

1 Classifying Earthquake Damage to Buildings 2 Using Machine Learning

3 Sujith Mangalathu^{*1}, Han Sun¹, Chukwuebuka C. Nweke.¹, Zhengxiang Yi¹
4 and Henry V. Burton¹ M. EERI

5
6 The ability to rapidly assess the spatial distribution and severity of building damage is
7 essential to post-event emergency response and recovery. Visually identifying and
8 classifying individual building damage requires significant time and personnel resources
9 and can last for months after the event. This paper evaluates the feasibility of using
10 machine learning techniques such as discriminant analysis, K-nearest neighbors, decision
11 trees and random forests, to rapidly predict earthquake-induced building damage. Data
12 from the 2014 Napa earthquake is used for the study where building damage is classified
13 based on the assigned ATC-20 tag (red, yellow and green). Spectral acceleration at a period
14 of 0.3 s, fault distance and several building specific characteristics (e.g. age, floor area,
15 presence of plan irregularity) are used as features or predictor variables for the machine
16 learning models. A portion of the damage data from the Napa earthquake is used to obtain
17 the forecast model and the performance of each machine learning technique is evaluated
18 using the remaining (test) data. It is noted that the random forest algorithm can accurately
19 predict the assigned tags for 66% of the buildings in the test dataset.

20 INTRODUCTION

21 Following a major earthquake, the ability to rapidly assess the spatial distribution and
22 severity of damage to the built environment is critical to the success of the emergency response
23 and repair/reconstruction phases of recovery (Ranf et al. 2007; Earl et al. 2009; Mangalathu,
24 2017). This situational awareness is an important part of the decision-making process for facility
25 owners, users, emergency responders and local and state officials. A general lack of knowledge
26 about the functional state of impacted infrastructure systems can lead to a disorganized societal
27 response (Ranf et al. 2007; Wald et al. 2008).

28 Buildings provide support for a range of activities and are central to the economic and
29 social development of a community. Timely awareness of the spatial impacts to building clusters
30 following a seismic event is key to the overall response because of the implications to the
31 functionality of all sectors of the economy (e.g. housing, education, businesses, healthcare). A
32 holistic view of the extent and distribution of damage can be obtained from the post-earthquake
33 inspection process, whereby teams of volunteer engineers are assembled and tasked with
34 inspecting and “tagging” each building. Buildings that are apparently safe to reoccupy are
35 assigned a green tag and those that are obviously unsafe to occupy are assigned a red tag. In
36 cases where the extent to which the potential danger of reoccupancy is less obvious, the
37 buildings are assigned a yellow tag (ATC, 1995). Depending on the scale of earthquake damage,
38 this can be a lengthy, resource intensive process. For example, it took more than 2 months to
39 assess the damage and tags to the more than 100,000 buildings impacted by the 1994 Northridge

¹ School of Civil and Environmental Engineering, University of California, Los Angeles, CA 90095, USA

* Corresponding author, E mail: sujithmangalath@ucla.edu

40 earthquake (Trifunac et al. 1994). While invaluable detailed knowledge about the seismic
41 performance of the affected building stock is gained from this type of effort, the overall pace of
42 progression is much too slow to support emergency response and early recovery planning.

43 Tools such as ShakeMap (Wald et al. 1999; Wald et al. 2005) and ShakeCast (Wald et al.
44 2008), which have been developed by the United States Geological Survey (USGS), are much
45 more feasible for providing timely information related to the impact of an earthquake on the built
46 environment. The former is used to generate real-time estimates of the extent and intensity of
47 ground shaking caused by an earthquake. ShakeMap combines knowledge from ground motion
48 recordings with information about the earthquake (e.g. magnitude, location and fault type) and
49 geology of the affected area to automatically generate shaking intensity maps. Within minutes of
50 the event, products such as peak ground acceleration and velocity maps, are made available to
51 the public through an interactive online interface. ShakeCast, which was developed as a
52 complement to ShakeMap, is a tool that allows stakeholders to obtain real-time estimates of the
53 damage to spatially distributed infrastructure. Through an online portal, ShakeCast users can
54 map facility locations, categorize structures based on their vulnerability and establish ground
55 motion intensity-based triggers for different states of damage. As an alternative to the user-
56 defined motion-damage relationships, the estimated shaking intensities can be coupled with the
57 pre-established damage fragilities that are available in loss modeling platforms such as HAZUS
58 (HAZUS-MH 2003) and OpenQuake (Silva et al. 2014). There is documented evidence of the
59 utility of opensource platforms that provide real-time estimates of earthquake impacts. For
60 example, Ranf et al. (2007) posited that the post-earthquake response and inspection of the 2001,
61 Nisqually, Washington (M6.7) earthquake, could have benefitted from the ground motion
62 intensities generated by ShakeMap. Wald et al. (2008) noted that the California Department of
63 Transportation has been instrumental in the development of the ShakeCast system.

64 Portfolio-scale predictions of earthquake-induced building impacts have historically
65 relied on the use of fragility functions, which probabilistically link ground shaking intensities to
66 degrees or categories of physical damage. As noted earlier, these damage fragility functions can
67 be implemented into ShakeCast to obtain early post-earthquake estimates of the spatial
68 distribution of damage. Realizations of damage on an individual building scale can then be
69 obtained using Monte Carlo simulation. The fragility relationships in HAZUS and OpenQuake
70 are categorized based on factors affecting earthquake vulnerability such as a building's age and
71 construction type. With the current advancements in information technology, there has been an
72 increase in the availability and access to building-specific data that are correlated with the
73 likelihood of seismic damage to buildings. These developments have made it possible to explore
74 the application of artificial intelligence (AI) to predict earthquake damage to buildings at the
75 portfolio scale. This approach can be used as an alternative or supplement to assessments based
76 on fragility functions. While prior researchers have investigated the use of AI for infrastructure
77 risk assessment (e.g. Nateghi et al. 2011), we know of no published works related to predicting
78 portfolio-scale building damage following an earthquake.

79 This paper compares the accuracy of different machine learning techniques for predicting
80 portfolio-scale building damage caused by earthquake shaking. Damage is described using the
81 tagging categories defined in the ATC 20 (ATC, 1995) document. Several predictor variables are
82 used as input parameters for the machine learning algorithms including the spectral acceleration
83 corresponding to a period of 0.3 s ($Sa_{0.3s}$) at the site of interest and the closest distance to the
84 surface projection of fault rupture (R_{jb}). Building characteristics such as the age (in years),

85 number of stories (NS), the presence of plan or stiffness irregularities (IR), floor area (BS) and
86 replacement cost (BV), are also used as predictor variables. Four different machine learning
87 algorithms are evaluated including discriminant analysis, K-nearest neighbors, decision trees and
88 random forests. The models are developed using the damage data from the 2014 South Napa
89 earthquake. The dataset consists of 2,276 buildings that were inspected and tagged by volunteer
90 engineers during the months following the event. Note that the ATC-20 tag is one of several
91 proxies that can be used to describe building damage in terms of discrete states. Similar but more
92 detailed damage proxies are used in HAZUS and OpenQuake. The characteristics and impacts
93 of the 2014 South Napa, California earthquake are summarized in the next section, which is
94 followed by a description of the dataset and preliminary trend analysis. An overview of the
95 considered machine learning techniques and their ability to predict earthquake-induced building
96 damage (as classified by ATC-20 tags) is then presented followed by the conclusions drawn from
97 the study.

98 **SUMMARY OF 2014 SOUTH NAPA EARTHQUAKE AND IMPACTS TO THE BUILT 99 ENVIRONMENT**

100 On August 24, 2014, an Mw 6.0 earthquake occurred in South Napa, causing damage to
101 buildings and lifeline systems in the cities of Napa, American Canyon and Vallejo. The epicenter
102 of the earthquake was located approximately 8 km south-southwest of Napa at N 38.22
103 W122.12. The event produced peak ground accelerations (PGAs) as high as 0.61g (Brocher et al.
104 2014) and the shaking was felt as far south as Salinas (USGS, 2014). While only one fatality
105 (Yune, 2014) and three hundred injuries have been attributed to the earthquake, more than 2,000
106 structures suffered damage that could be characterized as moderate to severe. Most of this
107 damage occurred in residential buildings. Estimates of the economic losses due to the earthquake
108 range from \$500 million to \$1 billion (Johnson and Mahin 2016).

109 Following the earthquake, the city of Napa assembled a group of volunteers to inspect
110 and tag the damaged buildings. The tagging was based on the ATC-20 (ATC 1995) guidelines
111 for assessing the post-earthquake safety of buildings, which is widely used in the United Sates
112 and other parts of the world. A total of 3, 677 buildings were inspected and the resulting dataset
113 was made available on the Earthquake Engineering Research Institute (EERI) clearinghouse
114 website (<http://eqclearinghouse.org/map/2014-08-24-south-napa/>). Of the 3,677 inspected
115 buildings, 1,729 were assigned green tags, 1,749 were yellow tagged and 199 buildings were
116 assigned red tags. Some buildings received multiple tags based on more than one inspection and
117 some undamaged buildings were deemed unsafe because of damage to adjacent structures
118 (Boatwright et al. 2015).

119 Most of the damage to woodframe residential buildings occurred because of unbraced
120 chimneys and unbraced, unbolted cripple wall foundations. Damage to chimneys included minor
121 cracking, spalling of bricks and complete collapse. Several houses suffered partial or complete
122 collapse of the cripple wall foundation. Other types of reported damage included breaks and
123 leaks in gas lines and collapsed car ports (Johnson and Mahin 2016). In commercial buildings,
124 both structural and non-structural damage have been documented with the former occurring in
125 unreinforced masonry and non-ductile concrete frames structures. The documented non-
126 structural and content damage included broken storefront glazing, damage to facades and breaks
127 in sprinkler pipes. Non-structural damage in wineries included broken wine bottles, collapsed
128 wine-barrel stacks and buckled tank walls (Fischer et al. 2016).

129 Damage to lifelines was minimal and did not have a significant and lasting impact. The
130 systems affected included bridges, roadways and water and power networks (Johnson and Mahin
131 2016). A wine spill into major sewers caused a 3-day disruption of water treatment operations
132 and power outages caused by damage to various electrical components lasted less than 24 hours.

133 DATA DESCRIPTION

134 BUILDINGS, DAMAGE CLASSIFICATION AND SPATIAL DISTRIBUTION

135 A building damage dataset for the 2014 South Napa earthquake was obtained from the
136 EERI clearinghouse website. As noted earlier, 3,677 buildings are included in this dataset, which
137 also includes several types of building-specific information that are relevant to this study
138 including the location (latitude, longitude and address), occupancy type (residential and
139 commercial), inspection date, ATC-20 placard (green, yellow and red) and a short description of
140 the damage caused by the earthquake. Only a subset of the EERI dataset, 2,276 residential
141 buildings, is used for the current study. Of the 2,276 buildings, 41% were assigned green tags,
142 55% were yellow tagged and 4% received red tags. The remaining buildings are excluded
143 because of missing or low-quality information (e.g. missing tag assignment, geographical
144 coordinates, year of construction). Among the red- and yellow-tagged buildings, approximately
145 62% of the damage occurred in chimneys and 10% in walls (superstructure and foundation).
146 Fireplace damage, stucco cracking and front porch collapses are also included in the residential
147 buildings. Typical descriptions of the more prominent types of damage associated with each of
148 the three tagging categories is presented in Table 1.

149 A map of the city of Napa with the location of the 2,276 damaged buildings considered in this
150 study and the epicenter of the earthquake is shown in Fig. 1. The ATC-20 inspection tags are
151 identified using color-coded markers (green, yellow and red). Fig. 1 shows that the building
152 damage was spread throughout the city, extending as far west as the Browns Valley District to
153 the Shurtleff neighborhood in the east. Moreover, despite the concentration of damage in the
154 southern and central region (e.g. Downtown, Fuller Park), buildings in the Springwood Estates
155 neighborhood, which is located approximately 8 kilometers north of the central business district,
156 were also damaged.

157 SELECTION OF PREDICTOR VARIABLES

158 The predictor variables used to train the machine learning algorithms are selected based
159 on knowledge of their relationship to the seismic vulnerability of buildings and correlation with
160 earthquake damage. The ease of access to the relevant information (e.g. can be obtained from
161 online resources with little or no need for field investigation) and ability to directly or indirectly
162 capture the spatiotemporal variation in construction practices and site information are also
163 considered in the predictor variable selection. Since most (approximately 80%) of the buildings
164 in the assembled dataset have three stories or less, $Sa_{0.3s}$ is used as the shaking-intensity-related
165 predictor variable. The $Sa_{0.3s}$ corresponding to each building site is obtained by interpolating the
166 values computed using the recorded ground shaking intensities. The details of the interpolation
167 procedure are described in the next section. The closest distance to surface projection of the fault
168 rupture (R_{jb}) and time-averaged shear-wave velocity to 30 m depth (V_{S30}), which is commonly
169 used in ground motion prediction equations, are also considered as predictor variables. It is worth
170 noting that some predictor variables may be perceived as being redundant. For example, it can be
171 argued that R_{jb} and V_{S30} are already reflected in $Sa_{0.3s}$. However, it is not obvious that these two

172 predictor variables are “fully” captured by $Sa_{0.3s}$. The redundancy of the chosen predictor
173 variables will be investigated later in the paper.

174 It is well-recognized that the likelihood of earthquake-induced damage has a strong
175 association with the number of stories and age of a building (e.g. both are incorporated either
176 directly or indirectly in damage fragility parameters). Therefore, these two building properties
177 are included as predictor variables. Building age is used in lieu of “code era” to consider the
178 effect of construction practices (e.g. two buildings may have the same code era but were
179 constructed based on different practices) and avoid discrepancies in permitting and construction
180 dates (e.g. construction permit obtained in non-ductile code era and actual construction started in
181 ductile code era). The number of stories were obtained from Google Maps and Google Street
182 View, and the age of each building was obtained from the real-estate website Zillow
183 (<https://www.zillow.com/>). The presence of a plan or vertical irregularity is known to increase
184 the seismic vulnerability of a building (ASCE, 2016) and is therefore also included as a predictor
185 variable. A building is considered irregular if there is a significant geometric variation in the
186 elevation or plan (e.g. plan projection that is more than 20% of the maximum plan dimension).
187 The irregularity information was also obtained from Google Street View. Because of its
188 correlation with overall size and seismic weight, the building floor area is also included as a
189 predictor variable. Building value (in dollars) is also considered as a predictor variable based on
190 the recognition that it may reflect some “hidden” regional information. For instance, similarly
191 valued buildings may be clustered in the same area and therefore constructed using similar
192 practices. Building floor areas and values were obtained from the real-estate website Zillow
193 (<https://www.zillow.com/>). As noted earlier, the effect of redundancies arising from the selection
194 of potentially correlated predictor variables (e.g. building size and value) is investigated later in
195 the paper.

196 A summary of the distribution of predictor variables and classifications (regular versus
197 irregular plan) as well as the tag assignments is shown in Fig. 2. The recorded and interpolated
198 $Sa_{0.3s}$ range from 0.55g to 1.33g. The maximum and minimum R_{jb} are 0.35 km and 9.15 km,
199 respectively. All buildings have a R_{jb} that is less than 10 km and 41% have a R_{jb} that is less than
200 5km. The age of buildings included in the dataset range from 9 years to 157 years.
201 Approximately 15% of the buildings are over a century old and only 1% were constructed in the
202 last 20 years. The number of stories range from one to six stories, however, more than 80 % of
203 the buildings have between one and three stories. 1078 buildings (slightly less than 50%) have
204 been classified as having either vertical or plan irregularities. More than 95% of the buildings
205 have a floor area of 5,000 ft² or less, which is consistent with most buildings having single-
206 family occupancy.

207 SPATIAL INTERPOLATION OF SPECTRAL INTENSITIES

208 A total of 381 pairs of horizontal ground motion component recordings from the 2014
209 South Napa earthquake were obtained from the Center for Engineering Strong Motion Data
210 (CESMD) (www.strongmotioncenter.org). The spatial distribution of the recordings is shown in
211 Fig. 3, most of which are from locations outside of the city of Napa. The recording stations have
212 R_{jb} values ranging from approximately 4 km to 130 km. The median R_{jb} is 47.5 km and the 14th
213 and 86th percentile values are 32.1km and 87.9 km, respectively. For each observation, the $Sa_{0.3s}$
214 values are computed and the maximum of the two components is used as the variable. Note that
215 the ground motion directionality is not considered in this study. The maximum $Sa_{0.3s}$ values
216 based on the recorded ground motions is approximately 1.322g and the median is 0.037g. The

217 low median value is consistent with the fact that approximately half of the recordings are from
218 stations with R_{jb} values greater than 50 km. The station corresponding to the maximum $Sa_{0.3s}$ has
219 an R_{jb} of approximately 14.07 km. It is worth noting that recordings from several other stations
220 with smaller fault distances had lower corresponding $Sa_{0.3s}$ values (e.g., Green Valley Road
221 station with $R_{jb} = 9.19$ km produced an $Sa_{0.3s}$ of 0.249g). This confirms that the $Sa_{0.3s}$ is not only
222 dependent on R_{jb} but also the site parameters such as basin depth and V_{S30} .

223 The kriging algorithm is used to interpolate the $Sa_{0.3s}$ values from the recorded sites to the
224 locations of the buildings considered in the current study. Originally developed as a geostatistical
225 technique, kriging has been used in prior studies to interpolate ground motion intensities and
226 engineering demand parameters (e.g. Kwak et al. 2016; Sun et al. 2018). First, the within-event
227 residual, which represents the difference between the predicted and recorded $Sa_{0.3s}$ values are
228 computed. The Boore et al. (2014) ground motion prediction equation (GMPE) model is used to
229 obtain the predicted $Sa_{0.3s}$ values at the sites corresponding to the recorded motions. Within-
230 event residuals are computed as the difference between the log-mean (from the GMPE) and
231 recorded $Sa_{0.3s}$ values. Semivariograms are used to describe the empirical spatial dependence
232 between the pairs of within-event residuals corresponding to all stations, which is then fitted to a
233 continuous exponential variogram model. Using this continuous variogram model, the kriging
234 algorithm is implemented to reconstruct within-event residual values of $Sa_{0.3s}$ at the building
235 sites of interest. Finally, the Boore et al. GMPE is used to compute the log-mean $Sa_{0.3s}$ values at
236 the sites of interest, which are added to the interpolated within-event residual to obtain the total
237 value.

238 To ensure adequate performance of the kriging model, the 381 $Sa_{0.3s}$ values associated
239 with the recorded ground motions are randomly separated into training and testing datasets that
240 comprise of 90% and 10% of the complete dataset, respectively. The interpolation model is
241 formulated using the training dataset and the testing dataset is used for evaluation. The mean
242 absolute relative deviation (MARD) is used as the performance measure, which is computed as
243 the mean of the absolute difference between the actual and predicted $Sa_{0.3s}$ within-event residual
244 values normalized by the actual value (Sun et al. 2018). This process is repeated by performing
245 non-replacement Bootstrap so a distribution of the MARD is obtained. More details on the
246 kriging procedure can be found in Kwak et al. (2016). The median MARD obtained from 20
247 repeated non-replacement Bootstrap is 34% and the 25th and 75th percentile MARD values are
248 28% and 35%, respectively. The spatial distribution of interpolated $Sa_{0.3s}$ values at the building
249 sites of interest is shown in Fig. 4 where the median and maximum values are 0.96g and 1.24g,
250 respectively.

251 PRELIMINARY DATA ANALYSIS

252 Scatter plots showing the assigned building tags as a function of pairs of predictor
253 variables are presented in Fig. 5 to reveal any obvious trends. Fig. 5a shows the clearest trend
254 between the level of damage (as reflected in the assigned tags) and the $Sa_{0.3s}$ and age of the
255 building where most of the red tags are clustered at sites with higher $Sa_{0.3s}$ levels and in older
256 buildings. The strong correlation of age and ground motion intensity with the level of building
257 damage has been reported in other studies (e.g. Boatwright et al. 2015). It is interesting that Fig.
258 5b does not show a strong trend between building damage and fault distance or between $Sa_{0.3s}$
259 and fault distance. One possible reason is the small R_{jb} range that is embedded in the dataset
260 (0.35 km to 9.15 km). As such, the differences in the level of shaking at the various sites is

261 attributed to site- and path-related factors such as basin and directivity effects (Baltay and
262 Boatwright 2015). Figs 5c to 5i show no specific pattern in the assigned tag as a function of the
263 building size and building value.

264 MACHINE LEARNING TECHNIQUES

265 To evaluate the efficacy of machine learning techniques in predicting assigned building tags,
266 the entire data set is divided into training and testing subsets. The training subset is used to
267 construct the predictive model, and testing subset is used to evaluate the performance of those
268 models. In this study, 70% of the data is used as the training set and 30% is used as the testing set
269 following the recommendation of Friedman et al. (2018). The division of data into testing and
270 training sets is assigned randomly. The evaluation of the model using the testing set helps inform
271 its ability to make predictions using unknown data and prevents overfitting (Friedman et al.
272 2001; Mangalathu et al. 2018; Mangalathu and Jeon 2018). The machine learning techniques
273 considered in this study include regression trees, linear discriminant analysis, k-nearest
274 neighbors and Random forests.

275 For all four methods, the predictor variables are denoted as $X = (X_1, X_2, \dots, X_8)$, where
276 $X_1 = S_{a0.3s}$, $X_2 = R_{jb}$, $X_3 = Age$, $X_4 = NS$, $X_5 = V_{s30}$, $X_6 = BR$, $X_7 = BV$, $X_8 = BS$, whose values are
277 placed in an $N \times p$, where $N = 2,276$ is the number of datapoints or damaged buildings and
278 $p = 8$ is the number of predictors. Y is an $N \times 1$ array of the categorical outputs (green, yellow
279 or red tag) from the prediction models. The methods adopted in the current study were chosen
280 because they have different inherent assumptions, which are explained in detail in the next
281 section. One of these assumptions is related to the decision boundaries (the hypersurface used to
282 separate the observations into the various categorical outputs, in this study, red, yellow and green
283 tags) that are adopted by the different algorithms. Discriminant analysis is parametric (i.e. the
284 form or shape of decision boundaries between the tags is assumed) and K-nearest neighbors,
285 decision trees, and random forest are non-parametric i.e. the form or shape of the decision
286 boundary is chosen to maximize the model performance. Fig. 6 shows a schematic representation
287 of the decision boundaries for the case where only two predictors (depicted as X_1 and X_2 for
288 illustration purposes) are considered. Fig. 6a shows an example of linear decision boundaries,
289 which is assumed by the linear discriminant analysis algorithm. Whereas, the non-parametric
290 methods can consider both linear and nonlinear (Fig. 6b) decision boundaries. Note that the non-
291 linear decision surfaces are different for the different non-parametric algorithms considered in
292 this study.

293 DISCRIMINANT ANALYSIS

294 Discriminant analysis is a machine learning algorithm used to classify two or more response
295 variables. The discriminant analysis can be of two types (1) linear discriminant analysis (LDA),
296 and (2) quadratic discriminant analysis (QDA). The type depends on the inherent assumptions in
297 the algorithm. In discriminant analysis, the mean and variance of X are calculated for each tag
298 assignment, and is assumed to be drawn from a multivariate Gaussian distribution with a class-
299 specific mean vector and covariance matrix, i.e., $X \sim N(\mu, \Sigma)$ where μ is the mean vector of X
300 (size $1 \times p$), and Σ is the $p \times p$ covariance matrix of X . The multivariate Gaussian density is
301 defined as

302
$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right) \quad (1)$$

303 By applying Bayes theorem to the density function, (Eq. 2), the probability that an observation x
 304 belongs to a class k (denoted by δ_k), can be estimated using Eq. 3.

305
$$P(Y = k | X = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^k \pi_l f_l(x)}; k \geq 2 \quad (2)$$

306

307
$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log(\pi_k) \quad (3)$$

308 where π_k represents the prior probability that a randomly chosen observation comes from the k th
 309 class. LDA assigns the variable to the class k that produces the largest δ_k . Eq. 3 shows that δ_k is a
 310 linear function of x , which stems from the assumed linear decision boundaries between the
 311 classes. Unlike LDA, QDA assumes that each class has its own covariance matrix, and hence the
 312 decision boundary is quadratic. In other words, QDA estimates a quadratic decision boundary
 313 between the green, yellow and red tags.

314 **K-NEAREST NEIGHBORS**

315 K-nearest neighbor (KNN) is a non-parametric method for classifying data into groups. For
 316 a positive integer K , and an observation x , KNN first identifies K points in the training data that
 317 are closest to x , which is represented by N_k . Hence KNN can be viewed as assigning the k -
 318 nearest neighbors a weight $1/k$ and all others zero weight. The conditional probability that
 319 observation x belongs to class k is then estimated as

$$p_k(X) = \Pr(Y = k | X = x) = \frac{1}{k} \sum_{i \in N_k} I(y_i = k) \quad (4)$$

320 KNN assigns x to the class k that has the largest probability. As given in Eq. 4, the choice of K
 321 determines the decision boundary between the classes and the accuracy of the model. A lower K
 322 corresponds to a low-bias (boundaries between output classes are less distinct) and high-variance
 323 (more sensitive to noise) classifier and a higher K corresponds to the low-variance and high-bias
 324 classifier.

325 **DECISION TREE**

326 Decision tree (DT) is a non-parametric test for classifying data and the algorithm generates a
 327 tree-like graph using the training data. The decision tree partitions the data into distinct and non-
 328 overlapping regions comprised of a root node (formed from the entire data), interior nodes, and
 329 terminal nodes. Each node in a decision tree has only one parent node and binary splits. The two
 330 main steps involved in building the decision tree are:

- 331 1. Dividing the training set space (X_1, X_2, \dots, X_p) into J distinct and non-overlapping regions
 332 R_1, R_2, \dots, R_J . The binary splitting to divide the regions is carried out based on a Gini
 333 Index (GI), which is defined as (Breiman 1984)

334

$$GI = \sum_{j=1}^k \hat{p}_{mj} (1 - \hat{p}_{mj}) \quad (5)$$

335 where \hat{p}_{mj} represents the portion of training observations in the m^{th} region from the j^{th}
 336 class. GI is measure of the total variance across k classes. Starting with the root node, all
 337 the predictors are assembled and almost all their possible split points are tested by
 338 computing the GI for each split. The predictor and split point with the lowest GI is then
 339 used to create two child interior nodes. This process is repeated to recursively partition
 340 the dataset until the entire training space is divided into J regions. The growth of the tree
 341 can be controlled appropriately by imposing a penalty factor on GI as

342

$$GI = \sum_{j=1}^k \hat{p}_{mj} (1 - \hat{p}_{mj}) + \alpha |T| \quad (6)$$

343 where α is the tuning parameter, and $|T|$ is the number of terminal nodes.

- 344 2. Determining whether there is over fitting or if the tree is unnecessarily complex. This is
 345 achieved by cost complexity pruning, which considers a sequence of trees indexed by a
 346 nonnegative tuning parameter, α . The optimal value of α can be obtained by k -fold cross
 347 validation where the trees that reduce the GI of the holdout data are removed (Friedman
 348 et al. 2001).

349 Fig. 7 shows a schematic representation of a simple DT (for illustrations purposes only) where
 350 two predictor variables ($Sa_{0.3s}$ and age) are used to predict green (G), yellow (Y), and red (R)
 351 tags. The tree initially divides the training dataset at the “root” (top) node based on the age of the
 352 building. Buildings that are less than or equal to 70 years in age are placed in one subset (shown
 353 as left branch) and buildings that are more than 70 years in age are placed in another (shown as
 354 the branch on the right). Subsequent splits are based on the value of $Sa_{0.3s}$ until the stopping
 355 criteria are satisfied. At the “leaf” nodes (the ones lowest in the tree), the observations are
 356 classified based on the tag that is most represented. An alternative representation of the decision
 357 tree with $Sa_{0.3s}$ and age as predictor variable is shown in Fig. 7b. Note that the parameters in the
 358 DT are learned from the data and there are no inherent assumptions in the decision boundaries.

359 **RANDOM FOREST**

360 Random forest (RF) is a learning method that consists of an ensemble of tree-structures. RF
 361 takes advantage of two powerful machine learning techniques: bagging and random feature
 362 selection (Breiman 2001). In bagging, each tree is independently constructed using a bootstrap
 363 sample of the training data, and the mean value of the outputs of the trees is used for prediction
 364 (Breiman 1996). RF is a revised version of bagging. Instead of using all predictor variables as in
 365 DT, RF randomly selects a subset of predictor variables to be split at each node when growing a
 366 tree. The main steps in the RF algorithm can be described as follows:

- 367 1. Generate n_t bootstrap samples from the training dataset.
 368 2. Generate a decision tree from each bootstrap sample by selecting the best split among the
 369 dataset.
 370 3. Predict the output of a new dataset by averaging the aggregate of predictions of n_t
 371 decision trees.

372 The output of the RF prediction can be expressed as:

373

$$\hat{f}_{RF}^{n_t}(x) = \frac{1}{n_t} \sum_{i=1}^{n_t} f_{n_t}(x) \quad (7)$$

374 where $\hat{f}_{RF}^{n_t}(x)$ denotes the outcome of the RF prediction (average value) from a total of n_t trees,
375 and $f_{n_t}(x)$ is the individual prediction of a tree for an input vector x . The variance of the average
376 of n_t random variables with a correlation coefficient ρ and standard deviation σ is (Friedman et
377 al. 2001):

378

$$\text{var}_{n_t} = \rho\sigma^2 + \frac{1-\rho}{n_t}\sigma^2 \quad (8)$$

379 An estimate of the error rate can be obtained from RF using the following procedure:

- 380 1. Generate predictions using the original dataset excluding the bootstrap sample (*out-of-bag*,
381 or OOB, data) using the tree grown with the bootstrap sample for each iteration.
- 382 2. Aggregate the OOB predictions and calculate the error rate, which is denoted as the OOB
383 estimate of the error rate.

384 Random forest consists of an ensemble of simple tree predictors, each capable of producing a
385 response when presented with a set of values of input variables. The training algorithm generates
386 random forests by bootstrap aggregating or bagging. After training, the prediction of a vector x
387 can be achieved by averaging the predictions from the individual regression trees (B) using Eq.
388 (9).

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B f_b(x) \quad (9)$$

389 where f_b is an individual regression tree.

390 Fig. 8 illustrates the decision boundaries and classification regions obtained from each of the
391 machine learning algorithms when age and $Sa_{0.3s}$ are used as the predictors (for illustration
392 purposes, models constructed using all predictors are described in the next section) on the actual
393 dataset. The observed data points are also shown as circles whose colors match the actual tag.
394 Fig. 8a shows that the linear decision boundary assumption that is made by LDA results in only
395 green and yellow tags. Whereas, the complex nonlinear decision boundaries adopted by the other
396 three algorithms (KNN, DT and RF shown on Figs. 8b, 8c and 8d, respectively) are better able to
397 capture the patterns in the real data.

398 **IMPLEMENTATION OF MACHINE LEARNING TECHNIQUES TO PREDICT
399 BUILDING TAG ASSIGNMENTS**

400 The machine learning techniques described in the previous section are used to predict the
401 assigned building tags from the 2014 South Napa earthquake. Note that all the predictor
402 variables, $X = (X_1, X_2, \dots, X_8)$, described in the previous section are used to establish the
403 prediction models. The performance of each machine learning model is evaluated using a
404 confusion matrix, which is a plot of the observed versus the predicted tag assignment. In other
405 words, each element in a confusion matrix C_{ij} ($i = 1:3$, $j = 1:3$) is equal to the number of

406 observations known to be in tag i , but predicted to tag j . Hence, each diagonal element in the
407 confusion matrix represents the tag assignments that are classified correctly by the machine
408 learning algorithm. Each off-diagonal elements represents the tags that are incorrectly predicted.
409 The confusion matrices for the assigned tags based on the training and testing sets are shown in
410 Figs. 9 and 10, respectively. The performance of the machine learning technique in the tag
411 assignment can be evaluated through precision and recall. The percentage of predicted tags that
412 are correctly assigned by the machine learning algorithm is called as precision, which is given in
413 the fourth row of the confusion matrix. The percentage of actual tags that is correctly assigned by
414 the machine learning algorithm is known as the recall and is given as the fourth column in the
415 confusion matrix. A high recall and precision rate indicate the ability of the machine learning
416 algorithm to accurately assign tags. The ratio of the failure modes that are accurately predicted to
417 the total set is defined as the accuracy of the prediction model. For example, Fig. 9a shows the
418 confusion matrix for the training set obtained from LDA. It is observed that LDA can correctly
419 predict 226 out of 661 green tag assignments, 724 out of 884 yellow tag assignments, and 5 out
420 of 49 red tag assignments. Hence, LDA has a total accuracy of 60% (955/1594), which is
421 reported in element (4,4) of the confusion matrix. Fig. 9a shows that LDA has low recall for
422 green and red tag, i.e., out of 661 actual green tags, LDA was able to assign 226 as green (34%
423 recall for the green tag). Similarly, the recall by LDA for the red tag is only 10%. The following
424 inferences can be drawn from Figs 9 and 10.

- 425 • For the training set, KNN and RF have the highest accuracy rate (99%). Amongst KNN
426 and RF, KNN has a precision of 100% for the red tag, i.e. there is a 100% chance that the
427 red tag predicted by KNN corresponds to an actual red tag. Also, the KNN has the
428 highest recall for the green tag (100%). In other words, KNN was able to assign green
429 tags with an accuracy of 100%.
- 430 • LDA has the lowest overall accuracy, precision and recall of all the machine learning
431 methods under consideration. This indicates that the decision boundary between the tags
432 are highly non-linear and complex. As mentioned before, LDA assigns a linear decision
433 boundary between the different tag classes.
- 434 • Although KNN has high overall accuracy for the training set, it has less accuracy than RF
435 for the testing set. This is due to the overfitting potential of KNN which has been
436 reported in other studies (Friedman et al. 2001).
- 437 • As noted before, the testing set is the data that is unknown to the model and is used to
438 evaluate the performance of the fitted model. It is observed in Fig. 10 that RF performs
439 the best amongst the machine learning models considered in the study based on the
440 overall accuracy for the testing set. RF has an overall accuracy of 66%.
- 441 • For the test set, RF has a recall of 79% for yellow tag. This suggests that RF was able to
442 assign yellow tags with a reasonable accuracy. However, it has a low recall of 13% for
443 red tag. The poor red tag prediction by RF is due to the lack of a specific pattern for red
444 tags compared to yellow tags as shown in Fig. 5. As shown in Fig. 5, for a same set of
445 input parameters, a red or yellow tag is possible, and RF is not able to clearly distinguish
446 between these tags as a function of the input parameters. Also, of the 2276 building
447 considered in the study, less than 4% of the buildings have a red tag. It is worth noting
448 that, while the misclassification of the red-tags as yellow is not ideal, in the context of
449 rapid assessment of building damage to guide personnel resource allocation for detailed

450 inspection (the intended application of the machine learning models), it is not too
451 concerning. The reason stems from the fact that yellow tags are often used, even during
452 in-person inspections, for cases where the safety concerns are not obvious and further
453 assessment is required. This additional assessment (or ameliorative action) often leads to
454 the tag assignment being re-assigned as green or red. If the machine learning models
455 were incorrectly assigning actual red tags as green, that would be much more of a
456 concern. It is also worth noting that the type of predictive capability assessment
457 conducted in the current paper (i.e. evaluating prediction accuracy at the building level
458 based on a training-testing) has not been performed for other damage assessment
459 approaches (e.g. HAZUS). As such, the accuracy of the machine learning models relative
460 to the status quo is still unknown and should be explored in future studies.

- 461 • Three strategies were considered to address the skewed distribution of the red tags (i.e.
462 the low numbers relative to the red and green tag): under sampling the majority class
463 (e.g. reducing the green tagged building sample size to match the red-tag buildings) or
464 over sampling the minority class (sample red and yellow tagged buildings to match the
465 green tag buildings), and adjusting misclassification costs to represent realistic penalties.
466 Because red tagged buildings are a significant minority within the inventory, under
467 sampling is not feasible. Moreover, over sampling approaches such as Synthetic Minority
468 Over Sampling Technique (SMOTE) are not reliable when the training data is not
469 linearly separable (Wang et al. 2006). As seen from the scatter plot in Fig. 5, the tag
470 distribution does not have a clearly distinguishable pattern with the input variables.
471 Hence, adjusting misclassification costs to represent realistic penalties, is applied to
472 address the challenge of the skewed distribution of the red tags. However, the overall
473 accuracy was increased by only 0.2% with the adjustment of misclassification costs.
- 474 • Overall, once the training model is available, the machine learning methods can rapidly
475 predict the level of building damage (as classified by assigned ATC-20 tags) with a
476 reasonable level of accuracy.

477 As RF has superior performance for the testing set compared to the other machine learning
478 models, further discussion in this paper is restricted to the RF-based prediction models.

479 FURTHER INVESTIGATION OF RF-BASED TAG ASSIGNMENT METHODOLOGY.

480 The performance of the RF-based tag assignment methodology is explored further by
481 conducting a sensitivity study. Fig. 11a shows the variation of the overall accuracy with the
482 number of trees, using different combinations of the various input parameters. It is observed that
483 the RF model that is constructed using all six predictors has an upper bound overall accuracy. In
484 other words, the overall accuracy does not increase significantly when the number of trees (N_t)
485 exceeds 100. Note that the RF results presented in the above section corresponds to 100 trees. It
486 is also noted from Fig. 11a that reasonably accurate predictions (around 65%) can be achieved by
487 using only $Sa_{0.3s}$, fault distance and age as predictor variables. As such, even if the geometric
488 configuration of the buildings is unknown, the machine learning model can be employed for tag
489 assignment if these three variables are available. As noted in the Introduction section, ShakeCast
490 provides estimates of the shaking intensity and fault distance within minutes following an
491 earthquake. The current study shows that, using a third parameter, building age, the machine
492 learning model can be deployed for rapid damage assessment. From Fig. 8a, it is interesting to
493 note that, when only $Sa_{0.3s}$ and age are used as the predictor variables in the RF model, the

494 predictions are significantly less accurate compared to when $Sa_{0.3s}$, age and fault distance are
495 used. This, despite the earlier observation in Fig. 5, which did not show a clear visual trend
496 between the tag assignment and the fault distance and between the fault distance and $Sa_{0.3s}$.
497 There are two key points here. First, this series of observations points to the fact that the fault
498 distance captures site-specific factors that are not reflected in $Sa_{0.3s}$. Note that this conclusion is
499 specific to the Napa earthquake data set and cannot be generalized for all earthquakes. Moreover,
500 even though there is no clear visual evidence of a relationship between fault distance and the
501 assigned tag (Fig. 5), the RF algorithm can reveal a subtle association between the two (Fig.
502 11a).

503 RF also yields the relative importance of each of the input parameters. This is achieved
504 by noting the increase in the OOB error of the variable for different permutations while the other
505 variables are kept constant. Fig. 11b shows the relative importance of each input parameter
506 considered in the current study. It also underscores the point that $Sa_{0.3s}$, fault distance and age are
507 the critical parameters for the building tag assignment.

508 To evaluate the spatial performance of RF-based tag assignment methodology, Figs. 12
509 and 13 compare the geographic distribution of the RF-based and actual tags for training and
510 testing sets, respectively. Consistent with earlier observation, the spatial distribution of the RF-
511 based and actual tags are comparable for the training set. For the testing set, RF predicts the
512 spatial distribution of yellow tags with a high level of accuracy. However, the RF-based red tags
513 are not predicted as accurately as the yellow tags.

514 As noted in the Introduction section, the tag assignments are based on different types of
515 damage. Fig. 14 shows the percentage distribution of damage noted from the actual data and RF-
516 based prediction. As noted earlier, most of the yellow tagged buildings experienced chimney
517 damage and the error between actual chimney damage and RF-prediction is less than 10%.
518 Similarly, for damage to other types of components such as superstructure walls, cripple walls
519 and porches, there is a good agreement between the actual and RF-based yellow tags.

520 CONCLUSION

521 Assessing building damage after an earthquake is a critical step in emergency response and
522 recovery planning. The damaged state of building can vary from superficial cracking to complete
523 collapse depending on the building properties, soil conditions and earthquake and ground motion
524 characteristics. The spatial distribution of impacts in an earthquake-affected area makes building
525 damage assessment a complex and time-consuming process. Tools such as ShakeMap and
526 ShakeCast provide rapid assessments of earthquake damage. However, their ability to predict the
527 spatial distribution of building damage with reasonable accuracy can vary depending on the
528 methodology used for the assessment.

529 This paper explores the effectiveness of using machine learning techniques to predict the
530 earthquake damage to buildings. The considered algorithms include discriminant analysis, K-
531 nearest neighbors, decision trees and random forests. Data from the 2014 Napa earthquake is
532 used for the evaluation where building damage is classified by their ATC 20 tags (red, yellow
533 and green). The predictor variables used as input parameters in the machine learning models
534 include the spectral acceleration corresponding to a period of 0.3s ($Sa_{0.3s}$) at the site of interest,
535 the fault distance (R_{jb}), the time-averaged shear-wave velocity to 30 m depth (V_{S30}), as well as
536 building characteristics such as the age (in years), number of stories, presence of irregularities,

537 value (in dollars) and floor area. The collected 2,276 building damage dataset is randomly
538 divided into training set and test subsets. The training set is used to establish the prediction
539 model and the test set is used to evaluate the model performance.

540 The results from the study showed that predictions based on the random forest (RF)
541 algorithm have an overall accuracy of 66% in assigning the tags and correctly assigned 79% of
542 the actual yellow tags in the test set. It is also noted that a good prediction can be obtained even
543 when $Sa_{0.3s}$, fault distance and age are the only predictor variables used in the model. The spatial
544 evaluation of the data set also underscores the ability of RF-based prediction model in accurately
545 assigning the yellow tags.

546 More work is needed to enable widespread adoption of machine learning algorithms for
547 earthquake-induced building damage predictions. The current study utilized data from a single
548 earthquake that was considered moderate in terms of the physical damage caused to
549 infrastructure. Moreover, only low-rise (one to three stories) residential buildings with specific
550 types of damage (e.g. chimney, cripple wall, porch damage) was included in the dataset. As such,
551 the findings may not apply to other construction types, building heights and damage classes.
552 Future research should also benchmark the performance of the machine learning methods
553 relative to the status quo (e.g. the damage fragility approach used by HAZUS). Despite all these
554 limitations, the current study has revealed the great potential for using artificial intelligence to
555 inform post-disaster emergency response and recovery planning. Future research should focus on
556 collecting additional data that is diverse in terms of the type of construction and scale and classes
557 of damage. By incorporating larger datasets (e.g. from multiple events) and additional site- (e.g.
558 soil conditions) and building-specific (e.g. presence of known vulnerabilities such as a soft-story)
559 predictor variables, the overall accuracy and generalizability of prediction models can be
560 improved in future works. Lastly, the application of machine learning algorithms can be
561 extended to other types of hazards (e.g. hurricanes) and building impact measures such as
562 economic losses and downtime.

563 ACKNOWLEDGEMENTS

564 The research presented in this paper is supported by the National Science Foundation CMMI
565 Research Grant No. 1538747.

566 REFERENCES

- 567 ASCE (2016), ASCE/SEI 7 Minimum Design Loads For Buildings and Other Structures, ASCE 7 & SEI
568 Standards, Reston, VA.
- 569 ATC, ATC-20 Procedures for Post-earthquake building safety evaluation procedures. Redwood, CA:
570 Applied Technology Council, 1995.
- 571 Baltay, A. S., Boatwright, J. (2015) Ground-motion observations of the 2014 South Napa earthquake.
572 *Seismological Research Letters*. **86**, 355-60.
- 573 Boatwright, J., Blair, J. L., Aagaard, B. T., Wallis, K. (2015) The distribution of red and yellow tags in
574 the City of Napa. *Seismlogical Research Letters*, **86**, 361-8.
- 575 Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A. (1984). Classification and regression trees CRC
576 press.
- 577 Breiman, L., (2001). Random forests. *Machine learning* **45**, 5-32.

- 578 Breiman, L., (1996). Bagging predictors. *Machine learning* **24**, 123-40.
- 579 Brocher, T. M., Baltay, A. S., Hardebeck, J. L., Pollitz, F. F., Murray, J. R., Llenos, A. L., Schwartz, D.
580 P., Blair, J. L., Ponti, D. J., Lienkaemper, J. J., Langenheim, V. E., (2015) The M w 6.0 24 August
581 2014 South Napa earthquake. *Seismological Research Letters*, **86**, 309-26.
- 582 Fischer, E. C., Liu, J., Varma, A. H. (2016) Investigation of cylindrical steel tank damage at wineries
583 during earthquakes: Lessons learned and mitigation opportunities. *Practice Periodical on Struct.
584 Dsgn Constr.*, **21**, 04016004.
- 585 Friedman, J., Hastie, T., Tibshirani, R. (2001) The elements of statistical learning *Springer series in
586 statistics* Springer, Berlin;
- 587 HAZUS-MH (2003) Multi-Hazard Loss Estimation Methodology: Earthquake Model. *Department of
588 Homeland Security, FEMA*, Washington, DC. 2003.
- 589 Johnson, L. A., & Mahin, S. A. (2016). The Mw 6.0 South Napa Earthquake of August 24, 2014: A
590 Wake-up Call for Renewed Investment in Seismic Resilience across California. California Seismic
591 Safety Commission, *Pacific Earthquake Engineering Research Center (PEER)*, CSSC Publication,
592 16-03.
- 593 Kwak, D. Y., Stewart, J. P., Brandenberg, S. J., Mikami, A. (2016) Characterization of seismic levee
594 fragility using field performance data. *Earthquake Spectra*. **32**, 193-215.
- 595 Mangalathu, S., Jeon, J-S., DesRoches, R. (2018) Critical uncertainty parameters influencing seismic
596 performance of bridges using Lasso regression. *Earthquake. Engineering and Structural Dynamics*
597 2018 (in press).
- 598 Mangalathu, S., Jeon, J-S. (2018) Classification of failure mode and prediction of shear strength for
599 reinforced concrete beam-column joints using machine learning techniques. *Engineering Structures*
600 **160**, 85-94.
- 601 Nateghi, R., Guikema, S. D., Quiring, S. M. (2011) Comparison and Validation of Statistical Methods for
602 Predicting Power Outage Durations in the Event of Hurricanes." *Risk Analysis.*, **31**, 1897-906.
- 603 Ranf, R. T., Eberhard, M. O., & Malone, S. (2007). Post-earthquake prioritization of bridge inspections.
604 *Earthquake Spectra*, **23**(1), 131-146.
- 605 Silva, V., Crowley, H., Pagani, M., Monelli, D., Pinho, R. (2014) Development of the OpenQuake engine,
606 the Global Earthquake Model's open-source software for seismic risk assessment. *Natural Hazards.*
607 **72**, 1409-27.
- 608 Sun, H., Burton, H., Zhang, Y., & Wallace, J. (2018). Interbuilding interpolation of peak seismic response
609 using spatially correlated demand parameters. *Earthquake Engineering & Structural Dynamics*,
610 **47**(5), 1148-1168
- 611 Trifunac, M., Todorovska, M. (1997) Northridge, California, earthquake of 1994: density of red-tagged
612 buildings versus peak horizontal velocity and intensity of shaking. *Soil Dynamics and Earthquake.
613 Engineering*, **16**, 209-22.
- 614 USGS (2014), <https://earthquake.usgs.gov/earthquakes/eventpage/nc72282711/executive#dyfi>
- 615 Wald, D. J., Quitoriano, V., Heaton, T. H., Kanamori, H., Scrivner, C. W., Worden, C. B. (1999) TriNet
616 "ShakeMaps": Rapid Generation of Peak Ground Motion and Intensity Maps for Earthquakes in
617 Southern California. *Earthquake Spectra*, **15**, 537-55.
- 618 Wald, D. J., Worden, B. C., Quitoriano, V., Pankow, K. L. (2005) *ShakeMap manual: technical manual,
619 user's guide, and software guide*. 2005.

- 620 Wald, D., Lin, K-W, Porter, K., Turner, L. (2008) ShakeCast: Automating and Improving the Use of
 621 ShakeMap for Post-Earthquake Decision-Making and Response. *Earthquake Spectra.*, **24**, 533-53.
- 622 Worden, C. B., Thompson, E. M., Baker, J. W., Bradley, B. A., Luco, N., & Wald, D. J. (2018). Spatial
 623 and Spectral Interpolation of Ground-Motion Intensity Measure Observations. *Bulletin of the*
 624 *Seismological Society of America*, 108(2), 866-875.
- 625 Yune, H. "Family Makes the Case for a Second Napa Earthquake Fatality," Napa Valley Register,
 626 September 24, 2014, http://napavalleyregister.com/news/local/family-makes-the-case-for-a-second-napa-earthquake-fatality/article_0ed1554d-bf785776-82ce-e247ad7cff54.html

628

629 **List of Tables**

630 **Table 1.** Typical damage state descriptions associated with each tagging category

631

632 **Table 1.** Typical damage state descriptions associated with each tagging category

Field Inspection Damage Description	ATC-20 Tag
No visual damage; minor cracking inside home (lath and plaster); cracking along ceiling walls; no structural damage observed; damaged retaining wall; A/C unit broken; detached fence	Green
Top of brick chimney fell; wall cracks throughout the apartment; chimney collapsed; restricted access near the chimney; foundation cracks; restricted use of stair case; damage to garage	Yellow
Racking; cracking of cripple walls; collapse imminent; unstable house; foundation tilting; leaning structure; walls out of plumb	Red

633

634

635 **List of Figures**

636 **Fig. 1.** Spatial distribution on the damage of buildings.

637 **Fig. 2.** Summary of predictor variables and assigned tags

638 **Fig. 3.** Locations of ground motion recordings from 2014 South Napa earthquake

639 **Fig. 4.** Interpolation result of geometric mean PGA during 2014 southern Napa earthquake

640 **Fig. 5.** Scatter plot showing the assigned building tags as a function pairs of predictor variables.

641 **Fig. 6.** A schematic representation of decision boundaries (a) linear, and (b) non-linear for the
 642 case where only two predictors are considered

643 **Fig. 7.** Illustration of a decision tree

644 **Fig. 8.** Decision boundaries for machine learning algorithms with $S_{ao.3}$ (g) and age as input
645 variables and the actual tag assignments: (a) linear discriminant analysis, (b) k-nearest neighbors,
646 (c) decision trees, and (d) random forests

647 **Fig. 9.** Performance of various machine learning techniques for the training set.

648 **Fig. 10.** Performance of various machine learning techniques for the test set.

649 **Fig. 11.** a) Variation of overall accuracy with the number of trees for various combination of
650 input parameters and b) relative importance of the input parameters by RF.

651 **Fig. 12.** Comparing the spatial distribution of a) actual and b) RF-based tags using training data.

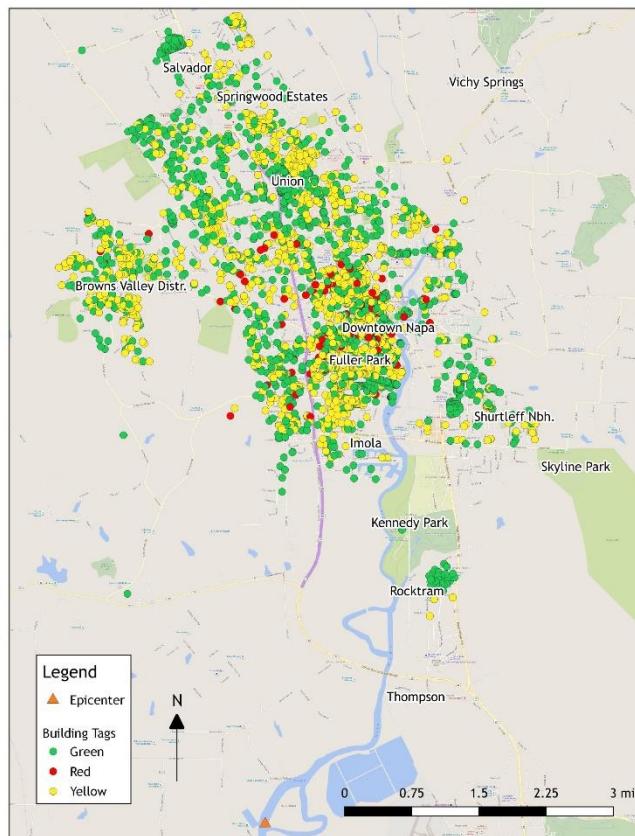
652 **Fig. 13.** Comparing the spatial distribution of a) actual and b) RF-based tags using testing data

653 **Fig. 14.** Comparing the observed and RF-predictions for specific types of yellow tagged damage
654 within the test set

655

656

657



658

659 **Fig. 1.** Spatial distribution on the damage of buildings.

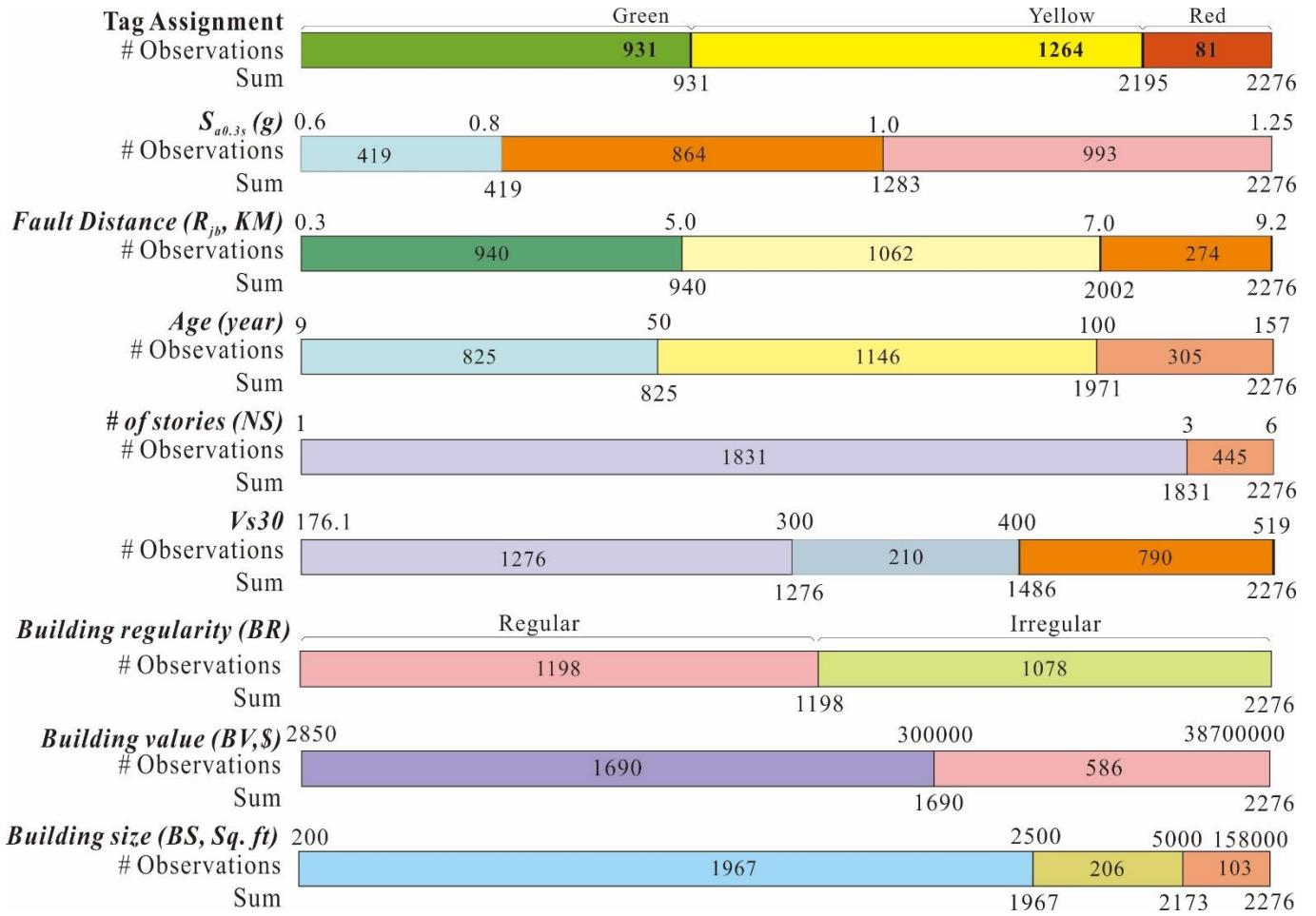
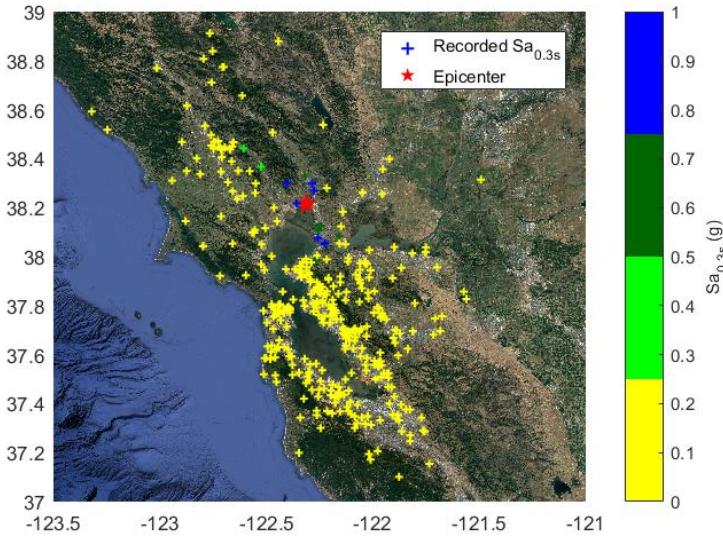


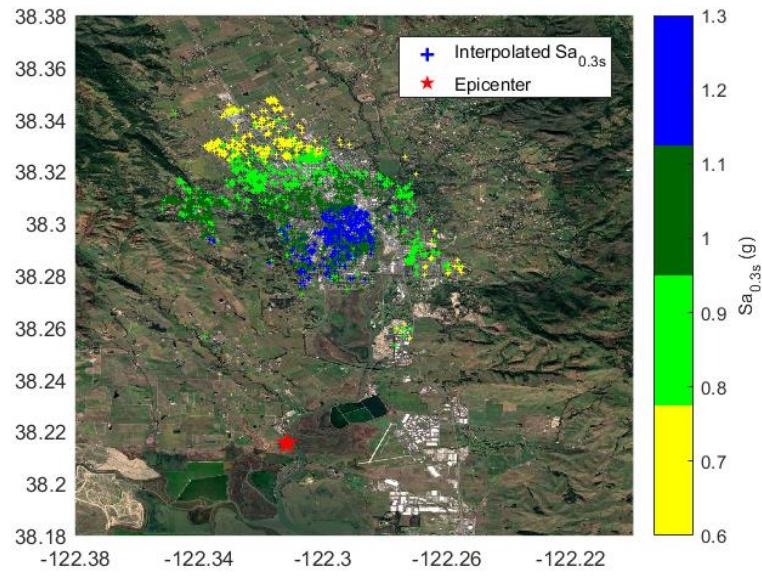
Fig. 2. Summary of predictor variables and assigned tags



662

663

Fig. 3. Locations of ground motion recordings from 2014 South Napa earthquake



664

665 **Fig. 4.** Interpolation result of geometric mean PGA during 2014 southern Napa earthquake

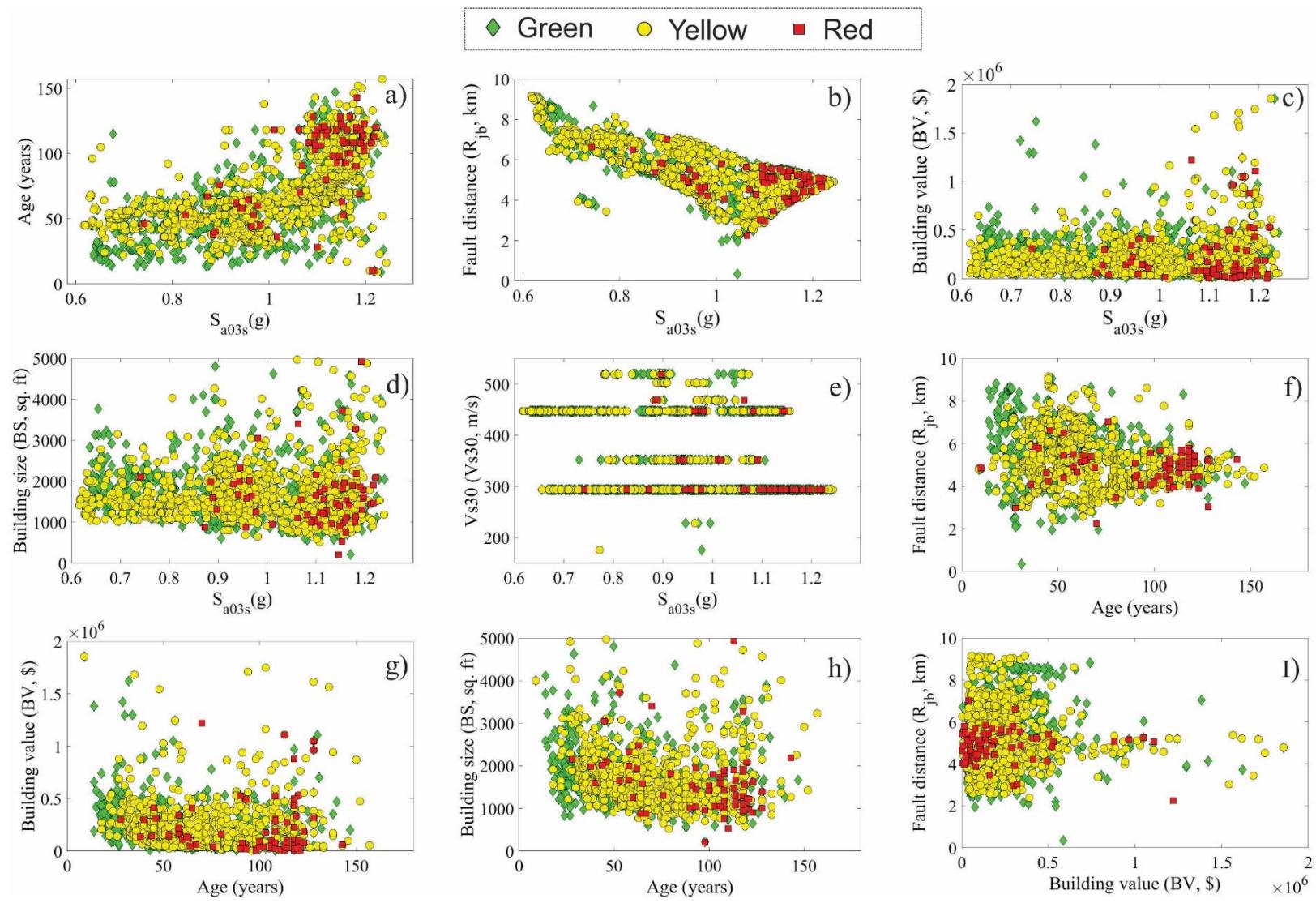
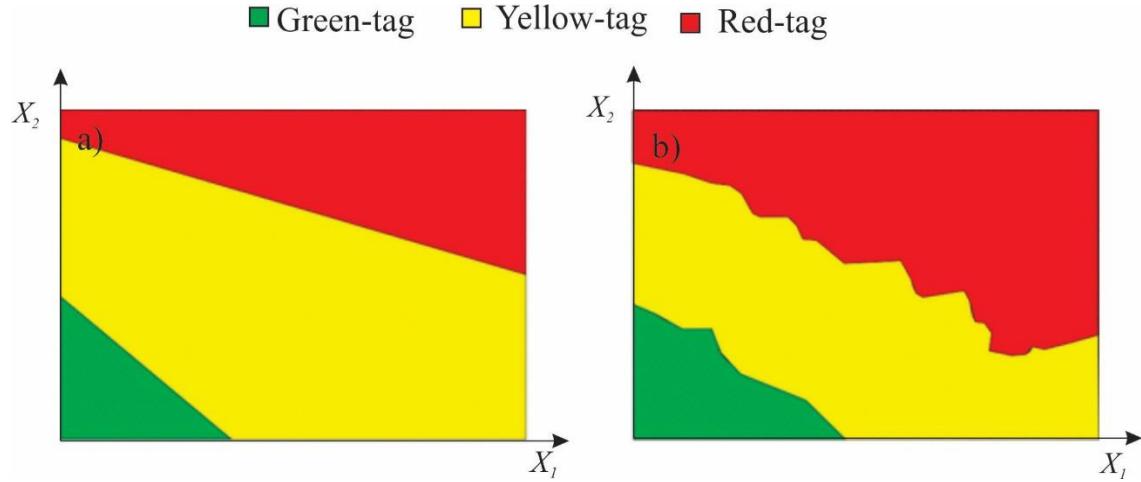


Fig. 5. Scatter plot showing the assigned building tags as a function pairs of predictor variables.



670 **Fig. 6.** A schematic representation of decision boundaries (a) linear, and (b) non-linear for the
671 case where only two predictors are considered

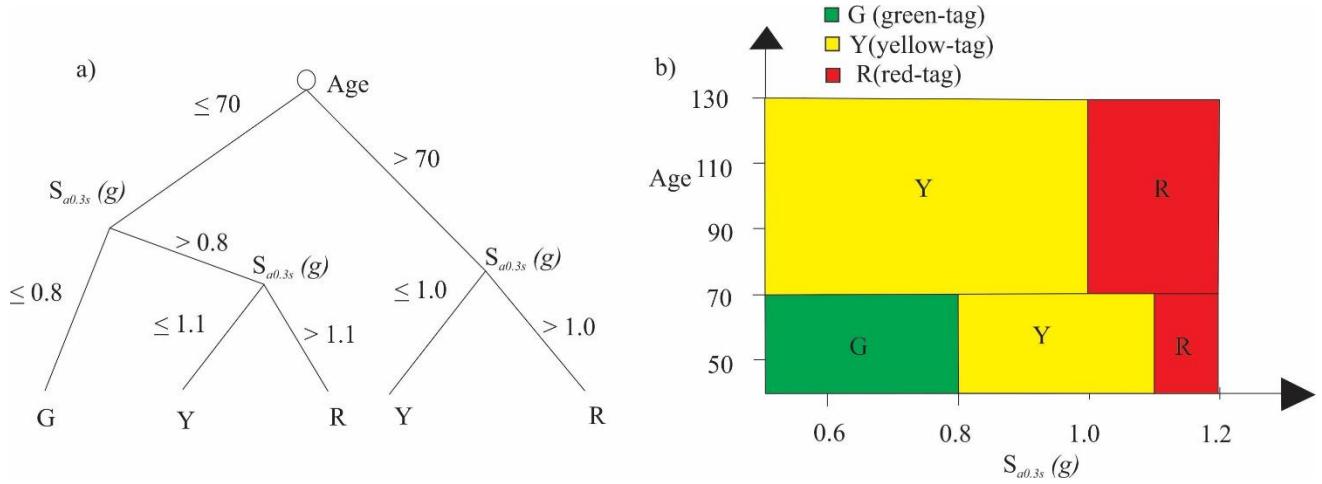
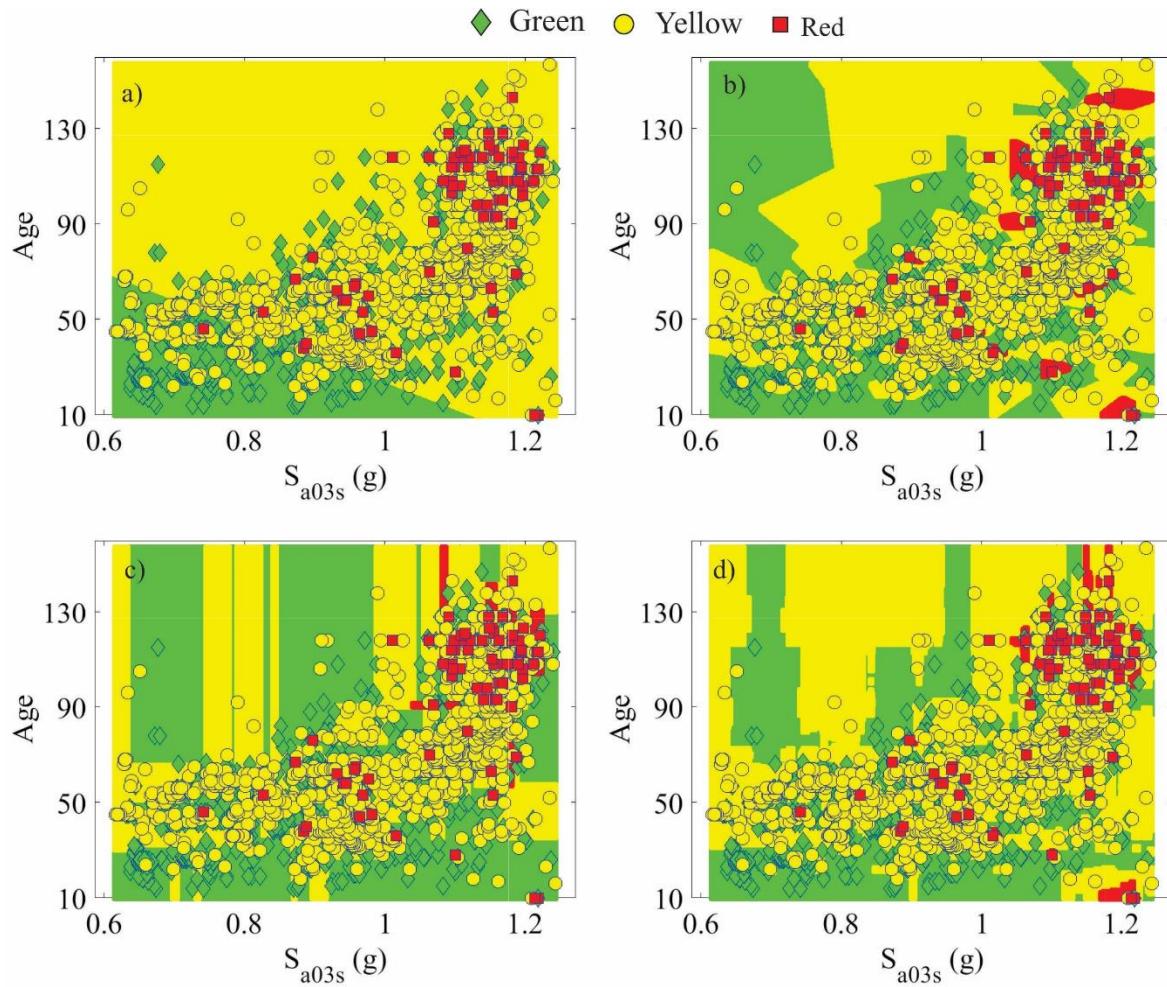


Fig. 7. Illustration of a decision tree



675

676 **Fig. 8.** Decision boundaries for machine learning algorithms with $S_{a0.3}$ (g) and age as input
 677 variables and the actual tag assignments: (a) linear discriminant analysis, (b) k-nearest neighbors,
 678 (c) decision trees, and (d) random forests

679

		Predicted class			
		G	Y	R	
Actual class	G	226 (14%)	434 (27.2%)	1 (0.1%)	34%
	R	153 (9.6%)	724 (45.4%)	7 (0.5%)	82%
		4 (0.3%)	40 (2.5%)	5 (0.3%)	10%
		59%	60%	38%	60%

(a) LDA

		Predicted class			
		G	Y	R	
Actual class	G	661 (41.4%)	0 (0%)	0 (0%)	100%
	R	0 (0%)	884 (55.5%)	0 (0%)	100%
		1 (0.1%)	1 (0.1%)	47 (2.9%)	96%
		99%	99%	100%	99%

(b) KNN

		Predicted class			
		G	Y	R	
Actual class	G	562 (35.2%)	97 (6.0%)	2 (0.1%)	85%
	R	61 (3.8%)	820 (51.4%)	3 (0.2%)	93%
		8 (0.5%)	15 (0.9%)	26 (1.6%)	53%
		89%	88%	84%	88%

(c) DT

		Predicted class			
		G	Y	R	
Actual class	G	660 (41.1%)	0 (0%)	1 (0.1%)	99%
	R	0 (0%)	884 (55.5%)	0 (0%)	99%
		0 (0%)	2 (0.1%)	47 (3.0%)	96%
		100%	99%	98%	99%

(d) RF

		Predicted class			
		G	Y	R	
Actual class	G	99 (14.5%)	169 (24.8%)	2 (0.3%)	37%
	R	60 (8.8%)	315 (46.2%)	5 (0.7%)	83%
		2 (0.3%)	29 (4.3%)	1 (0.1%)	3%
		62%	61%	13%	61%

(a) LDA

		Predicted class			
		G	Y	R	
Actual class	G	140 (20.5%)	125 (18.3%)	5 (0.7%)	52%
	R	107 (15.7%)	261 (38.2%)	12 (1.8%)	69%
		13 (1.9%)	16 (2.4%)	3 (0.4%)	9%
		54%	65%	15%	59%

(b) KNN

		Predicted class			
		G	Y	R	
Actual class	G	149 (21.8%)	117 (17.2%)	4 (0.6%)	55%
	R	120 (17.6%)	253 (37.1%)	7 (1.0%)	67%
		13 (1.9%)	17 (2.5%)	2 (0.3%)	6%
		53%	65%	15%	59%

(c) DT

		Predicted class			
		G	Y	R	
Actual class	G	141 (20.7%)	129 (18.9%)	0 (0%)	52%
	R	74 (10.9%)	302 (44.3%)	4 (0.6%)	79%
		7 (1.0%)	21 (3.1%)	4 (0.6%)	13%
		64%	67%	50%	66%

(d) RF

680

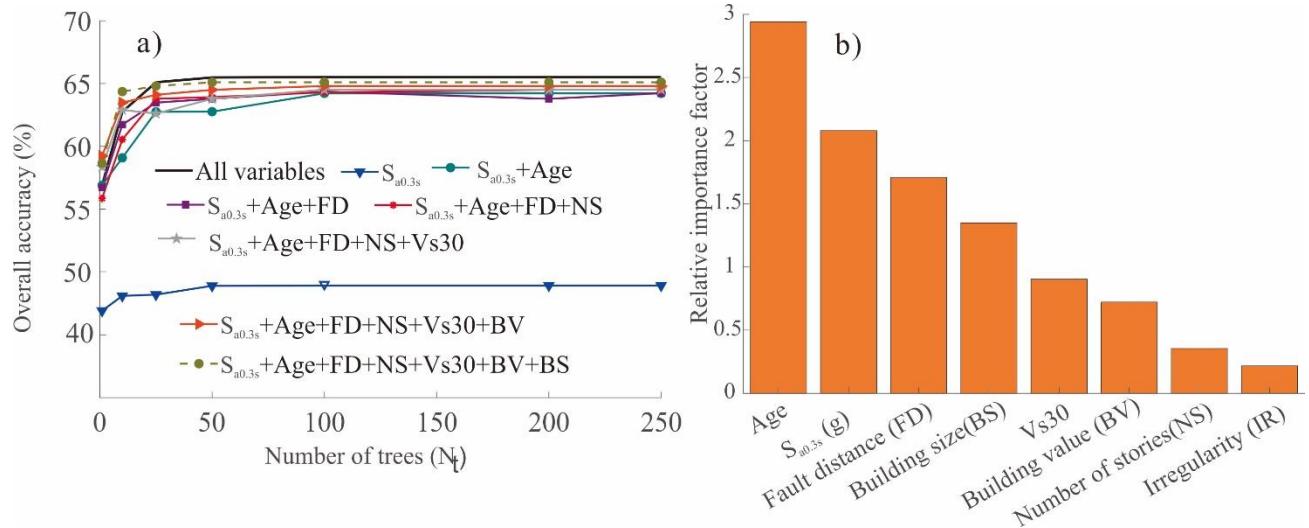
Fig. 9. Performance of various machine learning techniques for the training set.

682

683

684 **Fig. 10.** Performance of various machine learning techniques for the test set.

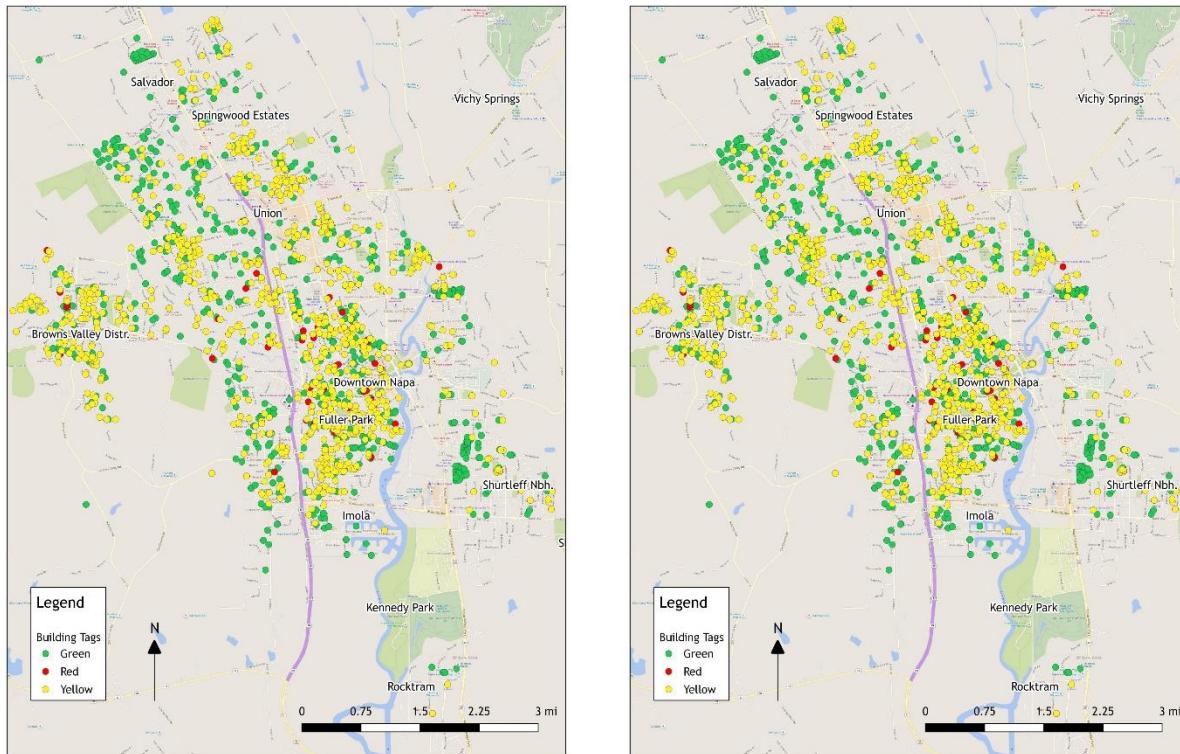
685



686

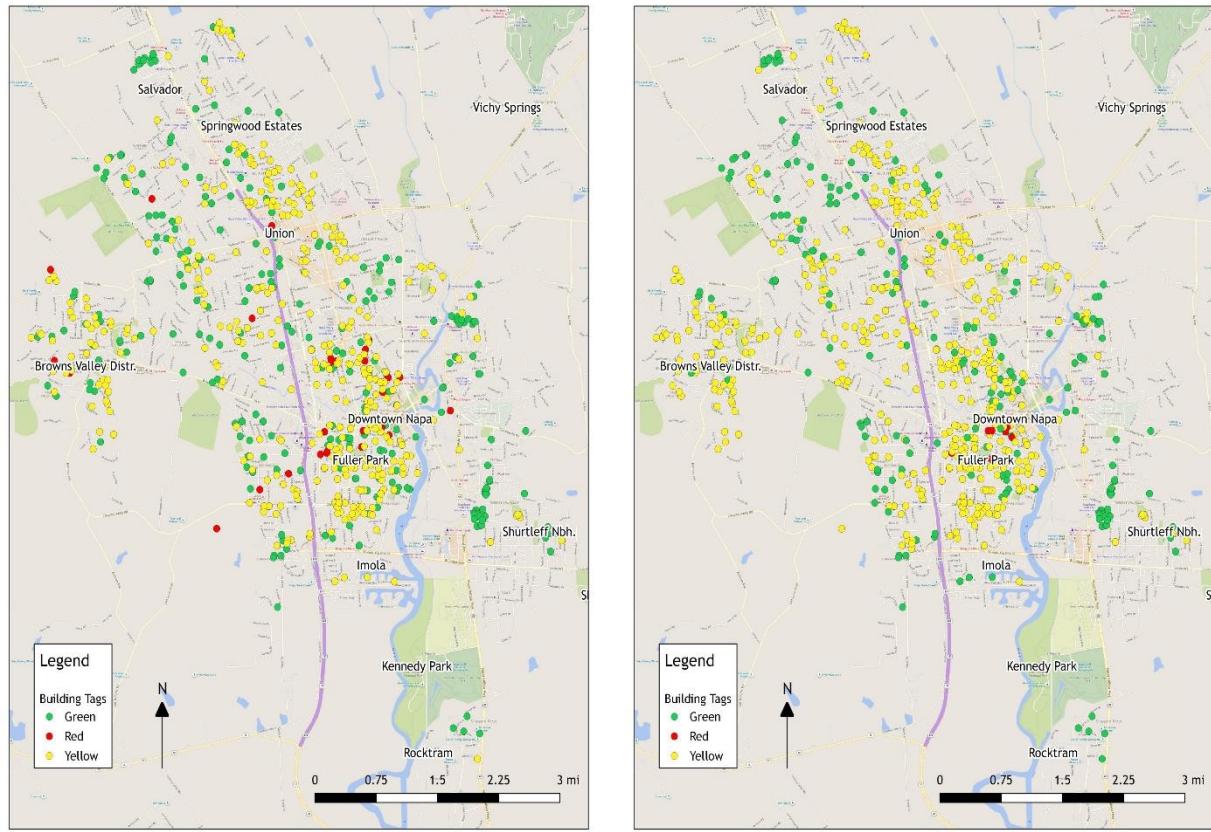
687 **Fig. 11.** a) Variation of overall accuracy with the number of trees for various combination of
688 input parameters and b) relative importance of the input parameters by RF.

689

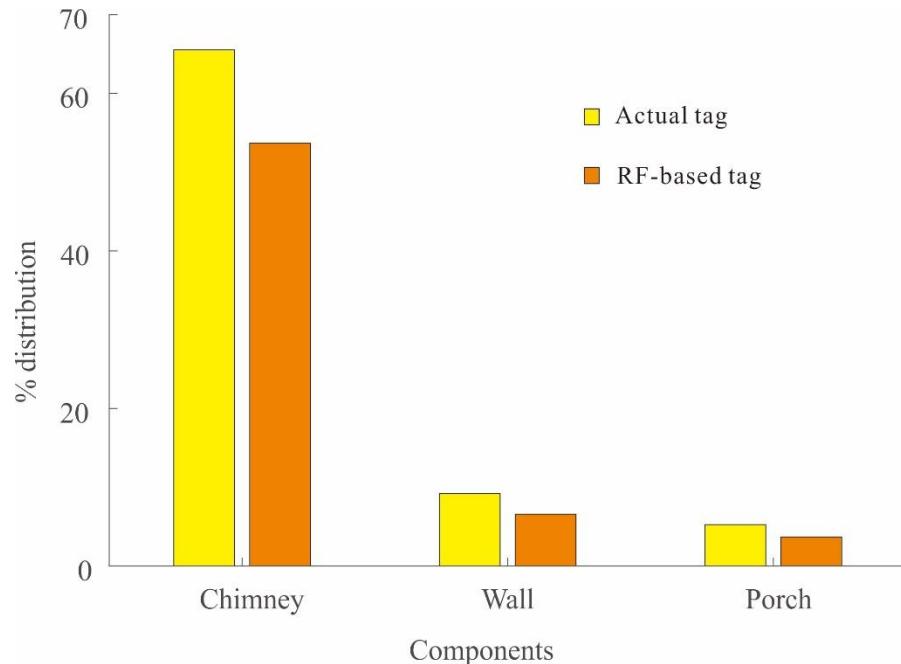


690 **Fig. 12.** Comparing the spatial distribution of a) actual and b) RF-based tags using training data.

691



692 **Fig. 13.** Comparing the spatial distribution of a) actual and b) RF-based tags using testing data
 693



694
 695 **Fig. 14.** Comparing the observed and RF-predictions for specific types of yellow tagged damage
 696 within the test set

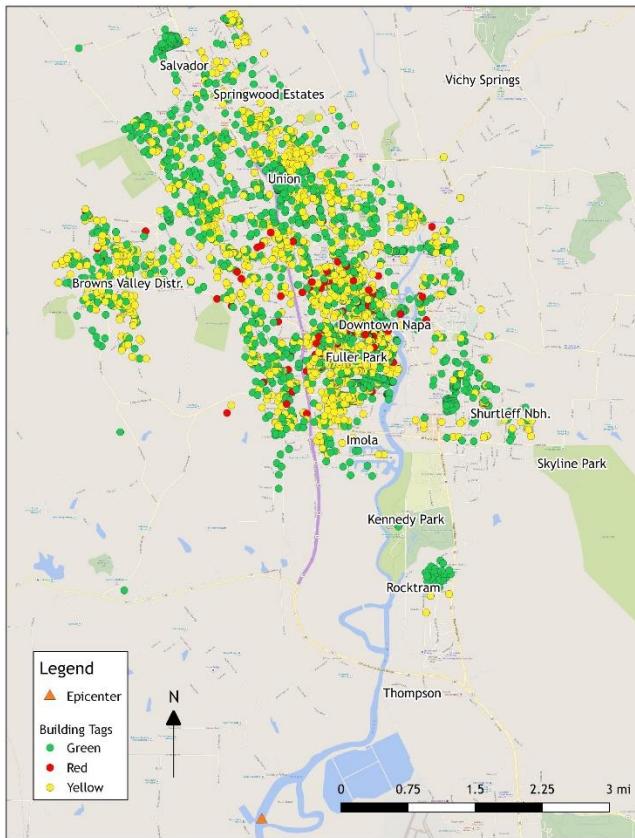


Fig. 1. Spatial distribution on the damage of buildings.

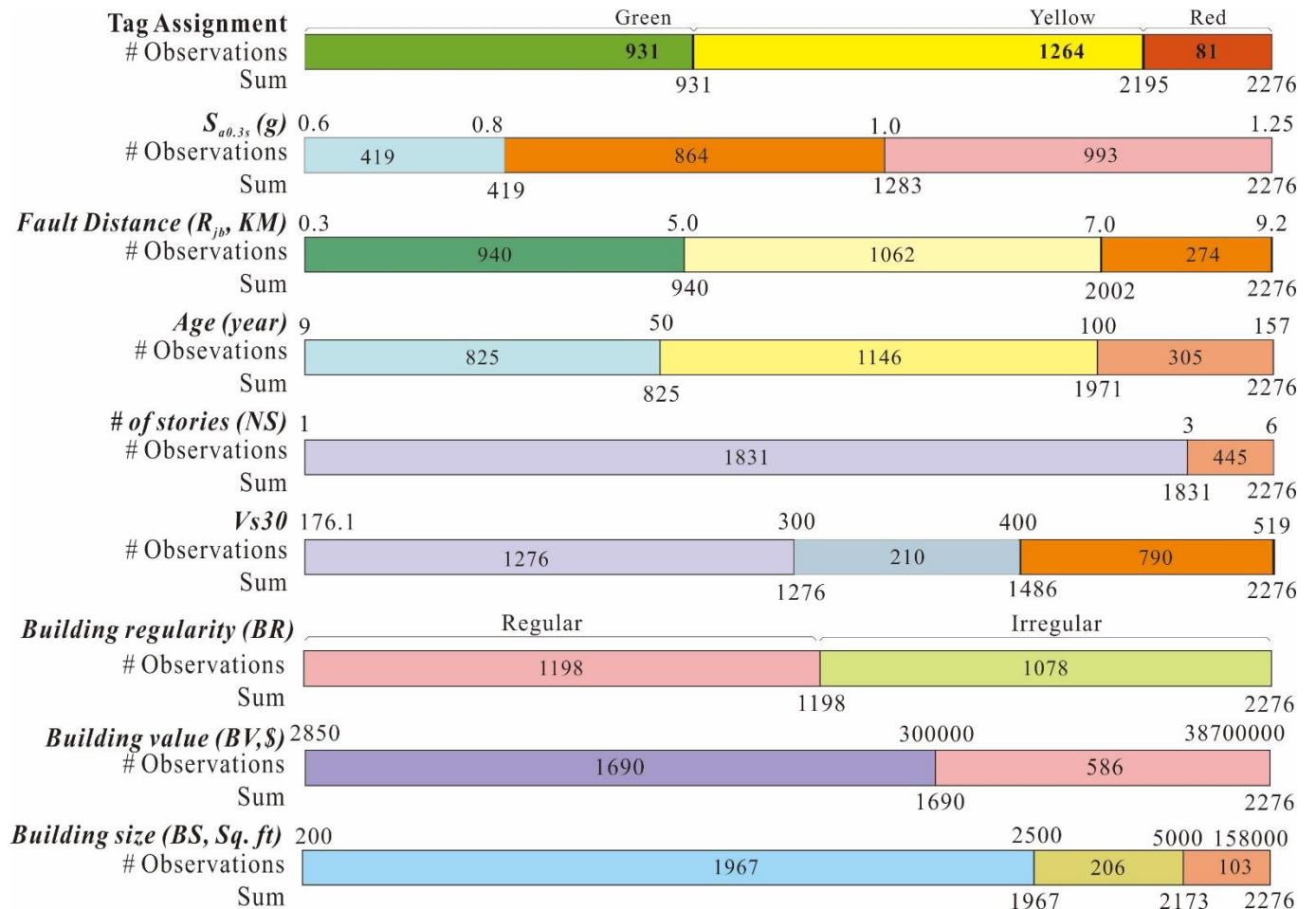


Fig. 2. Summary of predictor variables and assigned tags

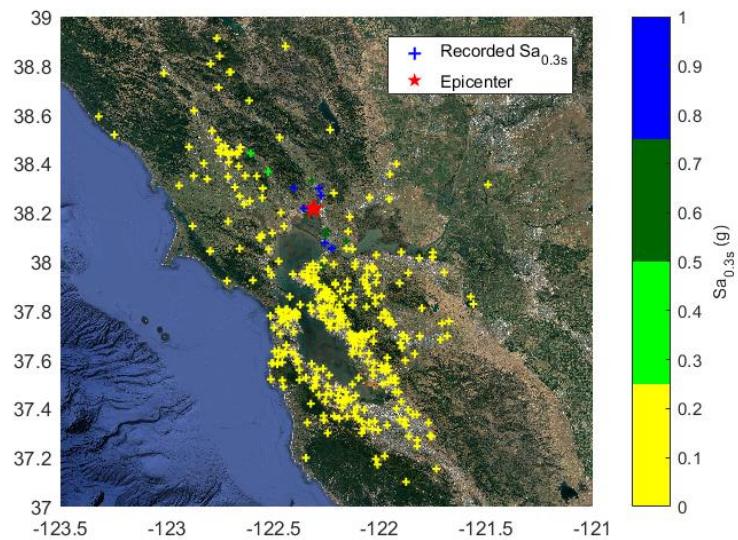


Fig. 3. Locations of ground motion recordings from 2014 South Napa earthquake

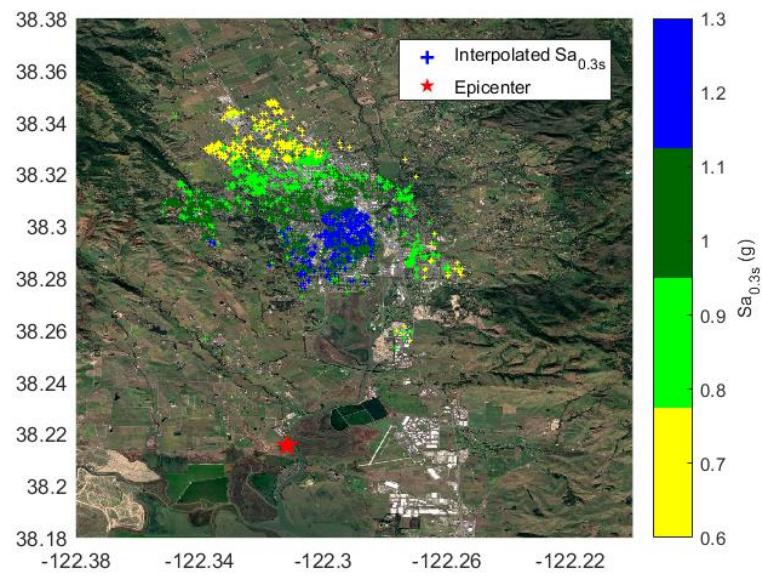


Fig. 4. Interpolation result of geometric mean PGA during 2014 southern Napa earthquake

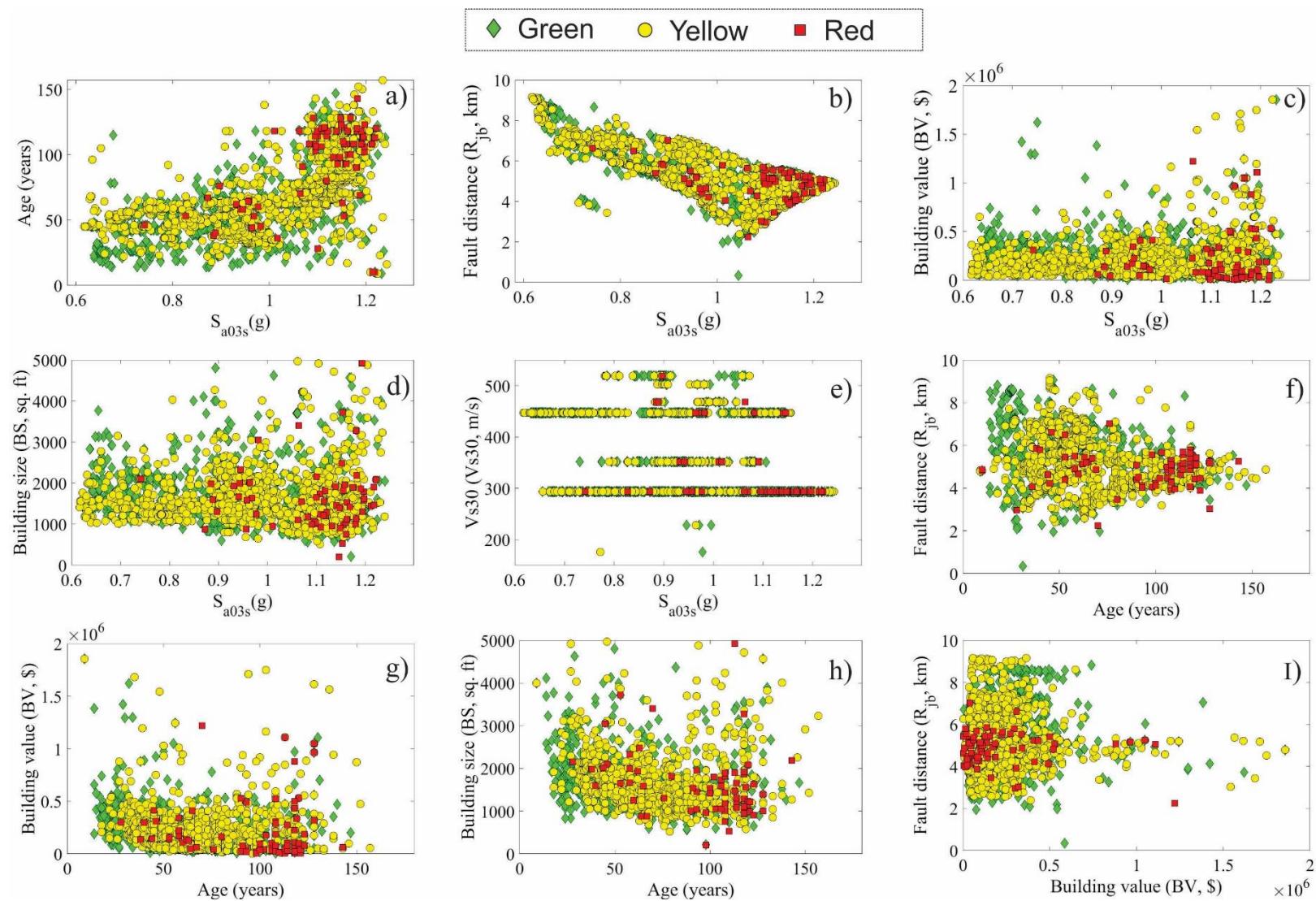


Fig. 5. Scatter plot showing the assigned building tags as a function pairs of **predictor variables**.

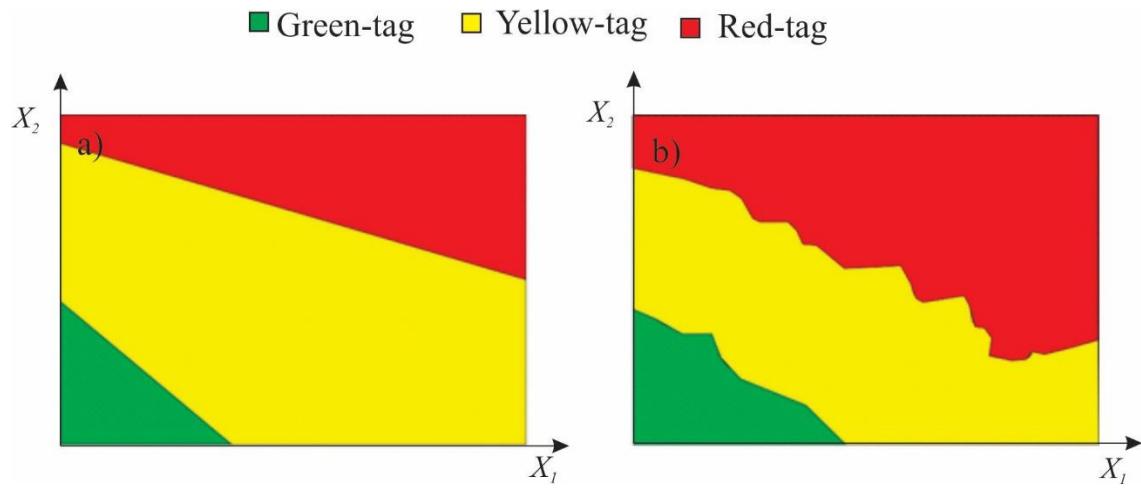


Fig. 6. A schematic representation of decision boundaries (a) linear, and (b) non-linear for the case where only two predictors are considered

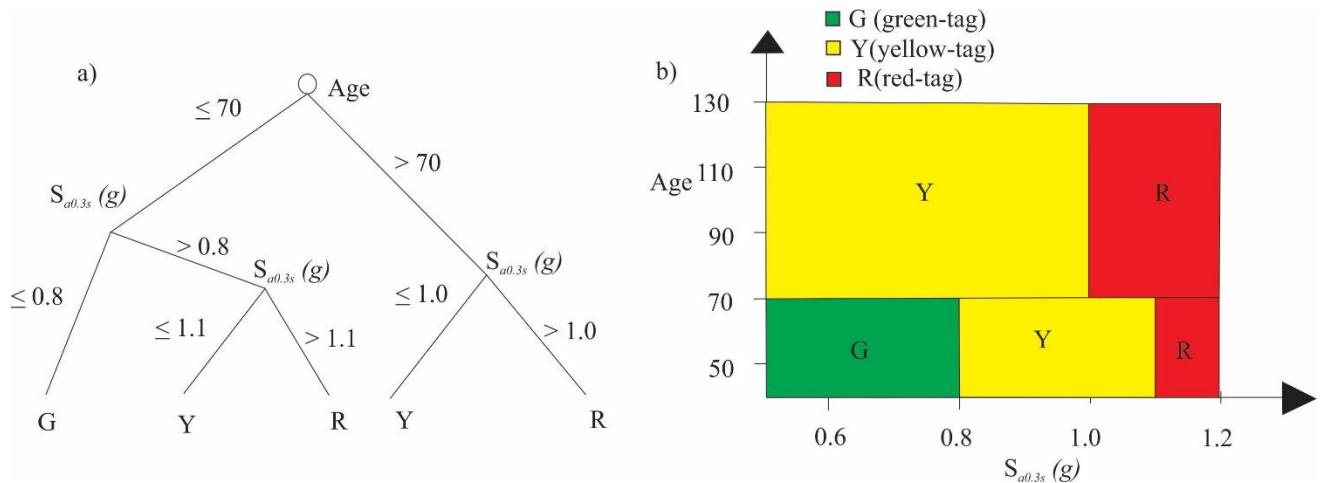


Fig. 7. Illustration of a decision tree

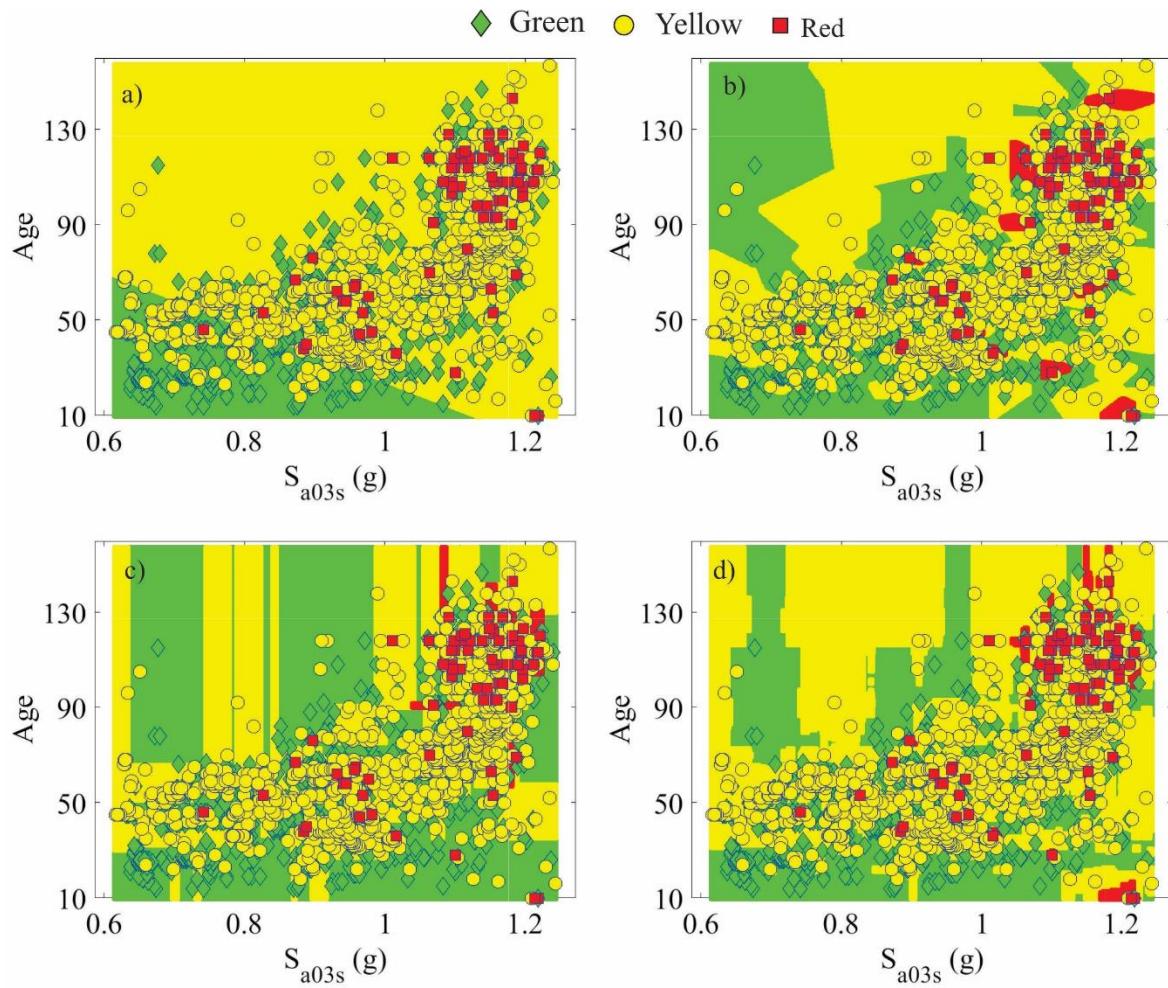


Fig. 8. Decision boundaries for machine learning algorithms with $S_{a0.3}$ (g) and age as input variables and the actual tag assignments: (a) linear discriminant analysis, (b) k-nearest neighbors, (c) decision trees, and (d) random forests

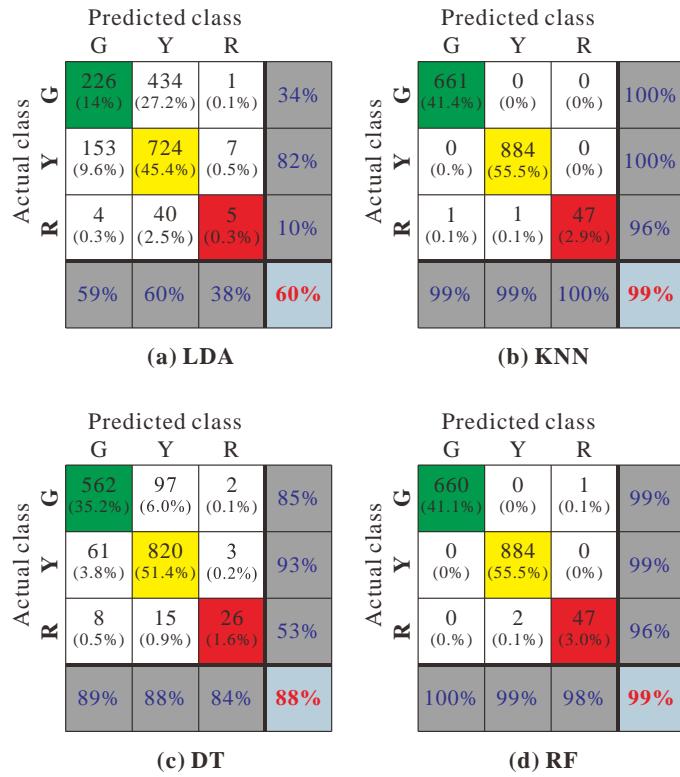


Fig. 9. Performance of various machine learning techniques for the training set.

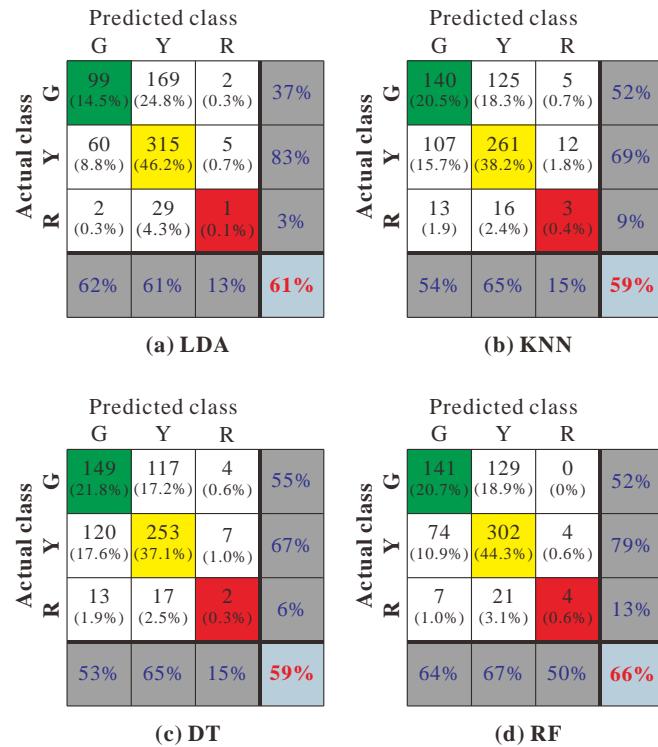


Fig. 10. Performance of various machine learning techniques for the test set.

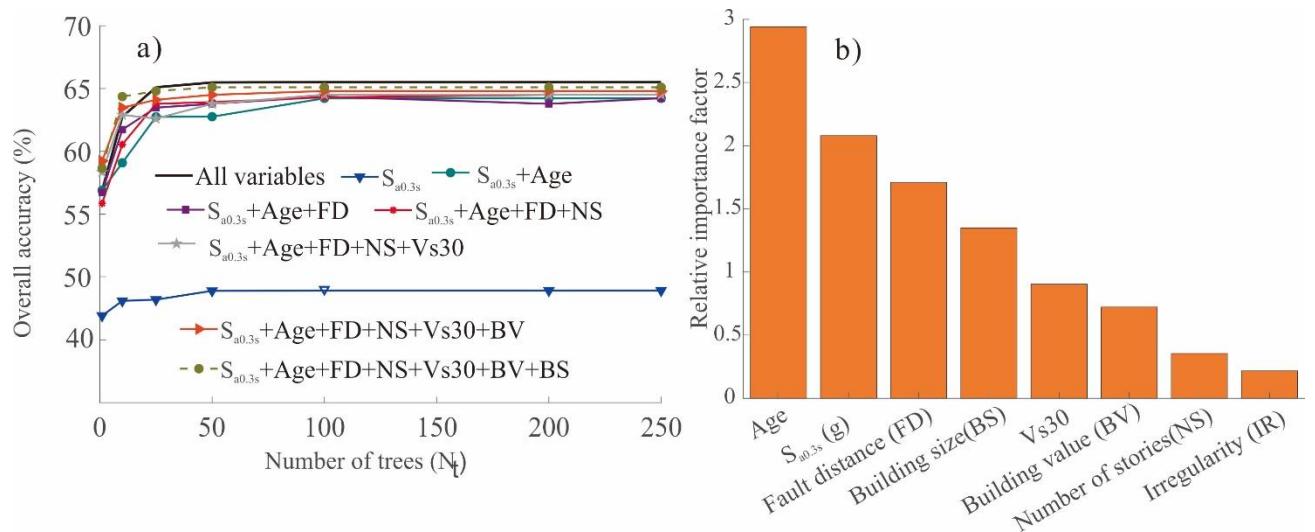


Fig. 11. a) Variation of overall accuracy with the number of trees for various combination of input parameters and b) relative importance of the input parameters by RF.

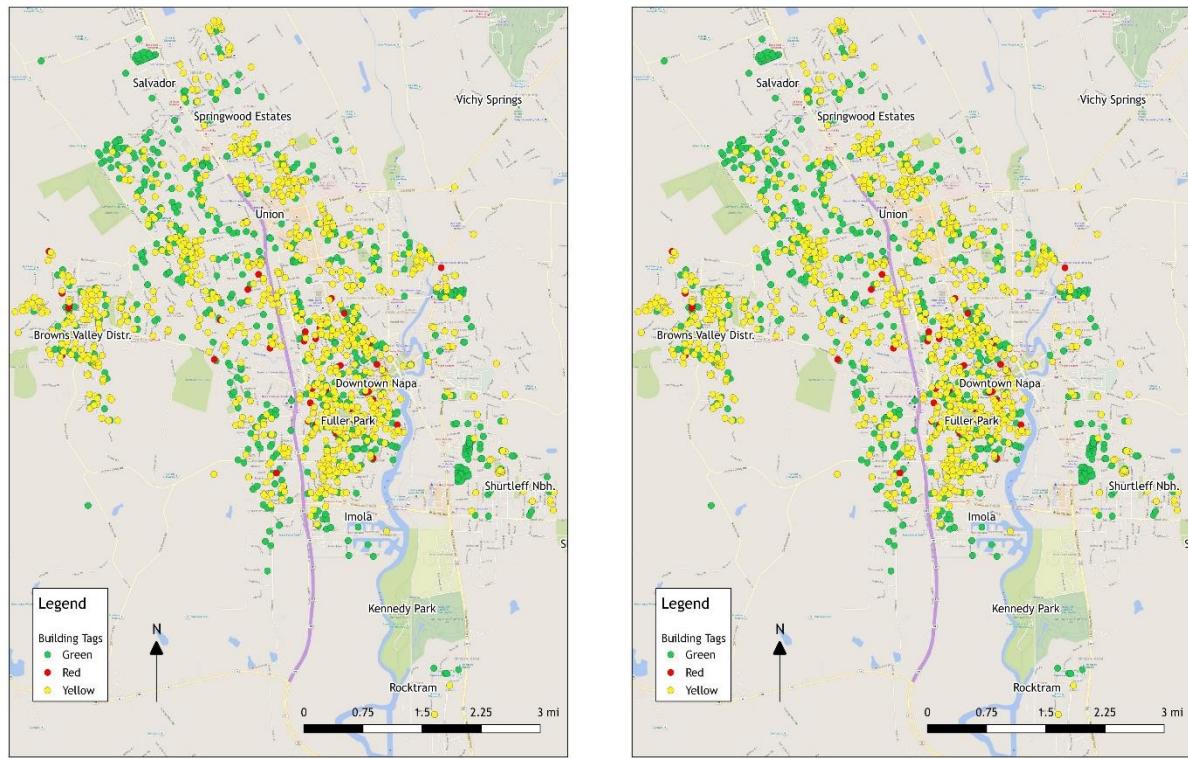


Fig. 12. Comparing the spatial distribution of a) actual and b) RF-based tags using training data.

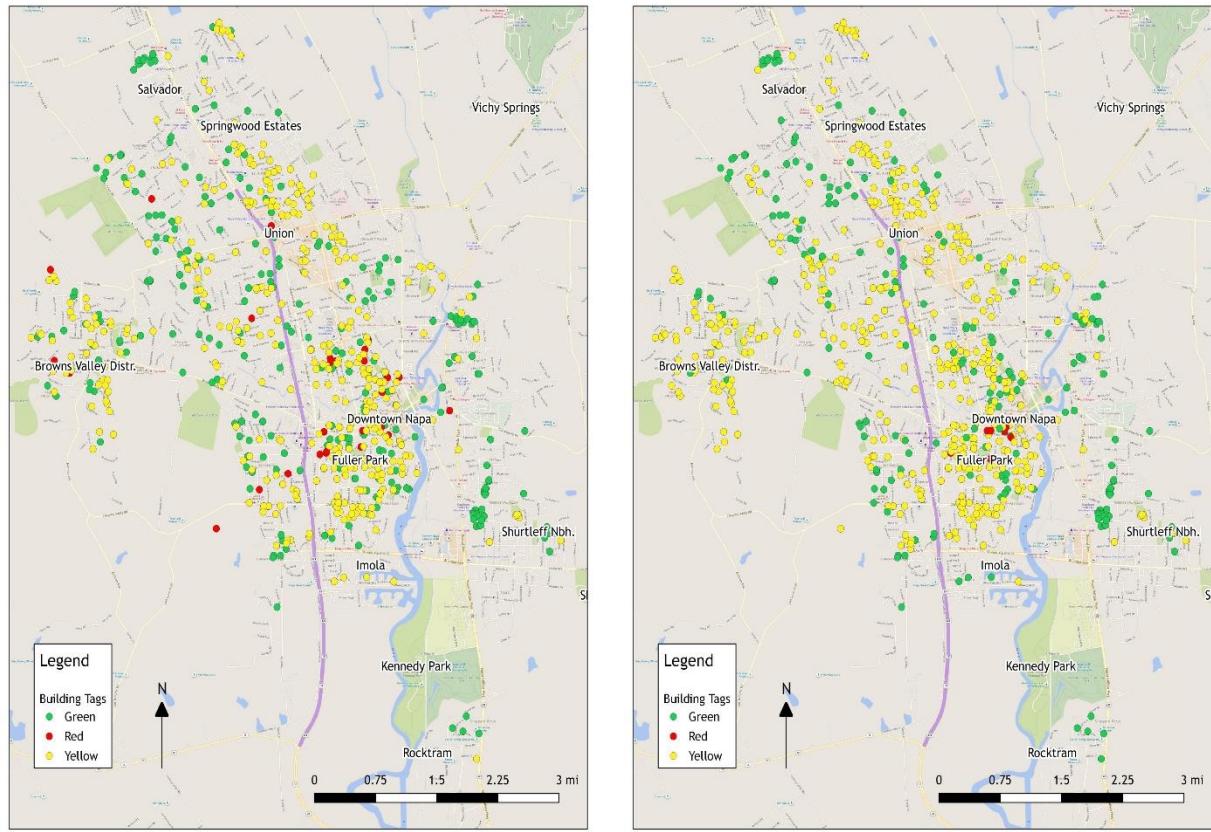


Fig. 13. Comparing the spatial distribution of a) actual and b) RF-based tags using testing data

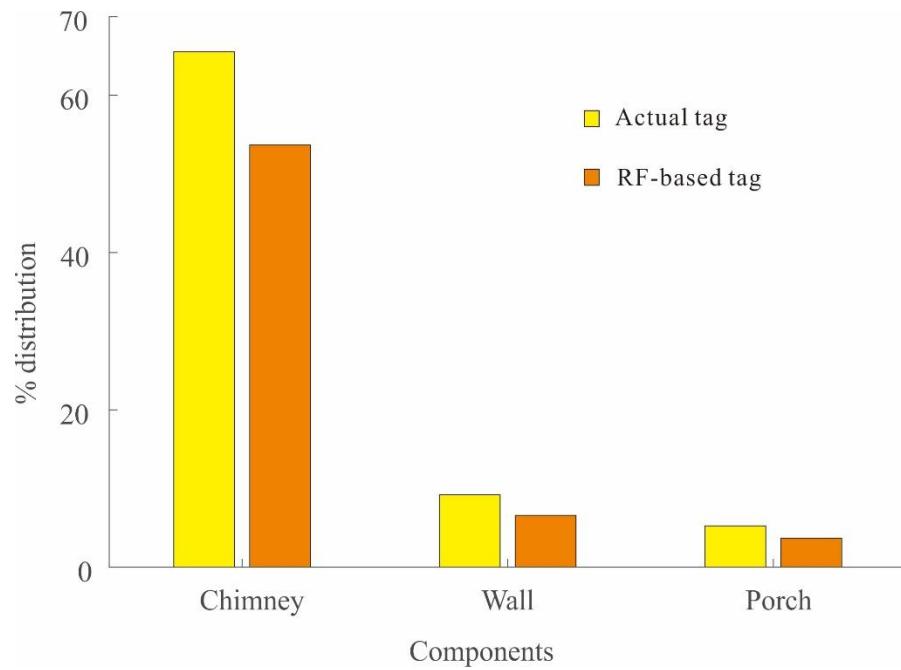


Fig. 14. Comparing the observed and RF-predictions for specific types of yellow tagged damage within the test set