Earthquakes and Structures, *Vol. 30, No. 6 (2026) 835-858*
https://doi.org/10.12989/eas.2026.30.6.835

Techno-Press

A near-fault ground-motion prediction equations for constant-strength relative input energy based on machine learning

Juanjuan Gao[1], Qinghui Lai*[1], Hui Chen[1], Daehyeon Kim[2]

[1]*Wenzhou Key Laboratory of Intelligent Lifeline Protection and Emergency Technology for Resilient City, Wenzhou University of Technology, Wenzhou 325035, China*
[2]*Department of Civil Engineering, Chosun University, Gwangju, 61452, Republic of Korea*

(Received January 5, 2026, Revised April 15, 2026, Accepted June 9, 2026)

Abstract. Input energy is the critical theoretical foundation for performance-based seismic design, as it directly reflects the energy demands imposed on structures during seismic events, thereby offering more physically meaningful metrics for structural performance assessment and design. However, relatively little attention has been devoted to the development of ground motion prediction equations (GMPEs) for near-fault elastoplastic input energy. Based on the NGA-West2 ground motion database, this study selects moment magnitude (M_w), average velocity of shear waves in the uppermost 30 m (V_{S30}), fault type, and rupture distance as key characteristic variables. A machine-learning-based support vector regression (SVR) framework is employed to predict constant-strength relative input energy, with model hyperparameters globally optimized using the particle swarm optimization (PSO) algorithm. These methodological choices aim to enhance the generalization capability and prediction accuracy of PSO-SVR machine-learning-based GMPEs. The rationality of the PSO-SVR machine-learning-based GMPEs fitting was verified through residual analysis and the influence of explanatory variables on the prediction results. The results show that the PSO-SVR machine-learning-based GMPEs for constant-strength relative input energy proposed demonstrates higher accuracy and stability in the prediction of near-fault ground motion data. The findings provide reliable references for performance-based seismic design and contribute valuable insights to the application of machine learning methods in earthquake engineering.

Keywords: constant-strength relative input energy; constant-strength; ground-motion prediction equations; machine learning; near-fault ground motions

1. Introduction

Compared to far-field ground motions, near-fault ground motion typically exhibit greater destructiveness due to the combined effects of rupture propagation and fault geometric characteristics. This type of ground motion is characterized by short duration, concentrated energy, and pronounced velocity pulses, which can readily induce strong dynamic responses in engineering structures [1-2]. However, despite its significant engineering importance, existing studies indicate that systematic research on near-fault ground motions remains relatively limited, and the corresponding models and design methodologies have not yet been fully established [3-4]. In near-fault regions, pulse-type ground motions warrant particular attention. These ground

*Corresponding author, Associate Professor, E-mail: 18845117968@163.com

http://www.techno-press.org/eas/5

ISSN: 2092-7614 (Print), 2092-7622 (Online)

motions are typically characterized by prominent velocity pulses and release their energy over a short duration. The structural impact they induce is significantly greater than that of non-pulse ground motions [5-6], potentially amplifying the displacement and energy demands of structures and thereby reducing the design safety margin [7-8]. Therefore, compared to far-field ground motions, near-fault ground motions impose more stringent seismic design requirements on buildings and engineering structures.

Under strong earthquake loading, the structural response is essentially a continuous process of energy input and dissipation. The input energy not only comprehensively accounts for the frequency content, duration, and amplitude of ground motion, but also reflects the energy dissipation characteristics within the structure, thereby enabling a more comprehensive assessment of the destructive potential of ground motion [9-10]. Compared to a single intensity measure, input energy provides a more direct link between seismic action and the structural damage process. Therefore, in structural seismic design and seismic risk assessment, it is considered one of the most rational indicators of potential damage [11-12].

In recent years, numerous scholars have devoted themselves to developing energy demand models to advance energy-based seismic design methodologies. In the early stage, Decanini and Mollaioli [13] proposed an input energy prediction equation based on 296 ground motion records and developed a simplified design expression. Subsequently, several researchers [14-15] conducted in-depth studies on related issues. Cheng et al. [16] theoretically established the relationship between the input energy spectrum and the pseudo-acceleration spectrum and proposed a corresponding prediction equation. Cheng et al. [17] developed constant-strength and constant-ductility input energy spectra and refined the existing models. The research results indicate that, except for extremely short periods, the difference between elastic and inelastic strength spectra is small and further decreases as the damping ratio increases. Research on special ground motions remains relatively limited, with only a few studies addressing pulse-type, near-fault, and mainshock-aftershock sequences. Alıcı and Sucuoğlu [18] proposed an attenuation model for near-fault ground motions and analyzed the effects of damping and intensity on energy dissipation. Zhai et al. [19] investigated the input energy spectrum characteristics of both near-fault and mainshock-aftershock ground motions and explored the influences of key parameters—such as period, magnitude, rupture distance, damping ratio, and stiffness—on the energy factor. These studies not only theoretically reveal the intrinsic connection between energy demand and ground motion characteristics, but also provide several methodological suggestions for translating energy-based theories into engineering applications [20-21].

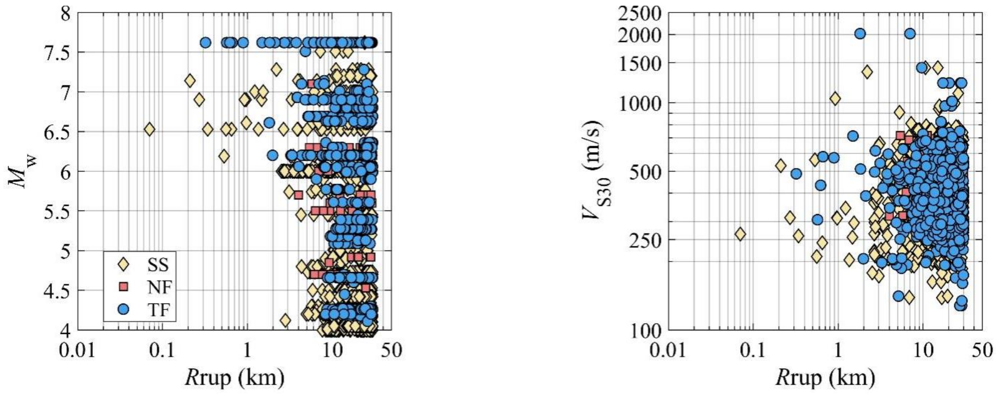
Meanwhile, the energy spectrum, as a tool capable of quantifying energy demands across different periods, has gradually gained recognition and application within the engineering community. For instance, the Japanese seismic code has officially incorporated the energy spectrum [22] and applied it to guide seismic design practices. Apart from Japan, other countries and regions are also making efforts to incorporate energy spectra into building codes, thereby promoting the gradual engineering implementation and standardization of energy-based design methods [23-25]. These explorations indicate that input energy and its related indicators not only hold significant theoretical value but also possess broad application potential in practical seismic design and code development. As a foundation for performance-based seismic design, the input energy-based ground motion model has attracted growing attention. Decanini et al. [13] investigated the attenuation relationship of input energy with focal distance and found that assessments based on peak input energy (EI) yielded significantly different results compared to those derived from peak ground acceleration and peak ground velocity. Yang et al. [26] developed a GMPE incorporating horizontal and vertical elastic input energy using a deep learning approach; Hu and Tan [27] established an elastic input energy model for ground motions in the Sagami Bay offshore region of Japan. Cheng et al. [17] developed an empirical elastoplastic input energy prediction equation. Although some progress has been made, research on input energy-based GMPEs remains limited compared to traditional parameters such as response spectra. Therefore, developing a GMPE incorporating elastoplastic input energy is essential for advancing performance-based seismic design.

NGA-West2 ground motion database is employed in this study, and key characteristic variables—including moment magnitude (M_w), average velocity of shear waves in the uppermost 30 m (V_{S30}), fault type, and rupture distance (R_{rup})—are selected to develop an elastoplastic input energy prediction model. Methodologically, machine-learning-based support vector regression (SVR) is adopted to establish the predictive framework, while the particle swarm optimization (PSO) algorithm is applied to globally optimize the PSO-SVR machine-learning-based GMPE's hyperparameters, thereby enhancing its generalization capability and prediction accuracy. Finally, the PSO-SVR machine-learning-based GMPE's rationality is validated through residual distribution analysis and examination of explanatory variables, and the influence of major explanatory variables on prediction performance is discussed, providing a reference for improving energy prediction models for near-fault ground motions and advancing performance-based seismic design.

2. Near-fault ground motion database

The near-fault ground motion records used in this study are screened from the NGA-West2 database, which was developed by the Pacific Earthquake Engineering Research Center. The NGA-West2 project represents a large-scale, interdisciplinary collaborative research program encompassing seismology, engineering, and statistics, with the primary objective of developing GMPEs for shallow crustal earthquakes [28]. Compared with earlier ground motion databases, the NGA-West2 database represents significant improvements in data volume, data quality, and structural comprehensiveness. It contains a large number of ground motion records that have undergone strict baseline correction and filtering, while systematically providing source parameters (e.g., earthquake magnitude, focal mechanism), site conditions, and several source-to-site distance metrics (e.g., epicentral distance, hypocentral distance and rupture distance). Since its release, the NGA-West2 database has become a fundamental resource in international earthquake engineering research due to its comprehensiveness and high data quality. It is widely used in the analysis of ground motion characteristics, the development of GMPEs, and the assessment of structural seismic responses.

In this study, to develop a PSO-SVR machine-learning-based GMPEs of constant-strength relative input energy suitable for near-fault regions, a total of 1,785 near-fault ground motion records from 99 earthquake events (1190 stations) with a rupture distance $R_{rup} \leq 30$ km were selected from the NGA-West2 database. Meanwhile, the moment magnitude M_W ranges of the selected earthquake events from 4.00 to 7.62, covering typical earthquake scenarios from small to large earthquakes, thereby ensuring the statistical representativeness of the near-fault ground motion database and the reliability of developed PSO-SVR machine-learning-based GMPEs. Fig. 1 shows the distributions of M_W versus R_{rup}, and the V_{S30} versus R_{rup} for the selected near-fault ground motion database, respectively. The selected 99 earthquake events were classified based on



(a) M_W-R_{rup} distribution (b) V_{S30}-R_{rup} distribution

Figure 1. Near-fault ground motion database

the fault mechanism as follows: strike-slip SS (56), normal fault NF (13), thrust fault TF (30). The selected near- fault ground motion records have a V_{S30} value ranging from approximately 100-2100 m/s, with merely 10% of entries exceeding 600 m/s. Moreover, 112 pulse-like ground motion records account for 6.3% of the near-fault ground motion database in this study. This near-fault ground motion database will be used for the subsequent development of PSO-SVR machine-learning-based GMPEs for constant-strength relative input energy in this study. The ground motion records were randomly divided into a training set and a testing set at a ratio of 7:3 [29].

3. Development of ground motion prediction equations

3.1 Calculation of constant-strength relative input energy

Under seismic loading, the equation of motion for a single-degree-of-freedom system (SDOF) is given by Eq. (1) [30].

$$\int m\ddot{v}dv + \int c\dot{v}dv + \int f_s dv = -\int m\ddot{v}_g dv, \tag{1}$$

m is the mass of SDOF, c is the viscous damping coefficient, f_s is the restoring force, $\ddot{v}$ is the absolute acceleration, $\dot{v}$ is the absolute velocity, v is the absolute displacement, and $\ddot{v}_g$ is the relative acceleration of ground motion.

The term on the right side of Eq. (2) represents the relative input energy, that is:

$$E_I = -\int m\ddot{v}_g dvm \tag{2}$$

Here, E_I represents the relative input energy.

It should be pointed out that the equivalent velocity here is the maximum input energy during the seismic action process. The reason is that when studying the input energy of ground motion, if the energy at the end of the seismic action is selected, it usually leads to an underestimation of the input energy [31-32]. To eliminate the influence of mass, the equivalent velocity V_E is usually adopted instead of the input energy [21, 30, 33], and the transformation equation is shown in Eq. (3). Therefore, the equivalent velocity is also used instead of the input energy for analysis.

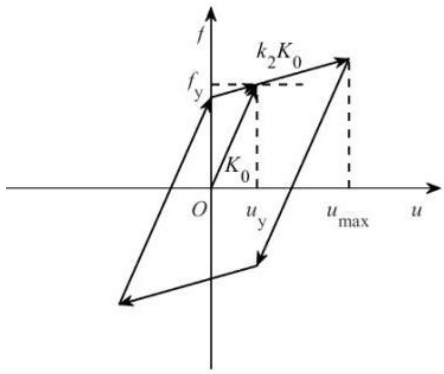


Figure 2. Bilinear hysteresis curve

$$V_E = \sqrt{2E_I/m} \tag{3}$$

3.2 Constitutive relation

For the development of PSO-SVR machine-learning-based GMPEs based on constant-strength relative input energy, the rational selection of the structural constitutive relationship is of crucial significance, as it not only determines the force-displacement behavior of the structure under seismic loading, but also directly influences its energy dissipation capacity and ductility demands [17]. In existing research, hysteretic skeleton curves are commonly used as the fundamental framework to characterize the force-displacement relationship of structures under cyclic loading. They effectively capture the stiffness evolution of structures from the elastic to the plastic stage and provide essential mechanical parameters for energy input and dissipation analyses.

This study adopts the bilinear model as the primary form of the hysteretic skeleton curve (see Fig. 2). This model concisely captures the initial stiffness, yield point, and post-yield stiffness characteristics of the structure through an idealized bilinear relationship, and effectively reproduces the fundamental hysteresis behavior under strong seismic excitations. While ensuring computational efficiency, the bilinear model also meets the accuracy requirements for developing a PSO-SVR machine-learning-based GMPE for constant-strength relative input energy. Therefore, it is selected as the main constitutive relationship representation in this study.

3.3 Ground motion prediction model based on PSO-SVR

Support Vector Regression (SVR), a supervised learning regression algorithm, exhibits strong generalization capability and is capable of efficiently handling high-dimensional data and nonlinear relationships. Its core principle involves mapping input features into a high-dimensional feature space through a kernel function, and subsequently finding a linear regression hyperplane that allows for a specified range of tolerance (Eq. (4)), thereby addressing nonlinear problems in the original feature space [34]. Assume that the training subset used to develop the PSO-SVR machine-learning-based GMPE for constant-strength relative input energy in this study is denoted as (X, y_i), i=1, 2, 3, …, n. X is a vector composed of multiple characteristic parameters, such as

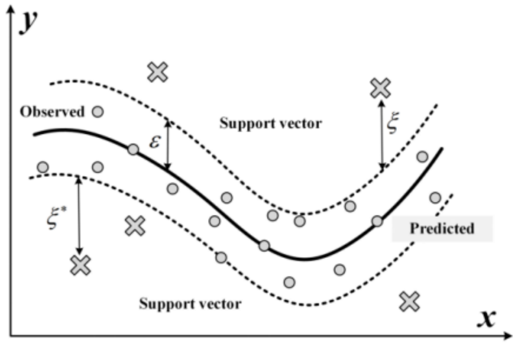


Figure 3. Fundamental concept of SVR algorithm

M_W, R_{rup}, and V_{S30}, which represent the key influencing factors of the constant-strength relative input energy, including source, path, and site conditions. y_i denotes the actual value of the constant-strength relative input energy, computed from the i-th near-fault ground motion record.

$$f(X,\omega) = \omega^T \cdot X + b \tag{4}$$

where ω represents the weight vector; b is the intercept.

In SVR, classical loss functions (e.g., absolute error or squared error) are not used. Instead, the ε-insensitive loss function $L(y, f(x, \omega), \varepsilon)$ (Eq. (5)) is introduced to minimize the prediction error of the regression model [35].

$$L(y_i, f(x,\omega), \varepsilon) = \begin{cases} |y_i - f(x,\omega)| - \varepsilon & |y_i - f(x,\omega)| > \varepsilon \\ 0 & |y_i - f(x,\omega)| \leq \varepsilon \end{cases} \tag{5}$$

where ε is the tolerance error.

In addition, to account for training sample data that fall outside the ε-insensitive zone, the slack variable ξ is introduced in SVR, enabling the transformation of the hyperplane optimization problem into a convex quadratic programming problem, as shown in Eq. (6).

$$\min_{\omega,\xi} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \text{ s.t.} \begin{cases} f(x,\omega) - y_i \leq \varepsilon + \xi_i \\ y_i - f(x,\omega) \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1,2,\ldots,n \end{cases} \tag{6}$$

where C is the penalty factor that controls the tolerance errors of the regression model. If C is too large, the regression model is prone to overfitting, which compromises its generalization ability. Conversely, if C is too small, the regression model becomes overly simplified and smoothed, leading to underfitting. The tolerance error ε determines the width of the ε-insensitive zone (see Fig. 3). A smaller ε value increases the number of support vectors, enabling the regression model to fit data details more closely but simultaneously raising the risk of overfitting. The fundamental concept of SVR algorithm is illustrated in Fig. 3. For all sample points within the ε-insensitive zone, the error is considered zero.

Whether a regression model with good generalization performance can be achieved using SVR primarily depends on the appropriate selection of hyperparameters, including the penalty factor C, the kernel parameter γ, and the tolerance error ε. In this study, the radial basis function (RBF) is selected as the kernel function, where γ controls the influence range of each training sample and

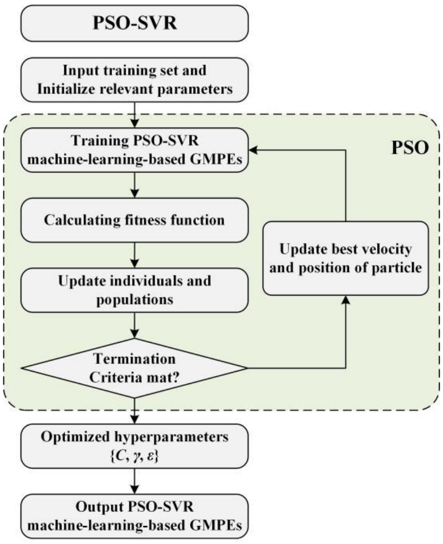


Figure 4. Flowchart of developed machine-learning-based GMPEs based on PSO-SVR

determines the complexity of the decision boundary. To ensure that the resulting regression model exhibits strong generalization capability, this study incorporates the particle swarm optimization algorithm (PSO) into the optimization of hyperparameters, and a support vector regression algorithm based on Particle swarm optimization (PSO-SVR) is proposed.

PSO was developed by Eberhart [36] and Kennedy [37] as an evolutionary algorithm. This algorithm offers high accuracy and fast convergence speed, enabling efficient execution of intelligent global optimization tasks. In the development of the PSO-SVR machine-learning-based GMPEs, each particle in the population represents a potential solution consisting of a set of hyperparameters $\{C, \gamma, \varepsilon\}$. The spatial position and velocity of each particle are updated iteratively based on the evaluation of the fitness function. Ultimately, the optimal combination of hyperparameters is identified according to a predefined maximum tolerance error or maximum number of iterations. The complete workflow for developing the PSO-SVR machine-learning-based GMPEs is shown in Fig. 4.

3.4 The prediction equations of constant-strength relative input energy

Previous studies on GMPEs mainly established the functional relationship between input energy and source, path, and site parameters through regression analysis of random effects models [17, 27]. Among them, the model research object of Hu and Tan [27] is the elastic input energy of offshore ground motion, and the research object of Cheng et al. [17] is the elastoplastic input energy of terrestrial ground motion. Since the research focus of this study is the input energy of onshore ground motions, the development of the input energy GMPE based on SVR incorporates the types and forms of input parameters used in Cheng et al. [17] (hereinafter referred to as CLM20). The input variables include M_w, the natural logarithm of rupture distance ($\ln(R_{rup})$), the natural logarithm of V_{S30} ($\ln(V_{S30})$), and a categorical variable representing fault type, as defined in Eq. (7). To facilitate the determination of hyperparameters for the subsequent regression model, all input parameters are normalized to the range [-1, 1]. Additionally, the model output is defined as

Table 1. Hyper-parameters and prediction results calculated for linear elastic systems (Cy=1.0)

T (s)	C	γ	$\varepsilon_{\text{training}}$	R^2_{training}	R^2_{testing}	MAE_{training}	MAE_{testing}	$RMSE_{\text{training}}$	$RMSE_{\text{testing}}$	τ_{training}	σ_{training}	τ_{testing}	σ_{testing}
0.1	2.755	1.414	0.408	0.790	0.755	0.511	0.576	0.654	0.741	0.329	0.565	0.416	0.613
0.2	2.720	1.315	0.097	0.817	0.774	0.468	0.525	0.631	0.678	0.287	0.563	0.359	0.576
0.3	15.148	0.690	0.070	0.831	0.823	0.460	0.501	0.622	0.650	0.322	0.532	0.374	0.532
0.4	14.463	1.326	0.467	0.861	0.841	0.465	0.510	0.595	0.659	0.292	0.519	0.409	0.516
0.5	18.945	1.279	0.436	0.878	0.844	0.454	0.515	0.589	0.659	0.307	0.503	0.375	0.543
0.6	8.081	1.255	0.474	0.878	0.859	0.464	0.520	0.596	0.675	0.308	0.510	0.435	0.514
0.7	4.648	0.854	0.354	0.877	0.864	0.479	0.521	0.617	0.677	0.322	0.526	0.434	0.518
0.8	24.959	0.501	0.134	0.886	0.874	0.460	0.503	0.612	0.655	0.323	0.519	0.392	0.524
0.9	3.424	1.229	0.331	0.898	0.870	0.457	0.506	0.599	0.656	0.318	0.507	0.404	0.516
1.0	64.106	0.047	0.187	0.892	0.875	0.474	0.511	0.617	0.670	0.340	0.515	0.442	0.504
1.2	6.718	0.832	0.189	0.901	0.899	0.454	0.461	0.609	0.604	0.345	0.501	0.350	0.493
1.4	8.639	0.654	0.042	0.907	0.893	0.440	0.469	0.601	0.624	0.328	0.503	0.416	0.466
1.6	22.650	0.432	0.399	0.906	0.901	0.454	0.487	0.598	0.635	0.334	0.496	0.381	0.508
1.8	4.316	1.111	0.184	0.915	0.891	0.434	0.484	0.581	0.651	0.299	0.498	0.419	0.498
2.0	43.149	0.521	0.142	0.907	0.902	0.447	0.470	0.611	0.615	0.327	0.516	0.397	0.469
2.5	20.349	0.868	0.477	0.909	0.897	0.467	0.495	0.603	0.654	0.300	0.523	0.415	0.506
3.0	47.426	0.496	0.525	0.906	0.889	0.471	0.523	0.614	0.669	0.320	0.525	0.385	0.547
3.5	17.661	0.687	0.370	0.911	0.882	0.459	0.517	0.597	0.685	0.294	0.519	0.417	0.544
4.0	11.599	0.651	0.288	0.904	0.896	0.468	0.496	0.614	0.653	0.310	0.530	0.396	0.518
4.5	19.806	0.879	0.388	0.909	0.886	0.459	0.516	0.599	0.673	0.287	0.526	0.406	0.538
5.0	10.327	1.083	0.327	0.901	0.904	0.472	0.484	0.622	0.621	0.309	0.540	0.359	0.507
6.0	13.553	0.841	0.207	0.904	0.885	0.464	0.513	0.613	0.668	0.296	0.537	0.422	0.518
7.0	7.617	1.219	0.352	0.907	0.873	0.470	0.511	0.609	0.661	0.303	0.528	0.357	0.557
8.0	3.890	0.751	0.143	0.894	0.884	0.470	0.517	0.625	0.667	0.305	0.545	0.401	0.533
9.0	9.726	1.317	0.250	0.899	0.887	0.459	0.498	0.604	0.652	0.286	0.532	0.401	0.513
10.0	6.757	1.436	0.333	0.898	0.890	0.457	0.504	0.594	0.654	0.272	0.528	0.388	0.526

the natural logarithm of the predicted elastoplastic input energy values, as given in Eq. (8).

$$X_i = \left[M_w, ln(R_{\text{rup}}), ln(V_{s30}), FM\right]_i \tag{7}$$

where FM is a dummy variable related to the fault type. When the fault type is strike-slip fault, normal fault or reverse fault, the FM values are 0,1 and -1, respectively.

$$Y_i = ln(IM)_i(IM = E_i(T)) \tag{8}$$

Tables 1-5 present the hyperparameter values and model performance metrics of the PSO-SVR machine-learning-based GMPEs for constant-strength relative input energy corresponding to strength coefficients Cy=1.0, 0.9, 0.7, 0.3 and 0.1 respectively. In this study, the coefficient of determination (R^2) and the standard deviation of the residuals were used to evaluate the validity of each PSO-SVR machine-learning-based GMPE and the accuracy of its predictions. Overall, for

Table 2. Hyper-parameters and prediction results calculated for linear elastic systems (Cy=0.9)

T (s)	C	γ	$\varepsilon_{\mathrm{training}}$	R^2_{training}	R^2_{testing}	MAE_{training}	MAE_{testing}	$RMSE_{\mathrm{training}}$	$RMSE_{\mathrm{testing}}$	τ_{training}	$\sigma_{\mathrm{training}}$	τ_{testing}	$\sigma_{\mathrm{testing}}$
0.1	2.114	1.775	0.382	0.797	0.760	0.500	0.568	0.641	0.731	0.319	0.556	0.410	0.605
0.2	4.247	1.651	0.401	0.824	0.790	0.473	0.511	0.614	0.669	0.273	0.550	0.373	0.556
0.3	2.843	1.382	0.274	0.838	0.821	0.469	0.495	0.613	0.645	0.298	0.536	0.368	0.529
0.4	6.890	1.150	0.424	0.863	0.828	0.461	0.527	0.598	0.667	0.304	0.515	0.416	0.522
0.5	16.332	0.898	0.425	0.873	0.845	0.463	0.518	0.596	0.671	0.309	0.510	0.414	0.528
0.6	6.619	0.938	0.405	0.874	0.866	0.475	0.497	0.620	0.624	0.323	0.529	0.385	0.492
0.7	15.764	1.116	0.375	0.888	0.860	0.461	0.521	0.600	0.662	0.320	0.508	0.371	0.546
0.8	6.456	1.248	0.236	0.890	0.877	0.453	0.486	0.608	0.630	0.325	0.513	0.370	0.510
0.9	31.816	1.051	0.295	0.905	0.869	0.437	0.511	0.569	0.682	0.291	0.489	0.427	0.525
1.0	98.180	0.023	0.001	0.890	0.878	0.482	0.500	0.622	0.663	0.350	0.514	0.424	0.510
1.2	18.316	0.635	0.008	0.899	0.903	0.443	0.464	0.613	0.597	0.337	0.511	0.351	0.482
1.4	19.352	0.533	0.171	0.901	0.909	0.451	0.461	0.610	0.598	0.352	0.497	0.378	0.464
1.6	5.640	1.003	0.045	0.910	0.900	0.426	0.491	0.586	0.641	0.310	0.497	0.381	0.517
1.8	4.898	0.983	0.199	0.907	0.908	0.447	0.468	0.603	0.608	0.328	0.506	0.362	0.489
2.0	51.865	0.388	0.319	0.911	0.888	0.453	0.493	0.601	0.646	0.333	0.500	0.398	0.509
2.5	75.499	0.179	0.385	0.903	0.886	0.484	0.498	0.632	0.665	0.342	0.532	0.423	0.509
3.0	9.759	0.603	0.161	0.902	0.896	0.469	0.497	0.629	0.643	0.324	0.539	0.405	0.500
3.5	6.413	1.292	0.265	0.914	0.876	0.446	0.526	0.590	0.695	0.289	0.515	0.422	0.552
4.0	8.208	0.847	0.282	0.909	0.888	0.460	0.496	0.607	0.650	0.301	0.527	0.393	0.518
4.5	23.537	0.642	0.328	0.908	0.882	0.463	0.522	0.604	0.678	0.311	0.518	0.420	0.533
5.0	9.376	0.897	0.361	0.906	0.884	0.469	0.508	0.614	0.659	0.286	0.544	0.391	0.531
6.0	8.158	1.475	0.311	0.905	0.886	0.464	0.522	0.609	0.665	0.291	0.535	0.389	0.531
7.0	25.230	0.548	0.355	0.898	0.885	0.481	0.509	0.627	0.658	0.331	0.532	0.392	0.526
8.0	22.130	0.799	0.372	0.899	0.886	0.473	0.497	0.616	0.649	0.317	0.528	0.351	0.546
9.0	12.021	1.118	0.379	0.898	0.885	0.464	0.503	0.610	0.645	0.301	0.530	0.376	0.524
10.0	14.450	0.936	0.382	0.894	0.885	0.468	0.520	0.608	0.658	0.301	0.528	0.340	0.563

both the training and test subsets, the R^2 values are almost exceed 0.8, which indicates an excellent model fit for the current dataset, with strong correlation between predicted and observed values. The mean absolute error (MAE) and root mean squared error (RMSE) are also selected as additional statistical performance metrics to conduct a more rigorous assessment of model performance. As listed in Tables 1-5, the MAE and RMSE values of the training subset are close to those of the test subset, demonstrating that the PSO-SVR machine-learning-based GMPEs do not exhibit overfitting. In addition, the MAE values are lower than 0.6 and the RMSE values are less than 0.75, which verifies that the proposed machine-learning-based GMPEs achieves favorable predictive accuracy. The training set yields maximum standard deviations of 0.352 (inter-event) and 0.565 (intra-event) for the residuals, whereas the test set exhibits corresponding values of 0.442 and 0.613 indicating that the developed model maintains relatively low predictive uncertainty.

Table 3. Hyper-parameters and prediction results calculated for linear elastic systems (Cy=0.7)

T (s)	C	γ	$\varepsilon_{\text{training}}$	R^2_{training}	R^2_{testing}	MAE_{training}	MAE_{testing}	$RMSE_{\text{training}}$	$RMSE_{\text{testing}}$	τ_{training}	σ_{training}	τ_{testing}	σ_{testing}
0.1	10.622	0.854	0.435	0.820	0.780	0.463	0.533	0.604	0.693	0.297	0.526	0.396	0.569
0.2	10.030	1.471	0.327	0.831	0.803	0.467	0.469	0.615	0.615	0.284	0.546	0.302	0.531
0.3	4.989	1.257	0.185	0.853	0.807	0.446	0.511	0.588	0.668	0.297	0.508	0.362	0.559
0.4	7.016	1.835	0.409	0.868	0.848	0.449	0.498	0.589	0.629	0.292	0.511	0.338	0.530
0.5	4.434	1.084	0.514	0.871	0.857	0.477	0.490	0.607	0.639	0.316	0.518	0.383	0.511
0.6	71.329	0.042	0.248	0.867	0.854	0.494	0.507	0.631	0.673	0.339	0.532	0.459	0.492
0.7	6.000	1.202	0.197	0.889	0.865	0.445	0.513	0.591	0.665	0.320	0.496	0.378	0.548
0.8	6.872	1.385	0.379	0.891	0.885	0.459	0.487	0.600	0.627	0.318	0.509	0.381	0.499
0.9	85.767	0.260	0.507	0.892	0.875	0.475	0.494	0.611	0.654	0.332	0.513	0.409	0.510
1.0	86.828	0.174	0.293	0.890	0.893	0.468	0.487	0.622	0.626	0.342	0.519	0.418	0.466
1.2	74.203	0.351	0.298	0.905	0.886	0.444	0.507	0.593	0.648	0.331	0.491	0.410	0.502
1.4	13.525	0.459	0.425	0.906	0.894	0.454	0.498	0.599	0.641	0.341	0.492	0.389	0.510
1.6	17.199	0.580	0.315	0.913	0.890	0.445	0.478	0.587	0.643	0.318	0.493	0.417	0.489
1.8	6.393	0.975	0.095	0.912	0.891	0.428	0.497	0.585	0.662	0.318	0.490	0.384	0.538
2.0	13.813	0.601	0.128	0.909	0.896	0.443	0.484	0.598	0.646	0.327	0.501	0.362	0.535
2.5	42.685	0.500	0.315	0.906	0.899	0.468	0.482	0.616	0.637	0.319	0.527	0.397	0.499
3.0	9.939	0.786	0.161	0.911	0.883	0.442	0.528	0.598	0.678	0.342	0.490	0.365	0.568
3.5	52.569	0.465	0.470	0.901	0.901	0.471	0.507	0.620	0.644	0.333	0.524	0.354	0.539
4.0	6.280	0.857	0.263	0.906	0.891	0.460	0.504	0.608	0.662	0.299	0.530	0.413	0.517
4.5	14.507	0.653	0.342	0.901	0.892	0.473	0.509	0.619	0.664	0.299	0.542	0.429	0.503
5.0	15.442	0.877	0.464	0.898	0.902	0.477	0.502	0.624	0.636	0.312	0.541	0.359	0.526
6.0	14.201	1.014	0.281	0.904	0.884	0.464	0.502	0.613	0.656	0.295	0.537	0.389	0.524
7.0	6.857	0.692	0.204	0.896	0.887	0.471	0.515	0.623	0.663	0.317	0.537	0.379	0.545
8.0	16.161	0.683	0.169	0.897	0.888	0.464	0.508	0.611	0.662	0.302	0.532	0.374	0.546
9.0	5.216	1.625	0.327	0.902	0.874	0.457	0.507	0.598	0.665	0.286	0.525	0.396	0.535
10.0	8.572	1.316	0.298	0.899	0.881	0.455	0.509	0.596	0.655	0.278	0.527	0.417	0.505

4. Model performance evaluation

4.1 Correlation and uncertainty analysis

The SVR algorithm is employed to predict and analyze ground motion parameters, and the PSO algorithm is applied to optimize the hyperparameters for improved model performance. Fig. 5 presents a comparison of uncertainty between predicted and observed values. The results show that, except for a slight reduction in correlation at low Cy values and short periods, R^2 between predicted and true values exceeds 0.8 across all remaining training and test subsets. This indicates that the developed PSO-SVR machine-learning-based GMPEs can accurately capture the variation patterns of intensity measure, with prediction results exhibiting strong consistency relative to the observed data. Furthermore, these findings suggest that Cy exerts a measurable influence on the

Table 4. Hyper-parameters and prediction results calculated for linear elastic systems (Cy=0.3)

T (s)	C	γ	$\varepsilon_{\text{training}}$	R^2_{training}	R^2_{testing}	MAE_{training}	MAE_{testing}	$RMSE_{\text{training}}$	$RMSE_{\text{testing}}$	τ_{training}	σ_{training}	τ_{testing}	σ_{testing}
0.1	8.907	1.112	0.346	0.874	0.835	0.421	0.484	0.548	0.638	0.247	0.490	0.347	0.533
0.2	9.954	0.841	0.448	0.857	0.840	0.443	0.474	0.575	0.634	0.273	0.507	0.378	0.506
0.3	7.158	1.166	0.392	0.870	0.855	0.443	0.461	0.571	0.616	0.272	0.502	0.363	0.496
0.4	13.055	1.081	0.389	0.883	0.854	0.436	0.484	0.567	0.638	0.277	0.494	0.403	0.493
0.5	7.550	1.042	0.344	0.881	0.878	0.448	0.474	0.593	0.606	0.312	0.504	0.367	0.482
0.6	6.475	1.069	0.331	0.887	0.883	0.451	0.461	0.597	0.602	0.317	0.506	0.385	0.458
0.7	19.961	0.359	0.211	0.885	0.883	0.460	0.481	0.617	0.610	0.345	0.511	0.395	0.465
0.8	9.032	1.172	0.201	0.906	0.877	0.423	0.486	0.563	0.648	0.295	0.479	0.425	0.489
0.9	8.243	0.895	0.269	0.905	0.884	0.436	0.473	0.577	0.634	0.324	0.477	0.362	0.521
1.0	46.391	0.191	0.077	0.899	0.893	0.444	0.486	0.601	0.622	0.349	0.490	0.369	0.501
1.2	61.298	0.403	0.254	0.913	0.881	0.424	0.505	0.572	0.656	0.315	0.478	0.429	0.495
1.4	21.763	0.438	0.306	0.910	0.896	0.442	0.473	0.588	0.631	0.325	0.489	0.415	0.475
1.6	12.612	0.551	0.195	0.914	0.888	0.434	0.482	0.585	0.638	0.315	0.493	0.405	0.490
1.8	9.933	0.776	0.223	0.913	0.895	0.434	0.485	0.584	0.638	0.316	0.490	0.380	0.512
2.0	10.815	1.069	0.301	0.914	0.896	0.440	0.481	0.588	0.629	0.300	0.506	0.390	0.493
2.5	11.243	0.735	0.308	0.916	0.877	0.443	0.513	0.580	0.683	0.289	0.503	0.419	0.532
3.0	32.512	0.451	0.354	0.900	0.905	0.469	0.487	0.620	0.625	0.316	0.533	0.397	0.481
3.5	35.701	0.506	0.441	0.904	0.893	0.469	0.497	0.614	0.642	0.316	0.527	0.375	0.521
4.0	31.508	0.539	0.382	0.906	0.883	0.461	0.513	0.601	0.675	0.282	0.532	0.421	0.527
4.5	13.986	0.681	0.332	0.900	0.888	0.469	0.504	0.613	0.662	0.315	0.526	0.373	0.546
5.0	11.254	0.965	0.628	0.897	0.894	0.492	0.490	0.626	0.632	0.316	0.541	0.349	0.527
6.0	7.899	1.047	0.406	0.904	0.875	0.468	0.512	0.605	0.662	0.302	0.525	0.386	0.538
7.0	8.823	0.856	0.447	0.899	0.874	0.470	0.520	0.609	0.676	0.282	0.540	0.436	0.516
8.0	7.167	0.733	0.183	0.894	0.878	0.463	0.502	0.615	0.662	0.304	0.534	0.371	0.543
9.0	11.004	0.804	0.191	0.899	0.863	0.450	0.527	0.601	0.676	0.278	0.532	0.413	0.535
10.0	11.969	1.634	0.295	0.900	0.875	0.442	0.503	0.591	0.648	0.281	0.521	0.370	0.533

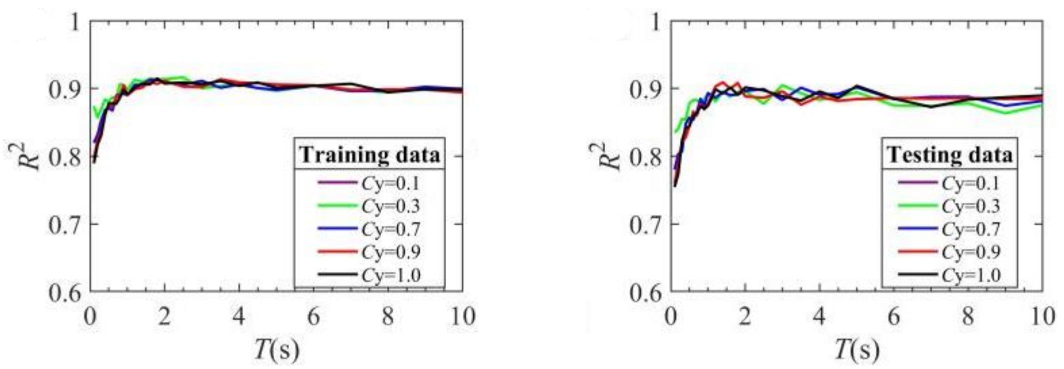
PSO-SVR machine-learning-based GMPE's predictive performance.

Furthermore, Fig. 6 illustrates the variation of the residual total standard deviation with respect to period. The standard deviation of the total residuals remains approximately 0.6 for both the training and test datasets. Notably, the test set exhibits a marginally higher value—a result consistent with expectations, as the model parameters were estimated exclusively from the training data and thus the test set was not involved in the regression fitting process. Moreover, comparison of the standard deviations of total residual between the present model and that of CLM20 reveals broadly comparable performance overall. Crucially, within the very short-period range, the present model exhibits no appreciable increase in residual standard deviation—indicating robust predictive stability. Notably, the dataset sizes employed in both studies are nearly identical, thereby supporting the validity of this cross-study comparison.

This indicates that the model exhibits low dispersion and minimal uncertainty across different datasets, with prediction results demonstrating good robustness and reliability. Overall, the PSO-

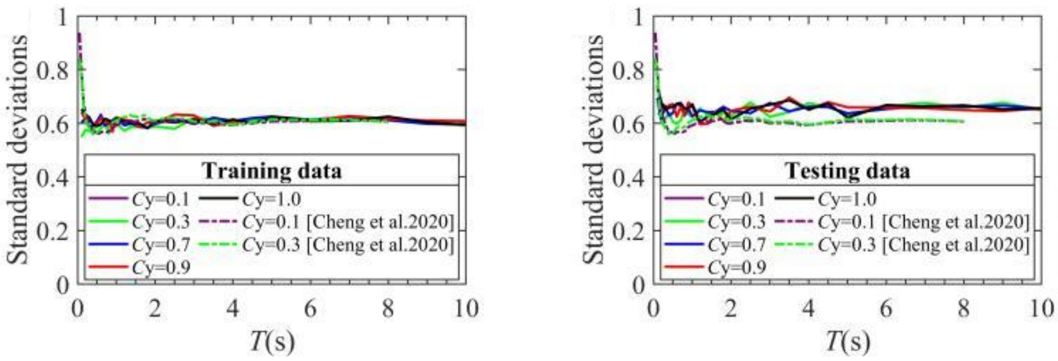
Table 5. Hyper-parameters and prediction results calculated for linear elastic systems (Cy=0.1)

T (s)	C	γ	$\varepsilon_{training}$	$R^2_{training}$	$R^2_{testing}$	$MAE_{training}$	$MAE_{testing}$	$RMSE_{training}$	$RMSE_{testing}$	$\tau_{training}$	$\sigma_{training}$	$\tau_{testing}$	$\sigma_{testing}$
0.1	9.580	1.078	0.204	0.899	0.862	0.389	0.454	0.518	0.614	0.244	0.457	0.344	0.504
0.2	10.842	1.517	0.412	0.885	0.885	0.427	0.428	0.558	0.553	0.268	0.490	0.311	0.458
0.3	11.432	0.969	0.314	0.893	0.869	0.412	0.474	0.546	0.612	0.270	0.475	0.388	0.475
0.4	67.726	0.032	0.338	0.881	0.875	0.455	0.467	0.587	0.633	0.328	0.487	0.388	0.499
0.5	14.695	0.822	0.404	0.906	0.865	0.415	0.494	0.543	0.653	0.307	0.447	0.400	0.517
0.6	47.986	0.401	0.276	0.890	0.903	0.443	0.442	0.588	0.586	0.318	0.495	0.361	0.462
0.7	80.403	0.052	0.322	0.898	0.880	0.455	0.472	0.600	0.605	0.337	0.496	0.384	0.467
0.8	22.274	0.603	0.046	0.907	0.889	0.415	0.463	0.572	0.607	0.306	0.483	0.384	0.470
0.9	7.767	0.698	0.186	0.908	0.895	0.423	0.472	0.569	0.618	0.309	0.478	0.402	0.470
1.0	4.521	0.989	0.313	0.911	0.900	0.428	0.463	0.566	0.608	0.300	0.480	0.401	0.457
1.2	92.967	0.111	0.201	0.911	0.883	0.431	0.498	0.571	0.664	0.303	0.484	0.462	0.475
1.4	37.463	0.104	0.168	0.903	0.902	0.448	0.473	0.606	0.602	0.334	0.506	0.404	0.446
1.6	26.676	0.802	0.237	0.916	0.897	0.424	0.469	0.565	0.623	0.301	0.477	0.376	0.492
1.8	69.814	0.188	0.334	0.907	0.890	0.456	0.482	0.596	0.646	0.313	0.508	0.430	0.482
2.0	10.038	0.291	0.331	0.902	0.895	0.464	0.480	0.611	0.637	0.324	0.517	0.413	0.486
2.5	46.696	0.502	0.116	0.905	0.892	0.448	0.472	0.608	0.616	0.321	0.516	0.378	0.486
3.0	4.569	0.688	0.001	0.898	0.895	0.458	0.476	0.625	0.613	0.329	0.531	0.352	0.502
3.5	10.908	1.013	0.302	0.907	0.881	0.447	0.501	0.584	0.673	0.265	0.520	0.427	0.520
4.0	15.092	1.007	0.381	0.906	0.878	0.452	0.497	0.589	0.658	0.281	0.518	0.402	0.520
4.5	6.856	1.133	0.331	0.897	0.891	0.465	0.479	0.606	0.631	0.294	0.530	0.366	0.514
5.0	7.269	0.966	0.284	0.896	0.890	0.462	0.487	0.609	0.631	0.301	0.530	0.358	0.521
6.0	24.588	1.003	0.311	0.898	0.885	0.457	0.482	0.595	0.646	0.280	0.525	0.386	0.517
7.0	15.262	1.017	0.400	0.895	0.876	0.463	0.505	0.601	0.662	0.291	0.527	0.388	0.535
8.0	15.755	0.962	0.387	0.894	0.877	0.457	0.495	0.599	0.652	0.279	0.531	0.396	0.519
9.0	8.115	1.115	0.418	0.893	0.873	0.465	0.509	0.601	0.654	0.299	0.522	0.374	0.537
10.0	6.576	1.414	0.293	0.893	0.882	0.450	0.490	0.596	0.636	0.293	0.520	0.335	0.542



(a) (b)

Figure 5. Correlation results of PSO-SVR machine-learning-based GMPEs



(a) (b)

Figure 6. The variation of the standard deviation of the residuals with the period

SVR machine-learning-based GMPEs achieves strong generalization ability and high prediction accuracy, providing a solid computational foundation for the development and application of GMPEs and offering a feasible technical approach for predicting intensity measures in related engineering contexts.

4.2 Model interpretability analysis

Although machine learning regression algorithms exhibit exceptional nonlinear fitting capabilities, the resulting non-parametric regression models are inherently black-box in nature, which reduces the transparency and credibility of the prediction process [38]. The shapley additive explanations (SHAP) method based on game theory has widely adopted to explain the internal mechanisms of various non-parametric GMPEs developed by data-driven machine learning algorithms [31, 39]. Accordingly, in this study, the SHAP method is employed to explain the influence of the input feature variables on the prediction results of the developed PSO-SVR machine-learning-based GMPEs (i.e., black-box models), thereby enhancing the interpretability of these black-box models and improving the credibility of the prediction results.

Fig. 7 shows the SHAP summary plot for PSO-SVR machine-learning-based GMPEs at Cy=0.1 across four representative periods (T=0.5 s, 1 s, 2 s, and 4 s). The SHAP value quantifies the marginal contribution of each input feature variables (Eq. (7)) to the PSO-SVR machine-learning-based GMPEs, where positive values indicate an increase in predicted V_E and negative values indicate a decrease. The color gradient represents the feature values from low (blue) to high (red), and four feature variables are ranked vertically by overall importance. Overall, the interpretability analyses reveal that earthquake magnitude (M_w) and source-to-site distance (R_{rup}) are consistently the two most dominant features influencing V_E across all representative periods, which aligns perfectly with the fundamental physical laws of seismic wave generation and propagation. The importance of site conditions (V_{S30}) and fault mechanisms (FM) is relatively lower than that of M_w and R_{rup}, and their importance ranking positions fluctuate with the periods. A pronounced positive correlation is observed between input variable M_w and output variable V_E: higher values (red points) are associated with positive SHAP values, indicating that larger earthquake events significantly increase the predicted V_E, while lower values (blue points) correspond to negative SHAP values, reducing the predicted V_E. The R_{rup} is inversely proportional to the M_w, which shows

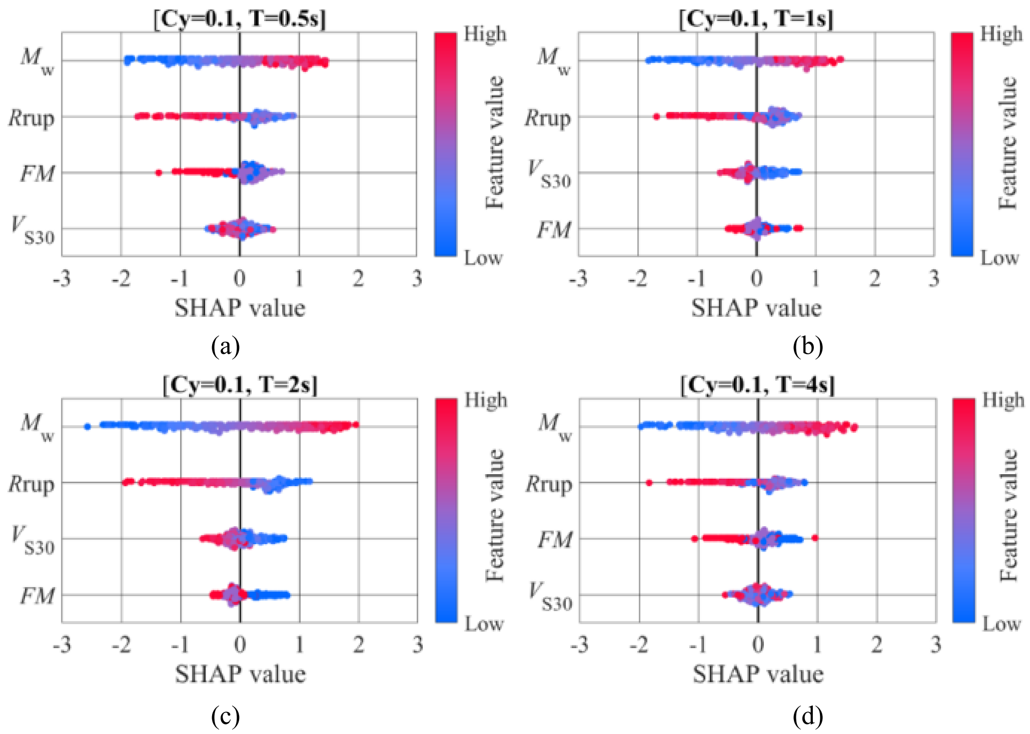


Figure 7. SHAP summary plot of PSO-SVR machine-learning-based GMPEs

a clear negative correlation with the output variable V_E. Low values of R_{rup} (red points) yield positive SHAP values, meaning that structures located closer to the seismogenic fault experience substantially higher input energy demands, whereas high values of R_{rup} (blue points) produce negative SHAP values, leading to lower energy demands.

The SHAP-value distributions for V_{S30} and *FM* are considerably narrower than those for M_{w} and R_{rup}, indicating a relatively weaker influence on the V_{E}. This hierarchy is consistent with the physical characteristics of seismic energy input. M_{w} primarily controls the amount of energy released at the source and the duration, while R_{rup} governs the attenuation of seismic-wave energy during propagation. Therefore, these two variables represent the dominant source and path effects and naturally account for most of the variability in V_{E}.

In contrast, V_{S30} characterizes site conditions and affects ground motions primarily through site amplification. Although the V_{S30} values in the database span a relatively broad range (approximately 100-2100 m/s), more than 90% of the records are associated with sites having V_{S30} values below 600 m/s. Consequently, the dataset is dominated by soft-to-medium stiffness site conditions, while stiff-site records are comparatively limited. This uneven distribution reduces the variability associated with site conditions and may partially limit the apparent contribution of V_{S30} in the SHAP analysis. From a physical perspective, V_{E} is a cumulative response parameter that depends on the combined effects of amplitude, frequency content, and duration, making it generally less sensitive to local site amplification than to source size and propagation distance.

Similarly, *FM* represents different faulting styles, including SS, NF, and TF. The influence of faulting style on ground motions is generally associated with differences in rupture geometry,

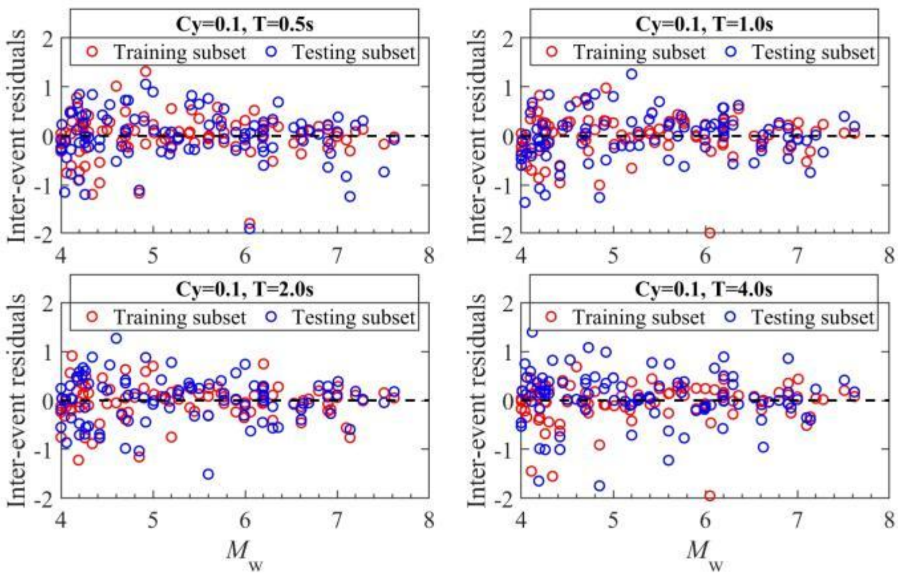


Figure. 8 Distribution of inter-event residuals and M_w

stress-release characteristics, and hanging-wall effects. However, these effects are typically smaller than the source-energy effect represented by M_w and the attenuation effect represented by R_{rup}. In addition, the faulting-style distribution in the database is unbalanced, consisting of 56 SS events, 30 TF events, and only 13 NF events. Such an imbalance reduces the statistical contrast among faulting categories and may further weaken the contribution of *FM* identified by the machine-learning model.

4.3 Residual analysis

Residual analysis is an important approach for evaluating the performance of GMPEs, and in prior studies, residuals are typically decomposed into inter-event and intra-event components [40-41]. Fig. 8 illustrates the distribution of inter-event residuals with respect to M_w, with four representative periods—T=0.5 s, T=1.0 s, T=2.0 s, and T=4.0 s—selected for analysis. The results show that residuals are approximately symmetrically distributed around the zero baseline, and the overall pattern indicates negligible bias. Numerically, the vast majority of residual values fall within the range of ±1. In the natural logarithmic coordinate system, this level of residual dispersion is considered highly favorable, indicating the absence of systematic prediction bias. In other words, the model maintains high fitting stability and low systematic error across different seismic magnitudes, demonstrating strong robustness and reliability in intra-event predictions.

Figs. 9-11 present the distributions of intra-event residuals with respect to R_{rup}, M_w and V_{S30}, respectively. In all cases, the residuals are generally symmetrically distributed around the zero baseline, with most values confined within the ±1.5 range and no evident systematic bias. Specifically, no discernible trend of residuals is observed with increasing rupture distance, and the distribution remains stable across different magnitude levels and site conditions. These results indicate that the proposed PSO-SVR machine-learning-based GMPEs maintain consistent predictive performance without introducing systematic errors associated with source distance,

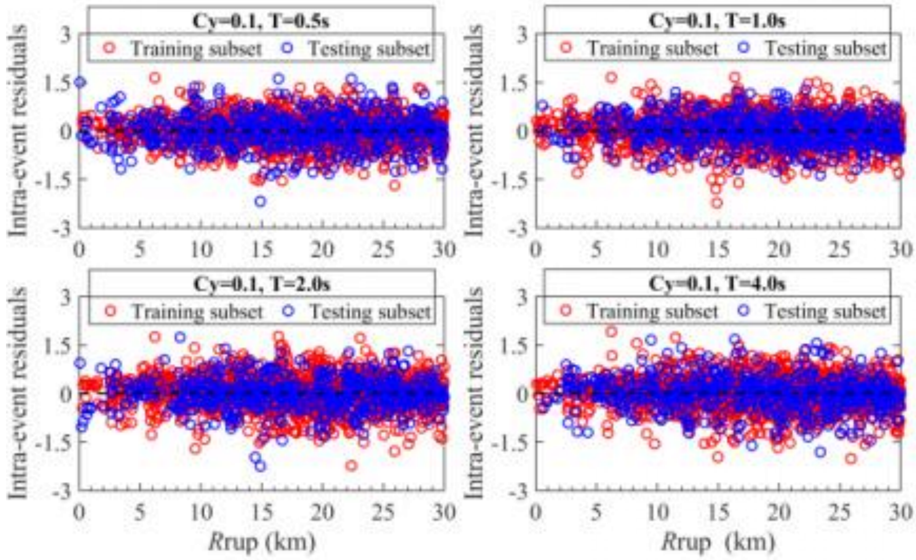


Figure 9. Distribution of intra-event residuals and R_{rup}

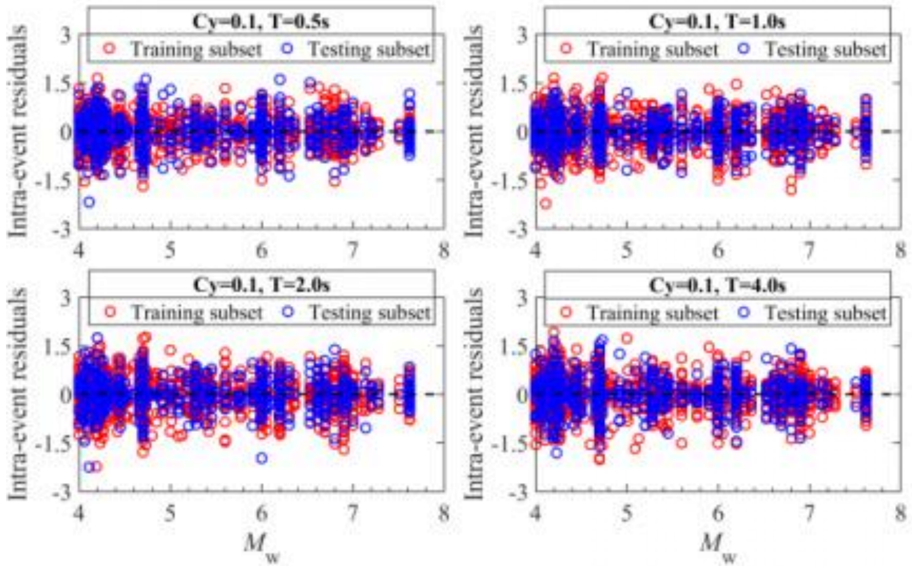


Figure 10. Distribution of intra-event residuals and M_{w}

earthquake magnitude, or site effect parameters, thereby further confirming their robustness and reliability.

Velocity pulses constitute a defining characteristic of near-fault ground motions; thus, evaluating the model's capability to capture such pulse features is essential. Within the NGA-West2 database, ground motion records are classified as pulse-like or non-pulse-like using established pulse identification criteria—primarily those proposed by Shahi and Baker [42]. Therefore, 112 pulse-like ground motion records from the near-fault ground motion database are

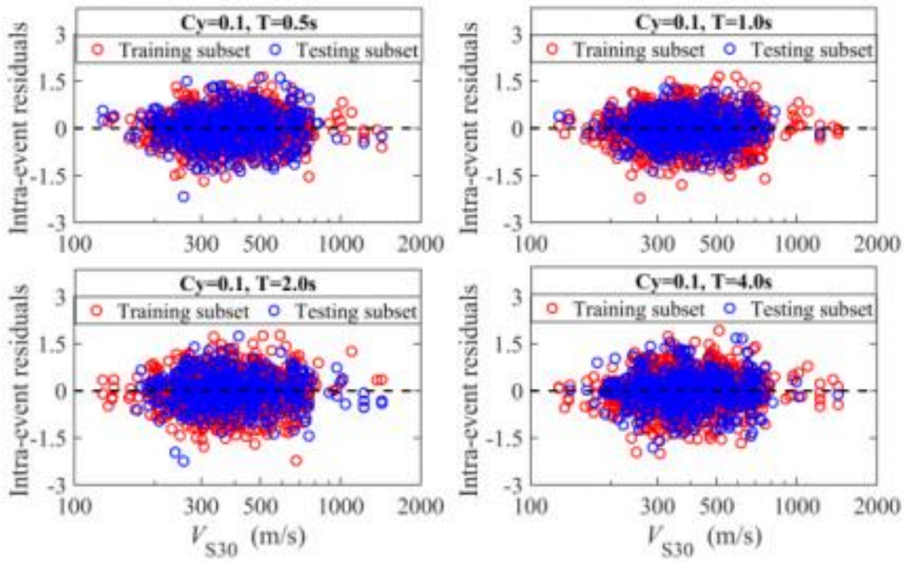


Figure 11. Distribution of intra-event residuals and V_{S30}

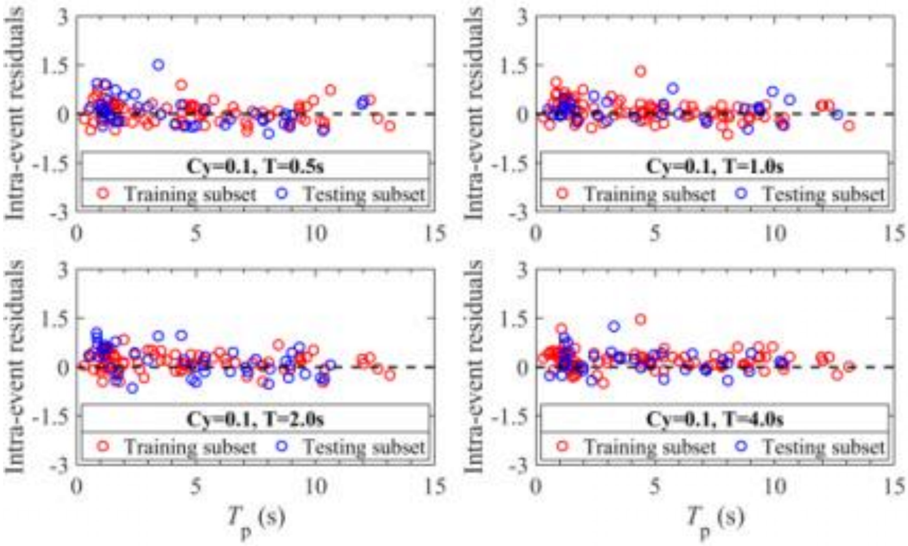
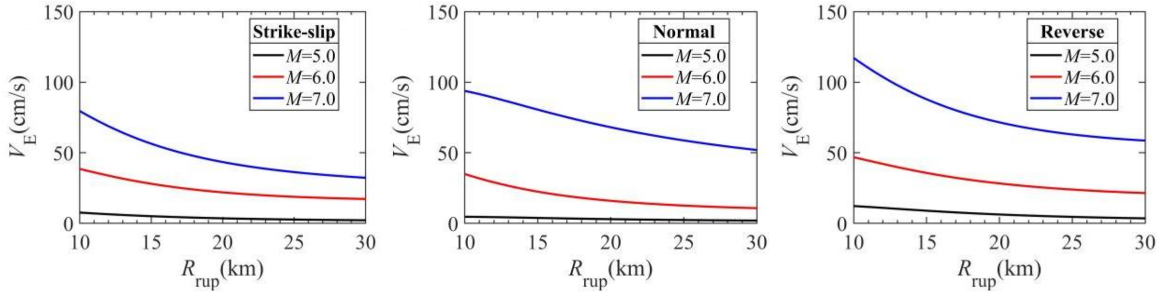


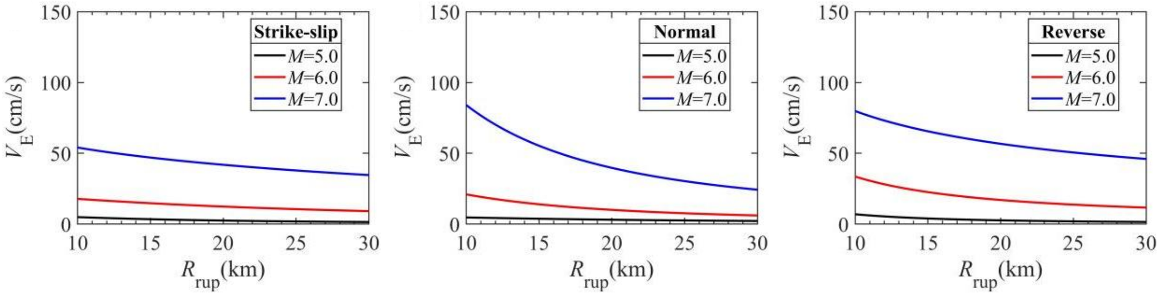
Figure 12. Distribution of within intra-event residuals and *T*p

selected as the pulse subset to verify the distribution of model residuals with respect to pulse period (T_p), as shown in Fig. 12. The results indicate that the residuals show an overall unbiased distribution across the range of T_p, without any clear systematic trend or dependence. This suggests that, although T_p is not explicitly included as an independent predictor in the model, its influence is implicitly captured through the existing explanatory variables and the data-driven learning process. As a result, no significant bias is observed when the model is applied to pulse-like ground motions.



(a) (b) (c)

Figure 13. Effect of M_w on prediction performance with T=0.5 s



(a) (b) (c)

Figure 14. Effect of M_w on prediction performance with T=2.0 s

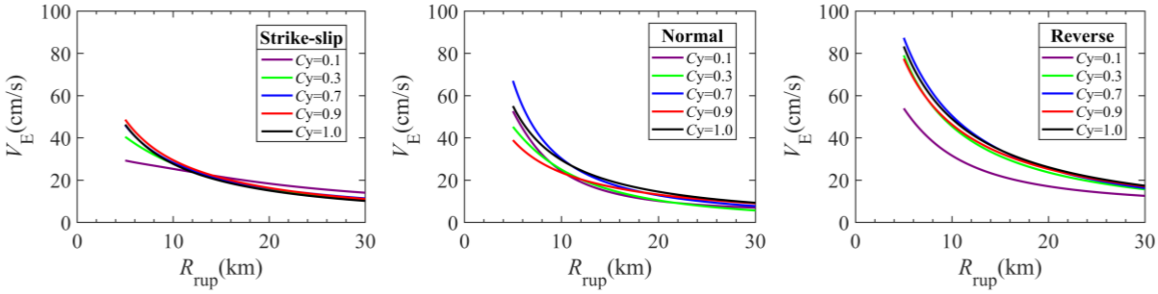
5. Model influencing factors

5.1 The influence of M_w and fault type

The intensity of ground motion results from the coupled effects of multiple factors, among which earthquake magnitude, fault type, and site parameters (e.g., V_{S30}) may significantly influence the level of input energy. Therefore, investigating the role of these parameters in the predictive outcomes is both necessary and scientifically meaningful. Fig. 13-14 illustrates the influence of M_w and fault type on the PSO-SVR machine-learning-based GMPEs prediction performance at representative periods (T=0.5 s and T=2.0 s). The results show that as earthquake magnitude increases, the V_E value exhibits a clear upward trend. This can be attributed to the fact that larger earthquakes release more energy, thereby increasing the level of input energy imparted to the ground.

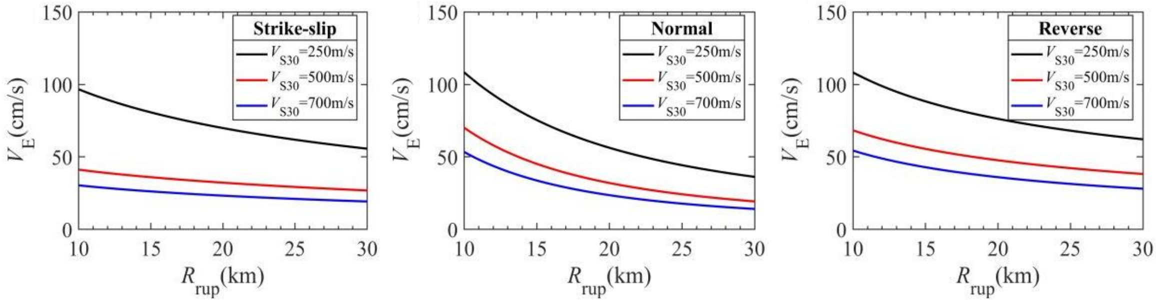
5.2 The influence of *Cy*

As illustrated in Fig. 15, V_E exhibits a monotonic decay with increasing R_{rup} across all three fault types. The influence of Cy on V_E displays a distinct near-field effect: it exerts a substantial impact only within the near-field region. With increasing R_{rup}, the V_E curves corresponding to different Cy values progressively converge; by R_{rup}=30 km, the curves become nearly



(a) (b) (c)

Figure 15. Influence of the *Cy* on the prediction results



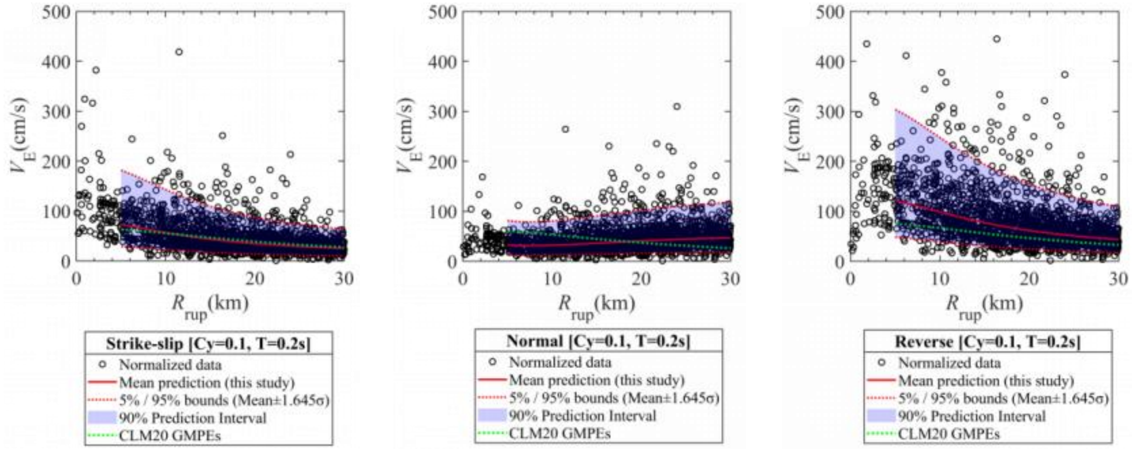
(a) (b) (c)

Figure 16. Effect of V_{S30} on model prediction results (T=2.0 s)

indistinguishable, indicating that *Cy*'s influence on V_E is effectively negligible at such distances. A comparative analysis across fault types reveals clear fault-type dependence in both the magnitude and spatial extent of *Cy*'s influence on V_E. Under strike-slip faulting, *Cy*'s influence is weakest: near-field differences in V_E among *Cy* values are minimal, and curve convergence occurs most rapidly. Under normal faulting, *Cy*'s influence is moderate, with near-field differentiation and convergence rates intermediate between those observed for strike-slip and reverse faults. Under reverse faulting, *Cy*'s influence is strongest: near-field V_E variations across *Cy* values are most pronounced, and curve convergence is slowest—resulting in the largest effective spatial range of *Cy*'s influence on V_E among the three fault types.

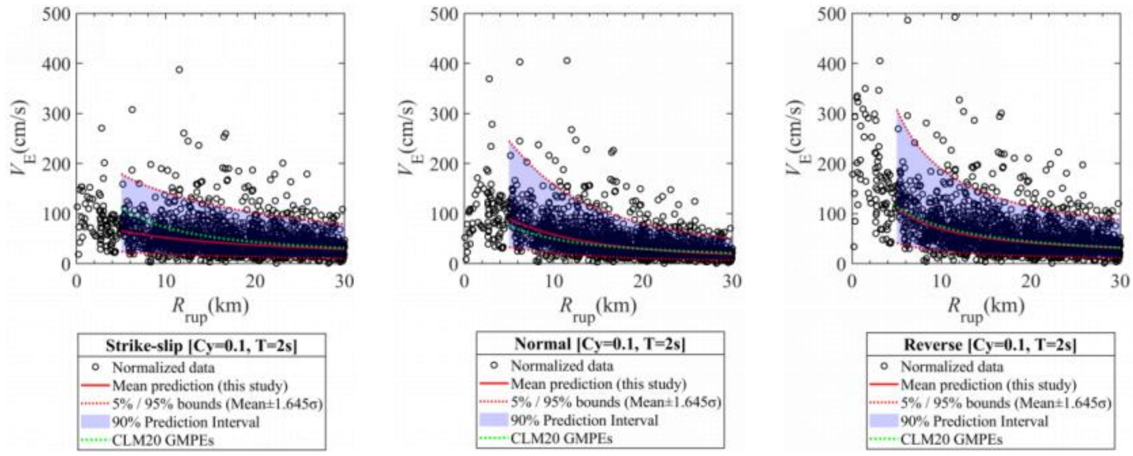
5.3 The influence of V_{S30}

Fig. 16 illustrates the influence of V_{S30} on the model's prediction results. It can be observed that as V_{S30} increases, the predicted V_E value exhibits a gradually decreasing trend. This indicates that under weak ground conditions (i.e., low V_{S30}), the amplification effect of ground motion is more pronounced, whereas in relatively stiff sites, this amplification effect is significantly reduced. This phenomenon reflects the critical role of site conditions in the propagation of ground motion, particularly the nonlinear response of low-velocity soil layers that enhances shaking and contributes to energy accumulation, thereby increasing ground motion amplitudes. Therefore, when developing or applying GMPEs, incorporating site-specific V_{S30} parameters is essential for



(a) (b) (c)

Figure 17. Relationship fitting between the observed values and the predicted values



(a) (b) (c)

Figure 18. Relationship fitting between the observed values and the predicted values

improving the reliability and applicability of prediction outcomes.

5.4 Predicted model

Evaluating the model's fit to observed values is a critical step in assessing its accuracy. Since predicted attenuation curves are typically derived under specific earthquake scenarios, actual ground motion records often do not fully satisfy these predefined conditions, limiting the validity of direct comparisons. To overcome this limitation, Kanno et al. [43] proposed an observation normalization method that enables the conversion of measured data from various earthquake scenarios into a common reference scenario, thereby facilitating more meaningful comparisons. In this study, the reference scenario was defined as follows: R_{rup}=0-30 km, V_{S30}=250 m/s, and M_w=6.5. As shown in Figs. 17-18, after applying this normalization procedure, most observed values fall within 90% prediction interval. This result indicates that the PSO-SVR machine-learning-based GMPEs developed in this study accurately captures the attenuation characteristics of constant-intensity relative input energy of near-fault ground motions, confirming the applicability and reliability of the PSO-SVR machine-learning-based GMPEs.

To further validate the theoretical soundness and predictive capability of the proposed model, a comparative assessment was performed against the CLM20 model—a widely accepted classical GMPE specifically developed for characterizing elastic-plastic input energy demand. The results indicate that, although discrepancies exist in certain regions, the overall trends are largely consistent. Such consistency provides indirect support for the physical plausibility of the proposed model, rather than indicating redundancy. A closer inspection reveals that the differences are concentrated in specific period and distance ranges. For example, at a period of T=0.2 s and at small R_{rup}, the proposed model exhibits a relatively slower attenuation rate. This behavior suggests an improved capability to capture the near-fault saturation effect, which is often insufficiently represented in traditional GMPEs developed using random-effects regression approaches.

It should be noted that for the normal-faulting scenario with Cy=0.1 and T=0.2 s, the attenuation trend of the V_{E} with increasing distance is less evident than that observed in other cases. This phenomenon is mainly attributable to the limited amount and distribution of available normal-fault ground-motion data. Specifically, only 13 normal-fault earthquake events satisfied the selection criteria adopted in this study, and 9 of these events had magnitudes smaller than 6.0. As a result, the dataset lacks sufficient coverage of moderate-to-large normal-fault earthquakes over a broad distance range, reducing its statistical representativeness and increasing the uncertainty of the regression results.

Furthermore, short-period ground-motion parameters (T=0.2 s) are generally more sensitive to event-to-event variability and local site effects. When combined with a limited number of records, these sources of variability can partially obscure the expected distance-attenuation behavior. Therefore, the attenuation relationship obtained for this specific case should be interpreted with caution.

Nevertheless, this phenomenon is confined to the normal-faulting case with Cy=0.1 and T=0.2 s and does not affect the attenuation patterns observed for most other faulting mechanisms, spectral periods, or exceedance probability levels. The overall trends derived by the proposed model remain physically reasonable and consistent with established ground-motion attenuation characteristics. Future updates incorporating additional moderate-to-large normal-fault records would further improve the robustness of the model for this specific condition.

6. Conclusions

Based on a near-fault ground motion database, this study develops PSO-SVR machine-learning-based GMPEs for constant-strength relative input energy using earthquake magnitude, fault type, rupture distance, and V_{S30} as predictor variables. By combining SVR-PSO, the proposed model captures the nonlinear relationship between seismic source characteristics, wave propagation effects, and site conditions. The main conclusions are summarized as follows:

1. Compared with conventional empirical models, the proposed PSO-SVR machine-learning-based GMPEs more effectively capture the attenuation characteristics of constant-strength relative input energy for near-fault ground motions, with reduced prediction uncertainty. This indicates that the PSO-SVR framework can better represent the nonlinear coupling effects of

source, path, and site factors.

2. Earthquake magnitude, rupture distance, and site conditions are the primary factors controlling constant-strength relative input energy. Larger magnitudes increase seismic energy input, while increasing rupture distance reduces energy demand because of geometric spreading and attenuation during wave propagation. Lower V_{S30} values correspond to higher relative input energy, indicating that soft soil sites tend to amplify seismic energy input.

3. Fault type significantly influences the distribution of constant-strength relative input energy. Different rupture mechanisms generate distinct radiation patterns and directivity effects, leading to different near-fault energy concentrations, particularly for pulse-like ground motions.

4. The proposed model captures the main physical characteristics of near-fault seismic energy input and provides a useful quantitative tool for energy-based performance seismic design.

The model is calibrated using records with moment magnitudes Mw ranging from 4.00 to 7.62 and rupture distances R_{rup} less than 30 km. Therefore, its applicability is limited to this parameter range, and extrapolation beyond the calibrated domain requires additional validation.

Acknowledgements

The work is supported by The National Natural Science Foundation of China (Grant no.52408535). The Hubei Provincial Natural Science Foundation of China (Grant no.2024AFB971).

Data availability statement

The source code and PSO-SVR machine-learning-based GMPEs that support the findings of this study will be publicly available on GitHub at https://github.com/l123qh/PSO-SVR-GMPEs after this study is published.

References

1. Cai, J., Bu, G.B., Zhou, J., Zuo, Z.L., Yang, C., Chen, Q.J. (2015). Modification of ductility reduction factor for strength eccentric structures subjected to pulse-like ground motions. Journal of Earthquake Engineering, 20(1), 12-38. https://doi.org/10.1080/13632469.2015.1053009.
2. Jäger, C., Adam, C. (2013). Influence of collapse definition and near-field effects on collapse capacity spectra. Journal of Earthquake Engineering, 17(6), 859-878. https://doi.org/10.1080/13632469.2013.795842.
3. Jalali, R.S., Jokandan, M.B., Trifunac, M.D. (2012). Earthquake response of a three-span simply supported bridge to near-field pulse and permanent-displacement step. Soil Dynamics and Earthquake Engineering, 43, 380-397. https://doi.org/10.1016/j.soildyn.2012.08.004.
4. Wen, W.P., Zhai, C.H., Li, S., Chang, Z., Xie, L.L. (2014). Constant damage inelastic displacement ratios for near-fault pulse-like ground motions. Engineering Structures, 59, 599-607. https://doi.org/10.1016/j.engstruct.2013.11.011.
5. Zamora, M., Riddell, R. (2011). Elastic and inelastic response spectra considering near-fault effects. Journal of Earthquake Engineering, 15(5), 775-808. https://doi.org/10.1080/13632469.2011.555058.
6. Yaghmaei-Sabegh, S. (2012). Application of wavelet transforms on characterization of inelastic

displacement ratio spectra for pulse-like ground motions. Journal of Earthquake Engineering, 16(4), 561-578. https://doi.org/10.1080/13632469.2011.640739.
7. Karalar, M., Padgett, J.E., Dicleli, M. (2012). Parametric analysis of optimum isolator properties for bridges susceptible to near-fault ground motions. Engineering Structures, 40, 276-287. https://doi.org/10.1016/j.engstruct.2012.02.023.
8. Masaeli, H., Khoshnoudian, F., Ziaei, R. (2015). Rocking soil-structure systems subjected to near-fault pulses. Journal of Earthquake Engineering, 19(3), 461-479. https://doi.org/10.1080/13632469.2014.990652.
9. Decanini, L.D., Mollaioli, F. (2001). An energy-based methodology for the assessment of seismic demand. Soil Dynamics and Earthquake Engineering, 21, 113-137. https://doi.org/10.1016/S0267-7261(00)00102-0.
10. Hu, J., Tian, H., Tan, J., Liu, M., Jin, C. (2025). Empirical correlation between elastic input energy and typical intensity measures for offshore ground motions. Earthquake Engineering and Engineering Vibration, 24(3), 653-673. https://doi.org/10.1007/s11803-025-2329-1.
11. Manfredi, G. (2001). Evaluation of seismic energy demand. Earthquake Engineering & Structural Dynamics, 30, 485-499. https://doi.org/10.1002/eqe.17.
12. Ordaz, M., Huerta, B., Reinoso, E. (2003). Exact computation of input-energy spectra from Fourier amplitude spectra. Earthquake Engineering & Structural Dynamics, 32, 597-605. https://doi.org/10.1002/eqe.240.
13. Decanini, L.D., Mollaioli, F. (1998). Formulation of elastic earthquake input energy spectra. Earthquake Engineering & Structural Dynamics, 27(12), 1503-1522. https://doi.org/10.1002/(SICI)1096-9845(199812)27:12<1503::AID-EQE797>3.0.CO;2-A.
14. Chai, Y.H., Fajfar, P. (2000). A procedure for estimation input energy spectra for seismic design. Journal of Earthquake Engineering, 4(4), 539-561. https://doi.org/10.1080/13632460009350382.
15. Shargh, G.B., Barati, R. (2020). Estimation of inelastic seismic input energy. Soil Dynamics and Earthquake Engineering, 142, 106505. https://doi.org/10.1016/j.soildyn.2020.106505.
16. Cheng, Y., Wang, J., Zhang, Y., Mollaioli, F. (2023). Empirical correlations between the spectral input energy and spectral acceleration. Journal of Earthquake Engineering, 27(6), 1514-1533. https://doi.org/10.1080/13632469.2022.2083727.
17. Cheng, Y., Lucchini, A., Mollaioli, F. (2020). Ground-motion prediction equations for constant-strength and constant-ductility input energy spectra. Bulletin of Earthquake Engineering, 18, 37-55. https://doi.org/10.1007/s10518-019-00725-x.
18. Alıcı, F.S., Sucuoğlu, H. (2016). Prediction of input energy spectrum: attenuation models and velocity spectrum scaling. Earthquake Engineering & Structural Dynamics, 45(13), 2137-2161. https://doi.org/10.1002/eqe.2749.
19. Zhai, C., Ji, D., Wen, W., Lei, W., Xie, L., Gong, M. (2016). The inelastic input energy spectra for main shock-aftershock sequences. Earthquake Spectra, 32(4), 2149-2166. https://doi.org/10.1193/121315EQS182M.
20. Kalkan, E., Kunnath, S.K. (2007). Effective cyclic energy as a measure of seismic demand. Journal of Earthquake Engineering, 11, 725-751. https://doi.org/10.1080/13632460601033827.
21. Dindar, A.A., Yalçin, C., Yüksel, E., Özkaynak, H., Büyüköztürk, O. (2015). Development of earthquake energy demand spectra. Earthquake Spectra, 31, 1667-1689. https://doi.org/10.1193/011212EQS010M.
22. Building Standard Law (BSL) (2015). Notification No. 631 of the Ministry of Land, Infrastructure, Transport and Tourism: Earthquake-resistant structural calculation based on energy balance. Ministry of Land, Infrastructure, Transport and Tourism, Tokyo, Japan.
23. Benavent-Climent, A., López-Almansa, F., Bravo-González, D.A. (2010). Design energy input spectra for moderate-to-high seismicity regions based on Colombian earthquakes. Soil Dynamics and Earthquake Engineering, 30, 1129-1148. https://doi.org/10.1016/j.soildyn.2010.04.022.
24. Amiri, G.G., Darzi, G.A., Amiri, J.V. (2008). Design elastic input energy spectra based on Iranian earthquakes. Canadian Journal of Civil Engineering, 35, 635-646. https://doi.org/10.1139/L08-013.

25. Tselentis, G.A., Danciu, L., Sokos, E. (2010). Probabilistic seismic hazard assessment in Greece-Part 2: acceleration response spectra and elastic input energy spectra. Natural Hazards and Earth System Sciences, 10, 41-49. https://doi.org/10.5194/nhess-10-41-2010.
26. Yang, Y., Cheng, Y., Yang, Y., Yuan, R., He, Y. (2024). A deep-neural-network-based prediction model for elastic input energy spectra of horizontal and vertical ground motions. Bulletin of the Seismological Society of America, 114(5), 2639-2653. https://doi.org/10.1785/0120240012.
27. Hu, J., Tan, J. (2022). GMPEs for elastic input energy spectra of horizontal and vertical offshore ground motions based on the ETMC database in Japan. Soil Dynamics and Earthquake Engineering, 155, 107198. https://doi.org/10.1016/j.soildyn.2022.107198.
28. Bozorgnia, Y., Abrahamson, N.A., Atik, L.A., Ancheta, T.D., Atkinson, G.M., Baker, J.W., Baltay, A., Boore, D.M., Campbell, K.W., Chiou, B.S.J., Darragh, R. (2014). NGA-West2 research project. Earthquake Spectra, 30(3), 973-987. https://doi.org/10.1193/072113EQS209M.
29. Zhu, J., Li, S., Song, J. (2022). Magnitude estimation for earthquake early warning with multiple parameter inputs and a support vector machine. Seismological Research Letters, 93(1), 126-136. https://doi.org/10.1785/0220210144.
30. Uang, C.M., Bertero, V.V. (1990). Evaluation of seismic energy in structures. Earthquake Engineering & Structural Dynamics, 19(1), 77-90. https://doi.org/10.1002/eqe.4290190108.
31. Hu, J., Ding, Y., Zhang, H., Jin, C., Wang, Z. (2023). A real-time seismic intensity prediction framework based on interpretable ensemble learning. Seismological Research Letters, 94(3), 1579-1602. https://doi.org/10.1785/0220220167.
32. Gong, M., Xie, L. (2005). Study on comparison between absolute and relative input energy spectra and effects of ductility factor. Acta Seismologica Sinica, 18(6), 717-726. https://doi.org/10.1007/s11589-005-0099-4.
33. Benavent-Climent, A., Pujades, L.G., López-Almansa, F. (2002). Design energy input spectra for moderate-seismicity regions. Earthquake Engineering & Structural Dynamics, 31, 1151-1172. https://doi.org/10.1002/eqe.153.
34. Cherkassky, V., Ma, Y. (2002). Selection of meta-parameters for support vector regression. International conference on artificial neural networks, Madrid, Spain, August.
35. Vapnik, V.N. (1995). The nature of statistical learning theory. Springer, New York, NY, USA.
36. Eberhart, R., Kennedy, J. (1995). A new optimizer using particle swarm theory. Proceedings of the sixth international symposium on micro machine and human science, Nagoya, Japan, October.
37. Kennedy, J., Eberhart, R. (1995). Particle swarm optimization. Proceedings of ICNN'95-International conference on neural networks, Perth, WA, Australia, November-December.
38. Tetik, T., Cihan, P. (2026). Non-parametric ground motion duration and energy-based intensity measure models: A novel stacking ensemble model using regional seismic data. Soil Dynamics and Earthquake Engineering, 203, 110135. https://doi.org/10.1016/j.soildyn.2026.110135.
39. Lundberg, S.M., Lee, S.I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774.
40. Yedulla, J., Sriwastav, R.K., Raghukanth, S.T.G. (2025). Non-parametric ground motion model for displacement response spectra and fling for Himalayan region using machine learning. Computers and Structures, 307, 107626. https://doi.org/10.1016/j.compstruc.2024.107626.
41. Sriwastav, R.K., Kammula, S.T., Yedulla, J., Raghukanth, S.T.G. (2025). Ground motion modeling and epistemic uncertainty using conditional variational autoencoders. Machine Learning: Earth, 1(1), 015005. https://doi.org/10.1088/3049-4753/adcfa5.
42. Shahi, S.K., Baker, J.W. (2014). An efficient algorithm to identify strong-velocity pulses in multicomponent ground motions. Bulletin of the Seismological Society of America, 104(5), 2456-2466. https://doi.org/10.1785/0120130191.
43. Kanno, T., Narita, A., Morikawa, N., Fujiwara, H., Fukushima, Y. (2006). A new attenuation relation for strong ground motion in Japan based on recorded data. Bulletin of the Seismological Society of America, 96(3), 879-897. https://doi.org/10.1785/0120050138.