

## Validation and application of empirical Shear wave velocity models based on standard penetration test

I. Shooshpasha<sup>a,\*</sup>, H. Mola-Abasi<sup>a</sup>, A. Jamalian<sup>b</sup>, Ü. Dikmen<sup>c</sup>, M. Salahi<sup>b</sup>

<sup>a</sup>Department of Civil Engineering, Babol University of Technology, Iran

<sup>b</sup>Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Guilan, Iran

<sup>c</sup>Department of Geophysical Engineering, Faculty of Engineering, Ankara University, 06100 Ankara, Turkey

Received 22 December 2012; accepted in revised form 28 October 2013

---

### Abstract

Shear wave velocity is a basic engineering tool required to define dynamic properties of soils. In many instances it may be preferable to determine  $V_s$  indirectly by common in-situ tests, such as the Standard Penetration Test. Many empirical correlations based on the Standard Penetration Test are broadly classified as regression techniques. However, no rigorous procedure has been published for choosing the models. This paper provides 1) a quantitative comparison of the predictive performance of empirical correlations; 2) a reproducible method for choosing the coefficients of previous empirical methods based on the particle swarm optimization and 3) taking into account the polynomial correlation, a new model proposed. Different empirical correlations are compared with different validation criteria. The best performing empirical correlations result in a new model and the unique coefficient associated determined by particle swarm optimization concluded. The more recent correlation only marginally improves prediction accuracy; thus, efforts should focus on improving data collection.

**Keywords:** Shear wave velocity; Standard penetration test; Least squares; Particle swarm optimization; Validation.

---

### 1. Introduction

Shear wave velocity ( $V_s$ ) is a principal geotechnical soil property for site response analysis. It is helpful for estimating an earthquake site response, liquefaction potential, soil density, site and soil classification since the dynamic soil modulus at small shear strain levels is directly related to  $V_s$ . In many cases, difficulties in soil sampling, and high costs of representative undisturbed specimens, in-situ investigations (e.g. seismic measurements) in lieu of laboratory element testing are preferred to determine  $V_s$  directly. Using seismic measurement techniques, a shear wave velocity profile can be established without boring and penetration [1]. These non-destructive and non-intrusive features make  $V_s$ -based approach a

---

\* Corresponding author. Tel: +98 1113232071; Fax: +98 111323107.  
E-mail address: shooshpasha@nit.ac.ir (I. Shooshpasha).

potentially attractive alternative for characterizing liquefaction susceptibility in sandy soils [2]. However, seismic in-situ tests are not always feasible, especially in urban areas due to space constraints and noise level limits. Therefore, it is necessary to determine  $V_s$  indirectly through methods such as the Standard Penetration Test (SPT) or the Cone Penetration Test (CPT), which are commonly used for conventional geotechnical site investigations. In geotechnical engineering, many soil parameters are associated with the Standard Penetration Test blow counts ( $N_{SPT}$ ).

The interdependency of factors involved in such problems prevents the use of regression analysis and demands a more extensive and sophisticated method resulted empirical correlations (ECs). The particle swarm optimization (PSO) can be used to evaluate ECs coefficients more accurate. In recent years, PSO has been used successfully in geotechnical practices (e.g. Zhanget al., [3], Khajehzadehet al., [4]).

This paper aims to 1) evaluate previous model coefficients; 2) quantify model improvement and 3) examine a new and more reliable model. The outline of this paper as follows, first reviews previous efforts in correlating  $N_{SPT}$  and  $V_s$ , then a brief explanation of the case histories under consideration. The phenomena of modeling with PSO are presented and finally the developed model is described and its accuracy is assessed through validation analysis.

## 2. Background to previously proposed correlations

Some researchers have proposed correlations mostly between  $N_{SPT}$  and  $V_s$  for different soils, e.g. sand, silt and clay (Ohba and Toriuma, [5], Fujiwara, [6], Ohsaki and Iwasaki, [7], Imai *et al.*, [8], Seed and Idriss, [9], Imai and Tonouchi, [10], Jinan, [11], Athanasopoulos, [12], Iyisan, [13]). Among of them, Imai and Yoshimura [14] studied 192 specimens and proposed empirical relationships between  $V_s$  and soil index properties. Sykora and Stokoe [15] asserted that geological age and soil type have little influence in predicting  $V_s$ . Jafari *et al.*, [16] presented a detailed historical review on statistical correlations between  $N_{SPT}$  and  $V_s$  for fine grained soils. Hasancebi and Ulusay [17] reported statistical correlations for sands and clays. Ulugergerli and Uyanık [18] investigated statistical correlations using 327 samples and delimited empirically a range for  $V_s$  values. Dikmen [19] investigated uncorrected SPT data and presented a correlation for all soil types and more recently, Akin *et al.*, [20] reported statistical correlation between  $V_s$  and  $N_{SPT}$  and proposed a site specific correlation as a function of depth and  $N_{SPT}$  valid up to a depth of 25 meters for alluvial and Pliocene type soils. Others have developed correlations accounting for stress-corrected  $V_s$ , energy-corrected  $N_{SPT}$  (e.g. Pitolakis *et al.*, [21], Kikue *et al.*, [22]),  $N_{SPT}$  and depth (e.g. Yoshida *et al.*, [23], Jamiolkowski *et al.*, [24]), fines content (e.g., Ohta and Goto [25]) and vertical effective stress ( $\sigma'_v$ ) (e.g. Brandenberger *et al.*, [26]). The  $V_s$  can also provide estimates of effective stress ( $\sigma'_v$ ) for clayey soils as suggested by Mayne and Martin [27]. Mayne [28] presented a relationship for the total unit weight ( $\gamma$ ) of saturated soils in terms of  $V_s$  and depth. Although almost all of the foregoing studies have focused on relationships between uncorrected  $N_{SPT}$  and  $V_s$ , the effect of fines content in order to restrict the disadvantage of SPT usage in clay soil type has been considered in this study. Presently existing 20 empirical relationships between  $V_s$  and  $N_{SPT}$  for all soil types in the literature is shown in Table 1.

## 3. Overview of database and case histories

A total of 42 CPT's, of which 11 included seismic profiling by means of a pair of geophones incorporated in the cone penetrometer, in combination with 30 exploration borings were performed at selected sites in the city of Adapazarı (Figure 1). These sites were selected

based on their performance after the devastating earthquake ( $M_w=7.4$ ) occurred in August 17, 1999 in Kocaeli (Turkey) [29].

Table 1. Inventory of proposed correlations between uncorrected  $NSPT$  and  $V_s$

Author(s)	Proposed correlation
Ohba and Toriuma (1970)	$V_s = 84N^{0.31}$
Fujiwara (1972)	$V_s = 92.1N^{0.337}$
Ohsaki and Iwasaki (1973)	$V_s = 91.4N^{0.39}$
Imai and Yoshimura (1975)	$V_s = 76N^{0.33}$
Imai et al (1975)	$V_s = 89.9N^{0.341}$
Ohta and Goto (1978)	$V_s = 85.35N^{0.348}$
Seed and Idriss (1981)	$V_s = 61.4N^{0.5}$
Imai and Tonouchi (1982)	$V_s = 96.9N^{0.314}$
Sykora and Stokoe (1983)	$V_s = 100.5N^{0.29}$
Jinan (1987)	$V_s = 116.1(N+0.3185)^{0.202}$
Yoshida et al (1988)	$V_s = 83N^{0.25}Z^{0.14}$ (Gravel type)
Jamiolkowski et al (1988)	$V_s = 69Z^{0.2}N^{0.17}$ (clay type)
Athanasopoulos (1995)	$V_s = 107.6N^{0.36}$
Iyisan (1996)	$V_s = 51.5N^{0.516}$
Kiku et al (2001)	$V_s = 68.3N^{0.292}$
Jafari et al (2002)	$V_s = 27N^{0.73}$ (clay type)
Hasancebi and Ulusay (2006)	$V_s = 90N^{0.309}$
Ulugergerli and Uyanık (2007)	$V_{SU} = 23.291Ln(N) + 405.61$ (upper bound) $V_{cr} = 52.9e^{-0.011N}$ (lower bound)
Dikmen (2009)	$V_s = 58N^{0.39}$
Akın et al (2011)	$V_s = 59.44N^{0.109}Z^{0.426}$

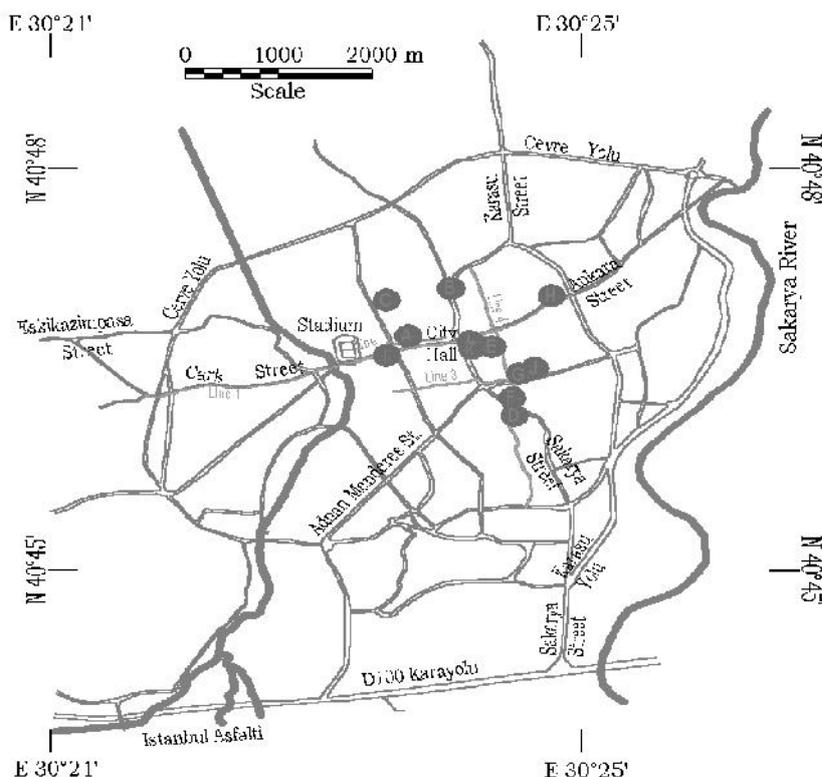


Figure 1. Location of in-situ tests in Adapazarı, Turkey [29]

The epicenter was located near the city of Izmit, and fault rupture was physically visible through most of the seismically impacted area (Marmara region); from Karamürsel to Akyazı. In the vicinity of Adapazarı, with peak ground accelerations recorded at approximately 0.4 g, as much as 70% of the buildings were subjected to large ground settlements, liquefaction, or subsidence and sea water inundation [30]. As illustrated in Figure 2, the southern shores of Izmit Bay are covered by Holocene deposits, these are principally fine-grained sandy sediments which become finer (more silty and clayey) northwards into the depths of Izmit Bay. Detailed geological and geotechnical properties of the site is available in PEER [29].

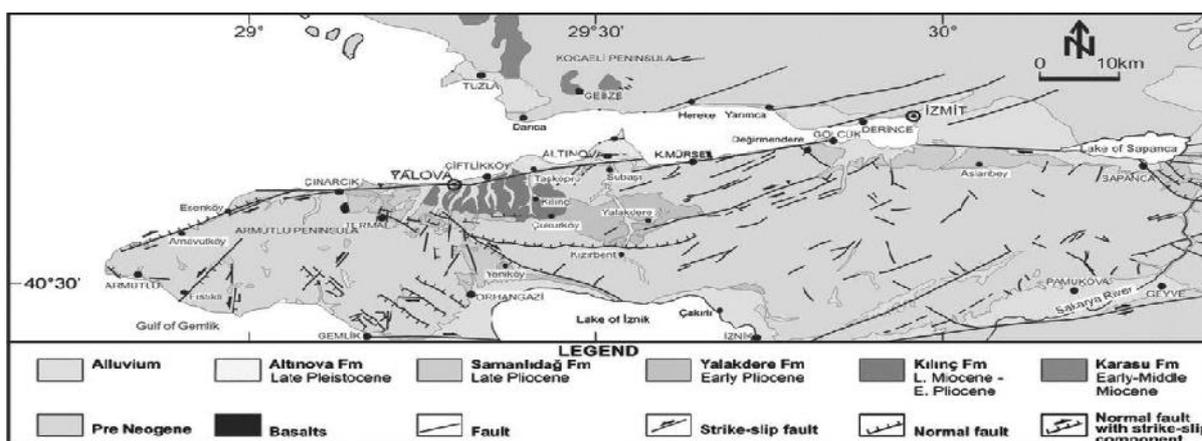


Figure 2. Simplified geological map of Armutlu peninsula (after Goncuoglu [30])

The database covers a wide spectrum of soils and seismic parameters, including soil layer depth (Z), energy corrected SPT blow count ( $N_{160}$ ), Fines Content (FC) (%  $>75\mu m$ ) and  $V_s$ . Further details regarding the measurement and interpretation of the foregoing parameters are available in PEER and Hanna *et al.*, [29, 30]. A sample of database is shown in Table 2. Figure 3 illustrates the percent distribution of descriptive variable characteristics for all case histories.

Table 2. Summary of in-situ tests performed in Adapazarı, Turkey [29]

Site Name	Latitude (°)	Longitude (°)	CPT	SCPT	Boring & SPT	SASW
Site A	40.77922	30.39487	√	√	√	√
Site B	40.78513	30.40024	√	√	√	√
Site C	40.78370	30.39221	√	√	√	√
Site D	40.76929	30.40828	√	√	√	√
Site E	40.77778	30.40518	√	√	√	×
Site F	40.77148	30.40795	√	√	√	×
Site G	40.77450	30.40896	√	√	√	√
Site H	40.78419	30.41295	√	√	√	×
Site I	40.77681	30.39223	√	√	√	×
Site J	40.77518	30.41077	√	√	√	√
Site K	40.77750	30.40340	√	×	√	×
Site L	40.77855	30.40272	√	×	√	×

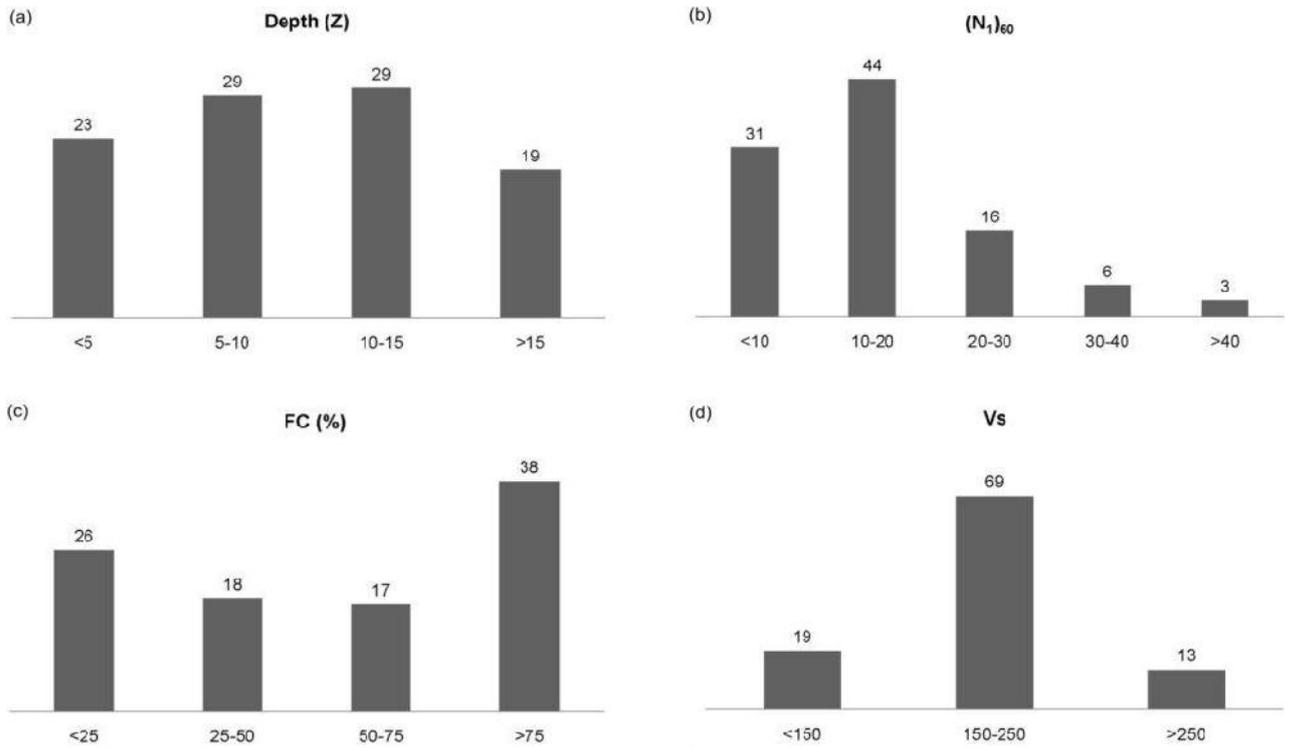


Figure 3. Distribution of descriptive variable characteristics for all case histories, a) Depth (m), b) energy corrected SPT blow count  $(N_1)_{60}$ , c) Fine contents (%), d) Shear wave velocity (m/s)

#### 4. Model Development

To investigate and compare the influence of parameters on the prediction of  $V_s$ , three velocity models were considered. The first model (Model-1) was of a form similar to the great majority of those listed in Table 1 in which the velocity was considered to be independent of any soil parameters except  $N_{SPT}$ , this model is of the form of:

$$V_s = a_1 N^{a_2} \quad (1)$$

The second model (Model-2),  $V_s$  was influenced by both  $N_{SPT}$  and  $Z$ , and proposed as the form,

$$V_s = a_1 Z^{a_2} N^{a_3} \quad (2)$$

In order to achieve better correlation, the third model (Model-3) presented that the  $V_s$  was influenced by  $N_{SPT}$  and  $Z$  and was of the form,

$$V_s = a_1 + a_2 Z + a_3 N + a_4 Z^2 + a_5 N^2 + a_6 NZ \quad (3)$$

Evaluation and validation of  $a_i$  coefficients in the three distinct models constitutes the main goal of this paper. Therefore, these models can be initially examined using the compiled database and simple regression analyses, containing least square approaches. The results are presented in Table 3.

Table 3. Correlation coefficients for three distinct models

Model	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>4</sub>	a <sub>5</sub>	a <sub>6</sub>
Model-1	108.5675	0.19849	-	-	-	-
Model-2	95.7194	0.18281	0.10063	-	-	-
Model-3	116.8281	5.7117	1.0228	-0.0733	-0.0085	0.0575

### 5. Principles of modeling by using particle swarm optimization

In this section, the PSO method of Kennedy and Eberhart[31] reviewed and discussed its main properties. Consider a special class of problems with bounded variables in the form of

$$\min(f(x)), \quad s. t. \quad l_k \leq x_k \leq u_k, k = 1, 2, \dots, n.$$

Particle swarm optimizers are optimization algorithms which are inspired of the behaviors of social models like bird flocking or fish schooling. It is introduced by Kennedy and Eberhart [31] as an optimization method for continuous nonlinear functions. Later, it has been applied to wide range of problems due to its simple concepts and easy implementation.

PSO is a stochastic and population-based optimization technique where individuals, referred to as particles are grouped into a swarm. Each particle has a position and velocity vector. Particles can initially be positioned randomly or to predetermined locations. The position of a particle is a candidate solution to the optimization problem. In a PSO system, each particle is moved through the multi-dimensional search space, adjusting its position in search space according to its own experience and that of neighboring particles. A particle therefore makes use of its neighbors to position itself toward an optimal solution. The effect is that particles move toward a minimum, while still searching a wide area around the best solution. The performance of each particle (i.e. the ‘‘closeness’’ of a particle to the global optimum) is measured using a predefined fitness function which encapsulates the characteristics of the optimization problem. Each particle in the swarm maintains the current position, current velocity and the personal best position. The personal best position associated with a particle *i* is the best position that the particle has visited, i.e. a position that yielded the highest fitness value for that particle. If *f* denotes the objective function then the personal best of a particle at a time step, *t* is updated as follows:

$$y_i(t + 1) = \begin{cases} y_i(t), & f(x_i(t + 1)) \geq f(y_i(t)) \\ x_i(t + 1), & f(x_i(t + 1)) < f(y_i(t)) \end{cases} \quad (4)$$

Two main approaches to PSO exist, namely lbest and gbest, where the difference is in the neighborhood topology used to exchange experience among particles. For the gbest model, the best particle is determined from the entire swarm. If the position of the best particle is denoted by the vector  $\hat{y}$ , then

$$\hat{y}(t) \in \{y_0, y_1, \dots, y_s\} = \min\{f(y_0(t)), f(y_1(t)), \dots, f(y_s(t))\} \quad (5)$$

wheres is the total number of particles in the swarm. For the lbest model, a swarm is divided into overlapping neighborhoods of particles. For each neighborhood,  $N_j$ , a best particle is determined with position  $\hat{y}_j$ . This best particle is referred to as the neighborhood best particle defined as

$$N_j = \{y_{i-l}(t), y_{i-l+1}(t), \dots, y_{i-1}(t), y_i(t), y_{i+1}(t), \dots, y_{i+l}(t), y_{i+l+1}(t)\} \quad (6)$$

$$\hat{y}_j(t + 1) \in N_j | f(\hat{y}_j(t + 1)) = \min\{f(y_i)\} \forall y_i \in N_j$$

where  $j$  stands for the number of neighborhood in the entire swarm. Neighborhoods are usually determined using particle indices [32], although topological neighborhoods have also been used [33]. The gbest PSO is a special case of lbest while  $l=s$ , that is, the neighborhood is the entire swarm. In the gbest PSO algorithm, the velocity,  $v_i$  and the position,  $x_i$  of a particle in the swarm are updated at the end of each iteration as follows:

$$v_i(t+1) = v_i(t) + c_1 r_1 [y(t) - x_i(t)] + c_2 r_2 [\hat{y}(t) - x_i(t)] \quad (7a)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (7b)$$

where  $r_1$  and  $r_2$  are uniformly distributed random numbers [31],  $c_1$  and  $c_2$  are the learning factors (weights) that controls the personal and global best respectively. Therefore, the second and the third terms in the right site of equation (7a) show personal and global (social) influences respectively. For the lbest version of PSO,  $\hat{y}(t)$  is replaced by, i.e. the corresponding neighborhood best particle. The PSO algorithm performs repeated applications of the updated equations above until a specified number of iterations have been exceeded or until velocity updates are close to zero. Figure 4 shows the original canonical PSO algorithm.

## 6. Improved PSO algorithms

Since the PSO method introduced by Kennedy and Eberhart [31], many techniques were proposed to refine and/or complement the standard elementary structure of the PSO algorithm, namely regarding its tuning parameters. A number of modifications to PSO have been introduced in the literature. The first modification was done by Shi and Eberhart [34]. They proposed a constant inertia weight,  $w$  which controls how much a particle tends to follow its direction as compared to the memorized personal best position (pbest),  $y_i(t)$  and the global best position (gbest),  $\hat{y}(t)$ . This version is referred to as PSO with constant inertia or PSO-CI. Hence, in PSO-CI, velocity update is given as

$$v_i(t+1) = w v_i(t) + c_1 r_1 [y(t) - x_i(t)] + c_2 r_2 [\hat{y}(t) - x_i(t)] \quad (8)$$

where the values of  $r_1$  and  $r_2$  are realized component-wise. Shi and Eberhart [35] also proposed a linearly varying inertia weight during the searching process. The inertia weight is linearly reduced during the search. This entails a more globally search during the initial stages and a more locally search during the final stages. This version is referred to as PSO with linear inertia or PSO-LI. Also, a linearly varying scaling factors,  $(c_1, c_2)$  of PSO during the search has been proposed. These factors are linearly reduced during the search. This version is referred to as PSO with linear acceleration weights or PSO-LA. Clerc and Kennedy [36] introduced another interesting modification to PSO in the form of a constriction coefficient  $K$ , which controls all the three components in velocity update rule (7a). This has an effect of reducing the velocity as the search progresses. In this modification, the velocity update is given as

$$v_i(t+1) = K (v_i(t) + c_1 r_1 [y(t) - x_i(t)] + c_2 r_2 [\hat{y}(t) - x_i(t)]) \quad (9)$$

where  $v_i$  is the velocity vector of  $i$ th person of population in time,  $t$  and  $K$  is given as

$$K = 2 / |2 - \varphi - \sqrt{\varphi^2 - 4\varphi}|, \quad \varphi = c_1 + c_2 > 4 \quad (10)$$

This version is referred to as PSO with constriction or CPSO. The structure of CPSO algorithm is similar to the structure of PSO algorithm (Figure 4) and the only difference is the

updating rule of velocity vector which is given in the Equation (9). Da and Xiurun [37] also modified PSO by introducing a temperature like control parameter as in the simulated annealing (SA) algorithm. This version of PSO probabilistically select the overall best,  $\hat{y}(t)$ , in the updating of  $v_i(t + 1)$  for each  $i$ . Unlike SA, the temperature increases from a low to a high value. This ascertains that PSO randomizes  $\hat{y}(t)$  in population for the low temperature. At the high temperature,  $\hat{y}(t)$  is selected deterministically as in the original PSO.

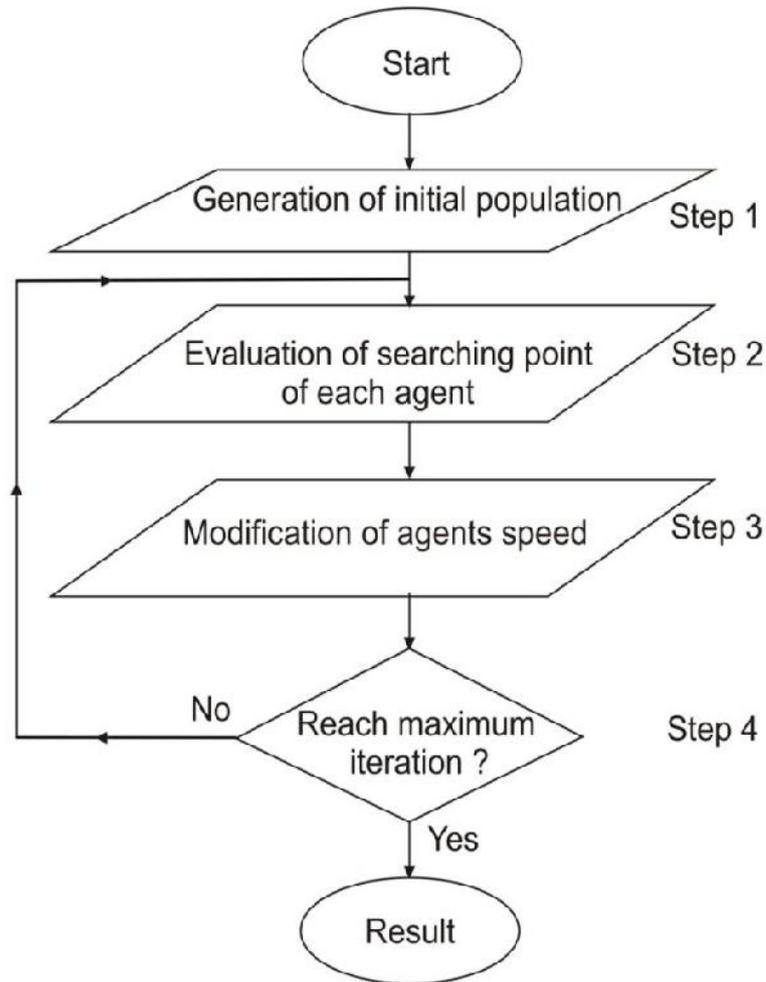


Figure 4. The standard elementary algorithm of PSO

### 7. Implementation of PSO algorithms

Four modifications of PSO used to compare the ability of these algorithms on solving the problem and choose the best model which best fit the data. In fact, the choice of swarm size depends on the size of observed data and the number of parameters. However in practice it is a common sense to fix its value 100, therefore in this study, in all PSO algorithms, the swarm size is fixed to 100. and the iteration process was continued until the total error fall below a specific value. For the boundaries of PSO parameters namely, weights and scaling factors are given as  $[w_{max}, w_{min}] = [1.2, 0.1]$  and  $[c_1, c_2] = [1.4, 0.4]$  respectively. All PSO algorithms have been tested using MATLAB7.8 on Intel Core2 Duo CPU 2 GHz with 2 GB of RAM and machine precision is  $2.22E-16$ . The fitness with CPU time for all models with algorithms is reported in Table 4.

Table 4. Comparison of CPU time, number of iterations and fitness of three models with PSO algorithms

	Model-1			Model-2			Model-3		
	fitness	CPU	iter	fitness	CPU	iter	fitness	CPU	iter
PSO	783.6821	0.525478	40	663.9724	0.968827	40	663.8156	0.245	70
PSO-LI	783.6798	0.388604	30	664.8387	1.185111	40	670.1749	0.59797	200
PSO-LA	788.6813	0.666018	30	1033.4795	0.707959	30	693.0513	0.63632	200
PSO-CI	783.6951	0.319481	23	676.4538	0.593899	24	672.2121	0.62314	200

Results show that PSO-CI scheme is almost better than the other three modifications of PSO algorithm in CPU time as well as the number of iterations. The minimum fitness value is obtained by PSO method after 70 iterations for Model-3. On the other hand, according to only fitness value, the standard PSO algorithm is better than the others for these models. In Table 5, computed parameters for all models with algorithms are reported. The comparison of predicted and measured data for all models is given in Figure 5. From the Figure 5b and Figure 5c, Model-2 and Model-3 show nearly similar presentation within the proposed models.

Table 5. Parameters of fitting three models with PSO algorithm

	Model-1		Model-2			Model-3					
	a1	a2	a1	a2	a3	a1	a2	a3	a4	a5	a6
PSO	106.5823	0.215205	88.0899	0.2019	0.1235	96.8403	9.9063	1.3515	-0.2175	-0.0010	-0.0062
PSO-LI	106.3793	0.215912	86.6807	0.1943	0.1373	119.3423	5.2044	1.0719	-0.0538	-0.0107	0.0608
PSO-LA	119.0149	0.173308	30.4111	0.4694	0.2054	117.3537	5.4819	1.0563	-0.0641	-0.0103	0.0618
PSO-CI	105.9853	0.217857	76.7284	0.2107	0.1732	116.3859	6.3021	0.9825	-0.0625	-0.0169	0.0369

## 8. Model validation

Model validation is possibly the most important step in model building sequence. There are many tools for model validation, but the primary tool for most process modeling applications is graphical residual analysis. Different types of plots of the residuals from a fitted model provide information on the adequacy of different aspects of the model. Numerical methods for model validation, such as the R-square measure are also useful but usually to a lesser degree than graphical methods. Graphical methods readily illustrate a broad range of complex aspects of the relationship between the model and the data. Numerical methods for model validation tend to be narrowly focused on a particular aspect of the relationship between the model and the data and often try to compress that information into a single descriptive number or test result. Numerical methods do play an important role as confirmatory methods for graphical techniques. There are also a few modeling situations in which graphical methods cannot easily be used. In these cases, numerical methods provide a fallback position for model validation. One common situation when numerical validation methods take precedence over graphical methods is when the number of parameters being estimated is relatively close to the size of the data set. In this situation residual plots are often difficult to interpret due to constraints on the residuals imposed by the estimation of the

unknown parameters. One area in which this typically happens is in optimization applications using designed experiments[38].

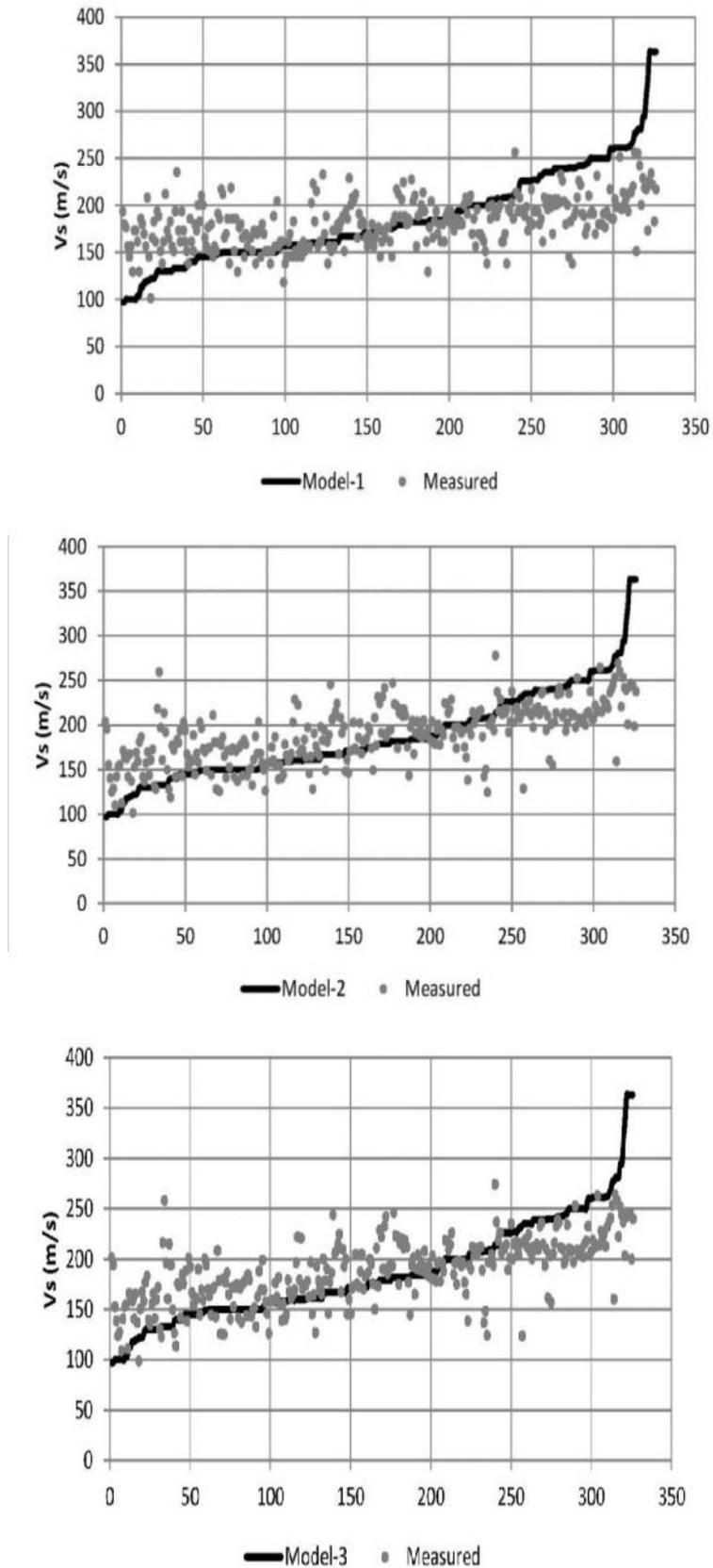


Figure 5. The fitness of the models, a) Model-1, b) Model-2, c) Model-3

Goodness of fit of a model describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question. Before the comparison of the developed models, brief information of basic formulas of error measures given. Some goodness of fit measures for parametric models are following [39,40]:

- *Sum of squares of error (SSE)*
- *R-square*
- *Degrees of freedom adjusted R-square*
- *Chi-square*
- *Reduced Chi-square*
- *Weighting methods*

### ***Sum of Squares of Error***

This measure measures the total deviation of the response values from the fit to the response values. It is also called the summed square of residuals and is usually labeled as SSE and defined as

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (11)$$

where  $y_i$  and  $\hat{y}_i$  denote measured and estimated values respectively. In this measure, a value closer to zero indicates that the model has a smaller random error component and that the fit will be more useful for prediction.

### ***R-Square***

This statistic measures how successful the fit is in explaining the variation of the data. Put another way, R-square is the square of the correlation between the measured and predicted values. It is also called the square of the multiple correlation coefficients and the coefficient of multiple determinations. R-square is defined as the ratio of the sum of squares of the regression (SSR) and the total sum of squares (SST). SSR is defined as

$$SSR = \sum_{i=1}^n (\bar{y} - \hat{y}_i)^2 \quad (12)$$

SST is also called the sum of squares about the mean, and is defined as

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (13)$$

where,  $SST = SSR + SSE$ . Given the last two definitions for SSR and SST, R-square is expressed as:

$$R - Square = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (14)$$

R-square can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of variance is accounted for by the model. When the number of coefficients in model is increased, the R-square increases although the fit may not improve in

a practical sense. To avoid this situation, the degrees of freedom adjusted R-square statistic which described below is needed.

### ***Degrees of freedom adjusted R-Square***

This measure uses the R-square measure defined above and adjusts it based on the residual degrees of freedom. The residual degrees of freedom is defined as the number of response values  $n$  minus the number of fitted coefficients  $m$  estimated from the response values.

$$v = n - m$$

where,  $v$  indicates the number of independent pieces of information involving the  $n$  data points that are required to calculate the sum of squares.

### ***Chi-square***

One way in which goodness of fit of a model can be measured, in the case where the variance of the measurement error is known, is to construct a weighted sum of squared errors:

$$\chi^2 = \frac{(O-E)^2}{\sigma^2} \quad (15)$$

where  $\sigma^2$  is the known variance of the observation,  $O$  and  $E$  stand for observed and estimated data respectively.

### ***Reduced Chi-square***

The reduced Chi-squared measure is simply the Chi-square divided by the number of degrees of freedom [38, 39, 40]:

$$\chi_{red}^2 = \frac{\chi^2}{v} = \frac{1}{v} \sum \frac{(O-E)^2}{\sigma^2} \quad (16)$$

where  $v$  is the number of degrees of freedom, usually given by  $N - n - 1$ , where  $N$  is the number of observations, and  $n$  is the number of fitted parameters, assuming that the mean value is an additional fitted parameter. The advantage of the reduced Chi-square is that it already normalizes for the number of data points and model complexity. As a rule of thumb, a large  $\chi_{red}^2$  indicates a poor model fit. However  $\chi_{red}^2 < 1$  indicates that the model is over-fitting the data. A  $\chi_{red}^2 > 1$  indicates that the fit has not fully captured the data. In principle, a value of  $\chi_{red}^2 = 1$  indicates that the extent of the match between observations and estimates are in accord with the error variance.

### ***Weighting methods***

Nonlinear regression is commonly done without weighting. If there is a similar trend between experimental and estimation, we can weight points differentially. There are two relative weighting methods:

- *weighting by  $1/y^2$*
- *weighting by  $1/y$*

### Weighting by $1/y^2$

This method minimizes the sum-of-squares of the relative distances of the data from the curve. The method is appropriate when we expect the average distance of the points from the curve to be higher when  $y$  is higher, but the relative distance (distance divided by  $y$ ) to be a constant. In this common situation, minimizing the sum-of-squares is inappropriate because points with high  $y$  values will have a large influence on the sum-of-squares value while points with smaller  $y$  values will have little influence. Minimizing the sum of the square of the relative distances restores equal weighting to all points.

### Weighting by $1/y$

Weighting by  $1/y$  is a compromise between minimizing the actual distance squared and minimizing the relative distance squared. The selection of the weighting methods in general depends on the distribution of data.

Now by applying above measures for error evaluation the results summarized in Table 6. According to the Table 6, results show that the Model-3 is somewhat better than the other two models based on the above fit measures for nonlinear models.

Table 6. Comparison of fit measures for three proposed models

Measure	Model-1	Model-2	Model-3
SSE	623215.6868	471281.9600	415501.2786
R-square	0.2348	0.4213	0.4898
R-square <sub>red</sub>	0.000725	0.001304	0.001535
$\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n x_i^2}$	248.6965	188.0668	165.8073207
$\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n x_i^2}$	0.7699	0.5841	0.519772165
Weighting by $1/y^2$	19.0937	15.6308	13.009615
Weighting by $1/y$	3167.3047	2511.0563	2148.49751

## 9. Comparison of the models

The accuracy of the proposed models, namely (1) to (3) without uncertainties in predicting  $V_s$  is compared with correlations presented previously by Hasancebi and Ulusay [17], Dikmen [19], Jamiolkowski *et al.* [24] and Akin *et al.* [20] (cf. Table 1). The statistical comparison is performed for all the 326 cases initially used for the model development. Figure 6 illustrates the scattering of predicted versus observed  $V_s$  values. The results obtained from the Figure 6 shows that Model-2 is virtually same as Model-3. These two models represent the measured data better. Figure 7, 8 and 9 shows the plots of fitness function that is optimized by PSO algorithms for Model-1, 2 and 3 (horizontal axis shows the iterations of algorithm). (a) illustrates plot of standard PSO algorithm, (b) shows plot of PSO-LI algorithm, (c) shows plot of PSO-CI algorithm and (d) shows plot of PSO-LA algorithm respectively.

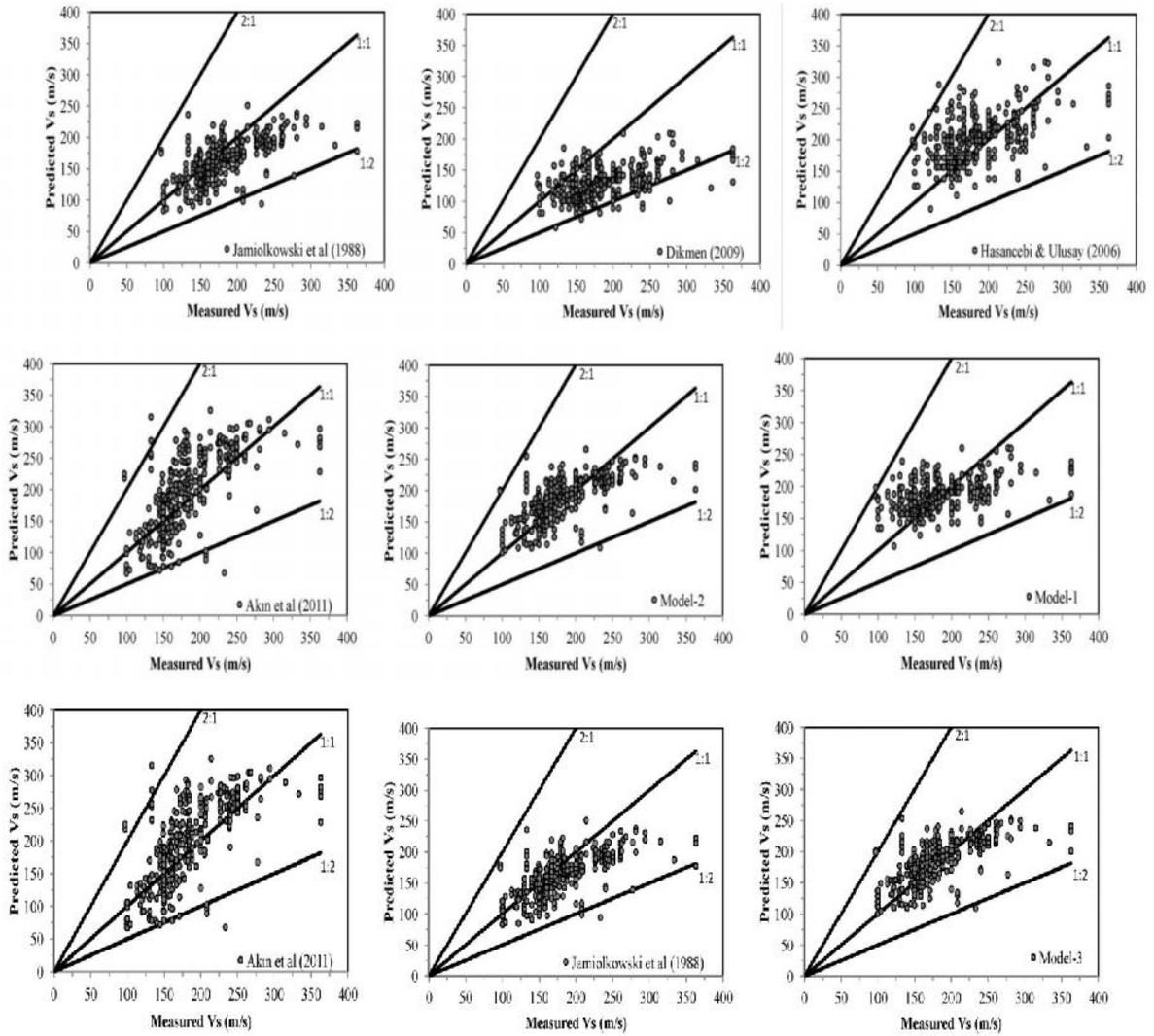


Figure 6. Comparison of measured versus predicted  $V_s$  by different methods, a) Hasancebi and Ulusay (2006), b) Dikmen (2009), c) Jamiolkowski *et al.* (1988), d) Akin *et al.* (2011), e) Model-1, f) Model-2 and g) Model-3

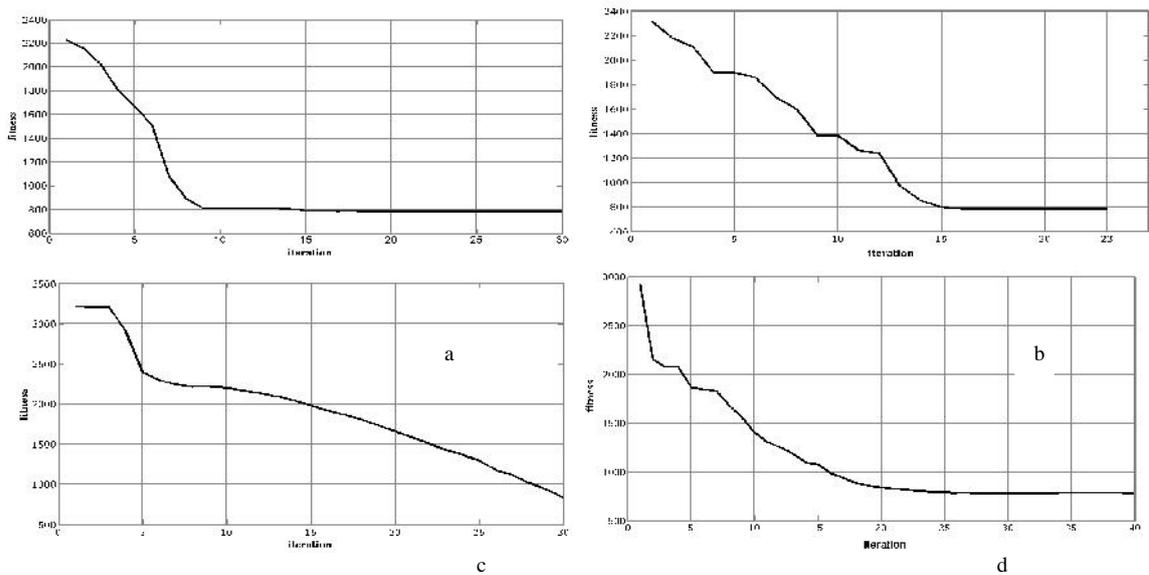


Figure 7. Plots of fitness function optimized by PSO algorithms for Model-1, a) PSO, b) PSO-LI, c) PSO-CI, d) PSO-LA

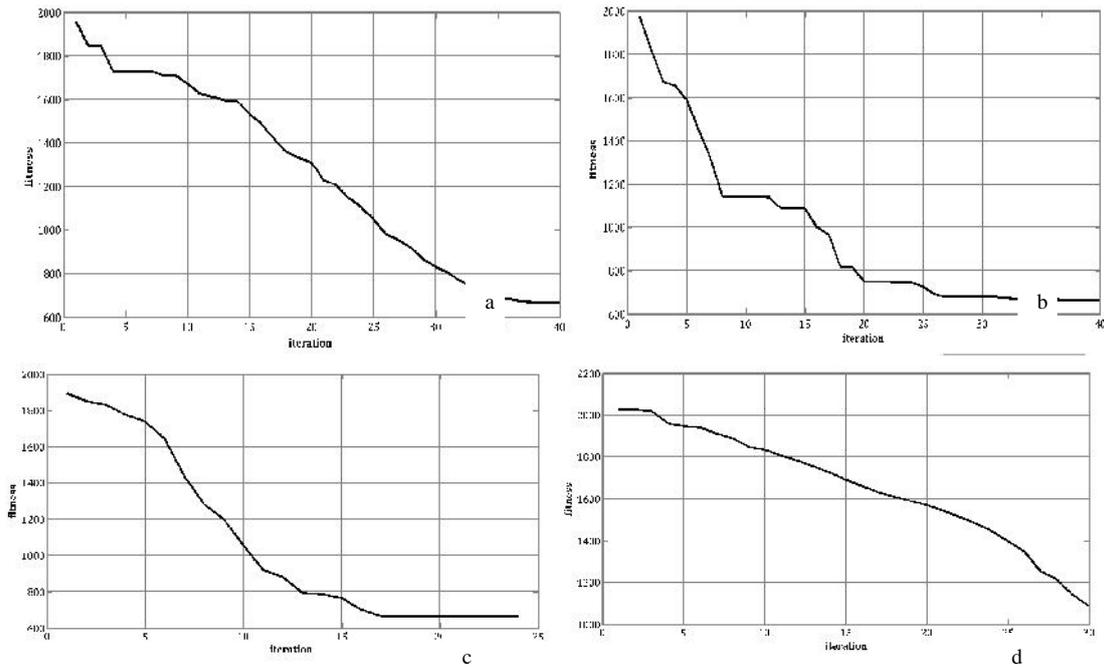


Figure 8. Plots of fitness function optimized by PSO algorithms for Model-2, a) PSO, b) PSO-LI, c) PSO-CI, d) PSO-LA

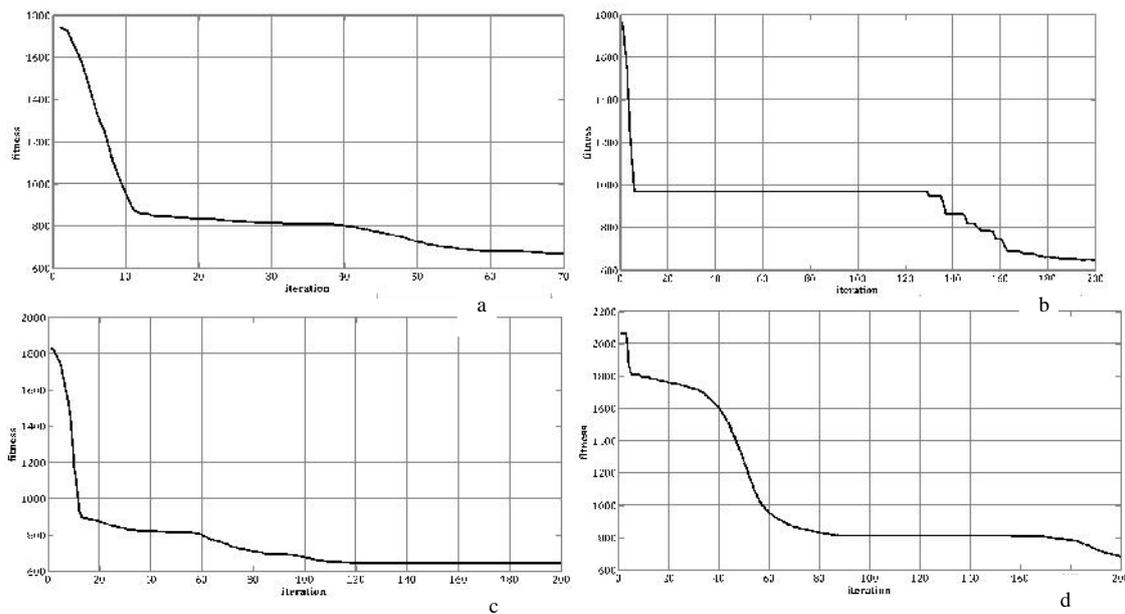


Figure 9. Plots of fitness function optimized by PSO algorithms for Model-3, a) PSO, b) PSO-LI, c) PSO-CI, d) PSO-LA

### 10. Discussion and Conclusions

In this paper, an attempt is made to deploy a system identification technique to develop a comparison over shear wave velocity. Common structure of empirical correlations, in the literature, is represented by Model-1 and Model-2, and a new model including polynomial correlation of  $N_{SPT}$  and  $Z$  (Model -3) is primarily investigated by using least square regression technique and then, various schemes of PSO technique is used to optimize the model parameters. Particle swarm optimization method is utilized for evaluation of the

coefficient of each parameter independently in the previous correlation. This is a major step forward in comparison with previous approaches that just consider the regression analysis. A new model is introduced to express the polynomial correlation content consideration of a database of case histories consisting of 326 dataset from a site in Adapazarı, Turkey.

Statistical comparison of the models shows that the accuracy of Model-3 is particular and nearly same as Model-2. However, in particular the aforementioned parameters, the new Model-3 performs somewhat better and based on this approach, it is proposed to better correlation. The most optimum values for each coefficient of each model was presented and fed into the model for prediction of  $V_S$ .

Results obtained from this study and previous researches reveal that empirical correlations derived from a local dataset should not implemented for different sites with significantly varying features. Therefore, it is strongly emphasis that these proposed relationships should be used with caution in geotechnical engineering and should be checked against measured  $V_S$ .

## References

- [1] S.L. Kramer, *Geotechnical Earthquake Engineering*, Prentice-Hall International Series in Civil Engineering and Engineering Mechanics, New Jersey, 1996.
- [2] R.D. Andrus, P. Piratheepanthan, B.S. Ellis, J. Zhang, C.H. Juang, Comparing liquefaction evaluation methods using penetration-  $V_s$  relationships, *Soil Dynamics and Earthquake Engineering*, Vol. 24 (2004) 713–721.
- [3] Y. Zhang, D. Gallipoli, C.E. Augarde, Simulation-based calibration of geotechnical parameters using parallel hybrid moving boundary particle swarm optimization, *Computers and Geotechnics* Vol. 36 (2009), 604–615.
- [4] M. Khajehzadeh, M.R. Taha, A. El-Shafie, M. Eslami, Modified particle swarm optimization for optimum design of spread footing and retaining wall, *Journal of Zhejiang University-SCIENCE A*, Vol. 12, 6 415–427.
- [5] S. Ohba, I. Toriumi, Dynamic response characteristics of Osaka Plain, *Proceedings of the Annual Meeting*, AIJ, (1970) (in Japanese).
- [6] T. Fujiwara, Estimation of ground movements in actual destructive earthquakes, *Proceedings of the Fourth European Symposium on Earthquake Engineering*, London, (1972) 125–132.
- [7] Y. Ohsaki, R. Iwasaki, On dynamic shear module and Poisson's ratio of soil deposits, *Soil Found.*, Vol. 13 (1973) 61–73.
- [8] T. Imai, H. Fumoto, K. Yokota, The relation of mechanical properties of soil to Pand S-wave velocities in Japan, *Proceedings of 4th Japan Earthquake Engineering Symp*, (1975) 89–96 (in Japanese).
- [9] H.B. Seed, I.M. Idriss, Evaluation of liquefaction potential sand deposits based on observation of performance in previous earthquakes, *ASCE National Convention (MO)* (1981) 481–544.
- [10] T. Imai, K. Tonouchi, Correlation of N-value with S-wave velocity and shear modulus, *Proceedings of the 2nd European Symposium of Penetration Testing*, Amsterdam (1982) 57–72.
- [11] Z. Jinan, Correlation between seismic wave velocity and the number of blow of SPT and depth, *Selected Papers from the Chinese Journal of Geotechnical Engineering*, (1987) 92–100.
- [12] G.A. Athanasopoulos, Empirical correlations  $V_s$  -N SPT for soils of Greece: a comparative study of reliability. In: Akmak, A.S.Ç (Ed.), *Proceedings of 7th International Conference on Soil Dynamics and Earthquake Engineering (Chania, Crete)*. Computational Mechanics, Southampton (1995) 19–36.
- [13] R. Iyisan, Correlations between shear wave velocity and in-situ penetration test results, *Teknik Dergi (Digest)* Vol. 7 (1996), 1187–1199 (in Turkish).
- [14] T. Imai, Y. Yoshimura, *The relation of mechanical properties of soils to P and S-wave velocities for ground in Japan*, Technical Note, OYO Corporation (1975).
- [15] D.E. Sykora, K.H. Stokoe, Correlations of in-situ measurements in sands of shear wave velocity, *Soil Dynamic Earthquake Engineering* Vol. 20 (1983) 125–36.
- [16] M.K. Jafari, A. Shafiee, A. Ramzkhah, Dynamic properties of the fine grained soils in south of Tehran, *J. Seismol. Earthq. Eng.* Vol. 4 (2002) 25–35.
- [17] N. Hasancebi, R. Ulusay, Empirical correlations between shear wave velocity and penetration resistance for ground shaking assessments, *Bulletin of Engineering Geology and the Environment*, Vol. 66(2006), 203–213.
- [18] U.E. Ulugergerli, O. Uyanık, Statistical correlations between seismic wave velocities and SPT blow counts and the relative density of soils, *Journal of Testing and Evaluation* Vol. 35 (2007) 1–5.
- [19] U. Dikmen, Statistical correlation of shear wave velocity and penetration resistant for soils, *Journal of Geophysical Engineering* Vol.6 (2009) 61–72.
- [20] M.K. Akin, S.L. Kramer, T. Topal, Empirical correlations of shear wave velocity ( $V_s$ ) and penetration resistance (SPT-N) for different soils in an earthquake-prone area (Erbaa-Turkey), *Engineering Geology* 119 (2011) 1–17.

- [21] K. Pitilakis, D. Raptakis, K.T. Lontzetidis, T. Vassilikou, D. Jongmans, Geotechnical and geophysical description of Euro-Seistests, using field and laboratory tests, and moderate strong ground motions, *Journal Earthquake Engineering*, Vol. 3 (1999), 381–409.
- [22] H. Kiku, N. Yoshida, S. Yasuda, T. Irisawa, H. Nakazawa, Y. Shimizu, A. Ansal, A. Erkan, In-situ penetration tests and soil profiling in Adapazari, Turkey, *Proc. ICSMGE/TC4 Satellite Conference on Lessons Learned from Recent Strong Earthquakes* (2001) 259–265.
- [23] Y. Yoshida, M. Ikemi, T. Kokusho, Empirical formulas of SPT blow counts for gravelly soils, *Penetration Testing-1988, ISOPT-1* Orlando, FL 2(1988) 381-387.
- [24] M. Jamiolkowski, V. Ghionna, R. Lancellotta, E. Pasqualini, New Correlations of Penetration Tests for Design Practice, *Proc. ISOPT-1, Orlando, FL* 1(1988) 263-296.
- [25] Y. Ohta, N. Goto, Empirical shear wave velocity equations in terms of characteristic soil indexes, *Earthquake Engineering Struct. Dyn.* Vol. 6 (1978) 167–87.
- [26] S.J. Brandenberg, N. Bellana, T. Shantz, Shear wave velocity as function of standard penetration test resistance and vertical effective stress at California bridge sites, *Soil Dynamics and Earthquake Engineering* Vol. 30 (2010) 1026–1035.
- [27] P.W. Mayne, G.K. Martin, Commentary on Marchetti flat dilatometer correlations in soils, *ASTM Geotechnical Testing Journal* Vol. 21 (1998) 222-239.
- [28] P.W. Mayne, Stress-strain-strength-flow parameters from enhanced in-situ tests, *Proceedings, International Conference on In-Situ Measurement of Soil Properties & Case Histories (In-Situ 2001)*, Bali, Indonesia (2001) 47-69.
- [29] PEER. 2000 Documenting Incidents of Ground Failure Resulting from the Aug 17, 1999, Kocaeli, Turkey Earthquake, <http://peer.berkeley.edu/publications/turkey/adapazari/index.html> (last accessed date 08.07.2011).
- [30] A. Hanna, D. Ural, G. Saygili, Neural network model for liquefaction potential in soil deposits using Turkey and Taiwan earthquake data, *Soil Dynamics and Earthquake Engineering*, Vol. 27 (2007) 521–540.
- [31] J. Kennedy, R.Eberhart, Particle swarm optimization:Proceedings of the IEEE Int. *Conference on Neural Networks, IV*, 1995, 1942-1948.
- [32] J. Kennedy, Small worlds and mega-minds: effects of neighborhood Topology on particle swarm performance, *Proceedings of the Congress on Evolutionary Computation*, 1999, 1931–1938.
- [33] P.N. Sugsnthan, Particle Swarm Optimizer with Neighborhood optimizer, *In: Proceedings of the Congress on Evolutionary Computation*, 1999, 1958–1961.
- [34] Y. Shi, R.C. Eberhart, A modified particle swarm optimizer, in: *Proceedings of the IEEE International Conference on Evolutionary Computation*, IEEE Press, Piscataway, NJ, 1998, 69–73.
- [35] Y. Shi, R.C. Eberhart, Parameter selection in particle swarm optimization, in: *Evolutionary programming VII*, in: *Lecture notes in Computer Science*, Vol.1447, Springer, Berlin, 1998, pp. 591–600.
- [36] M. Clerc, J. Kennedy, *The particle swarm-explosion, stability and convergence in a multidimensional complex space*, IEEE Trans. Evol. Comput. 6 (1) (2002) 58–73.
- [37] Y. Da, G. Xiurun, *An improved PSO-based ANN with simulated annealing technique*, Neurocomputing Vol. 63 (2005) 527–533.
- [38] C.Laub,L.K. Tonya, *Chi-Square Data Fitting*, University California, Davis.
- [39] J.R. Taylor, *An introduction to error analysis*, University Science Books, 1997.
- [40] D. M. Glover, W. J. Jenkins, S. C. Doney, *Least Squares and regression techniques, goodness of fit and tests, non-linear least squares techniques*, Woods Hole Oceanographic Institute, 2008.