LOCALIZED DAMAGE DETECTION OF CSMIP INSTRUMENTED BUILDINGS USING CUMULATIVE ABSOLUTE VELOCITY: A MACHINE LEARNING APPROACH

Sifat Muin[1] and Khalid M. Mosalam[1,2]

[1.] Department of Civil and Environmental Engineering, University of California, Berkeley
[2.] Pacific Earthquake Engineering Research (PEER) Center

Abstract

Post-earthquake damage assessment can be significantly expedited when machine learning (ML) algorithms are used. Recent earthquakes showed that even when a structure is operational and safe for occupancy, people chose to evacuate and not reoccupy it immediately. Such a behavior can be attributed to lack of knowledge about the structural conditions immediately following the event and the fear of being trapped in the building if aftershocks hit. Currently, there is a lack of rapid quantifiable methods to determine if buildings are safe for reoccupation after an extreme event. However, advances in remote sensing, computing technologies, and data science in the past few years paved the way to develop ML methods that can assess and quantify the conditions of structures in near-real time. This paper introduces a methodology to assess the severity of earthquake-induced damage using low dimensional, cumulative absolute velocity (CAV)-based feature and ML tools. The appropriate features and the ML tool are identified by analyzing a single degree of freedom (SDOF) model. The identified features are then applied to assess the severity and location of damage of two multi-degree of freedom (MDOF) systems and real structures instrumented by the California Strong Motion Instrumentation Program (CSMIP). Results show that the damage detection capability of the features is high.

Introduction

Structural Health Monitoring (SHM) is the process of developing automated and online damage detection and/or assessment capability for all types of engineered systems (aerospace, civil, mechanical, etc.). It has become an important field of engineering in the US as the civil infrastructure systems of the country are aging. According to ASCE infrastructure reports (ASCE 2017), most US infrastructures are rated between mediocre or poor. Many of them are nearing the end of their design life and show signs of deteriorations. Since replacing all these structures is not feasible, SHM is necessary to monitor the structural integrity and assess deterioration for the safe and continuous operation of these infrastructures and also for prioritizing to the decision makers their retrofit or replacement actions. Advances in remote sensing, computing technologies, and data science in the past few years paved the way to develop SHM techniques that can assess and quantify the conditions of structures in near-real time utilizing machine learning (ML) techniques.

Applications of ML in damage detection have been studied for a long time. One class classifier or novelty detection analysis is one of the most popular damage detection techniques among SHM researchers. Worden et al. (2000) applied such a technique to detect damage in a

three-degree of freedom spring system. They used Mahalanobis squared distance as the discordancy measure and detected outliers if the measure is greater than a threshold. Other researchers implemented auto-associative neural network (AANN) successfully to detect damage (Dua et al. 2001; Sohn et al. 2002). Cluster analysis has also been applied to detect damage in recent years (Kesavan and Kiremidjian 2012; Santos et al. 2014). However, this analysis technique is more frequently used by researchers to classify (rather than detect) damage (Tibaduiza et al. 2012; Palomino et al. 2012).

The studies mentioned above were conducted with simulated or experimental data. The damage features used to detect or classify damage are high dimensional which require a large amount of training data that were available from the simulations or the experiments. However, existing structures have limited data when it comes to earthquake response and particularly damage under extreme events. In this paper, low dimensional, cumulative absolute velocity (CAV)-based features are proposed that can be used with a limited dataset for seismic SHM.

Development of Single Dimensional Features

The CAV has been used as an earthquake intensity measure since a study conducted by the Electric Power Research Institute (EPRI) that dates back to 1990 (Reed and Kassawara 1990). In that study, it was found that ground motion CAV has a better correlation to damage than other intensity measures such as peak ground acceleration (PGA), spectral acceleration, effective durations, etc. However, CAV of floor accelerations has never been applied in SHM to assess damage. In a recent study by Muin and Mosalam (2017), CAV is introduced as a damage feature. CAV time series, its normalized version (NCAV), and other features extracted from it show distinct patterns in damaged structures which can be used to identify and locate damage. Since it is a waveform-based feature, this method does not contain any assumptions regarding the linear or nonlinear nature of the system that generates the waveform data. However, waveform-based damage features are high dimensional where the dimension is the number of scalar quantities that are necessary to describe the feature. The amount of training data required for accurate diagnostic grows explosively with the dimension of the feature (Farrar and Worden 2012). This is a challenge for damage classification problems as the amount of available data from damaged cases is low. In this paper, several CAV-based damage features are further studied where these features are low dimensional and therefore appropriate to be used in a ML computing environment with a limited dataset. Four proposed damage features are discussed below.

Proposed Features

The CAV is mathematically defined as follows:

$$CAV = \int_0^T |\ddot{u}(t)|dt \tag{1}$$

where $|\ddot{u}(t)|$ is the absolute value of acceleration at time t and T is the total duration of the recorded acceleration time history. For the CAV calculation, the considered acceleration is the floor accelerations of a building. Higher CAV value is expected in damaging events than in undamaged cases as damaging events are correlated to high amplitude motions (Wald et al. 1999; Hancock and Bommer 2006).

The R_{CAV} is mathematically defined as follows:

$$R_{CAV} = \frac{CAV_s}{CAV_l} \tag{2}$$

where CAV_s is the CAV of the floor acceleration representing the structural response and CAV_l is the CAV of the corresponding linear system excited by the same ground acceleration. For an undamaged case and accurate linear model, this value will be 1. With damage, acceleration amplitude will typically decrease compared to the linear case due to lengthening of the natural period. Therefore, R_{CAV} is expected to decrease with increasing damage states.

The definition of S_{CAV} is as follows:

$$S_{CAV} = \frac{D_{5-75,s} - D_{5-75,l}}{D_{5-75,l}} \times 100\% \tag{3}$$

It highlights the relative wave travel times between the actual structure (subscript *s*) and its linear counterpart (subscript *l*) providing insight into the change in the wave propagation behavior caused by damage. Higher S_{CAV} means a slower rate of change due to damage. The parameter D_{5-75} in Equation 3 is the effective duration of an earthquake defined by the time to achieve 75% of the final CAV value starting from the 5% of that value (Bommer and Martinez-Pereira 1999). For an undamaged case, this value will be zero. With damage, S_{CAV} is expected to increase.

The Δ_{CAV} is calculated by taking the absolute value of the difference between the area under the CAV plots of an actual event and the corresponding linear state. The change of pattern in CAV time series and its linear counterpart provides useful information about the damage. However, comparison at each point of observation makes it a very high dimensional feature. Hence, the area is calculated as a compact (low dimension) feature.

$$\Delta_{CAV} = \left(\frac{|A_s - A_l|}{A_l}\right) \times 100 \tag{4}$$

where A_s is the area under the CAV plot of structure and A_l is the area under the CAV plot of the corresponding linear system. This value is expected to increase with damage, while for an undamaged case, it will be zero.

Methodology

A single degree of freedom (SDOF) system is utilized in this study. This model is used to select appropriate features and ML technique for the problem at hand. The SDOF model is developed in OpenSEES (McKenna 2010) using Steel01 material which has a bilinear behavior with strain hardening of 1% as shown in Figure 1. The base shear coefficient (η), which is defined by the ratio of yield base shear (*Vy*) to the weight of the building (*W*), is assumed to be 0.2.

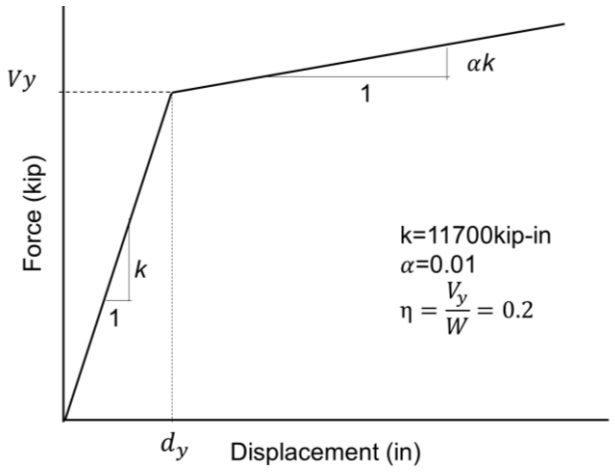

Figure 1 Story force-displacement relationship.

Two different sets of data have been used in this study. Set-1 includes responses of ground motions from the PEER NGA-West2 (Bozorgnia et al. 2014) database. This database has 21,539 records of shallow crustal earthquakes in active tectonic regions. Due to the possibility of anomalies from older data collection systems, only records from past 30 years have been selected. Moreover, records with PGA less than 1% g will not produce enough excitation useful for this study. Therefore, only records with PGA more than 1% g are considered. Lastly, to avoid homogeneity in response, not more than 20 records from a single event are selected. A total of 1,710 records matched these criteria. Set-2 comprises responses to site-specific ground motions that are used by Baker et al. (2011). These ground motions are selected by matching the uniform hazard spectrum and associated causal events for a site in Oakland, California. Set-2 consists of 120 ground motions representing three hazard levels, namely 2%, 10% and 50% probabilities of exceedance in 50 years. For the purpose of this study, damage states are defined based on displacement ductility, which is the most commonly used index to quantify structural damage. It is defined as the ratio of the maximum displacement sustained by the structure to its yield displacement as follows.

$$\mu = \frac{d_s}{d_y} \tag{5}$$

where d_s is the maximum absolute displacement of the structure and d_y is the yield displacement. If the displacement does not exceed this yield displacement, i.e. $\mu \leq 1$, then the structure is considered undamaged. The damage states are divided into three categories according to guidelines from FEMA P-58 (FEMA 2012) with $1 < \mu \leq 2$ as minor damage, $2 < \mu \leq 6$ as moderate damage, and $\mu > 6$ as major damage.

Nonlinear time history analysis (NTHA) is performed on the SDOF model with set-1 and set-2 ground motions. In the analysis, acceleration of the model is computed and the force-displacement of the spring is documented. Damage state is assigned by determining the maximum absolute displacement and consequently calculating μ according to Equation 5. For set-1, the structure remains undamaged for 1,215 (71%) records while 308 (18%) records cause minor damage, 143 (8%) records cause moderate damage, and 45 (3%) records cause major damage. This is representative of a database that collects data from real instrumented structures

where the majority of the collected data will be coming from undamaged structures and very few from severely damaged structures. For set-2, 5% causes minor damage, 29% moderate damage, and 66% major damage.

Results

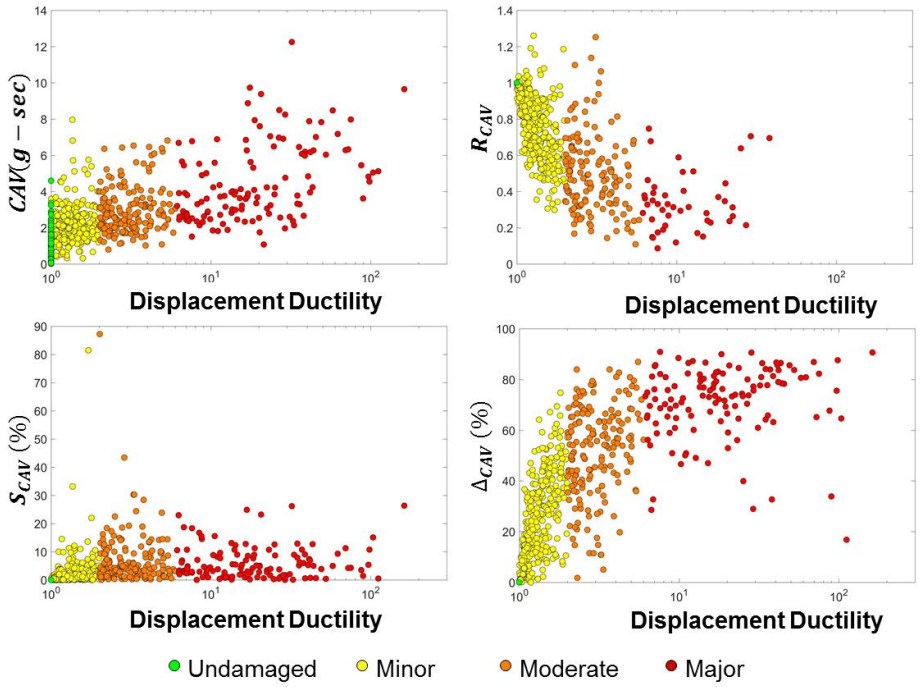

Figure 2 Plots showing relationship of proposed features and displacement ductility for four different damage states.

To identify suitable features, the relationship between each feature and the different damage states are observed. A good feature will demonstrate a certain pattern with increasing damage. Figure 2 shows the relationship of each of the four considered features with that of displacement ductility for the entire datasets (i.e. set-1 and set-2). Three of these features show trends with damage. The CAV value shows an increasing trend with the increase of displacement ductility. The R_{CAV} can distinguish minor damaged and undamaged cases as the R_{CAV} values other than 1 are damaged cases. Although for significant damage, acceleration of the structure and subsequently R_{CAV} will decrease, for certain ground motions, small damage may lead to an increase in acceleration and subsequently increase in the R_{CAV}. These are the motions for which the undamaged structure lies on the initial ascending part of the spectra.The Δ_{CAV} also shows a trend with displacement ductility. The S_{CAV} does not show any specific trend with displacement ductility. Therefore, S_{CAV} is not suitable as a damage feature.

A comparative analysis of four ML tools is performed using the three CAV-based features, namely CAV, R_{CAV}, and Δ_{CAV}, to determine the ideal feature and ML tool. The considered ML tools are logistic regression (LR), ordinal logistic regression (OLR), artificial neural network (ANN) with 10 (ANN_{10}) and 100 (ANN_{100}) neurons, and support vector machine (SVM). Two different training set and test set combinations are used for this purpose. The first

training set (TR-1) comprises of randomly sampled 85% of set-1 data as training set and the remaining 15% as the first test set (TE-1), i.e. both training and testing are performed with data from the same distribution. This is the usual practice in the ML field. The second training set (TR-2) comprises of the entire set-1 and second test set (TE-2) is set-2. The second combination is chosen in order to test the robustness of the features when tested against extreme values.

Table 1 Accuracy (%) achieved by ML models with different feature combinations with training and test sets from set-1.

Input Feature	OLR	LR	ANN_{10}	ANN_{100}	SVM
CAV	80.54	82.88	80.54	81.71	79.38
R_{CAV}	87.16	86.72	88.72	89.49	88.33
Δ_{CAV}	75.10	75.10	75.10	77.04	75.10
CAV, R_{CAV}	90.27	89.44	88.72	90.66	**91.05**
R_{CAV}, Δ_{CAV}	86.77	84.72	89.11	87.94	87.94
CAV, Δ_{CAV}	80.54	83.27	80.54	81.32	79.38
$CAV, R_{CAV}, \Delta_{CAV}$	90.27	89.05	90.27	90.66	89.88

Table 2 Accuracy (%) achieved by ML models with different feature combinations for set-1 as training set and set-2 as test set.

Input Feature	OLR	LR	ANN_{10}	ANN_{100}	SVM
CAV	36.67	12.50	18.33	15.83	8.33
R_{CAV}	60.00	42.50	30.83	37.50	20.83
Δ_{CAV}	61.67	45.00	42.50	40.00	21.67
CAV, R_{CAV}	**74.14**	61.67	18.33	40.00	25.00
R_{CAV}, Δ_{CAV}	65.83	45.00	60.00	40.00	22.50
CAV, Δ_{CAV}	70.00	60.00	51.67	36.67	24.17
$CAV, R_{CAV}, \Delta_{CAV}$	70.00	61.67	38.33	54.17	25.00

Tables 1 and 2 report the accuracy achieved by each model for the two combinations, respectively. Table 1 shows that the highest accuracy of 91.05% is achieved by the SVM model with CAV and R_{CAV} as features. ANN_{100} and OLR achieve comparable accuracies of 90.66% and 90.27%, respectively, with the same features. As expected, when the dataset is big enough in size and the data are representative of the population (i.e. training and test sets come from the same target unknown distribution), the results are not affected significantly by the choice of the ML algorithm. On the other hand, when the ML algorithms are tested over a different set of data, their predictive capabilities are significantly reduced (Table 2). For this case, the accuracy reduces for each model, where OLR achieves the highest accuracy of 74.14%, also with CAV and R_{CAV} features. This is due to the fact that ANN and SVM overfit the data of the training set and have poor generalization capabilities over the response of completely different ground motions. On the other hand, the simpler models LR and OLR have better generalization capabilities. From Table 2, it is seen that the OLR appears to be the most robust algorithm for making predictions

about events not available in the training set. For this reason, along with the fact that the computational demand of OLR is significantly smaller than ANN or SVM, OLR is used as the ML tool for the multi-degree of freedom (MDOF) analysis, as described in the next section. Moreover, it is seen that satisfactory performances are obtained using R_{CAV}, while the best ones are achieved using both CAV and R_{CAV} as features.

Damage Assessment of a MDOF Structure

Methodology

The identified features and the ML tool in the previous section are applied to MDOF systems representing a 5 story structure to evaluate the damage assessment performance. These models are also developed in OpenSEES (McKenna 2010) using the same bilinear steel model as the previously discussed SDOF system. The mass, stiffness, and damping of each story of the MDOF system are based on pushover and eigenvalue analyses reported by Mahin et al. (2015) and Günay and Mosalam (2017). The base shear coefficient (Figure 1) is taken as $\eta = 0.2$ which is representative of the value recommended by the code (ICC 2012) for regular structures designed for seismic risk category D.

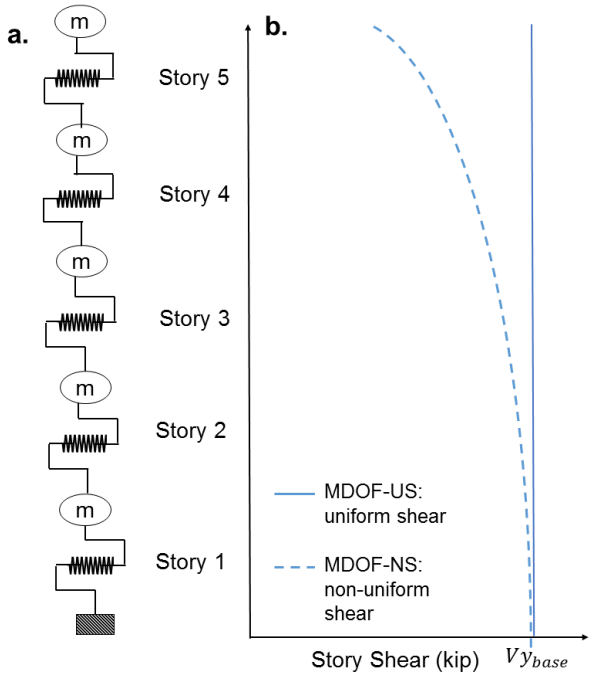

Figure 3: (a) MDOF model used in the study, (b) Story shear distribution along the height for the two models.

Figure 3a shows the MDOF representation of the 5 story structure and Figure 3b shows the story shear distribution of two different systems. One of the MDOF systems (MDOF-US) has uniform shear capacity along the height of the system which is equal to the calculated base shear (Vy_{base}). The other MDOF system (MDOF-NS) is designed to have non-uniform shear capacity distribution along the height of the structure. Designs of both systems are code conforming. However, MDOF-NS marginally meets the code and may be considered as the code minimum

design. Whereas MDOF-US has significantly higher value throughout the height and can be considered as the more conservative design which is often observed in low to medium rise structures. NTHA is performed on the models using both set-1 and set-2 ground motions, previously discussed. The acceleration is recorded at each degree of freedom along with force and displacement for each spring element. The worst damage states among the five stories are assigned as the damage state of the entire structure and its location as the worst damage location. However, for the cases when the worst damage state occurs simultaneously at several locations, the lowest story is identified as the damage location.

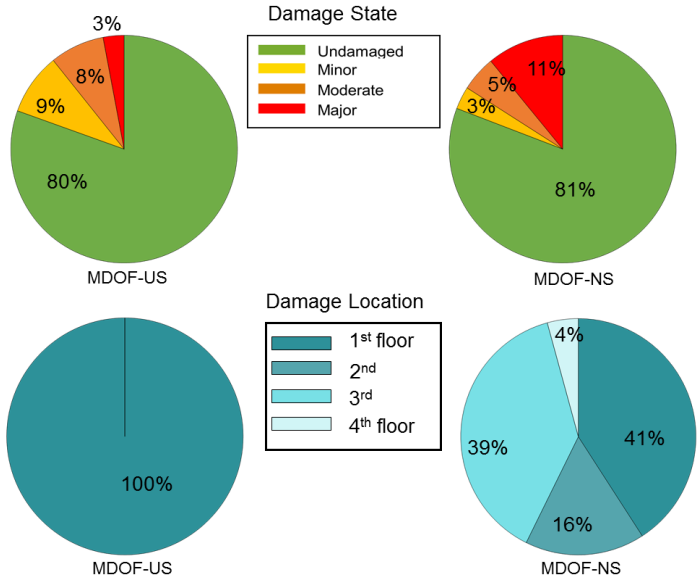

Figure 4: Worst damage state and its location distributions of the MDOF-US and MDOF-NS models.

For the MDOF-US model, out of the 1,710 cases of set-1, 1,376 (80%), 150 (9%), 133 (8%), and 51 (3%) cases are respectively undamaged, minor damaged, moderate damaged, and major damaged (Figure 4). For the MDOF-NS model, 1,382 (81%), 55 (3%), 86 (5%), and 187 (11%) cases are respectively undamaged, minor damaged, moderate damaged, and major damaged. For MDOF-US model, the worst damaged state of all 334 damage cases occurred at the first story. For MDOF-NS, 328 damaged incidents are distributed primarily between the first and third stories mainly due to the increased demand in the first story and reduced capacity of the third story.

The CAV of the first floor and the R_{CAV} of the top floor are used in this study. OLR is trained with set-1 and tested against TE-1 and TE-2 for both MDOF-US and MDOF-NS with damage categories defined similar to the case of the SDOF system. Using the damage state and location information obtained from the training set, the probability of the location of damage for a given damage state $P(n|y)$ is calculated where n is the story number, i.e. n =1 to 5 and y is the damage state, i.e. y =1, 2, and 3 for minor, moderate, and major, respectively. The probability of the worst damage location is subsequently determined using Equation 6 where the location with the highest probability is selected as that for the worst damage occurrence.

$$P(n) = P(n|y) \times P(y) \quad (6)$$

Results

Results reveal that MODF-US achieves damage state detection accuracy of 90.67% when tested with TE-1 and 84.00% when tested with TE-2. The higher accuracy with TE-1 is expected since the test set comes from the same distribution as the training set. Detection accuracy of damaged states for MDOF-NS (90.67%) is the same as that of the MDOF-US (90.67%) when tested with TE-1. However, the accuracy improves significantly to 96.67% when tested with TE-2. Table 3 presents the class-specific recall (i.e. fraction of relevant instances that have been retrieved over the total amount of relevant instances) values for the two MDOF systems for the two test sets TE-1 and TE-2. From this table, it is evident that for MDOF-US, the model predicts very well the undamaged class (99.3%) and the major damage class (92.2%). For moderate damage, it does fairly with 78.1% recall values. However, minor damage is mostly misclassified (28.6%). This is mainly due to the definition of the damage states. Since minor damage is defined as $1 \leq \mu \leq 2$, with this narrow range, the model performed poorly in detecting minor damages which was further worsened by the lack of data from minor damage states. Similar performance of the model is also observed for MDOF-NS where major damage class has high recall value (96.6%) together with the undamaged cases (99.3%). As the test cases for MDOF-NS has more data coming from the major damage and undamaged states, the results showed better overall accuracy for MDOF-NS than that of the MDOF-US.

Table 3: Class-specific recall values for the two MDOF models.

Class	MDOF-US	MDOF-NS
Undamaged	99.3	99.3
Minor	28.6	00.0
Moderate	78.1	46.3
Major	92.2	96.6

Damage locations are detected with 97.5% accuracy for MDOF-US with both TE-1 and TE-2. It is noted that three inaccurate cases are detected where the worst damage took place in both first and second stories, i.e. first story is labeled as the correct location per the definition stated above, but the model only identified second story as the worst damage location. The damage locations are detected with 93.0% accuracy and 95.0% accuracy with TE-1 and TE-2 for MDOF-NS. Damage location detection for this model is more critical since the non-uniformity of strength introduces significant uncertainty on the damage location. Thus, the results show that the location of damage can be identified with high confidence even for non-uniform structural properties by using this approach, i.e. using CAV and R_{CAV} as features with the OLR method.

Damage Detection of CSMIP Instrumented Buildings

In this section, damage assessment is conducted for real structures that are instrumented under the California Strong Motion Instrumentation Program (CSMIP). CSMIP was established in 1972 by California Legislation to obtain vital earthquake data for the scientific and the engineering communities through a statewide network of strong motion instruments. Although

the majority of instrumentations are installed at ground response stations, CSMIP also instrumented structures such as buildings, hospitals, bridges, dams, utilities, and industrial facilities. This study is focused primarily on buildings.

Building Portfolio

There is a significant number of CSMIP instrumented buildings studied in the literature, e.g. 151 instrumented CSMIP buildings studied by Fathali and Lizundia (2011), 64 instrumented CSMIP steel and reinforced concrete (RC) buildings in (Harris et al. 2015) and more than 40 CSMIP instrumented buildings in (Naeim et al. 2006), where data were collected from a significant number of earthquakes. Some of the previously studied buildings and the corresponding recorded strong motions are considered in this study. Moreover, buildings that captured responses of multiple earthquakes are ideal for this study. Several buildings selected from this large pool are investigated for the modeling of undamaged conditions. Table 4 provides a list of buildings selected in this study covering a wide range of primary lateral force resisting systems (PLFRS) and a variety of heights. The selection is influenced by having a reasonable number of earthquake data sets for each building.

Table 4 List of studied CSMIP instrumented buildings.

Index	**Station**		**PLFRS**	**Condition**	**# of EQs**	**# of sensors**
	Number	Name				
1	12267	Hemet – 4 story hospital	RCSW	U	10	10
2	58483	Oakland – 24 story residential building	RCSW	U	12	16
3	24579	Los Angeles – 9 story office building	RCMRF	U	4	18
4	24463	Los Angeles – 5 story warehouse	RCMRF	U	6	13
5	24322	Sherman Oaks – 13 story commercial building	RCMRF	U(R)	6	15
6	23634	San Bernardino – 5 story hospital	SMRF	U	5	12
7	57357	San Jose – 13 story government office building	SMRF	U(R)	3	22
8	24629	Los Angeles – 54 story office building	SMRF	U	7	20
9	3603	San Diego – 19 story commercial building	SEBF	U	3	16
10	58019	Stanford – 4 story residential building	WF	U	3	10
11	89494	Eureka – 5 story residential building	RM	U	7	13
12	58196	Berkeley- 5 Story parking structure	SCBF	U(R)	8	16
13	24386	Van Nuys- 7story hotel	RCMRF	D(R)	8	16
14	58354	Hayward - 13-story CSUH Admin Bld	RCMRF	Demo	2	16
15	01260	El Centro - Imperial Co Srvcs Bld	RCMRF	Demo	1	13

Among the selected fifteen buildings, nine of them are undamaged (U) and in operation at present, four are retrofitted (R) and currently in operation, and two have been demolished (Demo). One of the four retrofitted structures (station 24386) suffered severe damage during the 1994 Northridge earthquake while the other three were voluntarily retrofitted to seismically strengthen them. Out of the two demolished structures, one suffered from damage beyond repair (station 01260) and the other was found to be seismically unsafe (station 58354) by the California State University Seismic Review Board. The buildings have different occupancy types and PLFRS. Six of these structures have RC moment resisting frames (RCMRF), two have RC shear walls (RCSW), three have steel moment resisting frames (SMRF), two have steel braced frames (SBF), one has a wood frame (WF), and one has a reinforced masonry (RM) as the PLFRS. Each of these structures has at least 10 sensors installed at multiple floors.

H-MC framework for damage detection

The conventional post-earthquake damage assessment is a time consuming process. When relying on qualified inspectors alone, inspecting structures at the scale of a city can take weeks, if not months (Goulet et al. 2015). An alternative to using inspections alone is to use response data of the structure and ML tools to expedite the assessment process. However, a typical ML tool has some limitations when it comes to seismic damage detection due to the lack of data from damaged classes which may result in uncertainty in the detection. Human knowledge, domain expertise, and analytical skills can help to minimize these uncertainties. Therefore, an SHM framework called the human-machine collaboration (H-MC) is proposed herein which attempts to use advantages provided by ML in conjunction with the knowledge of humans. According to the National Research Council (2012), the H-MC is a framework in which humans co-work with artificial intelligence to complete specific tasks. The purpose of this framework is to use the particular strengths of both types of intelligence, and even physical capabilities, to fill in the weakness of one (e.g. the machine) by the intelligence of the other (e.g. the human).

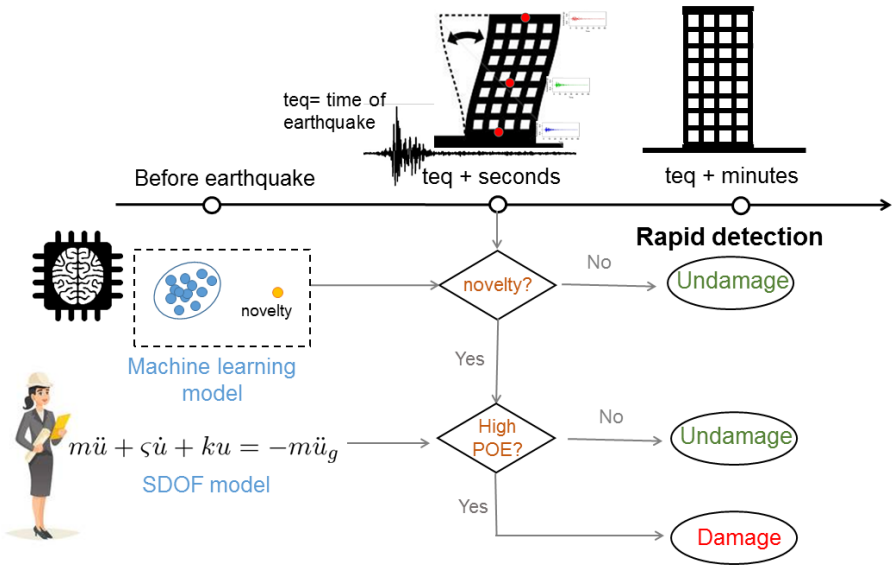

Figure 5 Human-machine collaboration (H-MC) framework for damage detection used in this study.

Figure 5 presents the H-MC framework for damage detection, which uses the responses from an undamaged structure and applies novelty detection as the ML tool for new data. Novelty detection is the identification of new or unknown data that a ML system is not aware of during training. It is similar to outlier detection, however, in the case of outlier detection the training dataset consists of outlier observations. The novelty model in this study develops non-parametric distribution using the training data and a distance measure of 1.5 times the interquartile range (IQR) to identify novelty, i.e. the new data is a novelty if it exists more than 1.5 interquartile range above the upper quartile or below the lower quartile. In the ideal case, where a large amount of data from the undamaged structure is available, considering data from low, moderate, and strong earthquakes, novelty detection alone could have indicated damage. However, for the buildings under considerations, data from strong but undamaging earthquakes are not available. Therefore, for these buildings, novelty detection may result in false positive detection for strong but undamaging events.

To overcome this limitation, the human aspect of this framework is introduced. A response envelope is developed by a domain expert representing the probability of exceedance (POE) of damage. This is performed by using a structure-specific SDOF model. NTHA is conducted using the 1,710 ground motions specified in set-1, and discussed above, having base shear coefficient recommended for the site of the building. Subsequently, the CAV and R_{CAV} of the damaging events are used to develop a joint cumulative distribution representing the POE of damage which is analogous to fragility curves but with two variables. Damage is identified when novelty is detected by the ML tool and the POE shows high probability values, as shown in Figure 5.

Results

Figures 6 to 8 show the damage detection results by the H-MC algorithm. In these figures, the colored contour plot is the POE envelope with higher probability shown with darker color. The dots are responses from all sensors installed at the roof level. When the dot is blue in color, it is detected as an undamaged event. When the dot consists of a red cross mark, it indicates damage has been identified. The damaged events will have a POE value greater than 0.5, i.e. they will be located in the darker region of the envelope. Figures 6 and 7 show that the undamaged buildings are correctly detected by the algorithm where all the responses are blue dots for these buildings. In some cases (such as station 58483, 24322, 58354, 57357, and 89494), novelty detection or POE envelope alone will result in false positive detection but when used together in the H-MC framework, the false positive results are successfully eliminated. Figure 8 shows the algorithm accurately detecting damage for station 24386 after the damaging event occurred. For station 01260, records are available from only the damaging earthquake. Therefore, the ML part of the framework is not possible to be applied for this building. However, by using the POE envelope alone, damage is detected for this case. In summary, the results show that the proposed H-MC framework uses two simple features, i.e. CAV and R_{CAV}, and able to obtain accurate detection although very limited dataset is used. This opens up the opportunity of rapid screening of existing structures after earthquakes with minimal computing time.

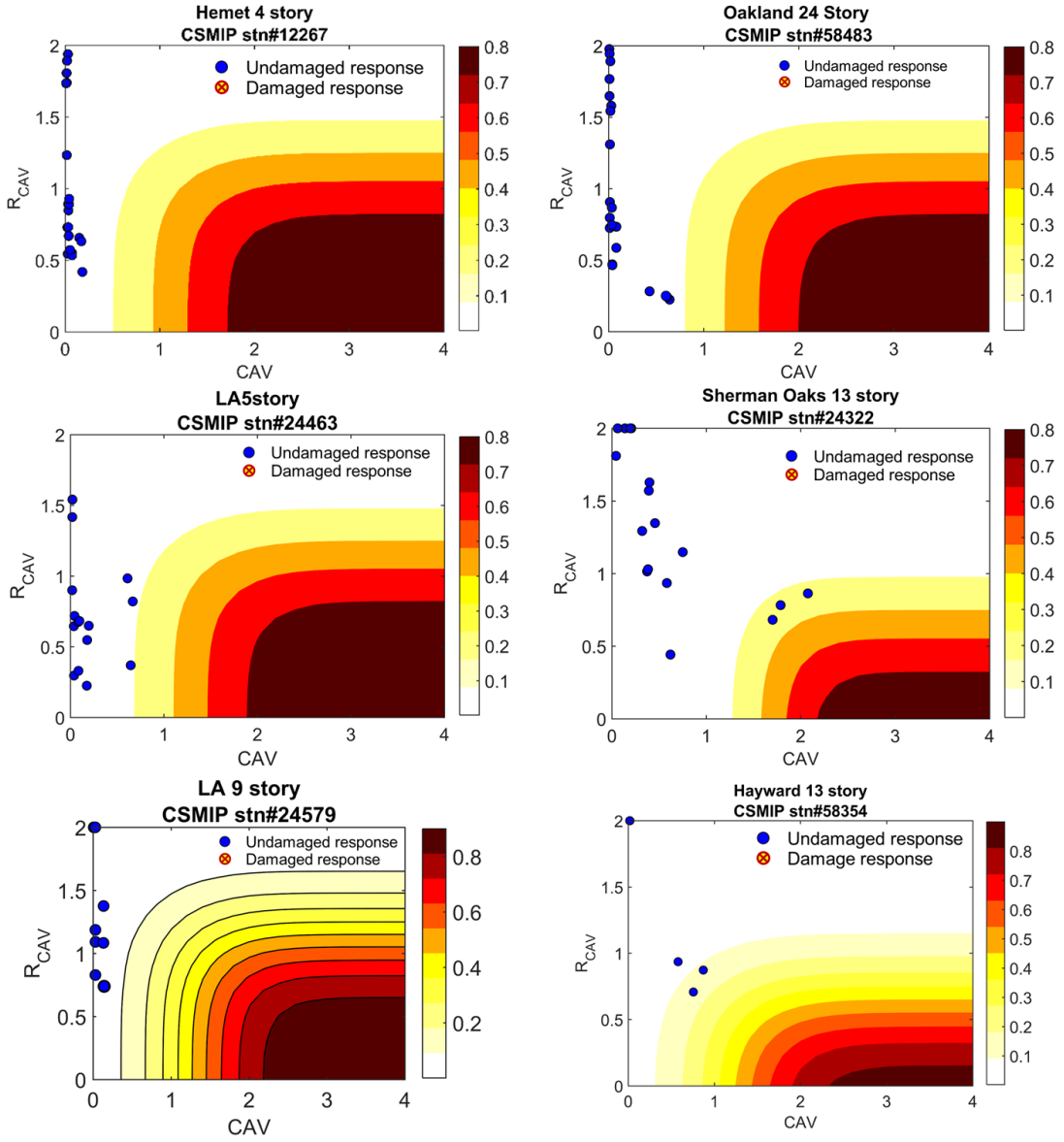

Figure 6 Plots generated by the damage detection algorithm showing accurate undamaged condition detection of the RC buildings.

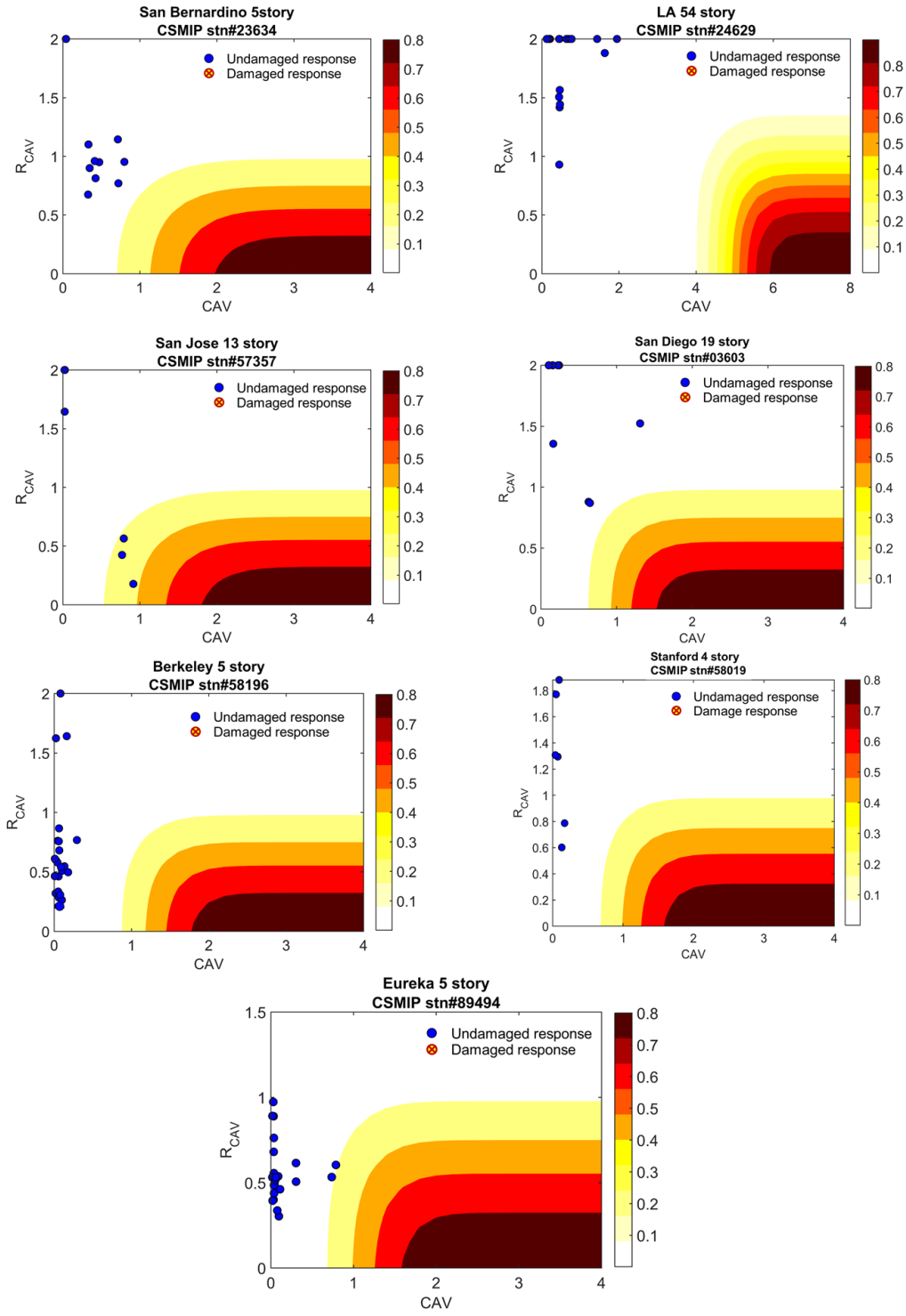

Figure 7 Plots generated by the damage detection algorithm showing accurate undamaged condition detection of steel, wood frame, and masonry buildings.

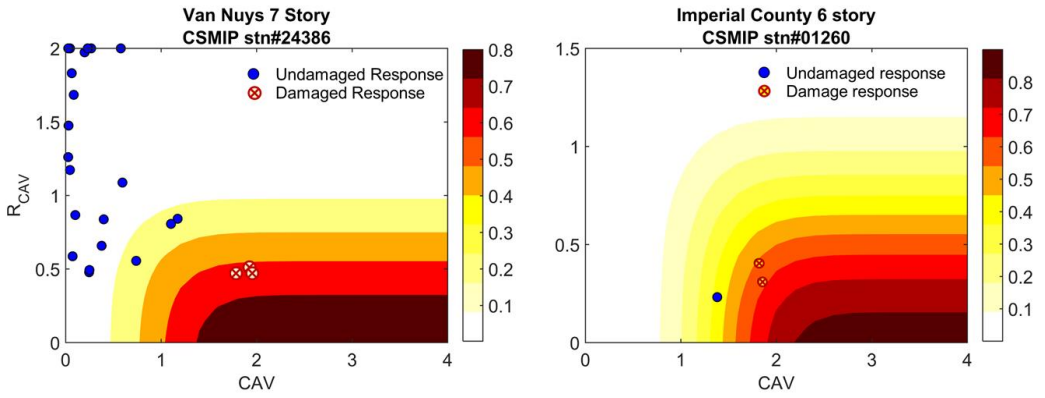

Figure 8 Plots generated by the damage detection algorithm showing accurate damaged condition detection of the two damaged buildings.

Conclusions

In this paper, a machine learning approach with single dimension features is presented to assess earthquake-induced damage in structures. Four cumulative absolute velocity (CAV)-based features are considered. A comparative study on a single degree of freedom (SDOF) system using these features and four separate machine learning tools reveals that CAV and R_{CAV} are the ideal features. These identified features are applied to assess the severity of damage of two multi-degree of freedom (MDOF) systems representing a five-story building with uniform (MDOF-US) and non-uniform (MDOF-NS) story shear capacity. Results show that this approach correctly detects the worst damage state with about 90.0% accuracy for both MDOF models when tested with NGA-West2 data. Moreover, this approach achieves damage state and location accuracies of 84.0% and 97.5%, respectively, for MDOF-US and damage state and location accuracies of 96.0% and 95.0%, respectively, for MDOF-NS for a test set significantly different from the training set.

Subsequently, the features are used in a human-machine collaboration (H-MC) framework to detect damage in selected fifteen California Strong Motion Instrumentation Program (CSMIP) instrumented buildings. The results showed that the H-MC algorithm correctly labeled the undamaged and damaged cases. This simple, low dimensional, and computationally efficient model can be reliably applied in structural health monitoring (SHM) scenarios with limited data. It opens up the opportunity to automate the process of damage detection and assess and notify the risk associated with each structure immediately after an earthquake. This is especially important for the residents of buildings in terms of having an idea about the condition of structures before conducting a formal tagging process which may take up to several weeks. Furthermore, such rapid damage assessment can be helpful for the immediate actions that need to be taken, such as deciding between leaving the building or staying in place and following the drop, cover and hold procedure. Overall, the model and corresponding results of the study can facilitate an efficient decision-making process regarding re-occupancy, emergency response, and future use of the structures following an earthquake event. This will, in turn, highlight the importance of building monitoring and encourage future building instrumentation efforts.

Acknowledgements

This research is mainly supported by the California Department of Conservation, California Geological Survey, Strong Motion Instrumentation Program agreement 1017-562. The authors would like to thank Dr. Selim Günay for his valuable inputs.

References

ASCE. 2017. “Infrastructure Report Card”. https://www.infrastructurereportcard.org/. Accessed 12 October 2017.

Baker, J.W., Lin, T., Shahi, S.K., Jayaram N. (2011) New ground motion selection procedures and selected motions for the PEER transportation research program. PEER Report 2011-3, Pacific Earthquake Engineering Research Center, UC Berkeley, CA.

Bommer, J. J. and Martinez-Pereira, A. (1999). The effective duration of earthquake strong motion. Journal of earthquake engineering, 3(02):127–172.

Bozorgnia, Y., Abrahamson, N.A., Atik, L.A., Ancheta, T.D., Atkinson, G.M., Baker, J.W., Baltay, A., Boore, D.M., Campbell, K.W., Chiou, B.S., Darragh, R., NGA-West2 research project. Earthquake Spectra, 2014, 30(3):973-987

Dua, R., Watkins, S. E., Wunsch, D. C., Chandrashekhara, K., and Akhavan, F. (2001). Detection and classification of impact-induced damage in composite plates using neural networks. In Proceedings, IJCNN’01. International Joint Conference on Neural Network. 1. 681–686.

Farrar, C. R. and Worden, K. (2012). Structural health monitoring: a machine learning perspective. Wiley, John & Sons.

Fathali, S. and Lizundia, B. (2011). Evaluation of current seismic design equations for nonstructural components in tall buildings using strong motion records. The Structural Design of Tall and Special Buildings, 20:30–46.

Günay, S. and Mosalam, K. M. (2017). Acceleration response spectrum: Revisited for higher mode effects.In 16th World Conference on Earthquake, Santiago Chile. January 9th to 13th 2017.

Goulet, J.A., Michel, C., and Kiureghian, A. D. (2015). Data-driven post-earthquake rapid structural safety assessment. Earthquake Engineering & Structural Dynamics, 44(4):549–562.

Hancock, J. and Bommer, J. J. (2006). A state-of-knowledge review of the influence of strong-motion duration on structural damage. Earthquake Spectra, 22(3):827–845.

Harris, A., Xiang, Y., Naeim, F., and Zareian, F. (2015). Identification and validation of natural periods and modal damping ratios for steel and reinforced concrete buildings in California. In SMIP15 Seminar on Utilization of Strong-Motion Data, pages 121–134.

IBC,(2012). International Code Council. International Building Code. International Code Council: Washington DC, United States.Kesavan, K. N. and Kiremidjian, A. S. (2012). A wavelet-based damage diagnosis algorithm using principal component analysis. Structural Control and Health Monitoring, 19(8):672–685.

Mahin, S., Lai, J. W., Wang, S., and Schoettler, M. (2015). Evaluating and improving the seismic performance of older tall buildings. In the second international conference on performance-based and life-cycle structural engineering. pages 25–37. School of Civil Engineering, The University of Queensland.

McKenna, F. (2010). OpenSees User's Manual.

Muin, S. and Mosalam, K. M. (2017). Cumulative absolute velocity as a local damage indicator of instrumented structures. Earthquake Spectra, 33(2):641–664.

Naeim, F., Hagie, S., Alimoradi, A., and Miranda, E. (2006). Automated post-earthquake damage assessment of instrumented buildings. In Advances in Earthquake Engineering for Urban Risk Reduction, pages 117–134. Springer.

National Research Council (NRC). 2012. Intelligent Human-Machine Collaboration: Summary of a Workshop. Edited by Ethan N. Chiang and Patricia S. Wrightson. Washington, DC: The National Academies Press. doi:10.17226/13479.

Palomino, L. V., Steffen, V., and Finzi, R. M. (2012). Fuzzy cluster analysis methods applied to impedance-based structural health monitoring for damage classification. In Topics in Modal Analysis II, Volume 6, pages 205–212. Springer.

Reed, J. W. and Kassawara, R. P. (1990). A criterion for determining exceedance of the operating basis earthquake. Nuclear Engineering Design, 123:387–396.

Santos, J., Calado, L., Orcesi, A., and Crémona, C. (2014). Adaptive detection of structural changes based on unsupervised learning and moving time-windows. In EWSHM-7th European Workshop on Structural Health Monitoring. Nantes, France.

Sohn, H., Worden, K., and Farrar, C. R. (2002). Statistical damage classification under changing environmental and operational conditions. Journal of Intelligent Material Systems and Structures, 13(9):561–574.

Tibaduiza, D., Mujica, L., Anaya, M., Rodellar, J., and Güemes, A. (2012). Independent component analysis for detecting damages on aircraft wing skeleton. In Proceedings of the 5th European Conference on Structural Control (EACS 2012), Genoa, Italy, pages 18–20.

Wald, D. J., Quitoriano, V., Heaton, T. H., and Kanamori, H. (1999). Relationships between peak ground acceleration, peak ground velocity, and modified Mercalli intensity in California. Earthquake Spectra, 15(3):557–564.

Worden, K., Manson, G., and Fieller, N. R. (2000). Damage detection using outlier analysis. Journal of Sound and Vibration, 229(3):647–667.