Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/EU08.20250411
Published: July 2, 2025

Optimum Shelter Location Problem Using Predicted Injuries

Şule Nur Sargın
Department of Industrial Engineering
Istanbul Medeniyet University
Istanbul, 34700, Turkey
sulenur.sargin@medeniyet.edu.tr

Ebru Angün
Department of Industrial Engineering
Galatasaray University
Istanbul, 34349, Turkey
eangun@gsu.edu.tr

Abstract

This paper focus on predicting the number of injured people in the aftermath of an earthquake, applying Gaussian process (also known as Kriging), which is a well-known method for machine learning. Different from neural networks, Gaussian process estimates also the variance of the predictor. The predicted injuries and their variances are used in a chance-constrained optimization problem to determine shelter locations from a given set of candidate locations. The resulting problem is a mixed-integer nonlinear programming optimization, which can be solved through commercial software in a reasonable amount of time for small scale problems.

Keywords
Disaster management, Gaussian process, chance-constraints, injury/casualty prediction

1. Introduction
Turkey is situated on the complex tectonic boundary of the Eurasian, Arabian, and African plates, making it one of the most seismically active regions in the world. Within the last three decades, Turkey was hit by three devastating earthquakes; the earthquake that struck Izmit in August 1999, resulted in significant loss of life and property. On February 6, 2023, two devastating earthquakes with magnitudes 8.0 and 7.9 in Richter scale struck southeastern region of Turkey, resulting in significant damage and loss of life (Erdik et al., 2023). As of April 5, the Turkish Ministry of Interior claimed that the earthquake on February 6, 2023, had taken the lives of 50,399 people in the country. This earthquake was the most powerful earthquake recorded since 1939 (Erdik et al., 2023). Disaster and Emergency Management Presidency of Turkey (AFAD, https://en.afad.gov.tr) said that rescue operations started immediately after the earthquake, and more than 2600 rescuers from 65 countries and regions had sent to assist Turkey in rescue operations on February 7.

Being a part of the disaster management, disaster preparedness refers to the knowledge, capacities, and actions taken by individuals, communities, organizations, and governments to anticipate, respond to, and recover from the impacts of disasters**.** It is a proactive approach to minimize the harm and disruption caused by disasters. Disaster preparedness consists of activities such as risk assessment, evacuation routes planning, determination of shelter locations, and stocking of supplies such as non-perishable food, first-aid kits and medication, and flashlights and batteries.

1.1 Objectives

This paper aims to propose a novel approach to predict the number of injured people in the aftermath of an earthquake using Gaussian process, which is one of the many machine learning methods. The predicted numbers and their variances are further used in a chance-constrained optimization problem to determine shelter locations from a given discrete set of candidate locations.

The rest of the paper can be summarized as follows. Section 2 reviews literature of the methods that have been used to predict injuries/casualties due to earthquakes. Section 3 introduces our methods to predict the injuries, validate our prediction method, and a chance-constrained optimization problem that we will solve to determine shelter locations. Section 4 presents our input/output data. Section 5 presents the results of our validation. Section 6 presents our numerical and graphical results, and provides further improvements. Finally, Section 7 presents our conclusions and future research areas.

2. Literature Review

The applications of machine learning (ML) and artificial intelligence (AI) methods in disaster management have been a growing research area as can be seen from the recent review articles Maguraushe et al. (2023), Guha et al. (2022), and Sun et al. (2020). Sun et al. (2020) explained that these applications focused on all four phases of the disaster management--namely, mitigation, preparedness, response, and recovery—where the emphasis is on the disaster response followed by disaster mitigation. Even though most of the papers that we have reviewed for this section did not mention ML or AI, they all applied statistical methods such as linear regression, principal component analysis, and several types of neural networks (NN); this enabled us to classify them as applications of the ML and AI methods to predict earthquake casualties and/or injuries.

Casualty/injury prediction methods can be classified into two major types: Empirical approaches and analytical approaches. Empirical approaches base fatality predictions on seismic data such as magnitude, intensity, population density, and the numbers of casualties/injuries from historical earthquakes data. Analytical approaches consider soil conditions, building inventories, demographic data of the disaster region in addition to seismic data. As a result, analytical approaches provide predictions with smaller variances; yet, they require the collection of more data for longer periods. Below, we summarize different prediction models from both empirical and analytical approaches.

Bastami and Soghrat (2017) predicted the number of earthquake fatalities using linear regression, where they used distance to the epicenter or peak ground acceleration (PGA) as explanatory variables; because the PGA's were not available for villages, they also estimated PGA using linear regression, and this time, distance to the epicenter and focal depth were the explanatory variables. Firuzi et al. (2022) modeled uncertainties in the number of fatalities of the past earthquakes and the ground motion parameters such as PGA, peak ground velocity, and modified Mercalli intensity by fuzzy regression and Monte Carlo simulation. To predict fatalities, this paper provides a simple model that multiplies the fatality rate with the exposed population. Tang et al. (2019) proposed a two-stage approach to rapidly predict the number of casualties in the aftermath of an earthquake; they showed that the cumulative number of casualties follows a logarithmic curve in time with a sharp increase in usually the first 72 hours.

Liu et al. (2025) proposed a Quick Rough Estimate (QRE) method to predict the final total number of fatalities. This method directly depends on the fatality numbers available so far and on the rescue efforts, and it indirectly depends on the earthquake magnitude. Furthermore, as more data become available, the method predicts with more accuracy. Li et al. (2021) proposed a fatality estimation method, multiplying the fatality rate with the population density exposed to the earthquake. Their proposed fatality rate uses fatality records normalized by the Human Development Index. Shi et al. (2024) proposed a casualty prediction model for Taiwan using macroseismic intensities, population exposure, and gross national income per capita as inputs. Their method assumes that the fatality ratio follows a lognormal cumulative distribution function. They further show that the prediction models employed for different geographical areas differ significantly.

Gul and Guneri (2016) estimated the number of injuries applying artificial neural network, where they used earthquake occurrence time, earthquake magnitude, and population density as predictors. Zhou et al. (2023) proposed a spatial evaluation model of earthquake-affected population based on the correlation characteristics of the influencing factors and the back propagation neural network. Of the ten influencing factors, they found that per capita Gross Domestic Product (GDP) and PGA had a stronger correlation with the earthquake-affected population.

Fang et al. (2020) proposed a casualty prediction and rescue model using a network model of the disaster area. The casualty prediction uses the damage matrix of a specific type of building with respect to the earthquake intensities. They further provide a matrix with the number of casualties with respect to the damage levels of the buildings. These matrices were obtained through sampling of the old data and simulation. Their approach requires the knowledge of all building types in the disaster area, which may not be available instantly. Furthermore, Xu et al. (2025) proposed a spatiotemporal casualty assessment method of earthquake-induced falling debris, accounting for human emergency behavior through agent-based simulation. They first obtained a high-resolution population distribution using Kriging interpolation, then they applied agent-based simulation for the emergency behaviors of the individuals during earthquake, and they used physics engine to simulate the distribution of falling debris in urban building cluster; so, they obtained a dynamic and spatiotemporal assessment method. Ren et al. (2024) proposed a probabilistic casualty assessment method under different seismic intensities applying Monte Carlo simulation; they considered uncertainties in earthquake occurrence time, building damage states, and occupancy numbers. They also constructed seismic casualty vulnerability curves for residential areas.

As a result of our literature review, we reached to the conclusion that ML/AI methods and simulation have been used successfully to predict earthquake injuries/casualties. However, these early works have not used the predictions for disaster management purpose, which is lacking in the current literature. Therefore, in the remaining of this paper, we will first explain a method to predict the number of injured people in the aftermath of an earthquake. Then, we will use this prediction to decide on the locations of the shelters.

3. Methods

In the current study, we apply Gaussian process (also known as Kriging) to predict injuries, and leave-one-out cross validation (LOO-CV) for model validation. Furthermore, the predicted model is used in the chance-constrained optimization problem to determine the locations of shelters and the allocations of various districts to these shelters, as we explain below.

3.1 Gaussian process

Gaussian process (GP) is one of the prediction methods used in ML and simulation-optimization, which exploits spatial correlations among input variables. GP can be preferred to neural networks because it provides both predictors and predictor variances; see Kleijnen (2015), and Rasmussen and Williams (2006).
GP models the unknown quantity $\boldsymbol{y}$ by

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{M}$$

where $\mu = E(y)$ and $\boldsymbol{M}$ follows a Gaussian process with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma_M}$. Different types of correlation functions model how the two inputs $\boldsymbol{x}$ and $\boldsymbol{x}'$ are correlated that result in different $\boldsymbol{\Sigma_M}$. Specifically, the *isotropic squared exponential* correlation function is defined by

$$\boldsymbol{\rho}(\boldsymbol{x}, \boldsymbol{x}') = \mathbf{exp}\,(-\boldsymbol{\theta} \sum_{\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{k}} (\boldsymbol{x_j} - \boldsymbol{x}'_{\boldsymbol{j}})^{\mathbf{2}}) \tag{1}$$

where $\theta \geq 0$ is a scalar parameter that is related to the length-scale parameter and k is the dimension of $\boldsymbol{x} = (x_1, \ldots, x_k)$. Then, the GP predictor at a new point $\boldsymbol{x}_*$ is

$$\widehat{\boldsymbol{y}}(\boldsymbol{x}_*) = \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\sigma}}_{\boldsymbol{M}}(\boldsymbol{x}_*)^{\boldsymbol{T}} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{M}}^{-\mathbf{1}} (\boldsymbol{y} - \mathbf{1}\widehat{\boldsymbol{\mu}}) \tag{2}$$

where $\hat{\sigma}_M(\boldsymbol{x}_*)$ is the covariance vector between the new point $\boldsymbol{x}_*$ and the old points $\boldsymbol{x}$, $\boldsymbol{y}$ is the vector of observations at the old points, and $\mathbf{1}$ is the vector of ones of appropriate dimension. Furthermore, the predictor variance, which is not given by neural networks, is as follows:

$$\boldsymbol{s}^{\mathbf{2}}[\widehat{\boldsymbol{y}}(\boldsymbol{x}_*)] = \widehat{\boldsymbol{\tau}}^{\mathbf{2}} - \widehat{\boldsymbol{\sigma}}_{\boldsymbol{M}}(\boldsymbol{x}_*)^{\boldsymbol{T}} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{M}}^{-\mathbf{1}} \widehat{\boldsymbol{\sigma}}_{\boldsymbol{M}}(\boldsymbol{x}_*) + \frac{[\mathbf{1} - \mathbf{1}^{\boldsymbol{T}} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{M}}^{-\mathbf{1}} \widehat{\boldsymbol{\sigma}}_{\boldsymbol{M}}(\boldsymbol{x}_*)]}{\mathbf{1}^{\boldsymbol{T}} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{M}}^{-\mathbf{1}} \widehat{\boldsymbol{\sigma}}_{\boldsymbol{M}}(\boldsymbol{x}_*)\mathbf{1}} \tag{3}$$

where $\tau^2 = Var(y)$. All parameters $\boldsymbol{\psi} = (\mu, \tau^2, \boldsymbol{\theta})$ are estimated through *maximum likelihood estimators* (MLE), and they are denoted by a head; i.e., $\widehat{\boldsymbol{\psi}} = (\hat{\mu}, \hat{\tau}^2, \widehat{\boldsymbol{\theta}})$. To estimate $\widehat{\boldsymbol{\psi}}$, we use free-of-charge software DACE which is well-documented in Lophaven et al. (2002) and available at http://www2.imm.dtu.dk/pubdb/pubs/1.

3.2 Leave-one-out cross-validation

We apply leave-one-out cross-validation (LOO-CV) for checking model validity. We summarize this method as follows; see also Kleijnen (2015, p. 95). This method deletes input $\boldsymbol{x_i}$ and its corresponding output $\boldsymbol{y_i}$ from the input/output data set, and re-estimate the parameters $\widehat{\boldsymbol{\psi}}_{-i}$. Then, LOO-CV uses (2) to estimate $\widehat{\boldsymbol{y}}_{-i}(\boldsymbol{x_i})$**.** The error in estimation is given by the difference between the observed and predicted values:

$$e_i = y_i - \widehat{y}_{-i}(x_i). \tag{4}$$

Then, we apply *paired-t test* to errors, where the null hypothesis is $\boldsymbol{H_0}: \boldsymbol{e} = \boldsymbol{0}$ and the alternative hypothesis $\boldsymbol{H_1}: \boldsymbol{e} \neq \boldsymbol{0}$. The test statistic is given by

$$t = \frac{\bar{e}}{s(\bar{e})} \tag{5}$$

with $\bar{\boldsymbol{e}}$ being the sample average and $\boldsymbol{s}(\bar{\boldsymbol{e}})$ the sample standard deviation of $\bar{\boldsymbol{e}}$ (the positive square root of sample variance $\boldsymbol{s^2}(\bar{\boldsymbol{e}})$) as follows with $\boldsymbol{n}$ denoting the number of input data

$$\bar{e} = \frac{\sum_{i=1}^{n} e_i}{n}, s^2(e_i) = \frac{\sum_{i=1}^{n} (e_i - \bar{e})^2}{n-1}, s^2(\bar{e}) = \frac{s^2(e_i)}{n}. \tag{6}$$

We then compute the $\boldsymbol{p}$ −value to conclude on the validity of our prediction model:

$$p = 2 * Prob\{t_{n-1} > t\} \tag{7}$$

where $\boldsymbol{t_{n-1}}$ has a $\boldsymbol{t}$-distribution with $\boldsymbol{n-1}$ degrees of freedom.

3.3 Optimization with chance-constraints

This section introduces an optimization problem with chance constraints based on Kınay et al. (2018). The aim is to select a number of shelter locations from a given set of locations, and allocate districts to their closest open shelters. This study is to be done before an earthquake occurs and can be considered as part of the preparation before the earthquake. We use the following notation (Table 1).

Table 1. Notation for the chance-constrained optimization problem

$\boldsymbol{I}$	:	Set of candidate shelter locations, indexed by $\boldsymbol{i}$
$\boldsymbol{J}$	:	Set of districts, indexed by $\boldsymbol{j}$
$\boldsymbol{w_i}$	:	Weight of candidate shelter location $\boldsymbol{i}$
$\boldsymbol{l_{ij}}$	:	Distance between shelter location $\boldsymbol{i}$ and district $\boldsymbol{j}$
$\boldsymbol{q_i}$	:	Capacity of shelter in location $\boldsymbol{i}$ in square meters
$\boldsymbol{\beta}$	:	Threshold for the minimum utilization rate of a shelter
$\boldsymbol{y_i}$	:	0-1 decision variables, takes value 1 if shelter in location $\boldsymbol{i}$ is open, and 0 otherwise
$\boldsymbol{z_{ij}}$	:	0-1 decision variables, takes value 1 if shelter in location $\boldsymbol{i}$ is allocated to district $\boldsymbol{j}$, and 0 otherwise
$\boldsymbol{w_{min}}$	:	Minimum weight among the open shelters

Maximize $$\boldsymbol{w_{min}} \quad (8)$$

Subject to
$$\boldsymbol{w_{min}} \leq \boldsymbol{w_i y_i} + (\boldsymbol{1} - \boldsymbol{y_i}) \qquad \boldsymbol{i} \in \boldsymbol{I} \quad (9)$$
$$\sum_{\boldsymbol{i}\in \boldsymbol{I}} \boldsymbol{z_{ij}} = \boldsymbol{1} \qquad \boldsymbol{j} \in \boldsymbol{J} \quad (10)$$
$$\boldsymbol{z_{i_j(1),j}} \geq \boldsymbol{y_{i_j(1)}} \qquad \boldsymbol{j} \in \boldsymbol{J} \quad (11)$$
$$\boldsymbol{z_{i_j(r),j}} \geq \boldsymbol{y_{i_j(r)}} - \sum_{\boldsymbol{s}=\boldsymbol{1}}^{\boldsymbol{r}-\boldsymbol{1}} \boldsymbol{y_{i_j(s)}} \qquad \boldsymbol{j} \in \boldsymbol{J},\ \boldsymbol{r} = \boldsymbol{2}, \ldots, |\boldsymbol{l}| \quad (12)$$
$$\boldsymbol{Prob}\left\{\sum_{\boldsymbol{j}\in \boldsymbol{J}} \left(\boldsymbol{3.5} * \hat{\boldsymbol{y}}_{\boldsymbol{j}}(\boldsymbol{x}_*)\right) \boldsymbol{z_{ij}} \leq \boldsymbol{q_i y_i}\right\} \geq \boldsymbol{1} - \boldsymbol{\gamma_i} \qquad \boldsymbol{i} \in \boldsymbol{I} \quad (13)$$
$$\boldsymbol{Prob}\left\{\sum_{\boldsymbol{j}\in \boldsymbol{J}} \left(\boldsymbol{3.5} * \hat{\boldsymbol{y}}_{\boldsymbol{j}}(\boldsymbol{x}_*)\right) \boldsymbol{z_{ij}} \leq \boldsymbol{\beta q_i y_i}\right\} \leq \boldsymbol{\delta_i} \qquad \boldsymbol{i} \in \boldsymbol{I} \quad (14)$$
$$\boldsymbol{y_i} \in \{\boldsymbol{0}, \boldsymbol{1}\} \qquad \boldsymbol{i} \in \boldsymbol{I} \quad (15)$$
$$\boldsymbol{z_{ij}} \in \{\boldsymbol{0}, \boldsymbol{1}\} \qquad \boldsymbol{i} \in \boldsymbol{I}, \boldsymbol{j} \in \boldsymbol{J} \quad (16)$$

The objective function (8) maximizes the minimum weight among all open shelters; these weights for the shelters are determined by their distance to healthcare facilities, electrical infrastructure, and sanitary system. The constraint (9) determines $\boldsymbol{w_{min}}$; i.e., it makes $\boldsymbol{w_{min}}$ equal to the minimum weight of all open shelters. The constraint (10) allocates every district $\boldsymbol{j}$ to exactly one shelter. The constraints (11) and (12) assign district $\boldsymbol{j}$ to the closest open shelter, where the notation $\boldsymbol{i_j(r)}$ means the $\boldsymbol{r}$th closest shelter to district $\boldsymbol{j}$; below, we provide further explanation for these two constraints. The constraints (13) and (14) are the chance constraints, where 3.5 is the minimum area required by one person; (13) implies that the capacity constraint for shelter $\boldsymbol{i}$ is to be satisfied with high probability, where $\boldsymbol{\gamma_i}$ is usually a small number. Furthermore, (14) implies that the under utilization of shelter $\boldsymbol{i}$ is to occur with low probability; i.e., $\boldsymbol{\delta_i}$ is also a small number. The constraints (15) and (16) define the domains of our decision variables. Note that the objective function (8), the constraints (9)-(12) and (15)-(16) are exactly the same as those in Kınay et al. (2018). The formulation of the chance constraints is new, and is one of the contributions of this paper.

In the constraint (13) and (14), instead of using the demand as in Kınay et al. (2018), we use our predictor (2) of injured people of district j, given the specific input data x_* for that district. Because we use GP to model the number of injured people, the predictor $\hat{y}_j(x_*)$ has also a normal distribution with mean in (2) and variance in (3). Furthermore, we assume that the predicted numbers of injuries $\hat{y}_j(x_*)$ for different districts $j \in J$ are independent, which implies that the left-hand side of the chance constraints (13) and (14) is normally distributed with mean $\sum_{j\in J}(3.5 * \hat{\boldsymbol{y}}_{\boldsymbol{j}}(\boldsymbol{x}_*) * z_{ij})$ and variance $\sum_{j\in J}(3.5^2 * s^2[\hat{\boldsymbol{y}}_{\boldsymbol{j}}(\boldsymbol{x}_*)] * z_{ij}^2)$. Due to the normality, these chance constraints can be rewritten as

$$q_i y_i\text{-}\sum_{j\in J}(3.5 * \hat{\boldsymbol{y}}_{\boldsymbol{j}}(\boldsymbol{x}_*) * z_{ij}) - z_{1-\gamma_i}\sqrt{\sum_{j\in J}\left(3.5^2 * s^2[\hat{\boldsymbol{y}}_{\boldsymbol{j}}(\boldsymbol{x}_*)] * z_{ij}^2\right)} \geq 0 \quad (17)$$

$$\beta q_i y_i\text{-}\sum_{j\in J}(3.5 * \hat{\boldsymbol{y}}_{\boldsymbol{j}}(\boldsymbol{x}_*) * z_{ij}) - z_{\delta_i}\sqrt{\sum_{j\in J}\left(3.5^2 * s^2[\hat{\boldsymbol{y}}_{\boldsymbol{j}}(\boldsymbol{x}_*)] * z_{ij}^2\right)} \leq 0. \quad (18)$$

Note that both constraints (17) and (18) are nonlinear, which makes the problem a mixed-integer nonlinear problem (MINLP). In the numerical experiments, we will solve the problem using (8)-(12), (17)-(18), and (15)-(16).

Now, we explain the constraints given in (11) and (12). Consider the data in Table 4, where the number 1 in each district row corresponds to the closest shelter to that district, the number 2 corresponds to the second closest shelter to that district, etc. Now, for District 1, constraint (11) can be written as

$$z_{61} \geq y_6.$$

This constraint assigns District 1 to Shelter 6 if that shelter is open. Furthermore, constraint (12) for District 1 for the next nearest four shelters (i.e., $r = 2, 3, 4, 5$) can be written as

$$z_{31} \geq y_3 - y_6;\ z_{71} \geq y_7 - y_6 - y_3;\ z_{81} \geq y_8 - y_6 - y_3 - y_7;\ z_{51} \geq y_5 - y_6 - y_3 - y_7 - y_8.$$

That is, if Shelter 6 is not open, and Shelter 3 is open, District 1 will be assigned to Shelter 3, which is the second nearest shelter, etc. Hence, the constraints (11) and (12) assign districts to their nearest open shelters.

4. Data Collection

To predict injuries, we collect and use the following input/output (I/O) data that correspond to previous earthquakes occurred in Turkey. As inputs $\boldsymbol{x}$, we use population density (i.e., population divided by the area of the province) given as the 5th column of Table 1 and the total number of undamaged, slightly damaged and moderately damaged buildings given as the 6th column of Table 1. As outputs $\boldsymbol{y}$, we use the total number of injuries, given in the 7th column of Table 1. We use different sources to collect these data. For example, we obtain population data from the Report of 2023 Kahramanmaraş and Hatay Earthquakes (Table 1: Population of Provinces based on 2022 data, page 9), available at https://www.sbb.gov.tr/wp-content/uploads/2023/03/2023-Kahramanmaras-ve-Hatay-Depremleri-Raporu.pdf. Furthermore, we collect areas of provinces (in square kilometer) from Wikipedia, available at https://tr.wikipedia.org/wiki/2023_Kahramanmara%C5%9F_depremleri. The number of undamaged, slightly damaged and moderately damaged buildings are from the earthquake report of Istanbul Technical University (Table 3.4., page 88), available at https://haberler.itu.edu.tr/docs/default-source/default-document-library/2023_itu_deprem_on_raporu.pdf. We present I/O data in Table 2.

Table 2. I/O data for estimating the number of injuries

No	Province	Population	Area	Population density	Number of buildings	Number of Injuries
1	Adana	2274106	13844	164.26	7305	7450
2	Adıyaman	635169	7337	86.57	23617	17499
3	Diyarbakır	1804880	15168	118.99	25482	902
4	Elazığ	591497	9383	63.03	2321	379
5	Gaziantep	2154051	6803	316.63	122924	25276
6	Hatay	1686043	5600	301.07	49227	30762
7	Kahramanmaraş	1177436	14525	81.06	47034	9243
8	Kilis	147919	1521	97.25	5194	754
9	Malatya	812580	12313	65.99	17368	9108
10	Osmaniye	559405	3767	148.50	30341	2606
11	Şanlıurfa	2170110	19242	112.77	33642	8919

5. Validation of the GP model

We do not present the MLE $\hat{\boldsymbol{\psi}}$ of the GP parameters estimated by DACE; yet, in Table 2, we present the GP estimates of the number of injuries in the corresponding province. To estimate $\hat{\boldsymbol{\psi}}$, we consider the logarithm of the number of injuries in Table 1 because this transformation gives us better estimates; we also show the logarithm of the number of injuries in Table 2, the 3rd column, denoted by $\boldsymbol{y}$. We apply LOO-CV, as we detailed in Section 3.2, and estimate the GP predictors $\widehat{\boldsymbol{y}}$ for every province. Furthermore, the last column in Table 2 illustrates the error observation for each province, defined by $\boldsymbol{e_i} = \widehat{\boldsymbol{y}}_{\boldsymbol{i}} - \boldsymbol{y_i}$. We summarize all LOO-CV results in Table 3.

Table 3. LOO-CV results of the GP model

No	Province	Log (# injuries), y_i	GP estimates of injuries $\hat{y}_i$	$e_i = \hat{y}_i - y_i$
1	Adana	8.9160	8.6228	-0.2932
2	Adıyaman	9.7699	8.5925	-1.1774
3	Diyarbakır	6.8046	8.7858	1.9812
4	Elazığ	5.9375	8.3891	2.4516
5	Gaziantep	10.1376	8.4580	-1.6796
6	Hatay	10.3340	8.3993	-1.9347
7	Kahramanmaraş	9.1316	8.7842	-0.3474
8	Kilis	6.6254	8.1650	1.5396
9	Malatya	9.1169	8.8386	-0.2783
10	Osmaniye	7.8656	8.0843	0.2187
11	Şanlıurfa	9.0959	7.6820	-1.4139

We applied paired-t test to the error observations $\boldsymbol{e_i}$ in Table 2; see equations (5), (6), and (7). The $\overline{\boldsymbol{e}}$ value is -0.0849 and $\boldsymbol{s(\overline{e}) = 0.4543.}$ The $\boldsymbol{t}$-statistic in (7) has ten degrees of freedom, so that the computed $\boldsymbol{p}$-value is equal to 0.8555. This high $\boldsymbol{p}$-value implies that we cannot reject $\mathbf{H_0}: \mathbf{e} = \mathbf{0}$. Hence, the GP process gives reasonable estimates of the number of injuries for the given I/O data.

6. Results and Discussion

This section is further divided into three subsections. The first subsection discusses the numerical results of the chance-constrained problem. The second subsection presents some graphical results. The third subsection discusses further improvements for both our prediction and optimization methods.

6.1 Numerical Results

We consider ten candidate shelter locations to be allocated to four districts. We randomly generate distances from the center of the districts to shelter locations using uniform distribution $\boldsymbol{U(5, 12).}$ Furthermore, for this hypothetical problem, we assume that the population densities in these four districts are again uniform $\boldsymbol{U(66, 301)}$, where we used the second smallest and second biggest population densities in Table 1, column 5, as the lower and upper bounds, respectively. Moreover, to generate the number of damaged buildings for the four districts, we again used uniform distribution $\boldsymbol{U(7305, 49227)}$, where we again obtained the lower and upper bounds from Table 1, column 6, third smallest and second biggest values. To estimate the parameters of the GP model, we used all I/O data (log of output) in Table 1. We generate the distance matrix only once, but the population densities and the number of damaged buildings for 25 times, and solved the problem defined in (8)-(12), (17)-(18), and (15)-(16) for each sample. We present the sampled data below in Table 4, where instead of distances between districts and shelters, we show the first, second, third, etc. nearest shelter to each district.

Table 4. Parameters of the optimization problem

District \ Shelter	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
D1	6	7	2	8	5	1	3	4	9	10
D2	3	7	9	2	6	1	10	5	8	4
D3	1	3	10	7	4	2	6	5	8	9
D4	3	6	5	8	10	2	1	9	4	7
Capacity	43575	37845	33424	44121	30637	35538	50586	47371	37927	53756
Weight	0.49	0.96	0.34	0.59	0.22	0.75	0.26	0.51	0.69	0.89

Furthermore, we set $\beta = 70\%$, $\gamma_i = 1\%$ for each district i, and $\delta_i = 5\%$ for each district $\boldsymbol{i}$. For each of the sampled population density and the sampled number of damaged houses, we solved the resulting MINLP using BARON solver of GAMS 24.8. Five of these 25 samples resulted in infeasible problems, suggesting that the current capacities of the shelters are not enough if more than 12,000 injuries occur in one of the districts. Furthermore, shelters S5, S6, and S9 were not open in any one of the samples. Shelter S8 was open in 18 out of 20 feasible samples, and shelter S2 was open 16 out of 20 feasible samples. Based on these results, S2 and S8 can be considered as two shelters that have the most potential to be open, and D1 and D2 are to be assigned to S8, and D3 and D4 are to be assigned to S2.

6.2 Graphical Results

Figure 1 shows that the GP predictor and the logarithm of the true value of the output (i.e., number of injuries in Table 2) are exactly the same; so, the sum of the training errors is zero because GP is an exact interpolator.

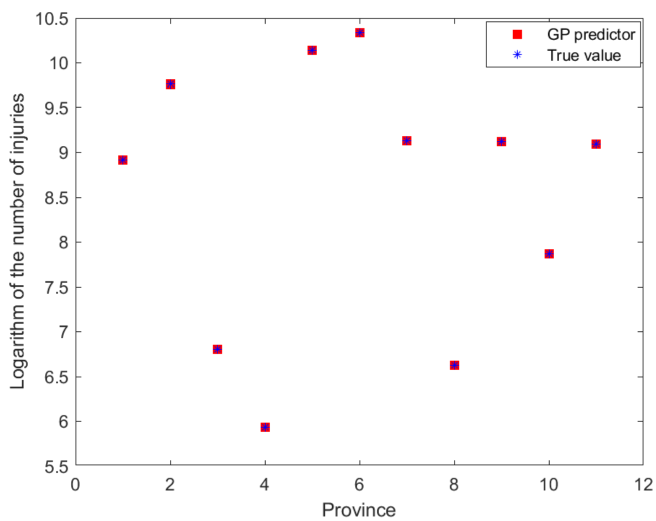

Figure 1. The GP predictors versus the logarithm of the number of injuries

6.3 Proposed Improvements

This paper can be further improved by comparing the GP predictor with other predictors in the literature. Furthermore, the optimization problem should be solved by using different values of the minimum utilization rate $\boldsymbol{\beta}$, which is set to 70% currently. We expect that the smaller this parameter is, the bigger the number of shelters will be open as the solution of the optimization problem.

7. Conclusion

This paper aims at modeling the unknown number of injured people in the aftermath of an earthquake through GP, which has not been used in fatality/injury prediction in the literature. Furthermore, the predicted injuries and their variances are used in a chance-constrained optimization problem to find the location of shelters from a finite number of candidate locations. We solved the resulting MINLP by sampling 25 population density and number of damaged houses using uniform distributions, where the parameters were selected from the real I/O. We have observed that if the number of injured people exceeded 12,000, the current capacities of the shelters were not enough; so, we could not find feasible solutions. This suggests that we should either increase the capacities of the shelters or divide the current districts into subdistricts, and assign them to different shelters. Furthermore, with the current setting of the parameters, only in one sample out of 20, there were three open shelters, and for the remaining 19 samples, there were only two open shelters. Finally, there were three shelters which were not open in any one of the sampled problems. Further research should consider different prediction methods, compare GP with these methods, and formulate new optimization problems that use the predictions with their variances.

References

Bastami, M., & Soghrat, M. R. An empirical method to estimate fatalities caused by earthquakes: the case of the Ahar–Varzaghan earthquakes (Iran). *Natural Hazards*, 86, 125-149, 2017.

Erdik, M., Tümsa, M. B. D., Pınar, A., Altunel, E., & Zülfikar, A. C. A preliminary report on the February 6, 2023 earthquakes in Türkiye. *Research Briefs*., 2023.

Fang, Z., Huang, J., Huang, Z., Chen, L., Cong, B., & Yu, L. An earthquake casualty prediction method considering burial and rescue. *Safety Science*, 126, 104670, 2020.

Firuzi, E., Hosseini, K. A., Ansari, A., & Tabasian, S. Developing a new fatality model for Iran's earthquakes using fuzzy regression analysis. *International Journal of Disaster Risk Reduction*, 80, 103231, 2022.

Guha, S., Jana, R. K., & Sanyal, M. K. Artificial neural network approaches for disaster management: A literature review. *International Journal of Disaster Risk Reduction*, 81, 103276, 2022.

Gul, M., & Guneri, A. F. An artificial neural network-based earthquake casualty estimation model for Istanbul city. *Natural Hazards*, 84, 2163-2178, 2016.

Kınay, Ö. B., Kara, B. Y., Saldanha-da-Gama, F., & Correia, I. Modeling the shelter site location problem using chance constraints: A case study for Istanbul. *European Journal of Operational Research*, 270(1), 132-145, 2018.

Kleijnen, J. P. Design and analysis of simulation experiments. *Springer International Publishing*, 2015.

Li, B., Gong, A., Zeng, T., Bao, W., Xu, C., and Huang, Z. A zoning earthquake casualty prediction model based on machine learning., *Remote Sensing*, 14(1), 30, 2021.

Liu, Y., Wang, Z., & Zhang, X. Estimating the final fatalities using early reported death count from the 2023 Kahramanmaraş, Türkiye, MS 8.0–7.9 earthquake doublet and revising the estimates over time. *Earthquake Research Advances,* 5(2), 100331, 2025.

Lophaven, S.N., Nielsen, H.B., Sondergaard, J.: DACE: a Matlab Kriging toolbox, version 2.0., IMM Technical University of Denmark, Kongens, Lyngby, 2002.

Maguraushe, K., Ndayizigamiye, P., & Bokaba, T. Trends and Developments in the Use of Machine Learning for Disaster Management: A Bibliometric Analysis. *In International Working Conference on Transfer and Diffusion of IT* (pp. 92-104). Cham: Springer Nature Switzerland, December, 2023.

Ren, Y., Ma, D., Wang, W., Wang, Z., Zhao, X., & Zhu, M. Probabilistic assessment of earthquake casualties in residential areas*. International Journal of Disaster Risk Reduction*, 113, 104902, 2024.

Shi, Y., Li, Y., & Zhang, Z. Reevaluating earthquake fatalities in the Taiwan Region: Toward more accurate assessments. *Seismological Research Letters*, 95(3), 1939-1948, 2024.

Sun, W., Bocchini, P., and Davison, B. D. Applications of artificial intelligence for disaster management., *Natural Hazards*, 103(3), 2631-2689, 2020.

Tang, B., Chen, Q., Liu, X., Liu, Z., Liu, Y., Dong, J., & Zhang, L. Rapid estimation of earthquake fatalities in China using an empirical regression method. *International Journal of Disaster Risk Reduction*, 41, 101306, 2019.

Xu, Z., Zhu, Y., Fan, J., Zhou, Q., Gu, D., & Tian, Y. A spatiotemporal casualty assessment method caused by earthquake falling debris of building clusters considering human emergency behaviors. *International Journal of Disaster Risk Reduction*, 105206, 2025.

Williams, C.K., Rasmussen, C.E. *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT press, 2006.

Zhou, H., Che, A., Shuai, X., and Zhang, Y. A spatial evaluation method for earthquake disaster using optimized BP neural network model., Geomatics, *Natural Hazards* and Risk, 14(1), 1-26, 2023.

Biographies

Şule Nur Sargın is a student in the Master Program of Industrial Engineering, Galatasaray University. Bachelor's degree in Industrial Engineering, Industrial Engineering Department, Istanbul University. She is currently research assisstant at Istanbul Medeniyet University. Her research interests are operation research, simulation, optimization.

Ebru Angün is Associate Professor of Operations Research in the Department of Industrial Engineering at Galatasaray University in Istanbul. Her research interests include simulation optimization, stochastic optimization, and applications in disaster management. She has been an author and reviewer for many international journals, and a project reviewer for the National Science Foundation of Turkey and the Research Foundation Flanders (FWO) in Belgium. She spent one year at the Georgia Institute of Technology, as a visiting scholar. Also see https://avesis.gsu.edu.tr/eangun and https://www.researchgate.net/profile/ Ebru-Angun