

# Predicting Fire-Following-Earthquake Ignition Probabilities in California: Validation Across Earthquake Events

Guochuan Wei

wgcbjut@163.com

Beijing University of Technology <https://orcid.org/0009-0006-8987-2922>

wei Wang

Beijing University of Technology

Donghui Ma

Beijing University of Technology <https://orcid.org/0000-0001-8971-3223>

---

## Research Article

**Keywords:** Fire following earthquake, Ignition probability, Nested cross-validation, Event-held-out validation, Priority screening

**Posted Date:** October 1st, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-10974783/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Predicting Fire-Following-Earthquake Ignition Probabilities in California: Validation Across Earthquake Events

Guochuan Wei<sup>1</sup>, Wei Wang<sup>2</sup>, and Donghui Ma<sup>2</sup>

<sup>1</sup> College of Architecture and Civil Engineering, Beijing University of Technology, Beijing 100124, China

<sup>2</sup> College of Architecture and Urban Planning, Beijing University of Technology, Beijing 100124, China

Corresponding authors

Correspondence to Wei Wang (ieeww@bjut.edu.cn) or Donghui Ma (donghuima2022@163.com).

ORCID

Guochuan Wei: <https://orcid.org/0009-0006-8987-2922>

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 52278472).

## Abstract

Fire following earthquake (FFE) is a secondary earthquake hazard that can require rapid spatial prioritization when response resources are constrained. This study developed a census-tract-level logistic regression model for the probability of at least one recorded FFE ignition, using 7,074 earthquake-census-tract observations from seven California earthquakes between 1983 and 2014. A nested leave-one-event-out cross-validation design was used to select model structures in the inner loop and assess the entire procedure on an earthquake event excluded from fitting in the outer loop. Across the seven outer test events, event-mean area under the receiver operating characteristic curve and average precision were 0.833 and 0.267, respectively. The event-specific Top20% priority screening sets contained 71.9% of observations with at least one recorded ignition. When refitted on all historical observations and applied to the U.S. Geological Survey HayWired earthquake scenario, the model identified screening sets concentrated mainly in Alameda, Contra Costa, and Santa Clara counties; the reported scenario ignition estimates were also concentrated in these counties. The results characterize event-held-out ranking performance across the seven historical events and support scenario-based priority screening under the stated data and modeling assumptions.

**Keywords** Fire following earthquake · Ignition probability · Nested cross-validation · Event-held-out validation · Priority screening

## 1 Introduction

Fire following earthquake (FFE) is an important secondary hazard in urban seismic risk assessment (Lee et al. 2008; Lu et al. 2024; Wang et al. 2023). During the 1906 San Francisco earthquake, fires accounted for approximately 80% of the city's damage and burned about 28,000 buildings over 12.2 km<sup>2</sup> (Scawthorn 1986). More than 200 fires occurred following the 1995 Great Hanshin earthquake, burning over 7,000 houses across more than 0.65 km<sup>2</sup> (Hokugo 1995). Following the 2024 Noto Peninsula earthquake, 17 fires were recorded, and the large fire in central Wajima burned approximately 240 buildings (Nishino 2025). Multiple simultaneous ignitions may coincide with damage to water-supply systems (Liang 2003; Yan et al. 2026), disrupted road access (Wang et al. 2024), and limited firefighting resources. These combined constraints can reduce the capacity for initial suppression. Individual fires may then spread and cause losses over a wider area (Sarreshtehdari and Elhami Khorasani 2021; Wang et al. 2026; Yoshioka et al. 2020). Estimating

ignition probabilities across the earthquake-affected area and identifying priority screening sets are therefore important for urban FFE risk reduction and emergency response.

Early FFE ignition models commonly estimated regional ignition counts using empirical relationships between ground-motion intensity and ignition rates per unit building floor area. The Hazus model (FEMA 1999) and the models proposed by Ren and Xie (2004) and Zhao et al. (2006) followed this general framework. Subsequent studies used geographic information systems to integrate ground-motion data, building inventories, and population data and develop multivariable statistical models at the census-tract level. Davidson (2009) used census-tract ignition counts as the response variable in generalized linear models and generalized linear mixed models. Elhami Khorasani et al. (2017) instead used a binary response indicating whether at least one ignition was recorded in a census tract. Using data from seven California earthquakes, Tong and Gernay (2022) applied a hierarchical Bayesian model to examine covariates associated with historical ignition counts and predict counts for future earthquake scenarios. Thus, FFE ignition research has expanded from regional count estimates derived from ignition rates per unit floor area to multivariable estimates of census-tract ignition counts and probabilities.

However, predictive performance for earthquakes excluded from model development remains insufficiently tested. Specifically, it remains unclear whether census tracts with at least one recorded ignition appear near the top of within-event rankings by predicted ignition probability. Previous multi-earthquake studies did not use all census-tract observations from one earthquake as an independent test set (Davidson 2009; Elhami Khorasani et al. 2017; Tong and Gernay 2022). When observations from the same earthquake are divided between training and test sets, that earthquake contributes data to both model fitting and testing. Performance estimates from such partitions do not directly characterize prediction for a complete held-out earthquake event (Roberts et al. 2017). Performance may also be overestimated when candidate model structures are selected and assessed using the same cross-validation results (Cawley and Talbot 2010; Krstajic et al. 2014; Varma and Simon 2006). A suitable validation design should therefore assign all census-tract observations from an earthquake to the same fold (Nowicki Jessee et al. 2018). Nested cross-validation then separates model selection in the inner loop from model assessment in the outer loop (Cawley and Talbot 2010; Krstajic et al. 2014; Varma and Simon 2006).

Accordingly, this study examines seven historical California earthquakes that occurred between 1983 and 2014. It evaluates the full modeling procedure, including model structure selection and regression coefficient estimation, on complete held-out earthquake events. Assessment focuses on within-event rankings by predicted ignition probability and calibration of outer-test predicted probabilities. It also compares performance estimates obtained using alternative ground-motion intensity measures and sample-splitting schemes. The final predictive model, selected and fitted using all historical observations, is then applied to the HayWired earthquake scenario to estimate census-tract ignition probabilities and identify Top-k% priority screening sets. The county-level distribution of these sets is compared with that of the ignition estimates reported for the HayWired earthquake scenario (Scawthorn 2018).

## 2 Data and Candidate Explanatory Variables

The study workflow is shown in Fig. 1 and comprises four components: (1) construction of the earthquake-census-tract database and specification of candidate model structures; (2) nested cross-validation with leave-one-event-out (LOEO) splitting in the inner and outer loops; (3) selection of the final model structure and sensitivity analyses; and (4) application to the HayWired earthquake scenario.

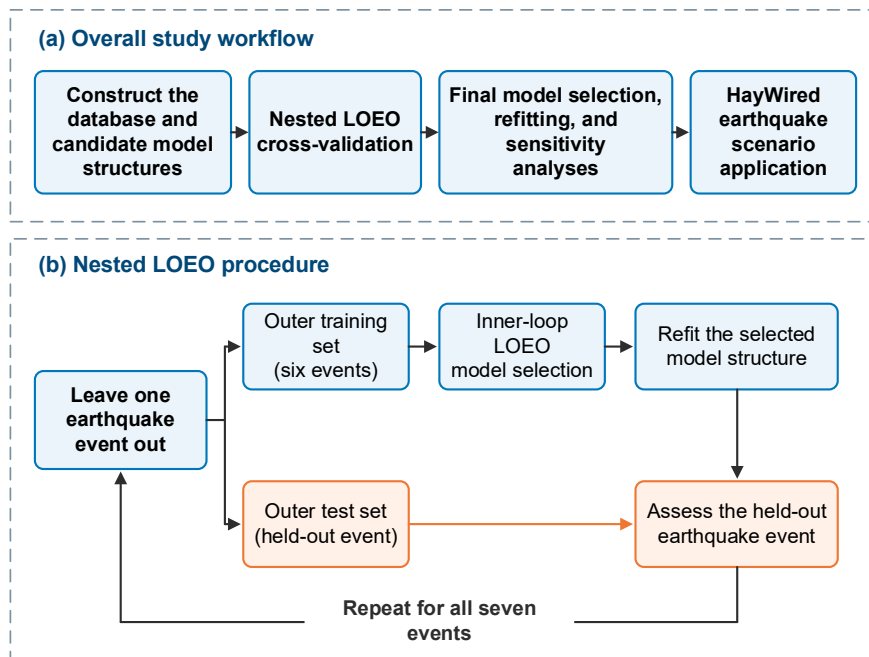


**Fig. 1** Overall study workflow and nested LOEO procedure

## 2.1 FFE ignition database

The FFE ignition database was compiled using California earthquake data assembled by Davidson (2009) and Elhami Khorasani et al. (2017). It contains ignition records from seven historical California earthquakes that occurred between 1983 and 2014. Eligible records were limited to ignitions associated with the mainshock that developed into structural fires, occurred within 10 days after the earthquake, and required a fire department response (Davidson 2009; Elhami Khorasani et al. 2017; Tong and Gernay 2022).

The study used 2000 Census tracts as the common spatial unit of analysis. Each historical ignition was assigned to the corresponding census tract based on its recorded location, and each earthquake-census-tract combination constituted one observation. Following the 0.08-g cutoff adopted by Elhami Khorasani et al. (2017), only census tracts with a census-tract area-weighted mean peak ground acceleration (PGA) of at least 0.08 g were included for each earthquake. When multiple ignitions were recorded in the same census tract during the same earthquake, they were consolidated into a single observation while retaining the ignition count. The remaining included census tracts were classified as census tracts with no recorded ignition. Table 1 summarizes the observations and ignitions in the database by earthquake.

As shown in Table 1, the database comprises 7,074 earthquake-census-tract observations and 155 recorded ignitions. Of these observations, 135 contained at least one recorded ignition, whereas 6,939 contained no recorded ignition. Observations with at least one recorded ignition therefore represented 1.91% of the database, resulting in an imbalanced binary response. The numbers of census-tract observations and recorded ignitions, as well as the proportion of observations with at least one recorded ignition, varied among the seven earthquakes. The 1994 Northridge earthquake contributed 2,696 observations and 82 recorded ignitions, accounting for 38.1% of all observations and 52.9% of all recorded ignitions, respectively. Although the 2014 Napa earthquake contributed only 74 observations, five contained at least one recorded ignition. The corresponding proportion, 6.76%, was the highest among the seven earthquakes.

**Table 1** Sample composition of the FFE ignition database for historical California earthquakes

| Earthquake            | Census-tract observations | Observations with at least one recorded ignition | Recorded ignitions | Proportion with at least one recorded ignition (%) | Magnitude |
|-----------------------|---------------------------|--------------------------------------------------|--------------------|----------------------------------------------------|-----------|
| 1983 Coalinga         | 44                        | 1                                                | 3                  | 2.27                                               | 6.7       |
| 1987 Whittier Narrows | 2202                      | 20                                               | 20                 | 0.91                                               | 5.9       |

| Earthquake          | Census-tract observations | Observations with at least one recorded ignition | Recorded ignitions | Proportion with at least one recorded ignition (%) | Magnitude |
|---------------------|---------------------------|--------------------------------------------------|--------------------|----------------------------------------------------|-----------|
| 1989 Loma Prieta    | 1399                      | 31                                               | 36                 | 2.22                                               | 6.9       |
| 1984 Morgan Hill    | 487                       | 6                                                | 6                  | 1.23                                               | 6.2       |
| 1986 N. Palm Spring | 172                       | 1                                                | 1                  | 0.58                                               | 6.0       |
| 1994 Northridge     | 2696                      | 71                                               | 82                 | 2.63                                               | 6.7       |
| 2014 Napa           | 74                        | 5                                                | 7                  | 6.76                                               | 6.0       |
| Total               | 7074                      | 135                                              | 155                | 1.91                                               | -         |

109 **2.2 Candidate explanatory variables**

110 Using the FFE ignition database described above, we compiled a census-tract-level database of candidate  
111 explanatory variables. Candidate variables were required to be relevant to FFE ignition and consistently  
112 available at the census-tract level. For this study, the candidate explanatory variables were organized into  
113 five categories: ground-motion intensity measures, exposure variables, proxy variables for building  
114 vulnerability, proxy variables related to potential ignition sources, and environmental variables. Table 2 lists  
115 the candidate explanatory variables and their respective categories.

116 **Table 2** Candidate explanatory variables and their categories

| Variable description                             | Variable code | Ground-motion intensity measure | Exposure variable | Proxy variable for building vulnerability | Proxy variable related to potential ignition sources | Environmental variable |
|--------------------------------------------------|---------------|---------------------------------|-------------------|-------------------------------------------|------------------------------------------------------|------------------------|
| Peak ground acceleration                         | PGA           | •                               |                   |                                           |                                                      |                        |
| Peak ground velocity                             | PGV           | •                               |                   |                                           |                                                      |                        |
| ShakeMap instrumental intensity                  | MMI           | •                               |                   |                                           |                                                      |                        |
| Total building floor area                        | TFA           |                                 | •                 |                                           |                                                      |                        |
| Population density                               | PD            |                                 | •                 |                                           |                                                      |                        |
| Pre-1970 housing-unit share                      | Pre70         |                                 |                   | •                                         |                                                      |                        |
| Wood-building floor-area share                   | WoodSh        |                                 |                   | •                                         |                                                      |                        |
| Unreinforced masonry building floor-area share   | URMSh         |                                 |                   | •                                         |                                                      |                        |
| Mobile-home building floor-area share            | MHSh          |                                 |                   | •                                         |                                                      |                        |
| Utility-gas-heated occupied housing unit density | GasHD         |                                 |                   |                                           | •                                                    |                        |
| Utility-gas-heated occupied housing unit share   | GasHSh        |                                 |                   |                                           | •                                                    |                        |

| Variable description                         | Variable code | Ground-motion intensity measure | Exposure variable | Proxy variable for building vulnerability | Proxy variable related to potential ignition sources | Environmental variable |
|----------------------------------------------|---------------|---------------------------------|-------------------|-------------------------------------------|------------------------------------------------------|------------------------|
| Commercial building floor-area share         | ComSh         |                                 |                   |                                           | •                                                    |                        |
| Industrial building floor-area share         | IndSh         |                                 |                   |                                           | •                                                    |                        |
| Residential building floor-area share        | ResSh         |                                 |                   |                                           | •                                                    |                        |
| Mean air temperature in the earthquake month | Temp          |                                 |                   |                                           |                                                      | •                      |

Ground motion is an important factor in FFE ignition (He and Elhami Khorasani 2022). Stronger shaking is more likely to displace building contents (Dhakal et al. 2016) and to damage gas and electrical connections or overturn equipment such as water heaters, thereby increasing FFE ignition probability (Elhami Khorasani et al. 2017; Tong and Gernay 2022). PGA, peak ground velocity (PGV), and ShakeMap instrumental intensity (MMI) were selected as candidate ground-motion intensity measures. All three measures were obtained from U.S. Geological Survey (USGS) ShakeMap grid.xml files (Worden et al. 2020). For each measure, the census-tract area-weighted mean was calculated using the area of intersection between each ShakeMap grid cell and census tract as the weight.

For census tracts with the same total building floor area, higher population density may correspond to more potential ignition sources (Elhami Khorasani et al. 2017). Total building floor area (TFA) and population density (PD) were therefore selected as exposure variables. TFA was obtained from the HAZUS-MH MR4 default building inventory and was used to represent the size of the building stock within each census tract. PD was calculated from Census 2000 population counts and census-tract land area and was used to represent population concentration within the tract.

Building age and structural type affect the extent of earthquake-induced building damage (Ismail and Khattak 2019; Yazgan et al. 2016). Structural damage may also cause damage to nonstructural systems and lead to ignition (Tong and Gernay 2022). The pre-1970 housing-unit share (Pre70), wood-building floor-area share (WoodSh), unreinforced masonry building floor-area share (URMSh), and mobile-home building floor-area share (MHSh) were therefore selected as proxy variables for building vulnerability (Davidson 2009). Pre70 was obtained from Census 2000 and was used to characterize the age composition of the housing stock. WoodSh, URMSh, and MHSh were obtained from the HAZUS-MH MR4 default building inventory and were used to characterize the structural-type composition of the building stock within each census tract.

Building occupancy affects the types of equipment and contents present within buildings and, consequently, the types of potential ignition sources (Tong and Gernay 2022). Gas leakage may also contribute to FFE ignition (Elhami Khorasani et al. 2017; Farahani et al. 2020). Five variables were selected as proxy variables related to potential ignition sources: the residential building floor-area share (ResSh), commercial building floor-area share (ComSh), industrial building floor-area share (IndSh), utility-gas-heated occupied housing unit density (GasHD), and utility-gas-heated occupied housing unit share (GasHSh). ResSh, ComSh, and IndSh were obtained from the HAZUS-MH MR4 default building inventory. GasHD and GasHSh were calculated from Census 2000 data on occupied housing units using utility gas as their principal house-heating fuel.

Mean air temperature in the earthquake month (Temp) was selected as the environmental variable. Because electric and gas appliances are more frequently used for heating in winter than in summer, Temp was used to represent seasonal differences among the earthquake events (Tong and Gernay 2022). Temperature data were obtained from Parameter-elevation Regressions on Independent Slopes Model monthly mean air-temperature grids (Daly et al. 2008). For each earthquake, the grid corresponding to the year and month of occurrence was selected, and the mean air temperature within each census tract was calculated.

### 3 Research Methodology

#### 3.1 Logistic regression model

A census-tract-level model of FFE ignition probability was developed using logistic regression. The model uses a binary response, and the logistic function maps a linear combination of explanatory variables to a conditional probability in the interval (0, 1) (Elhami Khorasani et al. 2017; Hosmer et al. 2013).

Let  $y_{ei}$  denote the binary ignition response for census tract  $i$  in earthquake event  $e$ :

$$y_{ei} = \begin{cases} 1, & \text{if at least one ignition was recorded,} \\ 0, & \text{if no ignition was recorded.} \end{cases} \quad (1)$$

Let  $X_{ei} = (X_{ei1}, X_{ei2}, \dots, X_{eiq})$  denote the explanatory-variable vector, where  $X_{eij}$  is the  $j$ th explanatory variable for census tract  $i$  in earthquake event  $e$ , and  $q$  is the number of explanatory variables. The corresponding conditional probability is

$$P_{ei} = P(y_{ei} = 1 | X_{ei}) \quad (2)$$

where  $P_{ei}$  denotes the conditional probability that at least one ignition is recorded in census tract  $i$  during earthquake event  $e$ , given  $X_{ei}$ . The logistic regression model is expressed as

$$P(y_{ei} = 1 | X_{ei1}, X_{ei2}, \dots, X_{eiq}) = \frac{1}{1 + \exp \left[ - \left( \beta_0 + \sum_{j=1}^q \beta_j X_{eij} \right) \right]} \quad (3)$$

Here,  $\beta_0$  is the intercept, and  $\beta_j$  is the regression coefficient associated with the  $j$ th explanatory variable, for  $j = 1, \dots, q$ . The regression coefficients were estimated using maximum likelihood estimation.

#### 3.2 Variable transformations and candidate model structures

PGA and PGV were right-skewed in the study dataset and were therefore transformed using the natural logarithm, yielding  $\ln(\text{PGA})$  and  $\ln(\text{PGV})$ , respectively. TFA and PD also exhibited right-skewed distributions and wide ranges of values. They were transformed as  $\ln(1+\text{TFA})$  and  $\ln(1+\text{PD})$ , respectively, so that zero values remained defined after transformation. GasHD was similarly right-skewed, with many observations concentrated near zero, and was therefore square-root transformed and denoted by  $\sqrt{\text{GasHD}}$ . No nonlinear transformation was applied to the remaining explanatory variables.

Existing census-tract-level FFE models include ground-motion intensity and total building floor area as explanatory variables and contain three to five explanatory variables (Davidson 2009; Elhami Khorasani et al. 2017; Tong and Gernay 2022). In this study, MMI and  $\ln(1+\text{TFA})$  formed the base model. The remaining 11 non-ground-motion candidate explanatory variables were treated as optional explanatory variables. To limit the complexity of each candidate model structure, no more than three optional explanatory variables were added to the base model. The candidate explanatory variables were combined according to their definitions and categories under the rules listed in Table 3, resulting in 211 candidate model structures.

**Table 3** Variable-combination rules for candidate model structures

| Variable group                             | Combination rule                   | Rationale                                                                         |
|--------------------------------------------|------------------------------------|-----------------------------------------------------------------------------------|
| $\ln(1+\text{PD})$ , $\sqrt{\text{GasHD}}$ | Not included in the same structure | Both are densities based on census-tract land area                                |
| $\sqrt{\text{GasHD}}$ , GasHSh             | Not included in the same structure | Both are derived from the same count of utility-gas-heated occupied housing units |
| WoodSh, URMSH, MHSh                        | At most two per structure          | All are building floor-area shares by structural type                             |
| ComSh, IndSh, ResSh                        | At most two per structure          | All are building floor-area shares by occupancy                                   |

### 3.3 Nested cross-validation with leave-one-event-out splitting

When the same cross-validation results are used both to select a model and to assess its predictive performance, the resulting performance estimates may be optimistically biased (Cawley and Talbot 2010; Krstajic et al. 2014; Varma and Simon 2006). This study therefore used nested cross-validation, in which the inner loop selected the model structure and the outer loop assessed the predictive performance of the full modeling procedure on held-out earthquake events. LOEO splitting was applied in both loops. Consistent with the event-wise holdout logic of Nowicki Jesse et al. (2018), each split held out one earthquake event in its entirety. All earthquake-census-tract observations associated with that event formed the inner validation set or the outer test set, as applicable. For each inner and outer split, the means and standard deviations used to standardize the explanatory variables were estimated exclusively from the corresponding training set and applied unchanged to the associated validation or test set. The procedure comprised the following five steps.

1. Outer-loop sample splitting. The outer loop used LOEO splitting, yielding seven outer folds. In each outer fold, one entire earthquake event was held out, with all its census-tract observations forming the outer test set. Observations from the remaining six earthquake events formed the outer training set. Model selection and regression coefficient estimation were performed entirely within the outer training set, whereas the outer test set was used only for model assessment in that fold.
2. Inner-loop sample splitting. Each outer training set contained six earthquake events and was further partitioned using LOEO splitting, yielding six inner folds. In each inner fold, all observations from one earthquake event formed the inner validation set, whereas observations from the remaining five events formed the inner training set.
3. Inner-loop model selection. In each inner fold, every candidate model structure was fitted using the inner training set, and the evaluation metrics were calculated from predictions for the inner validation set. After all six inner folds had been completed, the evaluation results for each candidate model structure were aggregated across the six inner validation events. The model-selection criterion was then applied to determine the model structure selected for that outer fold. The evaluation metrics and model-selection criterion are defined in Sect. 3.4.
4. Outer-loop model fitting and assessment. After inner-loop model selection, the regression coefficients of the selected model structure were re-estimated using the entire outer training set. Predicted ignition probabilities and evaluation metrics were then calculated for the census tracts in the outer test set. The metrics obtained from the seven outer folds constituted the outer-loop assessment results of the nested cross-validation.
5. Determination of the final model structure and fitting of the final predictive model. After completion of the outer-loop assessment, LOEO cross-validation was applied to all seven historical earthquakes. The evaluation results for each candidate model structure were aggregated across the seven validation events, and the same model-selection criterion was used to determine the final model structure. Its regression coefficients were then re-estimated using all observations, yielding the final predictive model used for mapping census-tract ignition probabilities and for earthquake-scenario applications (Lewis et al. 2023; Tsamardinos 2022).

### 3.4 Model assessment and selection

#### 3.4.1 Probability-ranking metrics and model-selection criterion

Both the inner-loop comparison of candidate model structures and the outer-loop model assessment used each complete held-out earthquake event as the unit of evaluation. Within each held-out event, census tracts were ranked in descending order of predicted ignition probability, and the area under the receiver operating characteristic curve (AUROC), average precision (AP), and Recall@Top-k% were calculated. Each metric was first calculated within the held-out event and then averaged across events with equal weight; the resulting value is hereafter referred to as the event mean.

AUROC was used to assess the model's overall ranking discrimination between census tracts with at least one recorded ignition and census tracts with no recorded ignition within a single held-out event (Saito and Rehmsmeier 2015). AP was calculated as the arithmetic mean of the precision values at the ranks occupied by census tracts with at least one recorded ignition (Manning et al. 2008). Recall@Top-k% was defined as the proportion of census tracts with at least one recorded ignition in an event that fell within the top k% of census tracts ranked by predicted ignition probability. The number of census tracts corresponding to each cutoff was rounded up to the next integer. The Top5%, Top10%, and Top20% cutoffs were specified a priori.

For each candidate model structure, these metrics were calculated separately for the six inner validation events, after which the event mean of each metric was obtained. Model selection considered both the overall ranking of census tracts with at least one recorded ignition and their recall at different screening proportions. Event-mean AP and the mean of event-mean Recall@Top5%, Recall@Top10%, and Recall@Top20% were each assigned a weight of 50%, and the candidate model structure with the highest composite score was selected. In the event of a tie, structures were compared sequentially by higher event-mean AP, higher event-mean AUROC, and fewer explanatory variables.

The evaluation metrics were calculated separately for the outer test event in each of the seven outer folds. The event mean of each metric was used to assess the performance of the full modeling procedure in ranking census tracts by predicted ignition probability for earthquake events held out in their entirety.

### 3.4.2 Calibration assessment of outer-test predicted probabilities

To assess the agreement between the outer-test predicted probabilities and the binary ignition responses, the outer-test predictions from the seven folds of nested cross-validation were pooled for calibration assessment. Because the Coalinga and N. Palm Spring earthquakes each contained only one census tract with at least one recorded ignition, calibration measures were not estimated separately for individual events. Instead, calibration was assessed using all pooled outer-test observations.

The calibration intercept and calibration slope were jointly estimated by fitting a logistic regression model with the binary ignition response as the dependent variable and the logit-transformed predicted ignition probability as the explanatory variable. Their ideal values are 0 and 1, respectively. A calibration curve was also constructed by applying local regression (loess) smoothing to the binary ignition response as a function of the predicted ignition probability. The 45° reference line represents perfect calibration (Austin and Steyerberg 2014; Van Calster et al. 2019).

### 3.5 Sensitivity analyses

After the final model structure was determined, two sensitivity analyses examined changes in rankings by predicted ignition probability and the corresponding evaluation metrics under alternative ground-motion intensity measures and sample-splitting schemes.

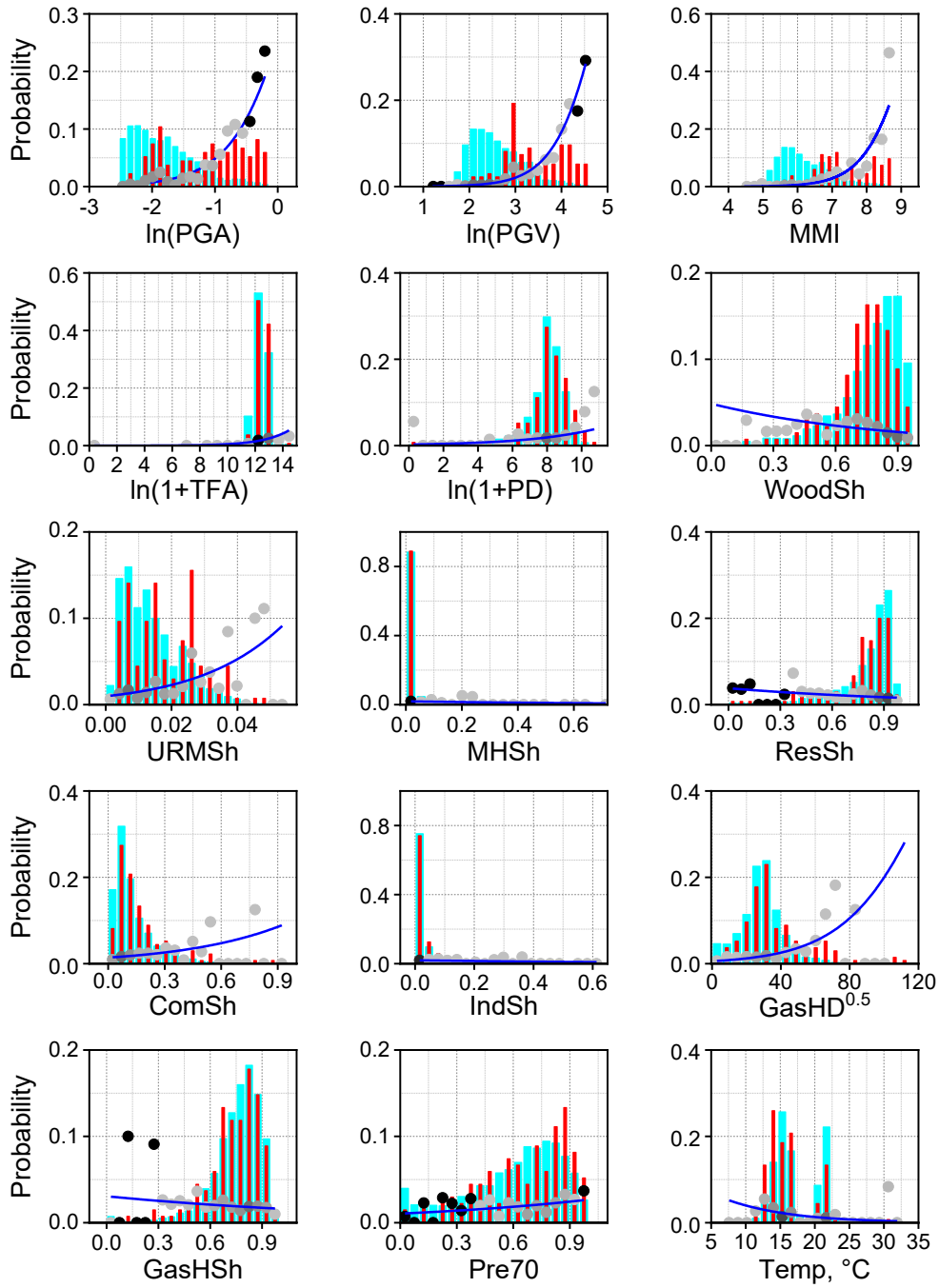
In the ground-motion intensity-measure sensitivity analysis, the non-ground-motion explanatory variables in the final model structure were held fixed. MMI, ln(PGA), and ln(PGV) were used separately to represent ground-motion intensity. The three models used the same seven LOEO splits and evaluation metrics. Their rankings by predicted ignition probability were compared within each held-out earthquake event.

The sample-splitting sensitivity analysis used the final model structure and repeated stratified five-fold cross-validation by earthquake event and response class. In each repeat, earthquake-census-tract observations were randomly allocated among five folds within each earthquake event and response class, with fold sizes balanced as closely as possible. For each cross-validation iteration, four folds from each of the seven earthquake events were pooled to form the training set, and the remaining fold from each event formed the test set. The training and test sets therefore contained approximately 80% and 20%, respectively, of the observations from each event. After predictions from the five test folds were combined, AUROC, AP, and Recall@Top-k% were calculated separately for each earthquake event from its complete set of out-of-fold predictions. The event mean of each metric was then calculated for that repeat. The five-fold partitioning and evaluation procedure was repeated independently 100 times. For each metric, the resulting event means were compared with the corresponding event mean obtained from fixed-structure LOEO.

## 4 Results

### 4.1 Distributions and correlations of candidate explanatory variables

To describe the univariable relationships between the candidate explanatory variables and the binary ignition response, Fig. 2 presents the variable distributions together with binned observed ignition proportions and predicted-probability curves.



■ Non-ignition    ■ Ignition    — Pred, probability    ● Obs, probability

**Fig. 2** Distributions of candidate explanatory variables in census tracts with at least one recorded ignition and census tracts with no recorded ignition. Gray points show binned observed ignition proportions, darker points denote bins with more observations, curves show predictions from univariable logistic regression

For  $\ln(\text{PGA})$ ,  $\ln(\text{PGV})$ , and  $\text{MMI}$ , the binned observed ignition proportions generally increased with ground-motion intensity, and the corresponding predicted-probability curves from univariable logistic regression showed upward trends (Fig. 2).  $\text{MMI}$  values were generally higher among census tracts with at least one recorded ignition than among those with no recorded ignition. Among the non-ground-motion explanatory variables, the univariable predicted-probability curves showed upward trends for  $\ln(1+\text{TFA})$ ,  $\ln(1+\text{PD})$ ,  $\text{URMSH}$ ,  $\sqrt{\text{GasHD}}$ ,  $\text{ComSh}$ , and  $\text{Pre70}$ ; downward trends for  $\text{WoodSh}$ ,  $\text{ResSh}$ , and  $\text{Temp}$ ; and comparatively little variation for  $\text{MHS}$ ,  $\text{GasHSh}$ , and  $\text{IndSh}$ . Some bins contained few observations, and the corresponding observed ignition proportions were more variable.

Fig. 3 presents the Spearman rank correlation coefficients for  $\text{MMI}$ ,  $\ln(1+\text{TFA})$ , and the 11 optional explanatory variables. The largest absolute correlations were mainly found between density variables,

between variables describing building structural-type composition, and between variables describing building-occupancy composition. The coefficients were  $-0.934$  between ComSh and ResSh,  $0.915$  between  $\ln(1+PD)$  and  $\sqrt{\text{GasHD}}$ , and  $-0.875$  between WoodSh and URMSH. These correlations were consistent with the variable-combination rules specified in Sect. 3.2 on the basis of the variable definitions.

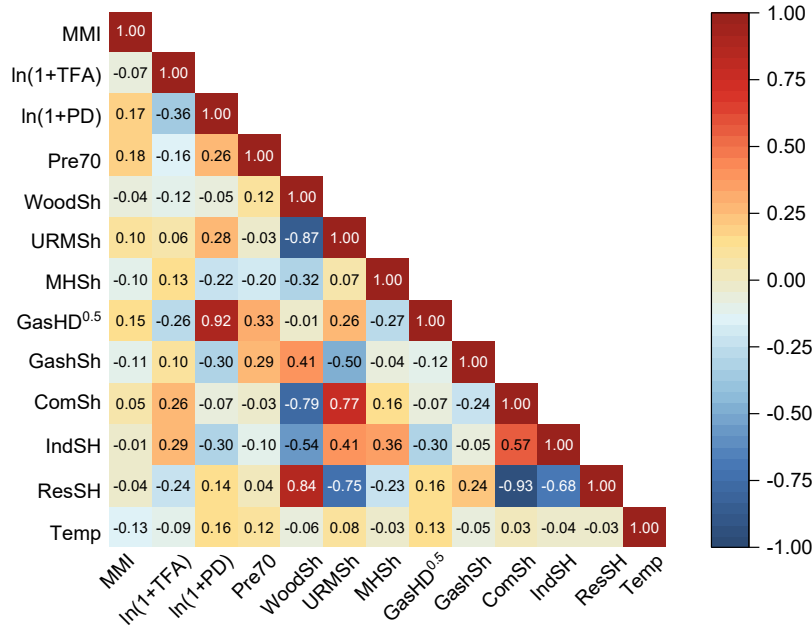


**Fig. 3** Spearman rank correlation matrix

## 4.2 Candidate model structure selection and assessment of the full modeling procedure

### 4.2.1 Model structures selected in the inner loops and outer-loop assessment results

Following the procedure described in Sect. 3.3, Table 4 reports the model structure selected by inner-loop cross-validation for each outer fold and the corresponding outer-loop assessment results. Across the seven outer folds, five distinct model structures were selected. Two structures were each selected in two outer folds, whereas the remaining three were each selected once. The selected structures differed across outer folds, indicating that model structure selection varied with the composition of earthquake events in the outer training set.

**Table 4** Model structures selected in the inner loops and outer-loop assessment results

| Outer test event    | Model structure selected in the inner loop                | Inner-loop event mean for AP | Mean of three inner-loop event-mean Recall@Top-k% values | AUROC | AP    | Recall@Top5% | Recall@Top10% | Recall@Top20% |
|---------------------|-----------------------------------------------------------|------------------------------|----------------------------------------------------------|-------|-------|--------------|---------------|---------------|
| 1983 Coalinga       | MMI + $\ln(1+TFA)$ + Pre70 + $\sqrt{\text{GasHD}}$        | 0.154                        | 0.466                                                    | 1.000 | 1.000 | 1.000        | 1.000         | 1.000         |
| 1984 Morgan Hill    | MMI + $\ln(1+TFA)$ + Pre70 + URMSH                        | 0.292                        | 0.581                                                    | 0.660 | 0.048 | 0.167        | 0.333         | 0.500         |
| 1986 N. Palm Spring | MMI + $\ln(1+TFA)$ + Pre70 + $\sqrt{\text{GasHD}}$ + Temp | 0.315                        | 0.579                                                    | 0.813 | 0.030 | 0.000        | 0.000         | 1.000         |

| Outer test event      | Model structure selected in the inner loop                                    | Inner-loop event mean for AP | Mean of three inner-loop event-mean Recall@Top-k% values | AUROC | AP    | Recall@Top5% | Recall@Top10% | Recall@Top20% |
|-----------------------|-------------------------------------------------------------------------------|------------------------------|----------------------------------------------------------|-------|-------|--------------|---------------|---------------|
| 1983 Coalinga         | MMI + $\ln(1+\text{TFA}) + \text{Pre70} + \sqrt{\text{GasHD}}$                | 0.154                        | 0.466                                                    | 1.000 | 1.000 | 1.000        | 1.000         | 1.000         |
| 1984 Morgan Hill      | MMI + $\ln(1+\text{TFA}) + \text{Pre70} + \text{URMSh}$                       | 0.292                        | 0.581                                                    | 0.660 | 0.048 | 0.167        | 0.333         | 0.500         |
| 1986 N. Palm Spring   | MMI + $\ln(1+\text{TFA}) + \text{Pre70} + \sqrt{\text{GasHD}} + \text{Temp}$  | 0.315                        | 0.579                                                    | 0.813 | 0.030 | 0.000        | 0.000         | 1.000         |
| 1987 Whittier Narrows | MMI + $\ln(1+\text{TFA}) + \text{Pre70} + \sqrt{\text{GasHD}}$                | 0.308                        | 0.557                                                    | 0.797 | 0.037 | 0.150        | 0.350         | 0.650         |
| 1989 Loma Prieta      | MMI + $\ln(1+\text{TFA}) + \text{URMSh} + \sqrt{\text{GasHD}} + \text{IndSh}$ | 0.289                        | 0.499                                                    | 0.785 | 0.074 | 0.129        | 0.419         | 0.645         |
| 1994 Northridge       | MMI + $\ln(1+\text{TFA}) + \text{Pre70} + \text{URMSh} + \text{ResSh}$        | 0.283                        | 0.571                                                    | 0.885 | 0.191 | 0.324        | 0.577         | 0.775         |
| 2014 Napa             | MMI + $\ln(1+\text{TFA}) + \text{Pre70} + \sqrt{\text{GasHD}} + \text{Temp}$  | 0.237                        | 0.534                                                    | 0.890 | 0.492 | 0.400        | 0.600         | 0.800         |
| Mean                  | —                                                                             | 0.268                        | 0.541                                                    | 0.833 | 0.267 | 0.310        | 0.469         | 0.767         |

The outer-loop metrics reported in Table 4 reflect the predictive performance of the full modeling procedure, including model structure selection and regression coefficient estimation, for completely held-out earthquake events. The event means for AUROC and AP were 0.833 and 0.267, respectively. The event means for Recall@Top5%, Recall@Top10%, and Recall@Top20% were 0.310, 0.469, and 0.767, respectively. The AUROC values for Whittier Narrows and Loma Prieta were similar, at 0.797 and 0.785, respectively, whereas the AP for Loma Prieta was 0.074, approximately twice that for Whittier Narrows (0.037). The two events showed similar overall ranking discrimination, but census tracts with at least one recorded ignition were concentrated to different degrees near the top of the respective rankings.

Across the seven events, the event-specific Top20% priority screening sets contained 97 of the 135 earthquake-census-tract observations with at least one recorded ignition, corresponding to 71.9%. Coalinga and N. Palm Spring each had only one earthquake-census-tract observation with at least one recorded ignition; consequently, their AP and Recall@Top-k% values depended on the ranking position of that single observation. For the remaining five events, Recall@Top20% ranged from 0.500 to 0.800. In every outer test event, the Top20% priority screening set included at least half of the census tracts with at least one recorded ignition.

#### 4.2.2 Calibration results for outer-test predicted probabilities

After pooling the outer-test results from all seven outer folds, the mean predicted probability among the

outer-test observations was 0.0215, whereas the observed ignition proportion was 0.0191. The calibration intercept and calibration slope were  $-0.638$  and  $0.824$ , respectively.

The predicted probabilities for the outer-test observations were concentrated primarily in the low-probability range, with 95% of the observations having predicted probabilities no greater than 0.106. As shown in Fig. 4, the loess-smoothed calibration curve was close to the  $45^\circ$  line of perfect calibration in the low-probability range and gradually fell below the reference line thereafter.

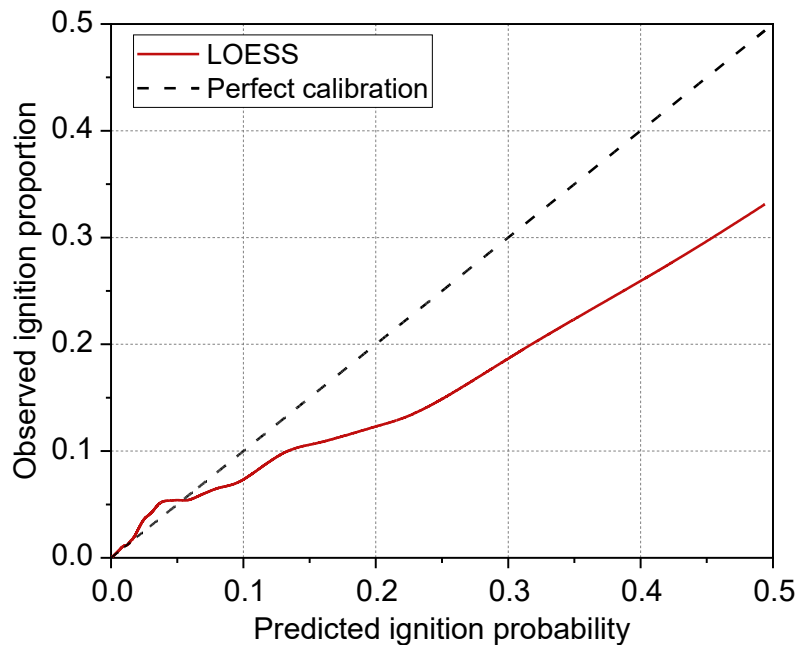


**Fig. 4** Calibration curve for outer-test predicted probabilities

The mean predicted probability was slightly higher than the observed ignition proportion, indicating slight overall overprediction. A calibration slope below 1 indicated that the predicted probabilities were too widely dispersed, such that higher predicted probabilities may have been too high and lower predicted probabilities may have been too low (Van Calster et al. 2019). Because relatively few observations fell in the higher-probability range, the loess curve was also more sensitive to the influence of a small number of observations (Austin and Steyerberg 2014).

#### 4.2.3 Rankings by predicted ignition probability for the outer test events

Fig. 5 shows the rankings by predicted ignition probability for the seven outer test events. The blue ranking curves generally declined steeply near the top of the rankings and gradually flattened thereafter. Relatively high predicted ignition probabilities were concentrated among a small number of top-ranked census tracts, whereas the predicted probabilities for most census tracts lay in a lower range. The probability ranges and rates of decline varied across events, indicating that the within-event distributions of predicted probabilities were not identical across the outer test events.

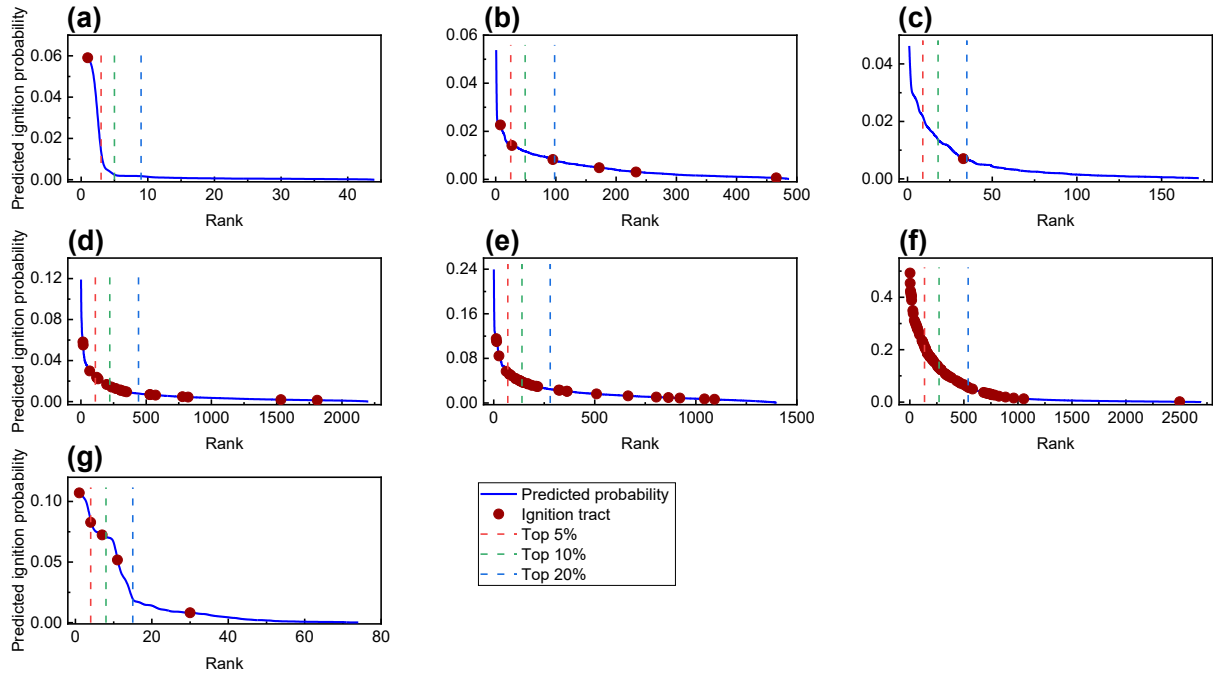


**Fig. 5** Rankings by predicted ignition probability for the seven outer test events. **a** 1983 Coalinga. **b** 1984 Morgan Hill. **c** 1986 N. Palm Spring. **d** 1987 Whittier Narrows. **e** 1989 Loma Prieta. **f** 1994 Northridge. **g** 2014 Napa. Blue curves show census tracts ordered by predicted probability; red points denote tracts with at least one recorded ignition; vertical dashed lines mark the Top5%, Top10%, and Top20% cutoffs

Overall, the red points were more densely distributed near the top of the rankings and became progressively sparser at lower ranking positions. However, red points extended beyond the Top20% cutoff in several events. Recall@Top20% was 0.775 for Northridge and 0.800 for Napa. In Northridge, 16 census tracts with at least one recorded ignition fell outside the Top20% priority screening set, with some extending into the lower-ranked portion of the ranking; in Napa, only one such tract fell outside this set. Thus, even when events had similar Recall@Top20% values, the distributions of census tracts with at least one recorded ignition across the predicted-probability rankings could differ markedly.

#### 4.2.4 Spatial distributions of Top-k% census tracts in the outer test events

Fig. 6 shows the spatial distributions of the Top-k% census tracts for the seven outer test events.

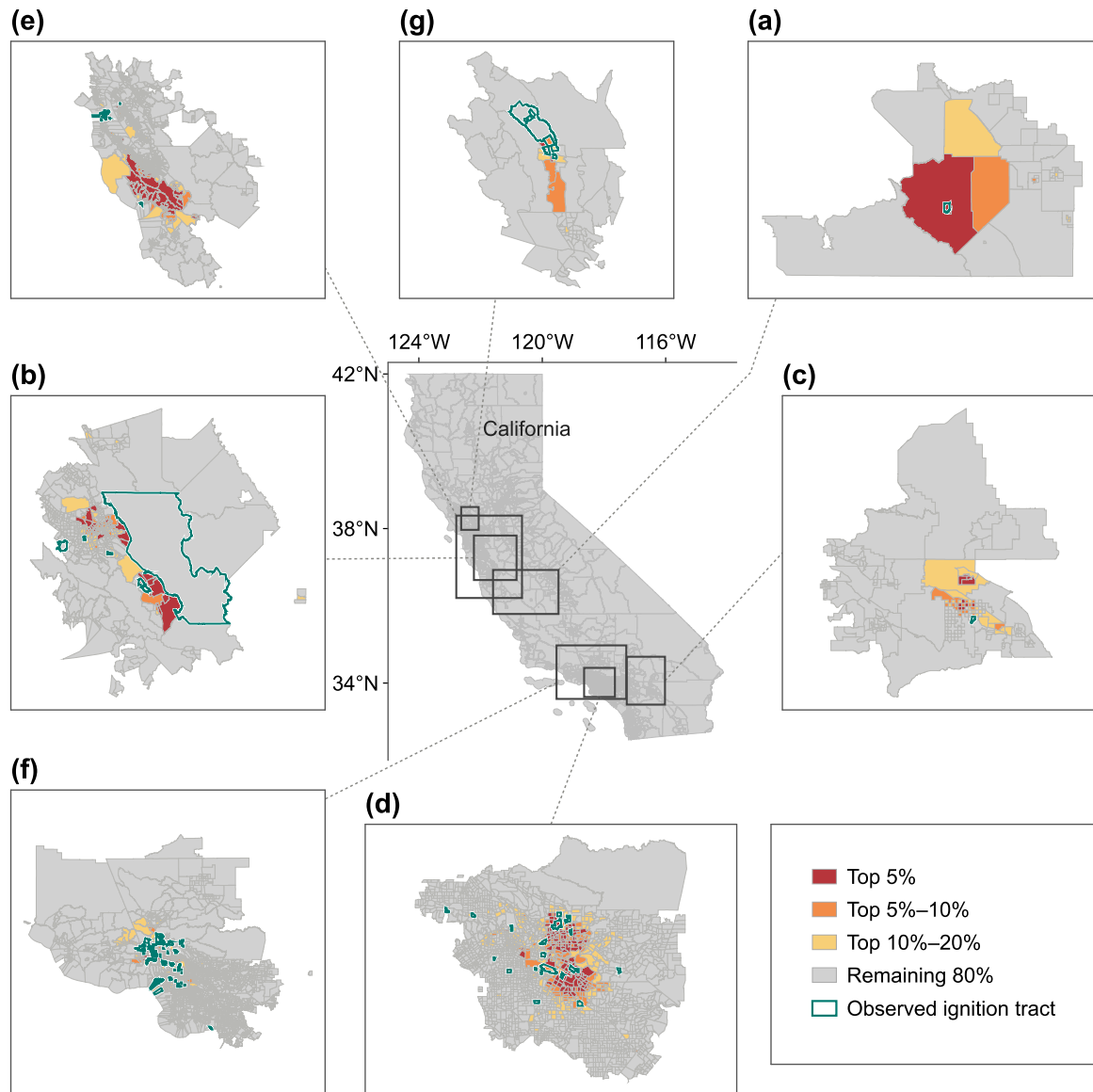


**Fig. 6** Spatial distributions of Top-k% census tracts for the seven outer test earthquake events. **a** 1983 Coalinga. **b** 1984 Morgan Hill. **c** 1986 N. Palm Spring. **d** 1987 Whittier Narrows. **e** 1989 Loma Prieta. **f** 1994 Northridge. **g** 2014 Napa. Colors show the Top5%, 5%–10%, 10%–20%, and remaining 80% ranking bands; green outlines denote tracts with at least one recorded ignition

The top-ranked census tracts were spatially contiguous in most events, forming local clusters or elongated zones. In some events, the top-ranked census tracts were distributed across several spatially separated areas. As the screening proportion increased from Top5% to Top20%, the newly included census tracts were generally located around the existing clusters.

### 4.3 Selection of the final model structure

After the seven outer-fold assessments were completed, model selection was repeated using all seven historical earthquakes, the candidate model structures described in Sect. 3.2, and the model selection criterion specified in Sect. 3.4. The model structure comprising MMI,  $\ln(1+\text{TFA})$ , Pre70, and  $\sqrt{\text{GasHD}}$  achieved the highest composite score, with an event mean for AP of 0.275 and a mean of 0.542 across the three Recall@Top-k% values. Accordingly, this structure was selected as the final model structure.

## 4.4 Sensitivity analysis results

### 4.4.1 Ground-motion intensity-measure sensitivity analysis

The non-ground-motion explanatory variables in the final model structure were held fixed, while MMI,  $\ln(\text{PGA})$ , or  $\ln(\text{PGV})$  was used to represent ground-motion intensity. All three models were evaluated using LOEO, with AUROC, AP, and  $\text{Recall@Top-k\%}$  calculated separately for each held-out earthquake event. Table 5 reports the event means for these metrics and the numbers of earthquake-census-tract observations with at least one recorded ignition contained in the Top-k% ranges.

**Table 5** LOEO assessment results for models using three ground-motion intensity measures

| Model             | Event | Event |                        |                         |                         | Observations  | Observations  | Observations  |
|-------------------|-------|-------|------------------------|-------------------------|-------------------------|---------------|---------------|---------------|
|                   | mean  | mean  | Event mean for         | Event mean for          | Event mean for          | with at least | with at least | with at least |
|                   | for   | for   | $\text{Recall@Top5\%}$ | $\text{Recall@Top10\%}$ | $\text{Recall@Top20\%}$ | one recorded  | one recorded  | one recorded  |
|                   | AUROC | AP    |                        |                         |                         | ignition in   | ignition in   | ignition in   |
|                   |       |       |                        |                         |                         | Top5%         | Top10%        | Top20%        |
| MMI               | 0.830 | 0.275 | 0.352                  | 0.506                   | 0.767                   | 43            | 72            | 98            |
| $\ln(\text{PGA})$ | 0.809 | 0.285 | 0.315                  | 0.472                   | 0.684                   | 27            | 55            | 79            |
| $\ln(\text{PGV})$ | 0.828 | 0.273 | 0.326                  | 0.507                   | 0.734                   | 37            | 66            | 91            |

As shown in Table 5, the MMI model yielded event means of 0.830, 0.352, and 0.767 for AUROC,  $\text{Recall@Top5\%}$ , and  $\text{Recall@Top20\%}$ , respectively. Its Top5%, Top10%, and Top20% ranges contained 43, 72, and 98 earthquake-census-tract observations with at least one recorded ignition, respectively, exceeding the corresponding counts for the other two models. The  $\ln(\text{PGA})$  model had the highest event mean for AP (0.285), but its event means for AUROC and all three  $\text{Recall@Top-k\%}$  metrics were lower than those of the MMI model. The  $\ln(\text{PGV})$  model yielded event means of 0.828 and 0.273 for AUROC and AP, respectively, which were close to those of the MMI model.

Using the MMI model as the reference, Fig. 7 shows the differences in the evaluation metrics for the  $\ln(\text{PGA})$  and  $\ln(\text{PGV})$  models in each held-out earthquake event. For the Loma Prieta and Northridge earthquakes, both models yielded lower values than the MMI model for all evaluation metrics. For the Napa earthquake, the AP values of the  $\ln(\text{PGA})$  and  $\ln(\text{PGV})$  models were higher by 0.169 and 0.103, respectively, and their  $\text{Recall@Top10\%}$  values were both higher by 0.200. The  $\ln(\text{PGA})$  model had the highest event mean for AP, mainly because of its higher AP for the Napa earthquake. The MMI model yielded higher  $\text{Recall@Top-k\%}$  values for the Loma Prieta and Northridge earthquakes and therefore contained the largest numbers of observations with at least one recorded ignition in all three Top-k% ranges.

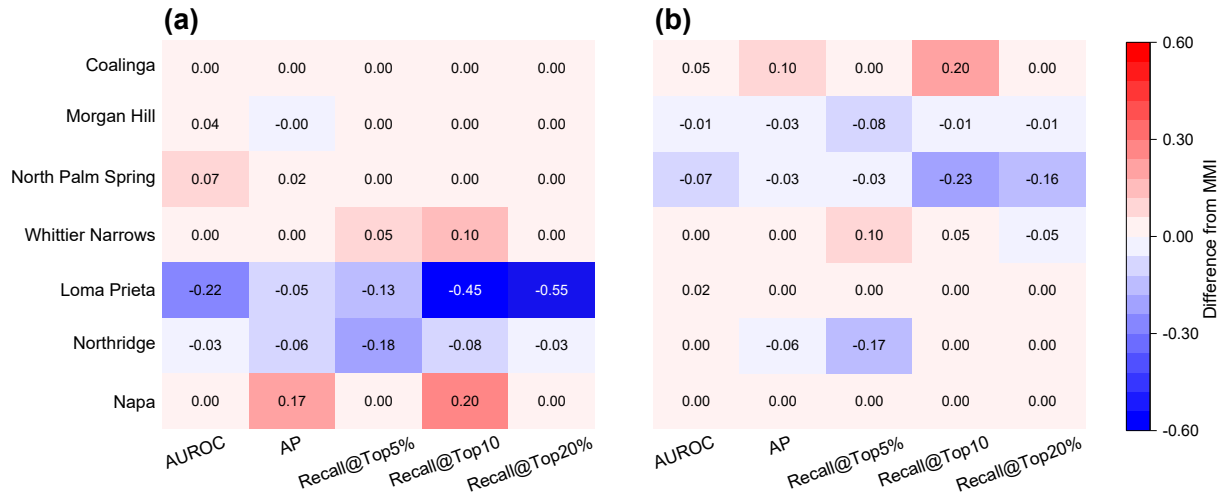


**Fig. 7** Differences in evaluation metrics among models using different ground-motion intensity measures across the held-out earthquake events. **a**  $\ln(\text{PGA}) - \text{MMI}$ . **b**  $\ln(\text{PGV}) - \text{MMI}$

Overall, with the non-ground-motion explanatory variables held fixed, changing the ground-motion intensity measure altered the rankings for some earthquake events. However, the event mean for AUROC exceeded 0.80 for all three models, and, for each model, the Top20% ranges across the seven events contained more than half of all earthquake-census-tract observations with at least one recorded ignition.

#### 4.4.2 Sample-splitting sensitivity analysis

Following the procedure described in Sect. 3.5, the final model structure was evaluated using repeated stratified five-fold cross-validation by earthquake event and response class, with 100 repetitions across the seven earthquakes. Table 6 summarizes the event means obtained in these repetitions. When averaged across the 100 repetitions, the event means for AUROC, AP, Recall@Top5%, Recall@Top10%, and Recall@Top20% were 0.838, 0.276, 0.394, 0.530, and 0.745, respectively. Compared with the fixed-structure LOEO results for the MMI model in Table 5, the first four metrics were slightly higher, whereas Recall@Top20% was slightly lower.

**Table 6** Event means from 100 repetitions of stratified five-fold cross-validation

| Evaluation metric | Mean  | Standard deviation | Median | 25th percentile | 75th percentile |
|-------------------|-------|--------------------|--------|-----------------|-----------------|
| AUROC             | 0.838 | 0.004              | 0.839  | 0.836           | 0.841           |
| AP                | 0.276 | 0.033              | 0.294  | 0.231           | 0.299           |
| Recall@Top5%      | 0.394 | 0.020              | 0.396  | 0.375           | 0.407           |
| Recall@Top10%     | 0.530 | 0.013              | 0.534  | 0.528           | 0.539           |
| Recall@Top20%     | 0.745 | 0.069              | 0.792  | 0.653           | 0.796           |

Both the training and test sets in repeated five-fold cross-validation contained observations from all seven earthquakes. Consequently, different census-tract observations from the same earthquake event contributed separately to model fitting and testing. By contrast, fixed-structure LOEO excluded all census-tract observations from the held-out earthquake when estimating the regression coefficients in each fold. The former assessed predictive performance for test census tracts within earthquake events that were also represented in the training set, whereas the latter assessed the predictive performance of the final model structure for a completely held-out earthquake event.

Fig. 8 shows the distributions of the event means across the 100 repetitions and compares them with the

fixed-structure LOEO results. AP and Recall@Top20% exhibited greater variability across the repetitions, primarily because the Coalinga and N. Palm Spring earthquakes each contained only one census tract with at least one recorded ignition. For Coalinga, this census tract ranked first in 74 repetitions and second in 26 repetitions. For N. Palm Spring, the only census tract with at least one recorded ignition fell within the Top20% range in 65 repetitions and outside it in 35 repetitions. These changes in ranking position for events containing only one such census tract widened the distributions of AP and Recall@Top20%.

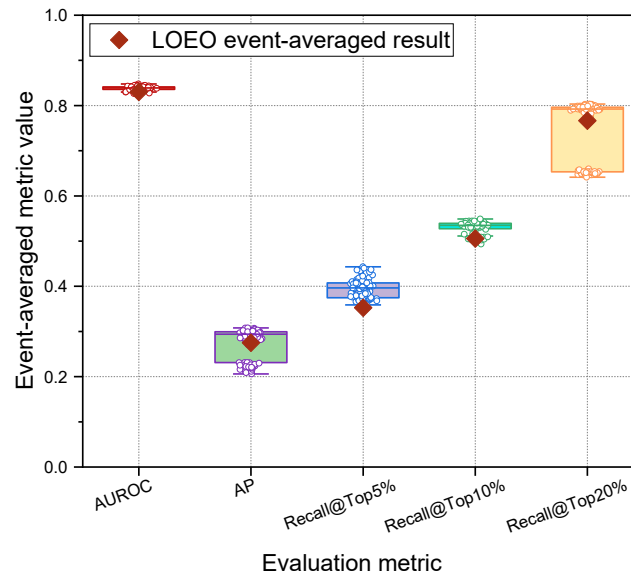


**Fig. 8** Distributions of event-mean metrics from 100 repetitions of stratified five-fold cross-validation. Diamonds show fixed-structure LOEO results

Overall, the event means for AUROC were at least 0.830 and those for Recall@Top20% exceeded 0.700 under both sample-splitting schemes. Under both schemes, the Top20% ranges contained most census tracts with at least one recorded ignition on average across the seven events. Because this study focuses on predictive performance for completely held-out earthquake events, assessment of the full modeling procedure remained based on the outer-loop results reported in Sect. 4.2.

#### 4.5 Regression coefficient estimates for the final predictive model

After the final model structure had been determined, the logistic regression model was refitted using all 7,074 earthquake-census-tract observations to obtain the final predictive model for probability mapping and scenario application (Lewis et al. 2023). The regression coefficient estimates for the final predictive model are presented in Table 7. The regression coefficients for MMI,  $\ln(1+\text{TFA})$ , Pre70, and  $\sqrt{\text{GasHD}}$  were all positive. Holding the other explanatory variables constant, higher values of each variable corresponded to higher predicted ignition probabilities.

**Table 7** Regression coefficient estimates for the final predictive model

| Variable              | Regression coefficient | Standard error |
|-----------------------|------------------------|----------------|
| Intercept             | -28.3373               | 2.7414         |
| MMI                   | 1.6485                 | 0.1127         |
| $\ln(1+\text{TFA})$   | 0.9369                 | 0.1979         |
| Pre70                 | 0.5564                 | 0.4346         |
| $\sqrt{\text{GasHD}}$ | 0.0365                 | 0.0057         |

## 5 Application to the USGS HayWired Earthquake Scenario

The USGS developed the HayWired earthquake scenario, which assumes an Mw 7.0 earthquake on the Hayward Fault, to examine damage and losses in the San Francisco Bay region under the scenario earthquake (Detweiler and Wein 2017). The final predictive model was applied to the HayWired earthquake scenario to calculate predicted ignition probabilities for individual census tracts and identify the Top5%, Top10%, and Top20% priority screening sets. The county-level ignition estimates reported by Scawthorn (2018) for the HayWired earthquake scenario were used as a regional spatial reference for comparing the counties in which the reported estimates and the Top-k% priority screening sets were primarily concentrated.

### 5.1 Distribution of predicted ignition probabilities and priority screening sets

Ground-motion data for the HayWired earthquake scenario were obtained from the USGS ShakeMap grid.xml file (Worden et al. 2020). After overlaying the ShakeMap grid with the Census 2000 tract boundaries, the area-weighted mean PGA and MMI were calculated for each census tract. The 1,721 census tracts with area-weighted mean PGA values of at least 0.08 g were included in the scenario analysis area.

The spatial distributions of predicted census-tract ignition probabilities and priority screening sets under the HayWired earthquake scenario are shown in Fig. 9. In Fig. 9(a), census tracts with higher predicted ignition probabilities were distributed mainly along the eastern shore of San Francisco Bay and in the South Bay. They were concentrated in Alameda and Santa Clara counties, with fewer such tracts in Contra Costa County and the northern Bay Area. Fig. 9(b) shows that, when census tracts were ranked in descending order of predicted ignition probability, the Top5%, Top10%, and Top20% priority screening sets comprised 87, 173, and 345 census tracts, respectively. As the priority screening set expanded from the Top5% to the Top20%, the newly included census tracts were located mainly around the previously included areas and extended progressively along the eastern shore of San Francisco Bay and into the South Bay.

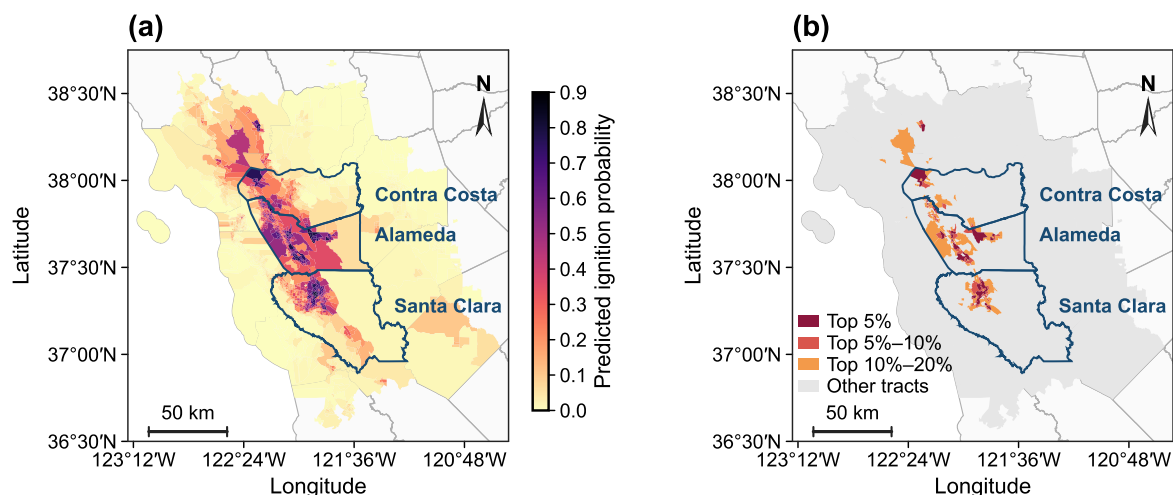


**Fig. 9** Spatial distributions of predicted ignition probability and Top-k% priority screening sets at the census-tract level under the HayWired earthquake scenario. **a** predicted ignition probability. **b** Top5%, Top10%, and Top20% priority screening sets

## 5.2 County-level comparison of priority screening sets and HayWired ignition estimates

The HayWired study estimated that 668 ignitions would occur following the scenario earthquake, including 352 in Alameda County, 123 in Contra Costa County, and 83 in Santa Clara County. Together, these three counties accounted for 558 estimated ignitions, or 83.5% of the total (Scawthorn 2018). Table 8 compares the county-level compositions of the ignition estimates reported for the HayWired earthquake scenario and the three priority screening sets defined in this study.

**Table 8** County-level composition of ignition estimates reported for the HayWired earthquake scenario and Top-k% priority screening sets

| Quantity                      | Unit          | Total | Alameda | Contra Costa | Santa Clara | Other counties | Three-county total | Three-county share (%) |
|-------------------------------|---------------|-------|---------|--------------|-------------|----------------|--------------------|------------------------|
| Reported ignition estimates   | Ignitions     | 668   | 352     | 123          | 83          | 110            | 558                | 83.5                   |
| Top5% priority screening set  | Census tracts | 87    | 41      | 4            | 36          | 6              | 81                 | 93.1                   |
| Top10% priority screening set | Census tracts | 173   | 86      | 9            | 70          | 8              | 165                | 95.4                   |
| Top20% priority screening set | Census tracts | 345   | 176     | 28           | 121         | 20             | 325                | 94.2                   |

## 6 Conclusions

Using 7,074 earthquake-census-tract observations from seven historical California earthquakes, this study developed a census-tract-level model of FFE ignition probability. The full modeling procedure was evaluated using nested cross-validation, and the final predictive model was applied to the HayWired earthquake scenario. The main conclusions are as follows.

(1) The event means for AUROC and AP across the seven outer test events were 0.833 and 0.267, respectively. The event-specific Top5%, Top10%, and Top20% priority screening sets contained 34, 67, and 97 earthquake-census-tract observations with at least one recorded ignition, respectively, accounting for 25.2%, 49.6%, and 71.9% of all such observations. In every outer test event, the Top20% priority screening set included at least half of the census tracts with at least one recorded ignition. Calibration assessment showed that the outer-test predicted probabilities were close to the observed ignition proportions in the low-probability range, where most observations were concentrated. Overall, the predicted probabilities were slightly overestimated, with larger deviations in the higher-probability range.

(2) The final model structure comprised MMI,  $\ln(1+\text{TFA})$ , Pre70, and  $\sqrt{\text{GasHD}}$ , representing a ground-motion intensity measure, an exposure variable, a proxy variable for building vulnerability, and a proxy

variable for potential ignition sources, respectively. This structure achieved the highest composite score among the 211 candidate model structures. Its event mean for AP was 0.275, and the mean of the three event-mean Recall@Top-k% values was 0.542. After the model was refitted using all historical observations, the regression coefficients for all four explanatory variables were positive.

(3) Changing the ground-motion intensity measure altered the rankings by predicted ignition probability for some earthquake events. For each of the three models, the Top20% ranges contained more than half of all earthquake-census-tract observations with at least one recorded ignition. Although the MMI model did not yield the highest event mean for AP, it yielded the highest event means for AUROC and Recall@Top20%. It also included the largest numbers of observations with at least one recorded ignition in all three Top-k% ranges. The MMI model therefore had a relative advantage in overall ranking discrimination and in the coverage of census tracts with at least one recorded ignition by the priority screening sets.

(4) The evaluation results from 100 repetitions of stratified five-fold cross-validation by earthquake event and response class were broadly similar to the LOEO results. Compared with the fixed-structure LOEO results, the event means for AUROC, AP, Recall@Top5%, and Recall@Top10% were higher by 0.008, 0.001, 0.042, and 0.024, respectively, whereas the event mean for Recall@Top20% was lower by 0.022. Both the training and test sets in repeated five-fold cross-validation contained census-tract observations from the same earthquake events. Therefore, the LOEO results remained the primary basis for evaluation in this study.

(5) The Top-k% priority screening sets and the ignition estimates reported for the HayWired earthquake scenario were concentrated mainly in Alameda, Contra Costa, and Santa Clara counties. Together, these counties accounted for 83.5% of the reported ignition estimates. Census tracts in these counties accounted for 93.1%, 95.4%, and 94.2% of the Top5%, Top10%, and Top20% priority screening sets, respectively. The Top-k% results identified the census-tract spatial extents corresponding to different screening proportions.

## References

- Austin PC, Steyerberg EW (2014) Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med* 33(3):517–535. <https://doi.org/10.1002/sim.5941>
- Cawley GC, Talbot NLC (2010) On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res* 11:2079–2107
- Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP (2008) Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int J Climatol* 28(15):2031–2064. <https://doi.org/10.1002/joc.1688>
- Davidson RA (2009) Modeling postearthquake fire ignitions using generalized linear (mixed) models. *J Infrastruct Syst* 15(4):351–360. [https://doi.org/10.1061/\(ASCE\)1076-0342\(2009\)15:4\(351\)](https://doi.org/10.1061/(ASCE)1076-0342(2009)15:4(351))
- Detweiler ST, Wein AM (eds) (2017) The HayWired earthquake scenario. U.S. Geological Survey Scientific Investigations Report 2017–5013. U.S. Geological Survey, Reston, VA. <https://doi.org/10.3133/sir20175013>
- Dhakal RP, Pourali A, Tasligedik AS, Yeow T, Baird A, MacRae G, Pampanin S, Palermo A (2016) Seismic performance of non-structural components and contents in buildings: an overview of NZ research. *Earthq Eng Eng Vib* 15(1):1–17. <https://doi.org/10.1007/s11803-016-0301-9>

543 Elhami Khorasani N, Gernay T, Garlock M (2017) Data-driven probabilistic post-earthquake fire ignition  
544 model for a community. *Fire Saf J* 94:33–44. <https://doi.org/10.1016/j.firesaf.2017.09.005>

545 Farahani S, Tahershamsi A, Behnam B (2020) Earthquake and post-earthquake vulnerability assessment of  
546 urban gas pipelines network. *Nat Hazards* 101(2):327–347. <https://doi.org/10.1007/s11069-020-03874-4>

547 FEMA (1999) Earthquake loss estimation methodology: HAZUS 99 technical manual. Prepared by the  
548 National Institute of Building Sciences for the Federal Emergency Management Agency, Washington, DC

549 He Z, Elhami Khorasani N (2022) Identification and hierarchical structure of cause factors for fire following  
550 earthquake using data mining and interpretive structural modeling. *Nat Hazards* 112(1):947 – 976.  
551 <https://doi.org/10.1007/s11069-022-05214-0>

552 Hokugo A (1995) Fires following the Great Hanshin earthquake in Kobe. In: Bulgakov VK, Karpov AI (eds)  
553 Fire science and technology: proceedings of the Second Asia-Oceania Symposium. Khabarovsk State  
554 University of Technology, Khabarovsk, pp 18–29

555 Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) Applied logistic regression, 3rd edn. Wiley, Hoboken,  
556 NJ. <https://doi.org/10.1002/9781118548387>

557 Ismail N, Khattak N (2019) Observed failure modes of unreinforced masonry buildings during the 2015  
558 Hindu Kush earthquake. *Earthq Eng Eng Vib* 18(2):301–314. <https://doi.org/10.1007/s11803-019-0505-x>

559 Krstajic D, Buturovic LJ, Leahy DE, Thomas S (2014) Cross-validation pitfalls when selecting and assessing  
560 regression and classification models. *J Cheminform* 6(1):10. <https://doi.org/10.1186/1758-2946-6-10>

561 Lee S, Davidson R, Ohnishi N, Scawthorn C (2008) Fire following earthquake—reviewing the state-of-the-  
562 art of modeling. *Earthq Spectra* 24(4):933–967. <https://doi.org/10.1193/1.2977493>

563 Lewis MJ, Spiliopoulou A, Goldmann K, Pitzalis C, McKeigue P, Barnes MR (2023) nestedcv: an R package  
564 for fast implementation of nested cross-validation with embedded feature selection designed for  
565 transcriptomics and high-dimensional data. *Bioinform Adv* 3(1):vb048.  
566 <https://doi.org/10.1093/bioadv/vb048>

567 Liang J (2003) Online location of seismic damage to a water distribution system. *Earthq Eng Eng Vib*  
568 2(2):181–187. <https://doi.org/10.1007/s11803-003-0002-z>

569 Lu XZ, Yue QR, Xu Z, Wang YX, Gu DL, Tian Y (2024) A review on novel seismic secondary disasters in  
570 urban dense building areas. *Industrial Construction* 54(2):25 – 34 (in Chinese).  
571 <https://doi.org/10.3724/j.gyzG23121501>

572 Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. Cambridge University  
573 Press, Cambridge. <https://doi.org/10.1017/CBO9780511809071>

574 Nishino T (2025) Fires following the 2024 Noto Peninsula earthquake: insights from ignition records, a fire  
575 damage survey, and a prospective regional fire risk assessment. *Earthq Spectra* 41(3):1998 – 2027.  
576 <https://doi.org/10.1177/87552930251335207>

577 Nowicki Jessee MA, Hamburger MW, Allstadt KE, Wald DJ, Robeson SM, Tanyas H, Hearne M, Thompson  
578 EM (2018) A global empirical model for near-real-time assessment of seismically induced landslides. *J*  
579 *Geophys Res Earth Surf* 123(8):1835–1859. <https://doi.org/10.1029/2017JF004494>

580 Ren AZ, Xie XY (2004) The simulation of post-earthquake fire-prone area based on GIS. *J Fire Sci*  
581 22(5):421–439. <https://doi.org/10.1177/0734904104042440>

582 Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ,  
583 Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF (2017) Cross-validation strategies for  
584 data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40(8):913 – 929.  
585 <https://doi.org/10.1111/ecog.02881>

586 Saito T, Rehmsmeier M (2015) The precision-recall plot is more informative than the ROC plot when  
587 evaluating binary classifiers on imbalanced datasets. *PLoS One* 10(3):e0118432.  
588 <https://doi.org/10.1371/journal.pone.0118432>

589 Sarreshtehdari A, Elhami Khorasani N (2021) Integrating the fire department response within a fire following  
590 earthquake framework for application in urban areas. *Fire Saf J* 124:103397.  
591 <https://doi.org/10.1016/j.firesaf.2021.103397>

592 Scawthorn C (1986) Fire following earthquake. *Fire Saf Sci* 1:971 – 979.  
593 <https://doi.org/10.3801/IAFSS.FSS.1-971>

594 Scawthorn C (2018) Fire following the HayWired scenario mainshock. In: Detweiler ST, Wein AM (eds) *The*  
595 *HayWired earthquake scenario—engineering implications*. U.S. Geological Survey Scientific Investigations  
596 Report 2017–5013–I–Q, chap P, pp 367–400. <https://doi.org/10.3133/sir20175013v2>

597 Tong Q, Gernay T (2022) A hierarchical Bayesian model for predicting fire ignitions after an earthquake with  
598 application to California. *Nat Hazards* 111(2):1637–1660. <https://doi.org/10.1007/s11069-021-05109-6>

599 Tsamardinos I (2022) Don't lose samples to estimation. *Patterns* 3(12):100612.  
600 <https://doi.org/10.1016/j.patter.2022.100612>

601 Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW (2019) Calibration: the Achilles  
602 heel of predictive analytics. *BMC Med* 17(1):230. <https://doi.org/10.1186/s12916-019-1466-7>

603 Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. *BMC*  
604 *Bioinformatics* 7(1):91. <https://doi.org/10.1186/1471-2105-7-91>

605 Wang HR, Xiao J, Li S, Zhai CH (2024) Resilience assessment and optimization method of city road network  
606 in the post-earthquake emergency period. *Earthq Eng Eng Vib* 23(3):765 – 779.  
607 <https://doi.org/10.1007/s11803-024-2254-8>

608 Wang W, Gao Y, Guo XD, Liu C (2023) Post-earthquake fire risk assessment based on multi-source data  
609 model. *Journal of Catastrophology* 38(3):192–197 (in Chinese)

610 Wang W, Chen BY, Zhang LM, Ma DH (2026) A study on the spread model of urban earthquake-induced  
611 fires based on variable neighborhood cellular automata. *Journal of Beijing University of Technology*  
612 52(7):764–777 (in Chinese). <https://doi.org/10.11936/bjutxb2024100005>

613 Worden CB, Thompson EM, Hearne M, Wald DJ (2020) ShakeMap manual online: technical manual, user's  
614 guide, and software guide. U.S. Geological Survey. <https://doi.org/10.5066/F7D21VPQ>

615 Yan PL, Guo ED, Huang Y, Zhao ZP, Lata A, Liu JY (2026) Seismic damage characteristics and restoration  
616 strategies of the Mandalay municipal water supply system following the 2025 M 7.9 Myanmar earthquake.  
617 *Earthq Eng Eng Vib* 25(1):13–25. <https://doi.org/10.1007/s11803-026-2360-x>

618 Yazgan U, Oyguç R, Ergüven ME, Celep Z (2016) Seismic performance of buildings during 2011 Van  
619 earthquakes and rebuilding efforts. *Earthq Eng Eng Vib* 15(3):591–606. <https://doi.org/10.1007/s11803-016-0346-9>

620

621 Yoshioka H, Himoto K, Kagiya K (2020) Large urban fires in Japan: history and management. Fire Technol  
622 56(5):1885–1901. <https://doi.org/10.1007/s10694-020-00960-8>

623 Zhao SJ, Xiong LY, Ren AZ (2006) A spatial–temporal stochastic simulation of fire outbreaks following  
624 earthquake based on GIS. J Fire Sci 24(4):313–339. <https://doi.org/10.1177/0734904106060786>

625

## 626 **Statements and Declarations**

627 Competing interests

628 The authors have no relevant financial or non-financial interests to disclose.