A Clustering Approach to Identification of Seismic Building Damage Patterns for Concrete Structures

Emily D. Elwood
Department of Civil, Environmental, and Architectural Engineering, Boulder, USA

Ross B. Corotis
Professor, Department of Civil, Environmental, and Architectural Engineering, Boulder, USA

ABSTRACT: Within the field of earthquake engineering, there is a desire to identify observed trends of building damage in response to seismic events. Regional loss models, post-earthquake safety evaluations, and rapid screening of structures for potential seismic hazards all utilize "patterns" of expected building behavior in some form. Due to the intrinsic uncertainties in the prediction of earthquake events and building response to these events, the identification of building damage patterns is a complex problem. Further, virtually every building is unique in configuration and characterized by varying site conditions. This research paper presents a unique application of fuzzy set theory within the domain of fuzzy classification (fuzzy clustering) to investigate whether seismic building damage patterns, often expressed linguistically, can be identified from empirical data. Information used in the classification analysis consists of building damage data from the 1994 Northridge Earthquake for concrete structures.

Many research issues related to the field of earthquake engineering are characterized by problems that are inherently uncertain. Sources of uncertainty range from the randomness that is frequently used to characterize earthquake intensity to limited information and knowledge about idealizations used in prediction and analysis tools (Wen et al. 2003). In spite of this uncertainty, various trends or patterns about the performance of buildings in response to seismic events are frequently listed in the literature. These trends often relate building characteristics that are believed to influence seismic behavior to expected levels of building damage. For example, buildings constructed to more recent designed guidelines are expected to have better seismic performance than older structures. It remains challenging, however, to characterize seismic building damage due to the uncertainty that surrounds earthquake engineering problems.

Traditionally, probabilistic and/or deterministic mathematical models have been used in widespread earthquake evaluation and damage prediction tools. Although both methodologies have given rise to significant advancements, they are limited in their applications. In regard to the former, the implications of a probabilistic framework for modeling uncertainty are many and diverse (see Reiter 1990, Wen et al. 2003). In the context of the latter, due to the variability and lack of precise knowledge of seismic building response, a deterministic approach to quantifying building damage patterns is difficult. Other uncertainty theories offer different philosophical approaches and mathematical formulations to modeling real-world phenomena. This research examines the use of fuzzy sets (or *linguistic variables*) for the purpose of identifying groups of buildings with similar damage characteristics. Central to this approach is recognizing the inherent uncertainty in verbal descriptors that are frequently used to characterize building damage.

A fuzzy clustering method introduced by Bezdek (1981) known as the *fuzzy c-means* (FCM) clustering method is presented. This is followed by implementation of this method for identifying building damage patterns in terms of important building features. Building damage data for concrete buildings from the 1994 Northridge earthquake are used. Finally, a discussion of the results is included and potential uses of identified patterns are proposed.

1. APPLICATIONS OF CLASSIFICATION FOR SEISMIC BUILDING DAMAGE ASSESSMENT

The use of mathematical classification techniques in the context of fuzzy set theory for applications of seismic damage assessment has been discussed since the theory first emerged within the earthquake engineering field. Naturally, researchers saw the link among physical, real-word, fuzzy abstractions of engineering parameters such as descriptions of damage states, levels of demand and building type classes, all of which are commonly used for estimating or quantifying seismic building damage. Initial research in this area includes Fu and Yao (1979), who investigated the use of pattern recognition in the context of building damage assessment. From their work, Ishizuka et al. (1980) reasoned that (i) the incorporation of experts' information is important and (ii) inference procedures that can handle uncertain and complex scenarios are necessary. For these reasons, much of their subsequent work was devoted to expert-based systems.

Watada et al. (1984) propose the use of fuzzy pattern recognition for classification of existing buildings into "safety" and "unsafety" based on visual inspection data and objective data (e.g., strong-motion data). Boissonnade and Shah (1985) and Dong (1986) illustrate how one can use classification and pattern recognition techniques for the purpose of systematically identifying regions of different seismic intensities (Ross 1995, Ross 2010). Boissonnade and Shah (1985) use the observed fraction of damaged buildings in varying damage states (*no damage*, *light damage*, *medium damage*, *severe damage* and *collapse*) within each region in their geographic-based approach.

More recently, Tesfamariam and Liu (2010) compare the applicability of different classical (crisp) classification techniques for the purpose of predicting different building performance levels. In this study, six building performance modifiers (or attributes) are considered in the classification: 1) number of stories above ground, 2) soft story index, 3) overhang ratio, 4) minimum normalized lateral stiffness index, 5) minimum normalized lateral strength index, and 6) normalized redundancy score. The classification schemes presented by Tesfamariam and Liu (2010) group observations into five damage states that are subsequently used to identify different levels of building performance. Two building performance levels are considered, *life-safety performance* and *immediate occupancy performance*. The study concludes that the classifiers have difficulty predicting *Damage* however, when the damage is grouped into more general criterion of *Life-Safety* or *Immediate Occupancy* classes, reasonably good results can be achieved (Tesfamariam and Liu 2010). Damage data for concrete buildings from the 1999 Duzce-Bolu region earthquake in Turkey is used in their study.

Despite early enthusiasm about the potential use of fuzzy classification techniques within the domain of earthquake damage assessment, these techniques have seldom been studied. Researchers have instead focused more on other fuzzy methodologies such as fuzzy systems and fuzzy decision making.

2. CLUSTER ANALYSIS - OVERVIEW

2.1. Why Classify?

One of the most fundamental capabilities of human intelligence is the ability to classify and recognize patterns (Dunn and Everitt 1982, Klir and Yuan 1995, Vick 2002). Without this ability, "experience would be nothing more than random accumulation of individual impressions" (Vick 2002, p. 97). If one accepts that human perceptions and the humanistic process of classifying and recognizing patterns are inherently fuzzy, the vision for the use of fuzzy models in the context of classification appears to be quite appropriate (Bezdek et al. 1999, Ross 2010). The connection between the areas of classification and fuzzy technologies has been recognized since fuzzy sets were first introduced (Zadeh 1977). Although the earliest reference to the use of fuzzy sets in numerical pattern recognition is credited to Bellman et al. (1966) (Bezdek et al. 1999, Ross 2010), it seems the inception of fuzzy models was harmonious with the motivations of classification. Suggested in Zadeh's (1965) seminal work "more often than not, the classes of objects in the real physical world do not have precisely defined criteria of membership" (Zadeh 1965, p. 338).

The objectives for finding classes are many (Kruse et al. 2007). Data groupings are informational, they are often easier to handle, and can aid in memory or prediction, or both (Dunn and Everitt 1982). Further, groups of similar observations have the ability to postulate a concise representation of the underlying data behavior (Chiu 1997, Kruse et al. 2007).

Although the literature offers an abundance of definitions for the terms, *classification* and *pattern recognition* (Bezdek et al. 1999), the general process aims to "search for structure in data" (Bezdek 1981, p. 1) and "classify these structures into categories" (Klir and Yuan 1995, p. 357). In finding structure, data are classified according to similar features, attributes, and other characteristics (Ross 2010). Such classification is the topic of *cluster analysis*.

2.2. Description of the Cluster Analysis Problem

Outlined by Bezdek et al. (1999), cluster analysis encompasses three problems: (1) tendency assessment, (2) clustering, and (3) validation. The first, *tendency assessment*, answers the question, should one look for clusters? The second, *clustering*, addresses the selection of the methods (algorithms) and models used for the cluster analysis. Finally, is the problem of *cluster validity*---the problem of identifying the appropriate number of clusters, or 'best' solution, that a clustering algorithm produces (Bezdek et al. 1999).

In the context of the contents presented in this paper, the *assessment* problem aims to investigate the appropriateness of clustering for the purpose of developing meaningful linguistic generalizations from individual seismic damage observations. However, Kruse et al. (2007) emphasize that in any case, one presupposes that identified groupings represent actual types or classes of objects. The classification analysis included herein is no exception. To identify patterns, the *clustering problem*, the fuzzy *c*-means method is selected. This clustering method is powerful for accommodating fuzzy information (Ross 2010) and therefore quite suitable for modeling building damage observations. A discussion of the resulting damage classifications and implications of the clustering results speaks to the *validity* problem.

3. FUZZY CLASSIFICATION

3.1. The c-Means Models

The *c-means* (commonly also referred to as *k-means*) models are the best known and most established clustering models (Bezdek et al. 1999). This family of clustering algorithms aims to classify a given set of data points into "homogeneous clusters" (Ross 2010, p. 362) such that the degree of similarity (or association) among data points is strong within the same cluster, and weak for data points in other clusters (Ruspini 1973, Klir and Yuan 1995). Cluster assignments for each data point are defined through geometric closeness (e.g., distance) in *n*-dimensional Euclidean space (Ross 2010). In this context, clusters partition the data and, as the name suggests, are also commonly referred to as *partitions*. A *fuzzy c-partition*, defines a family of fuzzy sets which are interpreted as the output fuzzy classes (or fuzzy clusters). Additional discussion about distance as a metric for similarity included in this paper is presented in a subsequent section.

The first crisp, or *hard c-means* (HCM), models date back to the 1950's (Bezdek et al. 1999). Years later, fuzzy sets and subsequent fuzzy clustering counterparts to the HCM models were formulated. Bezdek (1981) is credited with the development of a powerful fuzzy extension of the crisp hard *c*-means model---the fuzzy *c*-means (FCM) clustering method (Ross 2010). As mentioned previously, this is the clustering method used for the classification analysis presented in this paper.

3.2. The Fuzzy c-Means Clustering Method

To introduce the fuzzy *c*-means clustering method, let X represent a set of data (a set of *n* observations or *n* data samples) to be partitioned (Ross 2010):

$$X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n\} \quad (1)$$

Each observation, $\mathbf{x}_k$, is described by *m* features that can be expressed by an *m*-dimensional vector:

$$\mathbf{x}_k = \{x_{k1}, x_{k2}, x_{k3}, \ldots, x_{km}\} \quad (2)$$

Although fuzzy partitions share many of the same properties as crisp partitions, fuzzy classification differs in that a single data point can share membership with more than one cluster (Kruse et al. 2007, Ross 2010). In HCM, data points are assigned a membership value of unity to points assigned to that class (or cluster). This restriction guarantees that a single data point can belong to only one class. In FCM, data points are assigned a degree of membership (partial membership) in each fuzzy class (or fuzzy cluster). In this regard, fuzzy classes overlap. Partial membership is expressed in fuzzy set notation by (Ross 2010):

$$\mu_{ik}(x) = \mu_{Ç}(x_k) \in [0, 1] \quad (3)$$

where the membership value μ_{ik}, describes the degree of membership of the *k*th data point in the *i*th fuzzy class (Ç); {Ç, $i = 1, 2, \ldots, c$ where *c* denotes the number of clusters} (Ross 2010). Data points will generally have low membership values in the clusters that are remote ($\mu_{ik} << 1$), and high membership values ($\mu_{ik} \longrightarrow 1$) in clusters that are near (Kruse et al. 2007, Ross 2010). In this way, Ruspini (1973) explains that points between clusters can be mathematically classified as such. In the crisp case, the measure of isolation or separation of data points is not reflected by their mathematical membership (Ruspini 1973). Thus, fuzzy classification offers a much greater degree of detail of the class assignments (Kruse et al. 2007).

In FCM there is a mathematical restriction on the membership values assigned to each data point. It is required that the sum of membership values for a single data point over all of the classes equals unity, and the maximum membership assigned to any single data point be unity (Ross 2010). These requirements are described in Eq. (4) and Eq. (5), respectively.

$$\sum_{i=1}^{c} \mu_{ik} = 1, \text{ for all } k = 1, 2, \ldots, n \quad (4)$$

$$\bigvee_{i=1}^{c} = \mu_{Çi}(x_k) = 1 \quad (5)$$

Axiomatically, no class can be empty (i.e., one less cluster than the pre-specified number of clusters) and no class can contain all of the points (i.e., reject the hypothesis that clusters exist) (Bezdek et al. 1999, Ross 2010). This is expressed by:

$$0 < \sum_{k=1}^{n} \mu_{Çi}(x_k) < n, \text{ for all } i \quad (6)$$

The fuzzy *c*-means clustering method is defined by an algorithm that simultaneously minimizes the distance between points in any one cluster and maximizes the distance between clusters (Ross 2010). The within-cluster minimization (measured by the sum of the distances between data samples and cluster centers) is defined by a least squares objective function. The objective function, denoted here J_m, is given by (Klir and Yuan 1995, Ross 2010):

$$J_m(\underset{\sim}{U}, \mathbf{v}) = \sum_{k=1}^{n}\sum_{i=1}^{c}(\mu_{ik})^{m'}(d_{ik})^2 \quad (7)$$

where

$$d_{ik} = d(\mathbf{x_k} - \mathbf{v}_i) = \left[\sum_{j=1}^{m}(x_{kj} - v_{ij})^2\right]^{1/2} \quad (8)$$

and where $\underset{\sim}{U}$ is a *fuzzy c-partition matrix* ($c \times n$), $\mathbf{v}_i$ is a vector of cluster center coordinates, and d_{ik} measures the distance between the ith fuzzy cluster center ($\mathbf{v}_i$) and kth data point. Entries in the fuzzy c-partition matrix, μ_{ik}, describe the membership of the kth data point in the ith fuzzy cluster (or ith partition) of X (Bezdek et al. 1999).

The squared distance measure in Eq. (7), $(d_{ik})^2$, is weighted by the value $(\mu_{ik})^{m'}$. The *weighting exponent* (or *weighting parameter*) m' affects the amount of fuzziness in the membership assignments (Klir and Yuan 1995, Bezdek et al. 1999, Ross 2010). Larger values of m' ($m' \longrightarrow$ infinity) increase the fuzziness in the resulting classifications. Conversely, as m' decreases, the partitions become increasingly crisp and as $m' \longrightarrow 1$, the fuzzy c-partitions approach hard c-partitions (Bezdek et al. 1999, Ross 2010). Bezdek et al. (1999) report that the majority of the literature selects values in the range [1.1, 5] however, several authors agree that there is no theoretical 'best' selection for m' (Klir and Yuan 1995, Bezdek et al. 1999, Ross 2010). More narrowly, Pal and Bezdek (1995) report that m' in [1.5, 2.5] is commonly used by analysts.

Cluster center coordinates for the jth feature of the ith cluster, v_{ij}, are calculated by (Ross 2010):

$$v_{ij} = \frac{\sum_{k=1}^{n} (\mu_{ik})^{m'} \cdot x_{ki}}{\sum_{k=1}^{n} (\mu_{ik})^{m'}} \qquad (9)$$

for $j = 1, 2, \ldots, m;\ i = 1, 2, \ldots, n$. The resulting m-dimensional vector, $\mathbf{v}_i = \{v_{i1}, v_{i2}, v_{i3}, \ldots, v_{im}\}$ of coordinates for each cluster center are often referred to as *point-prototypes* or *prototypes*. As the name implies, these points are intended to serve as "prototypical representations" of the data from which they are derived (Hoppner et al. 1999, Bezdek et al. 1999, Kruse et al. 2007 p. 4).

As described previously in this section, the goal of the fuzzy c-means clustering method is to find a partition that minimizes J_m under the aforementioned simultaneous constraints. The smaller the value of $J_m(\underset{\sim}{U}, \mathbf{v})$ the 'better' the fuzzy partition (Klir and Yuan 1995). Thus, the 'best' solution is an optimization problem. To solve this, a popular alternating (or iterative) optimization procedure developed by Bezdek (1981), is used. In this procedure termination of the algorithm is defined by a pre-specified level of tolerance (i.e., level of acceptable accuracy) denoted by ε_L. Bezdek et al. (1999) state that in most practical applications, the optimization procedure can find reasonable results with ε_L = [0.01, 0.0001].

The step by step fuzzy c-means algorithm for solving the optimization is not reviewed here, however, it is provided in Klir and Yuan (1995), Bezdek et al. (1999), Ross (2010) as well as many other fuzzy and clustering text books.

The resulting optimum fuzzy c-partition can be expressed by:

$$J_m^*(\underset{\sim}{U}^*, \mathbf{v}^*) = \min_{\mathbf{M}fc} J_m(\underset{\sim}{U}, \mathbf{v}) \qquad (10)$$

where M_{fc} is a family of fuzzy c-partition matrices.

3.3. Hardening the Fuzzy c-Partition

Although fuzzy classification affords a much finer level of detail in the class assignments, hindsight suggests that an underlying motivation for many applications is to classify objects into actual (crisp) classes (Ross 2010). This is the process of *defuzzification*, that is, the process of reducing fuzzing information into a crisp, single-valued quantity (or class, set, etc.).

For fuzzy partitions, the most common method of defuzzification is the maximum membership method (Bezdek et al. 1999). This method assigns a membership value of unity to the largest entry in each column of the fuzzy partition matrix ($\underset{\sim}{U}$) and assigns a membership value of zero to all other entries in each column (Ross 2010). This is expressed by:

$$\begin{aligned} &\mu_{ik} = \max_{j \in c}\{\mu_k\},\ \text{then}\ \mu_{ik} = 1; \\ &\mu_{jk} = 0,\ \text{for all}\ j \neq i, \end{aligned} \qquad (11)$$

for $i = 2, \ldots, c$ and $k = 1, 2, \ldots, n$.

4. FUZZY CLASSIFICATION OF SEISMIC BUILDING DAMAGE DATA

4.1. Feature Selection

As part of any classification, the analyst must decide what features (or object criteria) to include (Dunn and Everitt 1982, Ross 2010). This is the subject of *feature selection*. For the analysis presented in this section, the criteria used for classification are drawn from the FEMA 154 methodology (FEMA 2002). This allows the analysis to systematically follow an established set of building features considered important for characterizing how a structure responds to seismically-induced loads.

General descriptions of the features (m = 5) considered in the classification are listed in Table 1. Building irregularities were not considered in the classification analysis (see Elwood (2014) for additional discussion).

Table 1. Building feature data.

Feature	Data (x_k)
Bldg. Height	Number of stories.
Bldg. Age	Year built.
Soil Type	Average shear wave velocity (Vs in feet/second).
Intensity	Horizontal peak ground acceleration (g).
Bldg. Damage	ATC-13 damage state (structural and nonstructural) (ATC 1985).[1]

[1] See Elwood (2014) for information about the reconciliation of ATC-38 (ATC 2001) reported structural and nonstructural damage.

4.2. Overview Building Damage Data

Information used for the fuzzy classification analysis consisted of building damage data reported in the *Database on the performance of structures near strong-motion recordings: 1994 Northridge, California, Earthquake, ATC-38 Report* (ATC 2001). Reported building locations were used to assign average shear wave velocity and peak ground acceleration values to each building record. Average shear wave velocity values were obtained from the United States Geological Survey (USGS) *Custom Vs Map Server* (USGS 2013) and horizontal peak ground acceleration values were taken from records at the station nearest to the reported location of each respective building (ATC 2001).

The analysis was limited to the subset of concrete structures (93 records). Buildings within this subset that were excluded from consideration include: three buildings for which information about building age was not reported and four additional buildings for which adequate damage information was not reported. All other building information outlined in Table 1 was provided for the remaining 86 records. One building was characterized by a significant difference in average shear wave velocity relative to all of the other building sites. Because the mathematical model used for the classification analysis in this paper is sensitive to outliers, this record was also excluded. Finally, since there was only one record that reported ATC-13 *Heavy* damage, this record was not included in the classification analysis. The subset of concrete buildings used in the classification analysis is referred to as ATC-38 Database CS (*Concrete Subset*) and denoted $X_{ATC\text{-}38,CS}$. Additional details about the damage data used for inclusion can be found in Elwood (2014).

4.3. Data Standardization

Since the feature data in Table 1 are measured in different units, data of this kind are typically transformed to a unified scale prior to classification (Dunn and Everitt 1982, Ross 2010). This is because it is generally not meaningful to calculate the Euclidean distance between raw data values from a geometric perspective. Although there are different methods for normalizing data, for purposes of the analysis presented in this application data are standardized by ranging in accordance with Eq. (12) (Dunn and Everitt 1982).

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{12}$$

Generally, ranging is one method that allows for more meaningful comparisons of the data in standardized space (Dunn and Everitt 1982).

The standardized data for building height and building age are illustrated in Figure 1. Data are colored according to reported ATC-13 damage state (ATC 1985): ATC-13 *None* (1) – green, ATC-13 *Slight* (2) – yellow, ATC-13 *Light* (3) – orange, and ATC-13 *Moderate* (4) – red.

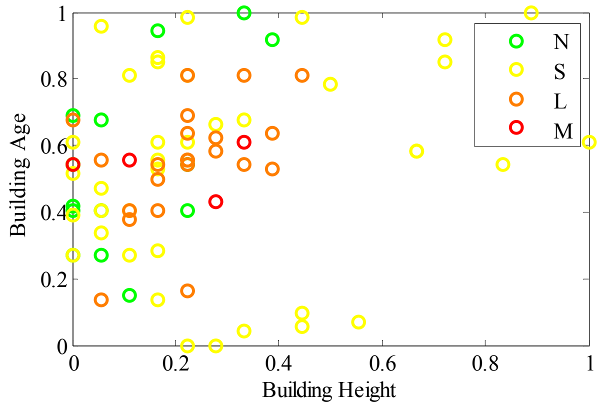

Figure 1. Standardized data in ATC-38 Database CS for building height and building age.

Figure 1 suggests that taller buildings were generally reported with ATC-13 *Slight* (2) damage. Other strong trends related to these particular building features are difficult to readily identify through visual observation of this two-dimensional plot. There appear to be few (if any) distinct linear trends between the building features. The classification analysis presented in the following section attempts to address this issue.

4.4. Discussion of Classification Results

For the results presented herein, Matlab Fuzzy Logic Toolbox functions were used to perform the fuzzy classification. The fuzzy *c*-means (FCM) clustering method was run with $m' = 1.75$ and $\varepsilon_L = 0.0001$. These values are within the respective ranges for each parameter reported previously in Section 3 and generally yielded good results. Of significance, the FCM assumes that the number of clusters (c) is pre-specified by the analyst. There are several proposed validity functions or validity indices that can be used to assess the appropriate number of clusters. In lieu of a mathematical approach to assess cluster validity (i.e., using a validity function for determination of the appropriate number of clusters), the number of clusters was selected through visually inspecting the simultaneous separation of reported damage class labels and the *terminal* partitions of the data (clusters at termination; when the iteration tolerance, $\varepsilon_L = 0.0001$, is satisfied). Running the FCM with seven clusters ($c = 7$) generally yielded reasonably good terminal separation of varying damage levels among the cluster groupings.

Terminal membership values were then hardened using Eq. (11). Hardened FCM partitions for building height and building age are shown in Figure 2. Cluster centers or class *prototypes* are denoted by crosses. Data are colored according to increasing terminal damage prototype values. The lowest terminal prototype value of damage is indicated in grey and the highest terminal prototype value of damage is indicated in red. Intermediate damage values are indicated by a color scale from green to orange for increasing levels of damage, respectively. This color scheme is strictly an ordinal level of representation.

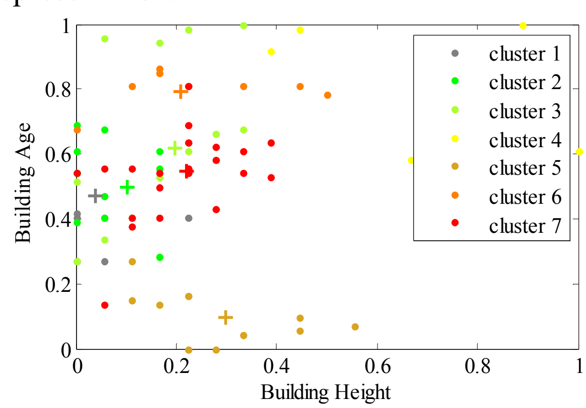

Figure 2. Hardened clustered data in ATC-38 Database CS for building height and building age.

The distribution of clustered buildings by damage labels (i.e., reported ATC-13 damage state) is shown in Figure 3. Again, data assignments have been hardened using Eq. (11).

It can be observed that buildings reported with no damage are successfully grouped into Cluster 1. There are five different clusters that represent buildings reported with ATC-13 *Slight* (2) damage, Clusters 2, 3, 4, 5, and 6. Thus, it appears that this damage state cannot efficiently be represented by one pattern, meaning there are multiple combinations of different building features that characterize this level of damage. Of these clusters, Cluster 6 (orange) has the most overlap with another damage state. This cluster contains relatively large fractions of records reported with ATC-13 *Slight* (2) and ATC-13 *Light* (3) damage. It appears that ATC-13 *Moderate* (4) damage is not well-represented by any of the clusters. Although all of the records reported with ATC-13 *Light* (3) damage are grouped (via hardening) into Cluster 7, this cluster also contains all of the records reported with ATC-13 *Moderate* (4) damage. This result is of little surprise considering that the data reported with ATC-13 *Moderate* (4) damage are sparse. It can be deduced that these four records simply do not have enough discriminatory power to generate a prototype that reflects the set of buildings reported as having sustained ATC-13 *Moderate* (4) damage. This also reveals a limitation of the methodology selected for this application in that the information content for records reported with ATC-13 *Moderate* (4) damage is limited by the data available for this class of buildings.

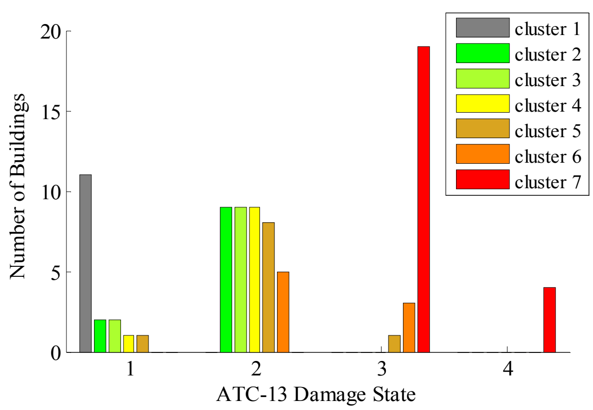

Figure 3. Distribution of clustered bldgs. by ATC-13 damage states (clusters have been hardened).

Terminal cluster center values produced from the FCM algorithm for $X_{ATC\text{-}38,CS}$ are listed in Table 2. The terminal center coordinates listed in this table correspond to each feature (m = 5): building height, building age, soil type, earthquake intensity, and degree of damage.

Once the data are grouped, linguistic descriptions can be assigned to each fuzzy cluster or fuzzy class based on the terminal cluster center values (see Table 2). Table 3 presents linguistic labels for the numerical values listed in Table 2. The verbal expressions in were generally drawn from linguistic labels contained in FEMA 154 (FEMA 2002) or FEMA 154 reference documents. Linguistic descriptions for perceived earthquake intensity were adapted from Dengler and Dewey 1998, 2003, reported in Wald et al. 2006. For some input features, intermediate linguistic assignments were added by the authors to provide for more refinement of the linguistic variables.

Table 2 $X_{ATC\text{-}38,CS}$ terminal cluster center values.

Cluster No.	No. Stories (v_{i1})	Year Built (v_{i2})	Vs (feet/sec) (v_{i3})	HPGA (g) (v_{i4})	ATC-13 DS (v_{iq})
Cluster 1	1.7	1955	1,052	0.44	1 - *None*
Cluster 2	2.9	1957	1,231	0.36	2 - *Slight*
Cluster 3	4.6	1966	1,207	0.47	2 - *Slight*
Cluster 4	14.1	1982	1,254	0.21	2 - *Slight*
Cluster 5	6.4	1927	1,247	0.19	2 - *Slight*
Cluster 6	4.8	1979	1,455	0.31	2 - *Slight*
Cluster 7	5.0	1961	1,240	0.27	3 - *Light*

The contents of Table 3 are referred to as *damage prototypes*. These prototypes serve to characterize the points (Hoppner et al. 1999, Bezdek et al. 1999, Kruse et al. 2007), or in this context, the buildings that share similar features to those described by the linguistic labels associated with each respective damage state. The labeling convention for each damage prototype is listed in the far left column. Observing the top row, the first damage pattern suggests that: a *Low-Rise* (L) building with a building age indicative of a *Moderate* (Mod) level of seismic design located at a site characterized by *Stiff Soils* (D) and subjected to *Severe* (SE) ground shaking resemble buildings that have been reported have sustained no damage (ATC-13 *None* (N)). Similar interpretations follow for the other prototypes.

4.5. Possible Uses for Damage Patterns

In addition to providing insight about linguistic patterns of building damage, damage prototype values could be used for analytical model calibration or validation. Damage prototypes identify a precise set of building features and associated observed damage level. In this way, the prototypes could help to investigate the correlation of observed building performance with the calculated response for a particular prototype. Using damage prototypes also has the benefit of answering the necessary and potentially challenging question of which building may be most appropriate for model calibration.

Additionally, the patterns identified herein can be used for the purpose of a pattern recognition system. In this context, new data samples (i.e., buildings) can be classified or assigned, either partially or completely, into established damage patterns (Ross 2010).

Possible extensions for this type of modeling include identifying patterns of losses that may be associated with various levels of observed damage, or developing patterns of post-earthquake building safety (see ATC 1995) based on building characteristics, both of which are often estimated and based primarily on subjective judgments (i.e., information that is nonrandom in nature).

5. CONCLUSIONS

The concepts presented herein hypothesize that clustering gives rise to general damage patterns. If one can understand these patterns, one can make observations about the behavior of the system. It is only from these observed patterns and understanding that finding ways to change the patterns is possible.

The systematic mapping of heuristic building damage patterns offers the capacity to "bridge the gap" between mathematical models, that often associate building performance to specified engineering parameters, and the real world, in which physical observations of damage are often expressed linguistically.

Table 3 Damage prototypes (damage patterns).

Prototype (Pattern)	Height	Age	Soil Type (Class)	Perceived Intensity
$N_{L,Mod,D,SE}$	Low (L)	Moderate (Mod)	Stiff Soil (D)	Severe (SE)
$NS_{L,Mod,C/D,SE}$	Low (L)	Moderate (Mod)	Dense Soil (C/D)	Severe (SE)
$S_{L/M,Mod,C/D,SE}$	Low/Mid (L/M)	Moderate (Mod)	Dense Soil (C/D)	Severe (SE)
$S_{H,Ben,C/D,VS/S}$	High (H)	Post-Bench. (Ben)	Dense Soil (C/D)	V. Strong/Strong (VS/S)
$S_{M,Pre,C/D,VS/S}$	Mid (M)	Pre-Code (Pre)	Dense Soil (C/D)	V. Strong/Strong (VS/S)
$S_{L/M,Ben,C,VS}$	Low/Mid (L/M)	Post-Bench. (Ben)	V. Dense Soil (C)	V. Strong (VS)
$L_{M,Mod,C/D,VS}$	Mid (M)	Moderate (Mod)	Dense Soil (C/D)	V. Strong (VS)

6. ACKNOWLEDGMENTS

The authors would like to thank Professor Keith Porter at the University of Colorado for research concepts and many useful conversations regarding this research. The authors also thank Charles Huyck of Imagecat Inc., for providing building inventory data.

The authors gratefully acknowledge the support of the National Science Foundation, grant number CMMI 1063790. The opinions expressed in this paper are those of the authors, and do not necessarily reflect the views or policies of NSF.

7. REFERENCES

Applied Technology Council (ATC) (1985). *Earthquake Damage Evaluation Data for California, ATC-13 Report*. Redwood City, California. ATC.

Applied Technology Council (ATC) (1995). Addendum to the ATC-20 Postearthquake Building Safety Evaluation Procedures, ATC-20-2 Report. Redwood City, California: ATC.

Applied Technology Council (ATC) (2001). *Database on the Performance of Structures Near Strong-Motion Recordings: 1994 Northridge, California Earthquake, ATC-38 Report*. Redwood City, California: ATC.

Bellman, R., Kalaba, R., and Zadeh, L. (1966). "Abstraction and Pattern Classification." *Journal of Mathematical Analysis and Applications* 13(1), 1-7.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. NY: Plenum.

Bezdek, J. C., Keller, J., Krisnapuram, R., and Pal, N. R. (1999). "Fuzzy Models and Algorithms for Pattern Recognition and Image Processing."

The Handbooks of Fuzzy Sets Series. Eds. D. Dubois and H. Prade. New York: Springer.

Boissonnade, A. C., and Shah, H. C. (1985). *Use of Pattern Recognition and Fuzzy Sets in Seismic Risk Analysis. Tech. no. 67*. Palo Alto: John A. Blume Earthquake Engineering Center.

Chiu, S. L. (1997). "Extracting Fuzzy Rules from Data for Function Approximation and Pattern Classification." *Fuzzy Information Engineering: A Guided Tour of Applications*. Eds. D. Dubois, H. Prade, and R. Yager. New York: John Wiley and Sons, 149-162.

Dong, W. (1986). *Applications of Fuzzy Set Theory in Structural and Earthquake Engineering*. Diss. Stanford Univ.

Dunn, G. and Everitt, B. S. (1982). *An Introduction to Mathematical Taxonomy*. Cambridge: Cambridge Univ. Press.

Elwood, E. D. (2014). *Fuzzy Classification and Fuzzy Pattern Recognition of Seismic Damage to Concrete Structures*. Diss. Univ. of Colorado Boulder.

Federal Emergency Management Agency (2002). *Rapid visual screening of buildings for potential seismic hazards: A handbook FEMA 154, Edition 2*. Washington D.C.: FEMA.

Fu, K. S. and Yao, J. T. P. (1979). "Pattern Recognition and Damage Assessment." *Proceedings from the third Engineering Mechanics Division Specialty Conference*, September 17-19, 1979, Univ. of Texas at Austin, Austin, Texas, 344-347.

Hoppner, F., Klawonn, F., Kruse, R., and Runkler (1999). *Fuzzy Cluster Analysis Methods for Classification, Data Analysis and Image Recognition*. New York: Wiley.

Ishizuka, M., Fu, K. S., and Yao, J. T. P. (1980). *Inference Method for Damage Assessment Systems of Existing Structures, Report CE-STR-80-17*. School of Civil Engineering, Purdue Univ., W. Lafayette, Indiana.

Klir, G. J. and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic Theory and Applications*. Upper Saddle River, NJ: Prentice Hall.

Kruse, R., Doring, C., and Lesot, M-J. (2007). "Fundamentals of Fuzzy Clustering." *Advances in Fuzzy Clustering and its Applications*. Eds. J. V. de Oliveira and W. Pedrycz. West Sussex, England: Wiley, 3-30.

Pal, N. R. and Bezdek J. C. (1995). "On Cluster Validity for the Fuzzy c-Means Model." *IEEE Transactions of Fuzzy Systems* 3(1), 370-379.

Reiter, L. (1990). *Earthquake Hazard Analysis: Issues and Insights*. Columbia Univ. Press.

Ross, T. J. (1995). *Fuzzy Logic with Engineering Applications*. West Sussex, England: Wiley.

Ross, T. J. (2010). *Fuzzy Logic with Engineering Applications, 3rd Edition*. West Sussex, England: Wiley.

Ruspini, E. H. (1973). "A Fast Method for Probabilistic and Fuzzy Cluster Analysis using Association Measures." *Proceedings from the Sixth Hawaii International Conference on System Sciences, January 1-11, 1973*. Ed. A. Lew. Honolulu, Hawaii, 56-58.

Tesfamariam, S. and Liu, Z. (2010). "Earthquake induced damage classification for reinforced concrete buildings." *Struct. Safety* 32(2), 154-164.

United States Geological Survey (USGS). *Custom Vs30 Mapping*. Web. 4 Oct. 2013. <http://earthquake.usgs.gov/hazards/apps/vs30/custom.php>.

Vick, S. G. (2002). *Degrees of Belief Subjective Probability and Engineering Judgment*. American Society of Civil Engineers (ASCE) Press, Reston, Virginia.

Wald, D. J, Worden, B. C., Quitoriano, V., and Pankow, K. L. (2006). *ShakeMap® Manual, Technical Manual, Users Guide and Software Guide. Version 1.0 619/06*. Golden, Colorado: United States Geological Survey (USGS).

Watada, J., Fu, K. S., Yao, J. T. P. (1984). *Damage Assessment using Fuzzy Multivariant Analysis. Report CE-STR-84-4*. School of Civil Engin., Purdue Univ., W. Lafayette, Indiana.

Wen, Y. K., Ellingwood, B. R., Veneziano, D. and Bracci, J. (2003). "Uncertainty Modeling in Earthquake Engineering," *MAE Center Project FD-2 Report,* February 12, 2003.

Zadeh, L. A. (1965). "Fuzzy Sets." *Information and Control* 8, 338-353.

Zadeh, L. A. (1975). "The Concept of a Linguistic Variable and its Application to Approximate Reasoning-I." *Information Sciences,*8,199-249.

Zadeh, L. A. (1977). "Fuzzy Sets and Their Application to Pattern Classification and Clustering Analysis." *Classification and Clustering*. Ed. J. Van Ryzin. New York: Academic Press, 251-282.