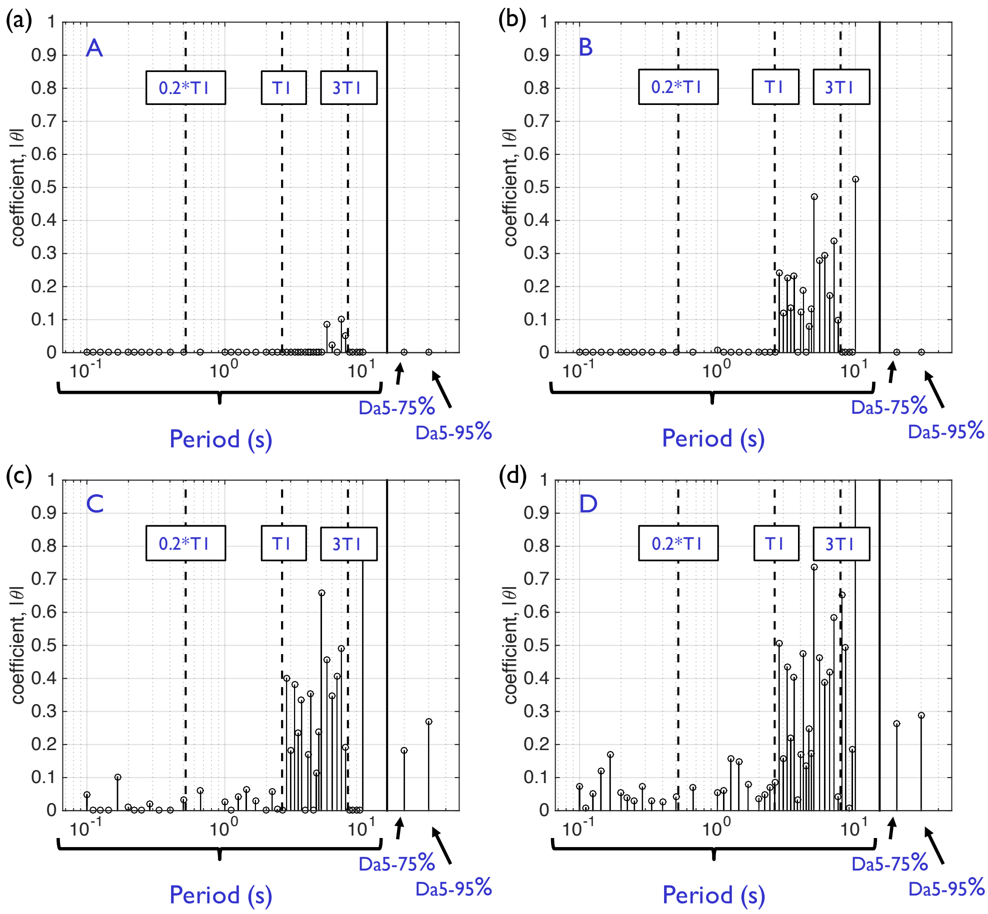

Figure 3: Feature selection using regularized logistic regression – importance of different intensity measures for prediction of collapse. Values of cost parameter associated with cases A-D are indicated in Figure 2.

tions are scaled, such as it is typically done using recorded motions.

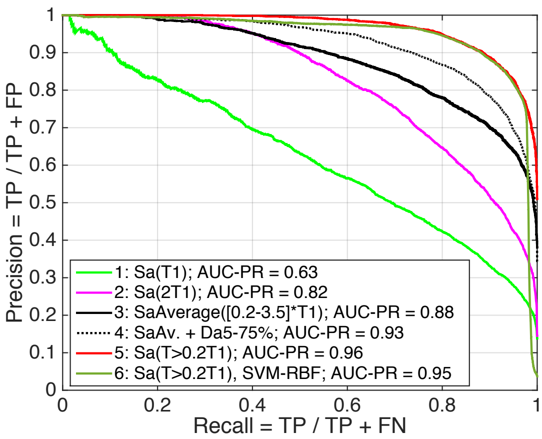

Figure 4: Collapse classification using logistic regression and SVM – precision-recall curves. TP - true positive, FN - false negative, FP - false positive.

The observations in the previous paragraph are all based on logistic regression models with linear features. To gauge the benefit to be gained from consideration of nonlinear combinations of predictive features, we considered the support vector machine (SVM) algorithm with Gaussian kernel (radial basis function, RBF) and trained with all Sa features (note: using the Gaussian kernel corresponds to training the model using an infinite dimensional feature mapping [Ng (2015)]). The SVM was implemented using LIBSVM [Chih-Chung and Chih-Jen (2013)] and can be described by the following optimization problem:

$$\begin{aligned} \min_{\theta,b,\xi} \quad & \frac{1}{2}\theta^T\theta + C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i(\theta^T\phi(\mathbf{x}_i)+b) \geq 1-\xi_i, \\ & \xi_i \geq 0, i=1,\ldots,l, \end{aligned} \tag{2}$$

where the used Gaussian kernel is of the form:

$$\mathbf{K}(\mathbf{x}_i,\mathbf{x}_j) = e^{-\gamma||\mathbf{x}_i-\mathbf{x}_j||^2} \tag{3}$$

Parameters C and γ were determined using a grid search with 70% of the data yielding final parameters $C=1$, and $\gamma=0.01136$; standardized features were used in the model. As shown in the Figure 4, the SVM model achieves essentially the same performance as its logistic regression counterpart. We also note that using the SVM algorithm is fairly more involved than using logistic regression models.

5. Test of collapse classification algorithm on benchmark sites

We next examine the question of how well do the collapse classifiers trained with data from one site generalize to other sites. Figure 5 shows the comparison of the predictions of mean annual frequency of collapse (λ_c) at case study sites obtained using an algorithm trained with data from WNGC site. It can be seen that the algorithm achieves a relatively good performance with differences between predicted values and benchmark results ranging from around 20% to 30% depending on the site. While these values seem relatively large, in the context of discrepancies in computing the λ_c using the well-established approaches (see for instance approximation of direct analysis results with multiple stripes analysis (MSA) combined with generalized conditional intensity measure (GCIM) based selection discussed in [Bradley et al. (2015)]) they are deemed acceptable. However, there is still room for improvement which could potentially be achieved by considering additional intensity measures as well as including source, rupture and site properties as predictive features. Such fine-tunings are left for future study. Additionally, one of the future goals is to predict entire seismic demand curves for instance by training neural networks on the developed structural response database. Finally, we mention that such algorithms are very computationally efficient – training takes around 4min on a contemporary personal computer while making predictions for hundreds of thousands of samples takes less than a second. This makes such algorithms particularly attractive for risk studies at a regional level as described in the next section.

6. Regional collapse risk estimation

CyberShake study area consists of 336 sites across Southern California (Figure 1) and around 500,000 seismograms with two horizontal components are simulated at each site. The simulated seismograms are available on SCEC servers, while the data on

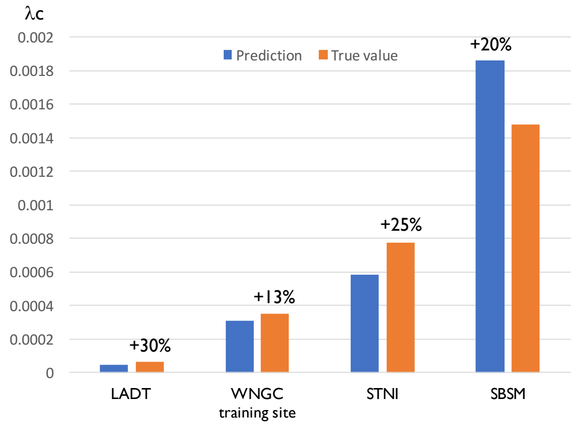

Figure 5: Predictions of mean annual frequency of collapse (λ_c) using a collapse classifier trained at the WNGC site.

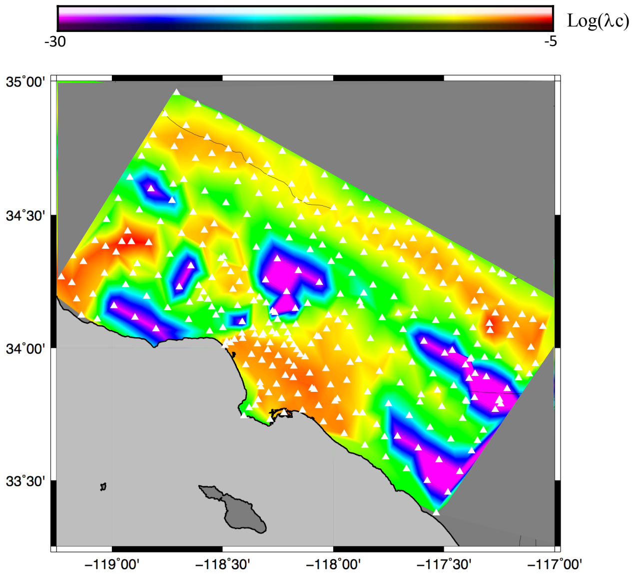

Figure 6: Collapse risk (λ_c) of an archetype 20-story tall building for Southern California.

intensity measures is stored in an SQL database which allows for easy access.

We train the collapse classifier with data from WNGC site and then use the trained classifier to make collapse predictions at each of the CyberShake sites while also computing the mean annual frequency of collapse (λ_c), as shown in Figure 6. Since the λ_c values are very small, the map plots the log(λ_c) to better distinguish between different values. As seen from the figure, higher values of collapse risk correlate well with locations of sedimentary basins (e.g. Los Angeles, San Bernardino and San Fernando basins). Additionally, higher values of risk are observed along the San Andreas fault predominantly reflecting the influence of rupture directivity. Areas of very small λ_c values are observed in mountainous regions.

7. Conclusions

With the broader goal of informing building design and improving seismic risk assessment of tall buildings, this work focused on development, testing, and application of efficient and reliable collapse classification algorithms for regional collapse risk assessment. A large structural response database was developed by performing around two million nonlinear response history analyses of an archetype 20-story tall building using CyberShake simulated seismograms. The database was used to contrast the utility of different intensity measures for collapse prediction, and to develop efficient and reliable classification algorithms for prediction of collapse. The developed collapse classifier was utilized to perform collapse risk estimation for Southern California using CyberShake data as inputs.

The analyses performed in this paper exemplify the possibility of physics-based simulated ground motions to provide novel insights to questions of engineering concern in ways not possible with recorded ground motions. Additionally, they open a number of directions for future exploration. First, given the success of the applied regularization, the efficiency of an arbitrary set of earthquake intensity measures can be analyzed. This can include previously unused features or end-to-end machine learning approaches which can prove to be particularly useful for generalization of the models to different geological settings. Furthermore, the developed structural response database can be utilized to characterize different damage states of the building prior to collapse using e.g. multinomial classification. Finally, analyses similar to the ones performed here can be used on a range of different building types; having building properties as predictors would enable development of models to simulate seismic performance of a diverse building stock allowing for loss assessment studies on community scale. Thus, the classification models developed in this study could ultimately be used for seismic risk mitigation.

8. REFERENCES

Bijelić, N., Lin, T., and Deierlein, G. (2018a). "Evaluation of building collapse risk and drift demands by nonlinear structural analyses using conventional hazard analysis versus direct simulation with CyberShake seismograms." *Bulleting of the Seismological Society of America*, (submitted for review).

Bijelić, N., Lin, T., and Deierlein, G. (2018b). "Validation of the SCEC Broadband Platform Simulations for Tall Building Risk Assessments Considering Spectral Shape and Duration of the Ground Motion." *Earthquake Engineering and Structural Dynamics*, 47, 2233–2251.

Bradley, B. A., Burks, L. S., and Baker, J. W. (2015). "Ground motion selection for simulation-based seismic hazard and structural reliability assessment." *Earthquake Engng Struct. Dyn.*, 44(13), 2321–2340.

Burks, L. S., Zimmerman, R. B., and Baker, J. W. (2014). "Evaluation of Hybrid Broadband Ground Motion Simulations for Response History Analysis and Design." *Earthquake Spectra*, 31(3), 1691–1710.

Chandramohan, R., Baker, J. W., and Deierlein, G. G. (2015). "Quantifying the influence of ground motion duration on structural collapse capacity using spectrally equivalent records." *Earthquake Spectra*, 32(2), 927–950.

Chih-Chung, C. and Chih-Jen, L. (2013). "LIBSVM: A Library for Support Vector Machines., <http://www.csie.ntu.edu.tw/ cjlin/papers/libsvm.pdf>.

Davis, J. and Goadrich, M. (2006). "The Relationship Between Precision-Recall and ROC Curves.." *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, PA.

Eads, L., Miranda, E., and Lignos, D. (2016). "Spectral shape metrics and structural collapse potential." *Earthquake Engng Struct. Dyn.*, 45(10), 1643–1659.

Fawcett, T. (2004). "ROC Graphs: Notes and Practical Considerations for Researchers." *Pattern Recognition Letters*, 31(8), 1–38.

Galasso, C., Zhong, P., Zareian, F., Iervolino, I., and Graves, R. W. (2013). "Validation of ground-motion simulations for historical events using MDoF systems." *Earthquake Engng Struct. Dyn.*, 42(9), 1395–1412.

Graves, R., Jordan, T. H., Callaghan, S., Deelman, E., Field, E., Juve, G., Kesselman, C., Maechling, P., Mehta, G., Milner, K., Okaya, D., Small, P., and Vahi, K. (2011). "CyberShake: A Physics-Based Seismic Hazard Model for Southern California." *Pure Appl. Geophys.*, 168(3-4), 367–381.

Haselton, C. B. and Deierlein, G. (2007). "Assessing seismic collapse safety of modern reinforced concrete moment frame buildings." *Technical Report 2007/08*, Pacific Earthquake Engineering Research Center, Berkeley, CA.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Jayaram, N. and Shome, N. (2012). "A Statistical Analysis of the Response of Tall Buildings to Recorded and Simulated Ground Motions." *15WCEE*, Lisboa, Portugal (September).

Koutsourelakis, P. S. (2010). "Assessing structural vulnerability against earthquakes using multi-dimensional fragility surfaces: A Bayesian framework." *Probabilistic Engineering Mechanics*, 25(1), 49–60.

Luco, N. and Cornell, C. A. (2007). "Structure-Specific Scalar Intensity Measures for Near-Source and Ordinary Earthquake Ground Motions." *Earthquake Spectra*, 23(2), 357–392.

McKenna, F., Fenves, G., and Scott, M. H. (2006). "OpenSees: Open system for earthquake engineering simulation, <http://opensees.berkeley.edu>.

Ng, A. (2004). "Feature selection, L1 vs. L2 regularization, and rotational invariance.." *Proceedings of the twenty-first international conference on Machine learning*, New York, NY, USA.

Ng, A. (2015). "CS229 Machine learning lecture notes: Support Vector Machines, <http://cs229.stanford.edu/notes/cs229-notes3.pdf>.

NIST (2011). "Selecting and Scaling Earthquake Ground Motions for Performing Response-History Analyses." *Report No. NIST GCR 11-917-15*, National Institute of Standards and Technology (November).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research*, 12, 2825–2830.

Rong-En, F., Kai-Wei, C., Cho-jui, H., Xiang-Rui, W., and Chih-Jen, L. (2008). "LIBLINEAR: A Library for Large Linear Classification.." *Journal of Machine Learning Research*, 9, 1871–1874.

Yazdi, A. J., Haukaas, T., Yang, T., and Gardoni, P. (2016). "Multivariate Fragility Models for Earthquake Engineering." *Earthquake Spectra*, 32(1), 441–461.