INTERNATIONAL JOURNAL OF MARITIME TECHNOLOGY
IJMT Vol.2/ Summer 2014 (15-28)
Available online at: http://ijmt.ir/browse.php?a_code=A-10-1-80&sid=1&slc_lang=en

Predicting ε_{50} for Lateral Behavior of Piles in Marine Clay Using an Evolutionary Based Approach

Babak Ebrahimian[1, 2*], Aida Nazari[3]

[1] *School of Civil Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran;*
[2] *The Highest Prestigious Scientific and Professional National Foundation, Iran's National Elites Foundation (INEF), Tehran, Iran; bebrahimian@ut.ac.ir, ebrahimian.babak@gmail.com*
[3] *School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran*

ARTICLE INFO

Article History:
Received: 12 Jan. 2014
Accepted: 1 Jul. 2014
Available online: 22 Sep. 2014

Keywords:
P-Y curve
Laterally loaded pile
Cone penetration test (CPTu)
Marine clay
South Pars field

ABSTRACT

Analyzing piles subjected to lateral loads significantly depends on soil resistance at any point along the pile as a function of pile deflection, known as p-y curve. On the other hand, the deformation characteristics of soil defined as "the soil strain at 50% of maximum deviatoric stress (ε_{50})" has considerable effect on the generated p-y curve. In this research, several models are proposed to predict ε_{50} specifically for designing very long pile foundations of offshore oil and gas platforms in South Pars field; Persian Gulf; Iran. Herein, ε_{50} is evaluated from extensive soil data of marine clays including in-situ and laboratory test results using evolutionary polynomial regression (EPR). It is demonstrated that the normalized cone tip resistance, which is an indication of soil undrained shear strength, leads to more realistic ε_{50} values compared with the laboratory-derived undrained shear strength parameter. Furthermore, the results of full scale lateral pile load tests in different sites are used in order to validate the performance of the proposed models in predicting lateral pile behavior. The results of a numerical study on lateral pile-soil system also show the efficiency of the proposed model in predicting lateral pile response.

1. Introduction

Pile foundations are often required to be designed against significant lateral in addition to vertical loads. These lateral loads can be imposed by wind, earth pressure, wave, tide, current and ship impact, mooring rope, earthquake, vehicle traction and etc. The performance of pile foundations is usually governed by either deflection or bearing capacity. Exceeding the maximum allowable lateral load may cause the failure of soil around the pile, or structural failure of the pile itself. In order to design a pile foundation safely and economically, accurate assessment of its behavior should be made using pile load tests data and/or the well-known analytical or numerical methods. As the full-scale load tests are very expensive and time consuming, analytical and numerical approaches are usually used to evaluate the lateral behavior of pile-soil systems.

The lateral pile-soil interaction behavior is commonly characterized by a series of uncoupled, nonlinear springs applied along the pile, known as p-y curves. Various formulations have been proposed to predict p-y curves in different site conditions (e. g., [1-7]). The American Petroleum Institute (API) method [7] is the widely used method based on Matlock's field research [1].

Pile geometry and soil properties are the key parameters in developing p-y curves. These curves mostly depend on the ultimate horizontal soil reaction (P_u) and the critical lateral displacement (y_c) corresponding to 50% mobilized P_u. y_c is defined as:

$$y_c = 2.5\varepsilon_{50} D \quad (1)$$

where, D is the pile diameter, and ε_{50} is the strain at one-half the maximum deviatoric stress in laboratory undrained compression tests on undisturbed cohesive soil samples. Typical p-y curves for cohesive soils is, shown in Figure 1, illustrate the role of the above mentioned parameters on developing such curves.

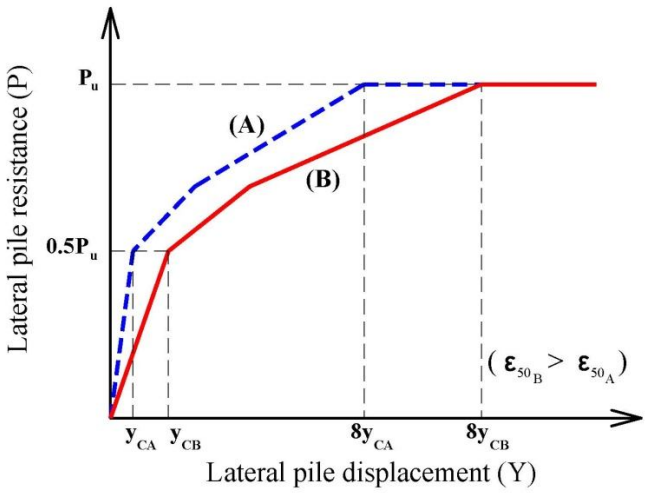

Figure 1. Typical p-y curves for pile in cohesive soil under static loading

Curves A and B in this figure, are schematic p-y curves for a soil with different ε_{50} values. As $\varepsilon_{50B} > \varepsilon_{50A}$, with the same pile geometry we have $y_{cB} > y_{cA}$. As shown in this figure, ε_{50} is an effective factor in generating p-y curves for clays. It is seen that higher ε_{50} values lead to softer soil behavior and higher pile lateral displacements for constant lateral load ratios (P/ P_u). Furthermore, the ultimate lateral load is obtained at higher levels of pile lateral displacements as ε_{50} increases. Hence, lateral stiffness and resistance of pile-soil system are affected by ε_{50}.

Sullivan et al. [8] recommended ε_{50} values for different clayey soils based on the undrained shear strength. However, such proposed ε_{50} values are not consistent with those obtained from experimental measurements conducted in different sites and do not result in accurate p-y curves in most soil conditions [9-10].

Hamilton et al. [10] performed some triaxial compression tests under isotropically consolidated undrained (CIU) and unconsolidated undrained (UU) conditions on Tilbrook Grange clays and measured the ε_{50} values. They realized that ε_{50} values obtained from CIU tests show less scatter than those of UU tests and found a trend line for ε_{50}. It was demonstrated that ε_{50} values, obtained from laboratory tests, were nearly five times greater than the values recommended by Sullivan for sites having similar undrained shear strengths. Afterwards, they compared different p-y curves derived based on laboratory ε_{50} values and those recommended by Sullivan. It was demonstrated that the uncertainty of predicted p-y curves decreases from 65% to 35% if laboratory ε_{50} values are used instead of those recommended by Sullivan. Additionally, they noted that the use of p-y curves based on Matlock method with ε_{50} values from CIU tests leads to a more reliable prediction of the lateral load-displacement response.

Hamilton et al. [10] discussed different methods to develop p-y curves for piles in stiff, overconsolidated clays. They compared measured values of ε_{50} derived from UU tests with those typically assumed from Sullivan recommendations and indicated that a slightly better prediction of load–displacement curves is achieved using measured ε_{50} values instead of those recommended by Sullivan.

Dunnavant [11] performed experimental and analytical investigations to predict the influences of pile and soil characteristics as well as loading conditions on lateral pile-soil interaction in saturated overconsolidated clays. It was shown that the overconsolidation ratio (*OCR*) of soil can affect the reference critical displacement (y_c) in p-y curves. In other words, for overconsolidated clays, the value of y_c would be smaller than those available in the literature. Degradation of p-y curves in overconsolidated clays begins at much smaller deflections than in soft clays. Also, it was recognized that pile stiffness and pile diameter could affect y_c.

Davies [12] and Robertson et al. [6] presented a preliminary semi-empirical method to evaluate p-y curves based on flat dilatometer test (DMT) data. They employed the DMT-based p-y curves to model the behavior of three full-scale lateral pile load tests. They showed that ε_{50} has an increasing trend versus depth in both clays and sands in the considered sites. It was found that the predicted deflections using DMT results agree well with those obtained from pile load tests. In all studied cases, the calculated bending moments from DMT-derived p-y curves were larger than those calculated from the measured pile deflection profiles.

The soil properties such as ε_{50} are very sensitive to soil disturbance due to coring procedure, and using ε_{50} values based on the tests on core samples may finally lead to considerable deviation in predicting the real pile behavior. On the other hand, in-situ testing methods such as flat dilatometer (DMT), pressuremeter (PMT) and cone penetration test (CPT) offer excellent means by which representative soil properties may be obtained [6, 12-13]. Therefore, such in-situ tests, with minimum soil disturbance, can be used for evaluating ε_{50} and developing p-y curves.

Cone penetration test (CPT) is a reliable in-situ test for its continuous sounding capability and good repeatability. It provides valuable geotechnical information in soil. Furthermore, the similarity between CPT penetration process and pile installation has led to its popularity in deep foundation analysis and design. Total cone tip resistance obtained from CPT has strong correlation with soil shear strength [14]; on the other hand, due to direct dependence of ε_{50} on shear strength, total tip resistance of CPT can be employed in evaluating ε_{50}.

Despite significant influence of ε_{50} on determining p-y curves, the prediction methods to evaluate this parameter are very rare in the literature. Therefore, this study investigates the use of CPT data to predict ε_{50} in clayey soils and examines the capability of

predicted ε_{50} values in generating realistic p-y curves for laterally loaded piles in different sites. The present calculations of ε_{50} are based on a comprehensive databank from laboratory and field tests, performed in South Pars field, Persian Gulf, south west of Iran. The field is an extremely strategic offshore area which contains world's largest gas resources. Many gas extraction facilities supported on long pile foundations have been constructed in this important region and a large number of such facilities is still under development. Hence, this research mainly focuses on accurate evaluation of ε_{50} as an influential parameter in analysis and design of piles against lateral loads in this region. In this regard, several statistical models, based on evolutionary polynomial regression (EPR) method, are proposed to evaluate ε_{50} values for clayey soils. The effects of cone tip resistance, undrained shear strength, overburden pressure as well as different index properties of soils such as overconsolidation ratio and plasticity index on ε_{50} are evaluated and discussed. In particular, the effect of undrained shear strength of cohesive soils obtained from field tests on ε_{50} is investigated and compared with the recommended values available in the literature. Finally, the validation of the proposed models is performed for full scale piles tested in two different sites with different soil conditions.

2. Site Description

The survey area, approximately 50×45 km^2, is located in Persian Gulf, Iran, between 27° 27' to 27° 28' Northing and 52° 27' to 52° 44' Easting (Figure 2).

Soil investigation activities comprised of sixteen boreholes; eight boreholes with 110 m depth and the rest with 80 m depth below the seabed. In-situ and laboratory tests were carried out to determine the geotechnical properties of sub-seabed soils. The in-situ tests included cone penetration and torvane. Laboratory tests such as unconsolidated undrained (UU) triaxial compression were performed which resulted in undrained shear strength of soil. The strain at 50% of maximum deviatoric stress (ε_{50}) and strain at failure were also obtained from stress-strain curves in UU tests. Atterberg limits and sieve tests were performed as well. Typical profiles of soil properties are illustrated in Figure 3 for a 110 m borehole within the considered survey area. The sub-seabed soils are generally clay, including very soft clay at top up to approximately 20 m which become stiffer with depth. Also lenses of sandy silt and gypsum are found in several depths.

Figure 2. Location of survey area in South Pars Field, south-west of Iran

3. Evolutionary Polynomial Regression

Evolutionary polynomial regression (EPR) is a useful toolbox developed on modeling methodology based on hybrid regression method by [15] and [16]. It is a symbolic data driven method which is used to create polynomial models to evolutionary compute based on input data [17]. The constitutive modeling of soil [18] and assessment of earthquake-induced soil liquefaction and lateral displacement [19] are some successful examples of using EPR in the field of geotechnical engineering.

EPR method includes two general stages; searching the model structures based on an integer Genetic Algorithm (GA) and evaluating each model parameters, such as numeric constant coefficients considering linear optimization [16]. The general symbolic expression derived from EPR is as follows:

$$\hat{Y} = \sum_{j=1}^{m} F(\mathrm{X}, f(\mathrm{X}), a_j) + a_0 \qquad (2)$$

where, $\hat{Y}$ is the estimated outputs of the system derived from EPR; F is the function constructed by the program; X is the input variables matrix; f is a user defined function; a_j is an adjustable parameter determined by the program; and m is the number of terms of the expression defined by user excluding bias a_0 if any.

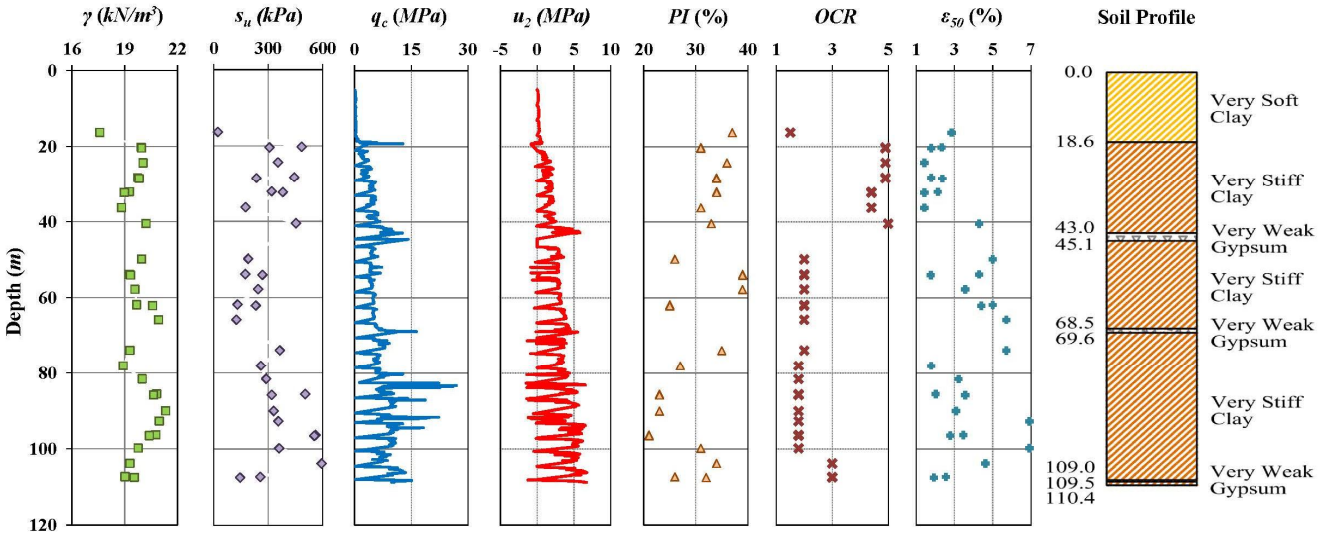

Figure 3. Soil profile and field and laboratory results for a typical 110 m borehole within survey area

It is noteworthy to mention that data-driven techniques like Genetic Programming (GP) and Artificial Neural Network (ANN) tend to reconstruct function *F* from input–output data. GP generates a population of expressions for *F*, coded in tree structures of variable size, and performs a global search of the best fit expression for *F*. ANN goal, on the other hand, is to map *F* rather than to find a feasible structure for it. However, both techniques have their own drawbacks. GP tends to search for mathematical expressions of *F* using an evolutionary approach, but the parameter values are generated as non-adjustable constants, referred to as ephemeral random constants. Therefore, the constants do not necessarily represent optimal values as in numerical regression methods and good structures of *F* can be missed in the process. Furthermore, the number of terms in GP-based expressions can greatly exceed, and the evolutionary search within GP can be quite slow.

Neural networks (NNs) suffer from some shortcomings as well. One of the drawbacks of a neural network is that the optimum structure of ANN (e.g., number of hidden layers, number of neurons, and transfer functions) should be identified a priori which is usually obtained using a time consuming trial and error procedure [16]. The other major shortcoming is related to the black box nature of an ANN model and the fact that the relationship between input and output parameters of the system is described in terms of a weight matrix and biases that are not easily accessible to users understanding. In fact, the black box nature and lack of interpretability have prevented ANNs from achieving their full potential in engineering applications. In other words, NN models give no information on the way the inputs affect the output and therefore are considered as a black box class of models. The lack of interpretability of NN models has stopped them from achieving their full potential in real world problems [18,20].

EPR is a new data mining technique that overcomes the shortcomings of ANNs and GPs. In the context of modeling classification, EPR is classified as a symbolic grey box technique, whose aim is identification and construction of clearly structured model expressions from observed data [18,21]. The approach integrates numerical and symbolic regression to perform evolutionary polynomial regression. The strategy uses polynomial structures to take advantage of their favorable mathematical properties. Indeed, the main idea behind the EPR is to use evolutionary search for exponents of polynomial expressions by means of a genetic algorithm (GA) engine while the parameters of the function are determined using the least square method. This allows (*i*) easy computational implementation of the algorithm, (*ii*) efficient search for an explicit expression (formula) and (*iii*) improved control of the complexity of the expression generated [22]. To avoid the problem of mathematical expressions growing rapidly in length with time associated with GP, in EPR the evolutionary procedure is conducted in the way that it searches for the exponents of a polynomial function with a fixed maximum number of terms, rather than performing a general evolutionary search as used in normal GP. Furthermore, during one execution it returns a number of expressions with increasing numbers of terms up to a limit set by the user, to allow the optimum number of terms to be selected. In addition, EPR overcomes the shortcomings of ANNs by providing a structured and transparent model representing the behavior of the system. EPR takes advantage of automatic model construction procedure that avoids the need to pre-select the functional form and the number of parameters in the model [23].

The general process in EPR can be rewritten based on vector form as:

$$\mathrm{Y}_{N\times 1}(\theta,\mathrm{Z}) = \left[\mathbf{I}_{N\times 1} \quad \mathrm{Z}^{j}_{N\times n}\right]\times\left[a_0 \quad a_1 \quad \ldots \quad a_n\right]^T = \mathrm{Z}_{N\times d}\times\theta^{T}_{d\times 1} \tag{3}$$

where, $\mathrm{Y}_{N\times 1}(\theta,\mathrm{Z})$ is the least squares estimate vector of the N target values; $\theta_{d\times 1}$ is the vector of d=m+1 parameters a_j and a_0 (θ^T is the transposed vector); $\mathrm{Z}_{N\times d}$ is a matrix formed by **I**, unitary vector for bias a_0, and m vectors of variables Z^j that for fixed j are a product of the independent predictor vectors of inputs, X=‹X_1 X_2… X_k›.

EPR performs evolutionary search of a model space using an analogy with stepwise regression [24] rather than by means of the traditional symbolic regression search based on parse tree structures. In this way, EPR performs a global search of input exponents and combination of input variables according to the user-defined cost function [16]. The adjustable parameters, a_j , can now be evaluated by means of the linear least squares (LS) method based on minimization of the sum of squared errors (SSE) as the cost function. The SSE function which is used to guide the search process toward the best-fit model is as follows:

$$SSE = \frac{\sum_{i=1}^{N}\left(y_a - y_p\right)^2}{N} \tag{4}$$

where, y_a are the target values in the training dataset and y_p are the model predictions.

The program search is based on pseudo-polynomial and true structures using a single and multi-objective genetic algorithm, with different general expression forms. The expression form considered in this research is defined as bellow:

$$Y = a_0 + \sum_{j=1}^{m}\left(\begin{array}{l} a_j.(X_1)^{ES(j,1)}.\cdots.(X_k)^{ES(j,k)}. \\ f\left((X_1)^{ES(j,k+1)}\right).\cdots.f\left((X_k)^{ES(j,2k)}\right)\end{array}\right) \tag{5}$$

In the above expression, X_i is the k candidate inputs vectors; a_j are constant values; ES is the matrix of unknown exponents which can be edited by user within the defined range of values; and m is the length of expressions defined by user, which represents the number of maximum terms in each set of results. Each monomial of polynomial models can contain user defined functions. For this purpose, $f(\,)$ is the function that can be selected by user based on available functions in the program. These functions may be logarithmic, exponential, tangent hyperbolic and etc. The inclusion of exponential functions in the general expression of Equations (2) allows EPR to explore a large space of formulae where the analyst's understanding of the physical process warrants their inclusion. However, if such functions are not naturally describing the phenomenon being modeled EPR search would find exponent values for such inputs to be equal to zero.

In order to determine all models corresponding to the optimal trade-off between the fitness and brevity of the model, EPR performs a multi-objective search exhibiting various mathematical models representing best fitness for possible models. For particular purpose, one can choose best models based on short gap reconstruction, gaining physical insight or forecasting the phenomenon. The fitness model defined in EPR is the Coefficient of Determination (CoD) which refers to how closely the regression expression fits the data points:

$$CoD = 1 - \frac{\sum_{n}(p-m)^2}{\sum_{n}(m-\overline{m})^2} \tag{6}$$

where, p is the predicted values by model derived from EPR; m is the measured values; m-bar is the average of measured values; and n is the number of data points. More details about EPR architecture for model representation as well as the method employed for parameter estimation can be found in [16].

4. Results and Discussion

The field and laboratory test results including 274 data series are considered as the databank for the numerical regression. In the present study, five variables are identified as primary input data of cohesive soils for evaluating ε_{50} as an output. The input data includes undrained shear strength (s_u), normalized cone tip resistance (q_c), total overburden pressure (σ_0), plasticity index (PI) and overconsolidation ratio (OCR). The input parameters variables may affect ε_{50} values have been selected based on a comprehensive literature survey as well as the principles of soil stress-strain behavior. The potentials and combinations of mentioned soil engineering parameters have been studied to generate EPR-based models with proper physical meanings.

In pattern recognition procedures, it is common practice to divide the available data into two subsets; training and testing. The model is firstly developed using the former and then tested using the latter one to ensure that the final obtained model has the ability to properly estimate ε_{50} for unseen or untrained cases. Here, the entire databank is divided into several random combinations of training and testing sets until a robust representation of the whole population, in terms of statistical properties, is achieved for both training and testing sets. The statistical properties of the parameters considered in this study including the values of maximum, minimum, mean, and standard deviation are presented in Table 1 for training, testing and all datasets. Training dataset includes 80% of all data (219) and the rest (55) are used as testing dataset. The statistical values of training, testing and all dataset, shown in Table 1, are close to each other.

Multiple runs are performed and the analyses are repeated with various combinations of different number of generations, different functions, different multi-objective optimization strategies and different number of terms in order to obtain the most suitable form for the model. The solutions are analyzed on the basis of the simplest generated model for each case. After analysis of different alternative models four relationships are developed for evaluating ε_{50}, which are presented in Table 2. It is noted that these models are not the only ones returned by EPR. Many other models can be obtained concerning their general expression forms, number of generations, terms and combinations of input parameters. Herein, the best models have been selected, based on the authors past experiences on the subject, to propose more sensible and practical equations with the sufficient physical meanings within the classical soil mechanics. To examine the robustness and assess the performance of EPR models, the following three statistical criteria have been used:

- Coefficient of determination (R^2), is a measure used to determine the relative correlation between two sets of variables, and defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(m_i - p_i)^2}{\sum_{i=1}^{n}(m_i - \overline{m})^2} \tag{7}$$

- Root mean square error (*RMSE*), is a measure of error, defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(m_i - p_i)^2}{n}} \tag{8}$$

The advantage of this criterion is that large errors receive greater attention than smaller ones.

- Mean absolute error (*MAE*), is another measure of error which eliminates the emphasis given to large errors, presented as:

$$MAE = \frac{\sum_{i=1}^{n}|m_i - p_i|}{n} \tag{9}$$

In the above relations, m_i and p_i are the i^{th} measured and predicted values of output parameter (ε_{50}), respectively; n is the number of data points; and m-bar indicates the average of measured output.

The suggested models to evaluate ε_{50} as well as the values of statistical criteria are presented in Table 2. It is seen that the performance of models improves from model 1 to 4 since R^2 value increases while *RMSE* and *MAE* values decrease. Based on the results summarized in Table 2, the EPR model 4 is chosen as the most appropriate one which is developed using four input parameters: q_c, σ_0, *PI*, *OCR*.

The first relationship is developed between undrained shear strength of soil and ε_{50} and the second one uses the normalized cone tip resistance (q_c) to predict ε_{50}, as shown in Table 2. By comparing the statistical characteristics of models 1 and 2, it can be found that the ε_{50} values predicted from field-based resistance property (q_c) are more accurate than those predicted from the laboratory-based resistance (s_u). By using q_c instead of s_u, R^2 increases from 6.6 for model 1 to 20.8 for model 2. However, R^2 value is not yet acceptable enough, and it seems that other influential parameters should be included in the model development process. Therefore, after several try and error procedures, it was found that the index properties of soil, e. g., *OCR* and *PI* have strong effects on the predicted ε_{50} values. According to Table 1, it is realized that model 3, which includes the above mentioned factors, predict ε_{50} more accurately than model 2. Furthermore, model 4 shows that the overburden pressure has also a notable positive influence on prediction accuracy.

Table 1. Statistical characteristics of databank

Subsets	Statistical characteristics	σ_0 (*kPa*)	s_u (*kPa*)	q_c (*kPa*)	*PI* (%)	*OCR*	Measured ε_{50} (%)
Testing data (55 data)	Minimum	216	19	162	14	0.9	0.9
	Maximum	1933	504	8767	40	4	9.2
	Mean	1081	241	4155	29	2.2	3.9
	Standard deviation	462	112	2190	6.8	0.74	2.1
Training data (219 data)	Minimum	217	19	139	12	0.9	0.7
	Maximum	2207	634	8943	47	5.3	9.3
	Mean	1077	274	4184	30	2.4	3.5
	Standard deviation	515	129	1996	7.4	1.1	2.0
All data (274 data)	Minimum	216	19	139	12	0.9	0.7
	Maximum	2207	634	8943	47	5.3	9.3
	Mean	1078	268	4178	30	2.4	3.6
	Standard deviation	505	126	2037	7.3	1.0	2.1

Table 2. Proposed models for estimating ε_{50}

No.	Equation	Involved parameters	R^2 (%)	*RMSE*	*MAE*
Model 1	$\varepsilon_{50} = -0.79 + 1.5 s_u^{0.2}$	s_u	6.6	1.99	1.65
Model 2		q_c	20.8	1.84	1.52
Model 3	$\varepsilon_{50} = 4.84 - 8.76\times10^{-2} q_c^{0.3} PI^{0.5} OCR^{-0.1} + 5.43\times10^{-6} q_c^{1.4} PI^{0.7} OCR^{-0.1} - 2.1\times10^{-3} q_c^{0.5} PI^{0.8} OCR^{0.7} - 1.24\times10^{-12} q_c^{3.3} OCR^{-0.3}$	q_c, *PI*, *OCR*	36.7	1.64	1.34
Model 4	$\varepsilon_{50} = -2.7\times10^{-13} \sigma^{1.5} q_c^{2.6} PI^{-1.3} OCR^{-0.2} - 1.8\times10^{-10} \sigma^{0.6} q_c^{1.3} PI^{1.6} OCR^{2} + 1.5\times10^{-6} \sigma^{1.5} q_c^{0.4} PI^{0.1} OCR^{0.6} + 1.55$	q_c, σ_0, *PI*, *OCR*	64.8	1.22	1.02

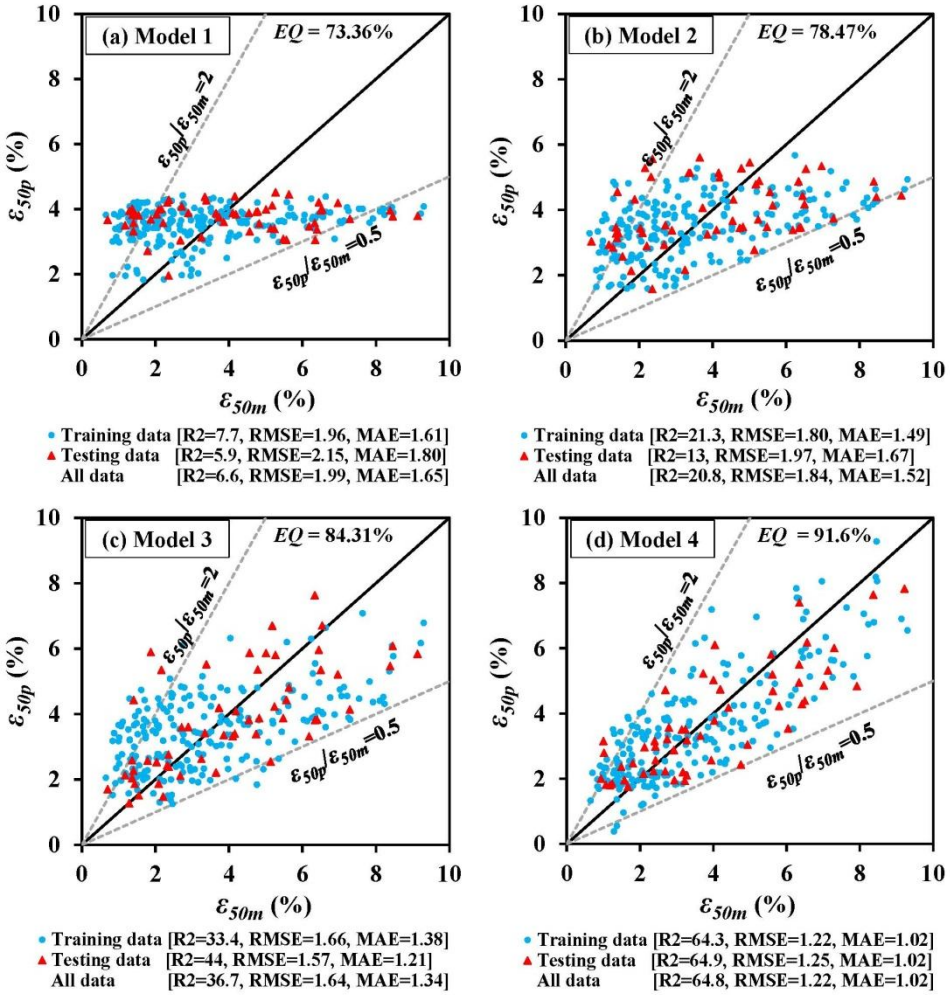

Figure 4. Predicted versus measured ε_{50} values for proposed EPR-based models

Figure 4 illustrates the predicting capability of models by plotting the predicted ε_{50} values against their corresponding measured values in training and testing datasets and their statistical characteristics are shown for quantitative comparison. Considering the data scatter in the graphs of Figure 4, the results of models for testing dataset are generally consistent with those for training dataset. The more the points are distributed uniformly around the ideal 45° line, and the less scattering around this line, the better the capability of the model in predicting ε_{50}. In this regard, it is clear that model 4 behaves better than the other ones. The upper and lower lines in Figure 4 show the boundaries for a zone that is characterized by the ratios of predicted to measured ε_{50} between 0.5 and 2.0.

The estimation quality of each model, defined as the number of the points that fall inside these two

boundaries as percent of the total points is shown in the figure. As the performance of models improves, the data show more concentration in the mentioned zone. While all models show acceptable estimation qualities, the estimation quality for model 4 has the highest value of 91.6% among the proposed models.

It is seen in Fig 4 that the predicted ε_{50} values from model 1, which was developed merely from undrained shear strength (s_u), are not well-distributed along the diagonal line and are concentrated in a narrow horizontal band. However, implementing the normalized cone tip resistance (q_c), instead of s_u, in model 2 has smoothed the above mentioned shortcoming.

The log normal distribution, used by [25], is an appropriate statistical criterion to further evaluate the performance of the proposed models. In this regard, the natural logarithm of the ratio of predicted to measured ε_{50}, ($\ln(\varepsilon_{50p}/\varepsilon_{50m})$), is calculated for each data point and then the mean and standard deviation of these values are determined as follows:

$$\mu_{\ln(\varepsilon_{50p}/\varepsilon_{50m})} = \frac{1}{n}\sum_{i=1}^{n}\ln(\varepsilon_{50p}/\varepsilon_{50m})_i \tag{10}$$

$$\sigma_{\ln(\varepsilon_{50p}/\varepsilon_{50m})} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\ln(\varepsilon_{50p}/\varepsilon_{50m})_i - \mu_{\ln(\varepsilon_{50p}/\varepsilon_{50m})}\right)^2} \tag{11}$$

where, the subscripts p and m denote "predicted" and "measured", respectively; n is the number of data considered in the analysis; $\mu_{\ln}$ and $\sigma_{\ln}$ are indicators for accuracy and precision of the models, respectively, which are used to identify the log normal distribution of the density function as:

$$f(\varepsilon_{50p}/\varepsilon_{50m}) = \frac{1}{\sqrt{2\pi}\sigma_{\ln}(\varepsilon_{50p}/\varepsilon_{50m})}\cdot \exp\left[-\frac{1}{2}\left(\frac{\ln(\varepsilon_{50p}/\varepsilon_{50m}) - \mu_{\ln(\varepsilon_{50p}/\varepsilon_{50m})}}{\sigma_{\ln(\varepsilon_{50p}/\varepsilon_{50m})}}\right)^2\right] \tag{12}$$

The better distribution is achieved when $\mu_{\ln(\varepsilon^{50p}/\varepsilon^{50m})}$ and $\sigma_{\ln(\varepsilon^{50p}/\varepsilon^{50m})}$ approach unity and zero, respectively. The log normal distributions of $\varepsilon_{50p}/\varepsilon_{50m}$ for the proposed models are presented in Figure 5.

The probability of predicting ε_{50} with 0 to 90% accuracy (10-100% absolute error) is calculated from Figure 5 and shown in Figure 6. Total area below each curve in Figure 5 is equal to one. Therefore, at a specified absolute error level, the probability of predicting ε_{50} is derived by calculating the total area below the log normal distribution curve within the accuracy limits. At a constant absolute error, a higher probability indicates the better ability of model in predicting ε_{50}. Based on this definition, the performance of the models improves from model 1 to 4 at all levels of absolute error.

The ability of different models to predict ε_{50} can be evaluated using cumulative probability, as used by [26]. They used the concept of cumulative probability as a criterion to evaluate the bias of their model. The cumulative probability for each $\varepsilon_{50p}/\varepsilon_{50m}$ can be obtained by the following definition:

$$CP_i = \frac{i}{n+1} \tag{13}$$

where, i is the data number, arranged in an ascending order. The cumulative probability versus the ratio $\varepsilon_{50p}/\varepsilon_{50m}$ for the proposed models is depicted in Figure 7. In order to assess the ability of each model in estimating ε_{50}, the 50% and 90% cumulative probabilities (CP50% and CP90%) are calculated. The difference between CP90 and CP50 (CP90%-CP50%) represents the discrepancy from accurate estimation. Ideally, if all data are predicted with no bias, the distribution of estimated to measured ε_{50} against CP will be a straight line with value of unity, indicating an exact estimation. In reality, the better performance of the model is achieved when $\varepsilon_{50p}/\varepsilon_{50m}$ is closer to unity at CP50%. Lower (CP90%-CP50%) for each model indicates the better prediction accuracy of the proposed model.

According to this criterion, it is observed in Figure 7 that model 4 leads to the optimum value of CP50% equal to unity and lower value of (CP90%- CP50%) compared with the other models.

In statistical analysis, a model would behave better, if residual values, i.e. the difference between the measured and predicted values of ε_{50}, are concentrated more uniformly around the mean value of residuals. The mean value of residuals is calculated by:

$$MR = \frac{1}{n}\sum_{i=1}^{n}(\varepsilon_{50m} - \varepsilon_{50p})_i \tag{14}$$

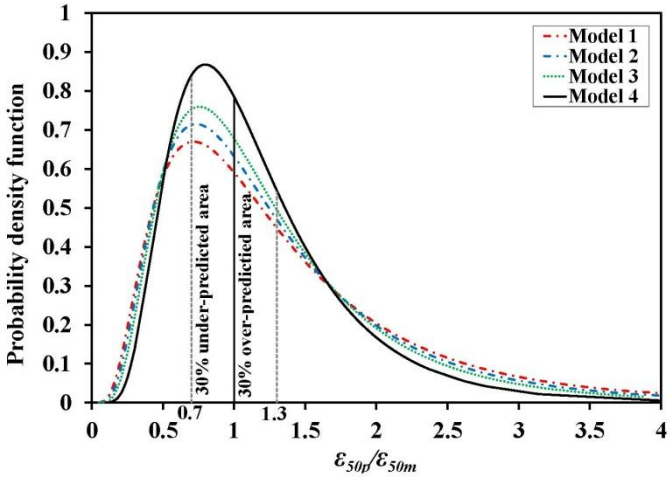

Figure 5. Log normal distributions of $\varepsilon_{50p}/\varepsilon_{50m}$ for proposed models

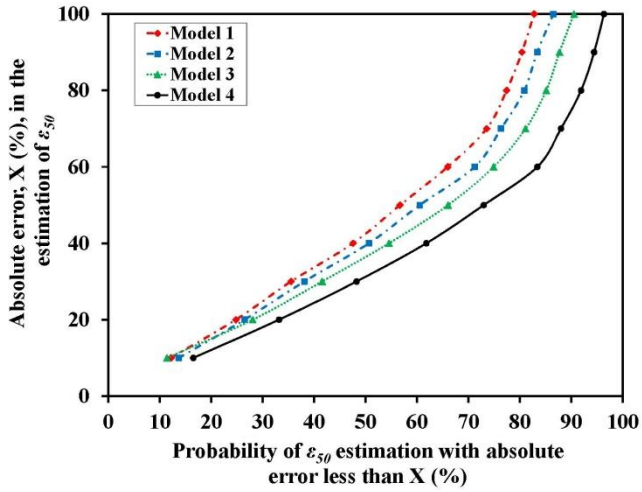

Figure 6. Probability of ε_{50} estimation with absolute error less than a given error, x (%)

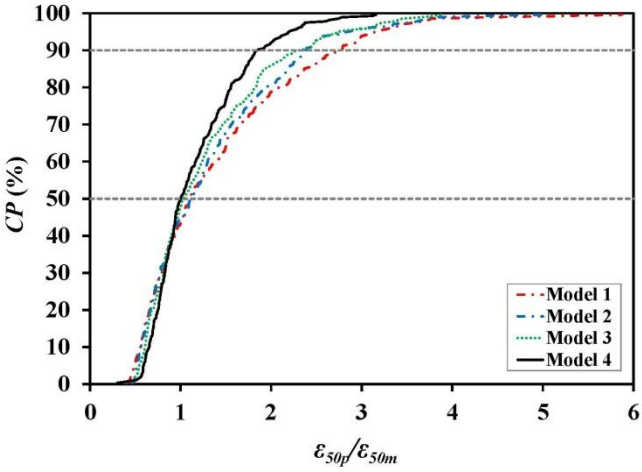

Figure 7. Cumulative probability plot of $\varepsilon_{50p}/\varepsilon_{50m}$ for proposed models

Figure 8 depicts the residuals of the training and testing sets for all presented models versus data number. In this figure, the residuals are scattered along a line indicating the mean (*MR*). In addition, the upper and lower bounds of residual scattering ($MR \pm \sigma$; σ is standard deviation of residuals) are shown in the figure. The ideal performance of each model is achieved by *MR* and σ equal to zero. In general, the lower absolute values of these two parameters represent the better performance of the model. A comparison between the proposed models in Figure 8, with respect to the above parameters, shows the improvement of the models from 1 to 4 by decreasing absolute *MR* and σ values.

5. Validation of the Proposed Models

In order to validate the proposed models, the test results in three different sites are considered. The first site is located at South Pars field, Persian Gulf, Iran, outside the survey area, shown as site 1 in Figure 2. The soil is very soft clay overlying a sandy silt or silty sand layer at shallow depths. Stiff to very stiff clay dominates at deeper parts. The profile of soil properties in 3 boreholes within this site are presented in Figure 9. Figure 10 shows ε_{50} values predicted by different models as well as the measured values obtained from UU tests in borehole depths. In all figures, the recommended ε_{50} values by Sullivan [8] are significantly lower than the measured ones. However, ε_{50} values predicted by models 1 and 4 compare relatively well with the measured ones in the full range of values along borehole depths, as shown in Figure 10. Generally, ε_{50} values show an increasing trend with depth from both laboratory measurements and the currently proposed models predictions. This result is in contradiction with the values of ε_{50} recommended by Sullivan.

Herein, it is attempted to validate the current models using the p-y curve results obtained from pile load tests conducted in two different sites (sites 2 and 3). General information about the considered sites are given in Table 3.

Figures 11(a) and (b) show the p-y curves generated based on ε_{50} values from different models as well as Sullivan recommendations for two different depths in sites 2 and 3, respectively. The figures also include p-y curves obtained from full-scale tests. It is noted that the procedure for generating p-y curves is based on [27].

The figures show that the calculated p-y curves from EPR-based models agree relatively well with the measured p-y curves. However, the p-y curves calculated from Sullivan recommendations show lower values of lateral displacement at all lateral load levels. This implies that using ε_{50} parameter from Sullivan recommendations in generating p-y curve leads to a stiffer behavior of pile-soil system against lateral load in comparison with the real behavior. It is observed that the predicted lateral displacements at 50% of maximum lateral load from the proposed models are 1.5-3.5 and 2-4 times as large as those obtained from Sullivan recommendations for sites 2 and 3, respectively.

In addition, the ratios of predicted to measured lateral pile displacements at maximum lateral load levels for the generated p-y curves at both sites are summarized in Table 6. It is seen from Table 6 that the generated p-y curves based on ε_{50} values from the proposed models yield lateral pile displacements very close to the measured ones with maximum error of 6%. However, using ε_{50} values from Sullivan's recommendation in generating p-y curves leads to very unconservative lateral pile displacements in both sites.

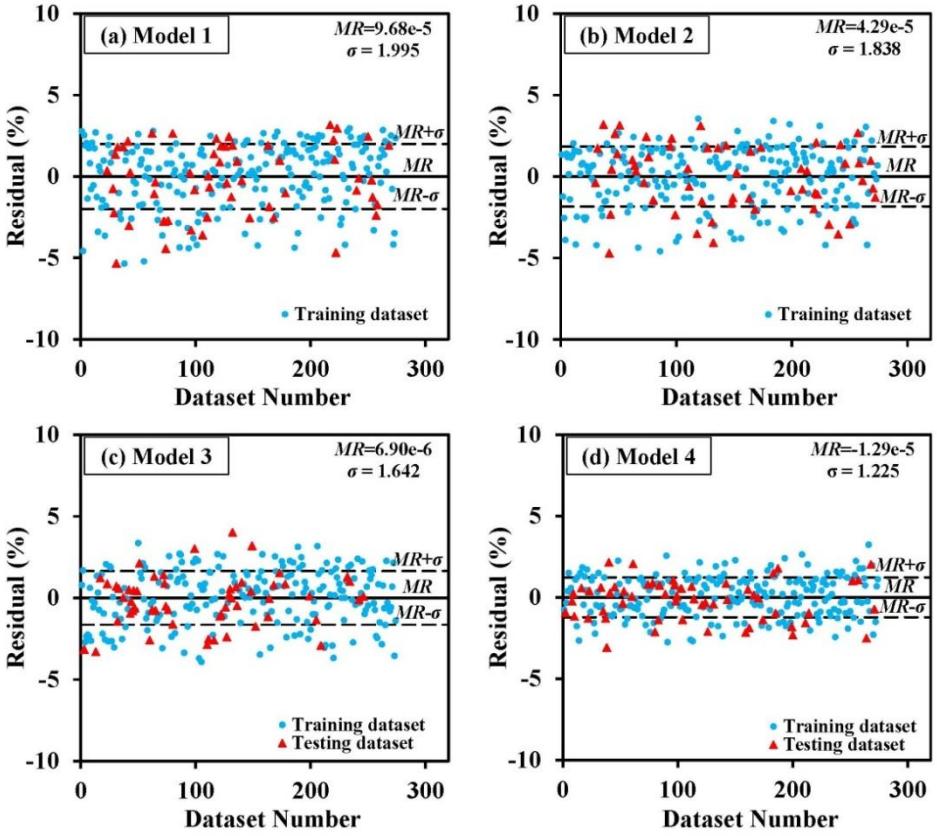

Figure 8. Distribution of residuals for EPR-based models

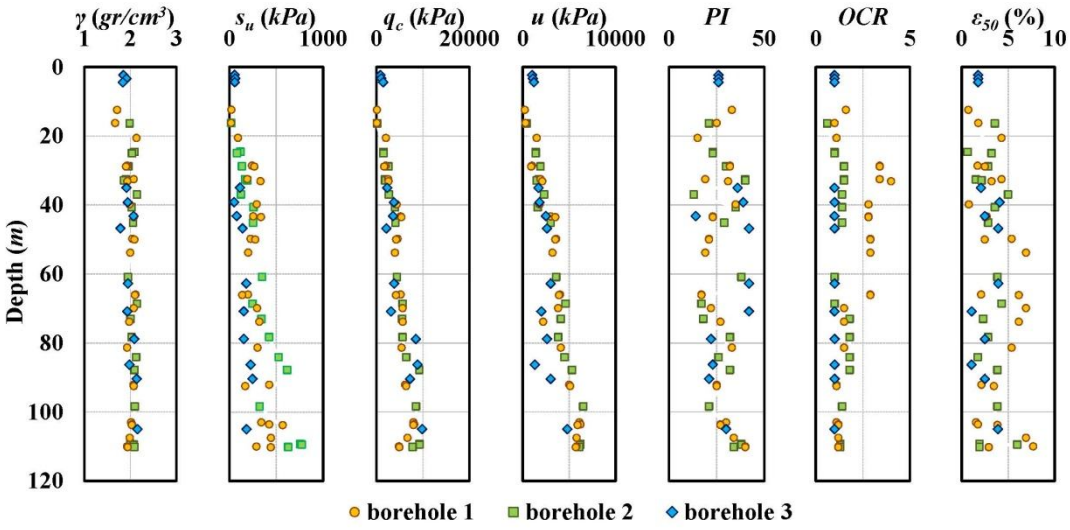

Figure 9. Geotechnical characteristics of soil in the boreholes of site 1

6. Numerical Simulation

In order to show the influence of ε_{50} on the lateral response of pile-soil systems, a numerical study was performed. In this regard, the lateral response of a given pile-soil system with 100 m length, 152.4 cm diameter, was numerically modeled. For considering the nonlinear pile soil interaction, PSI Module of commercial computer program SACS 5.1 is utilized [28,29]. The program uses a finite difference solution to solve the pile-soil model which is represented by a beam-column resting on a nonlinear elastic foundation. The nonlinear pile-soil-interaction is represented in a linear equivalent system. PSI analysis is also used to generate the equivalent foundation stiffness matrix in SACS. PSI models the connections of the soil and the piles as springs. The soil response to lateral loading is nonlinear. To model this nonlinearity, pile deformation can be related to soil resistance through nonlinear transfer curves (p-y curves). The PSI module uses the soil parameters such as ε_{50} to generate the lateral load-deflection (p-y) curves based on API method [7]. In this regard, ε_{50} values, used in generating p-y curves, was firstly substituted by measured ε_{50} values and then by those predicted from Model 4. For comparison purposes, further simulations were also carried out by ε_{50} values

obtained from recommendations of Sullivan et al. [8] and those from Robertson et al. [6].

The soil profile modeled in the numerical study is generally consists of clayey soil with undrained shear strength ranged from 70-650 *kPa*, effective unit weight of 5.8-12.6 kN/m^3 and ε_{50} ranged from 0.05-8.5 .

Table 3. General information about sites 2 and 3

Site No.	Location	Source of p-y curve data	Pile section	Pile section dimension (m)	Depth of measurement (m)	Pile length (m)	Relevant geotechnical properties	Reference
2	Incheon Bridge, Korea	Full-scale field load tests on piles	Circular	Diameter = 1.016 Thickness= 0.016	4D from ground surface (D= pile diameter)	26.6	Table 4	[30-32]
3	Bridge construction site near Oskaloosa, Iowa	Lateral load tests on steel and concrete piles	UHPC* H-shaped	0.254×0.254	5D from ground surface (D=equivalent diameter, 0.287)	10.7	Table 5	[33]

*Ultrahigh-performance concrete

Table 4. Soil properties of site 2 [30,31]

Type	Depth (m)	Unit weight (kN/m^3)	Moisture content (%)	s_u (kPa)	Friction angle (φ) (°)	Recommended ε_{50} in the literature
Upper clay	0-6.3	17.5	21.2	15-30	-	0.02
Lower clay	6.3-16.5	17.5	7.6	30-50	-	0.01
Silty clay	16.5-22.0	17.8	12.7	70	-	0.005
Residual soil	22.0-24.0	18.0	4.33	-	34	-

Table 5. Soil properties of site 3 [33]

Depth (m)	classification	Unit weight (kN/m^3)	Moisture content (%)	LL (%)	PI (%)	s_u (kPa)	Friction angle (φ) (°)	Recommended ε_{50} in the literature
0-1.5	ML	18.8	21.2	42.1	10.4	60	-	0.007
1.5-2.8	CL	18.5	7.6	44.4	17.9	60	-	0.007
2.8-4.9	CL	18.5	12.7	27.9	7.4	136	-	0.005
4.9-5.8	SC	20.5	4.33	32.5	17.7	-	41	-
5.8-7.7	CL	20.4	4.83	36.7	19.2	-	35	-
7.7-9.2	SW	20.6	20.6	-	-	-	42	-
9.2-10.5	CL	20.4	-	-	-	800	-	0.004
10.5-12.0	SW	20.4	-	-	-	-	-	-

Table 6. Ratio of predicted to measured lateral pile displacement

Site No.	At maximum lateral load level		
	Depth (m)	Average of proposed models	Sullivan et al. [8]
2	4D	1.05	0.36
3	5D	1.06	0.31

The numerically modeled pile-soil was laterally loaded and its lateral response containing lateral displacement, moment and shear were compared through various methods of ε_{50} estimation. The variations of lateral response of pile-soil along depth corresponding to various methods are depicted in Figure 12. The lateral responses of pile-soil system using measured ε_{50} values are also shown in this figure. As can be seen, lateral response of pile-soil is considerable only in the upper 40 m soil below the mudline and after that the pile is thoroughly fixed into the soil. It is in accordance to previous results reported in the literature that the most of lateral capacities and deformations of the piles are dependent to the characteristics of the upper part of soil, depending on their diameters [34,35]. However, it is clearly observed that the lateral response of pile in upper portion of the soil is dependent to what value of ε_{50} included in the numerical simulation. Compared to the three response curves obtained from measured ε_{50}, the curve corresponding to Model 4 of this study is of the best performance.

Figure 12 shows that the use of Sullivan et al. [8] recommendations resulted nonconservative displacements while the use of Robertson et al. [6] model introduce conservative to the obtained results. The results of the numerical study also proved the preference of proposed model to other available recommendations in the literature about the prediction of ε_{50}.

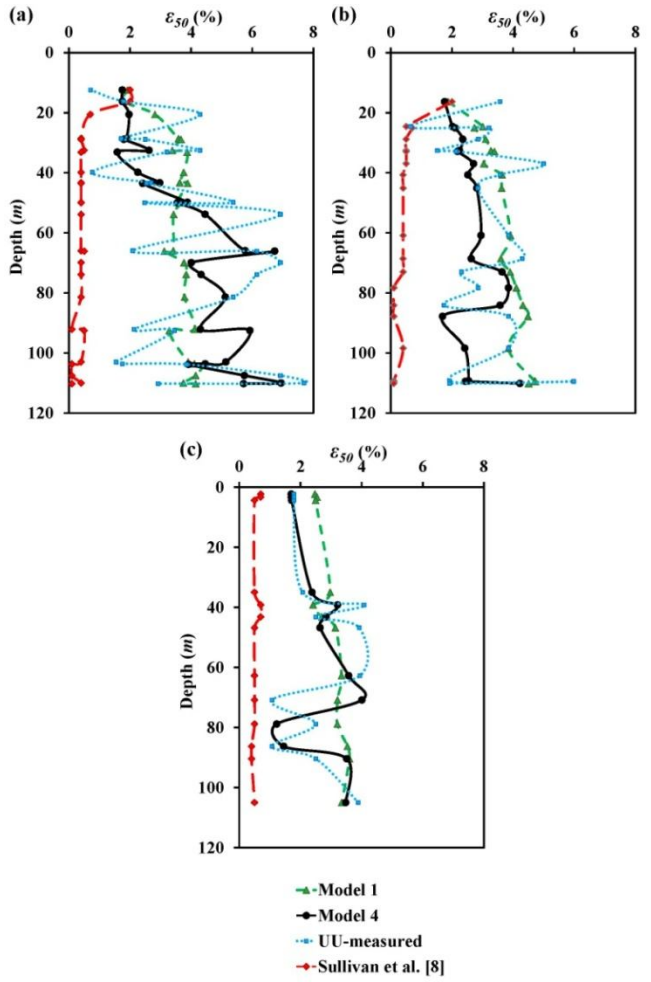

Figure 10. Profiles of predicted and measured ε_{50} values in site 1 from (a) borehole 1, (b) borehole 2, and (c) borehole 3

7. Summary and Conclusions

In this research, the results of field and laboratory tests data in South Pars field, Persian Gulf, Iran, are used to develop models for evaluating ε_{50} using EPR. In this regard, cone tip resistance of CPT and several parameters of cohesive soils (s_u, σ_0, *OCR* and *PI*) are considered in developing models. The conclusions are drawn as follows:

- According to the statistical analyses, the models developed using cone tip resistance (q_c) yield more accurate ε_{50} values than those developed using undrained shear strength of soils (s_u) obtained from UU tests. In general, ε_{50} is more realistically predicted using field-based, instead of laboratory-based, resistance of soil.
- The index properties of soil, e.g. *OCR* and *PI*, significantly improve the performance of the proposed models in predicting ε_{50}.
- According to statistical criteria, the models which are developed considering the effect of overburden pressure (σ_0) lead to better predicted ε_{50} values.
- The models are validated with the field data of a site, located outside the survey area. The predicted ε_{50} values are in relatively well agreement with the measured ones in the full range of values along all boreholes depths in this site. It is found that the predicted ε_{50} values from the proposed models increase with soil depth which agrees with the laboratory measurements.

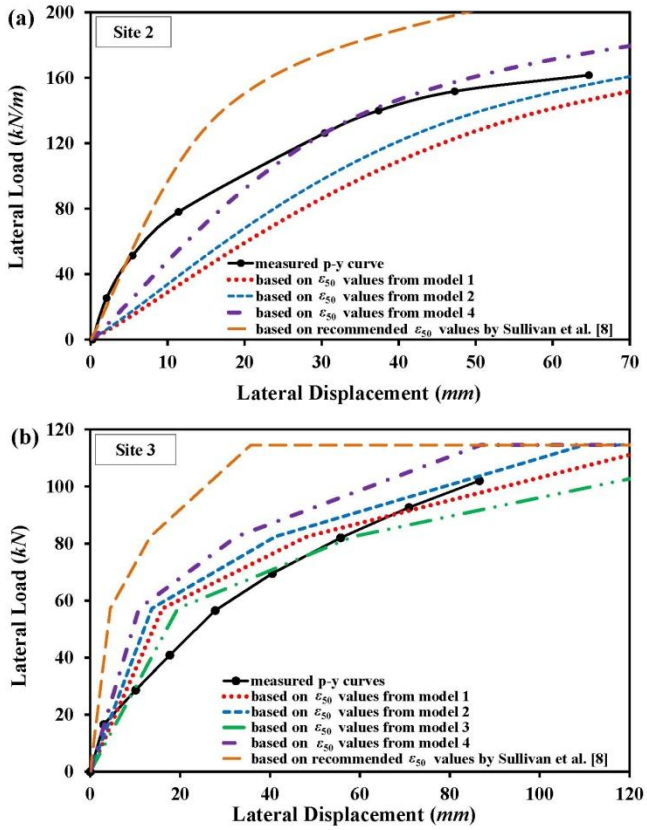

Figure 11. Measured and calculated p-y curves: (a) site 2 at depth = 4D; and (b) site 3 at depth = 5D, (D = pile diameter)

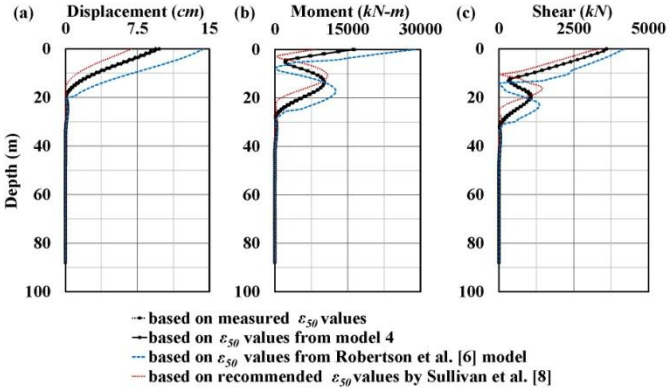

Figure 12. Numerical analysis results of lateral response of typical pile-soil system

- The models are validated with the field data of a site, located outside the survey area. The predicted ε_{50} values are in relatively well agreement with the measured ones in the full range of values along all boreholes depths in this site. It is found that the predicted ε_{50} values from the proposed models increase with soil depth which agrees with the laboratory measurements.

- Further model validation, with full scale lateral pile load tests data in two different sites, demonstrates the models capability in providing ε_{50} parameter to generate p-y curves consistent with the real behavior of pile-soil system measured in the field. Particularly, the results indicate that the p-y curves generated based on ε_{50} values from the proposed models are in better agreement with field data rather than the p-y curves obtained from previously recommended ε_{50} values in the literature.
- The numerical results show the importance of ε_{50} which is incorporated into the numerical model. Herein, it is demonstrated how different ε_{50} values obtained from different ways affect the lateral response of pile-soil system. While the use of Sullivan recommendations result in unconservative response and the use of Robertson model leads to conservativeness to the obtained results, the proposed model can be used to obtain more accurate results.

8. Acknowledgement

The authors wish to thank Mrs. Omrani for her support regarding the numerical modeling.

9. References

1- Matlock, H., (1970), *Correlations for design of laterally loaded piles in soft clay*, Proceedings of Offshore Technology Conference, H.M. Coyle and R.E. Bartoskewitz, eds., Houston, Texas, p.577-594.

2- Reese, L.C. and Welch, R.C., (1975), *Lateral loading of deep foundations in stiff clay*, Journal of Geotechnical and Geoenvironmental Engineering, Vol.101(7), p.633-49.

3- Reese, L.C., Cox, W.R. and Koop, F.D., (1975), *Field testing and analysis of laterally loaded piles in stiff clay*, Proceedings of Offshore Technology Conference, Houston, Texas, 2312. ~~P~~p.671-690.

4- O'Neill, M.W. and Gazioglu, S.M., (1984), *Evaluation of p–y relationships in cohesive soils, Proceedings of Symposium on Analysis and Design of Pile Foundations*, ASCE, 1-5 October, p.192-213.

5- Gabr, M.A. and Borden, R., (1988), *Analysis of load deflection response of laterally loaded piers using DMT*, Proceedings of 1st International Symposium on Penetration Testing, Balkema, Rotterdam, The Netherlands, p.513-520.

6- Robertson, P.K., Davies, M.P. and Campanella, R.G., (1989), *Design of laterally loaded driven piles using the flat dilatometer*, Geotechnical Testing Journal, Vol.12(1), p.30-38.

7- American Petroleum Institute, (1993), *Recommended practice for planning, designing and constructing fixed offshore platforms*, API recommended practice RP-2A, Washington, D.C.

8- Sullivan, W.R., Reese, L.C. and Fenske, C.W., (1980), Unified method for analysis of laterally loaded piles in clay, Numerical Methods in Offshore Piling, ICE, London, p.135-146.

9- Hamilton, J.M. and Dunnavant, T.W., (1993), *Analysis of behavior of the Tilbrook Grange lateral test pile, Large-scale pile tests in clay*, Edited by J. Clarke, Thomas Telford, London, U.K. p.448-461.

10- Hamilton, J.M., Long, M.M., Clarck, J. and Lambson, M.D., (1993), *Cyclic lateral loading of an instrumented pile in overconsolidated clay at Tilbrook Grange*, Large-scale pile tests in clay, Edited by J. Clarke, Thomas Telford, London, U.K. p 381-403.

11- Dunnavant, T.W., (1986), *Experimental and analytical investigation of the behavior of single piles in overconsolidated clay subjected to cyclic lateral loads*, Ph.D. Dissertation, University of Houston, Houston, Texas.

12- Davies, M.P., (1987), *Predicting axially and laterally loaded pile behavior using in-situ testing methods*, M.Sc. Thesis, Department of Civil Engineering, The University of British Columbia, Vancouver, B.C.

13- Schmertmann, J., (1978), *Guidelines for cone penetration test, performance and design, Federal highway administration*, Report No. FHWA-TS-78-209, U.S. Department of Transportation, Washington, D.C.

14- Lunne, T., Robertson, P.K. and Powell, J.J.M., (1997) *Cone penetration testing in geotechnical practice*, Blackie Academic, E.F. Spon/Routledge Publication, New York.

15- Giustolisi, O., Savic, D.A. and Doglioni, A., (2004), *Data reconstruction and forecasting by evolutionary polynomial regression*, Proceedings 6th International Conference on Hydroinformatics, Singapore, Vol.2, p.1245-1252. ISBN 981-238-787-0.

16- Giustolisi, O. and Savic, D.A., (2006), *A symbolic data-driven technique based on evolutionary polynomial regression*, Journal of Hydroinformatics, Vol.8(3), p.207-222.

17- Koza, J.R., (1992), *Genetic Programming: On the Programming of Computers by Natural Selection*, MIT Press, Cambridge, MA. ISBN 0-262-11170-5.

18- Javadi, A.A. and Rezania, M., (2009*), Intelligent finite element method: An evolutionary approach to constitutive modeling*, Advanced Engineering Informatics, Vol.23, p.442-451.

19- Rezania, M., Faramarzi, A. and Javadi, A.A., (2011), *An evolutionary based approach for assessment of earthquake-induced soil liquefaction and lateral displacement*, Engineering Applications of Artificial Intelligence, Vol.24, p.142-153.

20- Lu, M., AbouRizk, S.A. and Hermann, U.H., (2001), *Sensitivity analysis of neural networks in spool fabrication productivity studies*. ASCE Journal of Computing in Civil Engineering, Vol.15(4), p.299-308.

21- Doglioni, A., (2004), *A Novel Hybrid Evolutionary Technique for Environmental 14*

Hydraulic Modelling, PhD Thesis, Technical University of Bari, Italy.

22- Rezania, M., Javadi, A.A. and Giustolisi, O., (2008), *An evolutionary-based data mining technique for assessment of civil engineering systems*, Journal of Engineering Computations, Vol.25 (6), p.500–517.

23- Giustolisi, O. and Savic, D.A., (2009), *Advances in data-driven analyses and modelling using EPR-MOGA*, Journal of Hydroinformatics, Vol.11(3-4), p.225-36.

24- Draper, N.R. and Smith, H., (1998), *Applied Regression Analysis*, Third Edition, John Wiley and Sons, NewYork.

25- Briaud, J.L. and Tucker, L.M., (1988), *Measured and predicted axial response of 98 piles*, Journal of Geotechnical Engineering, Vol.114(9), p.984-1001.

26- Long, J.H. and Wysockey, M.H., (1999), *Accuracy of methods for predicting axial capacity of deep foundations*, Proceedings of OTRC' 99 Conference, Analysis, Design, Construction, and Testing of Deep Foundation, ASCE, Geotechnical Special Publication, Vol.88, p.190-195.

27- American Petroleum Institute, (2000), *Recommended practice for planning, designing and constructing fixed offshore platforms-working stress design*, API recommended practice RP-2A, 21st edition, Washington, D.C.

28- Engineering Dynamics Incorporated (EDI), (2008), *Structural Analysis Computer Systems (SACS)*, Version 5.2.

29- Engineering Dynamics Incorporated (EDI), (2007), *Structural Analysis Computer Systems (SACS) User Manual*.

30- Jeong, S.S., Kim, Y.H., Kim, J.H. and Shin, S.H., (2007), *Cyclic lateral load tests of offshore large diameter piles of Incheon Bridge in marine clay*, Proceedings of 17th International Offshore and Polar Engineering Conference, Lisbon, Portugal, 1-6 July, p.1353–61.

31- Kim, Y.H., Jeong, S.S. and Won, J.O., (2009), *Effect of lateral rigidity of offshore piles using proposed p–y curves in marine clay*, Marine Georesources and Geotechnology, Vol.27(1), p.53-77.

32- Kim, Y. and Jeong, S., (2011), *Analysis of soil resistance on laterally loaded piles based on 3D pile-soil interaction*, Computer and Geotechnics, Vol.38, p.248-257.

33- Suleiman, M., Vande Voort, T.h. and Sritharan, S., (2010), *Behavior of driven ultrahigh-performance concrete H-piles subjected to vertical and lateral loadings*, Journal of Geotechnical and Geoenvironmental Engineering, ASCE, Vol.136(10), p.1403-1413.

34- Briaud, J.L., (1997), *SALLOP: simple approach for lateral loads on piles*, Journal of Geotechnical and Geoenvironmental Engineering, ASCE, Vol.123(10), p.958-964.

35- Chen, J.Y., Matarek, B.A., Carpenter, J.F., and Gilbert, R.B., (2010), *Analysis of Potential Conservatism in Foundation Design for Offshore Platform Assessment*, Final Project Report Prepared for the American Petroleum Institute Under Contract, No. 2007-103130, 280p.