

Article



Support Vector Regression for the Relationships between Ground Motion Parameters and Macroseismic Intensity in the Sichuan–Yunnan Region

Dongwang Tao ^{1,2}, Qiang Ma ^{1,2,*}, Shuilong Li ³, Zhinan Xie ^{1,2}, Dexin Lin ^{1,2} and Shanyou Li ^{1,2}

- ¹ Institute of Engineering Mechanics, China Earthquake Administration, Harbin 150080, China
- ² Key Laboratory of Earthquake Engineering and Engineering Vibration, China Earthquake Administration, Harbin 150080, China
- ³ Earthquake Administration of Fujian Province, Fuzhou 350003, China
- * Correspondence: maqiang@iem.ac.cn

Received: 24 March 2020; Accepted: 24 April 2020; Published: 28 April 2020



Featured Application: Prediction of macroseismic intensity from ground motion parameters based on a support vector regression model is better than that based on linear regression model.

Abstract: In this paper, a nonlinear regression method called a support vector regression (SVR) is presented to establish the relationship between engineering ground motion parameters and macroseismic intensity (MSI). Sixteen ground motion parameters, including peak ground acceleration (PGA), peak ground velocity (PGV), Arias intensity, Housner intensity, acceleration spectrum intensity, velocity spectrum intensity, and others, are considered as candidates for feature selection to generate optimal SVR models. The datasets with both useable strong ground motion records and corresponding investigated MSIs in the Sichuan–Yunnan region, China, are all collected, and these 125 pairs of datasets are used for selecting features and comparing regression results. Nine ground motion parameters are selected as the most relevant features: PGA is the first fundamental one and PGV is the fifth relevant feature. Based on performance measures on the testing dataset, the best SVR model is given when the number of features is one all the way up to nine. According to predicted accuracy, SVR models with Gaussian kernel give much better MSI prediction than linear kernel SVR models and linear regression models. In particular, the Gaussian kernel SVR of PGA gives much higher MSI prediction accuracy than the linear regression model of PGV and PGA. The proposed SVR models are valid for MSI values from VI to IX, and they can be used for rapid mapping damage potential and reporting seismic intensity for this high-seismic-activity region.

Keywords: macroseismic intensity; ground motion parameters; support vector regression; Sichuan–Yunnan region; linear regression; PGA; PGV; instrumental seismic intensity

1. Introduction

Macroseismic intensity (MSI) is a local measure of the degree of earthquake damage and ground motion shaking in an earthquake, as evidenced by observed damages and human responses. MSI is crucial for seismic hazard, seismic design, and seismic loss, and has been widely used in the seismological, engineering, and loss-modeling communities [1–5]. MSI can also provide guidelines on seismic retrofitting of existing structures after strong ground motion shaking, in order to reduce the seismic risk of the structures in future earthquakes. There are many different MSI scales, and the most used ones are the modified Mercalli intensity (MMI) scale, European macroseismic scale (EMS), Japan

Meteorological Agency (JMA) intensity scale, and Chinese macroseismic intensity scale (CMSIS) [1–7]. Similar to the MMI scale, CMSIS has twelve levels ranging from I to XII, but the referred values of peak ground acceleration (PGA) and peak ground velocity (PGV) are quite different from those values adopted in MMI scale [6,7]. One main reason for this difference is that MSI has strong regional dependence and a source mechanism of seismic fault, velocity structures of crust and soil layers, and local site conditions in different regions, resulting in different ground motion prediction equations for both MSI and peak ground motion parameters. Thus, regional regression models for MSIs are generated in different countries and regions, e.g., California, eastern North America, Italy, Greece, and Japan. Another main reason is that the investigated MSIs are primarily based on the assessments of structural damage, which have large dispersion and randomness due to the subjective judgements by the investigating seismologists and structural experts. Thus, the MSI given by objective ground motions is another option, and regression models between MSI and ground motion parameters have recently been generated in different regions. These instrumental seismic intensities from regression models can provide rapid seismic intensity reporting after a destructive earthquake, as long as the ground motion records are available from seismic and strong ground motion instruments. The ShakeMap system in California and Italy, as well as the JMA's seismic intensity rapid reporting system in Japan, are all based on instrumental seismic intensities.

Among engineering ground motion parameters, PGA and PGV are widely used to establish a relationship with the MSI. Since PGA is fundamental in seismic ground motion parameter zonation and seismic design of structures, the relationship between PGA and the MSI primarily has been analyzed and established [1–3,7–14]. PGV is more indicative of earthquake damage in more flexible structures, and linear regression models are generated based on PGV [1-3,7-15]. Wald et al. [3] found that the MMI values from the relationship with PGV were better than the value based on PGA if the actual seismic intensity were no less than VII. Akansel et al. [16] stated that whether MMI was more related to PGA or PGV depended on the structural stiffness. Bilal and Askan [13] concluded that MMI had better linear correlation with PGA for brittle structure and PGV for ductile structure. Besides PGA and PGV, other ground motion parameters are also used to establish the relationships. Since the structural response is derived from acceleration response spectrum, acceleration spectrum intensity was used by Tong and Yamazaki [17] and Karim and Yamazaki [11], and they both found that acceleration spectrum intensity had a higher correlation coefficient than PGA and PGV. Velocity spectrum intensity was also used for establishing relationships [18,19]. Shabestari and Yamaziki [20] derived equivalent peak acceleration from band passed three component accelerations and obtained a new relationship between JMA intensity scale and the derived acceleration. This relationship has been implemented in the JMA seismic intensity rapid reporting system. Other ground motion parameters, such as ground motion duration [21], cumulative absolute velocity [22], and Arias intensity [23], were also introduced to generate relationships with MSI. Each ground motion parameter has its own advantage over others in certain cases. For example, PGA and acceleration spectrum intensity are better parameters for estimating lower MSIs, and PGV and velocity spectrum intensity are better parameters for estimating higher MSIs. It should be noted that the above findings are based on the linear regression method. If another, nonlinear regression method is used to establish the relationship, the findings could be different. The linear regression method handles multiple variables using ordinary least square or weighted least square, but the regression results show large scatter due to two main problems. One main problem is that the true relationship is not just linearly related to each variable, and cross-terms of multiple variables may appear. For example, McCann [24] added quadratic terms in regression equation. The true relationship is complicated and nonlinear, and linear regression is not good for handling the hidden nonlinear cross-terms. For example, Alvarez et al. [25] used neural networks to set up the complicated and nonlinear relationship. The other main problem is that MSI is a discrete integer, while the ground motion parameters are continuous real variables and are not consistent with one another. There should be some steps to handle this inconsistency. One possible choice is to round the function of the continuous real variables, so that the regression value is correct if

it is within MSI \pm 0.5. For example, if the output of regression function is 7.3, then round(7.3) = 7 (VII). In this case, the loss between regression output and this MSI value should be zero, but linear regression counts 0.3 as a loss. The nonlinear regression method is a choice to handle the above two problems.

2. Motivation and the Study Objective

Support vector regression (SVR) is a powerful nonlinear regression approach. It belongs to the support vector machine (SVM) and is also called support vector machine regression. SVR has been proven to be effective and powerful in real value function estimation [26-28], and it has also been widely used for model regression in engineering applications [29–34]. In earthquake engineering, Alvarez et al. [25] predicted MMI from PGA, PGV, moment magnitude, and epicentral distance. Hsu et al. [32] estimated on-site PGA from P wave features, such as cumulative absolute velocity, peak amplitudes of acceleration, velocity and displacement, integral of squared velocity, and predominant period. As is well-known, linear regression and neural networks employ empirical risk minimization to force the regression model to converge to sample target values as much as possible, without consideration of the structural characteristics of the regression model, so they both have generalization problems if the sample size is small. On the other hand, SVR considers structural risk minimization besides empirical risk minimization, so that the regression model matches the available training dataset reasonably well, and also generalizes well to the new testing dataset [35]. In this regard, SVR is also suitable for small sample size regression. For example, the qualitative structure activity relationships dataset, which contains only 74 samples with 27 features, and Boston housing data, which contains 506 samples with 13 features, are widely used for regression, and SVR perform well on these dataset [28]. Similarly, the number of samples with both measured ground motion parameters and investigated MSIs is limited for a specific region, so the SVR is used here for model regression. Meanwhile, SVR introduces nonlinear kernel to model the nonlinear function of multiple variables, and uses insensitive loss function to accept small deviations from objective function. Thus, SVR is quite suitable for establishing the relationship between multiple continuous ground motion parameters and discrete MSI.

3. Study Area and Datasets

The Sichuan-Yunnan region is located in the area where the Eurasian plate and the Indian plate collide with each other and squeeze strongly. It covers the Sichuan–Yunnan diamond block, the southern Yunnan block, the western Yunnan block, and the eastern part of the Bayanhar block. This region consists of main faults like the Longmen Mountain fault, Anning River fault, Zemu River fault, Xiaojiang River fault, Honghe River fault, and Xiaojinhe River fault, and is the most significant area for strong earthquakes in western China [36]. The area is about 865,000 square kilometers, and is twice as large as California and almost three times larger than Italy. The seismicity in the Sichuan–Yunnan region is at a high level. Up to now, more than 30 earthquakes larger than magnitude 7.0 have occurred, and these earthquakes have caused significant casualties and property losses. For example, the great Wenchuan earthquake killed more than 69,000 people and caused 852.3 billion Yuan direct economic losses [37]. After a destructive earthquake occurs, a reconnaissance team will be sent to conduct field investigation and produce an MSI map that reflects the scope and degree of the ground impact caused by the earthquake. These MSI maps are very valuable for earthquake emergency response and post-disaster rehabilitation. To measure the ground motions caused by earthquakes, there are now 400 permanent, strong ground motion observation stations mounted in this region: 224 stations in Sichuan Province and 176 in Yunnan Province. These stations have recorded high-quality, strong ground motions in the past few years [38].

The MSI for a station in an earthquake is determined by adding the location of the station to the MSI map and determining which isoseismic line encircles it. Nine moderate-to-large earthquakes that have both investigated MSI maps and ground motion records are analyzed in this study. The locations of these earthquakes, as well as the spatial distribution of the strong motion stations, are displayed

in Figure 1. There are 106 different strong motions stations, and some stations record more than one earthquake event. The detailed information of the nine earthquakes and the numbers of strong motion records are shown in Table 1. The surface magnitude (Ms), which is used in China to measure earthquake magnitude, of these earthquakes varies from 5.8 to 8.0, and the depths are from 5 km to 33 km. The epicentral distances of stations are from 6 km to 312 km, and the MSIs are from VI to IX. In total, 125 pairs of ground motion records and MSIs are used to analyze the relationship. The complete information of the 125 sets of data on MSIs, station names, site conditions, and epicentral distances is shown in Supplementary Materials Table S1.



Figure 1. Distribution of earthquakes and strong motion stations.

	TT /1	1	1 01	\ /	•
I ahia i	Harthalls	VO OVONTO 11	the Sichiian_	Viinnan	romon
Iavie I.	Laruqua	ικε ενεπισ π	i ule siciluali-	Tunnan	ICEIOIL.
					- 0 -

Earthquake Event	Date (dd-mon-yyyy)	Latitude (°)/Longitude (°) Magnitude Max Number /Depth (km) (Ms) Intensity of Records		Number of Records	Epicentral Distance Ranges (km)	
Ninger	03-Jun-2007	23.00/101.10/33	6.4	VIII	3	13-34
Wenchuan	12-May-2008	30.99/103.36/19	8.0	XI	79	21-312
Panzhihua	30-Aug-2008	26.30/102.06/14	6.1	VIII	4	25-65
Lushan	20-Apr-2013	30.30/103.00/13	7.0	IX	12	19–99
Ludian	03-Aug-2014	27.10/103.30/12	6.5	IX	8	8-175
Jinggu	07-Oct-2014	23.4/100.5/5	6.6	VIII	5	10-60
Kangding01	22-Nov-2014	30.26/101.69/18	6.3	VIII	3	30-35
Kangding02	25-Nov-2014	30.18/101.73/16	5.8	VIII	8	6-43
Jiuzhaigou	08-Aug-2017	33.20/103.82/20	7.0	IX	3	11–41

4. Ground Motion Parameters

Ground motion generated by an earthquake is complicated, and multiple parameters rather than a single parameter are used to quantitatively reflect the characteristics of ground motion. The amplitude, frequency content, and duration are the most significant characteristics in the engineering community [39]. Some ground motion parameters, such as PGA and PGV, provide information on amplitude, while other parameters, such as acceleration spectrum intensity and Arias intensity, reflect the above two or three characteristics. A total of 16 ground motion parameters are used to characterize the recorded ground motion in the relationship study. The single amplitude parameters include PGA and PGV. Peak ground displacement (PGD) is not included, due to its sensitiveness to long period noise, and different choices of baseline correction and filtering of acceleration may give quite different displacements. The individual frequency content parameters include central frequency, which measures the frequency where the power spectral density is most concentrated, and the ratio $v_{\text{max}}/a_{\text{max}}$, which gives the period where the ground motion is most significant. The individual duration parameters include bracketed duration and significant duration. Bracketed duration is the total time elapsed between the first and the last excursions of a given level. Absolute level 5 gal and relative level 5% PGA are both considered. Significant duration is defined as the interval time when a proportion of the total Arias intensity is accumulated, and the interval between 5% and 95% thresholds is chosen here. The ground motion parameters reflecting more than one characteristic include derivations of accelerations, such as the root mean square of acceleration, cumulative absolute velocity, Arias intensity, characteristic intensity, JMA equivalent peak acceleration, and destructive index. Spectrum-based intensities, such as the acceleration spectrum intensity, velocity spectrum intensity, and Housner intensity, are considered. For the completeness and readability of the paper, the definitions, explanations, and calculation formulas of the above ground motions are given as follows [39-41].

The root mean square of acceleration (a_{RMS}) stands for the effective average acceleration in the significant duration, given by

$$a_{\rm RMS} = \sqrt{\frac{1}{T_s} \int_{t_1}^{t_2} [a(t)]^2 dt}$$
(1)

where T_s is significant duration and t_1 and t_2 are the start and end time instants, respectively.

The cumulative absolute velocity (CAV) is proposed by U.S. Electrical Power Research Institute for indicating the onset of structural damage caused by an earthquake, and is given by

$$CAV = \int_{t_1}^{t_2} |a(t)| dt$$
⁽²⁾

The Arias intensity (AI) is proposed for indicating the damage potential to nuclear power plants, and is given by

$$AI = \frac{\pi}{2g} \int_{t_1}^{t_2} [a(t)]^2 dt$$
(3)

The characteristic intensity (I_c) is proposed to indicate structural damage caused by maximal deformation and dissipative hysteretic energy, and is given by

$$I_c = a_{\rm RMS}^{1.5} T_s^{0.5} \tag{4}$$

The JMA equivalent peak acceleration is used to calculate JMA seismic intensity. It is the value of vector composition of three component band-pass accelerations, each of which is filtered by a compound filter composed of a amplitude filter, a high-cut filter, and a low-cut filter, such that the total duration when the vector composite acceleration is larger than this value is longer than 0.3 s, as shown in Equation (5). The schematic diagram of JMA equivalent peak acceleration, compound filter, and

total duration with respect to peak values are shown in [11]. Since the JMA seismic intensity scale of 0–VII is quite different than the MSI I-XII in China, the JMA equivalent peak acceleration ($A_{0.3}$) rather than JMA seismic intensity is used to characterize ground motion:

$$A_{0.3} = a_0 |\tau(a_0) \ge 0.3 \tag{5}$$

where a_0 is the vector composite acceleration, and $\tau(a_0)$ is the duration of composite acceleration larger than a_0 .

The destructive index (DI) has been proposed by Nakamura [42] to estimate the damage potential of ground motion by calculating the logarithm of the product of vertical acceleration and velocity, and is given by

$$DI = \max(lg(|a(t)v(t)|))$$
(6)

The acceleration spectrum intensity (ASI) is proposed to analyze ground motion effect on short period structures like concrete dams, and is given by

$$ASI = \int_{0.1}^{0.5} S_a(\xi = 0.05, T) dT$$
(7)

where $S_a(\xi = 0.05, T)$ is the acceleration response spectrum with damping ratio $\xi = 0.05$.

The velocity spectrum intensity (VSI) is proposed to indicate ground motion damage potential on most structures whose fundamental periods are between 0.1 and 2.5 s:

$$VSI = \int_{0.1}^{2.5} PSV(\xi = 0.05, T) dT$$
(8)

where PSV($\xi = 0.05, T$) is the pseudo-velocity response spectrum with damping ratio $\xi = 0.05$.

The Housner intensity (HI) is quite similar to velocity spectrum intensity, except that the damping ratio is selected as 0.2, since the damping ratio will become larger when the structure is damaged by an earthquake:

$$HI = \frac{1}{2.4} \int_{0.1}^{2.5} PSV(\xi = 0.2, T) dT$$
(9)

Three component accelerations are used to calculate JMA equivalent peak acceleration, while only up–down (UD) component acceleration is used for the destructive index. For the remaining ground motion parameters, the geometric means of the two horizontal component accelerations are used. The natural frequency of middle- and high-rise buildings is mainly within 0.1–2.0 Hz, and in low-rise buildings is within 5.0–10.0 Hz. The corrected acceleration is filtered by a second-order Butterworth bandpass filter with a passing band of 0.1–10.0 Hz. A complete list of the 16 calculated ground motion parameters from the set of 125 ground motion records is shown in Supplementary Materials Table S1. The scatter plots of MSI versus ground motion parameter are shown in Figure 2, and the corresponding absolute values of Pearson correlation coefficients are shown in Figure 3. It can be seen that ASI, $A_{0.3}$, Ic, PGA, DI, HI, PGV, VSI, AI, a_{RMS} , and CAV have higher linear correlations with MSI, and duration parameters $Td_{b,5}$, Td_s , $Td_{b,5\%}$ and frequency parameters central frequency (CF) and v_{max}/a_{max} almost have no linear correlation.



Figure 2. Scatter of macroseismic intensity (MSI) versus ground motion parameters.



Figure 3. Absolute correlation coefficient for ground motion parameters.

5. Support Vector Regression

For a given dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, where $x_i = (x_{i1}, x_{i2}, \dots, x_{id})^T \in \mathbb{R}^d$ is the *i*th sample with feature dimensionality of *d*, and x_{ij} is the value for the *j*th feature, $y_i \in \mathbb{R}$ is the corresponding target value of the sample and *m* is the number of samples. As shown in Figure 4, the linear kernel SVR is to find a function $f(x) = \omega^T x + b$, where $\omega = (\omega_1, \omega_2, \dots, \omega_d)^T \in \mathbb{R}^d$ is a normal vector of the hyperplane and $b \in \mathbb{R}$ is the offset between the hyperplane and the coordinate origin, such that the function is as flat as possible and has at most ϵ deviation from each sample target value y_i . The optimization objective is given by [26]:

$$\min_{\boldsymbol{\omega}, \boldsymbol{b}} \frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\omega} + C \sum_{i=1}^m |\xi|_{\epsilon} (f(\boldsymbol{x}_i) - y_i)$$
(10)

where *C* is a regularization constant and $|\xi|_{\epsilon}$ is an ϵ -insensitive loss function, given by

$$|\xi|_{\epsilon} = \begin{cases} 0, \text{ if } |\xi| \le \epsilon \\ |\xi| - \epsilon, \text{ otherwise} \end{cases}$$
(11)



Figure 4. Sketch diagram for the support vector regression (SVR) method.

The first term of Equation (10) is to describe the flatness of the function, and it is also called "structural risk". The second term of the equation, without *C*, is to describe the fitness between the function and actual sample target values, and is also called "empirical risk". The constant *C* is a compromise between the two terms. In traditional linear regression, function loss is counted as long as the function value is not equal to the sample target value. This rule is too strict, and will result in overfitting. Since the determination of the sample target value could be disturbed by some subjective or objective factors, the sample target value contains a certain level of noise. To overcome this disadvantage, SVR counts function loss only when the difference between the function value and sample target value is larger than a given threshold ϵ . The condition *C* = 0 means only considering flatness, and SVR goes back to linear regression, while *C* approaching infinity means that every sample target value is within the ϵ deviation of the corresponding function value. To describe the real deviation, two slack variables are introduced, and the primal objective function can be deduced from Equation (10) as

$$\min_{\boldsymbol{\omega}, b, \xi, \xi^*} \frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\omega} + C \sum_{i=1}^m \xi_i + C \sum_{i=1}^m \xi_i^*$$
subject to
$$\begin{cases} \boldsymbol{\omega}^T \boldsymbol{x}_i + b - y_i \leq \epsilon + \xi_i \\ y_i - \boldsymbol{\omega}^T \boldsymbol{x}_i - b \leq \epsilon + \xi_i^* \\ \xi_i, \ \xi_i^* \geq 0, i = 1, 2, \dots, m
\end{cases}$$
(12)

To efficiently solve the above optimization problem with inequality constraints, the dual problem is obtained by using the Lagrange function method, given by

$$\max_{\alpha,\alpha^*} -\frac{1}{2} (\alpha - \alpha^*)^T x_i^T x_j (\alpha - \alpha^*) - \epsilon \sum_{i=1}^m (\alpha_i + \alpha_i^*) + \sum_{i=1}^m y_i (\alpha_i - \alpha_i^*)$$
subject to
$$\begin{cases} \sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0\\ 0 \le \alpha_i, \alpha_i^* \le C \end{cases}$$
(13)

Solve the above quadratic programming problem to get α , α^* , and the solution of the SVR is given by

$$f(\mathbf{x}) = \sum_{i=1}^{m} \left(-\alpha_i + \alpha_i^* \right) \mathbf{x}_i^T \mathbf{x}_j + b$$
(14)

where *b* is the mean of all possible *b* values for the support vectors when $-\alpha_i + \alpha_i^* \neq 0$, using Karush–Kuhn–Tucker (KKT) conditions [26–28].

$$b = \frac{1}{|S|} \sum_{s \in S} \left(y_s + \epsilon - \sum_{i=1}^m \left(-\alpha_i + \alpha_i^* \right) \mathbf{x}_i^T \mathbf{x}_s \right)$$
(15)

Here, *S* is the subscript set such that α_j is greater than 0, given by $S = \{j | \alpha_j\} 0, j = 1, 2, ..., m\}$.

For real-world problems, it is impossible to find such a hyperplane that satisfies both good flatness and fitness simultaneously. One possible approach is to find a hypersurface $f(x) = \omega^T \phi(x) + b$, which preprocesses sample x into a feature space by a mapping $\phi(x)$. With the help of the kernel function method, the solution of nonlinear SVR is given by

$$f(\mathbf{x}) = \sum_{i=1}^{m} \left(-\alpha_i + \alpha_i^*\right) K(\mathbf{x}, \mathbf{x}_i) + b$$
(16)

where $K(\mathbf{x}, \mathbf{x}_i)$ is the kernel function, and *b* is similar to Equation (15), although $\mathbf{x}_i^T \mathbf{x}_s$ is substituted by $K(\mathbf{x}_i, \mathbf{x}_s)$. Linear kernel $K(\mathbf{x}, \mathbf{x}_i) = \mathbf{x}^T \mathbf{x}_i$ and Gaussian kernel $(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma ||\mathbf{x} - \mathbf{x}_i||^2)$, where γ is a parameter representing the width of Gaussian kernel, are widely used as kernel functions for most SVRs. To choose the best SVR model, parameter *C* for linear kernel and parameters (C, γ) for a Gaussian kernel need to be selected. A simple grid search, with grids from 2^{-9} to 2^9 for *C* and 2^{-8} to 2^2 for γ , both in interval 2^1 , is used to select model parameters. To reduce overfitting, there are several strategies for model selection, such as hold-out, bootstrap, and *n*-fold cross-validation [27]. For a dataset with small size samples, the simple and powerful strategy is *n*-fold cross validation, which divides the dataset into *n* mutually exclusive and complementary subsets, and each time uses n - 1 subsets as training sets and the remaining subset as the testing set. The best parameters are selected by choosing the model that gives the minimum average mean squared error (MSE) for the all *n* subsets. Besides the MSE and correlation coefficient, accuracy percentage is proposed to evaluate the performance of regression model. Accuracy percentage (P_a) is defined as the number of correctly predicted data, when the rounded value of predicted MSI equals to actual MSI, divided by the number of testing data:

$$P_a = \frac{|\{i|round(f(\mathbf{x}_i)) = y_i, i = 1, 2, \dots, l\}|}{l} \times 100\%$$
(17)

where *round* is a function that rounds the element to the nearest integer, and *l* is the total number of testing data. Since the accuracy percentage is based on the condition that the prediction value is within MSI \pm 0.5, the deviation ϵ is set to 0.5 for the SVR.

The procedure of SVR for establishing the relationship between MSI and ground motion parameters are summarized as follows: (1) choosing some ground motion parameters as features; (2) making logarithmic transformation on ground motion parameters, except for the destructive index; (3) scaling the chosen features linearly to the range of [-1, +1]; (4) selecting optimal regularization constant *C* and kernel parameter γ for the regression model, using 10-fold cross-validation on the training dataset; and (4) assessing the performance of regression on the testing dataset. Support vector machine toolbox LIBSVM, developed by Chang and Lin [28], was used to perform the training and testing.

6. Results and Discussion

The observations in earlier earthquakes, such as the Ninger, Wenchuan, Panzhihua, and Lushan earthquakes, are used as a training set, and the trained model is tested on the observations in later earthquakes, such as the Ludian, Jinggu, Kangding01, Kanagding02, and Jiuzhaigou earthquakes. The training set contains 98 observations (78.4%), and the testing set has 27 observations (21.6%). The numbers of MSI VI, VII, VIII, and IX for the training and testing sets are shown in Figure 5. After optimal model parameters were obtained in cross-validation, the final regression model was

trained for the whole observations of occurred earthquakes, and will be used to predict the MSIs of future earthquakes.



Figure 5. Histogram of the macroseismic intensity dataset.

6.1. Feature Selection

As mentioned in Section 4, some ground motion parameters have no linear correlation with MSI, it is important to focus on the most relevant features and eliminate the irrelevant ones. The inclusion of irrelevant features in the SVR gives bad prediction results, due to the overfitting problem in the irrelevant information. Each ground motion parameter should be checked for relevancy, and each time one ground motion parameter should be selected as the sole feature. The performance of Gaussian kernel SVR is shown in Figure 6, where the optimal model parameters are C = 256 and $\gamma = 1/16$, using 10-fold cross-validation on the 98-observation training dataset. It can be seen that parameters ASI, $A_{0.3}$, Ic, PGA, DI, HI, PGV, VSI, and AI all have MSEs smaller than 0.5 and accuracy percentages more than 50%. PGA gives the highest accuracy percentage, followed by $A_{0.3}$ and HI, and a_{RMS} , CAV, $Td_{\text{b},5}$, Td_{s} , $Td_{\text{b},5\%}$, CF, and $v_{\text{max}}/a_{\text{max}}$ have MSEs greater than 0.5 and accuracy percentages less than 50%.



Figure 6. Performance of SVR with a sole feature for the 16 ground motion parameters.

In this regard, the first nine ground motion parameters are considered as relevant features, and the latter seven are irrelevant features. Here, only the first nine ground motion parameters are used

for further regression study. There are $\sum_{i=1}^{9} C_{9}^{i} = 511$ possible combinations of those nine features, and the best performances for SVRs having up to nine features are shown in Figure 7. It can be seen that seven features, including PGA, $A_{0.3}$, ASI, HI, PGV, VSI, and *I*c, give the highest accuracy percentages, followed by six features, then by one feature (PGA). It is noted that one feature, PGA, gives almost the same level accuracy as a combination of seven features, meaning PGA is the most fundamental feature for all cases. One reason for this is that the other six ground motion parameters have relative high cross-correlation coefficients with PGA, and are partially linearly dependent on one another. When the number of features becomes larger than seven, the accuracy percentage drops below 50%, and this means more features do not necessarily give better prediction. With the development of strong ground motion observation network, many more stations will be constructed. In the future, a large number of stations will be triggered in an earthquake event. Calculating PGA, $A_{0.3}$, ASI, HI, PGV, VSI, and *I*c from ground motion requires much more time than just calculating PGA, and the time for calculating them for many stations will be even more. Since every millisecond is important in rapid seismic intensity reporting, the SVR of PGA will be more effective than the SVR of this seven ground motion parameters.



Figure 7. Performance of best SVRs with 1–9 features.

6.2. Gaussioan Kernel Versus Linear Kernel and Linear Regression Method

To demonstrate the advantage of Gaussian kernel SVR, the prediction performances of SVR with linear kernel and linear regression are also calculated. For brevity, only the results of best models with one, two, and seven features are shown here. The linear regression using least square on the training dataset for the three cases are as follows:

$$MSI = 1.330 \log(PGA) + 3.863$$
(18)

$$MSI = 0.416 \log(PGA) + 1.024 \log(A_{0.3}) + 3.980$$
⁽¹⁹⁾

$$MSI = 0.196 \log(PGA) + 0.945 \log(A_{0.3}) + 1.007 \log(ASI) - 2.638 \log(HI) + 1.471 \log(PGV) + 0.898 \log(VSI) - 0.260 \log(Ic) + 3.019$$
(20)

The predicted MSIs versus actual ones on the testing dataset are shown in Figure 8a–c for one, two, and seven features, respectively. To show the scatter for the third method more clearly, the result of the linear regression is off to the left side of a Gaussian kernel, with a linear kernel SVR to the right side. It can be seen from Figure 8a that most of predicted values are within a ± 0.5 range of the actual values for the Gaussian kernel. The predicted points of MSI VI are well concentrated in a smaller range, and even MSI IX is well predicted. On the other hand, more points are out of the ± 0.5 range for the linear

SVR and linear regression. This means that Gaussian kernel SVR has better prediction performance than the other two methods. Figure 8b for two features and Figure 8c for seven features show similar results. Since Gaussian kernel SVR gives the lowest MSE and highest accuracy percentage, it is the best of the three regression models. Comparing Figure 8a with Figure 8c, it can also be seen that the performance of predicted MSIs using PGA is almost the same as the one using seven features.



Figure 8. Performance comparison of regression models with Gaussian kernel and linear kernel SVRs and linear regression. (a) One feature; (b) two features; (c) seven features.

Present empirical relationships for the MMI or MSI are mainly based on PGA and PGV [1–5,7–10], and the Gaussian SVR versus linear regression of PGA and PGV are also studied. As the dataset is not the same as those used in previous studies [1–5,7], the linear regression equation should be obtained on this training set again. The model of the PGA is given by Equation (18), and the model of the PGV is given by

$$MSI = 1.442 \log(PGV) + 5.299$$
(21)

The performance results are shown in Table 2. The linear regression model of PGV is better than that of PGA, and the accuracy percentage increases from 44.3% to 66.7%. The MSE decreases from 0.374 in the linear regression of PGV to 0.214 in the Gaussian kernel SVR of PGA, and the predicted accuracy increases from 66.7% to 74.1%. It is noted that the Gaussian kernel SVR of PGV is not better than that of PGA, which can also be seen from Figure 6. The well-accepted conception that PGV is better than PGA for estimating MSI assumes linear regression. The Gaussian SVR of PGA and PGV has almost the same MSE and correlation coefficient as that of PGA, but the accuracy percentage decrease from 74.1% to 68.6%. From the comparison, it was found that the Gaussian kernel SVR of PGA gave the best regression model for predicting MSI, and was much better than the linear regression of PGV or PGA.

Performance	Linear Regression of PGA	Linear Regression of PGV	SVR of PGA	SVR of PGV	SVR of PGA and PGV
MSE	0.452	0.374	0.214	0.284	0.227
Correlation coefficient	0.599	0.715	0.819	0.759	0.814
Accuracy percentage	44.3%	66.7%	74.1%	67.1%	68.6%

Table 2. Performance measures for Gaussian kernel SVR and the linear regression of peak ground acceleration (PGA) and peak ground velocity (PGV).

6.3. Gaussian Kernel SVR of PGA Versus Models from Previous Studies

The final Gaussian kernel SVR of PGA is obtained by training the whole available dataset, and it can be used for predicting MSIs in future earthquakes. The final SVR model was compared with regression models from previous studies to check regression performance. Models from three previous studies [3,5,7] are compared here. These three models have regression equations based on both PGA and PGV. The results are shown in Figure 9a,b, and the performance measures are shown in Table 3. It is clear from Figure 9 that the performance of the SVR model is much better than that of the other three models, especially at MSI VI and VII. The predicted points of MSI VI and VII are well concentrated in a much smaller range in the SVR model, while the points in the other three models have much larger scatter. It is interesting that all models have relatively good behavior at MSI IX. From Table 3, though the correlation coefficients of the four models are at the same level, the accuracy percentage of the SVR model is much higher than the other three. The reason for this is because these three models have too much prediction dispersion at MSI VI and VII. It should be noted that one study [3] was based on California data, and another [5] on global data. Regional variation and differences in datasets result in bad performance for the Sichuan–Yunnan earthquakes. As the regression equation was obtained for the same dataset as that in SVR model, the MSEs of the linear regression in Table 2 are much smaller than those in Table 3. Thus, for a specific region, one should be very careful using the regression model of another region. The accuracy percentage of this study is also better than the third other model [7], and there are two reasons. One is that this study contained datasets from other areas of western China besides the Sichuan-Yunnan region, and did not contain Jiuzhaigou earthquake records. The other is that the filtering process was different from this paper, which leads to the condition that the PGA and PGV are not exactly the same for the two datasets. It is suggested that to have comparable results of different regression methods, not only should the earthquake records be the same, but also the ground motion parameters after the filtering process, as much as possible.



Figure 9. Performance comparison of the Gaussian kernel SVR with previous models. (**a**) Reference model based on PGA; (**b**) reference model based on PGV.

	Reference [3] Model		Reference [5] Model		Reference [7] Model		SVR
Performance	PGA	PGV	PGA	PGV	PGA	PGV	PGA
MSE	1.479	2.024	0.986	0.590	1.162	1.096	0.300
Correlation coefficient	0.704	0.705	0.708	0.691	0.656	0.676	0.768
Accuracy percentage	32.0%	13.6%	42.4%	47.2%	47.2%	35.2%	71.2%

Table 3. Performance measures for nonlinear SVR and linear regressions based on PGA and PGV.

6.4. Disscussion of Earthquake Magnitude and Epicentral Distance

Since the MSI at a location is related to the earthquake magnitude and epicentral distance, the SVR model with and without these two parameters are also discussed. As shown in Figure 10, the performance of the SVR of PGA is almost the same as those SVRs considering earthquake magnitude and epicentral distance. This means that it is enough to use ground motion parameters for predicting MSI, and it is not necessary to include magnitude and distance terms in the SVR model.



Figure 10. Performance comparison of the SVR model with and without earthquake magnitude and epicentral distance.

7. Conclusions

In this study, SVR was used to model the relationship between discrete MSI and continuous ground motion parameters. MSI is treated as sample target, and the 16 ground motion parameters are considered as feature candidates. In the Sichuan–Yunnan region, 125 sets of ground motion records with corresponding investigated MSIs were used as a complete dataset for analysis. Based on the limited dataset, the main conclusions are as follows:

(1) During the single-feature scanning test, PGA, JMA equivalent acceleration, acceleration spectrum intensity, Housner intensity, PGV, velocity spectrum intensity, Arias intensity, characteristic intensity, and damage index are the most relevant features. Unlike the linear regression method, PGA is better than PGV for predicting MSI in an SVR model.

(2) The best model parameters for Gaussian kernel SVRs with one all the way up to nine features are provided. The SVR of PGA gives almost the same performance as that of SVR with nine features. According to the performance measures of MSE, the correlation coefficient, and accuracy percentage, the Gaussian kernel SVRs perform much better than the liner kernels and linear regressions.

(3) Gaussian kernel SVRs perform much better than previous models [3,5,7], especially with regard to the accuracy percentage. The comparison results also suggest that regression should better be done with a regional dataset.

(4) Gaussian kernel SVRs with or without earthquake magnitude and epicentral distance give similar prediction performance.

Since MSI and ground motion parameters have strong regional dependence, and the number of datasets for establishing the relationship in the studied area is limited, the conclusions may not be true

anymore when another dataset is used for regression. However, a Gaussian kernel SVR of PGA is a good initial start for the regression.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-3417/10/9/3086/s1. Supplement Table S1.

Author Contributions: Conceptualization, S.L. (Shanyou Li) and Q.M.; methodology, D.T. and D.L.; software, D.T. and S.L. (Shuilong Li); writing—original draft preparation, D.T. and Z.X.; writing—review and editing, D.T.; project administration, D.T.; funding acquisition, Q.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scientific Research Fund of Institute of Engineering Mechanics, China Earthquake Administration (Grant Nos. 2014B08, 2016A03), National Key Research and Development Program of China (Grant No. 2017YFC1500802), and Shandong Co-Innovation Center for Disaster Prevention and Mitigation of Civil Structures (Grant No. XTZ201901).

Acknowledgments: We thank the China Strong Motion Networks Center, Institute of Engineering Mechanics, China Earthquake Administration for providing the ground motion records and macroseismic intensity maps. We also thank Paul Wessel and Walter H.F. Smith for providing the Generic Mapping Tools software.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Gutenberg, B.; Richter, C.F. Earthquake magnitude, intensity, energy, and acceleration. *Bull. Seismol. Soc. Am.* **1956**, *46*, 105–145.
- 2. Trifunac, M.D.; Brady, A.G. On the correlation of seismic intensity scales with the peaks of recorded strong ground motion. *Bull. Seismol. Soc. Am.* **1975**, *65*, 139–162.
- 3. Wald, D.J.; Quitoriano, V.; Heaton, T.H.; Kanamori, H. Relationships between peak ground acceleration, peak ground velocity, and modified Mercalli intensity in California. *Earthq. Spectra* **1999**, *15*, 557–564. [CrossRef]
- 4. Musson, R.M.W.; Grünthal, G.; Stucchi, M. The comparison of macroseismic intensity scales. *J. Seismol.* 2009, 14, 413–428. [CrossRef]
- Caprio, M.; Tarigan, B.; Worden, C.B.; Wiemer, S.; Wald, D.J. Ground motion to intensity conversion equations (GMICEs): A global relationship and evaluation of regional dependency. *Bull. Seismol. Soc. Am.* 2015, 105, 1476–1490. [CrossRef]
- 6. Sun, B.T.; Yan, J.Q.; Li, S.Y. The development of macroseismic intensity and the evolution of its use. *Earthq. Eng. Vib.* **2019**, *39*, 1–8. (In Chinese) [CrossRef]
- 7. Du, K.; Ding, B.; Luo, H.; Sun, J. Relationship between peak ground acceleration, peak ground velocity, and macroseismic intensity in western China. *Bull. Seismol. Soc. Am.* **2018**, *109*, 284–297. [CrossRef]
- 8. Kaka, S.I.; Atkinson, G.M. Relationships between instrumental ground-motion parameters and modified Mercalli intensity in eastern North America. *Bull. Seismol. Soc. Am.* **2004**, *94*, 1728–1736. [CrossRef]
- 9. Tselentis, G.A.; Danciu, L. Empirical relationships between modified Mercalli intensity and engineering ground-motion parameters in Greece. *Bull. Seismol. Soc. Am.* **2008**, *98*, 1863–1875. [CrossRef]
- 10. Faenza, L.; Michelini, A. Regression analysis of MCS intensity and ground motion parameters in Italy and its application in ShakeMap. *Geophys. J. Int.* **2010**, *180*, 1138–1152. [CrossRef]
- 11. Karim, K.R.; Yamazaki, F. Correlation of JMA instrumental seismic intensity with strong motion parameters. *Earthq. Eng. Struct. Dyn.* **2002**, *31*, 1191–1212. [CrossRef]
- 12. Worden, C.B.; Gerstenberger, M.C.; Rhoades, D.A.; Wald, D.J. Probabilistic relationships between ground-motion parameters and modified Mercalli intensity in California. *Bull. Seismol. Soc. Am.* **2012**, 102, 204–221. [CrossRef]
- 13. Bilal, M.; Askan, A. Relationships between felt intensity and recorded ground-motion parameters for Turkey. *Bull. Seismol. Soc. Am.* **2014**, *104*, 484–496. [CrossRef]
- 14. Panza, G.F.; Cazzaro, R.; Vaccari, F. Correlation between macroseismic intensities and seismic ground motion parameters. *Ann. Geophys.* **1997**, *40*, 1371–1382.
- 15. Wu, Y.M.; Teng, T.L.; Shin, T.C.; Hsiao, N.C. Relationship between peak ground acceleration, peak ground velocity, and intensity in Taiwan. *Bull. Seismol. Soc. Am.* **2003**, *93*, 386–396. [CrossRef]

- Akansel, V.; Ameri, G.; Askan, A.; Caner, A.; Erdil, B.; Kale, Ö.; Okuyucu, D. The 23 October 2011 Mw7.0 Van (Eastern Turkey) earthquake: Interpretations of recorded strong ground motions and post-earthquake conditions of nearby structures. *Earthq. Spectra* 2014, *30*, 657–682. [CrossRef]
- Tong, H.; Yamazaki, F. A relationship between seismic ground motion severity and house damage ratio. In Proceedings of the Fourth U.S. Conference on Lifeline Earthquake Engineering, San Francisco, CA, USA, 10–12 August 1995.
- Housner, G.W.; Martel, R.; Alford, L. Spectrum analysis of strong-motion earthquakes. *Bull. Seismol. Soc. Am.* 1953, 43, 97–119.
- Chiauzzi, L.; Masi, A.; Mucciarelli, M.; Vona, M.; Pacor, F.; Cultrera, G.; Emolo, A. Building damage scenarios based on exploitation of Housner intensity derived from finite faults ground motion simulations. *Bull. Earthq. Eng.* 2012, *10*, 517–545. [CrossRef]
- 20. Shabestari, K.T.; Yamazaki, F. A Proposal of instrumental seismic intensity scale compatible with MMI evaluated from three-component acceleration records. *Earthq. Spectra* **2019**, *17*, 711–723. [CrossRef]
- 21. Trifunac, M.D.; Westermo, B. A note on the correlation of frequency-dependent duration of strong earthquake ground motion with the Modified Mercalli Intensity and the geologic conditions at the recording stations. *Bull. Seismol. Soc. Am.* **1997**, *67*, 917–927.
- 22. Cabanas, L.; Benito, B.; Herraiz, M. An approach to the measurement of the potential structural damage of earthquake ground motions. *Earthq. Eng. Struct. Dyn.* **1997**, *26*, 79–92. [CrossRef]
- 23. Margottini, C.; Molin, D.; Serva, L. Intensity versus ground motion: A new approach using Italian data. *Eng. Geol.* **1992**, *33*, 45–58. [CrossRef]
- 24. McCann, M.W.; Sauter, F.; Shah, H.C. A technical note on PGA-intensity relations with applications to damage estimation. *Bull. Seismol. Soc. Am.* **1980**, *70*, 631–637.
- Alvarez, D.A.; Hurtado, J.E.; Bedoya-Ruíz, D.A. Prediction of modified Mercalli intensity from PGA, PGV, moment magnitude, and epicentral distance using several nonlinear statistical algorithms. *J. Seismol.* 2012, 16, 489–511. [CrossRef]
- 26. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* 1999, 10, 988–999. [CrossRef]
- 27. Smola, A.J.; Lkopf, B.S. A tutorial on support vector regression. Stat. Comput. 2004, 14, 199–222. [CrossRef]
- Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. ACM. Trans. Intell. Syst. Technol. 2011, 2, 27. [CrossRef]
- 29. Clarke, S.M.; Griebsch, J.H.; Simpson, T.W. Analysis of support vector regression for approximation of complex engineering analyses. *J. Mech. Des.* **2005**, *127*, 1077–1087. [CrossRef]
- 30. Benkedjouh, T.; Medjaher, K.; Zerhouni, N.; Rechak, S. Health assessment and life prediction of cutting tools based on support vector regression. *J. Intell. Manuf.* **2013**, *26*. [CrossRef]
- 31. Peng, X.; Shi, T.; Song, A.; Chen, Y.; Gao, W. Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods. *Remote Sens.* **2014**, *6*, 2699–2717. [CrossRef]
- Hsu, T.Y.; Huang, S.K.; Chang, Y.W.; Kuo, C.H.; Lin, C.M.; Chang, T.M.; Wen, K.L.; Loh, C.H. Rapid on-site peak ground acceleration estimation based on support vector regression and P-wave features in Taiwan. *Soil Dyn. Earthq. Eng.* 2013, 49, 210–217. [CrossRef]
- 33. Santamaria-Bonfil, G.; Reyes-Ballesteros, A.; Gershenson, C. Wind speed forecasting for wind farms: A method based on support vector regression. *Renew. Energy* **2015**, *85*, 790–809. [CrossRef]
- 34. Li, L.; Zheng, W.; Wang, Y. Prediction of moment redistribution in statically indeterminate reinforced concrete structures using artificial neural network and support vector regression. *Appl. Sci.* **2019**, *9*, 28. [CrossRef]
- 35. Al-Anazi, A.F.; Gates, I.D. Support vector regression to predict porosity and permeability: Effect of sample size. *Comput. Geosci.* **2012**, *39*, 64–76. [CrossRef]
- 36. Deng, Q.; Zhang, P.; Ran, Y. Basic characteristics of active tectonics of China. *Sci. China Ser. D* 2003, *46*, 356–372. [CrossRef]
- 37. Wang, D.; Xie, L. Attenuation of peak ground accelerations from the great Wenchuan earthquake. *Earthq. Eng. Vib.* **2009**, *8*, 179–188. [CrossRef]
- 38. Wen, R. A review on the characteristics of Chinese strong ground motion recordings. *Acta Seismol. Sin.* **2016**, *38*, 550–563. (In Chinese) [CrossRef]
- 39. Kramer, S.L. *Geotechnical Earthquake Engineering*, 1st ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1996; pp. 65–84.

- 40. Bhargavi, P.; Raghukanth, S.T.G. Rating damage potential of ground motion records. *Earthq. Eng. Eng. Vib.* **2019**, *18*, 233–254. [CrossRef]
- 41. Ma, Q.; Li, S.; Li, S.; Tao, D. On the correlation of ground motion parameters with seismic intensity. *Earthq. Eng. Eng. Vib.* **2014**, *34*, 83–92. (In Chinese) [CrossRef]
- 42. Nakamura, Y. Research and development of intelligent earthquake disaster prevention system UrEDAS and HERAS. *J. Struct. Mech. Earthq. Eng. Jpn. Soc. Civ. Eng.* **1996**, *531*, 1–33. (In Japanese) [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).