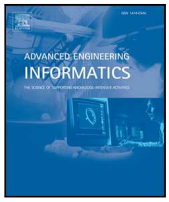
Advanced Engineering Informatics 60 (2024) 102414

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei



Full length article

# Efficient post-earthquake reconnaissance planning using adaptive batch-mode active learning

Amirhossein Cheraghi [a,*], Yinhu Wang [b], Nikola Marković [b], Ge Ou [a]

[a] Department of Civil and Coastal Engineering, University of Florida, Gainesville, 32611, FL, USA
[b] Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, 84112, UT, USA

ARTICLE INFO

*Keywords:*
Batch active learning
Gaussian process regression
Route optimization
Mutual information
Post-earthquake damage and loss assessment

ABSTRACT

In the aftermath of earthquakes, obtaining a timely and accurate estimate of infrastructure damage is crucial for effective emergency response and recovery planning. Previous research has shown that active learning (AL) methods enable the inference of regional infrastructure damage from sparse on-site inspections through post-earthquake reconnaissance surveys. However, the efficiency of the surveys depends on two key factors: appropriate identification of candidate buildings for inspection and optimized routing between the sequential inspections. This study presents an AL framework that integrates learning and routing objectives, aiming to enhance the efficiency of post-earthquake damage data collection. The proposed approach employs an adaptive batch-mode AL method, using a Gaussian process regression model and an information-theoretic criterion to suggest a batch of candidate buildings for inspection. These buildings are then adaptively visited based on an optimized route schedule. The results demonstrate that the proposed framework reduces the required number of building inspections by up to 53%, decreases resource demands, and improves the predictive performance of the model compared with a baseline method.

## 1. Introduction

Earthquakes are among the most catastrophic natural hazards that imperil public safety and infrastructure integrity. Recently, the magnitude 7.8 Turkey-Syria earthquake led to over 40,000 fatalities and severe damage to hundreds of thousands of buildings. In the immediate aftermath of an earthquake, decision-makers seek reliable information about the extent of the damage for two crucial purposes: identifying unsafe buildings to protect residents [1] and fairly allocating resources to mitigate socio-economic consequences of the disaster [2,3].

Several methods have been developed for assessing building damage after an earthquake, which can be divided into three main categories: model-based estimations, assessments based on remotely-sensed information, and field data collection. Damage forecast models such as Prompt Assessment of Global Earthquakes for Response (PAGER) [4] rapidly maps the intensity of ground motion (GM) to a damage level. However, the estimation of damage is based on vulnerability models that rely on past events, which implies the estimation is not completely event-specific [5]. This potentially leads to inflated levels of uncertainty for locations lacking sufficient historical or recent events [6]. Remote sensing-based methods provide broad coverage for the region of interest [7–9]; however, they are typically limited to binary estimates, such as collapsed or non-collapsed. Satellite imagery cannot accurately provide information about refined damage states due to technical limitations and, sometimes, due to inclement weather conditions [10–13]. Consequently, decision-making solely based on the aforementioned methods does not provide a clear picture of the damage in the affected region although they can be utilized as auxiliary sources of information. On-ground surveys, on the other hand, can provide detailed information on both structural and non-structural components whose damage data cannot be captured through other methods. Nevertheless, due to the time-consuming nature of these tasks, only sparse building inspections are typically plausible [11,14,15].

Researchers have employed data-driven approaches to overcome the limitations of post-earthquake data acquisition, resulting in improved accuracy and efficiency in inferring regional damage [16–18]. Lallemant and Kiremidjian [19] used a cokriging method to integrate damage data from remote-sensing and limited field surveys, which yielded damage estimates that closely aligned with actual damage. Loos et al. [5] proposed a Geospatial Data Integration Framework (G-DIF) that assimilates damage information from diverse sources, namely damage forecast models, remote-sensing data, and field surveys. It was shown that even with limited coverage, field observations contribute to improving damage predictions. Sheibani and Ou [13] adopted a supervised learning approach using a Gaussian process regression (GPR) model to predict regional-scale damage. Considering a

* Corresponding author.
*E-mail address:* a.cheraghi@ufl.edu (A. Cheraghi).

https://doi.org/10.1016/j.aei.2024.102414
Received 16 June 2023; Received in revised form 4 February 2024; Accepted 9 February 2024
Available online 13 February 2024
1474-0346/© 2024 Elsevier Ltd. All rights reserved.

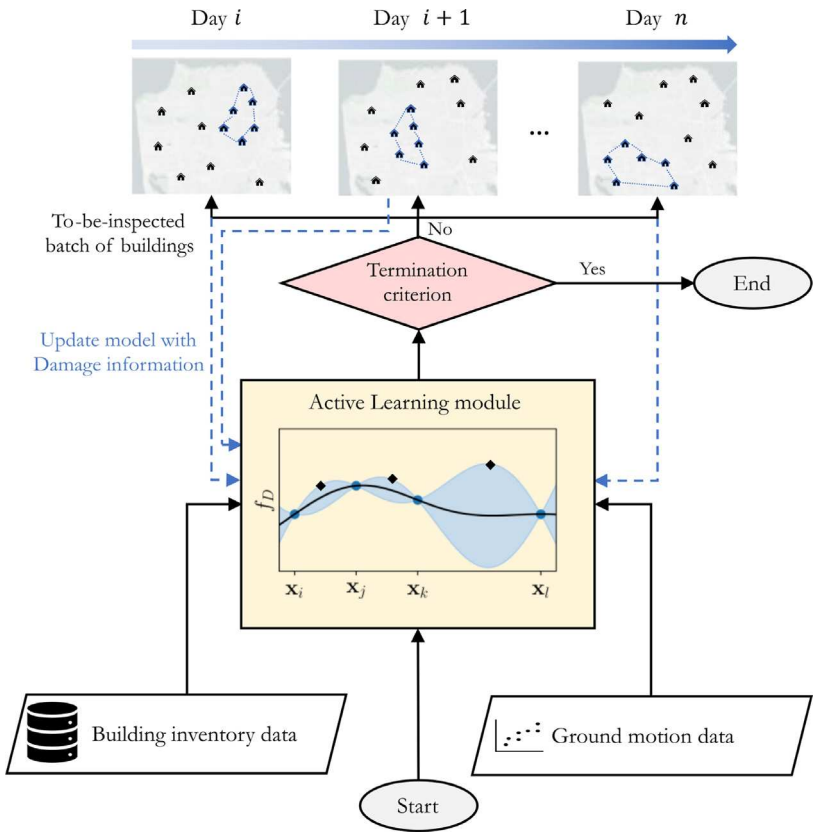



**Fig. 1.** Schematic overview of the adaptive batch-mode AL framework using Gaussian process regression. Black diamonds denote informative queries that form the batch of buildings for inspection (depicted as black house icons). The house icons with blue outlines highlight the subset of the batch that is adaptively chosen for damage inspections.

set of building attributes and GM indices as inputs and the observed damage as outputs, they showed that a GPR model could accurately infer the input–output relationship using data from only a limited number of inspected buildings. Bodenmann et al. [20] enhanced the predictive accuracy of traditional risk models by updating vulnerability function parameters with damage data obtained from field inspections, emphasizing the value of incorporating early damage inspection data. Given the impact of early post-earthquake surveys on regional damage estimation and limitations in the number of inspections, a systematic approach is needed to streamline these tasks.

Efficient post-earthquake damage assessment surveys require careful selection of candidate buildings. To accurately predict damage with limited observations, candidate buildings should be representative of the majority of structures in the affected area and cover a broad spectrum of building attributes and input GM intensities. Sheibani and Ou [13] used unsupervised learning to select candidate buildings, demonstrating that regional inference accuracy improves when employing $k$-medoids clustering. Other studies have also emphasized the importance of spatial coverage and strategic placement of inspections [5, 21]. While selecting representative and diverse candidate buildings increases predictive performance, this approach depends solely on the input space (building and GM characteristics) and is referred to as passive sampling [22,23]. In contrast, as the survey proceeds and more buildings are inspected, the gradual inflow of information about the buildings' damage can be leveraged to select future samples more efficiently. In machine learning research, this sampling paradigm is referred to as active learning (AL) [24]. In fact, the model can determine the next to-be-labeled data point (query) using the information gained through the previously collected samples.

The AL framework allows us to improve prediction accuracy with fewer training instances by selecting the most informative data points for the model to learn from [24]. This is particularly useful when labeling data is prohibitive or requires expert knowledge so that the model can achieve better performance using fewer training examples.

In the context of post-earthquake building reconnaissance, given the time-consuming nature of the inspections, the perishable quality of damage data that is prone to deterioration over time [25], and the immediate need for a regional damage estimate, the least number of building inspections should be demanded, and AL can be applied to identify the most informative buildings. AL has been used to efficiently select the candidate buildings and guide the post-earthquake reconnaissance missions [26]. It was shown that the sequential selection of buildings based on mutual information (between the inspected and uninspected buildings) led to better prediction performance for a GPR learner than entropy-based sequential design and random sampling [26].

Although sequential AL methods can effectively identify the most informative buildings for inspection and improve learning performance, it is essential to consider the logistics of the task to ensure an efficient post-earthquake survey. Factors such as road disruptions or blockages [27] and potential hazards to the survey teams can make particular buildings suggested by a sequential AL framework inaccessible. In such cases, arbitrarily replacing an inaccessible candidate building with another that is feasible to inspect can result in suboptimal information gain, which reduces the accuracy of damage and loss predictions in practice. Moreover, the sequential AL framework has no knowledge of the geographical location of buildings, which can lead to queries for inspections in distant locations, making the survey task more time-consuming and less efficient. To address these limitations of the current state-of-the-art methods in survey planning and increase the practicality of AL in natural hazard engineering, we propose an AL framework that allows for querying buildings in batches and considers the geographical location of candidate buildings.

As illustrated in Fig. 1, the framework employs a GPR model to infer a damage function $f_D$ that maps each building (such as $\mathbf{x}_i$) to

a numerical representation of the damage or loss that the building has sustained. Each day, the active learning module returns a batch of to-be-inspected buildings (depicted as black house icons) by identifying the buildings that knowing their damage status effectively contributes to learning the underlying damage function $f_D$, by strategically placing queries at informative points (as indicated by the black diamonds). The objective function governing the query placement is further elaborated in Section 3. An integrated routing module adaptively selects a subset of the batch (house icons with blue outline) such that in each workday the number of inspections is maximized and thus favors the informative candidate buildings in close vicinity of the surveying team. The results of the inspections are returned to the AL module for updating the GPR model with new inspection results (as indicated by the blue dots), and the AL module provides the batch of candidate buildings for the next day accordingly. The loop continues until a stopping criterion is satisfied. Having a batch of candidate buildings instead of querying one building at a time enables flexibility in the survey process and provides a buffer in favor of more efficient routing. Besides, it obviates the need for sequential model updating after each inspection which can be time-consuming and cause additional strain on the team's resources in terms of time and effort required for data processing and re-training the model.

The manuscript is organized as follows: Section 2 provides a brief overview of GPR for regional damage inference. The proposed framework is presented in Section 3. Section 4 outlines two earthquake testbeds – a hypothetical scenario and a real-world event – used to assess passive sampling and active learning algorithms. Section 5 examines sampling effectiveness and predictive performance. Time and resource requirements are evaluated in Section 6. Section 7 explores the limitations of the study and potential future directions. Finally, Section 8 summarizes the findings and draws conclusions.

## 2. Post-earthquake regional damage inference

The post-earthquake damage inference is formulated as a supervised learning problem. Each building is treated as a data point in a multidimensional space representing different features, and the observed damage and loss values are considered as labels [13]. Assuming a set of candidate buildings for inspection has been identified, the trained GPR model enables regional damage and loss estimation by predicting the labels for those buildings whose damage state is not known. We briefly describe GPR and its components in the context of the regional damage inference application.

### 2.1. Gaussian process regression

Gaussian process regression is a non-parametric Bayesian approach that infers unknown functions that map input–output data [28]. Non-parametric models do not assume a parametric form for the unknown function; instead, they directly estimate the function values from the data [29].

The function $f$ represents an unknown mapping between a vector $\mathbf{x} \in \mathbb{R}^D$ that encapsulates the building attributes and ground motion characteristics and a quantifiable measure of building damage or loss $f(\mathbf{x}) \in \mathbb{R}$ (i.e., $f : \mathbb{R}^D \to \mathbb{R}$).

We assume the function values $f(\cdot)$ form a Gaussian process (GP). A Gaussian process is a collection of random variables, any finite number of which has a multivariate normal distribution. A GP is defined by its mean $m$ and covariance $k$ functions:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \tag{1}$$

where

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})], \tag{2}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]. \tag{3}$$

When there is limited or unreliable information about the function being learned, it is common to assume a prior with zero mean, denoted by $m(\cdot) = 0$.

The observations obtained from building inspections are often affected by noise. The estimates of building loss made by the survey team are therefore considered to be noisy observations of the true damage. A zero-mean Gaussian noise term $\epsilon$ is thus added to the true function $f$. As such, the relation between the damage estimates provided by the survey team $y$ and the true damage value for a certain building $f(\cdot)$ can be formalized as follows:

$$y = f(\mathbf{x}) + \epsilon, \tag{4}$$

where $\epsilon$ denotes a zero-mean Gaussian noise with variance $\sigma_n^2$.

Extending the observations to cover multiple buildings, if $X_{\mathcal{L}}$ and $\mathbf{y}$ represent the set of inspected buildings (or training data) and their corresponding damage observations respectively, we are interested in inferring the function values at a set of buildings $X_*$ (or testing data), as denoted by $\mathbf{f}_*$. The conditional probability $\mathbf{f}_*|X_{\mathcal{L}}, \mathbf{y}, X_*$ characterizes the posterior distribution. Given that $\mathbf{f}_*$ is a Gaussian process, this probability results in a multivariate Gaussian distribution [30], represented by the mean vector $\boldsymbol{\mu}_*$ and covariance matrix $\boldsymbol{\Sigma}_*$:

$$\mathbf{f}_*|X_{\mathcal{L}}, \mathbf{y}, X_* \sim \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*), \tag{5}$$

where

$$\boldsymbol{\mu}_* = K_{*\mathcal{L}}[K_{\mathcal{L}\mathcal{L}} + \sigma_n^2 I]^{-1}\mathbf{y}, \tag{6}$$

$$\boldsymbol{\Sigma}_* = K_{**} - K_{*\mathcal{L}}[K_{\mathcal{L}\mathcal{L}} + \sigma_n^2 I]^{-1}K_{\mathcal{L}*}, \tag{7}$$

where $K_{ij}$ is a covariance matrix containing kernel function values $k(\cdot, \cdot)$ for all pairs of training and/or testing data points.

### 2.2. Kernel function and hyperparameter optimization

An important component of GPR is the kernel or covariance functions $k$. Kernels encode the correlation between function values for each pair of data points such that for $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^D$, $\mathrm{cov}[f(\mathbf{x}), f(\mathbf{x}')] = k(\mathbf{x}, \mathbf{x}')$. The choice of a suitable kernel function is part of the model selection process. Sheibani and Ou [13] conducted a comparative study on different kernel functions for regional damage inference problem and identified that rational quadratic (RQ) kernel with Automatic Relevance Determination (ARD) feature, effectively captures the data patterns in different dimensions, resulting in smaller generalization error. ARD functions enable the model to better express the structure of the data by identifying the dimensions that are less effective [30]. The effectiveness of each dimension is inversely related to the values of lengthscales that can be determined after the model is trained. The RQ kernel function is defined as follows:

$$k_{RQ}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \left[1 + (\mathbf{x}_i - \mathbf{x}_j)^\top \frac{\mathbf{M}}{2\alpha}(\mathbf{x}_i - \mathbf{x}_j)\right]^{-\alpha}, \tag{8}$$

where $\sigma_f^2$ is the signal variance and $\mathbf{M} = \mathrm{diag}(\boldsymbol{l}^{-2})$ is a diagonal matrix containing the lengthscales $l > 0$ associated with each dimension. Finally, $\alpha$ is a positive parameter that controls the shape of the function by adjusting the weight of the lengthscales.

Determining the posterior covariance is dependent on kernel functions and the noise variance $\sigma_n^2$. In fact, a set of hyperparameters $\boldsymbol{\theta} = \{\sigma_f^2, \boldsymbol{l}, \alpha, \sigma_n^2\}$ defines the model, and model selection is performed by maximizing the marginal likelihood of observations given a set of hyperparameters

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\mathrm{argmax}} \ \log p(\mathbf{y}|X_{\mathcal{L}}, \boldsymbol{\theta}), \tag{9}$$

where $\boldsymbol{\theta}^*$ denotes the optimal values of hyperparameters [30]. The log marginal likelihood of noisy observations given the set of hyperparameters $\boldsymbol{\theta}$ is calculated as follows:

$$\log p(\mathbf{y}|X_{\mathcal{L}}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^\top \left(K_{\mathcal{L}\mathcal{L}} + \sigma_n^2 I\right)^{-1}\mathbf{y} - \frac{1}{2}\left|K_{\mathcal{L}\mathcal{L}} + \sigma_n^2 I\right| - \frac{n}{2}\log(2\pi), \tag{10}$$

where $n$ is the size of the training set.

Thus far, we have assumed that a set of particular buildings with their associated damage labels is available. However, in practice, training data pairs $(X, \mathbf{y})$ are scarce. In fact, in a post-event scenario, only a pool of unlabeled data points is available, and the survey team should first identify candidate buildings and then obtain the labels through damage inspection. In the following, we describe an efficient inspection framework that expedites the survey process by selecting a well-chosen training set while considering the logistics of post-earthquake field surveys.

## 3. Efficient post-earthquake survey planning using adaptive batch-mode active learning

Post-earthquake building inspections are crucial for making informed disaster management decisions but can be time-consuming, hindering prompt data acquisition through surveys. AL prioritizes inspections based on the potential usefulness of the collected data for training machine learning models, minimizing data acquisition and speeding up decision-making. However, existing AL methods for post-earthquake survey planning select one building at a time. This process, known as sequential active learning [23], selects queries in the feature space where the model is least certain about its predictions. Although effective, this approach only considers uncertainty in model predictions and does not account for the logistics of the surveys, such as geographical location and accessibility of the candidate buildings. Thus, sequential AL may select inaccessible buildings, leading to inefficient survey planning.

If we readily determine a set of buildings to be inspected, $\mathcal{Q}_0$, then we can improve efficiency by sequencing inspections such that the total travel time is minimized. If the set $\mathcal{Q}_0$ is not updated with incoming information from inspections, then the described process represents passive learning, which suffers from not leveraging the labels' information. Alternatively, if $\mathcal{Q}_0$ is a singleton, then the process falls back to sequential active learning, which leaves no flexibility to efficiently sequence inspections. Therefore, we hypothesize that there exists a trade-off between planning efficiency and learning performance in post-earthquake building damage inspection. Based on the hypothesis, we use the size of set $\mathcal{Q}_0$ as a proxy for the trade-off between the route planning efficiency and learning performance to propose an active learning framework, herein referred to as Adaptive Batch-Mode Active Learning (A-BMAL).

The design logic of A-BMAL is as follows:

1. Each day, a batch of to-be-labeled points $\mathcal{Q}_0$ will be determined to enable route planning.
2. $\mathcal{Q}_0$ should include points whose inspection will maximize the regression model's inference performance.
3. The size of $\mathcal{Q}_0$ should exceed the total number of buildings that the reconnaissance team can inspect daily within the given time constraint.
4. Increasing the size of $\mathcal{Q}_0$ provides greater flexibility for route planning, but making it too large could result in the inclusion of less relevant points.

Let $\mathcal{S}$ be the data pool from which candidate buildings are to be selected for inspection. If $\mathcal{L}$ and $\mathcal{U}$ denote the sets of labeled and unlabeled data points, such that $\mathcal{S} = \mathcal{L} \cup \mathcal{U}$, we aim to label a subset $\mathcal{Q}_0$ of $\mathcal{U}$, given information gained from the previously labeled data points in $\mathcal{L}$, to maximize the predictive performance of the GPR model. This problem is known as batch-mode active learning in the machine learning literature [24,31]. A simple heuristic is to select a batch containing data points that the model is least certain about or data points with the highest entropy [32]. However, it is likely that the model is uncertain about similar data points [33]. To maximize the utilization of the training samples, the diversity within each batch should be explicitly considered to avoid similar data points in the batch [24]. Let $\mathcal{U}_I$ be a subset of $\mathcal{U}$, which comprises informative data points. Informativeness measures the potential for a given sample to reduce the model's uncertainty [31]. Assuming the Euclidean distance as the metric, the batch-mode AL objective can be formulated as follows:

$$\mathcal{Q}_0^* = \underset{\mathcal{Q}_0 \subset \mathcal{U}_I \subset \mathcal{U}}{\operatorname{argmin}} \sum_{x_i \in \mathcal{Q}_0} \sum_{\substack{x_j \in \mathcal{U}_I \\ j \neq i}} \|x_i - x_j\|^2, \tag{11}$$

where $\mathcal{Q}_0^*$ comprises data points that minimize the total distance to other informative data points in the unlabeled data pool. Zhdanov [33] reformulated the aforementioned objective by employing $k$-means clustering and integrating weights that reflect the informativeness of individual data points for classification applications. To partition the unlabeled data $\mathcal{U}$ into a set of disjoint clusters, $C$, where $C = \{C_1, C_2, \ldots, C_{|\mathcal{Q}_0|}\}$, the weighted $k$-means objective is as follows:

$$C^* = \underset{C}{\operatorname{argmin}} \sum_{i=1}^{|\mathcal{Q}_0|} \sum_{x_j \in C_i} w_j \|x_j - \mu_i\|^2, \tag{12}$$

where $\mu_i = \frac{\sum_{x_j \in C_i} w_j x_j}{\sum w_j}$ denotes the centroid of cluster $C_i$, and $w_j$ is the corresponding weight for each data point $x_j$, which is proportionate with its informativeness.

In our proposed framework, we extend this formulation to GP regression by customizing the computation of $w_j$. Specifically, $w_j$ will be computed based on the AL framework using the information-theoretic criterion proposed in Krause et al. [34]. Let $Y_{\mathcal{U}}$ denote a set of random variables representing the predicted function values at unlabeled data points. The entropy of $Y_{\mathcal{U}}$ given a set of labeled data points is denoted as $H(Y_{\mathcal{U}}|Y_{\mathcal{L}})$. Ideally, a desirable design leads to minimizing entropy,

$$\mathcal{L}^* = \underset{\mathcal{L} \subset \mathcal{S}}{\operatorname{argmin}}\ H(Y_{\mathcal{U}}|Y_{\mathcal{L}}), \tag{13}$$

where $\mathcal{L}^*$ denotes the optimal set of queries that minimizes the entropy of predictions. However, identifying an optimal set was shown to be NP-hard [35]. Krause et al. [34], instead, proposed a greedy approach that selects points that myopically maximize gain in mutual information (MI) between the labeled and unlabeled data points:

$$x^* = \underset{x \in \mathcal{U}}{\operatorname{argmax}}\ H(Y_x|Y_{\mathcal{L}}) - H(Y_x|Y_{\bar{\mathcal{L}}}), \tag{14}$$

where $\bar{\mathcal{L}} = \mathcal{U} \setminus \{x\}$.

Rather than selecting the next query at $x^*$, we propose computing the MI criterion for all the data points in the unlabeled set $\mathcal{U}$ and assigning the weight $w$ proportional to the computed MI value associated with each data point. Therefore, for every $x \in \mathcal{U}$:

$$w \propto H(Y_x|Y_{\mathcal{L}}) - H(Y_x|Y_{\bar{\mathcal{L}}}) \tag{15}$$

where $w$ is the assigned weight according to Eq. (12). This objective biases the cluster centroids toward the data points that maximally increase the MI between labeled and unlabeled data points (i.e., informative points). To compute the weights, we use a result showing that the entropy of a Gaussian random variable is a monotonic function of its variance [36]:

$$H(Y_x|Y_{\mathcal{L}}) = \frac{1}{2} \log(2\pi e \sigma^2_{Y_x|Y_{\mathcal{L}}}). \tag{16}$$

Therefore, for all $x \in \mathcal{U}$ given a labeled set of data $\mathcal{L}$, the weights in Eq. (15) can be expressed as follows:

$$w = \frac{\sigma^2_{Y_x|Y_{\mathcal{L}}}}{\sigma^2_{Y_x|Y_{\bar{\mathcal{L}}}}} \tag{17}$$

where the posterior variance at data point $x$ given sets $\mathcal{L}$ and $\bar{\mathcal{L}}$ is calculated according to Eq. (7),

$$\frac{\sigma^2_{Y_x|Y_{\mathcal{L}}}}{\sigma^2_{Y_x|Y_{\bar{\mathcal{L}}}}} = \frac{k(x,x) - K_{x\mathcal{L}} \left(K_{\mathcal{L}\mathcal{L}} + \sigma_n^2 I\right)^{-1} K_{\mathcal{L}x}}{k(x,x) - K_{x\bar{\mathcal{L}}} \left(K_{\bar{\mathcal{L}}\bar{\mathcal{L}}} + \sigma_n^2 I\right)^{-1} K_{\bar{\mathcal{L}}x}}. \tag{18}$$

The $k$-means algorithm minimizes within-cluster $L^2$-norm of data and the corresponding centroid. As such, due to within-cluster similarity induced by the objective, the centroids can be representative of data points in the clusters. Further, due to the convexity of clusters [37], the centroids tend to be distant from each other, resulting in diversity among the samples within a batch. It has been shown by Wu [23], Demir et al. [38], and Xu et al. [39] that clustering-based sampling methods lead to representativeness and diversity of the samples. At the same time, the weights calculated based on the MI criterion incline the centroids toward more informative points. Therefore informativeness, representativeness, and diversity are considered in the AL framework. Since $k$-means suggests centroids that are not a member of $\mathcal{U}$, we select the closest point to the centroid (i.e., medoid) for label query.

### 3.1. Adaptive batch-mode active learning for earthquake reconnaissance surveys

Starting with $\mathcal{L} = \emptyset$, we initialize the model by labeling a randomly selected batch of data points and updating $\mathcal{L}$ accordingly. Then, for each unlabeled data point $x_i \in \mathcal{U}$ where $\mathcal{U} = \mathcal{S} \setminus \mathcal{L}$, the calculated weight $w_i$ is assigned as defined in Eq. (17), proportional to information gain if the data point $x_i$ is labeled. Due to the diminishing returns property of MI criterion [34], the amount of information-gain $w_i$ is larger in the first cycles of the AL process. The values of $w_i$ gradually decrease as the model gains more information about the underlying unknown function. To account for this effect and have interpretable weights, we normalize $\boldsymbol{w}$, as denoted by $\bar{\boldsymbol{w}}$ in Algorithm 1, line 6.

Next, the data points in $\mathcal{U}$ along with their corresponding weights $\bar{\boldsymbol{w}}$ are clustered using Eq. (12). The closest data points to centroids are then selected to form an initial batch $\mathcal{Q}_0$, where $|\mathcal{Q}_0| = \beta_0$. Considering a time constraint as a daily work-time limit, an integrated optimization module maximizes the number of queries within $\mathcal{Q}_0$ and eventually selects $\mathcal{Q}$ as a subset of $\mathcal{Q}_0$. $\mathcal{Q}$ represents the set of informative candidate buildings that are in close geographical distance to the surveying team. The set of labeled data points $\mathcal{L}$ is appended with the newly labeled set $\mathcal{Q}$, and the model is updated with the augmented set of labeled data accordingly. The loop continues until reaching a maximum number of inspections $B$ or satisfying a termination criterion.

The computational complexity of the algorithm involves operations in training the GPR model, route optimization algorithm, and AL loop. According to Eq. (18), calculating the MI criterion requires the inversion of covariance matrix $K_{\bar{\mathcal{L}}\bar{\mathcal{L}}}$. If $|\mathcal{U}| = N_{\mathcal{U}}$, the time complexity of the algorithm for the selection of queries is in $\mathcal{O}(\frac{B}{\beta} N_{\mathcal{U}}^4)$, where $\beta = |\mathcal{Q}|$. It should be noted that $\beta$ is adaptively changed according to the spatial distribution (i.e., geographical location) of data points. Compared with the greedy approach for sequential AL [34], the time complexity reduces by a factor of $\beta$ due to the batch-mode framework. However, this is still a considerable time complexity for large data sets (e.g., $N_{\mathcal{U}} > 10,000$). A simple heuristic to reduce the time complexity can be pre-filtering the initial data pool using clustering-based methods to limit the size of $\mathcal{S}$ and $\mathcal{U}$ accordingly so that $N_{\mathcal{U}}$ can directly be reduced. The choice for the size of the truncated data pool amounts to a trade-off between the optimality of the batch design and computation efficiency.

### 3.2. Termination criterion

In experiments, a test set with ground truth values is commonly used to track the model's predictive performance during the labeling process. However, in practical scenarios where unlabeled data is being annotated, a labeled test set is rarely available beforehand. A fixed labeling budget is typically used to terminate the labeling process. However, using a fixed labeling budget to terminate the AL process restricts the potential to fully benefit from AL's objective of reducing labeling costs through fewer yet more informative queries [40]. A fixed budget may result in either overexploitation of queries, leading to a highly accurate model that has converged with fewer training instances, or a naive model lacking sufficient predictive performance [41].

**Algorithm 1:** Adaptive batch-mode active learning (A-BMAL)

**Input:** Data pool $\mathcal{S}$, Labeled data points $\mathcal{L} = \emptyset$, Budget ($B$), Initial batch size ($\beta_0$)
**Output:** Model $\mathcal{M}$

1 $\mathcal{L} \leftarrow$ Initialize $\mathcal{M}$ by labeling a randomly selected batch;
2 **while** *Budget or termination criterion is not met* **do**
  `/* termination criterion according to Section 3.2 */`
3  **for** $x_i \in \mathcal{S} \setminus \mathcal{L}$ **do**
4  $w_i \propto H(Y_{x,i}|Y_{\mathcal{L}}) - H(Y_{x,i}|Y_{\bar{\mathcal{L}}})$;
5  **end**
6  $\bar{\boldsymbol{w}} \leftarrow$ Normalize $\boldsymbol{w}$;
7  Perform weighted $k$-means with $\beta_0$ clusters, $\bar{\boldsymbol{w}}$ being the associated weights for data points in $\mathcal{S} \setminus \mathcal{L}$; `/* weights are assigned according to Eq. (12) */`
8  $\mathcal{Q}_0 \leftarrow$ Find the closest data points to centroids;
9  $\mathcal{Q} \leftarrow$ route optimization module $\leftarrow \mathcal{Q}_0$; `/* Optimization objective and constraints according to Section 3.3 */`
10  $\mathcal{L} \leftarrow \mathcal{L} \cup \mathcal{Q}$; `/* update the set of labeled data */`
11  Train $\mathcal{M}$ using the data points in $\mathcal{L}$;
12 **end**

In the post-earthquake survey scenario, it is necessary that the inspection team have an estimation of the model's convergence, and ideally know when to terminate the inspection task. This enables expedited decision-making using a dependable model that requires fewer training instances. To this end, we propose a simple termination criterion based on the stability of the prediction confidence of the GPR model. Since GPR is a probabilistic model, in each training cycle it provides a notion of uncertainty about the predictions through a posterior multivariate Gaussian distribution according to Eq. (5). The prediction confidence for a set of uninspected buildings can be calculated considering the diagonal elements of the posterior covariance matrix $\Sigma_*$ (i.e., prediction variance).

Let $V_i$ be a variable that denotes the prediction variance on the $i$th day of inspection for a given set of unlabeled buildings. A simple approach is to compare the mean values for prediction variance as more buildings are inspected and the training data increase each day (i.e., $\bar{V}_i$, $\bar{V}_{i+1}$, etc.). The mean values, however, may not fully capture the variation in the distribution of prediction uncertainty. We use Jensen–Shannon (JS) distance [42] as a metric to quantify the similarity between the distributions of $V$ in consecutive days. Unlike Kullback–Leibler (KL) divergence [43,44], JS divergence is symmetric and bounded, making it a suitable measure when a predefined threshold is considered as a stopping criterion. We use base 2 logarithm such that $D_{JS} \in [0, 1]$, where $D_{JS} = 0$ implies similar distributions. We assume that convergence occurs provided the model shows stable confidence over its predictions. If $p_i$ denotes the probability distribution of $V_i$, we define the termination criterion as follows:

$$\begin{aligned} &D_{JS}(p_i \| p_{i-1}) < \varepsilon_t, \\ &D_{JS}(p_{i-1} \| p_{i-2}) < \varepsilon_t \end{aligned} \tag{19}$$

where $\varepsilon_t$ denotes a small positive value representing a threshold, and for any pair of probability distributions such as $q_1$ and $q_2$, $D_{JS}(q_1 \| q_2)$ is:

$$D_{JS}(q_1 \| q_2) = \sqrt{\frac{D_{KL}(q_1 \| \frac{q_1+q_2}{2}) + D_{KL}(q_2 \| \frac{q_1+q_2}{2})}{2}} \tag{20}$$

where $D_{KL}$ denotes KL divergence. According to Eq. (19), on the $i$th ($i > 2$) day of the survey, the convergence criterion compares the distribution of prediction uncertainty with those of the preceding two

days. If after training the model each day, the confidence of the model does not considerably improve, the convergence criterion is met, and the labeling process is terminated.

### 3.3. Orienteering problem statement and formulation

Efficient planning for inspections involves designing an optimal route for inspecting candidate buildings. These buildings, represented as nodes with different geographical locations, need to be visited within a specified time frame, typically a workday. We aim to determine routes that maximize the number of buildings visited while ensuring the trip duration remains within the allocated time constraint.

The inspection routing problem is modeled as an orienteering problem (OP), which is an integer programming model [45]. The problem involves a set of nodes denoted as $V = 0, 1, 2, \ldots, N$, where node 0 represents the depot where the inspection team starts and ends their trip. Each node $i \in V \setminus 0$ takes $e_i$ time to be inspected and can be visited at most once due to a limited time budget denoted by $T_{max}$. The travel time between any pair of nodes $(i, j)$ is denoted as $t_{ij}$. The objective of the OP is to find a route that visits a maximum number of nodes while adhering to the time limit of $T_{max}$.

To formulate the OP as an integer program, we define a decision variable:

$$x_{ij} = \begin{cases} 1, & \text{if node } i \text{ is followed by node } j \text{ in the route} \\ 0, & \text{otherwise.} \end{cases}$$

Also, we define an auxiliary variable $u_i \in \mathbb{Z}^+$ denoting the position of node $i$ in the route, which will be used to eliminate any subtours. With the above notations and decision variables, the mathematical model is presented below:

$$\max \sum_{i=1}^{N} \sum_{j=1}^{N} x_{ij} \tag{21}$$

$$\text{s.t.,} \sum_{j=1}^{N} x_{0j} = \sum_{i=1}^{N} x_{i0} = 1, \tag{22}$$

$$\sum_{i=0}^{N} x_{ik} = \sum_{j=0}^{N} x_{kj} \leq 1; \forall k = 1, \ldots, N, \tag{23}$$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} t_{ij} x_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{N} e_i x_{ij} \leq T_{max}, \tag{24}$$

$$1 \leq u_i \leq N; \forall i = 1, \ldots, N, \tag{25}$$

$$u_i - u_j + 1 \leq (N-1)\left(1 - x_{ij}\right); \forall i, j = 1, \ldots, N, \tag{26}$$

$$x_{ij} \in \{0, 1\}, u_i \in \{1, 2, \ldots, N\}; \forall i, j = 0, \ldots, N. \tag{27}$$

The objective of the optimization problem described by Eq. (21) is to maximize the number of visited nodes. The constraints in Eq. (22) ensure that the route starts and ends at node 0, while Eq. (23) guarantees that each node is visited at most once and the route is connected. Constraint Eq. (24) sets an upper bound on the total time spent on travel and inspection for the entire route to ensure that it does not exceed the time budget. Constraints Eq. (25) and (26) are subtour elimination constraints. Finally, Eq. (27) specifies the domain of the variables $x_{ij}$ and $u_i$.

This section proposes a heuristic approach to find feasible solutions for the OP, which is known to be an NP-hard problem. The approach firstly optimizes a route to visit all nodes in a predefined node set using an insertion algorithm and a 2-opt exchange procedure (see Kay [46] for details). The insertion algorithm is a commonly used heuristic algorithm for solving routing problems due to its simplicity and effectiveness in producing good solutions [47,48]. To further improve the quality of the solution, the 2-opt exchange procedure can also be applied in our work. After applying the insertion algorithm and 2-opt exchange, the resulting route $\mathcal{R}$ is evaluated by calculating the total travel time plus inspection time, denoted by $T_{\mathcal{R}}$. This value is then compared with the time budget $T_{max}$ to ensure that the solution satisfies the time constraint. If the time budget is exceeded, a drop-node operator is applied to sequentially remove nodes from the route. The nodes to be dropped are selected based on their contribution to the objective function, which is determined by the score collection process. Specifically, the contribution value of node $i$ is calculated as $\phi_i = \frac{1}{t_{pi} + t_{ik} - t_{pk}}$, where $p$ and $k$ are the previous and next nodes to node $i$ in the route, respectively. The drop-node operator is applied repeatedly until the time constraint is satisfied. The details of this proposed heuristic approach for solving routing problems are summarized in Algorithm 2.

**Algorithm 2:** A heuristic to solve the inspection routing problem

**Input:** node set $V$; travel time $t_{ij}$ between $(i, j)$; time constraint $T_{max}$
**Output:** A route $\mathcal{R}$ with the maximum number of buildings inspected

Run insertion algorithm for an initial route visiting every node $i \in N$
Run 2-opt exchange procedure for route improvement and obtain $\mathcal{R}$
**while** $T_{\mathcal{R}} > T_{max}$ **do**
    Find the node
    $$i^* = \underset{i \in \mathcal{R}}{\operatorname{argmin}} \frac{1}{t_{pi} + t_{ik} - t_{pk}}$$
    with the least value of $\phi_{i^*}$
    Update $\mathcal{R} \leftarrow \mathcal{R} \setminus \{i^*\}$ and apply 2-opt exchange procedure on $\mathcal{R}$
**end**

## 4. Regional earthquake simulation testbeds and learning approaches

The proposed framework is evaluated using both hypothetical and real-world earthquake events. This section outlines the characteristics of the testbeds and enumerates the influential features to infer regional-scale response in a post-earthquake setting. Additionally, alternative passive sampling and active learning algorithms for earthquake survey planning are discussed.

### 4.1. San Francisco, CA testbed

San Francisco Bay Area is surrounded by San Andreas and Hayward faults, both having the precedent of generating major seismic events. With a population of over 7 million people and diverse buildings constructed over different years, the Bay Area is considered a densely inhabited and potentially seismically vulnerable region [49]. As a region with the highest population density, San Francisco County is herein considered for a post-earthquake reconnaissance survey.

The proposed method is evaluated using the regional Workflow for Hazard And Loss Estimation (rWHALE) [50]. rWHALE is a modular tool that estimates building loss on a regional scale by simulating various hazards. A hypothetical $M_w$ 7.0 earthquake is considered along the Hayward fault with the epicenter being in Oakland, CA. The simulation is based on a deterministic physics-based model using SW4 finite difference program [51] developed by Lawrence Livermore National Laboratory [52]. The simulation results in a grid of seismograms across the region at each 2 km, which amounts to 39 seismograms evenly distributed in the region of interest. Then, each building in the region is assigned the nearest seismogram using the k-dimensional tree algorithm [50]. The assigned GM time histories are applied to MDOF shear models proposed by Lu and Guan [53]. It should be noted the primary horizontal component is considered since the shear model has the same structural properties in $x$ and $y$ directions. Each model is defined based on the building information database provided by NHERI SimCenter [54] and UrbanSim [55], containing general attributes of

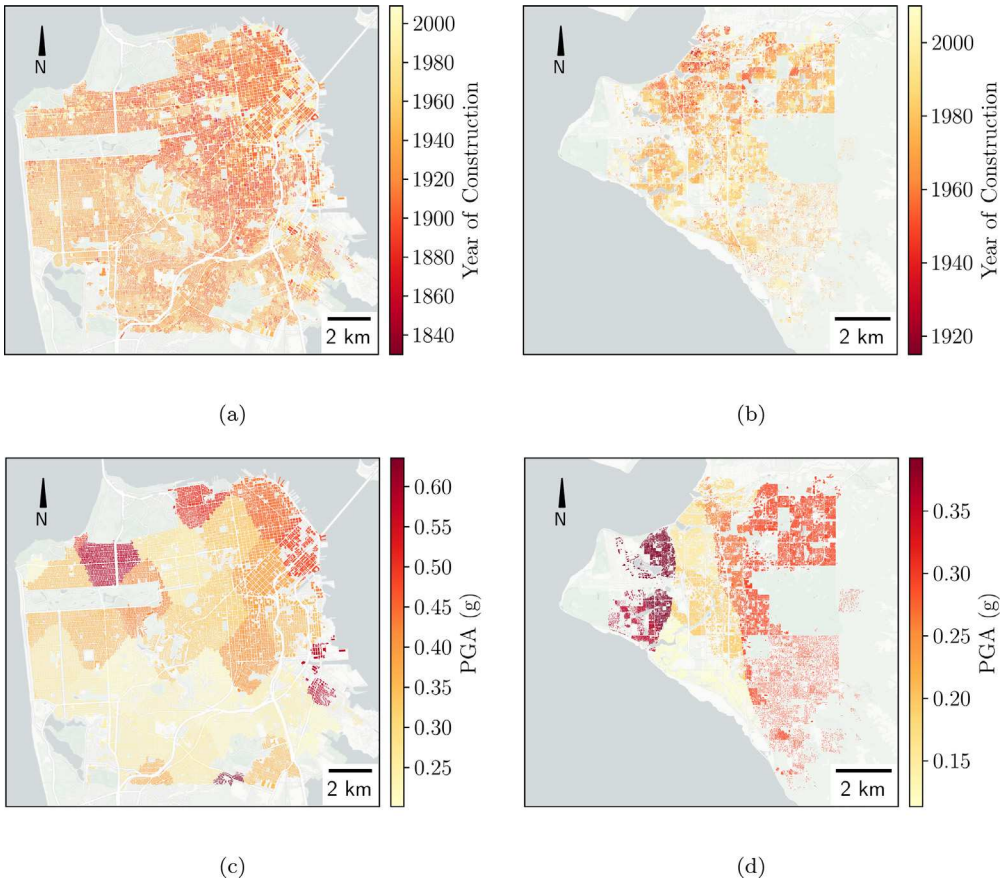



**Fig. 2.** Geospatial distribution of building attributes and ground motion intensities. Distribution of buildings' construction year for the (a) San Francisco and (b) Anchorage testbeds. Distribution of the assigned ground motion intensities based on PGA for the (c) San Francisco and (d) Anchorage testbeds. PGA refers to peak ground acceleration.

the buildings (e.g., number of stories, occupancy type, structure type, construction year, etc.). The uncertainty in modeling is considered by defining random variables for the story height, initial stiffness, and damping ratio.

The engineering demand parameters (EDP) such as maximum acceleration, inter-story drift ratio, and residual drift are calculated through time history analysis. The economic loss estimation is then performed using the proposed method by Zeng et al. [56] based on FEMA-P58 [57]. The loss ratio is considered as repair to replacement costs ratio, where unity denotes complete loss. In addition, red tags specify whether a building is safe for re-occupancy [58]. This damage measure can inform owners and residents about the safety of buildings. Based on the possible EDPs and the associated damage states, each building is assigned a value between 0 and 1, representing the probability of receiving a red tag.

### 4.2. Anchorage, AK testbed

On November 30, 2018, a powerful $M_w$ 7.1 earthquake took place in South-central Alaska, followed by an $M_w$ 5.8 aftershock and a tsunami warning. The earthquake's epicenter was located 12 km north of Anchorage and 19 km west of the Eagle River. This seismic event affected around 400,000 residents in the Anchorage Metropolitan Area, causing extensive damage to roads, infrastructure, and older buildings that were constructed before building codes were established [59].

To further evaluate the proposed method in a real-world scenario, we use the Anchorage testbed under the 2018 $M_w$ 7.1 earthquake event. Building information is extracted from the DesignSafe data repository [60], and GMs are obtained from seismograph stations available at the Center for Engineering Strong-Motion Data (CESMD) [61]. A set of 12 seismograms is used to generate seismic excitation across the region. The rWHALE workflow is then used to calculate the regional loss [50]. Each building is assigned a seismic excitation based on the nearest GM record. The horizontal component in the east–west direction (HNE) is then applied to shear building models. In line with our prior testbed, we employed consistent damage indicators (i.e., the loss ratio and red tag probability) to assess the regional-level damage sustained by each building. Therefore, using the workflow, the damage indices across the region can be computed given the buildings' information and the GM characteristics.

### 4.3. Dataset characteristics

Prediction of damage through a supervised learning model involves training the model with pairs of inputs and outputs. Sheibani and Ou [13] showed that a combination of readily obtainable building information attributes and certain GM indices can be used to accurately represent building inventory information for regional damage and loss inference using a GPR model. Several intensity measures, including spectral acceleration at the fundamental period $Sa(T_1)$, Arias intensity [62], Fajfar intensity [63], and spectral intensity [64] were identified as highly correlated with structural damage by Sheibani and Ou [13] and Ye et al. [64]. Furthermore, the statistical characteristics of GMs in the time domain can be represented by the interquartile range (IQR) and kurtosis of GM signals. Table 1 lists the set of 12 influential features comprising the buildings' attributes and GM characteristics. On the other hand, loss ratio and red tag probability are both dependent on the set of features and are considered outputs or labels.

Fig. 2 shows the distribution of building construction years and peak ground acceleration (PGA) as a fraction of the gravitational acceleration. Fig. 3 illustrates the building heights based on their number of

**Table 1**
The list of features representing building information and GM indices.

| Building information attributes | Ground motion indices |
|---|---|
| Construction year | $Sa(T_1)$ |
| No. of stories | Arias |
| Floor area | Fajfar |
| Occupancy type | Spectral intensity |
| Latitude | Signal IQR |
| Longitude | Signal kurtosis |

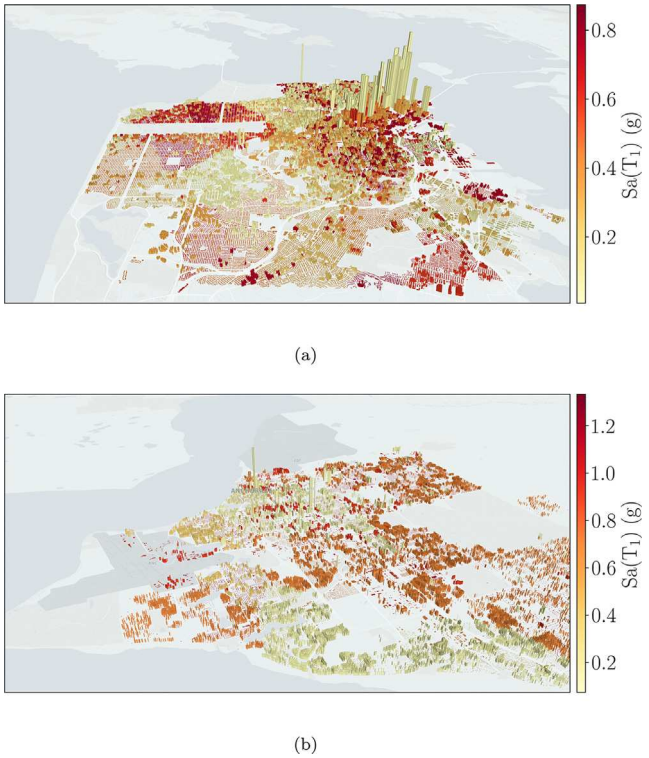



**Fig. 3.** Spectral acceleration $Sa(T_1)$, with building height proportional to the number of stories. (a) San Francisco testbed (b) Anchorage testbed.

stories and spectral acceleration as an intensity measure that accounts for the characteristics of GMs and structures alike.

Fig. 2 shows a noticeably broader range of construction years for the San Francisco testbed. Similarly, the map indicates more dispersion in construction years for San Francisco compared to Anchorage, suggesting a diverse range of buildings complying with different design codes. Additionally, the San Francisco testbed, with over three times the number of seismograms, showcases greater diversity in the assigned PGA for each building. Likewise, the increased heterogeneity is shown in Fig. 3, with the San Francisco testbed displaying a wider range of building heights and spectral acceleration. These testbeds exemplify distinct scenarios: the San Francisco testbed reflects a complex building stock with a heterogeneous data structure, while the Anchorage testbed represents a more homogeneous data structure. Examining both testbeds enables a better comparison to evaluate the robustness of the proposed methodology in different scenarios.

As for the data preparation step, 20k buildings are randomly selected from the testbed. The data is then split into two disjoint sets $\mathcal{S}$ and $\mathcal{S}_*$, each containing 10k buildings. $\mathcal{S}$ denotes the data pool from which the candidate buildings are selected for inspection. The calculated damage indices based on loss assessment are considered as the observed damage that is available after building inspections, thereby representing the labels in $\mathcal{S}$. $\mathcal{S}_*$ denotes a held-out test set by which the predictive performance of the model is evaluated. To reduce the dependency of prediction results on a specific data set, 80% of $\mathcal{S}$ ($\mathcal{S}_{80}$) is randomly selected to generate 64 different realizations of the data pool. In each realization, after the model is trained with the data in $\mathcal{S}_{80}$, the predictions of damage measures for buildings in $\mathcal{S}_*$ are compared against the corresponding test labels (i.e., ground truth). To rule out the effect of different dimensions, input data in $\mathcal{S}_{80}$ is normalized. The occupancy type is categorical data, with each category treated as a one-hot encoded feature. In the GPR formulation, since a noisy observation of the underlying function has been considered, as detailed in Section 2, Gaussian white noise is added to the unknown function $f(\cdot)$; however, it should be noted that the labels (loss ratio and red tag probability) are bounded within $[0, 1]$, and it is not reasonable to add Gaussian noise when the training labels are bounded [65]. As such, a transformation using probit function $\Phi^{-1}(\cdot)$ is applied to training labels, and the model is trained in warped space. It should be noted that the inverse of the transformation is later applied to the predicted values to calculate the error measure and evaluate the prediction performance of the model.

### 4.4. Learning approaches

This study compares different algorithms representing passive sampling and AL methods with the proposed framework. Assumptions about inspection time, travel speed, daily work-time limit, and other factors are made to account for the logistics of a post-earthquake survey. Some assumptions are based on current practice, while others simplify the problem formulation. The following assumptions apply to all sampling approaches:

- Based on the ATC-20-1 procedures [66], the required time for conducting a rapid evaluation is 30 min for each building [58].
- An average travel speed of 30 km/h is assumed for the survey team.
- Daily work-time limit is considered to be 10 hours—this includes the time the survey team spends inspecting the candidate buildings and traveling to reach different locations.
- A fixed location, also known as depot location, is assumed from which the survey team starts the inspections, and at the end of each workday, the survey team returns to the same location.
- The maximum budget for inspections, denoted as $B$, is set to 300. The training process will continue until the stopping criterion Eq. (19) is met, or until the budget is exhausted.

As a baseline method, random sampling of candidate buildings is considered. Since budget $B$ is the maximum number of inspections for the survey, a set containing $B$ buildings is randomly selected from $\mathcal{S}_{80}$. Next, the OP heuristic is applied to maximize the number of inspected buildings within the allowed daily time limit. For each day, a list of to-be-inspected buildings is returned by grouping the randomly selected buildings at a close geographical distance. This sampling approach is hereafter referred to as Random-OP.

As a more systematic sampling strategy, clustering the input data is performed using the $k$-means algorithm. The density of data points varies across the input domain; as a result, each cluster may have a different number of members, and thus, each medoid is representative of a different number of data points. We partition the data pool into $B$ clusters and greedily select the buildings that are representative of larger clusters. This method is hereafter referred to as Greedy Representative Sampling (GRS).

GRS, however, does not account for the optimality of routes since the candidate buildings are prioritized based on the size of the clusters. We alter this order of sampling by applying the routing objective on the same $B$ candidate buildings (i.e., the selected medoids). This way, each day of the survey, the representative buildings are instead sequenced such that the number of daily inspections is maximized. Thus, the order of inspection of buildings is subject to change and is no longer based on

the cluster size. We refer to this sampling approach as Representative Sampling with the OP algorithm (RS-OP).

Further, as the state-of-the-art AL method for post-earthquake survey planning, the sequential design AL algorithm proposed in Sheibani and Ou [26] is considered. This framework, also known as mutual information with adaptive local kernels (MI-ALK), sequentially selects the candidate buildings according to the MI criterion Eq. (14) and exploits the locality of kernel functions to adaptively update the MI score in the neighborhood of the previously selected queries. Since labeling a data point leads to updating the information in the neighborhood of that data point, the costly calculations of the MI criterion should be limited to those data points in each iteration. The neighborhood of an inspected building is determined according to the kernel function and a predefined threshold $\varepsilon$. Only those data points in $\mathcal{U}$ that are highly correlated with the current query (i.e., correlations greater than $\varepsilon$) are used to calculate the MI criterion. In case there exist numerous data points that are highly correlated with the current query, an upper limit $d$ is set to control the number of data points for which the MI criterion is to be calculated. Further details about this approximation method can be found in Sheibani and Ou [26] and Krause et al. [34]. In this manuscript, we consider $\varepsilon = 0.9$ and $d = 250$. As a result, the time complexity is reduced from $\mathcal{O}(BN_{\mathcal{U}}^4)$ to $\mathcal{O}(BN_{\mathcal{U}}d^3)$, where $N_{\mathcal{U}}$ denotes the number of uninspected buildings in the data pool $S_{80}$.

Lastly, the proposed method (A-BMAL) suggests a batch containing $\beta_0$ candidate buildings for each day of inspection. Considering the depot location, the locations of buildings in the suggested batch, and the 10-hr daily time constraint, the routing module identifies a subset of the batch such that each day the number of inspections is maximized. As the main objective of A-BMAL is providing extra flexibility while maintaining the informativeness of the candidate buildings, $\beta_0$ should ideally be larger than the number of buildings that the team is capable of inspecting each day. The capacity of the daily inspections is dependent on numerous factors such as travel speed, the density of buildings in the region of interest, the distribution of informative buildings, and so forth. As such, a judicious choice of $\beta_0$ is based on domain knowledge. In this study, considering the aforementioned assumption, we assume $\beta_0$ to be 30, and the proposed algorithm is referred to as A-BMAL$_{30}$.

## 5. Results and discussion on learning performance

This section evaluates the proposed method for post-earthquake regional damage inference, comparing it to passive and sequential active learning algorithms. The evaluation measures sampling effectiveness and prediction accuracy. In this section, the termination criterion is not applied so that each algorithm can reach the full budget of 300 inspections for a complete comparison.

### 5.1. Sampling diversity and representativeness

To ensure accurate inference results, it is important to consider a diverse and representative set of candidate buildings for inspection that are reflective of the broader population of buildings in the region of interest. To assess the diversity and representativeness of the queried data points, we have defined two measures that are monitored throughout the training process.

Diversity accounts for the coverage of the training data across the input domain. As more training data are gradually obtained, the diversity of the training set evolves. To examine this evolution, we measure the average Euclidean distance between training and testing data points. Let $x_i$ and $x_j$ be a pair of data points in the test set $S_*$ and training set (i.e., labeled data points) $\mathcal{L}$, respectively. The average distance for the testing point $x_i \in S_*$ to the training data points in the set $\mathcal{L}$ is calculated as follows:

$$\bar{d}(x_i) = \frac{1}{|\mathcal{L}|}\sum_{j=1}^{|\mathcal{L}|}\|x_i - x_j\| \tag{28}$$

where $\|\cdot\|$ denotes the Euclidean distance between each pair of data points, and $|\mathcal{L}|$ is the number of labeled data points. As such, for every testing data point $x_i$, $\bar{d}(x_i)$ indicates the average distance to a given training set $\mathcal{L}$. If more testing data points find the training set in close proximity, it can be inferred that the training data has better coverage with respect to the test data, suggesting a higher diversity. Conversely, if training data is concentrated in the input domain (limited diversity), the majority of testing data finds the training set at greater distances $\bar{d}(\cdot)$. If $\bar{d}'(\cdot)$ denotes the normalized values of $\bar{d}(x_i)$ for all $x_i \in S_*$, to examine the distribution of the calculated $\bar{d}'(\cdot)$ across $S_*$, we define $X_D = 1 - \bar{d}'(\cdot)$ such that $X_D \in [0,1]$. $X_D$ is a variable that describes the normalized distance of each test data point to a given training set. Large values of $X_D$ imply that testing points find the training set at a close distance. If the distribution of $X_D$ is examined across $S_*$ and the majority of test data show higher values of $X_D$, it can be concluded that the training set has more diversity when compared to a distribution where the test data have a high concentration of lower values of $X_D$. Therefore, the diversity of the training set $\mathcal{L}$, can be quantified based on the distribution of $X_D$. The diversity measure $D_{\mathcal{L}}$ for a given training set $\mathcal{L}$ is formulated as follows:

$$D_{\mathcal{L}} = \frac{1}{|S_*|}\sum_{i=1}^{|S_*|} X_{D,i} \tag{29}$$

where $|S_*|$ is the size of the test set.

On the other hand, the kernel function in GPR encodes the correlation between the function values of a pair of input data points. A higher correlation between the testing points and the training set indicates the testing points are better represented in the training set. The average correlation between each test data point $x_i \in S_*$ and the training set $\mathcal{L}$ can be measured as follows:

$$\bar{r}(x_i) = \frac{1}{|\mathcal{L}|}\sum_{j=1}^{|\mathcal{L}|}\bar{k}(x_i, x_j) \tag{30}$$

where $\bar{k}(\cdot,\cdot) = k(\cdot,\cdot)/\sigma_f^2$ represents the correlation function of a pair of input data points. We define the variable $X_R$ to account for various possible values of $\bar{r}(\cdot)$, such that $X_R \in [0,1)$. As with the diversity measure, a higher concentration of test data at larger $X_R$ implies the test data is better represented by the training set. The measure $R_{\mathcal{L}}$ is defined to quantify the representativeness of the training set $\mathcal{L}$ as follows:

$$R_{\mathcal{L}} = \frac{1}{|S_*|}\sum_{i=1}^{|S_*|} X_{R,i}. \tag{31}$$

To illustrate the density of test data for different values of $X_D$ and $X_R$, a joint probability density function (PDF) is used. If $X_D$ and $X_R$ are divided into equidistant bins with widths $b_D$ and $b_R$, a joint PDF of $X_D$ and $X_R$ can be formulated as:

$$f_{D,R}(d,r) = \frac{\sum_{i=1}^{|S_*|}\delta_i(d,r)}{|S_*|\, b_D b_R} \tag{32}$$

where $\delta(d,r)$ is assigned based on an indicator function that is 1 for $i$th test point if it falls into the bin associated with the joint outcome $(d,r)$ and 0 otherwise; $b_D$ and $b_R$ are both set to be 0.01. Therefore, the joint PDF of the test data serves as a two-dimensional indicator for evaluating sampling effectiveness.

Fig. 4 and Fig. 5 illustrate the sampling performance evolution of each learning algorithm, for the San Francisco and Anchorage testbeds, respectively. Each row represents an algorithm, and the columns depict the algorithm's progress as more training data is acquired until reaching the budget of 300 inspections. The joint PDFs are computed as weighted averages of 64 realizations. The weights are assigned based on the cumulative number of inspected buildings up to a given day, considering that each realization may cover a different number of buildings per workday.

Comparing the columns in Figs. 4 and 5, as more training data is added, the density $f_{D,R}$ shifts toward the upper right corner, indicating

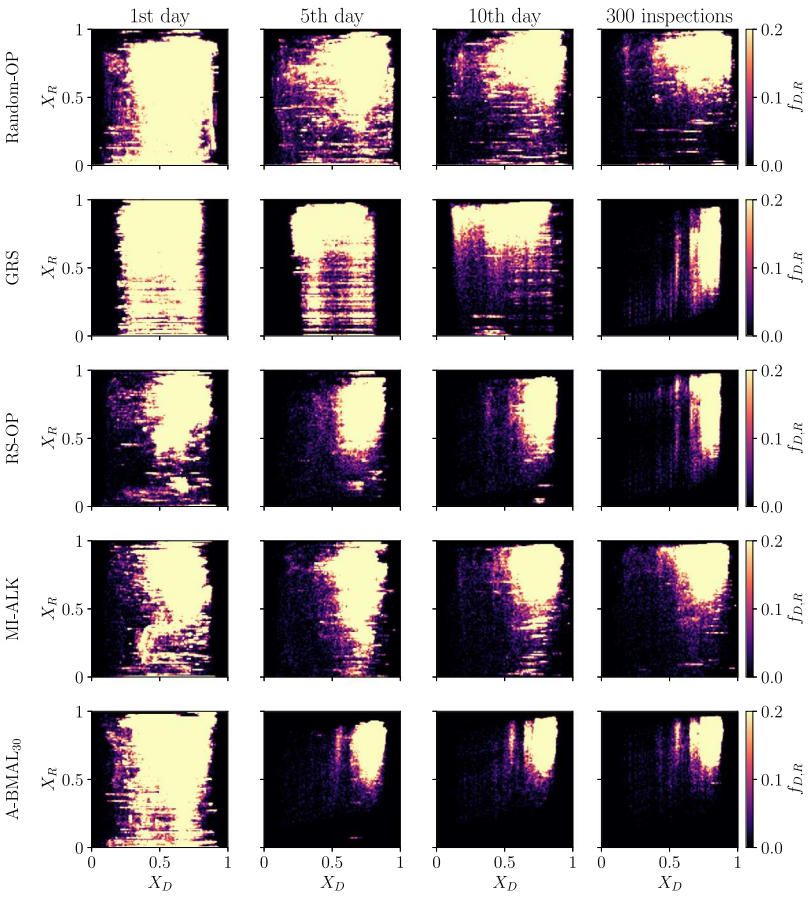



**Fig. 4.** Evolution of joint probability density of $X_D$ and $X_R$ as building inspections progress for loss ratio inference across the San Francisco testbed.

an increased likelihood of finding more diverse and representative training data. However, Random-OP exhibits a slower density shift compared to other algorithms, maintaining higher density at small $X_D$ values even after 300 inspections. This implies limited diversity in the training set with respect to the testing data. Moreover, Random-OP shows a high density ($f_{D,R} > 0.1$) of testing data at $X_R < 0.2$ after 300 inspections, suggesting the presence of training data points that are weakly correlated with testing data.

Compared to Random-OP, GRS and RS-OP exhibit higher densities at larger values of $X_D$ in the last column of Figs. 4 and 5, indicating more diverse final training sets. This is expected since GRS and RS-OP employ clustering to determine the training sets. In the initial stages of the survey (first two columns in Figs. 4 and 5), GRS samples training sets that are further from the majority of the testing data, resulting in a higher density of testing data at $X_D < 0.5$ compared to RS-OP. This suggests that training samples from larger clusters are concentrated and exhibit limited diversity with respect to the test data. However, the distribution gradually becomes more diverse after the 10th day. In contrast, RS-OP consistently shifts the density toward higher values of $X_D$ as more training data is acquired.

The MI-ALK approach sequentially selects buildings that maximize the MI criterion. MI-ALK demonstrates quicker convergence of diversity in training data points compared to random sampling, as indicated by darker regions with lower $f_{D,R}$ values for smaller $X_D$ values. However, clustering-based methods eventually achieve more diverse training sets. This highlights the effectiveness of clustering-based approaches in preventing excessive similarity among data points, making them suitable for avoiding similarity within a batch in the batch-mode AL framework.

A-BMAL$_{30}$ combines clustering-based and information-based methods. In the last row of Figs. 4 and 5, by the 5th day of inspection, densities shift toward higher values of $X_D$, indicating A-BMAL$_{30}$ has identified a diverse set of training data. Similarly, augmenting the training set gradually shifts densities toward higher values of $X_R$. The absence or low density of testing points at $X_R < 0.5$ after the 10th day confirms that A-BMAL$_{30}$ efficiently selects training samples that strongly correlate with the test data and better represent the underlying data.

Table 2 shows diversity ($D$) and representativeness ($R$) measures for the training sets as more data is acquired. Each algorithm corresponds with 2 rows of data where the upper row denotes the calculated measures for loss ratio, and the lower row denotes the scores for red tag probability. A-BMAL$_{30}$ achieves the highest diversity value in the initial phase (5th day) and maintains the highest diversity throughout the survey (300 inspections) compared to other algorithms. Similarly, an early convergence pattern is observed in the representativeness measure. The calculated measures, along with the small dispersions in the joint PDF, suggest that the proposed method effectively identifies diverse and representative samples in the initial phase of the survey.

In the Anchorage testbed, it is worth noting that GRS achieves the highest representativeness score values by reaching $R$-score of up to 0.95 as early as the 10th day, as specifically underscored in Table 2. This indicates that exclusively selecting the largest clusters helps create a training set that effectively represents the testing data. This suggests a concentration of data points in the multidimensional feature space, where a high representativeness score predominantly relies on sampling larger clusters. In fact, one can observe $R$-score reduces after the 10th day by adding the smaller clusters. This scenario implies a simple structure of the data as a result of a homogeneous building stock in the Anchorage testbed, especially when compared to a more complex and heterogeneous data set such as the San Francisco building stock, which encompasses a greater diversity of building occupancy types, various numbers of stories, and other attributes.

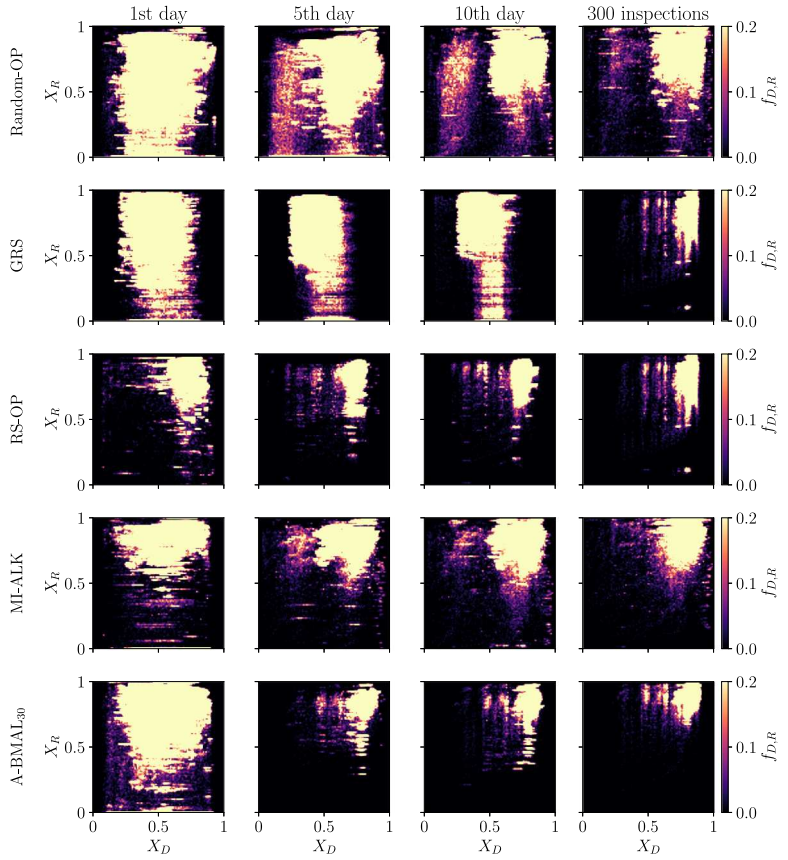



**Fig. 5.** Evolution of joint probability density of $X_D$ and $X_R$ as building inspections progress for loss ratio inference across the Anchorage testbed.

**Table 2**
Measures of diversity and representativeness for the considered algorithms as building inspections progress. Each algorithm is represented by two rows: the upper row displays the sampling measures for loss ratio, while the lower row pertains to red tag probability.

| Testbed | | San Francisco | | | | Anchorage | | | |
|---|---|---|---|---|---|---|---|---|---|
| Measure | Algorithm | 1st day | 5th day | 10th day | 300 inspections | 1st day | 5th day | 10th day | 300 inspections |
| *D* | Random-OP | 0.63 ± 0.10<br>0.62 ± 0.10 | 0.72 ± 0.09<br>0.72 ± 0.09 | 0.76 ± 0.08<br>0.76 ± 0.08 | 0.80 ± 0.07<br>0.79 ± 0.07 | 0.52 ± 0.10<br>0.55 ± 0.10 | 0.64 ± 0.08<br>0.68 ± 0.08 | 0.72 ± 0.07<br>0.72 ± 0.07 | 0.80 ± 0.05<br>0.81 ± 0.05 |
| | GRS | 0.55 ± 0.13<br>0.56 ± 0.14 | 0.51 ± 0.13<br>0.51 ± 0.13 | 0.65 ± 0.09<br>0.65 ± 0.10 | 0.82 ± 0.03<br>0.82 ± 0.03 | 0.47 ± 0.1<br>0.47 ± 0.10 | 0.41 ± 0.07<br>0.41 ± 0.07 | 0.39 ± 0.06<br>0.39 ± 0.06 | 0.83 ± 0.03<br>0.83 ± 0.03 |
| | RS-OP | 0.67 ± 0.05<br>0.67 ± 0.05 | 0.75 ± 0.04<br>0.75 ± 0.04 | 0.79 ± 0.04<br>0.79 ± 0.04 | 0.82 ± 0.03<br>0.82 ± 0.03 | 0.72 ± 0.04<br>0.70 ± 0.04 | 0.75 ± 0.03<br>0.75 ± 0.03 | 0.76 ± 0.03<br>0.76 ± 0.03 | 0.83 ± 0.03<br>0.83 ± 0.03 |
| | MI-ALK | 0.65 ± 0.05<br>0.66 ± 0.05 | 0.73 ± 0.05<br>0.74 ± 0.05 | 0.78 ± 0.05<br>0.77 ± 0.05 | 0.81 ± 0.05<br>0.81 ± 0.05 | 0.65 ± 0.06<br>0.68 ± 0.05 | 0.75 ± 0.05<br>0.76 ± 0.04 | 0.78 ± 0.05<br>0.79 ± 0.04 | 0.82 ± 0.04<br>0.82 ± 0.04 |
| | A-BMAL$_{30}$ | 0.63 ± 0.10<br>0.61 ± 0.10 | **0.81 ± 0.03**<br>**0.81 ± 0.03** | **0.82 ± 0.03**<br>**0.82 ± 0.03** | **0.82 ± 0.03**<br>**0.83 ± 0.03** | 0.56 ± 0.09<br>0.54 ± 0.09 | **0.82 ± 0.03**<br>**0.82 ± 0.03** | **0.83 ± 0.03**<br>**0.83 ± 0.03** | **0.84 ± 0.03**<br>**0.84 ± 0.03** |
| *R* | Random-OP | 0.74 ± 0.15<br>0.30 ± 0.10 | 0.78 ± 0.08<br>0.46 ± 0.09 | 0.82 ± 0.06<br>0.64 ± 0.07 | 0.84 ± 0.04<br>0.83 ± 0.03 | 0.55 ± 0.14<br>0.49 ± 0.10 | 0.68 ± 0.10<br>0.68 ± 0.05 | 0.77 ± 0.08<br>0.78 ± 0.06 | 0.85 ± 0.06<br>0.80 ± 0.05 |
| | GRS | 0.74 ± 0.17<br>0.67 ± 0.17 | 0.79 ± 0.12<br>0.66 ± 0.08 | 0.82 ± 0.07<br>0.77 ± 0.09 | 0.75 ± 0.05<br>0.83 ± 0.03 | 0.72 ± 0.14<br>0.72 ± 0.08 | 0.75 ± 0.11<br>0.82 ± 0.05 | 0.82 ± 0.09<br>0.95 ± 0.01 | 0.85 ± 0.03<br>0.81 ± 0.02 |
| | RS-OP | 0.75 ± 0.06<br>0.43 ± 0.07 | 0.77 ± 0.05<br>0.75 ± 0.03 | 0.80 ± 0.04<br>0.84 ± 0.02 | 0.84 ± 0.03<br>0.87 ± 0.02 | 0.82 ± 0.03<br>0.72 ± 0.05 | 0.83 ± 0.03<br>0.79 ± 0.03 | 0.87 ± 0.02<br>0.81 ± 0.02 | 0.85 ± 0.03<br>0.81 ± 0.02 |
| | MI-ALK | 0.75 ± 0.08<br>0.50 ± 0.09 | 0.76 ± 0.06<br>0.66 ± 0.04 | 0.83 ± 0.05<br>0.81 ± 0.03 | 0.88 ± 0.04<br>0.90 ± 0.02 | 0.82 ± 0.07<br>0.75 ± 0.05 | 0.84 ± 0.04<br>0.75 ± 0.04 | 0.84 ± 0.05<br>0.79 ± 0.03 | 0.90 ± 0.04<br>0.86 ± 0.01 |
| | A-BMAL$_{30}$ | 0.73 ± 0.16<br>0.60 ± 0.14 | **0.81 ± 0.03**<br>**0.72 ± 0.05** | **0.83 ± 0.03**<br>**0.85 ± 0.02** | **0.86 ± 0.03**<br>**0.90 ± 0.01** | 0.76 ± 0.13<br>0.78 ± 0.08 | **0.86 ± 0.03**<br>**0.79 ± 0.03** | **0.86 ± 0.03**<br>**0.80 ± 0.02** | **0.91 ± 0.02**<br>**0.83 ± 0.02** |

### 5.2. Predictive performance

To assess the model's predictive performance, various measures evaluate prediction accuracy and quantify prediction uncertainty. The Standardized Mean Squared Error (SMSE) measures the deviation between the predicted mean values of the unknown function and the test labels (ground truth). SMSE is calculated as follows:

$$\mathrm{SMSE} = \frac{1}{N}\frac{\sum_{i=1}^{N}(y_{*,i}-\mu_{*,i})^2}{\mathrm{Var}(\mathbf{y}_*)} \tag{33}$$

where $y_{*,i}$ is the label value for the $i$th test point in $S_*$, $\mu_{*,i}$ is the prediction at $i$th test point according to Eq. (6), and $N$ is the size of

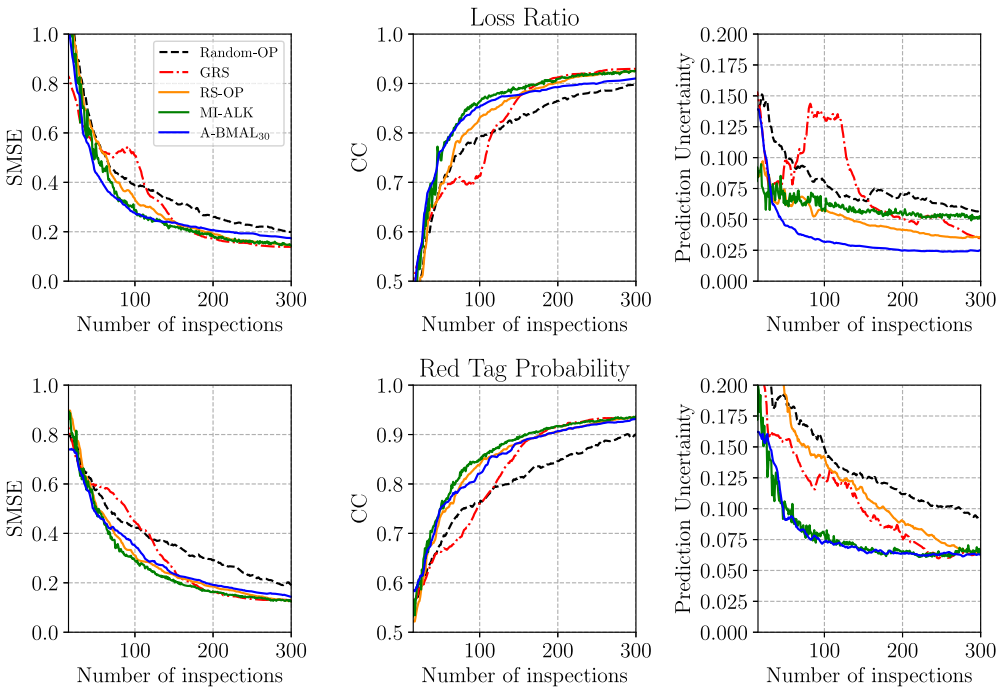



**Fig. 6.** Predictive performance comparison for the San Francisco testbed.

the test set $\mathcal{S}_*$. Besides, to standardize the metric, the total value of the residuals is divided by the variance of the test data labels $\mathbf{y}_*$.

The correlation between the test data labels and predictions is measured as another learning performance metric:

$$\mathrm{CC} = \frac{\sum_{i=1}^{N}(y_{*,i} - \bar{y}_*)(\mu_{*,i} - \bar{\mu}_*)}{\sqrt{\sum_{i=1}^{N}(y_{*,i} - \bar{y}_*)^2 \sum_{i=1}^{N}(\mu_{*,i} - \bar{\mu}_*)^2}} \tag{34}$$

where $\bar{y}_*$ and $\bar{\mu}_*$ are the mean values for test data labels and predictions, respectively. A higher correlation coefficient (CC) between the test data labels and the prediction values indicates the model better captures the trend in the underlying data, and thus it implies a superior learning performance.

As a probabilistic model, GPR yields a posterior multivariate Gaussian distribution. The posterior covariance $\Sigma_*$ in Eq. (7) captures the prediction uncertainty. The prediction uncertainty for a set of testing data points can be quantified by calculating the mean prediction variance across the test set $\mathcal{S}_*$ given training data set $\mathcal{L}$:

$$\frac{1}{N}\sum_{i=1}^{N}\left[\sigma_f^2 - K_{x_i\mathcal{L}}\left(K_{\mathcal{L}\mathcal{L}} + \sigma_n^2 I\right)^{-1} K_{\mathcal{L}x_i}\right] \tag{35}$$

where $x_i$ represents the $i$th testing point in $\mathcal{S}_*$, and $\sigma_f^2$ and $\sigma_n^2$ denote the signal and noise variances, as specified in Eq. (9).

Fig. 6 and Fig. 7 show predictive performance for loss ratio and red tag probability in the San Francisco and Anchorage testbeds, respectively. Results are median values from 64 realizations. More training data decreases SMSE and increases CC, enhancing predictive accuracy. However, Random-OP converges slower than other algorithms. For example, in San Francisco with 150 training samples, Random-OP shows approximately 75% and 90% higher SMSE for predicting loss ratio and red tag probability, respectively, compared to other algorithms. This trend extends to prediction uncertainty, where Random-OP's uncertainty reduction is more limited than other methods. Thus, random selection of candidate buildings requires more inspections to achieve comparable accuracy to other sampling approaches.

Regarding passive sampling methods, GRS and RS-OP achieve similar levels of SMSE, CC, and prediction uncertainty at 300 training data. However, their performance differs before reaching 300 inspections. GRS samples buildings from larger clusters, while RS-OP prioritizes inspections to maximize the number of daily inspections. Although the final performance is the same, RS-OP tends to achieve better predictive performance with smaller training data sets. In predicting the loss ratio in both testbeds, for instance, GRS exhibits significantly higher SMSE and lower CC values compared to RS-OP at 100 training samples. This can be attributed to the limited diversity of samples selected based on cluster size. However, after adding 150 training samples, GRS shows improved learning performance, matching the SMSE and CC values of RS-OP. As outlined in Section 5.1, within the context of the Anchorage testbed's less complex data structure, sampling larger clusters demonstrates higher representativeness, which directly influences learning performance when predicting the red tag probability. Despite the sampled points' limited diversity compared to RS-OP, GRS outperforms RS-OP in the Anchorage testbed.

MI-ALK demonstrates superior learning performance with the lowest SMSE and highest CC values. When comparing MI-ALK's prediction uncertainty with that of passive sampling methods, it becomes evident that MI-ALK's uncertainty stabilizes swiftly, needing only 100 training samples for stability in the San Francisco scenario. In the Anchorage testbed, characterized by a simpler data structure, the uncertainty in the predictions decreases even more rapidly, requiring only 50 training samples. On the other hand, passive methods require additional samples to reduce the uncertainty. The convergence pattern of MI-ALK's prediction uncertainty indicates that sampling informative data points significantly reduces uncertainty initially, but the reduction rate plateaus. This observation aligns with the diminishing returns property of the MI criterion [34]. However, when comparing the prediction uncertainty for the loss ratio in both testbeds, MI-ALK exhibits approximately 50% higher levels of prediction uncertainty compared to RS-OP, indicating that relying solely on informativeness does not guarantee high prediction confidence.

On the contrary, A-BMAL$_{30}$ incorporates both informativeness and diversity in its sampling strategy. It uses the MI criterion to select informative samples and clustering to ensure diversity. The results presented in Figs. 6 and 7 show that A-BMAL$_{30}$ achieves comparable SMSE and CC values to MI-ALK and RS-OP, with a slight decline in accuracy in the San Francisco testbed. However, in terms of prediction uncertainty, A-BMAL$_{30}$ consistently outperforms other methods.

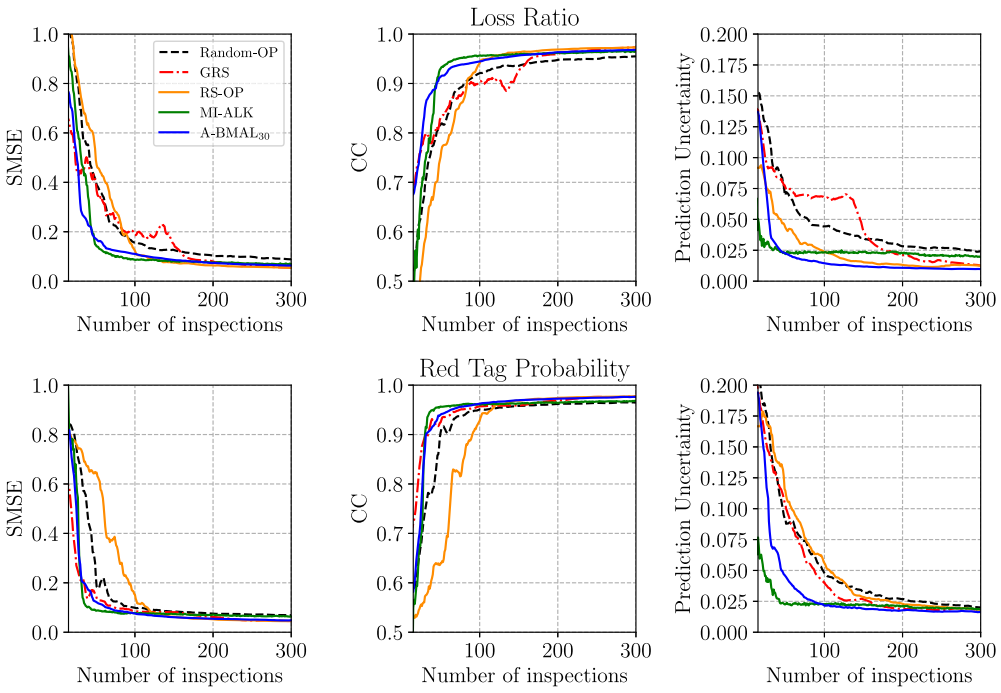



**Fig. 7.** Predictive performance comparison for the Anchorage testbed.

It exhibits a distinctive convergence pattern, with stable prediction uncertainty after 100 training samples, and achieves smaller values of prediction uncertainty at the end of the training (as shown for the loss ratio in Figs. 6 and 7). This suggests that, while sampling informative data points leads to early convergence, explicit inclusion of diversity through the clustering algorithm further reduces prediction uncertainties. In general, A-BMAL$_{30}$ benefits from both diversity and informativeness in its query selection, resulting in higher prediction confidence.

In practice, the absence of a test set poses challenges in precisely assessing model error to determine the termination of the survey. Consequently, as highlighted in Section 3.2, the stability of uncertainties becomes crucial in decision-making and the reliance placed on inference outcomes. Of specific interest are models exhibiting notable convergence patterns based on prediction uncertainty while maintaining sufficiently accurate predictions. Based on the learning performance presented in Figs. 6 and 7, AL models (i.e., MI-ALK and A-BMAL$_{30}$) fulfill these criteria.

## 6. Results and discussion on post-earthquake reconnaissance efficiency

In this section, we apply the termination criterion specified in Section 3.2 to different algorithms. The threshold $\varepsilon_t$ in Eq. (19) is universally set to 0.05. Algorithms failing to meet this criterion are assigned a budget of 300 inspections. We discuss how the termination of the survey affects the selection of the buildings. Then we assess the resource requirement and predictive performance of each algorithm. Moreover, this section delves deeper into evaluating the effectiveness of the proposed framework in earthquake engineering by comparing predictions with recorded data from instrumented buildings during the seismic event in Anchorage.

### 6.1. Influence of termination criterion on the selection of candidate buildings

Following each survey day, we compute predictive variance across the 10,000 buildings in the test set $S_*$. Extracting the diagonal elements of $\Sigma_*$ in Eq. (7), one can obtain the daily distribution of prediction variance. The termination criterion assesses the JS distance between distributions of prediction uncertainty for successive days. As described in Section 3.2, according to Eq. (19), if two consecutive JS distances fall below the threshold $\varepsilon_t = 0.05$, it signals that additional inspections no longer significantly enhance the model, leading to the termination of the survey.

This section examines the influence of each algorithm on building selection, using three attributes – floor area, construction year, and spectral acceleration – to represent building characteristics and ground motion intensity. Fig. 8 visually demonstrates how each algorithm samples various buildings until the termination point.

In each block, the last column, labeled (T), presents the actual (true) distribution of the feature. The remaining five columns depict the distribution of samples, for random selection, passive, and active learning methods. The effective range is constrained within the 5th to 95th percentile for all three features, and we divide each feature into 50 bins. We utilize 64 realizations of $S_{80}$ to run the different algorithms and obtain samples, and the figure reflects the average result. The number above the five columns indicates the median number of inspections required until the survey termination.

Darker-colored bins indicate a higher likelihood of sampling within the respective bin. Specifically, MI-ALK (A1) and A-BMAL$_{30}$ (A2) demonstrate an ability to capture patterns resembling the true distribution with significantly fewer inspections compared to the other three methods. Meanwhile, certain regions are over-represented in alternative algorithms, while A-BMAL$_{30}$ (A2) yields a balanced number of samples, closely mirroring the true distribution even with fewer inspections. This pattern is particularly evident in the evaluation of spectral acceleration, as depicted in Fig. 8.

This performance is because samples obtained from A-BMAL$_{30}$ (A2) exhibit greater diversity, effectively covering a broader range with fewer inspections. This observation aligns with the previous discussion on diversity measures in Section 5.1. With the sampling performance considered, we will now proceed to the discussion on the time and resource requirements for each algorithm.

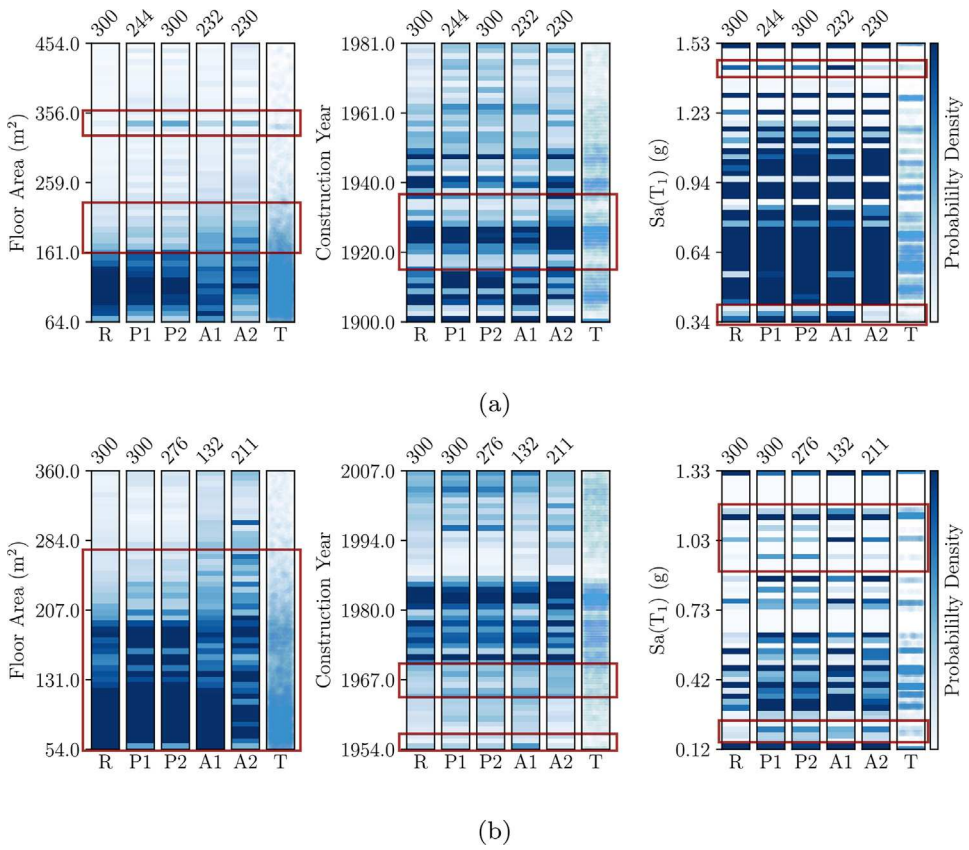



**Fig. 8.** Sampling performance until the survey termination for predicting loss ratio in the (a) San Francisco and (b) Anchorage testbeds. R:Random-OP, P1: GRS, P2: RS-OP, A1: MI-ALK, A2: A-BMAL$_{30}$, and T: true distribution.

### 6.2. Time and resource requirements

Fig. 9 displays the overall time required by each algorithm to satisfy the termination criterion. The bar plots represent the median values of 64 realizations, with error bars indicating the 25th and 75th percentiles. Notably, Fig. 9(a) demonstrates that A-BMAL$_{30}$ achieves convergence with 22% and 26% less time compared to Random-OP and RS-OP for predicting loss ratio and red tag status, respectively. This reduction in time demand is more pronounced in the Anchorage scenario, displaying a 26% reduction for loss ratio and a 51% reduction for red tag probability compared to Random-OP (Fig. 9(b)). This accelerated convergence expedites decision-making while reducing costs and resource requirements for the survey task.

The findings in the Anchorage testbed exhibit a similar trend. Although MI-ALK registered the least time for the Anchorage scenario, an observation from Fig. 9(b) indicates that the time spent traveling between buildings is roughly twice that of A-BMAL$_{30}$. This potentially leads to a higher total time in realistic scenarios where road blockages occur due to the presence of debris, or the travel speed is much lower than our current assumption (i.e., 30 km/h). Additionally, the simpler nature of the data in Anchorage building stock has further accentuated the convergence. However, this early convergence in MI-ALK might not persist in a more complex data set characterized by a more balanced feature space.

Furthermore, increasing the batch size postpones convergence. A larger batch size corresponds to an increased number of clusters $\beta_0$, resulting in smaller clusters, reduced distance among medoids, and less sample diversity. Given a fixed data pool $S$ for selecting candidate buildings, increasing $\beta_0$ increases the likelihood of selecting less informative samples in each iteration of A-BMAL. Hence, there is a trade-off between efficient routing and the model's learning performance when considering larger batch sizes.

Fig. 10 presents the total travel distance and number of inspections for each sampling approach. GRS and MI-ALK require significantly greater travel distances compared to other methods. On the other hand, A-BMAL variants with larger $\beta_0$ offer more routing flexibility and achieve reduced travel distances. This is more noticeable in the San Francisco case due to the higher density of the building stock compared to the Anchorage testbed. In the San Francisco scenario, A-BMAL$_{300}$ converges with 32% and 27% less travel distance than A-BMAL$_{30}$ for loss ratio and red tag probability prediction, respectively. However, larger $\beta_0$ values lead to delayed convergence due to the trade-off between routing efficiency and learning performance.

It is noteworthy that both Random-OP and RS-OP barely converge within the budget limit (i.e., 300 inspections) as can be seen in Fig. 10(a), with the error bars showing no variation in the number of inspections. Similarly, examining the error bars in Fig. 10(b), we observe that Random-OP consistently fails to converge in the Anchorage case.

Different algorithms require varying numbers of inspections to reach convergence due to the termination criterion. However, terminating the survey early with fewer inspections affects the model's performance. To assess the performance at their termination points, the model error and prediction uncertainty of each algorithm are compared in Fig. 11.

As illustrated in Fig. 11, Random-OP shows higher SMSE and prediction uncertainty compared to other approaches, while GRS and RS-OP have the lowest SMSE values. Nevertheless, in the San Francisco testbed, both GRS and RS-OP display a lack of convergence. This is indicated (in Fig. 11(a)) by the numbers above the bar charts, denoting the number of inspections at the stopping point for each algorithm. Additionally, this absence of convergence is visible in the prediction uncertainty plots depicted in Fig. 6, where a consistent negative slope is observed until reaching the budget limit. In the

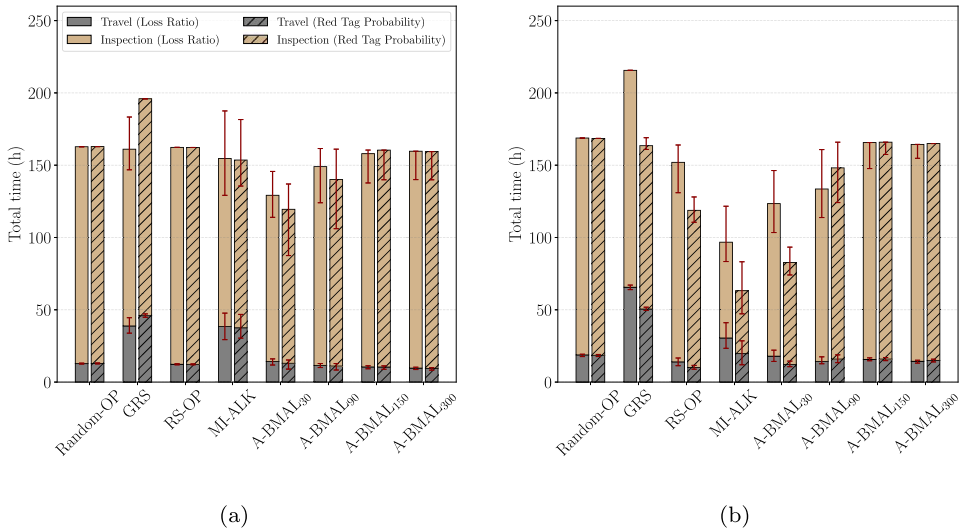



**Fig. 9.** The total time required until reaching the termination criterion for the (a) San Francisco and (b) Anchorage testbeds. The total time is split into the travel time and inspection time. The error bars show the values between the 25th and 75th percentiles.

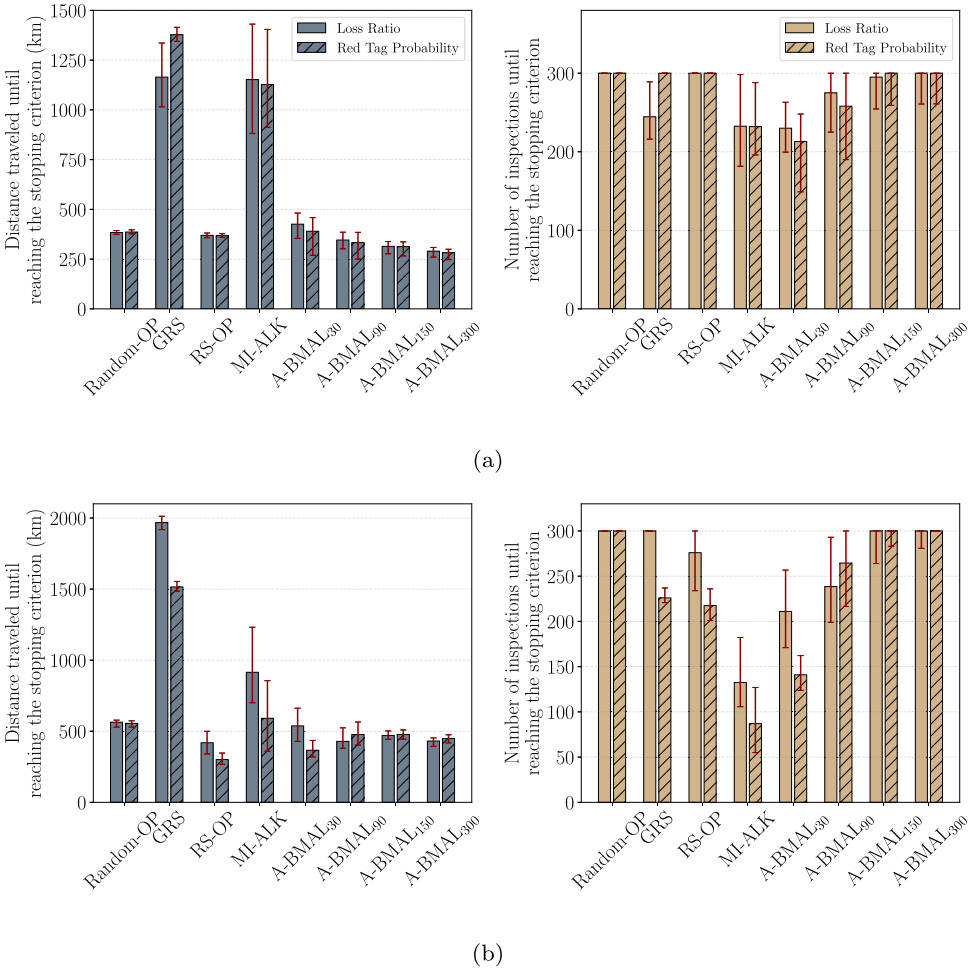



**Fig. 10.** Cumulative distance traveled and the total number of building inspections until reaching the termination criterion for the (a) San Francisco and (b) Anchorage testbeds.

Anchorage testbed, passive sampling algorithms indicate meeting the convergence criterion before reaching the 300-inspection budget. This behavior is strongly influenced by the simpler building stock observed in the Anchorage testbed, where the concentration of data in the feature space contributes to a faster convergence by solely relying on sampling representative buildings. However, it is important to note that convergence occurs at appreciably higher inspection numbers compared to MI-ALK and A-BMAL$_{30}$. Thus, for passive sampling approaches, increasing the number of inspections reduces the model's uncertainty and improves its performance through additional data acquisition.

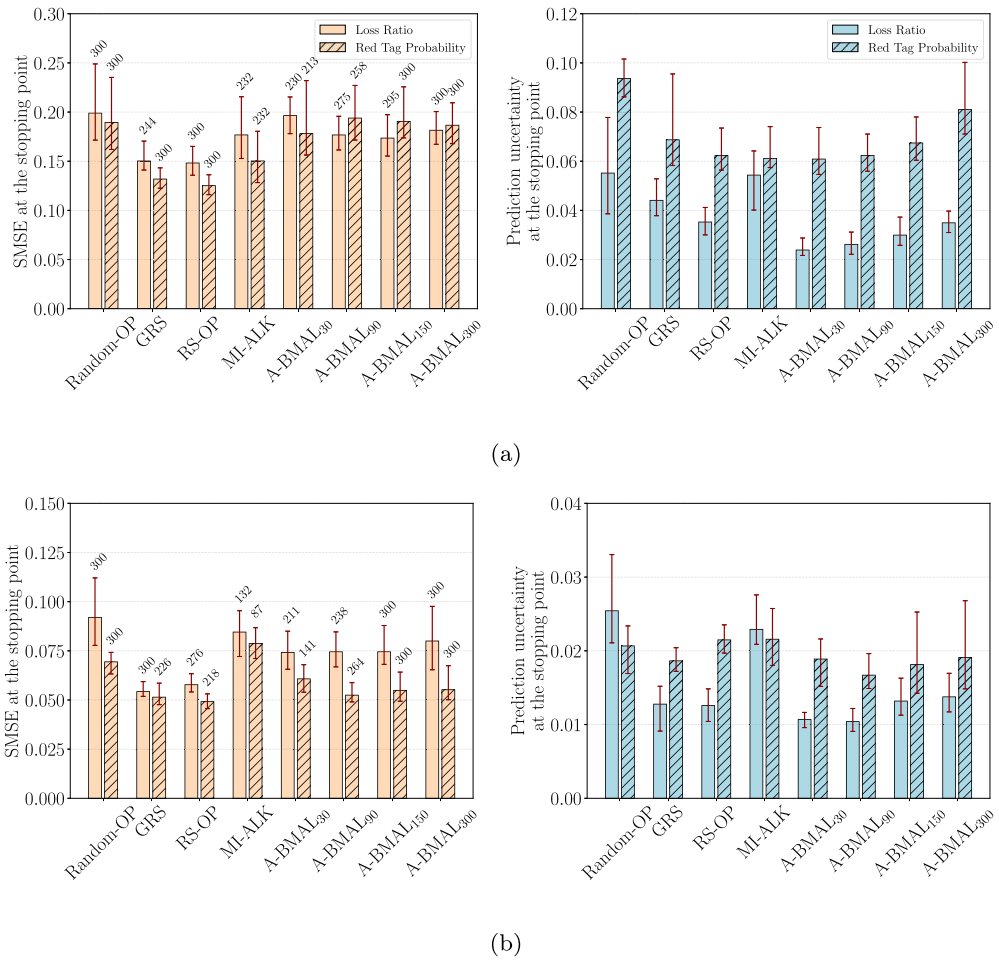



**Fig. 11.** Performance comparison of various algorithms at their termination points for the (a) San Francisco and (b) Anchorage testbeds. The numbers above each bar chart represent the median number of inspections needed to meet the stopping criterion.

As such, clustering-based passive sampling methods are suitable for guiding extended reconnaissance missions as they do not exhibit an early convergence pattern that is crucial for rapid decision-making in time-sensitive scenarios.

On the other hand, MI-ALK and A-BMAL$_{30}$ demonstrate an early convergence pattern as discussed in Section 5.2, allowing the survey team to terminate the inspection process and rely on the model's predictions.

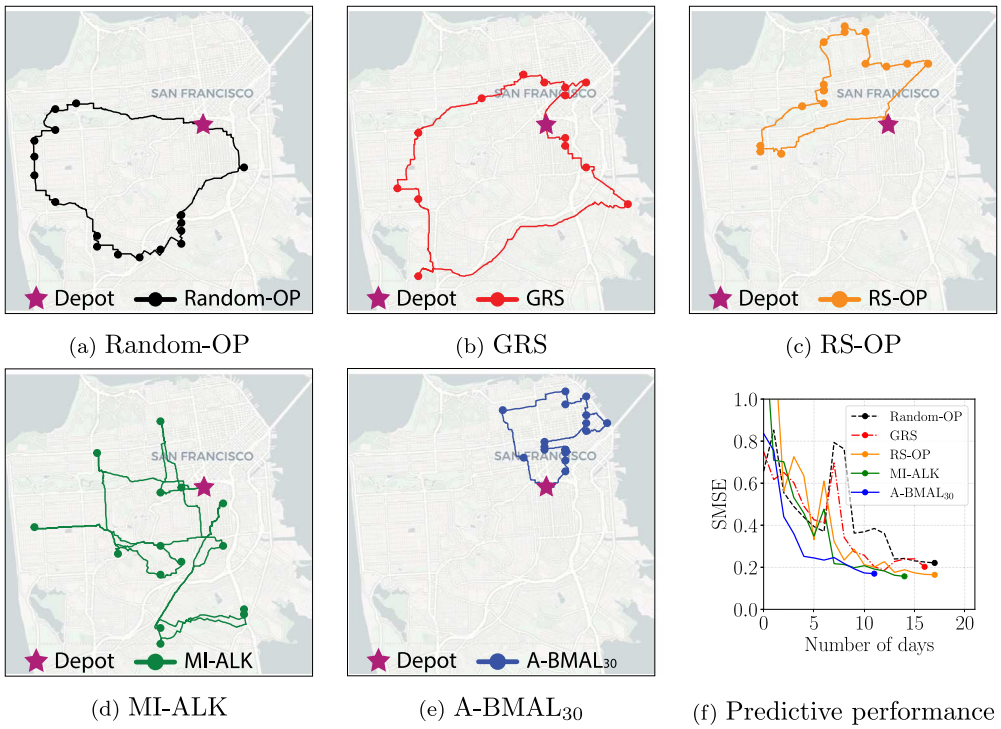
As illustrated in Fig. 11, comparing MI-ALK and A-BMAL$_{30}$, MI-ALK requires fewer inspections to meet termination criterion in Anchorage but with higher prediction error and model uncertainty than A-BMAL$_{30}$. In San Francisco, A-BMAL$_{30}$ converges with slightly fewer inspections but shows higher prediction error. However, MI-ALK's dependency on query sequence can limit its practicality for geographically distant queries. Additionally, the homogeneous nature of the data set in the Anchorage testbed further leads to rapid convergence using MI-ALK, while a more heterogeneous building stock as that in the San Francisco favors A-BMAL$_{30}$, showing faster convergence. Therefore, A-BMAL$_{30}$ proves advantageous for more diverse building inventories that broadly span the feature space, enabling faster convergence due to its capability to utilize both sampling mechanisms based on diversity and informativeness of the queries.

Examining the variants of A-BMAL with different batch sizes, Fig. 11 shows a direct correlation between batch size and prediction uncertainty. As previously mentioned, a larger $\beta_0$ favors the routing efficiency but undermines the learning performance by limiting the diversity and informativeness of the samples. It can be inferred that the value of $\beta_0$ can be deemed as a proxy for the trade-off between efficient routing and learning performance.

To demonstrate the algorithms' performance, Fig. 12 displays the routes created by different approaches during a one-day building damage inspection in the San Francisco testbed along with the associated predictive performance for a single realization. A-BMAL$_{30}$ exhibits a convergence pattern before the 5th day of the survey. Accelerated convergence allows A-BMAL$_{30}$ to meet the termination criterion approximately six days earlier than passive sampling methods. Furthermore, A-BMAL$_{30}$ achieves the same level of accuracy as MI-ALK, but three days ahead. Even if the survey is halted within the initial days, A-BMAL$_{30}$ maintains a more stable predictive accuracy compared to other algorithms. Results suggest that A-BMAL$_{30}$ is robust to incomplete data and can deliver reliable predictions if the survey is prematurely interrupted.

### *6.3. Evaluation of model predictions against measured data from instrumented buildings*

Experience from major earthquakes has shown that non-structural losses collectively exceed the economic losses caused by structural damage [67,68]. Meanwhile, non-structural damage is often correlated with acceleration demands [67]. As a result, effectively predicting acceleration demand projects expected non-structural damages, enabling a faithful loss assessment. To showcase the effectiveness of the proposed model, we assess its performance by comparing its predictions with the measured roof accelerations recorded during the 2018 Anchorage earthquake. This comparison will help us determine the model's accuracy in predicting acceleration demand and non-structural damage.




(a) Random-OP (b) GRS (c) RS-OP

(d) MI-ALK (e) A-BMAL$_{30}$ (f) Predictive performance

**Fig. 12.** Inspection routes generated by various algorithms for the San Francisco testbed alongside the prediction accuracy as the survey progresses. Inspection routes and model performance for more realizations are available at https://youtu.be/DM5PyKU61hU.

To monitor the response of the structures in regions with high seismicity, the United States Geological Survey (USGS) has equipped selected buildings with sensors and accelerometers [69]. This section focuses on two prominent structures in Anchorage, chosen to evaluate our proposed approach and demonstrate its potential application in the earthquake engineering context. The Atwood building, a 20-story steel moment frame with an in-plan steel shear-walled core, was selected for analysis. Its sensor array comprises 32 channels of accelerometers located across 10 levels of the structure [69]. Channel 32 was selected since it corresponds to the primary component of the horizontal accelerogram employed in our testbed, specifically monitoring roof acceleration along the east–west direction. Similarly, we examined the measured response of the Frontier Building, a 14-story reinforced moment frame. We utilized channel 33 among an array consisting of 36 accelerometers. Further information about the configuration of sensors can be found in CESMD [61] and Çlebi [69]. Fig. 13 shows the selected buildings and the extracted records from the roof accelerometers.

We utilized three algorithms that have consistently demonstrated superior overall performance: RS-OP, MI-ALK, and A-BMAL$_{30}$. The models were trained using 64 realizations of different $S_{80}$ as the available data pool. Throughout the training process, the models were fed with the specific features of the Atwood and Frontier buildings, as outlined in Table 1. As only a few buildings are instrumented, to train the model, we rely on the calculated roof node acceleration, extracted from the simulation workflow. Similarly, in a real-world setting, such a structural response (i.e., roof-level acceleration) may not be available during post-earthquake reconnaissance. However, given the strong correlation between nonstructural damage and acceleration demand, alternative indicators of nonstructural damage can be employed to train the model.

Fig. 14 depicts the evolving median values of predicted roof acceleration throughout the training process. A-BMAL$_{30}$ accurately captures the target response (calculated acceleration) for both buildings. When comparing the AL-based methods MI-ALK and A-BMAL$_{30}$ with the passive sampling RS-OP, a notable convergence trend is observed for the AL methods. This amounts to inferring the response of different buildings by relying on the model in earlier stages of training, thereby demanding fewer resources and expediting decision-making. It is worth noting that while MI-ALK exhibits a similar convergence pattern, its reliance on the inspection sequence poses limitations in cases where the queries are located in distant geographical locations or are not accessible to obtain inspection data. As a result, the proposed method is a viable approach that can take into account the learning objective while remaining practical due to the extra flexibility that is available due to the batch query scheme.

## 7. Limitations and future work

This section discusses the limitations of the proposed framework and suggests areas for future research.

In our methodology, we assumed that the damage inspection data have noise levels that follow an independent and identical (Gaussian) distribution across different buildings (i.e., a homoskedastic noise). This assumption results in a Gaussian likelihood and enables a closed-form solution for inference. While this tractability is advantageous, it is essential to recognize the potential limitations of this assumption in real-world scenarios. In practice, real-world situations may induce noise levels with varying parameters (or heteroskedastic noise). For instance, specific regions or types of buildings might exhibit significantly different noise variances in the observed damage. This complexity in the real-world data warrants further exploration in future research, with an emphasis on capturing heteroskedastic noise patterns to enhance the inference results.

Additionally, in a real-world context, subjective interpretation of building damage by different survey team members could introduce an additional source of uncertainty to the model, a factor that our formulation does not explicitly account for. The subjective interpretation of building damage leads to varying noise parameters, thereby contributing to noise heteroskedasticity. Consequently, this aspect of

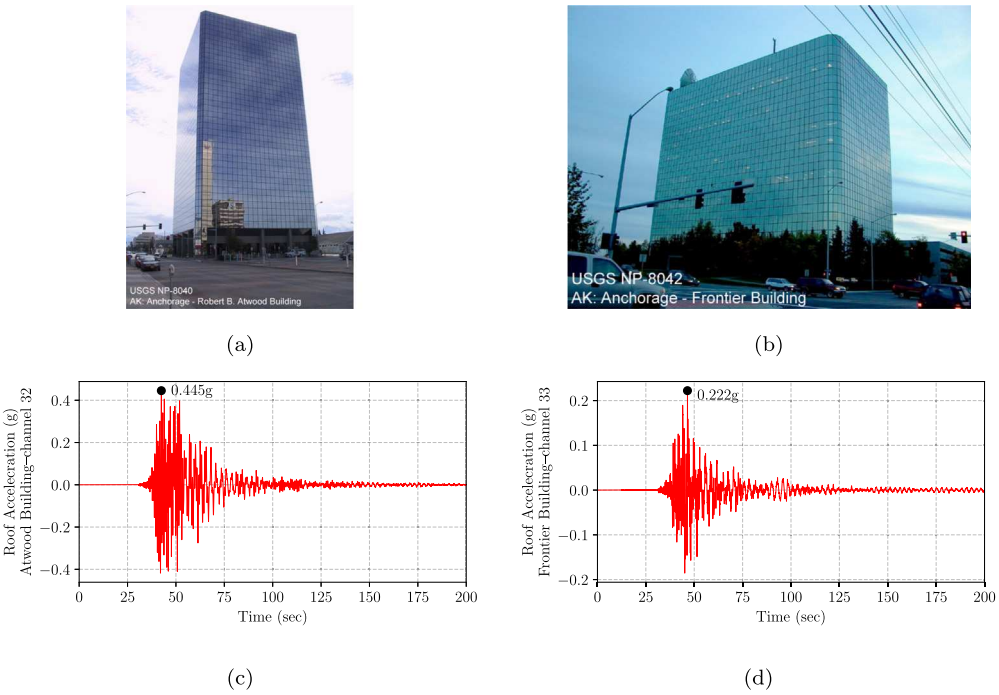



**Fig. 13.** (a) Atwood Building, (b) Frontier Building (building photos are adapted from CESMD [61]); (c) Roof acceleration in the east–west direction at Atwood Building, (d) Roof acceleration in the east–west direction at Frontier Building.

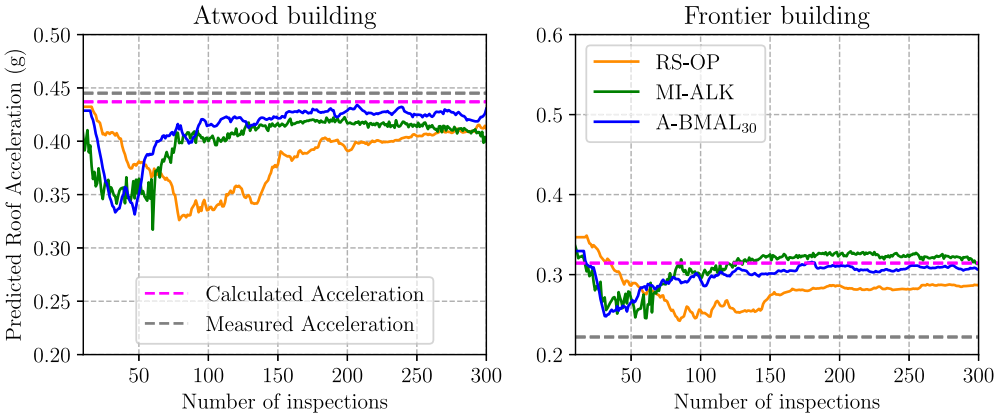



**Fig. 14.** Evolution of the predicted roof acceleration for instrumented buildings in the Anchorage testbed.

real-world data can also be addressed with a model that is able to capture heteroskedastic noise.

Our study utilized real-world data for the feature set; however, the labels used for training and testing our models were sourced from a regional simulation. While we ensure the presence of noise in the simulation results by introducing uncertain model parameters and our approach accounts for the presence of noise (as discussed in Section 2), it is important to note that real-world settings might induce considerably higher noise levels. This may impact the signal-to-noise ratio, potentially undermining the accuracy of the inference results.

The task of inferring damage indices was framed as a regression problem in our study. In rapid post-earthquake evaluations, damage assignment often employs discrete values or classes like different color tags (e.g., green, yellow, or red). While our approach allows for transforming and mapping these discrete classes to continuous values to be used in regression formulation, there exists an advantage in formalizing the problem within a classification framework.

Furthermore, our methodology assumes data collection by a single survey team. However, there is potential for extending this framework to encompass multiple teams concurrently collecting and sharing data, allowing for model updates as damage information becomes available from different teams. This extension remains an area for further exploration.

## 8. Conclusion

Post-earthquake surveys are time-consuming tasks, while the perishable nature of damage data in the disaster aftermath necessitates prompt acquisition of the data. Hence, a methodical approach is crucial to ensure streamlined data collection within time constraints and resource limitations, maximizing the utilization of the acquired data. This paper proposes a new batch-mode active learning framework for efficient identification of informative buildings in post-earthquake surveys. The framework selects informative candidate buildings in batches, allowing flexible route planning and improving survey efficiency. To overcome the challenge of scarce ground truth data in post-earthquake scenarios, a stopping criterion based on the stability of the model's prediction uncertainty is proposed. The framework is

compared with passive learning and state-of-the-art active learning methods in guiding post-earthquake surveys following both a hypothetical earthquake scenario in the San Francisco Bay area and the 2018 Anchorage earthquake. The key findings of the study are:

- Using a combination of an information-theoretic criterion and clustering, the proposed framework was able to expedite the coverage of the region by effectively identifying and selecting candidate buildings that were both informative and diverse. This approach allowed for more efficient exploration of the region.
- The batch active learning approach resulted in a distinct convergence pattern in prediction confidence, informing the survey team that additional inspections would yield minimal improvements in prediction quality. This is particularly advantageous in time-sensitive situations that necessitate a prompt evaluation of regional damage.
- The termination criterion employed in this study does not rely on the labels of the test data. As a result, the survey team can determine when to conclude inspections in the absence of ground truth damage data.
- The proposed batch query scheme enhances routing flexibility and resolves data acquisition inefficiencies observed in sequential AL methods. In comparison to the sequential approach, it reduces travel distance by 65%, thereby improving data acquisition efficiency.
- Compared with the baseline approach (random sampling with optimized routes), the proposed method reduced inspection time by up to 26% and 51% for the San Francisco and Anchorage testbeds, respectively, while enhancing prediction accuracy and confidence.
- Clustering-based passive sampling, sequential AL, and the proposed batch AL frameworks highlight specific strengths in varying scenarios. Passive sampling methods are particularly advantageous when time constraints are less critical, offering decent accuracy but slower convergence. Sequential AL suits scenarios characterized by a more homogeneous building stock that is concentrated in the feature space, effectively capturing the regional damage trends by depending solely on sampling informative buildings. Meanwhile, the proposed approach is ideal for complex and heterogeneous building inventories that widely span across a multitude of features, leveraging more advanced sampling mechanisms based on active learning principles. Additionally, it proves advantageous in scenarios with reduced travel speed or road obstructions due to its enhanced routing flexibility compared to sequential AL.

## CRediT authorship contribution statement

**Amirhossein Cheraghi:** Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Yinhu Wang:** Writing – original draft, Visualization, Methodology. **Nikola Marković:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Ge Ou:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Grammarly and ChatGPT in order to improve readability and language. After using these tools, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

## Acknowledgments

This material is based upon work supported by the National Science Foundation, United States under award number 2112758. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## References

[1] D. Lallemant, R. Soden, S. Rubinyi, S. Loos, K. Barns, G. Bhattacharjee, Post-disaster damage assessments as catalysts for recovery: A look at assessments conducted in the wake of the 2015 Gorkha, Nepal, earthquake, Earthq. Spectr. 33 (2017) 435–451.

[2] J. Cusicanqui, N. Kerle, F. Nex, Usability of aerial video footage for 3-d scene reconstruction and structural damage assessment, Nat. Hazards Earth Syst. Sci. 18 (2018) 1583–1598.

[3] J. Boakye, C. Murphy, P. Gardoni, R. Kumar, Which consequences matter in risk analyis and disaster assessment? Int. J. Disaster Risk Reduct. 71 (2022) 102740.

[4] K. Jaiswal, D.J. Wald, Rapid Estimation of the Economic Consequences of Global Earthquakes, US Department of the Interior, US Geological Survey Reston, VA, 2011.

[5] S. Loos, D. Lallemant, J. Baker, J. McCaughey, S.-H. Yun, N. Budhathoki, F. Khan, R. Singh, G-dif: A geospatial data integration framework to rapidly estimate post-earthquake damage, Earthq. Spectr. 36 (2020) 1695–1718.

[6] M. Erdik, K. Şşetyan, M. Demircioğlu, U. Hancılar, C. Zülfikar, Rapid earthquake loss assessment after damaging earthquakes, Soil Dyn. Earthq. Eng. 31 (2011) 247–266.

[7] D. Contreras, T. Blaschke, D. Tiede, M. Jilge, Monitoring recovery after earthquakes through the integration of remote sensing, gis, and ground observations: The case of l'aquila (Italy), Cartogr. Geogr. Inf. Sci. 43 (2016) 115–133.

[8] Z. Zhou, J. Gong, Automated residential building detection from airborne lidar data with deep neural networks, Adv. Eng. Inf. 36 (2018) 229–241.

[9] Y. Pi, N.D. Nath, A.H. Behzadan, Convolutional neural networks for object detection in aerial imagery for disaster response and recovery, Adv. Eng. Inform. 43 (2020) 101009.

[10] T.W. Gillespie, J. Chu, E. Frankenberg, D. Thomas, Assessment and prediction of natural hazards from satellite imagery, Prog. Phys. Geogr. 31 (2007) 459–470.

[11] H. Stone, V. Putrino, D. D'Ayala, Earthquake damage data collection using omnidirectional imagery, Front. Built Environ. 4 (2018) 51.

[12] D. Contreras, S. Wilkinson, P. James, Earthquake reconnaissance data sources, a literature review, Earth 2 (2021) 1006–1037.

[13] M. Sheibani, G. Ou, The development of gaussian process regression for effective regional post-earthquake building damage inference, Comput.-Aided Civ. Infrastruct. Eng. 36 (2021) 264–288.

[14] E.B. Anil, B. Akinci, J.H. Garrett, O. Kurc, Information requirements for earthquake damage assessment of structural walls, Adv. Eng. Inform. 30 (2016) 54–64.

[15] F. Marquis, J.J. Kim, K.J. Elwood, S.E. Chang, Understanding post-earthquake decisions on multi-storey concrete buildings in Christchurch, New Zealand, Bull. Earthq. Eng. 15 (2017) 731–758.

[16] S. German, I. Brilakis, R. DesRoches, Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments, Adv. Eng. Inform. 26 (2012) 846–858.

[17] J.-A. Goulet, C. Michel, A.D. Kiureghian, Data-driven post-earthquake rapid structural safety assessment, Earthq.e Eng. Struct. Dyn. 44 (2015) 549–562.

[18] H. Luo, S.G. Paal, Advancing post-earthquake structural evaluations via sequential regression-based predictive mean matching for enhanced forecasting in the context of missing data, Adv. Eng. Inform. 47 (2021) 101202.

[19] D. Lallemant, A. Kiremidjian, Rapid post-earthquake damage estimation using remote-sensing and field-based damage data integration, in: Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures, 2013, pp. 16–20.

[20] L. Bodenmann, Y. Reuland, B. Stojadinović, Dynamic post-earthquake updating of regional damage estimates using gaussian processes, Reliab. Eng. Syst. Saf. (2023) 109201.

[21] S. Loos, D. Lallemant, F. Khan, J.W. McCaughey, R. Banick, N. Budhathoki, J.W. Baker, A data-driven approach to rapidly estimate recovery potential to go beyond building damage after disasters, Commun. Earth Environ. 4 (2023) 40.

[22] H. Yu, S. Kim, Passive sampling for regression, in: 2010 IEEE International Conference on Data Mining, IEEE, 2010, pp. 1151–1156.

[23] D. Wu, Pool-based sequential active learning for regression, IEEE Trans. Neural Netw. Learn. Syst. 30 (2018) 1348–1359.

[24] B. Settles, Active Learning Literature Survey, 2009.

[25] R.M. Adams, C.M. Evans, L. Peek, Defining, collecting, and sharing perishable disaster data, Disasters (2023).

[26] M. Sheibani, G. Ou, Adaptive local kernels formulation of mutual information with application to active post-seismic building damage inference, Reliab. Eng. Syst. Saf. 215 (2021) 107915.

[27] Y.-C. Yu, P. Gardoni, Predicting road blockage due to building damage following earthquakes, Reliab. Eng. Syst. Saf. 219 (2022) 108220.

[28] E. Schulz, M. Speekenbrink, A. Krause, A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions, J. Math. Psych. 85 (2018) 1–16.

[29] K.P. Murphy, Probabilistic Machine Learning: An Introduction, MIT Press, 2022.

[30] C.K. Williams, C.E. Rasmussen, Gaussian Processes for Machine Learning, Volume 2, MIT press Cambridge, MA, 2006.

[31] O. Reyes, S. Ventura, Evolutionary strategy to perform batch-mode active learning on multi-label data, ACM Trans. Intell. Syst. Technol. 9 (2018) 1–26.

[32] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423.

[33] F. Zhdanov, Diverse mini-batch active learning, 2019, arXiv preprint arXiv: 1901.05954.

[34] A. Krause, A. Singh, C. Guestrin, Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies, J. Mach. Learn. Res. 9 (2008).

[35] C.-W. Ko, J. Lee, M. Queyranne, An exact algorithm for maximum entropy sampling, Oper. Res. 43 (1995) 684–691.

[36] C. Guestrin, A. Krause, A.P. Singh, Near-optimal sensor placements in gaussian processes, in: Proceedings of the 22nd International Conference on Machine Learning, 2005, pp. 265–272.

[37] P. Mitra, S.K. Pal, M.A. Siddiqi, Non-convex clustering using expectation maximization algorithm with rough set initialization, Pattern Recognit. Lett. 24 (2003) 863–873.

[38] B. Demir, C. Persello, L. Bruzzone, Batch-mode active-learning methods for the interactive classification of remote sensing images, IEEE Trans. Geosci. Remote Sens. 49 (2010) 1014–1031.

[39] Z. Xu, K. Yu, V. Tresp, X. Xu, J. Wang, Representative sampling for text classification using support vector machines, in: Advances in Information Retrieval: 25th European Conference on IR Research, ECIR 2003, Pisa, Italy, April (2003) 14–16. Proceedings 25, Springer, 2003, pp. 393–407.

[40] H. Ishibashi, H. Hino, Stopping criterion for active learning based on error stability, 2021, arXiv preprint arXiv:2104.01836.

[41] M. Bloodgood, K. Vijay-Shanker, A method for stopping active learning based on stabilizing predictions and the need for user-adjustable stopping, 2014, arXiv preprint arXiv:1409.5165.

[42] J. Lin, Divergence measures based on the shannon entropy, IEEE Trans. Inform. Theory 37 (1991) 145–151.

[43] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1951) 79–86.

[44] T.M. Cover, Elements of Information Theory, John Wiley & Sons, 1999.

[45] P. Vansteenwegen, W. Souffriau, D. Van Oudheusden, The orienteering problem: A survey, European J. Oper. Res. 209 (2011) 1–10.

[46] M.G. Kay, Matlog: Logistics Engineering MATLAB Toolbox, 2013.

[47] C. Archetti, F. Guerriero, G. Macrina, The online vehicle routing problem with occasional drivers, Comput. Oper. Res. 127 (2021) 105144.

[48] X. Zhao, K. Ji, P. Xu, W.-w. Qian, G. Ren, X.-n. Shan, A round-trip bus evacuation model with scheduling and routing planning, Transp. Res. A: Policy Practice 137 (2020) 285–300.

[49] X. Lu, F. McKenna, Q. Cheng, Z. Xu, X. Zeng, S.A. Mahin, An open-source framework for regional earthquake loss estimation using the city-scale nonlinear time history analysis, Earthq. Spectr. 36 (2020) 806–831.

[50] W. Elhaddad, F. McKenna, M. Rynge, J. Lowe, C. Wang, A. Zsarnoczay, Nheri-Simcenter/Workflowregionalearthquake: Rwhale (Version V1. 1.0), Zenodo, 2019.

[51] N. Petersson, B. Sjogreen, Sw4, Version 2.01 [Software]. Computational Infrastructure of Geodynamics, vol. 10, Zenodo, Switzerland, 2017, p. 5281.

[52] A.J. Rodgers, A. Pitarka, R. Pankajakshan, B. Sjögreen, N.A. Petersson, Regional-scale 3d ground-motion simulations of mw 7 earthquakes on the hayward fault, Northern California resolving frequencies 0–10 hz and including site-response corrections, Bull. Seismol. Soc. Am. 110 (2020) 2862–2881.

[53] X. Lu, H. Guan, Earthquake Disaster Simulation of Civil Infrastructures, Springer, 2017.

[54] A. Zsarnóczay, W. Elhaddad, B. Cetiner, K. Zhong, F. McKenna, G. Deierlein, SimCenter Earthquake Testbed, DesignSafe-CI, San Francisco, CA, 2023.

[55] P. Waddell, Urbansim: Modeling urban development for land use, transportation, and environmental planning, J. Am. Plan. Assoc. 68 (2002) 297–314.

[56] X. Zeng, X. Lu, T. Yang, Z. Xu, Application of the fema-p58 methodology for regional earthquake loss prediction, Nat. Hazards 83 (2016) 177–192.

[57] FEMA, Seismic Performance Assessment of Buildings Volume 1: Methodology, Technical Report, (FEMA-P58) Federal Emergency Management Agency, Washington, DC, 2012.

[58] FEMA, Post-Disaster Building Safety Evaluation Guidance (FEMA P-2055), 2019.

[59] W.M. Hassan, J. Rodgers, C. Motter, J. Thornley, Structural performance of buildings during the 30 november 2018 m7. 1 anchorage, alaska earthquake, Earthq. Spectr. 38 (2022) 200–234.

[60] E.M. Rathje, C. Dawson, J.E. Padgett, J.-P. Pinelli, D. Stanzione, A. Adair, P. Arduino, S.J. Brandenberg, T. Cockerill, C. Dey, et al., Designsafe: New cyberinfrastructure for natural hazards engineering, Nat. Hazards Rev. 18 (2017) 06017001.

[61] CESMD, Center for Engineering Strong-Motion Data, CESMD, 2023, https://www.strongmotioncenter.org/, [Accessed on November 30, 2023].

[62] A. Arias, A Measure of Earthquake Intensity. Seismic Design for Nuclear Power Plants, Massachusetts Institute of Technology, 1970.

[63] P. Fajfar, T. Vidic, M. Fischinger, A measure of earthquake motion capacity to damage medium-period structures, Soil Dyn. Earthq. Eng. 9 (1990) 236–242.

[64] L. Ye, Q. Ma, Z. Miao, H. Guan, Y. Zhuge, Numerical and comparative study of earthquake intensity indices in seismic analysis, Struct. Des. Tall Special Build. 22 (2013) 362–381.

[65] E. Snelson, Z. Ghahramani, C. Rasmussen, Warped gaussian processes, in: Advances in Neural Information Processing Systems, vol. 16, 2003.

[66] ATC, Field Manual: Procedures for Postearthquake Safety Evaluation of Buildings, ATC-20-1 Report, second ed., Applied Technology Council, Redwood City, California, 2005.

[67] A. Kazantzi, D. Vamvatsikos, E. Miranda, Evaluation of seismic acceleration demands on building nonstructural elements, J. Struct. Eng. 146 (2020) 04020118.

[68] A. Filiatrault, T. Sullivan, Performance-based seismic design of nonstructural building components: The next frontier of earthquake engineering, Earthq. Eng. Eng. Vibration 13 (2014) 17–46.

[69] M. Çlebi, Highlights of a cursory study of behavior of three instrumented buildings during the mw 7.1 anchorage, alaska, earthquake of 30 november 2018, Seismol. Res. Lett. 91 (2020) 56–65.