Asim, K. M., Javed, F., Hainzl, S., Iqbal, T. (2019): Fault Parameters-Based Earthquake Magnitude Estimation Using Artificial Neural Networks. - Seismological Research Letters, 90, 4, 1544-1551.

https://doi.org/10.1785/0220190051

Fault parameters-based earthquake magnitude estimation using Artificial Neural Networks

Khawaja M. Asim[a], Farhan Javed[a,b], Sebastian Hainzl[c], Talat Iqbal[a]
[a]Centre for Earthquake Studies, National Centre for Physics, Islamabad, Pakistan
[b]International Centre for Theoretical Physics, Trieste, Italy
[c]GFZ German Research Centre for Geosciences, Potsdam, Germany

Corresponding Author:
Khawaja M. Asim
Researcher,
Centre for Earthquake Studies,
National Centre for Physics
P.O. Box No. 2141,
Islamabad – 44000, Pakistan.
Tel: + 92-345-9591700
Email: asim.khawaja@ncp.edu.pk, asimkhawaja786@gmail.com
ORCID: 0000-0002-6196-4503

Abstract

In this study, a computer-aided methodology is proposed to estimate the earthquake magnitude based on fault parameters. So far, log-linear regression equations are separately employed for each fault parameter. However, this can lead to inconsistent magnitude predictions because non-linear parameter correlations are ignored and those parametric functions cannot take into account potential deviations from log-linear scaling. In order to address the aforementioned deficiencies, we employ Artificial Neural Network (ANN) to estimate the magnitude of earthquakes simultaneously using all available fault parameters such as rupture length and width, thereby excluding the chances of inconsistent estimations. Our evaluation of M>=5 earthquakes shows that the predictions from the proposed methodology outperform the regression equation-based predictions in terms of mean absolute error and root mean square error. Furthermore, the pictorial view of the performance also demonstrates the strength of ANN to identify and reproduce, without any initial assumption, systematic deviations from the log-linear scaling of earthquake magnitudes as a function of the fault parameters.

Keywords: Fault parameters, Artificial Neural Network, Earthquake magnitude estimation, Seismic hazard Assessment

Introduction

Seismic hazard assessments rely on estimates of the maximum possible magnitude (M_{max}) of earthquakes. However, instrumental earthquake catalogs are usually too short to cover full seismic cycles and thus do not include the largest possible events. Therefore, geologically mapped faults or paleo-earthquake studies are often used to estimate M_{max}. For that purpose, the relation between the earthquake magnitude (M, hereinafter refers to the moment magnitude) and fault parameters, such as rupture length (L), rupture width (W), area (A), and slip must be known.

In theory, the moment magnitude is simply a function of the shear modulus, the mean slip, and the rupture area. However, the average earthquake slip on the rupture area is usually not known and empirical estimations based only on the rupture dimensions might differ significantly from the true value. Many authors proposed empirical scaling relationships between seismic moment and fault area (Thatcher and Hanks 1973, Kanamori and Anderson 1975, Kanamori 1977) and fault length or width (Scholz 1982, Romanowicz 1992, Romanowicz and Rundle 1993). Previous research studies used regression analysis to develop such empirical relationships between fault rupture parameters and magnitude for large worldwide historic earthquakes (Wells and Coppersmith 1994, Mai and Beroza 2000, Henry and Das 2001, Leonard 2010). Currently, the empirical relations of Wells and Coppersmith (1994) (WC-94) are commonly employed to estimate earthquake magnitudes, but these relations are not self-consistent because the regression equations for the earthquake magnitude are estimated independently for the different fault parameters, which limit their applicability.

It is also noted that data for many recent large earthquakes were missing during the time of the aforementioned studies. Furthermore, the conventional methodologies based on regression equations cannot account for non-linear correlations between the rupture parameters. For a given fault parameter (e.g. L), these regression equations simply predict one magnitude value independently of the values of the other fault parameters (e.g. W) and the observed deviations are taken as random fluctuations. Figure 1 shows the fit of the WC-94 regression equation to actual earthquake magnitudes as function of L. The data set used in Figure 1 consists of the combined WC-94 and SRCMOD data sets described in Section 2. The predictions simply follow a line, while true values widely scatter around it with some systematic trends. For example, the WC-94 regression equation fails to correctly predict magnitudes $M \geq 8.0$, thus underestimating the seismic hazard in that range. In order to improve hazard estimations, a methodology capable of incorporating the non-linear dependence of earthquake magnitudes on fault parameters is highly desirable. We investigate the application of intelligent computing algorithms as one solution to this problem.

Machine learning is a branch of computer science that has the ability to identify and extract meaningful, hidden relations from data. These learned relations are then used to make predictions for unseen data (Reyes, et al. 2013). In the recent past, the use of machine learning techniques in the field of seismology and earth sciences has increased (Asim, et al. 2017, Asim, et al. 2018, Rouet-Leduc, et al. 2017, DeVries, et al. 2018, Bergen, et al. 2019, Asencio-Cortés, et al. 2016, Morales-Esteban, et al. 2013, Tareen, et al. 2019). This interdisciplinary approach has already provided new insights and increased predictability for different challenging data sets (Kong, et al. 2018). In this paper, we test its applicability to the problem of magnitude estimation based on sets of fault parameters. In particular, we employ Artificial Neural Network (ANN) for the mapping between fault parameters and the corresponding earthquake magnitude. We split the collected earthquake data, consisting of historical and instrumental earthquakes compiled by

Wells and Coppersmith (1994) and the finite-source rupture model database SRCMOD (see Section 2), into training and test data, showing that the proposed methodology provides improved, robust, and self-consistent estimations of earthquake magnitude by simultaneously taking into account the knowledge of all available fault parameters.

The analysis is divided into two parts. In Section 3.1, we estimate the magnitude of the target events by means of the regression equation proposed by Wells and Coppersmith (1994). Here we also analyze regression equations which are recomputed based on the new and increased data collection including many events occurred after 1993. In Section 3.2, Artificial Neural Network (ANN) is developed for training data and tested for unseen data. The results are then discussed and compared to the conventional regression-based methodologies in Section 4.

Earthquake Data

We analyze data from past earthquakes with magnitude $M \geq 5.0$, which are collected from both the WC-94 catalog and SRCMOD database.

WC-94 Catalog:

The WC-94 catalog of past large earthquakes was compiled by Wells and Coppersmith (1994). In this publication, a total of 244 events, which occurred until 1993, are listed with mixed focal mechanisms consisting of both strike-slip and dip-slip earthquakes. The fault parameters of these events were estimated either by paleoseismological and seismological studies, aftershock distributions, or geodetic modeling of surface deformations. We selected those events from the catalog which have both fault length L and width W information. If both subsurface and surface rupture lengths were provided for a single event, we chose the subsurface length. Our selection criterion yields a total of 180 earthquake ruptures with M≥5.0 for our analysis. Out of these cases, 95 include information about the maximum slip of the rupture (See Table S1(a)).

SRCMOD Database:

A data set consisting of more recent earthquakes is maintained by Martin Mai and his colleagues. The SRCMOD database collects finite-source rupture models (http://equake-rc.info/SRCMOD/searchmodels/allevents/) (Last accessed on April 10, 2019), which are delivered by different research teams. These slip models are obtained from the inversion of seismic, geodetic, tsunami and other geophysical techniques with variable resolutions. At the time of our access, the SRCMOD database contained 347 slip models for a total of 178 different earthquakes. Because of their limited resolution, we excluded slip models which were solely derived from tsunami data. The remaining 316 slip models are used in our analysis, after exclusion of M<5.0 earthquakes. For each case, the dimensions of the fault planes on which the slip had been inverted are provided in the SRCMOD database. However, the dimension for the inversion can be much larger than the actual dimension of the earthquake slip region. Therefore, each model and its associated fault parameters were manually inspected (see Table S1(b)). As an example, the slip model of Motagh, et al. (2010) for the 2007 M7.8 Tocopilla earthquake shows significant slip only in a discrete part of the assumed fault plane, which is 349 km long and 180 km wide. The recomputed rupture length and width of the actual slip area are 270 km and 110 km, respectively, reducing the area by a factor of approximately two in this case. Note that we calculate the rupture area simply by the product of L and W.

Research Methodology

To relate fault parameters to earthquake magnitudes, we firstly employ the traditional regression-based methodology and then apply the new ANN-based approach. This allows us to properly compare the results of both approaches in terms of consistency, accuracy and robustness.

Regression Analysis

Regression analysis is carried out to derive log-linear relations between fault parameters and earthquake magnitudes. However, these relations are separately derived for each individual fault parameter, i.e. independent regression equations relating either L or W to the earthquake magnitude are obtained. In this regard, the regression equations of Wells and Coppersmith (1994) are widely used, which were derived for the WC-94 data set (including all mechanisms) as follows:

$$M = 4.38 + 1.49 \log(L) \quad (1)$$

$$M = 4.06 + 2.25 \log(W) \quad (2)$$

To take advantage of the extended data set, we also perform our own regressions for the combined WC-94 and SRCMOD data collection described in Section 2. Our derived new empirical relations between M, L and M, W, respectively, are given by

$$M = 4.33 + 1.57 \log(L) \quad (3)$$

$$M = 4.7 + 1.7 \log(W) \quad (4)$$

Figure 2 shows the fit of these regression lines to the empirical earthquake data.

Artificial Neural Networks

Artificial Neural Networks (ANN) are inspired from biological neural networks consisting of neurons and weighted connections between different layers of neurons (Hassoun 1995). Earthquake magnitude estimation using fault features is treated as a regression problem and solved using ANN. The network is trained on a part of the known data by providing fault features on the input layer and the corresponding actual earthquake magnitudes on the output layer. The input layer leads to the hidden layer through weighted connections and is further passed to the output neurons. The weights of connections in neurons are either excitatory or inhibitory. An output value is received at the output neuron through the processing of fault parameters. The error between the value received at the output neurons and the actual earthquake magnitude is then calculated and propagated backwards in order to adjust and tune the weighted connections accordingly. This process of adjusting connection weights based upon known data is called “learning”. The explained topology of ANN is called feed forward neural network and the learning process is referred as back propagation.

The number of input neurons is equal to the number of fault parameters provided as input. In regression problems, only a single neuron is kept in the output layer. In addition to the input and output layers, a single hidden layer is used. In this case, we chose seven neurons in the hidden layer based upon performance during training and cross-validation. As activation function, we used tan-sigmoid,

$$f(x) = \tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \ , \quad (5)$$

which is widely used as an activation function in shallow neural networks. We also found that it performs better than linear and sigmoid activation functions in our case.

The trained ANN model maps the given set of fault parameters, such as L, W, A, and maximum slip S (if available) to the actual earthquake magnitude in an optimized way and is then capable to predict the earthquake magnitude for unseen data. In other words, ANN learns from the known available data and develops a relation between potential inputs (fault parameters) and corresponding output (earthquake magnitude).

In this paper, we use two different sets of fault input data for the ANN approach. In the first scenario, we restrict the fault parameter set to the geometrical values, namely L, W and the rupture area (L-W-A). In practical applications, only L-W-A might be available, while the maximum slip value is often not known. However, we also developed a second ANN model simultaneously utilizing all fault parameters including maximum slip (L-W-A-S). Although the second scenario might have fewer applications for hazard assessment, it is interesting to evaluate and compare the overall prediction performance of ANN for both cases, because such a comparison can highlight the information gain due to additional input values. However, for the comparison of our ANN model results with conventional regression results, we concentrate on the L-W-A approach as the most practical one.

In order to assess the performance of ANN, we predict earthquake magnitudes from the test data. Only a portion of the available data set (training data set) is fed to ANN for learning and predictions are obtained for unseen data (test data set). In this study, a cross validation strategy is employed to test the ANN predictions on the combined WC-94 and SRCMOD data set. The *k*-fold cross validation approach is widely applied in demonstrating the performance of classification and regression techniques (Wong 2015). In particular, it is expected to capture the general properties in cases of limited data samples. We choose the specific value of $k=10$, i.e. a 10-fold cross-validation (Idris, et al. 2017). In this procedure, we divided the entire data set into

10 non-overlapping subsets. One of these 10 subsets was reserved for independent standalone testing, while the remaining 9 subsets were used for model training. We repeated the process until all 10 subsets were separately employed once for testing. Therefore, we trained 10 different ANN models separately and obtained earthquake magnitude predictions for every sample available in the data set.

Results and discussion

In the following, we describe the results of the new ANN method and compare them with those obtained by the conventional methodologies to demonstrate that the computer-aided technique has the potential to improve seismic hazard assessments, especially magnitude predictions based on fault parameters.

The performance of the magnitude estimation is expressed using the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE). These errors are computed between actual and estimated magnitudes for the test data sets to quantify the overall performance of the ANN. In contrast, for the conventional regression equations, the errors are calculated partly (WC-94 relations) or fully (Eqs.3,4) for the same data for which the models have been developed. The calculated errors are defined as:

$$\mathrm{MAE}=\frac{\sum_{i=1}^{n} | M_{Predicted_i} - M_{Actual_i} |}{\mathrm{n}} \tag{6}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{n} (M_{Predicted_i} - M_{Actual_i})^2}{n}} \tag{7}$$

Below we show that the proposed ANN-based estimates are superior to the regression-based magnitude estimations with respect to consistency and performance.

Self-consistency of ANN-based methodology

Regression relations, as presented in Eq. 1-4, determine the earthquake magnitude as function of a single fault parameter, i.e. L or W in this case. A major issue in such approaches is the lack of self-consistency. Regression relations tend to yield different magnitude estimations for the same earthquake if either the regression equation for L or W is used. For example, for our data set, the RMSE-value of the difference $M(L_i)–M(W_i)$ is RMSE = 0.221 for the WC-94 and RMSE = 0.162 for the new regression equations, respectively. However, in the proposed ANN-based approach, all given fault parameters are employed simultaneously to estimate the earthquake magnitude. The simultaneous use of all fault parameters eschews different magnitude estimations for the same sample, thereby providing a self-consistent earthquake magnitude estimation methodology by taking potential non-linear parameter correlations into account.

Table 1 demonstrates once more the inconsistency of regression equation-based estimations, which show different errors for the equations based on L and W. On the contrary, a trained ANN simultaneously takes all given fault parameters into account and provides a single prediction.

Performance of ANN-based methodology

The individual magnitude predictions of the ANN-model and the regression equations for the whole data set are provided in the supplementary material (Please see Tables S2(a), S2(b)). A summary of the model performance is provided in Figure 3 in terms of histograms of the residuals and in Table 1 in terms of RMSE and MAE, computed between predicted magnitudes and actual earthquake magnitudes. The results for the test data sets highlight the robustness of ANN-based predictions, demonstrating that ANN has the ability to show decent performance across the whole data set.

It is evident from Figure 3 and Table 1 that the L-W-A scenario leads to significantly decreased errors compared to the regression equations. The availability of the maximum slip value S in addition to L, W and A further improves the results of the ANN-approach in terms of RMSE and MAE. Figure 4 shows scatter plots of the actual earthquake magnitudes and ANN-based predictions. We detect systematic deviations from the exponential relations assumed in the regression equations. The model clearly depicts the noticeable scale breaks in the relations between M, L and M, W. For $L <= 80$ km, the magnitudes scale approximately linearly with the logarithm of length. However, between 80 and 180 km, the slope becomes smaller, then increases again for $L \geq 180$ km (corresponding to $M \geq 7.7$). A quite similar behavior is observed for the dependence of magnitude on fault width. The ANN-predictions reproduces the two observed kinks in the scaling at rupture widths of approximately 25 km and 100 km.

We also analyzed the results of the ANN with respect to the predictions for different rupture mechanisms. In Figure 5, the results are separately shown for the strike-slip and dip-slip events in the data set. The general trend of both rupture types is well reproduced. However, some outliers with magnitudes significantly higher than the average value for the given rupture length are observable in the case of strike-slip events. Earthquakes with largely erroneous predictions (encircled in Figure 5a) are mostly related to historical earthquakes in the WC-94 catalog, in particular those which lack subsurface length information. Another example is the 1920 Ms8.5 Gansu, China, earthquake, one of two strike-slip earthquakes with magnitude >8.0 in the data set. The largest strike-slip event is the 2012 M8.7 Sumatra intra-plate earthquake. Besides the missing subsurface length information for the Gansu event, the erroneous prediction of the two largest events can be explained by the fact that both events can be rarely seen as a single fault rupture, because they ruptured several subfaults (Huan, et al. 1992). In particular, the Sumatra event consisted of an extraordinarily complex four-fault rupture lasting about 160 seconds (Yue, et al. 2012).

The quality, quantity and diversity of data hold crucial importance in training an ANN model. Larger and diversified data sets lead to a better trained model with robust prediction capabilities (Krizhevsky, et al. 2012). The catalog employed in this study contains earthquakes with magnitudes ranging from a minimum magnitude of 5.0 to a maximum of 9.1. However, the catalog is skewed towards higher magnitudes and has fewer events of low magnitudes. The abundant presence of a particular data class in the training set forces the ANN to better fit this class. Therefore, the model is able to identify that class over unseen data with higher accuracy. In our earthquake catalog (provided in the supplementary material), a varying number of samples are present in different magnitude classes. We define different ranges for earthquake magnitudes and analyze the performance of predictions as function of the sample size in the magnitude bins. The result in terms of MAE (Willmott and Matsuura 2005) is provided in Table 2. The MAE is highest for the least abundant magnitude class, while it decreases with increasing number of instances for a class. Thus, the result is in agreement with the expected relation to the sample size in each class. It also verifies the need of data diversity for an improved training of the machine learning model.

When an ANN is initialized, random weights are assigned to the connections between layers of neurons. During the learning process, the neurons continually adjust the connection weights until the model's performance reaches a maximum. It is noted that the sufficiency of training data and random initialization also play an important role in the performance of trained models. The performance of ANN may vary in different simulations if the training data set is not available in sufficient quantity. Therefore, it is important to analyze the robustness of the proposed methodology concerning the random initialization. For this purpose, we have carried out 10 independent simulation runs and the performance of every run is measured by the MAE-value (Asim, et al. 2018). We found that the MAE-value only varied by 1%, demonstrating the robustness of our ANN-model (See Table S3).

The performance of ANN has been compared to another well-known machine learning technique, namely the Random Forest (RF) method. RF is a decision-tree based algorithm, which provides an ensembled outcome of multiple decision-trees (Breiman 2001). We found that RF performs better than the regression equations in estimating earthquake magnitude in the case of the L-W-A-S scenario, but it is outperformed by ANN in all cases.

For practical applications of our proposed techniques, researchers can follow two different approaches: (a) They can use the ANN-model trained on our data set; or (b) they can run new simulation on updated data sets. Both options can be employed according to the feasibility of the potential users/researchers. The codes developed for this research study are shared publicly for the use of research community (available on: https://doi.org/10.6084/m9.figshare.8010608). The codes are developed in MATLAB and require neural network toolbox for successful execution.

Conclusion

We propose a new technique based on ANN for estimating the earthquake magnitude based on given fault information. This is often needed to relate geological fault information to potential maximum earthquake magnitudes for seismic hazard estimations. The ability of ANN to identify hidden patterns in data and its simultaneous use of fault parameters ensures the consistency of the approach. Our analysis of the predictions based on the parameters describing the geometrical dimension of faults (L, W, A) show a clearly improved performance in comparison to conventional regression methods. The proposed method has also the ability to simply integrate additional fault information in a consistent way. The addition of the maximum slip has been shown to further improve these estimations, thereby encouraging the use of additional fault parameters such as fault dip and rake in the future.

Data and Resources

The earthquake catalogs used in this research are taken from the Wells and Coopersmith (1994) and SRCMOD databases (http://equake-rc.info/SRCMOD/searchmodels/allevents/) (Last accessed on April 10, 2019). These data are also available in Table S1. The details regarding usage of data are explained in Section "Earthquake Data".

Acknowledgements

The authors are thankful to the Associated Editor Zefeng Li, electronic supplement editor Kim Olsen, and two anonymous reviewers for their valuable comments which helped to improve the manuscript. We are also grateful to Anne Strader for carefully reading and correcting our paper, as well as to friends and colleagues for their valuable suggestions and ideas.

References

Thatcher, W., and T. C. Hanks (1973). Source parameters of southern California earthquakes, Journal of Geophysical Research **78** 8547-8576.

Kanamori, H., and D. L. Anderson (1975). Theoretical basis of some empirical relations in seismology, Bulletin of the Seismological Society of America **65** 1073-1095.

Kanamori, H. (1977). The energy release in great earthquakes, Journal of Geophysical Research **82** 2981-2987.

Scholz, C. H. (1982). Scaling laws for large earthquakes: consequences for physical models, Bulletin of the Seismological Society of America **72** 1-14.

Romanowicz, B. (1992). Strike-slip earthquakes on quasi-vertical transcurrent faults: Inferences for general scaling relations, Geophysical Research Letters **19** 481-484.

Romanowicz, B., and J. B. Rundle (1993). On scaling relations for large earthquakes, Bulletin of the Seismological Society of America **83** 1294-1297.

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

Mai, P. M., and G. C. Beroza (2000). Source scaling properties from finite-fault-rupture models, Bulletin of the Seismological Society of America **90** 604-615.

Henry, C., and S. Das (2001). Aftershock zones of large shallow earthquakes: fault dimensions, aftershock area expansion and scaling relations, Geophysical Journal International **147** 272-293.

Leonard, M. (2010). Earthquake fault scaling: Self-consistent relating of rupture length, width, average displacement, and moment release, Bulletin of the Seismological Society of America **100** 1971-1988.

Reyes, J., A. Morales-Esteban, and F. Martínez-Álvarez (2013). Neural networks to predict earthquakes in Chile, Applied Soft Computing **13** 1314-1328.

Asim, K. M., F. Martínez-Álvarez, A. Basit, and T. Iqbal (2017). Earthquake magnitude prediction in Hindukush region using machine learning techniques, Nat Hazards **85** 471-486.

Asim, K. M., A. Idris, T. Iqbal, and F. Martínez-Álvarez (2018). Earthquake prediction model using support vector regressor and hybrid neural networks, PloS one **13** e0199004.

Rouet-Leduc, B., C. Hulbert, N. Lubbers, K. Barros, C. J. Humphreys, and P. A. Johnson (2017). Machine learning predicts laboratory earthquakes, Geophysical Research Letters **44** 9276-9282.

DeVries, P. M., F. Viégas, M. Wattenberg, and B. J. Meade (2018). Deep learning of aftershock patterns following large earthquakes, Nature **560** 632.

Bergen, K. J., T. Chen, and Z. Li (2019). Preface to the Focus Section on Machine Learning in Seismology, Seismological Research Letters **90** 477-480.

Asencio-Cortés, G., F. Martínez-Álvarez, A. Morales-Esteban, and J. Reyes (2016). A sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction, Knowledge-Based Systems **101** 15-30.

Morales-Esteban, A., F. Martínez-Álvarez, and J. Reyes (2013). Earthquake prediction in seismogenic areas of the Iberian Peninsula based on computational intelligence, Tectonophysics **593** 121-134.

Tareen, A. D. K., K. M. Asim, K. J. Kearfott, M. Rafique, M. S. A. Nadeem, T. Iqbal, and S. U. Rahman (2019). Automated anomalous behaviour detection in soil radon gas prior to earthquakes using computational intelligence techniques, Journal of environmental radioactivity 203 48-54.

Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, P. Gerstoft, and B. J. Meade (2018). Machine Learning in Seismology: Turning Data into Insights, Seismological Research Letters **90** 3-14.

Motagh, M., B. Schurr, J. Anderssohn, B. Cailleau, T. R. Walter, R. Wang, and J.-P. Villotte (2010). Subduction earthquake deformation associated with 14 November 2007, Mw 7.8 Tocopilla earthquake in Chile: Results from InSAR and aftershocks, Tectonophysics **490** 60-68.

Hassoun, M. H. (1995). Fundamentals of artificial neural networks, MIT press.

Wong, T.-T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation, Pattern Recognition **48** 2839-2846.

Idris, A., A. Iftikhar, and Z. u. Rehman (2017). Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling, Cluster Computing.

Huan, W., M. Ge, and X. Chang (1992). The features of multiple fractures associated with the great Haiyuan 8.5 magnitude earthquake of 1920, Acta Seismologica Sinica **5** 45-56.

Yue, H., T. Lay, and K. D. Koper (2012). En échelon and orthogonal fault ruptures of the 11 April 2012 great intraplate earthquakes, Nature **490** 245.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks, in *Advances in neural information processing systems*, 1097-1105.

Willmott, C. J., and K. Matsuura (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Climate research **30** 79-82.

Breiman, L. (2001). Random forests, Machine learning **45** 5-32.

List of Figure Captions

Figure 1: Earthquake magnitudes versus rupture length: Black triangles refer to the actual magnitudes, whereas red circles represent the predicted magnitudes from the regression equation of Wells & Coppersmith (1994).

Figure 2: Earthquake magnitude versus (a) rupture length and (b) rupture width: Symbols refer to observed values, while lines represent the regression lines (predicted values).

Figure 3: Histogram comparison of residuals from our ANN methodology for the L-W-A scenario with WC-94 regression equations (WC-L, WC-W) and new regression equations (New-L, New-W)

Figure 4: Actual and predicted earthquake magnitudes plotted against (a) length, (b) width for the case of the L-W-A scenario, while (c, d) show the same results for the L-W-A-S scenario.

Figure 5: (a) Scatterplot of actual and predicted earthquake magnitudes as function of fault length for (a) strike-slip and (b) dip-slip events in the case of the L-W-A scenario.

Table 1: Performance of ANN earthquake magnitude estimations for the whole earthquake data set acquired through 10-fold cross-validation.

Performance Measure	**ANN**		**Regression Methods**			
	L-W-A	L-W-A-S	L (WC-94)	W (WC-94)	L (new)	W (new)
RMSE	0.303	0.288	0.372	0.593	0.356	0.518
MAE	0.239	0.229	0.293	0.463	0.282	0.407

Table 2: Prediction quality for different available magnitude ranges in the earthquake data set.

Earthquake Magnitude Range	L-W-A		L-W-A-S	
	No. of Instances	MAE	No. of Instances	MAE
[5.0, 6.0)	77	0.242	32	0.258
[6.0, 7.0)	161	0.230	132	0.222
[7.0, 8.0)	187	0.228	176	0.220
>8.0	71	0.286	72	0.238