Spatial Evaluation of Earthquake Disaster Based On Correlation Characteristics And BP Neural Network

Hanxu Zhou

Shanghai Jiao Tong University

Ailan Che (✉ alche@sjtu.edu.cn)

Shanghai Jiao Tong University

Xianghua Shuai

Shenzhen Academy of Disaster Prevention and Reduction

Research Article

Keywords: Earthquake-hit population, Spatial evaluation, Correlation characteristic, BP neural network, Sample optimization, Generalization capability.

Posted Date: October 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-956348/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Rapid spatial evaluation of disaster after earthquake occurrence is required in the emergency rescue management, due to its significant support for decreasing casualties and property losses. The earthquake-hit population is taken as an example of earthquake disaster to construct the evaluation model using the data from the 2013 *Ms7.0* Lushan earthquake. Ten influencing factors are classified into environmental factors and seismic factors. The correlation analysis reveals characteristics that there is a nonlinear relationship between the earthquake-hit population and various factors, and per capita GDP and PGA factor have a stronger correlation with earthquake-hit population. Moreover, the spatial variability of influencing factors would affect the distribution of earthquake-hit population. The earthquake-hit population is evaluated using BP neural network with optimizing training samples based on the spatial characteristics of per capita GDP and PGA factors. Different number of sample points are generated in areas with different value intervals of influencing factors, instead of the random distribution of sample points. The minimum value of RMSE (Root Mean Square Error) from testing set is 18 people/km², showing good accuracy in the spatial evaluation of earthquake-hit population. Meanwhile, the optimizing samples considering spatial characteristics could improve the convergence speed and generalization capability comparing to random samples. The trained network was generalized to the 2017 *Ms7.0* Jiuzhaigou earthquake to verify the prediction accuracy. The evaluation results indicate that BP neural network considering the correlation characteristics of factors has the capability to evaluate the seismic disaster information in space, providing more detailed information for emergency service and rescue operation.

1. Introduction

Earthquake disasters have a profound impact on human living environment due to their suddenness and destructiveness. Severe casualties, house collapse and economic loss would be caused under the action of intense seismic ground motion. Strong earthquakes have continued to appear worldwide in the past two decades (Rossetto et al., 2007; Lara et al., 2016; Shimada, 2016). People injured or killed by the earthquake could range from a few to tens of thousands of people, distributed in different spatial locations (Zhao et al., 2018). Although the concerns of seismic problem continue to deepen and the seismic awareness of human is constantly enhanced, within the past decades, the active activities of geological structures are still affecting the environment of anthroposphere (Sun et al., 2016; Wu et al., 2020; Santos-Reyes and Gouzeva, 2020; Luo et al., 2021). Due to the unpredictability of earthquake occurrence, it is difficult to prepare before earthquake, so countries are committed to improve the emergency rescue ability after earthquake (Huang and Li, 2014). Among the various types of earthquake disaster information, modeling the earthquake casualty is particularly important for offering reference for emergency rescue and decision making. Casualty evaluation after earthquake is fast becoming an important issue increasingly responsible for significant economic, social, and environmental risk management (Huang and Huang, 2018).

The disaster under intense seismic motion is a complex result of various influencing factors. Seismic intensity, topography, population and economic level are all related to casualties due to earthquake to a certain extent. The traditional physical model or statistical regression model is difficult to reflect the nonlinear relation between earthquake-hit population and factors (Erdik et al., 2011). With the continuous improvement of computing speed in recent years, machine learning methods have been more widely used. More and more scholars apply them for disaster mapping under earthquake considering that machine learning methods could provide the ability to learn from historical data for producing insight into extreme events (Yang et al., 2015; Choubin et al., 2019; Pourghasemi et al., 2019; Jena et al., 2020; Hou et al., 2020; Si and Du, 2020; Luo et al., 2020). Aghamohammadi et al. (2013) used artificial neural network (ANN) for estimating the human loss of building damage under earthquake based on the data of the 2003 Bam earthquake. Huang et al. (2015) proposed a robust wavelet (RW) v-SVM (support vector machine) earthquake casualty prediction model. The factors including earthquake magnitude, intensity, population density, pre-warning level, in-building probability, location of occurrence, supply support and building collapse ratio were considered. It was concluded that RW v-SVM model had higher prediction accuracy and quicker learning than standard SVM and neural network. Gul and Guneri (2016) built up an ANN model for casualty prediction taking occurrence time, magnitude, population density as factors. Data of 21 earthquakes in Turkey were collected as samples for network training. Huang et al. (2020) established Extreme Learning Machine (ELM) network to predict earthquake casualty based on the data of 84 groups of earthquake victims in China. It was found that the ELM algorithm had better robustness and generalization capability than BP neural network and SVM. It can be noticed that the existing researches focused on the prediction of population numerical value

affected by earthquake considering factors of multiple dimensions. Moreover, the accuracy and performance of different machine learning methods were compared based on the evaluation results of earthquake casualty. However, the input layer of different machine learning methods used numerical data without spatial information, and the spatial characteristics of disaster information of output layer have not been evaluated effectively. For earthquake emergency management, the spatial distribution of disaster information within the earthquake affected area has a greater significance for the formulation of detailed rescue plans.

The generalization capability of network refers to the ability to obtain accurate output when inputting new data other than training samples. Generalization capability is the most important index to measure the performance of network. The complexity of structure and samples are the main factors affecting the generalization capability of model. Research of Partridge (1996) on three-layer neural network found that the influence of training set on generalization capability is great, even more than the influence of neural number. Many researchers combine principal component analysis (PCA), clustering analysis and other methods with machine learning to optimize the training set aiming to improve the generalization capability of the network (Basharat et al., 2016; Li et al., 2020). Lou et al. (2012) used PCA to reduce the dimension of assessment factors, disaster-formative environment and disaster-affected bodies, and established a BP neural network to assess the economic loss under tropical cyclones in Zhejiang Province. A combined use of PCA and ANN was adopted by Gao et al. (2020) to evaluate the personal exposure level to PM_{2.5}, and it was found that the combined use of PCA and ANN produced more accurate results than simple ANN method. It can be seen that optimizing the input samples of network could improve the generalization capability. Most of the existing sample optimization methods are based on statistical analysis on numerical dimensions. The distribution of influencing factors and training results in the spatial dimension are also related. Sample optimization based on spatial correlation characteristics might provide a novel solution to improve the generalization capability.

The study presented herein aimed at effective evaluating the spatial distribution of earthquake disaster information in each county. The earthquake-hit population spatial distribution was selected as the study content and evaluated based on correlation characteristics of influencing factors and BP neural network, using data from the 2013 *Ms7.0* Lushan earthquake. The selection of samples was optimized based on the spatial characteristics resulting from correlation analysis, to improve the generalization capability of network and accuracy of evaluation results.

2. Influencing Factors Of The Earthquake Disaster In The 2013 *Ms7.0* Lushan Earthquake

2.1 Earthquake-hit population in the 2013 *Ms7.0* Lushan earthquake

The Lushan *Ms7.0* earthquake occurred on April 20, 2013 and the epicenter was located at 30°18'N, 103°56'E, in Lushan county, Sichuan province, China. The focal depth of the Lushan earthquake was 13km. The affected-area of the Lushan earthquake was at the junction of Qinghai Tibet Plateau and Sichuan Basin. The Lushan earthquake was caused by the tectonic activity of Longmenshan fault zone, the same as the 2008 *Ms8.0* Wenchuan earthquake. The distance between Lushan earthquake and Wenchuan earthquake epicenter was about 85 km. A total of 196 people were killed, 21 were missing and 11470 injured in the Lushan earthquake. The Lushan earthquake affected an area of 12500 km² and caused a direct economic loss of about 185.4 billion yuan. After the earthquake, Sichuan Province immediately started the first level emergency procedures and sent out army to carry out emergency rescue work.

The total number of earthquake-hit population under the Lushan earthquake reached 3.7 million. The earthquake-hit population refers to the people who have suffered property or life losses due to the earthquake. The earthquake-hit population not only reflects the severity of natural disasters, but also reveals the impact of earthquake on people's lives. The earthquake-hit population also offers references in the formulation of emergency rescue plan, leading to a fact that the number of earthquake-hit population has become an important index to evaluate the damage caused by earthquake. Figure 1 illustrates the earthquake-hit population density (number of earthquake-hit people per square kilometer) in each county-level administrative region within the earthquake affected area. The data was collected and released by Sichuan provincial government after the earthquake on the Internet (Wang and Li, 2014).

In Figure1, the color represents the earthquake-hit population density. The earthquake-hit population density in Yucheng District, Danling County and Mingshan County was relatively higher. The earthquake-hit population density in Mingshan County was the highest, reaching approximately 432 people/km². It can be observed that the region with the highest earthquake-hit population density was not the region where the epicenter was located. It indicated that the impact of earthquakes on population is complicated in space, the epicenter is not necessarily the most severely affected area under seismic motion. Similar phenomenon had appeared in the 2008 *Ms8.0* Wenchuan earthquake (Yang et al., 2014). The casualties caused by earthquakes are related to many categories of influencing factors. The influence factors of earthquake-hit population are divided into environmental and seismic factors.

2.2 Environmental influencing factors

The environmental influencing factors refer to the environmental conditions in the study area, and there is no direct relationship between environmental factors and earthquake occurrence. The environmental influencing factors considered in the research contained elevation, slope angle, population density, per capita GDP, distance to fault, distance to river and Normalized Difference Vegetation Index (NDVI). The data details of environmental influencing factors are shown in Table 1. In the environmental factors, elevation, slope angle, distance to fault, distance to river and NDVI are the maps with data varying with spatial location. However, population density and per capita GDP have inconsistent gradation with other environmental factors. There is one attribute value for each county level administrative region for these two factors, because counties were used as the basis unit of statistics.

Table 1
Data details of the environmental influencing factors

Data	Sources	Scale/resolution	Update time
Elevation	Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences (http://www.gscloud.cn)	30×30m	2009
Slope angle	Calculated from digital elevation model	30×30m	2009
Population density	China Earthquake Network Center	County level	2011
Per capita GDP	China Earthquake Network Center	County level	2011
Distance to fault	China Earthquake Network Center	1:100000	2011
Distance to river	China Earthquake Network Center	1:100000	2011
NDVI	Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences (http://www.gscloud.cn)	30×30m	2012

2.2.1 Elevation

Elevation is considered to be the most important factor in the analysis of natural disaster susceptibility (Peng et al., 2014; Tehrany et al., 2015; Saha et al., 2020). There is also correlation between elevation and distribution of earthquake-hit population. On the one aspect, the population is concentrated on the plains with lower elevations; on the other aspect, there is slope amplification effect on seismic ground motion resulting in the more severe geological disasters in high elevation areas (Zhang et al., 2018). The digital elevation model with a resolution of 30×30m updated in 2009 (Figure 2(a)) was obtained from Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences.

2.2.2 Slope angle

The slope angle is a geomorphic parameter which has an important impact on seismic geological disasters such as landslide, debris flow, barrier lake, etc. In the field investigation of historical strong earthquakes, it was found that a large number of earthquake casualties were caused by geological disasters triggered by seismic motion (Xu et al., 2015). The slope angle map (Figure 2(b)) was derived from the digital elevation model using ArcMap software.

2.2.3 Population density

The population density is a key factor in risk assessment of natural disaster. It is calculated as the ratio of population to bare land area here. Some strong earthquakes occurred in mountainous areas with low population density, which posed a relatively small threat to people's lives and property (Ara, 2014). Because of the strong mobility of population, it is difficult to obtain the spatial distribution of population at the moment before earthquake occurrence. Therefore, the resident population of each county in census was applied to approximate the population distribution (Figure 2(c)). The population density is the ratio of population to area in each county. The population data were updated in 2011 and offered by China Earthquake Network Center.

2.2.4 Per capita GDP

The per capita GDP is another crucial factor in earthquake-hit population evaluation. The per capita GDP reflects the economic level of local people from one aspect, and the economic level would affect the seismic resistance ability of engineering constructions. Generally, the higher the economic level is, the stronger seismic resistance ability the constructions have. Similar to population density map, the per capita GDP of each county was applied as per capita GDP distribution map (Figure 2(d)). The per capita GDP data were updated in 2011 and offered by China Earthquake Network Center.

2.2.5 Distance to fault

The distance to fault is another significant factor related to seismic geological disasters. Generally, fractured or weak zones are located near fault bedding planes, which are susceptible to weathering and sliding (Conforti et al., 2014). The fault data were offered by China Earthquake Network Center and the distance to fault was calculated using buffers of ArcMap software (Figure 2(e)).

2.2.6 Distance to river

The distance to river can also influence the earthquake-hit population, as river erosion and soil saturation would decrease the seismic stability of slopes (Yalcin, 2008). The river data were offered by China Earthquake Network Center and the distance to river was calculated using buffers of ArcMap software (Figure 2(f)).

2.2.7 NDVI

The NDVI can play a feasible role in the earthquake-hit population evaluation under the Lushan earthquake. NDVI quantifies vegetation by measuring the difference between near infrared (vegetation strong reflection) and red light (vegetation absorption). The closer the NDVI is to +1, the better the vegetation coverage in the area is, and the lower the degree of urbanization is. The NDVI map was obtained from Landsat 7 ETM+ satellite images acquired in 2012 from Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences (Figure 2(g)).

2.3 Seismic influencing factors

The seismic influencing factors refer to the elements and characteristics that are directly related to earthquake occurrence, and can be rapidly reflected by seismic motion monitoring instruments after earthquake occurrence. The seismic influencing factors considered in this research contained peak ground acceleration (PGA), peak ground velocity (PGV) and distance to epicenter. The data of seismic influencing factors are shown in Table 2.

Table 2
Data details of the seismic influencing factors

Data	Sources	Scale/resolution
PGA	China Earthquake Network Center	30×30m
PGV	China Earthquake Network Center	30×30m
Distance to epicenter	Internet	/

2.3.1 PGA

The PGA distribution map is the most commonly used parameter to describe the seismic ground motion intensity of an earthquake (Boatwright et al., 2003; Yuan et al., 2013). PGA represents the peak value of acceleration time-history waveform recorded on the ground surface during earthquake occurrence. It can be considered as the maximum instantaneous force exerted by earthquake, and can effectively evaluate the intensity of seismic ground motion at different positions in space. The PGA data were recorded by strong-motion seismograph network which could be offered in a short time after earthquake occurrence. The data used in this research were supported by China Earthquake Network Center (Figure 3(a)).

2.3.2 PGV

PGV is also an important index to evaluate the intensity of seismic ground motion. The acceleration time-history wave would miss some information, such as low-frequency components. The velocity time-history wave can better record this information. Therefore, PGV data are considered to comprehensively evaluate the seismic ground motion intensity. The PGV data were supported by China Earthquake Network Center (Figure 3(b)).

2.3.3 Distance to epicenter

Distance to epicenter is a parameter to measure the relative distance between the study region and the epicenter. In previous researches, it was shown that with the increase of distance to epicenter, the impact of earthquake disaster was gradually reduced. The distance to epicenter was calculated using buffers of ArcMap software (Figure 3(c)).

3. Spatial Correlation Characteristics Of Influencing Factors

Earthquake-hit population is related to environmental and seismic influencing factors. The spatial distribution variability of influencing factors leads to the difference of earthquake-hit population in different counties. The Spearman rank correlation coefficients were calculated to analyze the relationship between earthquake-hit population and influencing factors, which can be expressed as follow in Eq. (1):

$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

1

where D_i denotes the ranking difference of earthquake-hit population and factor; n is the number of samples. The coefficient measures the degree of consistency and describes the strength of monotonicity between earthquake-hit population and influencing factors (Peng, 2015; Shahaki and Celikag, 2019). The coefficient ranges from 1 to -1. When the coefficient is positive, the data are positively correlated. The closer the absolute value of coefficient is to 1, the stronger correlation the data have.

Within the study region shown in Figure 1, 1000 sampling points were randomly generated and distributed using ArcMap software. The earthquake-hit population density data and influencing factor data at sampling points were extracted to construct the database for correlation analysis.

3.1 Correlation analysis between earthquake-hit population and environmental influencing factors

The Spearman correlation coefficient and significance test results between earthquake-hit population and environmental influencing factors are listed in Table 3. The results of significance test were all less than 0.05. It showed that the number of samples was reasonable and the value of correlation coefficient was effective. The highest correlation coefficient was -0.322 between earthquake-hit population density and per capita GDP, showing that in the regions with high per capita GDP, the earthquake-hit population was generally low. The minimum value of coefficient was -0.073. The results of correlation coefficient revealed that among various environmental influencing factors, none of them had a direct linear relationship with earthquake-hit population density. The earthquake-hit population density distribution is the result of multiple environmental factors.

Table 3
Spearman correlation coefficient results with environmental influencing factors

		Elevation	Slope angle	Population density	Per capita GDP	Distance to fault	Distance to river	NDVI
Earthquake-hit population density	Correlation coefficient	-0.187	-0.073	0.260	-0.322	0.181	-0.180	0.105
	Significance test	5.58×10^{-9}	2.36×10^{-2}	3.19×10^{-16}	2.08×10^{-24}	1.78×10^{-8}	2.09×10^{-8}	1.12×10^{-3}

Moreover, the spatial correlation between earthquake-hit population density and environmental factors are illustrated in Figure 4. The histogram statistics the average earthquake-hit population density within the different factor intervals, which can reflect the related characteristics between spatial distribution of earthquake-hit population density and influencing factors. It can be seen from Figure 4 that in various factors, the average earthquake-hit population density in different intervals was discrete. However, relatively high earthquake-hit population density would concentrate in the specific range of certain factors. For example, in the 2013 Lushan earthquake, area between 500m and 1200m in elevation had relatively high earthquake-hit population density (Figure 4(a)). As regards the relationship between earthquake-hit population density and distance to river, the result showed that the maximum earthquake-hit population density was observed in the interval closest to river (Figure 4(f)).

3.2 Correlation analysis between earthquake-hit population and seismic influencing factors

The Spearman correlation coefficient and significance test results between earthquake-hit population and seismic influencing factors are listed in Table 4. The results of significance test were all less than 0.05. The correlation coefficient between earthquake-hit population density and PGA had the highest value of 0.433. It showed that the earthquake-hit population density had remarkable PGA positive correlation, and the higher the PGA, the greater the earthquake-hit population.

Table 4
Spearman correlation coefficient results with seismic influencing factors

		PGA	PGV	Distance to epicenter
Earthquake-hit population density	Correlation coefficient	0.433	0.215	-0.186
	Significance test	1.05×10^{-44}	1.82×10^{-11}	7.60×10^{-9}

The spatial correlation between earthquake-hit population density and seismic influencing factors are illustrated in Figure 5. It can be observed that PGA had a stronger positive correlation with the earthquake-hit population density. The earthquake-hit population density in each interval generally increased with the value of PGA. However, when PGA value ranged from 750 m/s² to 850 m/s², the earthquake-hit population density had relatively lower value (Figure 5(a)). The reason was that the area with

great seismic motion intensity had a low population density. Therefore, the population density affected by the earthquake was relatively lower.

Correlation coefficient is a statistical index to reflect the close degree of linear correlation between variables. The results of correlation analysis indicated that there is a nonlinear relationship between the earthquake-hit population and various influencing factors in spatial distribution. The correlation coefficients imply the existence of spatial correlation between factors and disaster results. For example, in the area with greater seismic motion, the earthquake-hit population density was higher. Nevertheless, the numeric value of correlation coefficients indicates that the earthquake-hit population is a result of the complex interaction of multiple factors. It is difficult to evaluate the spatial earthquake-hit population with a linear relation.

4. Earthquake-hit Population Spatial Evaluation Using Bp Neural Network

4.1 BP neural network

By imitating the information transmission principle of biological neuron, artificial neural network has a powerful ability of data processing (Wang et al., 2016). BP (Back Propagation) neural network is a common model of artificial neural network for weight training, widely applied in the forecasting area. It can more perfectly reflect the mapping relationship between neurons and has the function of approximating non-linear functions arbitrarily (Wang et al., 2018). The BP neural network consists of input layer, hidden layer and output layer, and each layer comprises several neurons. The mapping relationship between input layer and output layer is jointly determined with the activation function and threshold in the hidden layer. In the research, the input layer contained environmental and seismic influencing factors of earthquake-hit population, and the output layer calculated the value of earthquake-hit population density (Figure 6). Moreover, the hidden layer was constructed through model training based on the sample data in the 2013 *M_s7.0* Lushan earthquake. The network building process was divided into three steps: forward calculation, error back propagation and weight update (Wen and Yuan, 2020). The Sigmoid was selected as the excitation function, gradient descent algorithm was used to reach the best solution of network and the learning rate was 0.02.

4.2 Sample optimization selection based on correlation characteristics

Generalization capability is an important indicator to measure the accuracy of the neural network for predicting data outside the sample. In order to effectively evaluate the spatial distribution of the earthquake-hit population in a newly occurred earthquake, it is necessary to ensure that the network has a preferable generalization capability. Training samples have a great impact on generalization capability (Partridge, 1996), so the selection of samples is optimized based on the results of correlation analysis.

In the process of network training, to uniform the format of influencing factor data and earthquake-hit population data, the vector images were transformed into raster layer data. The number of raster-based samples was far greater than the number of influencing indicators, which could result in the over fitting of neural network. In practical application, a part of all samples would be randomly selected as the training set. However, the selection of random samples might lose part of the spatial characteristics of all sample data in the training process, thus reducing the generalization capability and evaluation accuracy of the network. Based on the correlation characteristic between influencing factors and earthquake-hit population density, it was observed that compared with other factors, there was a stronger correlation between per capita GDP, PGA and the earthquake-hit population density. It implied that the area with lower values of per capita GDP and higher values of PGA had a greater number of earthquake-hit population, which was the key study area of earthquake casualty evaluation. The per capita GDP and PGA indicator indirectly reflected the spatial distribution characteristics of earthquake-hit population to some extent. Therefore, more samples were selected in the area with lower values of per capita GDP and higher values of PGA, and fewer samples were selected in the area with higher values of per capita GDP and lower values of PGA to consider the spatial variability of earthquake-hit population.

Figure 7 illustrates the frequency histograms of per capita GDP and PGA factor. It can be seen that the frequency of raster was approximately normal to per capita GDP value, and the frequency of raster decreased with the increase of PGA value. It indicated that the area of poor economy and intense seismic motion was much smaller than that of strong economy and weak motion. Although the area was small, it was the predominant area of disaster assessment and emergency rescue under earthquake. The

per capita GDP and PGA data were classified into five clusters using Natural Breaks Classification method according to the value, respectively. The Natural Breaks Classification method is an extensively applied clustering method to maximize the internal similarity of each cluster and the difference between clusters. The proportion of samples was determined on the basis of the average value of clusters as listed in Table 5. The proportion of samples was proportional to the average of clusters. A total number of 1000 sample points were generated in the study area to extract attribute values from raster layers. The distribution comparison between random sample points and optimizing sample points is showed in Figure 8. It can be seen that according to the result of clustering analysis, in the area with low per capita GDP values and high PGA values, the sample points were denser. The samples were optimized based on the numerical and spatial characteristics of per capita GDP and PGA indicator.

Table 5
Range of PGA clusters and number of sample points

Cluster	Factor	Range	Average	Proportion	Number of sample points
1	Per capita GDP ($\times 10^4$ RMB)	0-1.85	1.54	15.7%	157
2		1.85-2.73	2.35	9.7%	97
3		2.73-3.47	3.22	7.8%	78
4		3.47-5.40	3.99	5.7%	57
5		5.40-8.10	6.44	3.7%	37
1	PGA (cm/s^2)	0-114.66	62.55	2.1%	21
2		114.66-242.11	174.02	5.7%	57
3		242.11-402.22	311.52	10.3%	103
4		402.22-601.56	489.12	16.1%	161
5		601.56-840.11	700.74	23.2%	232
Total	–	–	–	100%	1000

4.3 Earthquake-hit population evaluation results

Two networks based on the random samples and optimizing samples were trained, respectively. One hidden layer is sufficient for most of the applications (Aghamohammadi et al. 2013). Therefore, a three-layer network containing one input layer, one hidden layer and one output layer was adopted. In order to ensure the scientific of the evaluation, 70% of the sample data were used as training set, and 30% of the sample data were used as testing set. The testing set was a sample set to test the classification ability of the trained network. The normalized goal error of training set was 0.0003. When the error of the training set during the iteration was less than 0.0003, the training process stops.

Figure 9 showed the results of iteration number and Root Mean Square Error (RMSE) of two networks based on random samples and optimizing samples. The mathematical expression of RMSE is shown in the following equation:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_i - y_i)^2}$$

2

where t_i is the evaluation result of earthquake-hit population; y_i is the actual data of earthquake-hit population. Networks with different numbers of neurons in the hidden layer was trained. In Figure 9, the horizontal axis was the number of neurons, the left vertical axis was the number of iterations required for the error of training set to reach the goal error, and the right vertical axis was the RMSE of testing set. Moreover, the blue curve represented the results from random samples and the red represented the

ones from optimizing samples. It can be observed that in networks with different numbers of neurons, the number of iterations based on the optimizing samples was less than that of random samples. It indicated that the optimizing samples accelerated the convergence speed. In the meanwhile, by comparing the RMSE of the networks, it can be noted that the RMSE of the optimizing samples was smaller than that of the random samples, except when the number of neurons was 13. When the number of neurons in the hidden layer was 13, the RMSE of the two networks was close to each other as illustrated in Figure 9. The RMSE measures the difference between estimation data and testing set, and reflects the generalization capability and evaluation accuracy of network for new data. It implied that the earthquake-hit population evaluation based on optimizing samples not only had faster convergence speed, but also had better generalization capability and prediction accuracy comparing to random samples.

4.4 Verification of network generalization capability

In order to verify the evaluation accuracy of the trained network obtained from the samples of the 2013 *Ms7.0* Lushan earthquake, they were applied to evaluate the earthquake-hit population in the 2017 *Ms7.0* Jiuzhaigou earthquake. The Jiuzhaigou *Ms7.0* earthquake happened on August 8, 2017 and its epicenter was located at 33°12'N, 103°49'E, in Jiuzhaigou county, Sichuan province, China. The epicenter of Jiuzhaigou earthquake was approximately 330 km away from that of Lushan earthquake. The focal depth of the Jiuzhaigou earthquake was 20 km. The Jiuzhaigou earthquake killed 25 people, injured 525 people and damaged about 70000 houses. A number of about 220000 earthquake-hit population was caused. The area severely affected by the earthquake mainly included Hongyuan County, Jiuzhaigou County, Pingwu County, Songpan County and Zoige County. Jiuzhaigou earthquake and Lushan earthquake had the same magnitude, and the earthquake affected areas of them had similar geographical conditions. Therefore, the Jiuzhaigou earthquake was selected to verify the generalization capability and effectiveness of the earthquake-hit population evaluation. The earthquake-hit population density in each county-level administrative region collected by Sichuan provincial government was illustrated in Figure 10.

The networks based on optimizing samples and random samples of Lushan earthquake were used to evaluate the earthquake-hit population under Jiuzhaigou earthquake separately. The output of network was a raster layer with earthquake-hit population density data varying with spatial position. The average value of the earthquake-hit population density raster data within the county area was calculated as the earthquake-hit population density of this county. The earthquake-hit population is the product of earthquake-hit population density and county area.

The actual data and evaluation results of earthquake-hit population were listed in Table 6. There were deviations between estimation value and actual value in each area. For the total number of earthquake-hit population, the evaluation result of optimizing samples was much closer to actual data than that of random samples. The actual data of earthquake-hit population in Jiuzhaigou earthquake was 216597. The evaluation results of earthquake-hit population were 198362 and 347207 for optimizing samples and random samples respectively. The mean absolute error (MAE) was calculated to assess the evaluation results and the expression of MAE is shown as following:

Table 6
Actual data and evaluation results of earthquake-hit population in the Jiuzhaigou earthquake

Area	Actual data	Evaluation result (optimizing samples)	Evaluation result (random samples)
Hongyuan County	3209	18937	36898
Jiuzhaigou County	139411	92806	191298
Pingwu County	38451	32962	49858
Songpan Coungty	24922	22004	28508
Zoige County	10604	31652	40642
Total	216597	198362	347207
Mean absolute error	–	18357	26121

$$MAE = \frac{1}{n} \left(\sum_{i=1}^n |t_i - y_i| \right)$$

3

The MAE of evaluation result of earthquake-hit population were 18357 and 26121 for optimizing samples and random samples respectively. The comparison of actual data and evaluation results in each county was showed in Figure 11. The histogram represented the value of actual data and evaluation results, and the curve represented the error rate of evaluation results. The expression of error rate is shown as follow:

$$\text{Error rate} = \frac{|t - y|}{t}$$

4

The error rate was defined as the ratio of the difference between actual data and result and actual data. It can be seen that for Jiuzhaigou County, Pingwu County and Songpan County, where the earthquake-hit population was relatively large, the error rate of evaluation results based on the BP neural network was less than 100%. For Hongyuan County and Zoige County, the error rate was relatively great. The maximum value of error rate based on the random samples reached 1049.84% and that based on the optimizing samples reached 490.14%. It can be noted that in each area the error rate of optimizing samples was smaller than that of random samples. The earthquake-hit population evaluation based on the optimizing samples had more accurate prediction for new data. It revealed that optimizing samples can effectively offer a more accurate evaluation of earthquake-hit population.

5. Conclusions

In the present study, a spatial earthquake-hit population distribution is evaluated based on correlation characteristics of factors and BP neural network. The main conclusions are as follows:

- 1) The influencing factors of earthquake-hit population are classified into environmental and seismic factors. Elevation, slope angle, population density, per capita GDP, distance to fault, distance to river and NDVI are considered as environmental factors, and PGA, PGV and distance to epicenter are considered as seismic factors. The correlation analysis between earthquake-hit population and influencing factors indicates that per capita GDP and PGA have stronger correlation relation with earthquake-hit population in the Lushan earthquake. There is a great nonlinear relationship between the earthquake-hit population and various influencing factors.
- 2) Samples have a significant impact on the generalization capability and evaluation accuracy of neural network. The samples are optimized according to the spatial distribution of per capita GDP and PGA based on the correlation characteristics. In the area with lower per capita GDP values and higher PGA values, more sample points are generated and distributed according to the correlation between per capita GDP, PGA and earthquake-hit population. By comparing to the random samples, the optimizing samples can effectively improve the convergence speed and generalization capability of the trained network. In networks with different numbers of neurons, the number of iterations based on the optimizing samples is less than that of random samples. The network trained by the optimizing samples considering the spatial characteristics has more accurate prediction ability.
- 3) A BP neural network is established using influencing factors as input indicators based on the data from Lushan earthquake. The trained network is applied for Jiuzhaigou earthquake to test the generalization capability and prediction accuracy. The results show that the neural network has good prediction accuracy on the spatial evaluation in the study area. The total evaluating earthquake-hit population of five counties affected by Jiuzhaigou earthquake based on optimizing samples is 198362 people, while the actual data is 216597 people. BP neural network has abilities to construct complex nonlinear relations to evaluate earthquake-hit population. The trained network can offer spatial evaluation of earthquake-hit population as well as other earthquake disaster information quickly after the occurrence of earthquake, providing significant reference for emergency rescue.

Declarations

Declaration of Competing Interest

None.

Acknowledgement

This work is supported by the National Key R&D Program of China (2018YFC1504504).

References

1. Aghamohammadi, H., Mesgari, M. S., Mansourian, A. & Molaei, D. Seismic human loss estimation for an earthquake disaster using neural network. *International Journal of Environmental Science and Technology*, **10**, 931–939 <https://doi.org/10.1007/s13762-013-0281-5> (2013).
2. Ara, S. Impact of Temporal Population Distribution on Earthquake Loss Estimation: A Case Study on Sylhet, Bangladesh. *International Journal of Disaster Risk Science*, **5**, 296–312 <https://doi.org/10.1007/s13753-014-0033-2> (2014).
3. Basharat, M., Ali, A., Jadoon, I. A. K. & Rohn, J. Using PCA in evaluating event-controlling attributes of landsliding in the 2005 Kashmir earthquake region, NW Himalayas, Pakistan. *Nat. Hazards*, **81**, 1999–2017 <https://doi.org/10.1007/s11069-016-2172-9> (2016).
4. Boatwright, J. *et al.* *Bulletin of the Seismological Society of America*, **93** (5), 2043–2055 <https://doi.org/10.1785/0120020201> (2003). The Dependence of PGA and PGV on Distance and Magnitude Inferred from Northern California ShakeMap Data
5. Choubin, B. *et al.* Earth fissure hazard prediction using machine learning models. *Environ. Res*, **179**, 108770 <https://doi.org/10.1016/j.envres.2019.108770> (2019).
6. Conforti, M., Pascale, S., Robustelli, G. & Sdao, F. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy)., **113**, 236–250 <https://doi.org/10.1016/j.catena.2013.08.006> (2014).
7. Erdik, M., Sesetyan, K., Demircioglu, M. B., Hancilar, U. & Zulfikar, C. Rapid earthquake loss assessment after damaging earthquakes. *Soil Dynamics and Earthquake Engineering*, **31** (2), 247–266 <https://doi.org/10.1016/j.soildyn.2010.03.009> (2011).
8. Gao, S. *et al.* Combined use of principal component analysis and artificial neural network approach to improve estimates of PM2.5 personal exposure: A case study on older adults. *Science of the Total Environment*, **726**, 138533 <https://doi.org/10.1016/j.scitotenv.2020.138533> (2020).
9. Gul, M. & Guneri, A. F. An artificial neural network-based earthquake casualty estimation model for Istanbul city. *Nat. Hazards*, **84**, 2163–2178 <https://doi.org/10.1007/s11069-016-2541-4> (2016).
10. Hou, P., Jolliet, O., Zhu, J. & Xu, M. Estimate ecotoxicity characterization factors for chemicals in life cycle assessment using machine learning models. *Environmental International*, **135**, 105393 <https://doi.org/10.1016/j.envint.2019.105393> (2020).
11. Huang, C. & Huang, Y. An information diffusion technique to assess integrated hazard risks. *Environ. Res*, **161**, 104–113 <https://doi.org/10.1016/j.envres.2017.10.037> (2018).
12. Huang, R. Q. & Li, W. L. Post-earthquake landsliding and long-term impacts in the Wenchuan earthquake area, China. *Eng. Geol*, **182**, 111–120 <https://doi.org/10.1016/j.enggeo.2014.07.008> (2014).
13. Huang, X., Song, J. & Jin, H. The casualty prediction of earthquake disaster based on Extreme Learning Machine method. *Nat. Hazards*, **102**, 873–886 <https://doi.org/10.1007/s11069-020-03937-6> (2020).
14. Huang, X., Zhou, Z. & Wang, S. The prediction model of earthquake casualty based on robust wavelet v-SVM. *Nat. Hazards*, **77**, 717–732 <https://doi.org/10.1007/s11069-015-1620-2> (2015).
15. Jena, R. *et al.* Earthquake hazard and risk assessment using machine learning approaches at Palu, Indonesia. *Science of the Total Environment*, **749**, 141582 <https://doi.org/10.1016/j.scitotenv.2020.141582> (2020).

16. Lara, A., Garcia, X., Bucci, F. & Ribas, A. What do people think about the flood risk? An experience with the residents of Talcahuano city. *Chile. Natural Hazards*, **85**, 1557–1575 <https://doi.org/10.1007/s11069-016-2644-y> (2016).
17. Li, X. *et al.* Forecasting of bioaerosol concentration by a Back Propagation neural network model. *Science of the Total Environment*, **698**, 134315 <https://doi.org/10.1016/j.scitotenv.2019.134315> (2020).
18. Lou, W. P., Chen, H. Y., Qiu, X. F., Tang, Q. Y. & Zheng, F. Assessment of economic losses from tropical cyclone disasters based on PCA-BP. *Nat. Hazards*, **60**, 819–829 <https://doi.org/10.1007/s11069-011-9881-x> (2012).
19. Luo, L., van Lombardo, L., Westen, C., Pei, X. & Huang, R. From scenario-based seismic hazard to scenario-based landslide hazard: rewinding to the past via statistical simulations. *Stochastic Environmental Research and Risk Assessment*. <https://doi.org/10.1007/s00477-020-01959-x> (2021)
20. Luo, Z., Huang, F. & Liu, H. PM2.5 concentration estimation using convolutional neural network and gradient boosting machine. *Journal of Environmental Sciences*, **98**, 85–93 <https://doi.org/10.1016/j.jes.2020.04.042> (2020).
21. Partridge, D. Network generalization differences quantified. *Neural Netw*, **9** (2), 263–271 [https://doi.org/10.1016/0893-6080\(95\)00110-7](https://doi.org/10.1016/0893-6080(95)00110-7) (1996).
22. Peng, L. *et al.* Landslide susceptibility mapping based on rough set theory and support vector machines: a case of the Three Gorges area, China., **204**, 287–301 <https://doi.org/10.1016/j.geomorph.2013.08.013> (2014).
23. Peng, Y. Regional earthquake vulnerability assessment using a combination of MCDM methods. *Annals of Operations Research*, **234**, 95–110 <https://doi.org/10.1007/s10479-012-1253-8> (2015).
24. Pourghasemi, H. R., Gayen, A., Panahi, M., Rezaie, F. & Blaschke, T. Multi-hazard probability assessment and mapping in Iran. *Science of the Total Environment*, **692**, 556–571 <https://doi.org/10.1016/j.scitotenv.2019.07.203> (2019).
25. Rossetto, T. *et al.* The Indian Ocean tsunami of December 26, 2004: observations in Sri Lanka and Thailand. *Nat. Hazards*, **42**, 105–124 <https://doi.org/10.1007/s11069-006-9064-3> (2007).
26. Saha, S. *et al.* Prediction of landslide susceptibility in Rudraprayag, India using novel ensemble of conditional probability and boosted regression tree-based on cross-validation method. *Science of the Total Environment*, <https://doi.org/10.1016/j.scitotenv.2020.142928> (2020).
27. Santos-Reyes, J. & Gouzeva, T. Mexico city's residents emotional and behavioural reactions to the 19 September 2017 earthquake. *Environ. Res*, **186**, 109482 <https://doi.org/10.1016/j.envres.2020.109482> (2020).
28. Shahaki, Kenari, M. & Celikag, M. Correlation of Ground Motion Intensity Measures and Seismic Damage Indices of Masonry-Infilled Steel Frames. *Arabian Journal for Science and Engineering*, **44**, 5131–5150 <https://doi.org/10.1007/s13369-019-03719-8> (2019).
29. Shimada, N.. Outline of the Great East Japan Earthquake. In: Urabe J., Nakashizuka T. (eds) Ecological Impacts of Tsunamis on Coastal Ecosystems. Ecological Research Monographs. Springer, Tokyo. https://doi.org/10.1007/978-4-431-56448-5_1 (2016)
30. Si, M. & Du, K. Development of a predictive emissions model using a gradient boosting machine learning method. *Environmental Technology & Innovation*, **20**, 10128 <https://doi.org/10.1016/j.eti.2020.101028> (2020).
31. Sun, L., Chen, J. & Li, T. A MODIS-based method for detecting large-scale vegetation disturbance due to natural hazards: a case study of Wenchuan earthquake stricken regions in China. *Stochastic Environmental Research and Risk Assessment*, **30**, 2243–2254 <https://doi.org/10.1007/s00477-015-1160-z> (2016).
32. Tehrany, M. S., Pradhan, B. & Jebur, M. N. Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. *Stochastic Environmental Research and Risk Assessment*, **29**, 1149–1165 <https://doi.org/10.1007/s00477-015-1021-9> (2015).
33. Wang, S. & Li, D. ArcGIS-based system analysis of building damage from the Ms7.0 Lushan earthquake. *Earthquake Research in Sichuan*, **2** (151), 1–5 (in Chinese) <https://doi.org/10.13716/j.cnki.1001-8115.2014.02.001> (2014).
34. Wang, S., Zhang, N., Wu, L. & Wang, Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method., **94**, 629–636 <https://doi.org/10.1016/j.renene.2016.03.103> (2016).
35. Wang, X. P. *et al.* Estimation of soil salt content (SSC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR), Northwest China, based on a Bootstrap-BP neural network model and optimal spectral indices. *Science of the Total*

- Environment*, **615**, 918–930 <https://doi.org/10.1016/j.scitotenv.2017.10.025> (2018).
36. Wen, L., Yuan, X. & PSO. Forecasting CO2 emissions in Chinas commercial department, through BP neural network based on random forest and. *Science of the Total Environment*, **718**, 137194. <https://doi.org/10.1016/j.scitotenv.2020.137194> (2020)
37. Wu, Q., Wu, J. & Gao, M. Correlation analysis of earthquake impacts on a nuclear power plant cluster in Fujian province. *China. Environmental Research*, **187**, 109689 <https://doi.org/10.1016/j.envres.2020.109689> (2020).
38. Xu, C. *et al.* Landslides triggered by the 20 April 2013 Lushan, China, Mw 6.6 earthquake from field investigations and preliminary analyses. *Landslides*, **12**, 365–385 <https://doi.org/10.1007/s10346-014-0546-1> (2015).
39. Yalcin, A. GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations., **72**, 1–12 <https://doi.org/10.1016/j.catena.2007.01.003> (2008).
40. Yang, J., Chen, J., Liu, H. & Zheng, J. Comparison of two large earthquakes in China: the 2008 Sichuan Wenchuan Earthquake and the 2013 Sichuan Lushan Earthquake. *Nat. Hazards*, **73**, 1127–1136 <https://doi.org/10.1007/s11069-014-1121-8> (2014).
41. Yang, Z. H. *et al.* Urgent landslide susceptibility assessment in the 2013 Lushan earthquake-impacted area, Sichuan Province, China. *Nat. Hazards*, **75**, 2467–2487 <https://doi.org/10.1007/s11069-014-1441-8> (2015).
42. Yuan, R. M. *et al.* Density Distribution of Landslides Triggered by the 2008 Wenchuan Earthquake and their Relationships to Peak Ground Acceleration. *Bulletin of the Seismological Society of America*, **103** (4), 2344–2355 <https://doi.org/10.1785/0120110233> (2013).
43. Zhang, Z., Fleurisson, J. A. & Pellet, F. The effects of slope topography on acceleration amplification and interaction between slope topography and seismic input motion. *Soil Dynamics and Earthquake Engineering*, **113**, 420–431 <https://doi.org/10.1016/j.soildyn.2018.06.019> (2018).
44. Zhao, J. *et al.* *International journal of environmental research and public health*, **15** (6), 1111 <https://doi.org/10.3390/ijerph15061111> (2018).

Figures

Figure 1

Earthquake-hit population density distribution in the 2013 Ms7.0 Lushan earthquake

Figure 2

Spatial distribution of environmental influencing factors

Figure 3

Spatial distribution of seismic influencing factors

Figure 4

Correlation between earthquake-hit population density and environmental influencing factors

Figure 5

Correlation between earthquake-hit population density and seismic influencing factors

Figure 6

BP neural network structure diagram

Figure 7

Frequency histogram of influencing factor

Figure 8

Comparison between random sample points and optimizing sample points

Figure 9

Results of iteration number and RMSE of network models

Figure 10

Earthquake-hit population density distribution in the 2017 Ms7.0 Jiuzhaigou earthquake

Figure 11

Comparison of actual data and evaluation results of earthquake-hit population