

# Modeling of Electrical Resistivity of Soil Based on Geotechnical Properties

Bandar Alsharari<sup>a</sup>, Andriy Olenko<sup>b,\*</sup>, Hossam Abuel-Naga<sup>c</sup>

<sup>a</sup>*Tabarjal Technical College Branch, Technical and Vocational Training Corporation, Riyadh, Saudi Arabia*

<sup>b</sup>*Mathematics and Statistics Department, La Trobe University, Melbourne, 3086, Victoria, Australia*

<sup>c</sup>*Department of Engineering, La Trobe University, Melbourne, 3086, Victoria, Australia*

<https://doi.org/10.1016/j.eswa.2019.112966>

---

## Abstract

Determining the relationship between the electrical resistivity of soil and its geotechnical properties is an important engineering problem. This study aims to develop methodology for finding the best model that can be used to predict the electrical resistivity of soil, based on knowing its geotechnical properties. The research develops several linear models, three non-linear models, and three artificial neural network models (ANN). These models are applied to the experimental data set comprises 864 observations and five variables. The results show that there are significant exponential negative relationships between the electrical resistivity of soil and its geotechnical properties. The most accurate prediction values are obtained using the ANN model. The cross-validation analysis confirms the high precision of the selected predictive model. This research is the first rigorous systematic analysis and comparison of difference methodologies in ground electrical resistivity studies. It provides practical guidelines and examples of design, development and testing non-linear relationships in engineering intelligent systems and applications.

*Keywords:* modeling, non-linear models, artificial neural network, electrical resistivity, compacted clays

*2010 MSC:* 62J02, 62M45

---

## Highlights

- Electrical resistivity of soil is modelled by its geotechnical properties.
- Several linear, non-linear and artificial neural networks models are developed.
- Guidelines in the development and quality analysis of the models are provided.

---

\*Corresponding author

Phone: +61-3-94792609

*Email addresses:* [Bandar2@tvtc.gov.sa](mailto:Bandar2@tvtc.gov.sa) (Bandar Alsharari), [A.Olenko@latrobe.edu.au](mailto:A.Olenko@latrobe.edu.au) (Andriy Olenko), [h.aboel-naga@latrobe.edu.au](mailto:h.aboel-naga@latrobe.edu.au) (Hossam Abuel-Naga)

- The models demonstrate significant exponential negative relationships.
- Artificial neural networks show much greater accuracy than other models.

## 1. Introduction

Investigations of the electrical resistivity of soil are important for various industrial applications, for example, estimation of corrosion of underground facilities, analysis of power distribution grounding systems, safe grounding design, see, for example, (Ackerman et al., 2013; Zhou Wang et al., 2015; Zastrow, 1979; Datsios et al., 2017; Clavel et al., 2018; Gomes et al., 2018), and the references therein.

The electrical resistivity of geo-materials has attracted the attention of the researchers since 1912 when Schlumberger used the electrical resistivity measurements to characterize the subsurface ground (Griffiths & King, 1965). Since then, the need to understand the behavior of electrical resistivity of soils has been extended to different areas such as design of electrical earthing system, monitoring moisture content for agriculture applications, electro-osmosis consolidation and drainage of soils, and assessing the homogeneity of compacted clay liner (Laver & Griffiths, 2001; Lim et al., 2013; Rhoades et al., 1976; Abu-Hassanein et al., 1996; Tabbagh et al., 2002; Brillante et al., 2015; Jones et al., 2011; Al Rashid et al., 2018).

In most of these areas the number of measurements can be substantial. Even for a few geotechnical soil properties their composition can often result in cases that are not reported in previous studies or available technical documentation. It requires an automatic computer system that can deal with these issues to determine a level of electrical resistivity to be used in practical decision-making solutions. This article discusses and compares several design approaches in this inference analysis.

For engineering design purposes, as the electrical conductivity of soils is a function of its geotechnical properties (soil mineralogy, particle size distribution, void ratio, pore size distribution, pore connectivity, degree of water saturation, pore water salinity, and temperature), it is important to determine a robust quantitative relationship between the electrical resistivity of soils and its geotechnical properties, see (Kibria & Hossain, 2012).

In the vast majority of applications determining soil electrical resistivity is done by trial and error. The results are often heavily based on empirical knowledge and can be influenced by subjective perceptions. There is a need to develop rigorous automatic methodologies allowing to estimate electrical resistivity with more certainty. To achieve this, it is important to evaluate various alternatives approaches.

Several laboratory studies have collected data on electrical resistivity of soils and their physical properties, see, for example, (Edlebeck & Beske, 2014; Zhou Wang et al., 2015; Southey et al., 2015; Al Rashid et al., 2018; Alipio & Visacro, 2013, 2014; Mokhtari & Gharehpetian, 2018). However, only a few electrical resistivity models were used to assess the role of different physical properties.

Unfortunately, none of these articles justified their model selection. No rigorous statistical evidence was found that can help to determine the best model. The aim of this study is to develop several statistical models of electrical resistivity and to provide general methodology for their comparison and performance evaluation. This systematic approach helps in selecting the most powerful predictive model.

It is worth mentioning that while artificial neural networks and other advanced computational methods were used for modeling characteristics of materials, see, for example, (Gençoğlu, M.T. & Cebeci, 2009; Gusel & Brezocnik, 2011; Hwang et al., 2010), this research is the first rigorous systematic analysis and comparison of difference methodologies in ground electrical resistivity studies. This paper provides practical guidelines and examples of modelling and investigating non-linear relationships in engineering applications.

Extensive experimental study was conducted in the Department of Engineering at La Trobe University to investigate the effect of soil mineralogy, moisture content, water salinity, and dry unit weight on the electrical resistivity measurements. These experimental data were used to develop a linear regression model, an exponential non-linear regression model, multivariate adaptive regression splines (MARS), and the artificial neural network model (ANN) for predicting electrical resistivity of soil. Several other models were investigated, but are not reported in this paper as their performances were not satisfactory.

While this paper concentrates on the design, development and testing of non-linear models in engineering soil studies the presented methodology can be applied to much wider classes of intelligent systems. In particular, the discussed ANN models can provide the overall trends in rather complex relationships even from the limited data.

All computations, data transformations, visualisations, statistical analysis, and models fitting were performed using R 3.5.0 software, in particular, the packages `plyr`, `nlme`, `car`, `hexbin`, `MASS`, `earth`, `nnet`, and `devtools`. The data and R codes used in this article are freely available in the folder "Research materials" from <https://sites.google.com/site/olenkoandriy/>.

## 2. Data

The electrical resistivity measurements were obtained using a soil box shown in Figure 1. The measurements process was conducted using the Wenner four-electrode method, (Wenner, 1915; ASTM G57, 2006). A DC test voltage with the magnitude 12V is applied between the outer electrodes, where the corresponding current,  $I$ , and the voltage drop between the inner electrodes,  $V$ , are measured. The resistivity,  $ER$ , is then  $ER = 2\pi aV/I$ . See more details about the measurement process in (Al Rashid et al., 2018).

The full data set consists of 864 observations with five variables. The variables of interest for this study are:

- Electrical resistivity (ER): This is a continuous variable that measures the electrical resistivity of soil in Ohm-m (the response variable).

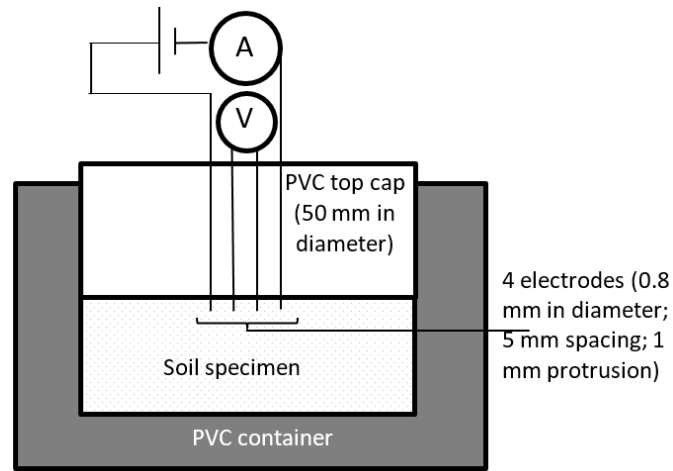


Figure 1: Test apparatus

- Soil water salinity (Mol): This is a continuous variable that measures the amount of NaCl salt in the water in mol unit and has only four values: 0, 0.5, 1 and 2 Mol.
- Soil moisture content (Moist): This is a continuous variable that measures the percentage of gravimetric water content. It is the ratio of weight of water to the weight of solid soil particles.
- Dry unit weight (Uw): This is a continuous variable that measures the dry unit weight in  $kN/m^3$ .
- Soil type (ST): This is a factor (categorical) variable that identifies the soil type. Nine different soil types were test in the conducted experimental study. They comprise different mixtures of Kaolin-Bentonite (K-B) and Kaolin-Sand (K-S) as shown in Table 3 which also provides the number of observations for each type of soil. The engineering properties of Kaolin, Bentonite, and Sand used in this study can be found in (Al Rashid et al., 2018).

Table 1 provides a snapshot of a part of the original data frame for only two soil types. Namely, pure kaolin (100%K) and kaolin with 10% of bentonite (90%K+10%B).

Table 2 shows the summary statistics of all continuous variables, namely electrical resistivity, molarity, moisture and unit weight.

Table 3 provides the number of observations for each type of soil.

A histogram of the electrical resistivity of soil is shown in Figure 2a. This distribution is highly skewed to the right, indicating that ER is abnormally distributed (potentially exponentially distributed). However, the histogram of

Table 1: Part of original data

	Molarity	Case 1		Case 2		Case 3		Case 4		Case 5							
		Moisture	ER	Moisture	ER	Moisture	ER	Moisture	ER	Moisture	ER						
100%K	0	3	1.016362	5934.6	5	1.027368	5197.259	10	1.02	3538.241	13	1.041136	2432.229	16	1.052393	1510.553	
			1.171358	1601.4		1.171244	1403.223		1.071858	957.3231		1.170817	660.0569		1.170675	412.335	
			1.348841	496.12		1.343553	435.2838		1.2606	298.4022		1.323722	207.1478		1.317112	131.1025	
	0.5	6	1.4	281.03	10	1.42	247.275	15	1.509148	171.3263	16	1.45	120.6938	20	1.47	78.5	
			1.027631	2064.55		1.021577	619.32		1.038425	345.9156		1.037725	225.3994		0.987859	55.615	
			1.163968	434.89		1.118193	175.1		1.14218	106.0556		1.148254	69.96938		1.096678	13.799	
	1.0	5	1.333892	208.025	11	1.303434	40.82	12	1.313324	25.98538	15	1.312816	18.92163	20	1.239812	7.227	
			1.635182	37.68		1.59972	19.625		1.619975	13.68105		1.607991	10.22605		1.622551	5.8718	
			1.028	2559.1		1.063478	681.5678		1.01	146.01		1.12	53.38		1.02	75.36	
	2.0	5	1.18	475.71	10	1.21663	100.4971	14	1.24	30.62	15	1.25	15.7	20	1.13	11.62	
			1.37	186.83		1.399014	34.94215		1.42	11.46		1.43	7.22		1.28	5.03	
			1.58	51.81		1.574892	18.66338		1.59	6.63		1.65	2.28		1.48	1.37	
	90%K+10%B	0.0	7	1.030352	192.4035	11	1.088916	124.1576	13	1.124055	83.21	18	1.09376	59.346	20	1.085783	47.09264
				1.124903	150.249		1.189365	94.21963		1.227882	60.602		1.193663	43.646		1.185712	33.84331
				1.238982	108.0945		1.310333	64.47794		1.353144	38.308		1.314284	28.0559		1.305619	23.81715
		0.5	5	1.378521	65.94	10	1.458622	34.54	13	1.456682	15.7	14	1.461368	12.403	20	1.452981	10.58769
1.140324				221.37	1.17859		36.267	1.179477		25.12	1.194166		13.816	1.168713		7.81075	
1.245628				162.495	1.286753		27.004	1.287731		18.6987	1.303776		10.362	1.278694		5.759938	
1.0		7	1.373522	103.62	11	1.417946	17.741	15	1.417953	10.205	22	1.434969	6.908	23	1.412282	3.709125	
			1.529188	44.745		1.57767	8.478		1.577582	6.2957		1.601824	3.454		1.51	3.1	
			0.829129	153.86		0.97998	28.26		0.979412	18.84		1.049472	5.181		1.086263	1.884	
2.0		7	0.904592	120.419	11	1.071823	21.98	12	1.068467	14.3969	15	1.14596	3.925	20	1.186142	1.57	
			0.996088	86.821		1.18321	15.7		1.175672	9.9381		1.261828	2.72238		1.305496	1.256	
			1.107322	53.38		1.319892	9.42		1.30736	5.495		1.404084	1.413		1.452586	0.942	
				1.104579	148.365		1.120985	55.69575		1.12437	22.98088		1.109781	10.205		1.106861	4.294766
				1.206119	108.33		1.223386	40.70618		1.226976	16.85788		1.211814	7.7087		1.191663	3.308156
				1.327457	67.667		1.346253	25.54783		1.350246	10.72114		1.334341	5.2124		1.292625	2.323594
				1.478992	28.26		1.497397	10.71525		1.501188	4.598138		1.484773	2.7161		1.416992	1.342297

Table 2: Summary statistics of the continuous variables

	Mol	Moist	Uw	ER
Min. :	0.000	3.00	0.7158	0.183
1st Qu.:	0.375	10.00	1.1570	2.890
Median :	0.750	15.00	1.3279	11.707
Mean :	0.875	14.93	1.3321	98.203
3rd Qu.:	1.250	20.00	1.4927	46.508
Max. :	2.000	27.00	1.9200	6232.900
NA's :		144	144	144

$\log(\text{ER})$ , which is shown in Figure 2b, is much closer to the normal distribution. Therefore, the log or similar transformations of the ER variable might be appropriate when fitting the linear regression model.

### 3. Multiple regression analysis

#### 3.1. Ordinary least squares approach (OLS)

First, the function `lm` of the R software was used to fit the multiple linear regression model to the data set, where the electrical resistivity of soil is the outcome variable, and the predictors are molarity, moisture and unit weight

Table 3: Sample sizes for each soil type

Soil type	n
100%K	92
90%K+10%B	92
80%K+20%B	96
70%K+30%B	104
60%K+40%B	92
90%K+10%S	92
80%K+20%S	104
70%K+30%S	96
60%K+40%S	96

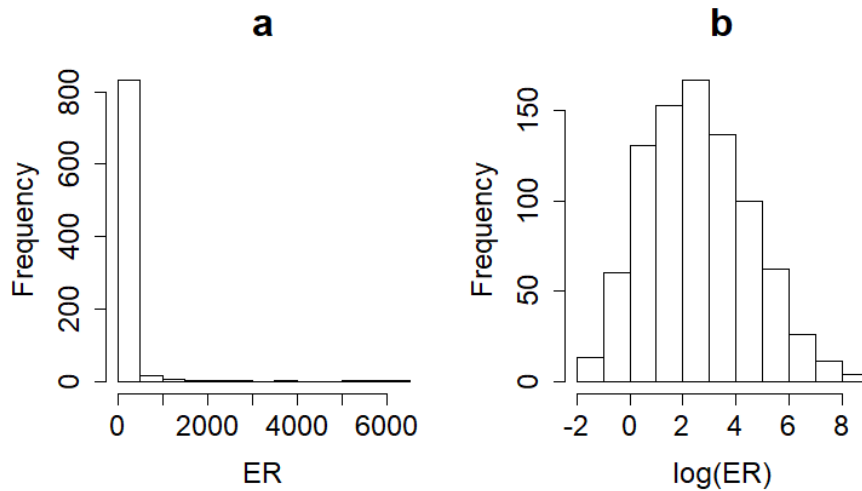


Figure 2: Histograms of the electrical resistivity and the log resistivity.

in addition to the soil type (where this variable is treated on the base of the percentage of kaolin that is mixed with bentonite and sand).

Figure 3 displays that the electrical resistivity for 100%K composition is likely to yield variation higher than those from other types of soil. There is a big difference in variability of ER for the soil that consists of pure kaolin and the variability of ER for soil that mixed with either bentonite or sand. This may indicate that treating the soil composition variable on the basis of pure and not pure kaolin leads to an improvement in our liner model.

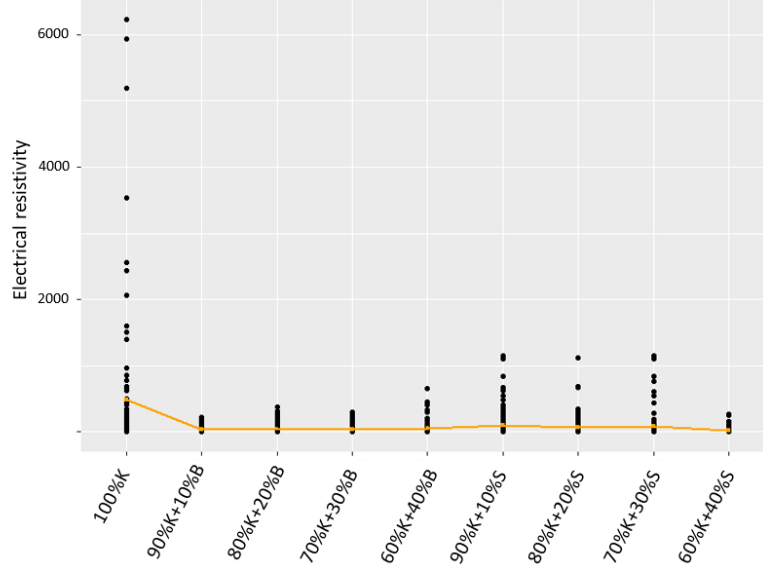


Figure 3: Scatter plot of ER vs soil type. The line represents the mean values.

Hence, we fit the following three models:

$$\widehat{ER} = 1101.2 - 643.7ST.KB - 247.3ST.KS - 59Mol - 14.8Moist - 473.3Uw$$

$$(Adj R^2 = 0.1623) \quad (1)$$

$$\widehat{ER} = 1202.5 - 373.45ST.K - 59.7Mol - 14.4Moist - 378.1Uw$$

$$(Adj R^2 = 0.2057) \quad (2)$$

$$\widehat{ER} = 4829.4 - 4316.8ST.K - 44.1Mol - 82.9Moist - 2523.7Uw$$

$$+ 75ST.K \times Moist + 2297.5ST.K \times Uw$$

$$(Adj R^2 = 0.4212) \quad (3)$$

It can be seen that the adjusted  $R^2$  for model (1) is 16.23%, where  $ST.KB$  and  $ST.KS$  represent the percentage of bentonite and sand that mixed with kaolin respectively. While in model (2), these two variables are replaced by  $ST.K$  which results in an increase in the adjusted  $R^2$  from 16.23% to 20.57%. This means it is better to treat the soil composition variable on the basis of whether it is mixed with bentonite or sand, or not. Moreover, the adjusted  $R^2$  increased from 20.57% to 42.12% (Model (3)). Hence, the interaction effect between soil type and Moisture ( $ST.K \times Moist$ ) and interaction effect between soil type and unit weight ( $ST.K \times Uw$ ) variables are significant predictors. Model (3) explains 42.12% of the variability in the electrical resistivity of soil using this data. However, the validity of inference from this model depends on

the linear regression assumptions being fulfilled. A violation of the underlying assumptions may lead to an invalid conclusion and an opposite conclusion can be obtained when using different samples (Montgomery et al., 2012). The main assumption is that the errors are from a normal distribution ( $\varepsilon \sim N(0, \sigma^2)$ ). So, we check error normality and constant error variance assumptions. Figure 4 illustrates the residuals vs fitted plot and normal Q-Q plot. A normal Q-Q plot raises concerns over the underlying normality. In addition, the residuals vs fitted plot shows that the constant error variance assumption is violated. There seems to be an increase in error variance as the fitted values increase. Hence, the detection of such problems undermines the validity of first simple linear models.

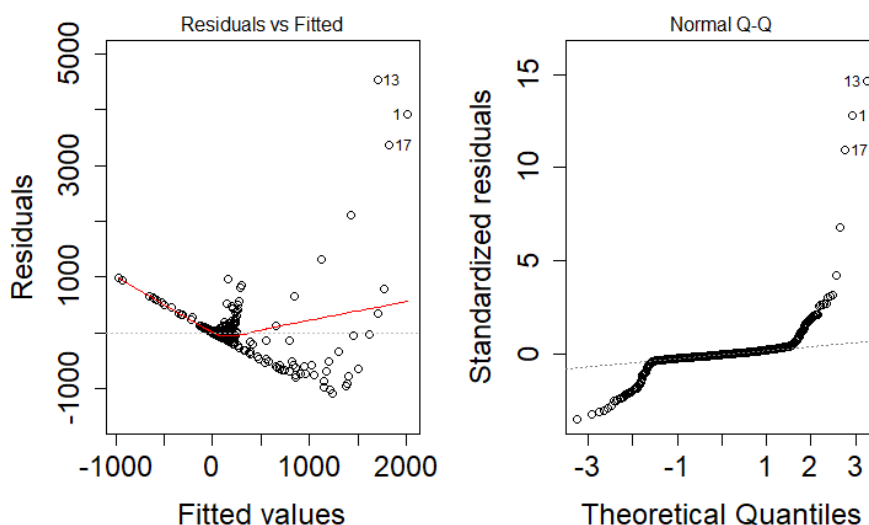


Figure 4: Residuals vs fitted and Normal Q-Q plots for Model (3).

### 3.2. Box-Cox transformation

The relationship between the dependent and independent variables is not a linear relationship. This may be the reason behind the problems that were found in the previous models in addition to the small value of  $R^2$ .

The non-normality in the response variable often causes issues in the OLS technique. A violation of this assumption may lead to a wrong conclusion. Therefore, this problem needs to be addressed before using the OLS technique. One way to deal with non-normality in the response variable is using Box-Cox transformation. In the data set, all values of the response variable are positive ( $ER > 0$ ). According to (Sakia, 1992), for positive values of the response variable, ( $y_i$ ), the family of the Box-Cox transformation is defined as:

$$y_i^\lambda = \begin{cases} (y_i^\lambda - 1)/\lambda, & \text{if } \lambda \neq 0, \\ \log(y_i), & \text{if } \lambda = 0. \end{cases}$$



The parameter  $\lambda$  can be estimated using the profile likelihood function. The dependent variable was transformed using the Box-Cox transformation. The estimated value of  $\lambda$  was very close to 0 ( $\lambda = -0.05$ ). Consequently, the log of the ER variable might be an appropriate transformation. So, the natural logarithm was applied to the dependent variable and the models were refitted again as it is shown below.

$$\log(\widehat{ER}) = 11.705 - 1.26ST.KB - 0.72ST.KS - 0.911Mol - 0.206Moist - 3.802Uw \quad (Adj R^2 = 0.8265) \quad (4)$$

$$\log(\widehat{ER}) = 12.14 - 0.98ST.K - 0.91Mol - 0.20Moist - 3.67Uw \quad (Adj R^2 = 0.8449) \quad (5)$$

$$\log(\widehat{ER}) = 16.05 - 5.23ST.K - 0.90Mol - 0.246Moist - 6.30Uw + 0.05ST.K \times Moist + 2.83ST.K \times Uw \quad (Adj R^2 = 0.8541) \quad (6)$$

The R-squared and adjusted R-squared values significantly increased for these three models which is a substantial improvement of the models.

To check the validity of the assumptions the residual vs fitted plot was examined. For model (6), the results are shown in Figure 5. The residual vs fitted plot shows that there is no dependency between residuals and fitted values which indicates that there are no violations to errors normality. Additionally, the normal Q-Q plot for model (6) shows that the points fall close to the 45-degree reference lines with a small deviation at the upper tails. Therefore, the normality is probably a reasonable assumption.

#### 4. Nonlinear regression model

Fitting a linear model provides a reasonable first approximation. However, the relationship between the electrical resistivity of soil and its geotechnical properties is not a linear relationship. Therefore, the use of a non-linear regression model may result in a better fit of the data. Furthermore, using a non-linear model makes the interpretation of the estimated parameters clearer.

According to (Fox & Weisberg, 2010), the general form of non-linear regression models can be written as

$$y = E(y|\mathbf{x}) + \varepsilon = f(\mathbf{x}|\boldsymbol{\theta}) + \varepsilon.$$

The best estimate of  $\boldsymbol{\theta}$  minimizes the residual sum of squares. Unlike linear models, one needs to use an iterative numerical procedure to estimate  $\boldsymbol{\theta}$ . The non-linear least-squares algorithm may not converge if inappropriate starting values are provided. Therefore, appropriate starting values need to be selected (Fox & Weisberg, 2010) to estimate the parameters.

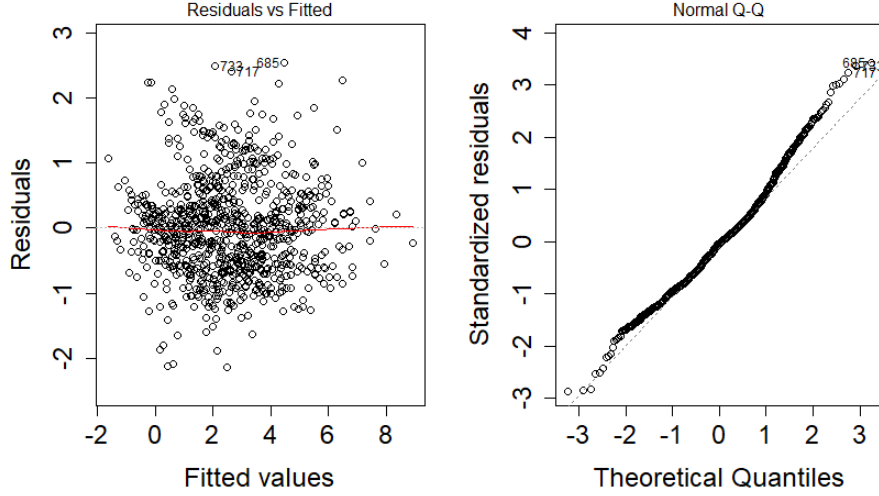


Figure 5: Residuals vs fitted and Normal Q-Q plots for Model (6).

Due to the nature of the data (see the discussions in the previous sections), the non-linear model that may fit the data well is

$$ER = \beta_0 + \beta_1 \exp(\beta_2 ST.KB) \exp(\beta_3 ST.KS) \times \exp(\beta_4 Mol) \exp(\beta_5 Moist) \exp(\beta_6 Uw) + \varepsilon. \quad (7)$$

This model was fitted using the `nls` function of the R computer package `nls`. The appropriate starting values can be obtained by fitting a linear model to the data and taking its coefficients as starting estimates. The starting values to estimate  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$  and  $\beta_7$  in model (7) were obtained from

$$\log(\widehat{ER}_i) = 11.70548 - 0.12629ST.KB_i - 0.07152ST.KS_i - 0.91053Mol_i - 0.20625Moist_i - 3.80215Uw_i.$$

The estimates of the coefficients, together with their corresponding standard errors, observed test statistics and p-values of Model (7) suggest that the predicted value of ER is given by

$$\widehat{ER} = 44.63 + 635736041 \exp(-5.074ST.KB - 4.042ST.KS + 0.09561Mol - 0.1968Moist - 10.76Uw) \quad (8)$$

Unfortunately, the predicted ER using this model cannot be less than the value of the intercept (44.63). However, we have some actual values of ER that are smaller than this value. As a result, this model will fail to accurately predict electrical resistivity when its actual value is less than 44.63. In addition, there

is a positive relationship between ER and Mol (since the estimate of  $\beta_4 > 0$ ) which contradicts to the previous models. Therefore, we refine the functional relationship between the response and the predictors and obtain the estimated equation

$$\widehat{ER} = 78.37 + 318382653 \exp(-48.25ST.KB) \exp(-39.082ST.KS) \times \exp(-40.64Moist) \exp(-0.1878Uw) - 10.08Mol. \quad (9)$$

We also refit the model using the ST.K variable instead of ST.KB and ST.KS. Namely, we fit the following model

$$ER = \beta_0 + \beta_1 \exp(\beta_2 ST.K) \exp(\beta_4 Moist) \exp(\beta_5 Uw) + \beta_3 Mol + \varepsilon. \quad (10)$$

The R output below demonstrates that all variables in model (10) are significant now.

```
Formula: ER~beta0+beta1*exp(beta2*ST.K+beta4*
Moist + beta5 * Uw) + beta3 * Mol
Parameters:Estimate Std. Error tvalue Pr(>|t|)
beta0      5.762e+01  1.057e+01   5.454 0.000
beta1     1.455e+08  6.189e+07   2.352 0.019
beta2     -3.602e+00  1.319e-01 -27.313 0.000
beta3     -3.297e+01  8.896e+00  -3.706 0.000
beta4     -1.865e-01  7.133e-03 -26.141 0.000
beta5     -9.330e+00  4.117e-01 -22.665 0.000

Residual standard error: 190.6 on 858 degrees
of freedom
Number of iterations to convergence: 17
Achieved convergence tolerance: 3.4e-06
```

Hence, the fitted non-linear model (10) gives the predicted value of ER by

$$\widehat{ER} = 57.62 + 145546403.36.67^{-ST.K} 1.20^{-Moist} 11275.04^{-Uw} - 32.97Mol. \quad (11)$$

Figure 6 shows the predicted vs actual plots for each non-linear model. Plots a, b and c are for models (8), (9) and (11) respectively.

We choose the best non-linear model based on the smallest value for AIC:

```
> AIC(Model.11,Model.13,Model.15)
df  AIC
Model.11  8 11654.63
Model.13  8 11642.82
Model.15  7 11532.53
```

The MSE values of models (8), (9) and (11) are 41471.40, 40908.61 and 36089.50 respectively. Hence, the best of these models is model (11) since it has the smallest value of AIC and the smallest value of MSE.

Figure 7 shows the residuals vs fitted values plot for model (11). Overall, there is concern about the dependency of errors and the magnitude of residuals increases with the increase of fitted values.

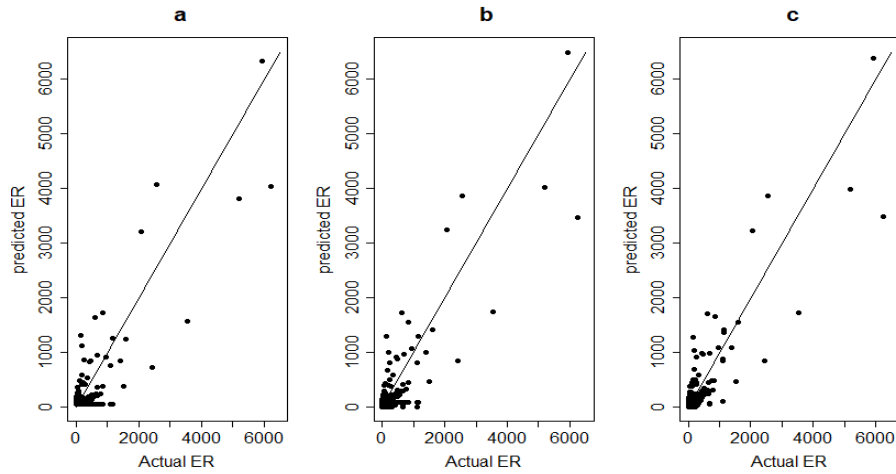


Figure 6: Predicted vs actual values plots for each model.

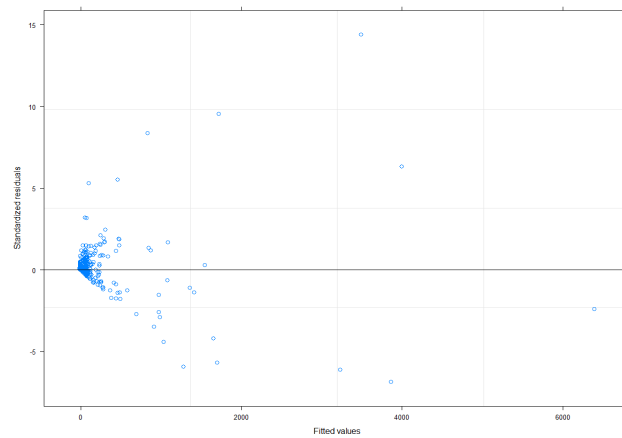


Figure 7: Residuals vs fitted values plot for model11.

## 5. Multivariate adaptive regression splines

The article (Friedman, 1991) introduced multivariate adaptive regression splines (MARS) which is a flexible regression model that can be used to handle both linear and non-linear relationships. The significant feature of MARS lies in its ability to generate simple and easy-to-interpret models that capture mapping in multi-dimensional data (Zhang & Goh, 2016).

The MARS model has the form (Hastie et al., 2009):

$$f(X) = \beta_0 + \sum_{m=1}^M \beta_m h_m(X),$$

where  $h_m(X)$  represents a basis function (hinge function) or product of two or more such functions. Each basis function consists of the pair  $\max(0, x - c)$  and  $\max(0, c - x)$  where  $c$  is a knot. Each pair is multiplied by a parameter which is estimated by minimizing the residual sum of squares (RSS).

The MARS model divides the data into regions and fits each region with an appropriate model.

Now, we refit models (4), (5) and (6) using the MARS approach. The MARS models were fitted by using the R package "earth". The results of the fits to the ER data are available in the folder "Research materials" from <https://sites.google.com/site/olenkoandriy/>.

Figure 8 shows the predicted ER vs the observed ER for the three MARS models.

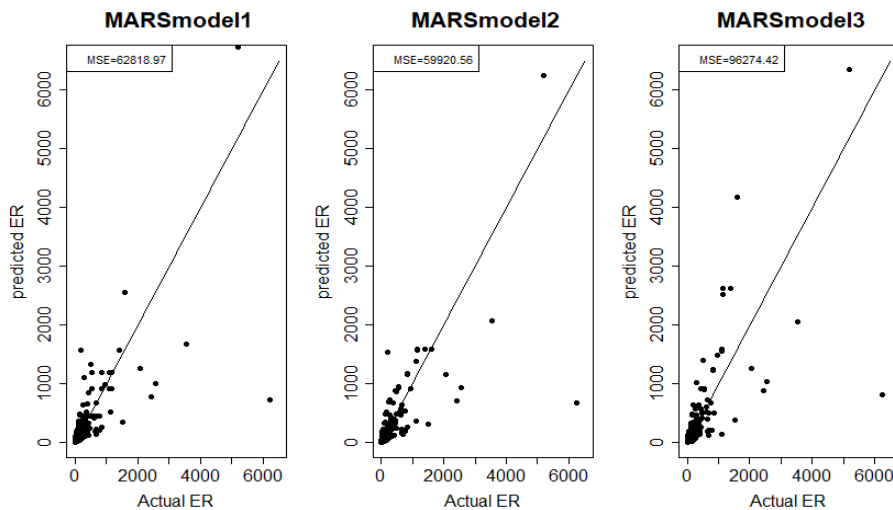


Figure 8: Predicted ER vs observed ER for the MARS models.

To compare these MARS models and to check their accuracy, the MSE for each model was calculated. Table 4 shows the obtained MSE values.

	MARS1	MARS2	MARS3
MSE	62818.97	59920.56	96274.42

Based on Figure 8 and the MSE values, the second MARS model gives the best fit of the three MARS models. Its estimated equation is

$$\begin{aligned} \widehat{ER} = & \exp[4.958 - 0.8739ST.K + 2.808 \max(0, 0.5 - Mol) - 0.9058 \\ & \times \max(0, Mol - 0.5) + 0.739 \max(0, Mol - 1) - 0.2447 \\ & \times \max(0, Moist - 5) + 0.07649 \max(0, Moist - 15) \\ & + 105.9 \max(0, Uw - 1.005) - 119 \max(0, Uw - 1.014) \\ & + 12.28 \max(0, Uw - 1.099) + 2.504 \max(0, 1.663 - Uw)]. \end{aligned}$$

Figure 9 shows the residual plots of the second MARS model. It can be seen that the residuals vs fitted plot shows no dependency in errors and they appear to have almost constant variance. Furthermore, the residuals QQ plot shows that the errors are normally distributed.

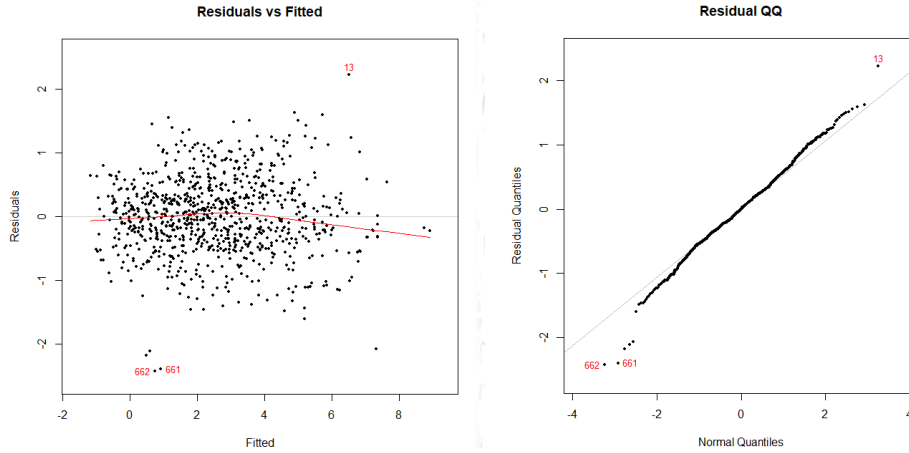


Figure 9: Residuals analysis of MARS model 2

## 6. Artificial neural network model

The artificial neural network (ANN) can learn from data and detect patterns and relationships in data (Agatonovic-Kustrin et al., 2000). ANNs can be used to predict the output from available information. The neural structure of an ANN consists of a number of neurons which are known as processing elements (Erzin et al., 2010), where each processing element represents an equation that consists of weighted inputs, transfer function and one output (Agatonovic-Kustrin et al., 2000).

We explore the use of ANNs to predict the electrical resistivity of soil based on its geotechnical properties. Three ANN models (ANN.model1, ANN.model2 and ANN.model3) were developed for predicting ER.

The first model, ANN.model1, takes into consideration the effect of ST.KB, ST.KS, Mol, Moist and Uw, whereas the second model, ANN.model2, takes into account the effect of ST.K, Mol, Moist and Uw. The third model, ANN.model3, takes into account the effects that are in ANN.model2 plus the effect of interaction between ST.K and Moist, and the interaction between ST.K and Uw. All three models have the output ER.

In most applications, a single hidden layer is enough for reasonable approximation (Agatonovic-Kustrin et al., 2000). So, one hidden layer is chosen for each model. (Erzin et al., 2010) pointed out that the upper limit for the number of neurons in the hidden layer is equal to  $2i + 1$  where  $i$  is the number of input parameters. Hence, the number of neurons in the hidden layer for ANN.model1, ANN.model2 and ANN.model3 should not be greater than 13, 11 and 15, respectively. To get the optimum number of neurons in the hidden layer of each model, we fitted each model with one neuron in the hidden layer and then increased the number of neurons to the upper limit. Moreover, we randomly divided the data into a training dataset (75% of the original data) and an independent validation dataset (25% of the original data) to estimate the performance of the models.

The neural network plots for each model are shown in Figure 10, 11 and 12. Note that each model has two bias layers (B1 and B2).

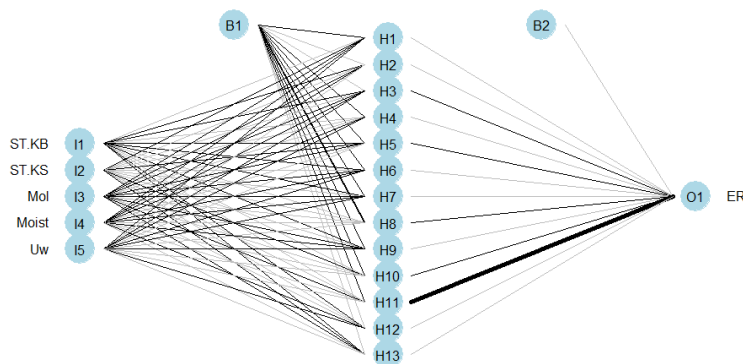


Figure 10: Neural network plot for ANN.model1.

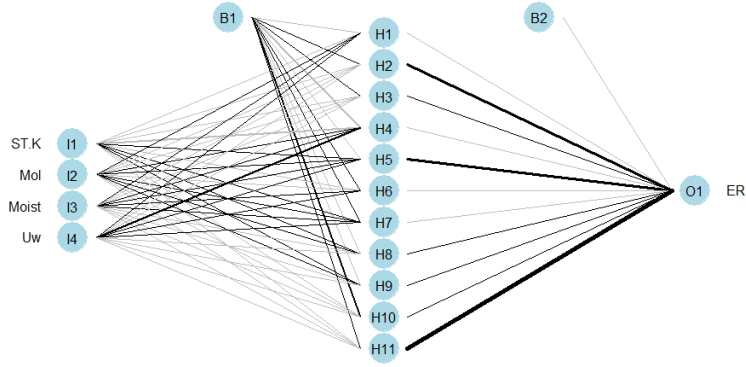


Figure 11: Neural network plot for ANN.model2.

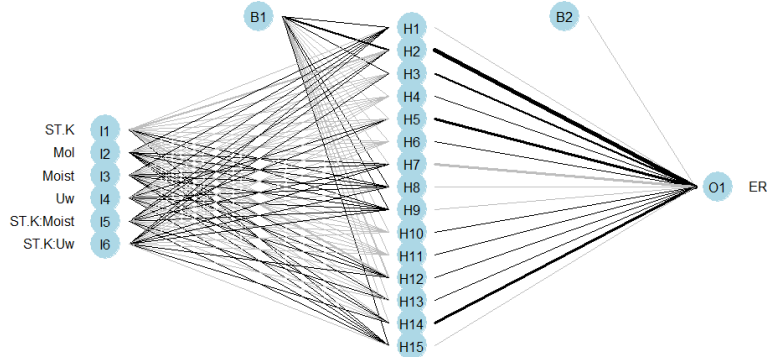


Figure 12: Neural network plot for ANN.model3.

Figure 13 shows the predicted value of ER vs actual values for each model. It can be seen that most of points lie very close to the reference lines, especially for ANN.model2. The  $\square$  symbols represent the training sample and the  $\circ$  symbols represent the validation sample. The MSE errors for ANN.model1, ANN.model2 and ANN.model3 are 182060.6, 49255.5 and 63702.84 respectively.

We can conclude that ANN.model2 is the best model and it can be used to reliably predict the electrical resistivity of soil.

## 7. Comparison and validation of models

We explored various statistical models to fit the electrical resistivity of soil and its geotechnical properties. The methods that were used include linear, non-linear regressions and artificial neural networks (ANN). Each model has its own advantages and disadvantages. Unfortunately, there is not single statistic that can be used to choose which one of these models is the best. Therefore, a particular statistic was used for each method. For example, the R-squared was



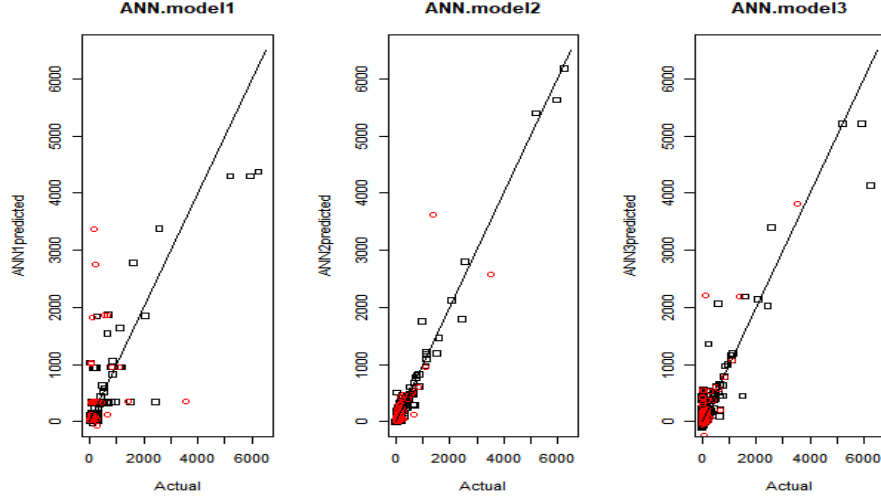


Figure 13: Predicted ER vs actual ER for the three ANN Models.

used to compare linear models fitted using OLS approaches, and MSE and AIC were used to compare non-linear models.

To confirm our findings and compare all models we used the cross-validation analysis. The data were randomly divided into a training dataset (75% of the original data) and a validation dataset (25% of the original data). Then, the best four models,

$$\begin{aligned} \log(\widehat{ER}) &= 16.05 - 5.23ST.K - 0.90Mol - 0.246Moist - 6.30Uw \\ &\quad + 0.05ST.K \times Moist + 2.83ST.K \times Uw; \\ \widehat{ER} &= 57.62 + 145546403 \cdot 36.67^{-ST.K} 1.20^{-Moist} 11275.04^{-Uw} - 32.97Mol; \\ \widehat{ER} &= \exp[4.958 - 0.8739ST.K + 2.808 \max(0, 0.5 - Mol) \\ &\quad - 0.9058 \max(0, Mol - 0.5) + 0.739 \max(0, Mol - 1) \\ &\quad - 0.2447 \max(0, Moist - 5) + 0.07649 \max(0, Moist - 15) \\ &\quad + 105.9 \max(0, Uw - 1.005) - 119 \max(0, Uw - 1.014) \\ &\quad + 12.28 \max(0, Uw - 1.099) + 2.504 \max(0, 1.663 - Uw)]; \end{aligned}$$

and the artificial neural network model ANN.model2, which input layer consists of ST.K, Mol, Moist and Uw, were refitted using only the training dataset. The new fitted models were applied to the validation dataset to predict ER. MSE was calculated for each model. The procedure was repeated 100 times. Each time, new training and validation data sets were randomly selected. The mean of MSEs for each model was calculated, see Table 5. The result shows that the most accurate prediction values of the electrical resistivity were obtained using the ANN model.

Table 5: MSEs for the best four models

	Lin	NonLin3	MARS2	ANN2
Mean MSE	67827.3	52639.23	84044.25	38831.16

Based on all the all above evidence we conclude that the neural network model (ANN.model2) is the best model for predicting electrical resistivity.

## 8. Conclusion and future work

The main objective of this study was to develop an accurate statistical model that can be used to estimate the electrical resistivity of soil based on the knowledge of soil type (ST), molarity (Mol), moisture (Moist) and unit weight (Uw).

The obtained results showed that there is a significant exponential negative relationship between electrical resistivity and all four predictors. Moreover, it was demonstrated that treating soil composition as a factor variable (denoted by ST.K) on the basis of pure kaolin or not pure kaolin is better than treating it as a continuous variable based on the percentage of bentonite (ST.KB) and sand (ST.KS). The interactions between ST.K and moisture, and between ST.K and Uw were found to be significant predictors for first two models. As a result, both these interactions were included in the final models. Moreover, the log transformation of the dependent variable yields better models. In particular, it substantially improved the value of R-squared in the linear models from about 0.20 to more than 0.80.

The analysis was performed to identify the optimal combination of the degree of interactions and the number of retained terms or complexity of neural networks. For MARS2 and ANN2 models it was shown that the interaction between ST.K and Uw does not contribute significantly to electrical resistivity prediction. A plausible explanation is that MARS2 and ANN2 models use more flexible non-linear transformations compare to the second model. They can reduce prediction errors directly via separately transformed ST.K and Uw.

Several useful models have been developed and the results from each model were compared with the actual values of electrical resistivity using the cross-validation. In addition, performance indices such as the coefficient of determination (R-squared), MSE and AIC were used to assess the performance of the models.

It was found that the ANN model is able to efficiently predict the electrical resistivity of soil and is better than the other models that were developed.

The developed ANN model in this study could contribute to the continuous research effort in the field of electrical resistivity imaging as it captures the influence of different geotechnical properties of soil on its electrical resistivity. Furthermore, it sheds light on what has been learnt throughout the study in terms of the sensitivity of different soil variables and how its incremental changes in the experimental program should be chosen to maximize the accuracy of the

developed model. In this regard, it is recommended that, in the future experimental studies, the incremental change in the percentage of bentonite and sand, salinity of pore water, moisture content, and dry unit weight should be smaller than the values used in this study. In fact, this recommendation can be justified by the observed exponential relationships between electrical resistivity of soil and its geotechnical properties which causes a large change in electrical resistivity even with a small change in geotechnical properties. However, experts in the field know that very small incremental changes in the geotechnical parameters of soil in laboratory experiments are often difficult to be achieved accurately. Therefore, their effects on measurement results could be lower than the true ones. It is desirable to implement experimental designs that balance the size of increments and the number of repeated measurements. The problem of balancing and assessing accuracy of such tests is an interesting direction for future research.

For fully automated intelligent system analysis based on the first two models, transformation of data is an area of potential improvement. The choice of transformation method is conventionally based on descriptive statistics results. This process can be semi-supervised if a finite number of admissible functional transformations and relationships is specified.

One of the main known ANN limitations is knowledge extraction from trained ANNs, i.e. interpreting ANN models similar to the other methods that provide functional relation to input effects. The Lek's profile and local interpretable model-agnostic explanations methods, see (Zhang et al., 2018), could be potentially useful approaches in interpreting ANN results.

For small randomly selected test or training sets ANN resulting models can substantially vary, nevertheless, providing rather similar predictions. Studies with a large number of explanatory parameters can also result in a large set of candidate models. In such studies it would be interesting to quantitatively investigate agreement levels between top selected models by using various agreement coefficients, see (Olenko & Tsyganok , 2016).

Finally, in the future research it is important to investigate other parameters like the proctor density of the soil, ambient temperature, two-layer or a multi-layer stratification to obtain finer electro resistivity models.

## **Acknowledgement**

The authors are grateful to the referees for their careful reading of the paper and suggestions that helped to improve the manuscript.

## **References**

Abu-Hassanein, Z. S., Benson, C. H., & Blotz, L. R. (1996). Electrical resistivity of compacted clays. *Journal of Geotechnical Engineering*, 122(5), 397-406.

- Ackerman, A., Sen, P. K., & Oertli, C. (2013). Designing safe and reliable grounding in AC substations with poor soil resistivity: An interpretation of IEEE Std. 80. *IEEE Transactions on Industry Applications*, 49(4), 1883-1889.
- Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, 22(5), 717-727.
- Al Rashid, Q.A., Abuel-Naga, H.M., Leong, E.-C., Lu, Y., & Al Abadi, H. (2018). Experimental-artificial intelligence approach for characterizing electrical resistivity of partially saturated clay liners. *Applied Clay Science*, 156, 1-10.
- Alipio, R. & Visacro, S. (2013) Frequency dependence of soil parameters: effect on the lightning response of grounding electrodes. *IEEE Transactions on Electromagnetic Compatibility*, 55(1), 132-139.
- Alipio, R. & Visacro, S. (2014). Modeling the frequency dependence of electrical parameters of soil. *IEEE Transactions on Electromagnetic Compatibility*, 56(5), 1163-1171.
- ASTM G57 (2006). *Standard Test Method for Field Measurement of Soil Resistivity Using the Wenner Four-Electrode Method*. West Conshohocken: ASTM International.
- Brillante, L., Mathieu, O., Bois, B., van Leeuwen, C., & Lévêque, J. (2015). The use of soil earthing system performance. *Rev. Energ. Ren.: Power Engineering*, 57-61.
- Clavel, E., Roudet, J., Guichon, J. M., Gouichiche, Z., Joyeux, P., & Derbey, A. (2018). A nonmeshing approach for modeling grounding. *IEEE Transactions on Electromagnetic Compatibility*, 60(3), 795-802.
- Datsios, Z. G., Mikropoulos, P. N., & Karakousis, I, (2017). Laboratory characterization and modeling of DC electrical resistivity of sandy soil with variable water resistivity and content. *IEEE Transactions on Dielectrics and Electrical Insulation*, 24(5), 3063-3072.
- Edlebeck, J. E., & Beske, B. (2014) Identifying and quantifying material properties that impact aggregate resistivity of electrical substation surface material. *IEEE Transactions on Power Delivery*, 29(5), 2248-2253.
- Erzin, Y., Rao, B., Patel, A., Gumaste, S., & Singh, D. (2010). Artificial neural network models for predicting electrical resistivity of soils from their thermal resistivity. *International Journal of Thermal Sciences*, 49(1), 118-130.
- Fox, J., & Weisberg, S. (2010). *Nonlinear Regression and Nonlinear Least Squares in R Appendix to an R Companion to Applied Regression*. <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion/appendix/Appendix-Nonlinear-Regression.pdf> Accessed 21 July 2018.

- Friedman, J. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1), 1-67.
- Gomes, T. V., Schroeder, M. A. O., Alipio, R., de Lima, A. C. S., & Piantini, A. (2018). Investigation of overvoltages in hv underground sections caused by direct strokes considering the frequency-dependent characteristics of grounding, *IEEE Transactions on Electromagnetic Compatibility*, to appear.
- Gençoğlu, M.T. & Cebeci, M. (2009). Investigation of pollution flashover on high voltage insulators using artificial neural network, *Expert Systems with Applications*, 36(4), 7338-7345.
- Griffiths, D. H. & King, R. F. (1965). *Applied Geophysics for Engineers and Geologists*. Oxford: Pergamon Press.
- Gusel, L. & Brezocnik, M. (2011). Application of genetic programming for modelling of material characteristics, *Expert Systems with Applications*, 38(12), 15014-15019.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Springer-Verlag.
- Hwang, R.-C. Chen, Yu-Ju, & Huang, H.-C. (2010). Artificial intelligent analyzer for mechanical properties of rolled steel bar by using neural networks, *Expert Systems with Applications*, 37(4), 3136-3139.
- Jones, C.J.F.P. Lamont-Black, & J. Glendinning S. (2011). Electrokinetic geosynthetics in hydraulic applications, *Geotextiles and Geomembranes*, 29, 381-390.
- Kibria, G., & Hossain, M. S. (2012). Investigation of geotechnical parameters affecting electrical resistivity of compacted clays. *Journal of Geotechnical and Geoenvironmental Engineering*, 138(12), 1520-1529.
- Laver, J. A., & Griffiths, H. (2001). The variability of soils in earthing measurements and earthing system performance. *Rev. Energy Ren. Power Engineering, Cardiff University, UK*, 57-61.
- Lim, S. C, Gomes, C., & Kadir, M. Z. (2013). Earthing in troubled environment. *International Journal of Electrical Power & Energy Systems*, 47, 117-128.
- Mokhtari, M. & Gharehpetian, G. (2018). Integration of energy balance of soil ionization in CIGRE grounding electrode resistance model. *IEEE Transactions on Electromagnetic Compatibility*, 60(2), 402-413.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. Hoboken, NJ: Wiley.
- Olenko, A., & Tsyganok, V. (2016). Double entropy inter-rater agreement indices. *Applied Psychological Measurement*, 40(1), 37-55.

- Rhoades, J., Raats P.A.C, & Prather, R. (1976). Effect of liquid-phase electrical conductivity, water content and surface conductivity on bulk soil electrical conductivity. *Soil Sci. Soc. of Am. J.*, 40, 651-655.
- Sakia, R. (1992). The Box-Cox transformation technique: A Review. *Journal of the Royal Statistical Society*, 41(2), 169-178.
- Southey, R. D., Siahrang, M., Fortin, S., & Dawalibi, F. P. (2015). Using fall-of-potential measurements to improve deep soil resistivity estimates. *IEEE Transactions on Industry Applications*, 51(6), 5023-5029.
- Tabbagh, A., Benderitter, Y., Michot, D., & Panissod, C. (2002). Measurement of variations in soil electrical resistivity for assessing the volume affected by plant water uptake. *European Journal of Environmental and Engineering Geophysics* 7, 229-237.
- Wenner, F. (1915) A method for measuring earth resistivity. *Bulletin of the Bureau of Standards*, 12, 469-478.
- Zastrow, O. W., (1979). Choices of jacketed or bare concentric neutral cable for effective grounding and corrosion control. *IEEE Transactions on Industry Applications*, IA-15(1), 80-84.
- Zhang, W., & Goh, A. T. (2016). Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geoscience Frontiers*, 7(1), 45-52.
- Zhang, Z., Beck, M. W., Winkler, D. A., Huang, B., Sibanda, W., & Goyal, H. (2018). Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Annals of translational medicine*, 6(11): 216.
- Zhou Wang, J., Cai, L., Fan, Y., & Zheng, Z., (2015). Laboratory investigations on factors affecting soil electrical resistivity and the measurement. *IEEE Transactions on Industry Applications*, 51(6), 5358-5365.