UNIVERSIDAD
NACIONAL
DE COLOMBIA

A comparison of alternative curve fitting techniques for different earthquake fault parameters of Iranian earthquakes

Serkan Öztürk[1],*, Mohammad R. Ghassemi[2], Mahmut Sarı[1]
[1]Gümüşhane University, Department of Geophysics, TR-29100, Gümüşhane, Turkey
[2]Research Institute for Earth Sciences, Geological Survey of Iran, Iran
*Corresponding author: serkanozturk@gumushane.edu.tr

ABSTRACT

In this study, we tried to estimate the optimum linear equations among the parameters associated with different earthquake fault mechanisms for Iranian earthquakes. For this purpose, we tested different curve fitting methods in order to present the most proper empirical relationships between several seismic parameters for different fault systems. In the present paper, 46 large and destructive Iranian earthquakes whose magnitudes change between 5.8 and 7.8 from 1900 to 2014 were used for the analyses. A comparison was made by using four types of curve fitting techniques. The estimation procedures are considered as (1) L_2 or Least Squares Regression, (2) L_1 or Least Sum of Absolute Deviations Regression, (3) Robust Regression and, (4) Orthogonal Regression. Confidence intervals were selected as 95% for all types of regression relationships. In the selection of the best probability distribution, we considered the correlation coefficients of the linear regressions as a powerful and conceptually simple method. Correlation coefficients of all relationships change between 0.299 and 0.986 with Orthogonal regression, between 0.168 and 0.792 with L_1 regression, between 0.059 and 0.829 with Robust regression. For Iranian earthquakes, the most suitable and reliable empirical relationships between moment magnitude (Mw) and surface wave magnitude (Ms), Mw and surface rupture length (SRL), Mw and maximum displacement (MD), and SRL and MD were obtained by Orthogonal regression since it supplies stronger correlation coefficients than those of the other regression techniques in most estimates. The results show that estimated empirical relationships among the different fault parameters by using the Orthogonal regression method can be accepted as more up-to-date and more appropriate in comparison with the other regression norms. Consequently, these equations were suggested as more reliable in the estimation of the maximum surface displacement, maximum surface rupture length and associated with the maximum credible earthquakes for different areas of Iran. Furthermore, obtained relationships can be statistically significant for the assessment of seismic, tectonic and geologic activities, and they can be used to evaluate the rupture hazard of the Iranian Plateau.

Comparación de técnicas de ajuste de curvas para diferentes parámetros de fallas en terremotos en Irán

RESUMEN

En este estudio, los autores se enfocaron en estimar las ecuaciones lineales óptimas para los parámetros asociados con diferentes mecanismos de fallas de terremotos en Irán. Con este propósito se evaluaron diferentes métodos de ajuste de curvas para presentar la relación empírica más apropiada entre varios parámetros sísmicos para diferentes sistemas de fallas. En el presente artículo se analizaron 46 terremotos ocurridos en Irán y con magnitudes entre 5.8 y 7.8, entre 1900 y el 2014. Se realizó una comparación al usar cuatro tipos de técnicas de ajuste de curvas. La estimación de los procedimientos se consideraron así (1) L_2 o Regresión de Mínimos Cuadrados, (2) L_1 o Suma de Mínimos en Regresión de Desviaciones Absolutas, (3) Regresión Robusta y (4) Regresión Ortogonal. Se seleccionaron intervalos de confianza del 95 por ciento para todos los tipos de relaciones de regresión. En la selección de la mejor distribución de probabilidades se consideraron los coeficientes de correlación de las regresiones lineales como un método fuerte y conceptualmente simple. Los coeficientes de correlación de todas las relaciones cambian entre 0.299 y 0.986 con regresión ortogonal; entre 0.168 y 0.792, con regresión L_1; y entre 0.059 y 0.829, con regresión robusta. Para los terremotos en Irán, las relaciones empíricas que más se ajustan y que son más confiables entre la magnitud de momento (Mw) y la magnitud de onda superficial (Ms), entre la magnitud de momento y la longitud de ruptura superficial (SRL), entre la magnitud del momento y el desplazamiento máximo (MD), y entre la longitud de ruptura superficial y el desplazamiento máximo se obtuvieron por la regresión ortogonal, ya que esta provee coeficientes de relación más fuertes que aquellos estimados por medio de otras técnicas de regresión. Los resultados muestran que las relaciones empíricas estimadas entre los diferentes parámetros de falla al usar el método de regresión ortogonal podría ser aceptado como el más actualizado y más apropiado con respecto a otras normas de regresión. Además, estas ecuaciones se sugieren como las más confiables en la estimación del desplazamiento máximo de superficie, de la máxima longitud de ruptura superficial y asociada con la fiabilidad máxima en terremotos para diferentes zonas de Irán. Adicionalmente, las relaciones obtenidas pueden ser estadísticamente significativas para la evaluación de las actividades sísmicas, tectónicas y geológicas, y pueden ser utilizadas para medir el riesgo de ruptura en el altiplano iraní.

Introduction

Fitting of data points by parametric curves and surfaces is considered in many scientific fields such as mathematics, statistics, earth and computer sciences. Several standard statistical software has been used by the practitioners to fit the data sets and to estimate the suitable relationships between some response or dependent variable and hypothesized predictor or independent variables (Öztürk, 2014). The researchers want to develop a physical model from an obtained relationship and therefore, they need to determine the value of the response variable based on the estimated relationships. In many situations, however, it is not possible that these types of relationships are valid just by supposing a mathematical model of the relationship. In addition, these variable selection of regression norms cannot be directly applied to contaminated data sets with an ordinary estimation technique. Thus, it is necessary to be collected the data from the population of available data of these variables and then, an empirical relationship between the dependent and independent variables may be estimated from the data (Giloni & Padberg, 2002).

In applied statistics, investigation of the validity of curve fitting methods to fit the observed data is a significant problem. Despite the differences of existing curve fitting techniques, the principal of the most techniques is based on the classical optimization theory and methods. However, the researchers encounter practical curve fitting problems in many quantitative oriented disciplines, and the solution needs to find the most reliable solution to an over determined system of linear equations (Cadzow, 2002). An effective and accurate alternative curve fitting technique has a great importance and serves as a basic module in practice. For this reason, the curve fitting is one of the most common tools in establishing the relationship between a response and an explanatory variable statistic for many social and engineering fields (Durio & Isaia, 2003). The ignoring of a significant component of variability and the equation error is one of the main problems with the usage of the curve fitting models for any data set. Therefore, this problem can lead to either over or under correction for measurement error depending upon the relative sizes of the variances included. The estimation of correlation coefficient for the curve fitting models is a simple tool, and the assessment of the curve fitting can be succeeded by means of the estimation of some criterion such as the calculation of coefficient for the curve fitting model. Many statistical analyzes show that it is not considered as a final model selection criterion. However, this calculation supplies an evidence of the conformity of the selected descriptive variables in estimating the response from the curve fitting model (Renaud & Victoria-Feser, 2010).

In many data processing applications, the linear equation systems under consideration are not consistent and researchers desired to estimate the most approximate solution. Many curve fitting techniques have been used in these types of analyses, for example: L_1 Norm or Least Sum of Absolute Deviations Regression (Giloni & Padberg, 2002), L_2 Norm or Least Squares Regression (Cadzow, 2002), Robust Regression (Huber, 1964), Total Least Squares or Orthogonal Regression (Carroll & Ruppert, 1996), Geometric Mean Regression (GMR, Leng et al., 2007), Principle Components Regression (PCR, Maronna, 2005), Covariate Adjusted Regression (CAR, Şentürk & Nguyen, 2006) and Least Trimmed Squares Regression (LTS, Rousseeuw & Leroy, 1987). The main purpose of this study is to estimate the most optimum and reliable curve fitting solutions for linear equations among the earthquake fault parameters of Iranian earthquakes. In this context, a comparison among the first four curve fitting techniques mentioned above was made in order to provide the optimum statistical model between different variables. We did not discuss the other methods such as PCR, GMR, LTS or CAR since they have been preferred in more distinctive applications and they have not generally been used in geophysical analyses. One can also find many details about all these curve fitting techniques in literature (e.g., Branham, 1982; Stefanski, 1991; Hartmann et al., 1997; Spiess & Hamerle, 2000; Shi & Lukas, 2005). Consequently, L_1 Norm, L_2 Norm, Robust and Orthogonal regressions were tested for Iranian earthquakes in order to provide newer and more suitable empirical relationships among different earthquake fault parameters. By using these different statistical regression techniques, we aimed to make a thorough and consistent appraisal of maximum surface rupture length, maximum surface displacement and associated with the maximum credible earthquakes for different parts in Iran. Also, regional differences in these faulting parameters can be evaluated in terms of geologic, seismic and tectonic variations as the preliminary results of the future studies for the study region.

Mathematical background of the curve fitting methods

Curve fitting problems are classified within the category of mathematical problems, and many different distance functions or metrics have been utilized to carry out linear curve fitting process. There are many mathematical techniques for the solution of these types of problems. However, since our aim is only to evaluate different models and to estimate the optimum relationships, we will not discuss the mathematical background in detailed. For this reason, we will give short and principal explanations for four different curve fitting models in this section, as well as their basic statistical features and quality of fit.

The curve fitting process for a given data set has a great importance in statistical data analysis. These types of data analyses are naturally related to distances in Euclidean geometry and by considering the linear algebra, an analytical result may be possible. Some metrics or distance functions may be useful to perform a linear equation to the data. In order to analyze a linear equation system, it is accepted m observations or measurements on the dependent variable y and some number $n \geq 1$ of independent variables $x_1,\ldots,x_n$. Each one of them is also known m values. It can be given as (Giloni et al., 2006):

$$y = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_m \end{pmatrix}, \quad X = \begin{pmatrix} \chi_1^1 & \cdots & \chi_n^1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \chi_1^m & \cdots & \chi_n^m \end{pmatrix} = \begin{pmatrix} x^1 \\ \cdot \\ \cdot \\ \cdot \\ x^m \end{pmatrix} = (x_1,\ldots,x_n) \quad (1)$$

where $y \in R$ is a vector of m measurements and X is an $m \times n$ matrix of real frequently termed as the design matrix. In addition, $x_1,\ldots,x_n$ are column vectors with m components and $x^1,\ldots,x^m$ are row vectors with n components corresponding to the columns and rows of X, respectively. Linear equation system can be formulated statistically as in the following (Durio & Isaia, 2003):

$$y = X\beta + \varepsilon \qquad (2)$$

where $\beta^T = (\beta_1,\ldots,\beta^n)$ is the vector of parameters of the linear system and $\varepsilon^T = (\varepsilon_1,\ldots,\varepsilon_m)$ a vector of m random variables corresponding to the error terms in the suggested equation. An upper index T denotes "transposition" of a vector or matrix. The dependent variable y is a random variable in the hypothesized model. The observations include some "noise" or observation errors and they are taken in the error terms ε. However, we can write it for the linear equation that we want to solve:

$$y = X\beta + r \qquad (3)$$

where given several randomly fixed parameter vector β, the components r_i of the vector $r^T = (r_1,\ldots,r_m)$ are the residuals that result, given the measurements y, a fixed design matrix X, and the chosen vector $\beta \in R^n$. Consequently, the residuals, r, are in terms of the hypothesized model, realizations of the random error terms ε given the particular measurements y and parameter settings β. Given y and X, the main purpose in linear fitting model is to estimate the parameter settings $\beta \in R^p$. This means that several suitable measure of the dispersion of the resulting residuals $r \in R^m$ must be as small as possible (Giloni & Padberg, 2002).

It has completely possible that, e.g., $x_1^j = 1$, for all $j \in \{1,\ldots,m\}$ in the design matrix X (Giloni & Padberg, 2002). In this situation, β_1 is attributed as the "intercept term" and it is corresponded to the situation in the two parameter case, i.e., when $n=2$. If $x_1^j = 1$, for all $j \in \{1,\ldots,m\}$ and $n=1$, the problem of estimating a "best" fitting scalar β_1 means that we desire several good measure of "centrality" of the measurements y.

Least Squares Method (L2 Norm)

Least squares (LS) method is the most preferred and the best well-known regression technique. This method is the most basic form of the LS optimization problems, and there are many applications in mathematics and statistical data analyses besides other scientific fields. The model estimates β are determined

by minimizing the sum of squared residuals under the Euclidean (or L_2) norm $\|x\|_2 = \sqrt{\sum x_j^2}$, i.e. The aim is to estimate the parameters $\beta \in R^p$ such that:

$$S = r^T \left(y - X\beta\right)^T \left(y - X\beta\right) = y^T y - 2\beta^T \left(X^T y\right) + \beta^T \left(X^T X\right)\beta \quad (4)$$

must be the minimum. This transformation is uniform and optimality is not affected from it. This minimized equation is positive semi-definite and thus, the first-order conditions achieve the process. For example, in order to minimize S, its gradient ∇S in accordance with β should be computed and adjusted as zero or can be formulated as (e.g., Cadzow, 2002; Giloni & Padberg, 2002):

$$\frac{1}{2}\nabla S = -X^T y + \left(X^T X\right) = 0\beta \quad (5)$$

The statement $\left(X^T X\right) = \beta X^T$ should be analyzed for β and is named as normal equations for L_2 norm (Giloni & Padberg, 2002). Supposing that the rank of X is n, i.e., that $r(X)=n$, it means that $\left(X^T X\right)^{-1}$ exist and the ideal $\beta = \beta^{LS}$ is stated as:

$$\beta^{LS} = \left(X^T X\right)^{-1} X^T y \quad (6)$$

i.e., this regression yields an excellent level. The solution process is focused on the inversion of X^TX, but a quantitative analyzer should accomplish the possible singularity of X^TX.

The statistical curve fitting techniques have been applied densely for a very long time. Although some significant progress has been made to evaluate the goodness of fit (GOF) besides the other statistical features of the linear curve fitting techniques, some features of the statistical curve fitting techniques require generally poor assumptions (Sen and Srivastava 1990). The estimates β^{LS} in the LS can be accepted as the best estimators when the error terms in the model have a normal distribution. However, the LS estimates depend generally on the presence of the second moment of the error distribution. As a result, this regression model is especially useful for all cases in which involve the analyses of great data sets and handle great samples with the constant numbers of outliers (Cadzow, 2002).

Least Sum of Absolute Deviations Method (L1 Norm)

It is well-known that the LS method is very sensitive to unusual measurements in a data. Hence, a large number of robust estimators have been suggested as alternatives, and the least sum of absolute deviations (L_1) technique was one of the earliest among them. In this method, coefficients are calculated by minimizing the sum of the absolute values of the residuals. Because L_1 norm can be highly affected by a single measurement, it has been mostly disregarded as a robust alternative to L_2 norm. It has many names in the literature such as LAD (least absolute deviation), LAE (least absolute error), LAR (least absolute residual), LAV (least absolute value), LSAD (least sum of absolute deviations), MAD (minimum absolute deviation), MSAE (minimum sum of absolute errors). L_1 norm has an increasing interest as an alternative to L_2 norm over the past 30 years or so, and it has been a new perspective since 1960s by different researches (e.g., Fama, 1965; Mandelbrot, 1967; Blattberg & Sargent, 1971; Huber, 1987). More descriptive or remarkably illustrative results were provided in many cases by using the L_1 norm. The estimates of the parameters $\beta \in R^p$ of the linear curve fitting technique are achieved by minimizing the sum of the absolute residuals (Giloni & Padberg, 2002):

$$\|y - X\beta\|_1 = \sum_{i=1}^{n} |y_i - x_i\beta| \quad (7)$$

Instead of the L_2 model, L_1 model $\|x\|_1 = \sum |x_j|$ is used to minimize dispersion of the residuals. This target function is obviously non differentiable and computation does not assist (Giloni & Padberg, 2002). A significant advantage of L_1 method in comparison with L_2 regression is its stability. It is well known that the L_1 technique can resist some large errors in the data y for unlimited problem (Shi and Lukas 2005). When a small part of the data being evaluated is not reliable (i.e., contains data outliers), L_1 model is suitable. L_1 model application may ignore some bad data points if it emphasizes the larger part of data points (Cadzow, 2002).

L_1 model calculation of β is the solution to the problem for the linear model which has linear constraints:

minimize $e_n^T r^+ - e_n^T r^-$

subject to $X\beta + r^+ - r^- = y$, $\quad$ (8)

β free, $r^+ \geq 0, \quad r^- \geq 0$

In the equation (8), the residuals r of the general formulation (3) are simply replaced by a difference r^+-r^- of positive variables. The non-differentiability of the exact value target function is dissembled by the target function of (8). However, it accurately captures the target function of L_1 norm. Because of the non-differentiability of the criterion function, L_1 estimation is more difficult than L_2 estimation. The statistical hypothesis improved for L_1 method is less advanced than it is for L_2 method (Rosenberg and Carlson 1977), whereas we have the formula $\beta^{LS} = \left(X^T X\right)^{-1} X^T$ for the L_2 norm estimates β^{LS}. It allow us to obtain the sample distribution of β^{LS} from the error distribution, and the dependence of the L_1 norm estimator on the errors is more complex. The asymptotic theory for L_1 norm is not as well-developed as for L_2 norm. This theory is true to a certain degree. It is also true for high-breakdown model estimators. In addition, L_1 norm estimator is not at all robust to measurements with unusual predictor values; that is, it has a low breakdown point (Giloni et al., 2006). Also, one can see Arthanari & Dodge (1981) for further mathematical background and discussion of the L_1 approach.

Orthogonal Method (Total Least Squares Method)

Orthogonal regression (OR) is one of the best known methods for errors-in-variables estimation in the simple linear curve fitting technique. This method is also sometimes named as the functional maximum likelihood estimator under the limitation of known error variance ratio. In an ordinary linear curve fitting model, the aim is to minimize the sum of the squared vertical distances between the y data values and the corresponding y values on the fitted line. However, the aim in OR is to minimize orthogonal (perpendicular) distances from the data points to the fitted line.

OR is derived a pure observation error perspective. It is supposed that there are theoretical constants y and X which are linearly related through (Carroll & Ruppert, 1996):

$$y = \beta_0 + \beta_1 X \quad (9)$$

Formulation (9) shows that y and X would be exactly linearly related if they could be measured. In a classical OR technique, instead of measuring (y, X), Carroll & Ruppert (1996) estimated them by corrupted by observation error:

$$Y = y + \varepsilon$$
$$W = X + U \quad (10)$$

where ε and U are independent mean zero random variables with variances σ_ε^2 and σ_u^2, respectively. Equations (9) and (10) are combined, the result can be written as:

$$y = \beta_0 + \beta_1 X + \varepsilon \quad (11)$$

The knowledge of error variance ratio is necessary for OR techniques:

$$\eta = \frac{Var\left(Y|X\right)}{Var\left(W|X\right)} = \frac{\sigma_\varepsilon^2}{\sigma_u^2} \quad (12)$$

Equation (12) is based on a sample of size n, $\left(Y_i, W_i, X_i\right)_{i=1}^{n}$ and the course of X's are unknown and unobserved. OR estimator is estimated by minimizing:

$$\sum_{i=1}^{n}\left\{\left(Y_i - \beta_0 - \beta_1 X_i\right)^2 / \eta + \left(W_i - X_i\right)^2\right\} \tag{13}$$

in the unknowns, namely $\beta_0, \beta_1, X_1, \ldots, X_n$. Equation (13) is just the ordinary total Euclidean or perpendicular distance of $\left(Y_i, W_i\right)_{i=1}^{n}$ from the line $\left(\beta_0, \beta_1, X_i, X_i\right)_{i=1}^{n}$ if $\eta = 1$. However, equation (13) is a weighted perpendicular distance if $\eta \neq 1$. If S_y^2, S_w^2 and S_{wy} are the sample variance of the Y's, the sample variance of the W's, and the sample covariance between the Y's and the W's, respectively, OR estimate of slope is (Carroll & Ruppert, 1996):

$$\hat{\beta}_1 = \frac{s_y^2 - \eta s_w^2 + \left\{\left(s_y^2 - \eta s_w^2\right)^2 + 4\eta^2_{wy}\right\}^{1/2}}{2s_{wy}} \tag{14}$$

This well-known OR estimator is explained in many studies such as Madansky (1959), Kendal & Stuart (1979), Weisberg (1985), Fuller (1987) and Leng et al. (2007). The use of OR must include a careful analysis of equation error and is not only the usual estimation of the ratio of observation error variance. Also, OR fitting technique is an acceptable norm as long as η in equation (4) is specified truly. If its assumptions hold, OR is a completely justifiable technique of estimation. Because OR method does not consider equation error, it frequently causes itself to misuse by careless as a technique (Öztürk, 2014).

Robust Method

The most important issue in the LS regression is non-robustness to outliers. If the error distribution is not normal, especially if the errors are heavy-tailed, linear LS calculations can produce unexpected results. Especially, if there are extremely incompatible data points, there can be a strong effect on the model parameters by these outliers and we can see it on the results. An easy solution can be obtained by repetitively weeding up these outliers and by calculating again the LS fit for the remaining points. Another method is to operate a different regression model which is not as vulnerable as LS to incompatible data and it is called as Robust Regression (RR). This technique was introduced by Huber (1964), and the best known and common technique for RR is M-estimation. If we take the general curve fitting model in equation (2) into account for the ith of m measurements:

$$yi = \alpha + \beta_1 \chi_{i1} + \beta_2 \chi_{i2} + \ldots\ldots + \beta_k \chi_{ik} + \varepsilon_i \\ = x_i'\beta + \varepsilon_i \tag{15}$$

The fitted model is formulated as:

$$yi = a + b_1 \chi_{i1} + b_2 \chi_{i2} + \ldots\ldots + b_k \chi_{ik} + e_i \\ = x_i'b + e_i \tag{16}$$

The target function is minimized by the general M-estimator as follow:

$$\sum_{i=1}^{b} \rho\left(e_i\right) = \sum_{i=1}^{n} \rho\left(y_i - x_i'b\right) \tag{17}$$

If the derivative of ρ is accepted as ψ=ρ', differentiating the target function according to the coefficients, b, and getting the partial derivatives to 0, yield a system of $k+1$ estimating equations for the coefficients:

$$\sum_{i=1}^{n} \psi\left(y_i - x_i'b\right)x_i' = 0 \tag{18}$$

If we describe the weight function $\omega(e) = \psi(e)/e$ and $\omega_i = \omega(e_i)$, then the estimating equations can be given as follow:

$$\sum_{i=1}^{n} \omega_i \left(y_i - x_i'b\right)x_i' = 0 \tag{19}$$

Solving the estimating equations is a weighted LS problem, minimizing $\sum \omega_i^2 e_i^2$. However, the weights depend upon the residuals, the residuals depend upon the estimated coefficients, and the estimated coefficients depend upon the weights. One can find many details and mathematical backgrounds in Huber (1964).

Nonlinear curve fitting techniques play a significant role in many fields. The classical LS method is generally preferred in many cases for calculating the parameters of a nonlinear model. However, it is well known that these classical techniques are frequently very sensitive to extreme values. Most robust improvements on the estimation of curve fitting model are based on the generalizations of the LS or maximum likelihood methods. RR technique is less influenced by outliers. However, the methods of small sample asymptotes may be very helpful in evaluating the behavior of RR estimates. Several of RR methods are also explained in many studies such as Huber (1981), Field (1997), Sinha et al. (2003) and Abdelkader et al. (2010).

Estimation of correlation coefficient for curve fitting methods

In order to fit mathematical models to measured data, curve fitting applications have been extensively preferred by researchers in many different disciplines. When the error terms identically and independently show a normal distribution, the classic regression method of the LS is efficient. Real data frequently include the extreme values and these data points are covered by a majority of the data while the non-measured random disturbances in a curve fitting method are generally accepted as normally distributed. Many extreme values are controlled by genuinely thick-tailed or asymmetric error distributions, whereas these errors may be the result of observation mistakes or human recording error. Thus, a useful solution for the problem is to minimize the sum of residual error magnitude. If data vector includes a small number of data outliers, the sum of error magnitudes is of particular use in many applications. In these situations, the sum of squared errors criterion is wrongly influenced by these data outliers and therefore often leads to a poor selection of the coefficient vector. In fact, separating the extreme values is not appropriate since they can represent the correct data generating process in these types of situations (Boyer et al., 2003).

The choice of an optimum probability function for a given data set is one of the most significant issues in curve fitting analyses. As seen in literature studies, different methods such as the chi-square test of homogeneity, the test of normality, Kendall and Spearman coefficients, Bayesian Interactive Model, Principal Component Analysis, Mutual Information Coefficient, Hoeffding method etc. can be applied and then, the most appropriate model can be selected. However, there is not any specific rule in the choosing the suitable distribution or parameter calculation methods. The choosing a suitable distribution in many cases is dependent on the GOF assessment. The GOF in a model can be defined as the technique for analyzing how well sample data compatible with an accepted probability distribution (Öztürk, 2014). Some tests for the GOF have been improved and one of the best known and the most frequently used criteria among them is called as correlation coefficient. This term is given as R^2 or sometimes r is used, and it has a significant role in curve fitting analyses. As stated in Heo et al. (2008), it is not recommended to be used as a single tool for the GOF, but it can supply an acceptable and quick solution. Although correlation coefficient is only based on the covariance penalty, it is accepted as a substantially simple and effective tool. Alternative models were not used in this study despite many techniques as suggested in literature since these types of applications have been used for more certain fields and they have not been used in geophysical applications.

Considering equation (2), we can assume that $\hat{y}_i = x_i^T \hat{\beta}$ is the predicted value for y_i in the selected model with $p < q$ explanatory variables (based on a LS) and $RSS = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2$ is the identical residual sum of squares. Thus, Γ_p can be given as (Renaud & Victoria-Feser 2010):

$$\Gamma p = \frac{E(RSS)}{n\sigma^2} + \frac{2}{n\sigma^2} \sum_{i=1}^{n} Cov(y_i, \hat{y}_i) - 1 \qquad (20)$$

As mentioned above, a very simple and practice tool for the GOF is also provided by determination of correlation coefficient for the linear curve fitting models. If we assumed that x is random, we can describe as the parameter which is the squared correlation between y and the best linear combination of the x (Anderson, 1984):

$$\phi^2 = \max Corr^2\left(y, x^T\beta\right) \qquad (21)$$

The R^2 is generally given as a quantity and it estimates the percentage of variance of the response variable explained by linear relationship. It is estimated as follow:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \qquad (22)$$

where ESS, TSS and RSS denote the explained, total and residual sum of squares, respectively. If we have an intercept term in the linear model, estimation of R^2 is essentially equal to the square of the R^2 between y_i and $\hat{y}_i$, i.e. (e.g., Greene, 1997):

$$R^2 = \left(\frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)(y_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}}_i)^2}}\right)^2 \qquad (23)$$

where $\bar{y}$ is the average values of the measurements y_i and $\bar{\hat{y}}$ is the fitted quantiles $\hat{y}_i$. This equation shows that R^2 measures the GOF of the linear model by its ability to predict the response variable and its ability measured by the correlation. Moreover, this formulation denotes that the response distribution does not need to be Gaussian to allow for the interpretation of R^2. It is also a direct estimator of the population parameter of equation (3). The scale and location of R^2 do not change and this is statistically independent of the standard deviation and average, $\bar{y}$. In fact, R^2 supplies a quantitative evaluation of fit and shows the linearity of the probability plot. As an important result, if R^2 is close to 1.0, it is assumed that the measurements can be constructed from the fitted distribution (Heo et al., 2008).

Data catalog of Iranian earthquakes and a brief definition of the seismotectonics

The first aim of this study is to prepare a uniform data catalog taking the types of surface wave magnitude (Ms) and moment magnitude (Mw) into consideration. Then, the second aim is to estimate the new empirical relationships among the seismicity parameters of different earthquake faulting mechanisms such as Mw and surface rupture length (SRL), Mw and maximum displacement (MD), and SRL and MD. The data set used in the scope of this work is updated form Ghassemi (2016). Table 1 shows the details of parameters for different faulting mechanisms and is modified from Ghassemi (2016). He used 46 Iranian earthquakes whose magnitudes vary from 5.8 to 7.8 and which occurred between the time intervals 1900 and 2014 (Figure 1). As an initial step in Ghassemi (2016), a catalog of Iranian earthquakes related to direct loses was used and the number of initial earthquake catalog was reduced to 41 events. Then, 5 more events were also added to complete the catalog and to extend it to 2014. The outcomes in the

Table 1. Data set of fault parameters for Iranian earthquakes between 1900 and 2014 (for details, see Ghassemi, 2016). RLSS: right lateral strike-slip, LLSS: left lateral strike-slip, NF: normal fault, TF: thrust or reverse fault, SRL: surface rupture length, MD: maximum displacement, VDC: vertical displacement component, HDC: horizontal displacement component

| No	Date	Longitude	Latitude	Rupture Mechanism	Mw	SRL	MD
1	23.01.1909	49.13	33.41	RLSS-NF	7.3	>40	>1
2	18.04.1911	57.03	31.23	TF	6.3	18	vdc. >0.5
3	01.05.1929	57.81	37.73	RLSS	7.2	74	2
4	06.05.1930	44.60	38.24	NF-RLSS	7.1	16-30	4
5	16.02.1941	58.87	33.41	RLSS	6.3	8-10	0.5-1
6	27.11.1945	63.47	25.02	TF	7.8		
7	23.09.1947	58.67	33.67	RLSS	6.8	20	vdc. 0.3-0.8 hdc. 1
8	05.07.1948	57.73	29.88	RLSS	6.1		
9	05.10.1948	58.55	37.88	RLSS	7.1		
10	12.02.1953	54.88	35.39	TF	6.5	>8	vdc. >1.4
11	02.07.1957	52.47	36.07	TF	6.8		
12	13.12.1957	47.82	34.58	TF	6.7		
13	16.08.1958	48.17	34.30	NF-RLSS	6.7	20	vdc. 1.5
14	01.09.1962	49.81	35.71	TF	7.1	80	vdc. 1.4 hdc. 0.6
15	31.08.1968	58.96	34.02	LLSS	7.1	80	vdc. 2.1 hdc. 4.5
16	01.09.1968	58.23	34.05	TF	6.4		
17	30.07.1970	55.89	37.67	RLSS	6.5		
18	10.04.1972	52.98	28.38	TF	6.9		

(*Continued*)

Table 1. Data set of fault parameters for Iranian earthquakes between 1900 and 2014 (for details, see Ghassemi, 2016). RLSS: right lateral strike-slip, LLSS: left lateral strike-slip, NF: normal fault, TF: thrust or reverse fault, SRL: surface rupture length, MD: maximum displacement, VDC: vertical displacement component, HDC: horizontal displacement component

No	Date	Longitude	Latitude	Rupture Mechanism	Mw	SRL	MD
19	02.07.1972	50.85	30.06	TF	5.5	1.5	vdc. 4
20	24.11.1976	44.02	39.12	RLSS	7.1	55	vdc. 0.5 hdc. 3.5
21	21.03.1977	56.45	27.59	TF	7.0		
22	06.04.1977	50.76	31.90	TF	5.9		
23	19.12.1977	56.61	30.90	RLSS	5.8	19.5	hdc. 0.2
24	16.09.1978	57.12	33.40	TF	7.3	85	< 1.7
25	16.01.1970	59.50	33.80	LLSS	6.5		
26	14.11.1979	59.81	33.91	RLSS	6.6	20	1
27	27.11.1979	59.63	34.05	LLSS	7.1	68	2.5-4
28	11.06.1981	57.68	29.85	RLSS-TF	6.6	15	
29	28.07.1981	57.77	29.97	TF	7.0	65	0.4
30	20.11.1989	57.72	29.90	RLSS	5.8	11	vdc. 0.01 hdc. 0.004
31	20.06.1990	49.23	37.00	LLSS	7.3	>80	vdc. 0.95 hdc. 0.6
32	06.11.1990	55.46	28.24	TF	6.4	15	vdc. 1.5
33	23.02.1994	60.54	30.78	TF	6.1	9.5	1.7
34	24.02.1994	60.51	30.79	TF	6.2		
35	04.02.1997	57.31	37.73	RLSS	6.4	15	0.5-1.0
36	28.02.1997	48.07	38.12	RLSS	6.0		
37	10.05.1997	59.81	33.85	RLSS	7.2	125	vdc. 0.9 hdc. 2.3
38	14.03.1998	57.59	30.14	RLSS-NF	6.6	23	hdc. 3
39	18.11.1998	57.58	30.33	RLSS	5.3	4	
40	22.06.2002	49.01	35.62	TF	6.4	3	0.16
41	26.12.2003	58.27	28.90	RLSS	6.5	5	0.2
42	28.05.2004	51.57	36.29	TF	6.2		
43	22.02.2005	56.79	30.71	TF	6.4	13	1
44	27.11.2005	55.83	26.75	TF	6.0		
45	20.12.2010	59.19	28.33	RLSS	6.5	14	0.05
46	11.08.2012	46.78	38.41	RLSS	6.4	13	0.5-1

present study may differ from specific earthquakes which are used to estimate the relationships between Mw and SRL, Mw and MD, and SRL and MD since this catalog is not complete for SRL and MD. In general, magnitude scale is given as Ms in the original Iranian earthquake database. This data set includes 166 events for Ms and Mw types between 1962 and 2004. In order to prepare a uniform catalog covering Mw scale, new relationships between Mw and Ms were firstly obtained and then, the other empirical relationships among the faulting parameters were formulated by using the most suitable curve fitting technique. One can see Ghassemi (2016) for many details such as the geometry and kinematics of surface ruptures, the secondary ruptures associated with earthquake faulting mechanism, seismic properties of earthquake faulting and the previous relationships among these variables.

Iran is located within the Alpine-Himalayan orogenic belt in the active collision zone between the Eurasia to the north and Arabia to the south. The Iranian Plateau is consisted of different tectonic systems such as transpressive fold-and-thrust mountain belts with active strike slip and reverse faulting, an active subduction zone and recent volcanic activity; as a result, the plateau shows different crustal blocks of variable thickness and rigidity interspersed by relatively stable aseismic blocks of different sizes (Berberian, 2014). The Iranian Plateau shows active deformation, high elevation and earthquake activity with complex interactions of active strike-slip faults and thrusts, and different destructive earthquakes recorded in historical and instrumental periods. The most significant tectonic structures are mostly accommodated by shortening and strike-slip faulting in the mountain belts and can be given as the Great Caucasus, Zagros, Alborz, Kopeh-Dagh, active Makran subduction zone and also Azerbaijan (northwest Iran) seismotectonic province (Javidfakhr et al., 2011). Along the Iranian Plateau, fault ruptures from earthquakes show a large-scale typical changes and evaluation of them supply many practical understanding for surface rupture hazard in Iran (Ghassemi, 2016). It is well known that Iran was struck with strong and destructive earthquakes in the past
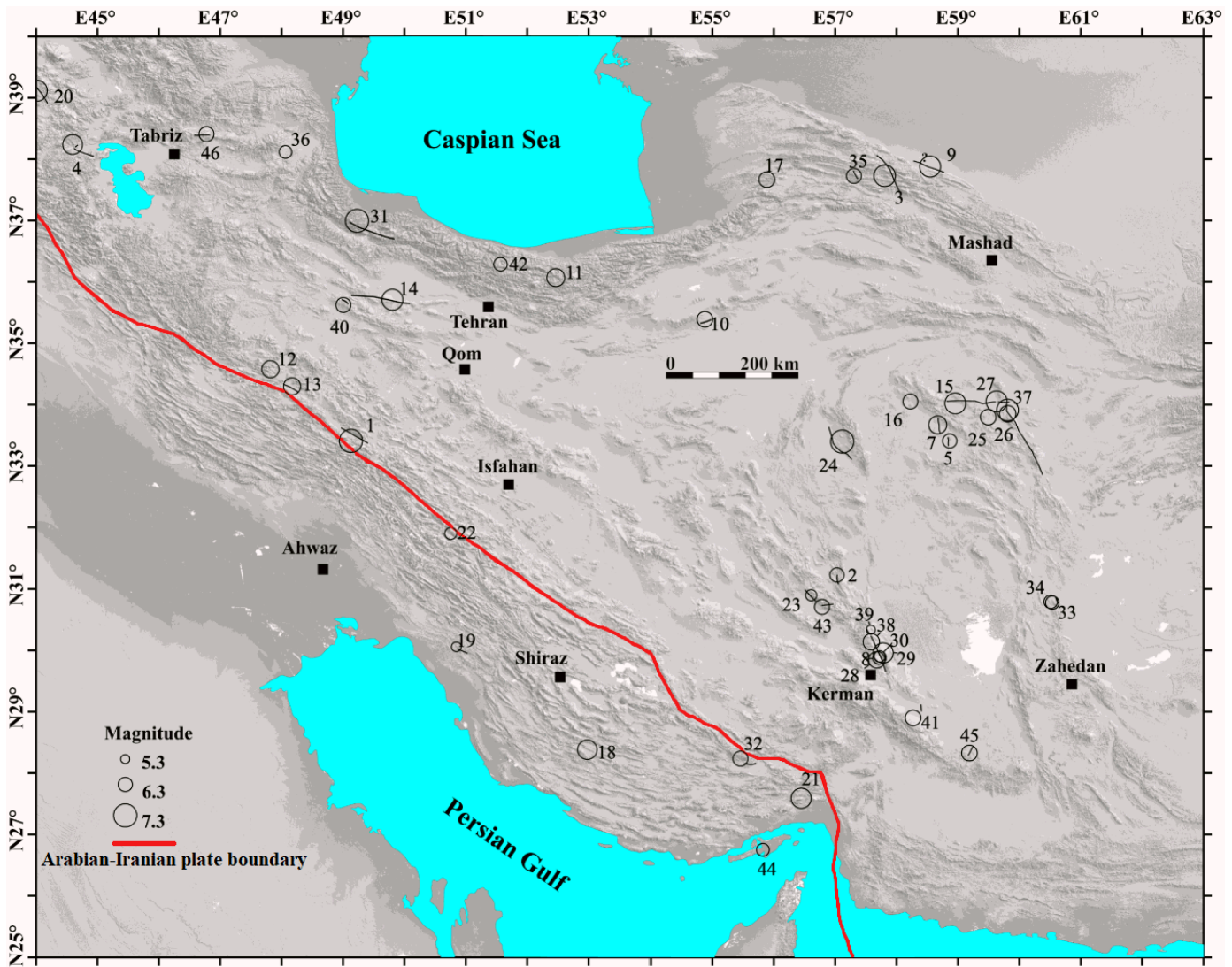
Figure 1. Map showing locations of the earthquake events discussed in this paper. Numbers refer to the event numbers in Table 1. Circles show instrumental epicenters of the events. Where present, surface ruptures are shown in black lines. Instrumental epicenter for the Event No. 6 in Makran region lies outside the map area in Pakistan's coast. Modified after Ghassemi (2016).

and recent years and these earthquakes caused large damages and high loses in human life. Some of these earthquakes are 1909 Silakhor earthquake (Mw7.4), 1945 Makran earthquake (Mw8.0), 1968 Ferdows earthquake (Mw7.4), 1990 Manjil-Rudbar earthquake (Mw7.4), 2013 Ghosht (Mw7.7) earthquake. As a remarkable result, since this study does not include a seismotectonic evaluation or discussion on the earthquake fault rupture hazards in Iran, we did not give many details for the tectonic structures and seismicity of Iran. Thus, a brief and general information is provided in this section. Many readers can reach the knowledge on these subjects from different studies such as Mirzaei et al. (1997), Masson et al. (2005), Djamour et al. (2010), Ghassemi (2016), Öztürk et al. (2017).

Results and discussions for the applications of alternative curve fitting methods

The main purpose of this study is to provide the new, up-to-date and reliable empirical relationships between Mw and Ms, Mw and SRL, MD and Mw, and MD and SRL considering the parameters of surface ruptures for Iranian earthquakes. In this scope, we used different regressions models such as L_1, Orthogonal and Robust for different earthquake faulting mechanisms (strike-slip and thrust or reverse faults) in Iran. As stated in Öztürk (2014),

measurements of surface ruptures generally accompany many large earthquakes and the logarithms of their length are linearly related to earthquake size. These empirical relationships can supply not only fault sizes for a specific magnitude, but also the maximum magnitude based on these sizes. Such types of results are also very useful for geotechnical, seismic risk and hazard, and seismotectonic studies. Many authors estimated these types of relationships between magnitude and earthquake faulting parameters for different parts of the world (e.g., Acharya, 1979; Wells & Coppersmith, 1994; Ambraseys & Jackson, 1998; Öztürk, 2014; Ghassemi, 2016).

Ghassemi (2016) preferred the LS method for the analyses and provided several empirical relationships among Mw, MD and SRL based on different earthquake faulting mechanisms in Iran. In addition to the LS method, three different regression applications were used in this study for the same data in Ghassemi (2016). A comparison between the results of Ghassemi (2016) and the results of present study were given in Table 2. Many researchers in various scientific and engineering disciplines have used these types of curve fitting methods in order to estimate the mathematical formulation for observed data points. If we have an unusual error distribution in a data set, linear LS method cannot provide a good estimate (Öztürk, 2014). Therefore, the minimizing the sum of residual error size will be a more preferable and proper definition of the solution. For this purpose, correlation coefficients of the curve fitting models can be preferred as a trustworthy and acceptable tool in the evaluation

Table 2. Empirical equations with LS method of Ghassemi (2016) and all regression fits derived in this study. SS: strike-slip, TF: thrust or reverse fault, All: including SS+TF. Standard errors for all equations are also given.

Ghassemi (2016) using LS method (L_2)	This study using L_1 method	This study using OR method	This study using RR method	Fault type
Mw=1.020+0.848Ms (R^2=0.852)	Mw=1.182(0.175)+0.818(0.014) Ms (R^2=0.792)	Mw=0.668(0.177)+0.913(0.014)Ms (R^2=0.986)	Mw=1.084(0.174)+0.838(0.014)Ms (R^2=0.829)	All
Mw=5.294(0.217)+0.966(0.151) Log(SRL) (R^2=0.873)	Mw=5.412(0.265)+0.887(0.084) Log(SRL) (R^2=0.660)	Mw=5.570(0.234)+0.822(0.077) Log(SRL) (R^2=0.945)	Mw=5.651(0.238)+0.758(0.075) Log(SRL) (R^2=0.802)	TF
Mw=5.658(0.152)+0.806(0.102) Log(SRL) (R^2=0.797)	Mw=5.289(0.338)+1.057(0.074) Log(SRL) (R^2=0.590)	Mw=4.630(0.359)+1.489(0.078) Log(SRL) (R^2=0.788)	Mw=5.275(0.335)+1.051(0.073) Log(SRL) (R^2=0.576)	SS
Mw=5.523(0.131)+0.870(0.089) Log(SRL) (R^2=0.800)	Mw=5.303(0.309)+0.997(0.056) Log(SRL) (R^2=0.693)	Mw=4.972(0.332)+1.248(0.060) Log(SRL) (R^2=0.778)	Mw=5.396(0.309)+0.950(0.056) Log(SRL) (R^2=0.630)	All
Log(MD)=-2.230(2.432)+0.320(0.364) Mw (R^2=0.114)	Log(MD)=-1.455(0.376) +0.217(0.119)Mw (R^2=0.168)	Log(MD)=-3.002(0.378) +0.419(0.120)Mw (R^2=0.299)	Log(MD)=-1.074(0.363) +0.162(0.115)Mw (R^2=0.059)	TF
Log(MD)=-7.435(1.345)+1.105(0.199) Mw (R^2=0.658)	Log(MD)=-5.767(0.318) +0.824(0.071)Mw (R^2=0.652)	Log(MD)=-6.545(0.324) +0.927(0.073) Mw (R^2=0.779)	Log(MD)=-5.101(0.307) +0.737(0.069)Mw (R^2=0.628)	SS
Log(MD)=-6.320(1.208)+0.938(0.179) Mw (R^2=0.532)	Log(MD)=-3.471(0.338) +0.523(0.062)Mw (R^2=0.308)	Log(MD)=-6.123(0.364) +0.870(0.066)Mw (R^2=0.650)	Log(MD)=-4.142(0.336) +0.623(0.061)Mw (R^2=0.436)	All
Log(MD)=-0.559(0.361)+0.352(0.256) Log(SRL) (R^2=0.240)	Log(MD)=-0.257(0.352) +0.230(0.111)Log(SRL) (R^2=0.186)	Log(MD)=-0.630(0.345) +0.421(0.109)Log(SRL) (R^2=0.411)	Log(MD)=-0.337(0.337) +0.243(0.107)Log(SRL) (R^2=0.132)	TF
Log(MD)=-0.927(0.242)+0.751(0.158) Log(SRL) (R^2=0.616)	Log(MD)=-1.184(0.356) +0.876(0.080)Log(SRL) (R^2=0.554)	Log(MD)=-1.909(0.402) +1.287(0.090)Log(SRL) (R^2=0.697)	Log(MD)=-1.173(0.351) +0.867(0.078)Log(SRL) (R^2=0.494)	SS
Log(MD)=-0.778(0.211)+0.609(0.141) Log(SRL) (R^2=0.457)	Log(MD)=-0.753(0.358) +0.564(0.065)Log(SRL) (R^2=0.294)	Log(MD)=-1.461(0.406) +1.001(0.074)Log(SRL) (R^2=0.588)	Log(MD)=-0.822(0.357) +0.620(0.159)Log(SRL) (R^2=0.355)	All

of the results. As given in Table 2, OR technique provides stronger correlation coefficients than those of the other regression techniques in most estimates expect a few ones. A few poor correlation coefficients, for example R^2=0.299 between MD and Mw, and R^2=0.411 between MD and SRL, were obtained in the estimations. However, these relationships can be suggested in the evaluation of seismic hazard in Iran. Ghassemi (2016) stated that these types of estimations are related to seismicity analyses and these relationships evidently present the probable range of versatility of faulting parameters. In addition, these relationships may supply an understanding to excessive limits for rupture hazard evaluations, and they may give practical insights in paleo-seismological surveys on active faults of the Iranian Plateau.

Table 2 shows the comparison of the empirical relationships calculated in this study with the relationships of Ghassemi (2016). Graphical representations of the whole empirical relationships obtained in this study with three regression models and their confidence intervals between Mw and Ms, Mw and SRL, MD and Mw, and MD and SRL were also plotted in Figures 2, 3 and 4. The first numbers in the equations show the error terms and standard deviations of them (in parentheses) and the second numbers show the slopes of the related distributions and their standard deviations (in parentheses), respectively. Considering three curve fitting methods in addition to the LS method by Ghassemi (2016), the suggested linear-linear, log-linear and log-log relationships were given as the more reliable and optimum by OR method with their R^2 in the following forms:

Mw=0.668(±0.177)+0.913(±0.014)*Ms, for all earthquake mechanisms (R^2=0.986)

Mw=5.570(±0.234)+0.822(±0.077)*$Log(SRL)$, for thrust and reverse faults (R^2=0.945)

Mw=4.630(±0.359)+1.489(±0.078)*$Log(SRL)$, for strike slip faults (R^2=0.788)

Mw=4.972(±0.332)+1.248(±0.060)*$Log(SRL)$, for all fault mechanism (R^2=0.778)

$Log(MD)$=-3.002(±0.378)+0.419(±0.120)*Mw, for thrust and reverse faults (R^2=0.299)

$Log(MD)$=-6.545(±0.324)+0.927(±0.073)*Mw, for strike slip faults (R^2=0.779)

$Log(MD)$=-6.123(±0.364)+0.870(±0.066)*Mw, for all fault mechanism (R^2=0.650)

$Log(MD)$=-0.630(±0.345)+0.421(±0.109)*$Log(SRL)$, for thrust and reverse faults (R^2=0.411)

$Log(MD)$=-1.909(±0.402)+1.287(±0.090)*$Log(SRL)$, for strike slip faults (R^2=0.697)

$Log(MD)$=-1.461(±0.406)+1.001(±0.074)*$Log(SRL)$, for all fault mechanism (R^2=0.588)

The selection of the confidence interval is quite optional and 90%, 95% or 99% are frequently selected in practice, however, 95% is used in general. As stated in Öztürk (2014), confidence intervals are designed to define the statistical properties of data and are may be used practically in many case. Thus, confidence intervals for all types of regression relationships are selected as 95% in this study. As seen in Table 2 and Figures 2 to 4, correlation coefficients of three regression fits are almost the same for the relationships between Mw and SRL in L_2 norm (R^2=0.797) and OR method (R^2=0.788) for strike slip faults, and between Mw and SRL in L_2 norm (R^2=0.800) and OR method (R^2=0.778) for all faulting types. Regional differences in SRL for a given magnitude can be commented in terms of the local changes in geologic, seismic and tectonic activities. These results suggest that seismic efficiency in a region can be dependent on the rupture length or magnitude. Also, if and when the error terms in the linear regression model follow a normal distribution, then the least squares curve fitting technique is better estimator under the most appropriate criteria. Traditional LS estimation is sufficient when error terms show an independently and identically distribution. However, the LS estimates can give bad fit when error distribution is abnormal. The usage of
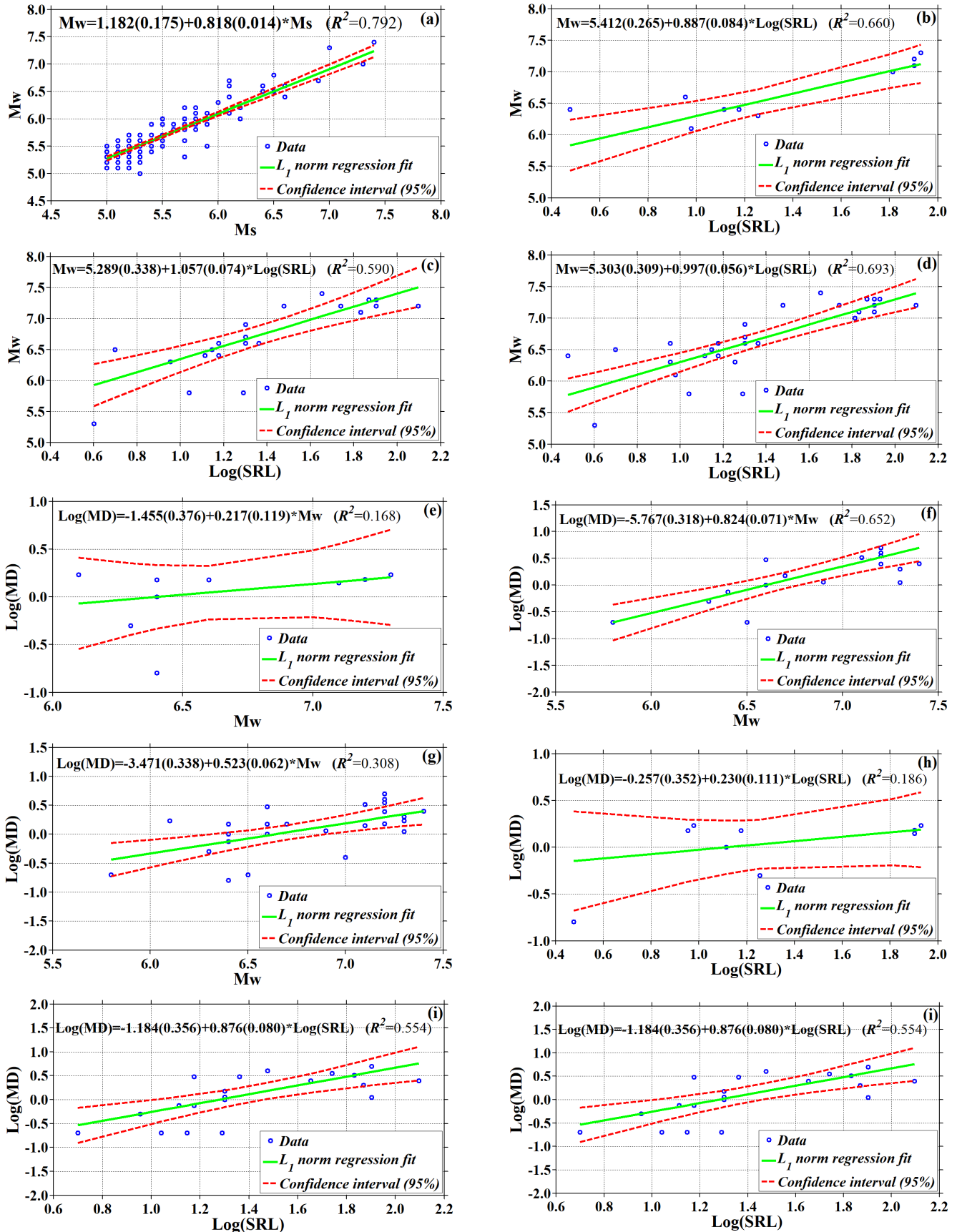
Figure 2. Estimated relationships with Least sum of absolute deviations (L_1) method between (a) Mw and Ms for all faulting types, (b) Mw and SRL for thrust and reverse faults, (c) Mw and SRL for strike-slip faults, (d) Mw and SRL for all faulting types, (e) MD and Mw for thrust faults, (f) MD and Mw for strike slip-faults, (g) MD and Mw for all faulting types, (h) MD and SRL for thrust faults, (i) MD and SRL for strike-slip faults, and (j) MD and SRL for all faulting types.

any curve fitting criterion is preferred when it was doubtful that a small part of the data which is analyzed includes data outliers and thus is not reliable. Curve fitting criterions have the capability of strongly ignoring some bad data points while selecting the majority of data points which put forth the true nature of the data more suitable. The minimizing the sum of residual error magnitude is a reasonable and helpful process for the solution of problem. If the dataset contains a small number of data, the sum of error magnitudes is of particular use and in these situations, these data outliers extremely influence the sum of squared errors criterion. However, the sum of error magnitudes tends to ignore the data outliers on condition that they have relatively few numbers (Öztürk, 2014). Also, Öztürk (2012) suggested for the statistical regression techniques on different dataset that representation of empirical relationships can be made as more appropriate and trustworthy with Least Sum of Absolute Deviations or Robust Regressions for the clustered data while it can be more suitable and reliable with Least Squares or Orthogonal Regressions.

To investigate the maximum surface rupture length (MSRL) and the related maximum credible earthquakes (MCE) in Iran, we studied major earthquake sources in different seismotectonic regions in conjunction with the empirical relationships by OR method of our research. The resulting data were summarized in Table 3. The values are different from those of Ghassemi (2016) in some respects as a result of differences in revised selected earthquake sources as well as difference in the applied regression methods.

As it is noticed for the analysis of data on surface ruptures of thrust (and reverse) faults using OR method (Figure 3b), the regression line has a small slope with a coefficient of 0.945. This may be partly because of rarity of data on surface ruptures associated with thrust events. As a result, it appears that the relationship may underestimate the earthquake magnitude for major (long) thrust faults, especially when value of SRL is used for blind and/or deep-seated faults. To compensate for such apparent deficiency, we have used the positive side of the standard errors in SRL-magnitude equations for estimation of MCE resulting from active thrust faulting in different seismotectonic regions of Iran. Considering our objective to estimate the MCE for each region, the biggest potential seismic sources are picked for each region. Therefore, it should be kept in mind that in dealing with any seismic hazard approach which does not involve such extreme seismic sources, the MCE would be less than the suggested values in this section (i.e. values in Table 3).

A major potential seismic source in the Kopeh-Dagh region is the Maraveh-Tappeh thrust fault (ca. 260 km long). Considering a rupture which involves half the length (130 km) along the geological fault, a maximum magnitude (MCE) of ca. 7.7 is suggested for the Kopeh-Dagh region. The Ashk-Abad fault, which forms the northern border of the Kopeh-Dagh geological province within the Turkmenistan territory, is apparently capable of producing earthquakes of magnitudes as large as 8.2 (e.g. the 1895 Krasnovodsk-Uzun Ada earthquake; Ambraseys, 1997).

North of Central Iran, Alborz and Azerbaijan region include major strike-slip and thrust faults which are capable of great earthquakes. These include the Doruneh, Meyamey, Torud, Astaneh, Mosha, Eivanekey, Ipak, Soltanieh and Tabriz faults, among many other. Length of the above-mentioned faults and/or

their segments attain lengths up to 170 km, and some of them have produced historical earthquakes as large as magnitude 7.9 (the 856 Kumesh earthquake; Ambraseys & Melville, 1982). Therefore, we consider a magnitude 8.0 as the MCE for the northern Central Iran, Alborz and Azerbaijan region.

Within the Zagros region (except for the High Zagros region) emergent geological faults are rare, and therefore most of attempts for earthquake hazard analyses use folds above blind thrust faults as the source for potential earthquakes. As a result, the surface rupture length used in the related calculations, in fact is not really on the surface but it is related to rupture of a blind fault at some depth (see Berberian, 1995). Following the same approach, we have considered the Kabir-Kuh anticline as the biggest seismic source in folded Zagros region with a length of 166 km, and anticipating rupture of half of the related thrust fault (83 km), an MCE of 7.5 is suggested for Zagros folded region. Therefore, we suggest that for the Zagros folds (and related faults) which are longer than the suggested value in Table 3, the related maximum SRL of 83 km to be used for deterministic evaluation of the magnitude attributed to the fold seismic sources.

The Eastern Iran seismotectonic region is dominated by active strike-slip faulting, and less frequent thrust faults. Surface ruptures are highly manifested in this part of the country as compared to other seismic regions. As a result, relationship between SRL and magnitude in this region shows a greater slope, therefore we have used OR method without including an upper standard error limit to estimate MCE. No historical or instrumental earthquake greater than 7.3 has been reported for the region, therefore we used a maximum length of ca. 100 km for OR of all fault type, suggesting a MCE of magnitude 7.5 for the Eastern Iran region.

Central Iran region has a very low seismicity as compared to other parts of the country, and no historical or instrumental record of earthquakes greater than magnitude 6.2 has been reported for this region. Therefore, a MCE of 6.5 appears to be appropriate for the region.

The greatest earthquakes in the Makran region may occur on the subduction zone at depth. The whole length of the subduction zone is ca. 850 km, which if half of the whole length (425 km) is combined with a width of 180 km and a related 20 m of average displacement, gives a MCE of magnitude 9.1.

We made a comparison for two important relationships suggested for global and Iranian earthquakes, including SRL-Mw (Figure 5) and SRL-MD (Figure 6) relationships. Slope of our regression line for SRL-Mw relationship greatly conforms to those of Nowroozi (1985) and Wells and Coppersmith (1994), however it is less than that of Ghassemi (2016) (Figure 5). A similar comparison between the graphs for SRL-MD indicates that slope calculated in this study is very similar to those of Nowroozi (1985) and Wells & Coppersmith (1994), however the regression has a lower intercept value than those of aforementioned studies. As seen in Figure 6, both slope and intercept values of our study are smaller than that of Ghassemi (2016).

Many studies were made for surface rupture hazard in Iran and some empirical relationships were provided among different earthquake faulting parameters by different researchers such as Berberian (1976), Nowroozi (1985), Wells & Coppersmith, (1994), Berberian (2014), Ghassemi (2016). The only comprehensive analysis of Iranian earthquakes was made by Nowroozi (1985). Also, more detailed quantitative assessments of surface rupture hazard of Iranian earthquakes including historical and instrumental period were provided with Berberian (2014) and Ghassemi (2016). When we compared the empirical relationships estimated in this study with the most up-to-date relationships of Ghassemi (2016), the results in this study present higher R^2 among magnitude and different faulting parameters in OR method. Correlation coefficients of the relationships in Ghassemi (2016) changes between 0.114 and 0.873 with L_1 norm, whereas correlation coefficients with OR technique in present study vary from 0.299 to 0.986. From the observations in this study and in Ghassemi (2016), we can conclude that these estimations for OR method can be thought as good and reliable relationships, and one can use them for the other seismicity analyses or rupture hazard evaluations of the Iranian Plateau. As a remarkable result, we did not make a detailed discussion on the earthquake fault rupture hazards in Iran since this study does not aim to evaluate and discuss this subject. Many details can be found in Ghassemi (2016) for the insights of earthquake surface rupture hazards. For this reason, a short and general comparison of three different curve fitting techniques is provided for the data of Ghassemi (2016). Furthermore, the effectiveness of the different estimation techniques was examined in this study and the relative predictive validity of four regression techniques was compared for different faulting parameters. These detailed analyses show that empirical
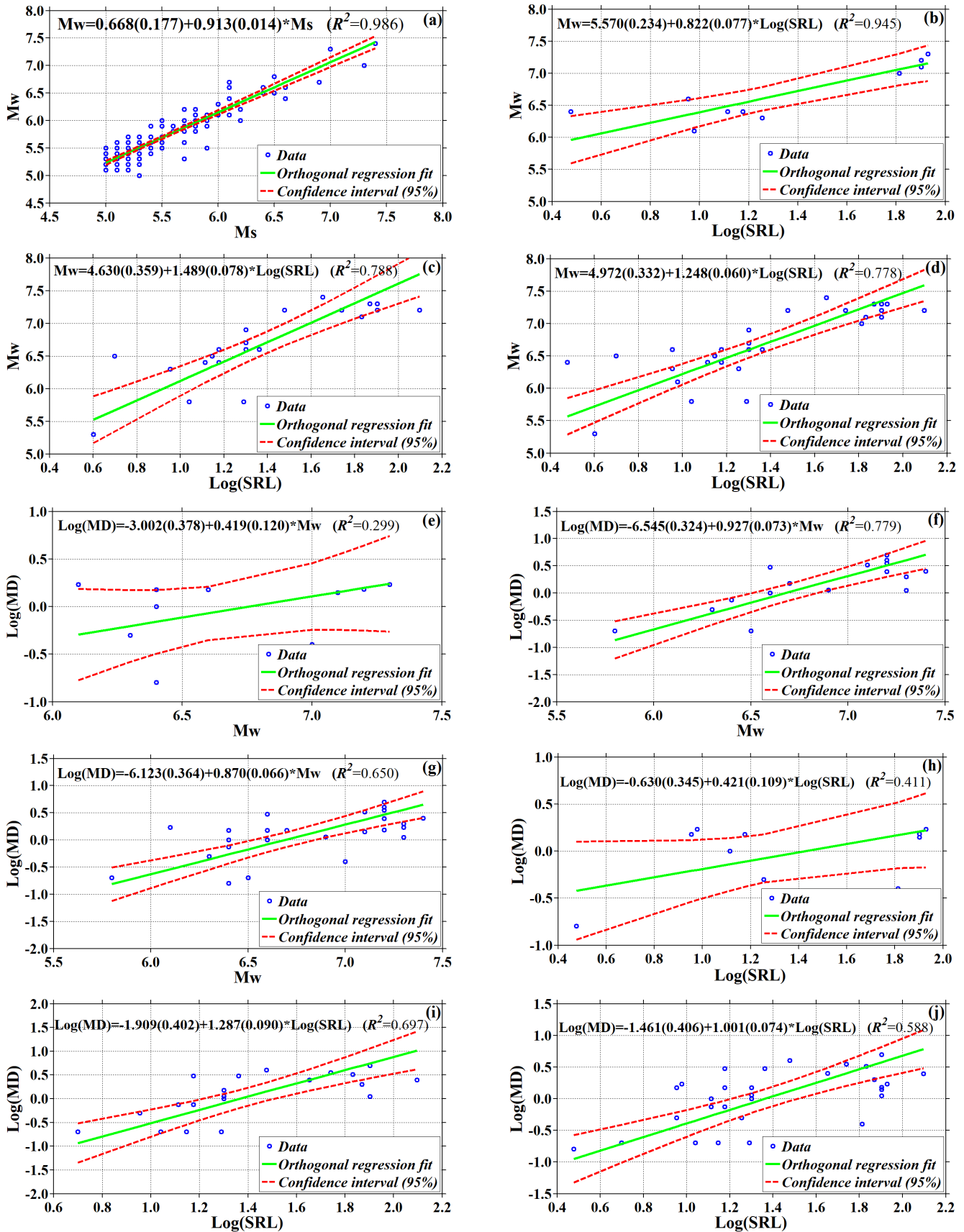
Table 3. Maximum surface rupture length (MSRL), maximum surface displacement (MSD), and related maximum credible earthquakes (MCE) suggested for different seismotectonic regions of Iran. See text for further explanation.

Region	MSRL (km)	MSD (m)			MCE (Mw)
		TF	SS	All	
Kopeh-Dagh	130	1.8	6.4	4.5	7.7
Northern Central Iran, Alborz, Azerbaijan	170	2.0	9.1	5.9	8.0
Zagros	83	1.5	3.6	2.9	7.5
Eastern Iran	100	1.6	4.6	3.5	7.5
Central Iran	16	0.8	0.4	0.6	6.5
Makran	425*	20	-	-	9.1

* Length quoted for Makran is rupture length at depth

Figure 3. Estimated relationships with Orthogonal regression (OR) method between (a) Mw and Ms for all faulting types, (b) Mw and SRL for thrust and reverse faults, (c) Mw and SRL for strike-slip faults, (d) Mw and SRL for all faulting types, (e) MD and Mw for thrust faults, (f) MD and Mw for strike slip-faults, (g) MD and Mw for all faulting types, (h) MD and SRL for thrust faults, (i) MD and SRL for strike-slip faults, and (j) MD and SRL for all faulting types.
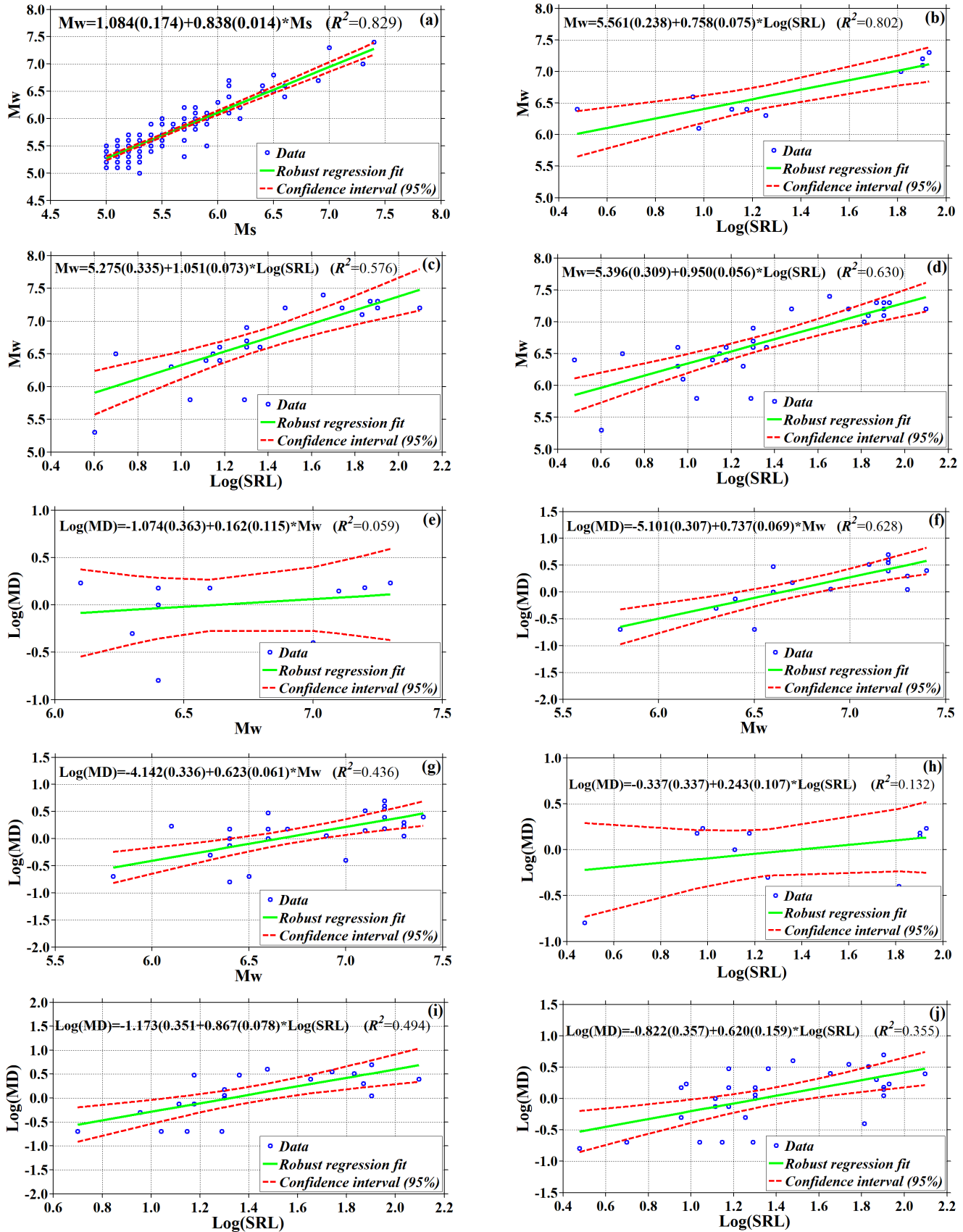
Figure 4. Estimated relationships with Robust regression (RR) method between (a) Mw and Ms for all faulting types, (b) Mw and SRL for thrust and reverse faults, (c) Mw and SRL for strike-slip faults, (d) Mw and SRL for all faulting types, (e) MD and Mw for thrust faults, (f) MD and Mw for strike slip-faults, (g) MD and Mw for all faulting types, (h) MD and SRL for thrust faults, (i) MD and SRL for strike-slip faults, and (j) MD and SRL for all faulting types.
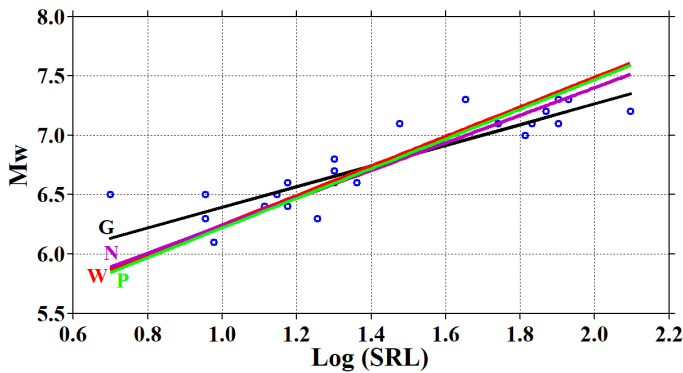
Figure 5. Comparison among different regressions of moment magnitude (Mw) versus surface rupture length (SRL in km) for Iranian earthquakes of all rupture mechanisms (26 events). P: present study, W: Wells and Coppersmith (1994), N: Nowroozi (1985), G: Ghassemi (2016).
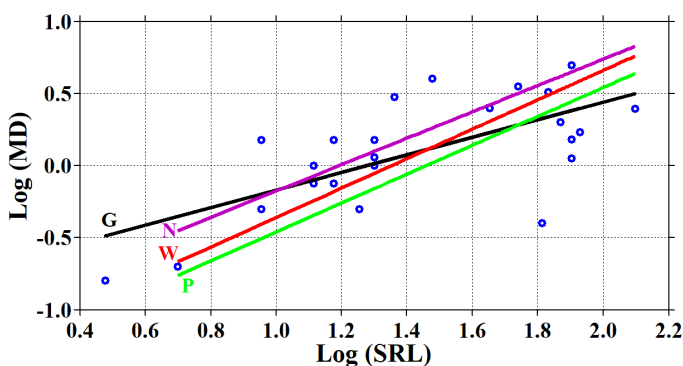
Figure 6. Comparison among different regressions of surface rupture length (SRL in km) versus maximum displacement (MD in m) for Iranian earthquakes of all rupture mechanisms (24 events). P: present study, W: Wells and Coppersmith (1994), N: Nowroozi (1985), G: Ghassemi (2016).

relationships estimated by OR method in the present study can be thought and used as more appropriate and more trustworthy for Iranian earthquakes and the other mentioned applications. We finally hypothesized that OR estimation technique can provide more stabilized interpretations for the rupture properties, variability of the earthquake fault rupture geometry and kinematics by using these empirical relationships among different faulting parameters.

Conclusions

This study focused on the applications of different regression techniques and a comparison of the empirical relationships of the data set on the surface ruptures of Iranian earthquakes was made. Using a database of Iranian earthquakes, some empirical relationships were estimated among different parameters such as moment magnitude, surface wave magnitude, surface rupture length and maximum displacement for different faulting types. For this purpose, three curve fitting methods are applied as (a) Least Sum of Absolute Deviations, (b) Robust Regression and (c) Orthogonal Regression. In order to decide how we can select the best curve fitting method for a given data set, we calculated separately the correlation coefficients for the thrust or reverse faults, strike slip faults and all faulting types since it is an easy applicable and a practical technique. For the analyses, 46 strong and large earthquakes with magnitudes ≥ 5.8 between 1900 and 2017 were used. In order to estimate the statistical properties of different data sets, 95% confidence intervals were used for each regression relation. Using the above-mentioned techniques, we derived linear relationships among different faulting parameters for Iranian earthquakes. Correlation coefficients varies from 0.299 to 0.986 with Orthogonal regression, from 0.168 to 0.792 with L_1 regression, from 0.059 to 0.829 with Robust regression. These results show that Orthogonal regression fits give more

powerful coefficients and can be thought more suitable equations as compared to the other curve fitting techniques.

Resulting equations are proposed as more suitable and reliable in the estimation of the maximum surface rupture length, maximum surface displacement and associated maximum credible earthquakes for different areas of Iran. As a general conclusion, detailed analyses for the data sets of different faulting mechanisms show that more appropriate, more trustworthy more up-to-date representations of empirical relationships can be done with Orthogonal regression. In addition, these types of relationships may supply significant insights in the estimation of the maximum surface rupture length, maximum surface displacement, and associated maximum credible earthquakes for different seismotectonic regions of Iran. We suggested that these relationships may also provide important outputs in the estimation of earthquake magnitudes in paleo-seismological studies.

References

Abdelkader, G., Ali, L., & Rachida, R. (2010). Robust nonparametric estimation for spatial regression. *Journal of Statistical Planning and Inference*, 140, 1656-1670.
Acharya, H. K. (1979). Regional variations in the rupture-length magnitude relationships and their dynamical significance. *Bulletin of the Seismological Society of America*, 69(6), 2063-2084.
Ambraseys, N. N. (1997). The Krasnovodsk (Turkmenistan) earthquake of 8 July 1895. *Journal of Earthquake Engineering*, 1(2), 293-317.
Ambraseys, N. N., & Jackson, J. A. (1998). Faulting associated with historical and recent earthquakes in the Mediterranean region. *Geophysics Journal International*, 133, 390-406.
Ambraseys, N. N., & Melville, C. P. (1982). *A history of Persian earthquakes*. Cambridge University Press, Cambridge.
Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Second ed., Wiley, New York.
Arthanari, T. S., & Dodge, Y. (1981). *Mathematical Programming in Statistics*. Wiley, New York.
Berberian, M. (1976). Documented earthquake faults in Iran. *Geological Survey of Iran*, 39, 143-186.
Berberian, M. (1995). Master "blind" thrust faults hidden under the Zagros folds: active basement tectonics and surface morphotectonics. *Tectonophysics*, 241, 193-224.
Berberian, M. (2014). *Earthquakes and coseismic faulting on the Iranian Plateau: A historical, social and physical approach*. Elsevier Series Development in Earth Surface Processes, 17, p. 628.
Blattberg, R. C., & Sargent, T. (1971). Regression with non-Gaussian stable disturbances: Some sampling results. *Econometrica*, 39, 501-510.
Boyer, B. H., McDonald, J. B., & Newey, W. K. (2003). A comparison of partially adaptive and reweighted least squares estimation. *Econometric Reviews*, 22, 115-134.
Branham, R. L. Jr. (1982). Alternatives to least-squares. *Astronomical Journal*, 87, 928-937.
Cadzow, J. A. (2002). Minimum ℓ_1, ℓ_2 and ℓ_∞ norm approximate solutions to an overdetermined system of linear equations. *Digital Signal Processing*, 12, 524-560.
Carroll, R. J., & Ruppert, D. (1996). The use and misuse of orthogonal regression estimation in linear errors-in-variables models. *The American Statistician*, 50, 1-6.
Djamour, Y., Vernant, Ph., Bayer, R., Nankali, H. R., Ritz, J. F., Hinderer, J., Hatam, Y., Luck, B., Moigne, N. L., Sedigh, M., & Khorrami, F. (2010). GPS and gravity constraints on continental deformation in the Alborz mountain range, Iran. *Geophysical Journal International*, 183, 1287-1301.

Durio, A., & Isaia, E. D. (2003). Parametric regression models by minimum L_2 criterion. A study on the risks of fire and electric shocks of electronic transformers. *Developments in Applied Statistics*, 19, 69-83.

Fama, E. F., (1965). Portfolio analysis in a stable Paretian market. *Management Science*, 11, 404-419.

Field, C. A. (1997). Robust regression and small sample confidence intervals. *Journal of Statistical Planning and Inference*, 57, 39-48.

Fuller, W. A. (1987). *Measurement Error Models*. John Wiley and Sons, New York.

Ghassemi, M. R. (2016). Surface ruptures of the Iranian earthquakes 1900-2014: Insights for earthquake fault rupture hazards and empirical relationships. *Earth Science Reviews*, 156, 1-13.

Giloni, A., & Padberg, M. (2002). Alternative methods of linear regression. *Mathematical and Computer Modeling*, 35, 361-374.

Giloni, A., Simonoff, J. S., & Sengupta, B. (2006). Robust weighted LAD regression. *Computational Statistics and Data Analysis*, 50, 3124-3140.

Greene, W. (1997). *Econometric Analysis*. Third ed. Prentice-Hall, Englewood Cliffs, NJ.

Hartmann, C., Vankeerberghen, P., Smeyers-Verbeke, J., & Massart, D. L. (1997). Robust orthogonal regression for the outlier detection when comparing two series of measurement results. *Analytica Chimica Acta*, 344, 17-28.

Heo, J. H., Kho, Y. W., Shin, H., Kim, S., & Kim, T. (2008). Regression equations of probability plot correlation coefficient test statistics from several probability distributions. *Journal of Hydrology*, 355, 1-15

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35, 73-101.

Huber, P. J. (1981). *Robust Statistics*. Wiley, New York.

Huber, P. J. (1987). *The place of the L_1 norm in robust estimation*. In: Dodge, Y. (Ed.), Statistical data analysis based on the L_1 norm and related methods. North-Holland, Amsterdam, pp. 23-33.

Javidfakhr, B., Bellier, O., Shabanian, E., Siame, L., Léanni, L., Bourlès, D., & Ahmadian, S. (2011). Fault kinematics and active tectonics at the southeastern boundary of the eastern Alborz (Abr and Khij fault zones): Geodynamic implications for NNE Iran. *Journal of Geodynamics*, 52, 290-303.

Kendal, M. G., & Stuart, A. (1979). *The advanced theory of statistics*. 4th edition, vol. 2, Hafner, New York.

Leng, L., Zhang, T., Kleinman, L., & Zhu, W. (2007). Ordinary Least Square Regression, Orthogonal Regression, Geometric Mean Regression and their Applications in Aerosol Science. *Journal of Physics*, Conference Series, 78 (1), article ID. 012084, 5 pp.

Madansky, A. (1959). The fitting of straight lines when both variables are subject to error. *Journal of the American Statistical Association*, 54, 173-205.

Mandelbrot, B. (1967). The variation of some other speculative prices. *Journal of Business*, XL, 393-413.

Maronna, R. (2005). Principal components and orthogonal regression based on robust scales. *Technometrics*, 47, 264-273.

Masson, F., Chery, J., Hatzfeld, D., Martinod, J., Vernant, P., Tavakoli, F., & Ghafory-Ashtiani, M. (2005). Seismic versus aseismic deformation in Iran inferred from earthquakes and geodetic data. *Geophysical Journal International*, 160, 217–226.

Mirzaei, N., Gao, M., & Chen, Y. T. (1997). Seismic source regionalization for seismic zoning of 565 Iran: major seismotectonic provinces. *Journal Earthquake Prediction Research*, 7, 465-495.

Nowroozi, A. A. (1985). Empirical relations between magnitudes and fault parameters for earthquakes in Iran. *Bulletin Seismological Society of America*, 75, 1327-1338.

Öztürk, S. (2012). An analysis on the application of statistical regression methods for different data set. *Gümüşhane University Journal of Science and Technology Institute*, 2(2), 55-67.

Öztürk, S. (2014). A new empirical relation between surface wave magnitude and rupture length for Turkey earthquakes. *Earth Sciences Research Journal*, 18(1), 15-26.

Öztürk, S., Ghassemi, M. R., & Sarı, M. (2017). Empirical relationships between the surface wave magnitude and moment magnitude for the Iranian earthquakes, International Conference on Mathematics and Engineering, 10-12 May, 2017, (ICOME 2017), 127 pp., İstanbul, Turkey.

Renaud, O., & Victoria-Feser, M. P. (2010). A robust coefficient of determination for regression. *Journal of Statistical Planning and Inference*, 140, 1852-1862.

Rosenberg, B., & Carlson, D. (1977). A simple approximation of the sampling distribution of least absolute residuals regression estimates. *Communications in Statistics-Simulu Computa*, 136(4), 421-437.

Rousseeuw, R. J., & Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

Sen, A., & Srivastava, M. (1990). *Regression analysis: Theory, methods, and applications*. Springer-Verlag, New York.

Shi, M., & Lukas, M. A. (2005). Sensitivity analysis of constrained linear L_1 regression: perturbations to response and predictor variables. *Computational Statistics and Data Analysis*, 48, 779-802.

Sinha, S. K., Field, C. A., & Smith, B. (2003). Robust estimation of nonlinear regression with autoregressive errors. *Statistics and Probability Letters*, 63, 49-59.

Spiess, M., & Hamerle, A. (2000). A comparison of different methods for the estimation of regression models with correlated binary responses. *Computational Statistics and Data Analysis*, 33, 439-455.

Stefanski, L. A. (1991). A Note on high-breakdown estimators. *Statistics and Probability Letters*, 11, 353-358.

Şentürk, D., & Nguyen, D. V. (2006). Estimation in covariate-adjusted regression. *Computational Statistics and Data Analysis*, 50, 3294-3310.

Weisberg, S. (1985). Applied Linear Regression. Second edition. John Wiley and Sons, New York.

Wells, D., & Coppersmith, J. K. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America,* 84(4), 974-1002.