Hierarchical Multiple Linear Regression for Fast Estimation of Subsurface Resistivity from Apparent Resistivity Measurements Based on Dipole-Dipole Array

S B Muhammad¹² and R Saad¹

¹ Geophysics Section, School of Physics, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia
² Department of Physics, Usmanu Danfodiyo University, Sokoto, PMB 2346 Sokoto, Nigeria
sabiubala@gmail.com

Abstract. Multiple linear regression (MLR) models for fast estimation of true subsurface resistivity from apparent resistivity field measurements are developed and assessed in this study, based on Dipole-dipole array. The parameters that have been investigated are apparent resistivity ($\rho_a$), horizontal location ($X$) and depth ($Z$) of measurement as the independent variables; and true resistivity ($\rho_t$) as dependent variable. To achieve linearity in both resistivities, datasets were first transformed into logarithmic domain following diagnostic checks of normality of the dependent variable and heteroscedasticity to ensure accurate models. Four MLR models namely; DD1, DD2, DD3 and DD4 were developed based on hierarchical combination of the independent variables. The generated MLR coefficients were applied to another dataset to estimate $\rho_t$ values for validation. The accuracy of the models was assessed using coefficient of determination ($R^2$), standard error (SE) and weighted mean absolute percentage error (wMAPE). It is concluded that the MLR models can estimate $\rho_t$ for the Dipole-dipole array with high level of accuracy, with the DD4 model being the best.

1. Introduction

The importance of resistivity surveying to solving many practical problems cannot be overemphasized. It plays vital role in water exploration [1,2], sites characterization [3,4] and mineral prospecting [5,6]. However, the measured apparent resistivity data from such surveys cannot be used directly to describe the subsurface [7]. Therefore, there is a need to develop resistivity models that can portray the true subsurface condition based on the apparent resistivity measurements, typically by inverse modeling exists [8]. Resistivity inversion has received enormous attention from various researchers over the years. Investigations are still ongoing to further mitigate the problems associated its speed and efficacy. Several researches have been conducted to develop and control resistivity inversion mechanisms.

The least – squares optimization based techniques are the most common and widely explored by various researchers [9,10] compared to other approaches. They have been applied to solve many practical inversion problems of both simple and complex structures with recorded remarkable success [11]. They are however not without problems. Some of their major drawbacks are: non-uniqueness of solutions, consume a lot of time and memory space for the iterative computations to converge at
minimum data misfit [12,13], inability to produce accurate models when the starting initial model is far from the actual solution, emergence of false anomalies at intermediate and later iterations [14] and difficulties in dealing with nonlinear relationships among the modeling (subsurface) parameters [15].

Multiple regression coefficients obtained from linear least squares method can provide alternative means for rapidly estimating the true subsurface resistivity whenever apparent resistivity site measurements are conducted. The nonlinearity in the nature underground resistivity distributions can be taken care of by using certain mathematical transformations, rather than using the data directly in the regression process. The objective of this research is to develop and assess mathematical models for fast estimation of true subsurface resistivity from apparent resistivity measurements. This will aid in cutting down the data processing time and computer memory capacity needed to analyze the apparent resistivity data using conventional algorithms.

2. Methodology

2.1 Resistivity Data

A typical electrical resistivity method measures underground potential contrast (V) upon injecting specific amount of current (I) into the ground (Fig.1). Resistance (R) is then automatically computed from Ohm’s law (Eq. 1) and subsequently the apparent resistivity when array geometrical factor is known [16]. The data utilized for the regression modelling of this research was acquired from ground resistivity survey conducted along two profiles, one for model calibration (Line 1) and the other for model validation (Line 2), using multi-electrode resistivity meter (ABEM Terrameter SAS4000 instrument). Each line consisted of 41 stainless steel electrodes at 1 m minimum spacing connected to a cable take–out using jumper with dipole-dipole array as the chosen protocol (Fig. 2). Dipole-dipole array configuration was particularly selected for its high sensitivity to lateral resistivity variations compared to the remaining arrays. The current was set between 1 mA and 100 mA. A computer–controlled system Terrameter SAS4000 was used to automatically select the four active electrodes for each voltage measurement and consequently compute the apparent resistivity (Eq. 2) using the array geometrical factor (k) defined in Figure 2. Apparent resistivity values are arranged systematically such that each datum point is specified with its horizontal (X) and vertical (Z) coordinates (Fig.3).

\[ R = \frac{V}{I} \]  \hspace{1cm} (1)
\[ \rho_a = k \frac{\Delta V}{I} \]  \hspace{1cm} (2)

For dipole-dipole array, \( k = \pi n(n + 1)(n + 2)a \), where \( a \) is the dipole length and \( n \) is the dipole separation factor.

For dipole-dipole array, \( k = \pi n(n + 1)(n + 2)a \), where \( a \) is the dipole length and \( n \) is the dipole separation factor.

Standard constrained least-squares inversion was applied to the apparent resistivity data with help of Res2Dinv software package to obtain the true resistivity subsurface model (at 2.4% RMS error), which was later digitized to obtain \( \rho_t \) values at same coordinates (X and Z) as the \( \rho_a \) values. Datasets (X, Z, \( \rho_a \) and \( \rho_t \)) were then arranged in columns in preparation for the multiple linear regression modelling.
2.2 Multiple Linear Regression

Least squares based multiple linear regression model with \( n \) independent variables (\( x \)) in the observations and a dependent variable (\( y \)) can be written as

\[
y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon
\]  

(3)

where \( \beta_0 \) is the effective intercept of the \( n \) regression lines and \( \varepsilon \) is the error term. In this approach, the model converges when the squared sum of the differences between the estimated and observed values is minimum. From Eq. 3, the values of the dependent variable can be estimate for given values of independent variable(s) if the error terms are assumed to be normally distributed with zero mean and constant variance. The distribution of the \( y_i \) is thus symmetrical and bell-shaped, with a constant standard deviation at \( x \) levels [17].

In this research, true resistivity (\( \rho_t \)) is considered as dependent variable, while apparent resistivity (\( \rho_a \)), horizontal location (\( X \)) and depth (\( Z \)) are the independent variables, with \( \rho_a \) being the most important variable. \( X \) and \( Z \) are only added to verify whether they can add any statistical significance to the resulting model. The independent variables were not entered into the regression modelling at once. Thus, four models namely; DD1, DD2, DD3 and DD4 were developed based on hierarchical combination of the independent variables to ascertain their contributions. Logarithmic transforms of \( \rho_a \) and \( \rho_t \), denoted as \( \log \rho_a \) and \( \log \rho_t \) respectively, were rather used instead of using the actual variables directly following pre-modelling diagnoses such as tests of normality, linearity, heteroscedasticity, axes balance and outliers. First datasets (Line 1) were imported into SPSS to run the regression modelling (model calibration). Regression coefficients obtained from model calibration were applied to the second datasets (Line 2) for validation. Estimated \( \rho_T \) values for the various models were transformed back to \( \rho_t \) and plotted using Golden Surfer 11 at same color scale and blanking as the observed true resistivity subsurface model \( \rho_t \) model obtained from the inversion (with Res2Dinv) for visual assessment.

The accuracy of each model was assessed using coefficient of determination (\( R^2 \)), standard error (SE) and weighted mean absolute percentage error (wMAPE) for model calibration; and SE and wMAPE for validation. According to [18], an added variable is statistically significant if it brings about an increase in \( R^2 \) (or decrease in SE) of at least 5% and above. The \( R^2 \) SE and wMAPE were calculated from Eq. 4, 5 and 6 respectively.

\[
R^2 = \frac{\sum (\rho_T - \hat{\rho}_T)^2}{\sum (\rho_T - \bar{\rho}_T)^2}
\]  

(4)

\[
SE = \left[\frac{1}{N} \sum (\rho_T - \hat{\rho}_T)^2\right]^{1/2}
\]  

(5)

\[
wMAPE = \frac{1}{N} \sum \left| \frac{\rho_T - \hat{\rho}_T}{\rho_T} \right| \times 100
\]  

(6)

where \( \rho_T, \hat{\rho}_T \) and \( \bar{\rho}_T \) are observed, estimated and average Log \( \rho_t \) values respectively; and \( N \) is the number of data points.

3. Results and Discussion

Table 1 presents summary of the four regression models developed in this study. The \( R^2 \) is found to increase considerably from DD1 to DD4. Similarly, the calibration SE and wMAPE are found to decrease from DD1 to DD4, even though the decrease in wMAPE is not all that significant. It can be deduced from the table that there is only about 1.4 % increase in \( R^2 \) and 1 % decrease in SE for DD2 model as result of adding X variable to the original (DD1) model, rendering it insignificant. When Z is added to original model (to create DD3) however, an increase of about 11 % in \( R^2 \) and 13 % drop in
SE occurred, making DD3 a statically significant model going by the assertion made by [18] that adding a variable into a regression model is only significant if bringing about at least 5% rise in $R^2$ or 5% drop in SE. When both $X$ and $Z$ variables were added to form the overall (DD4) model, there was about 19% increase in $R^2$ and 26% decrease in SE.

The validation SE and wMAPE showed comparable results (Table 2), but has slightly favoured DD3 (with 0.220 and 5.56%) to DD4 (with 0.224 and 5.68%) respectively. This might be for to the fact that outliers were not removed from the validation data (Line 2), as was the case for the model calibration data (Line 1). However, this should not be a problem, considering that these differences are just so small to warrant any significant preference between the two models. For this reason, DD4 is therefore regarded as most preferred among the four models (as pointed out previously) for the rapid estimation true subsurface resistivity whenever apparent measurements are conducted using Dipole–dipole array.

### Table 1. Models summary table

| Models | Estimated MLR Models | $R^2$ | SEc | wMAPEc [%] | SEC | wMAPEv [%] |
|--------|----------------------|------|-----|------------|-----|------------|
| DD1    | $\rho_T = 1.008 \rho_A + 0.065$ | 0.47 | 0.210 | 5.35 | 0.242 | 6.07 |
| DD2    | $\rho_T = 1.003 \rho_A - 0.002X + 0.124$ | 0.48 | 0.208 | 5.34 | 0.242 | 6.12 |
| DD3    | $\rho_T = 0.485 \rho_A + 0.108Z + 1.898$ | 0.58 | 0.183 | 4.61 | 0.220 | 5.56 |
| DD4    | $\rho_T = 0.432 \rho_A - 0.002X + 0.118Z + 2.125$ | 0.66 | 0.155 | 4.02 | 0.224 | 5.68 |

*subscripts c and v designate calibration and validation respectively

### 4. Conclusions

Four hierarchical multiple linear regression models for estimating true subsurface resistivity rapidly are developed and validated in this study. Numerical accuracy assessment carried out suggested DD4 as the best model having the highest $R^2$ of 0.659 and lowest SE and wMAPE of 0.155 and 4.02% respectively. It is therefore concluded that the DD4 MLR model can estimate $\rho_t$ for the Dipole-dipole array with within allowable accuracy limits. It is recommended that the procedure be applied to the remaining resistivity array configurations for the same objective as outlined in this study. Other subsurface parameters which were not investigated in this study are also recommended be examined to achieve better estimation models.

### References

[1] Saad R, Adli I and Mohamad A S 2012 The study of iron ore prospect using 2-d resistivity and induced polarization (IP) method Electron. J. Geotech. Eng. 17 2981–2988
[2] Saad R, Syukri M, Nordiana M M and Ismail N A 2013 Water table delineation for leachate identification using 2-D electrical resistivity imaging and seismic refraction at Gampong Jawa, Banda Aceh Electron. J. Geotech. Eng. 18 1529–1535
[3] Dahlin T, Wisén R and Zhang D 2007 3D effects on 2D resistivity imaging – modelling and field surveying results. Near Surface 2007, 13th European Meeting of Environmental and Engineering Geophysics, Istanbul, Turkey.
[4] Bery A A and Saad R 2012 Clayey sand soil’s behaviour analysis and imaging subsurface structure via engineering characterizations and integrated geophysicals tomography modeling methods Int. J. Geosci. 3 93–104
[5] Saad R, Nawawi M N M and Mohamad E T 2012 Groundwater detection in alluvium using 2-D electrical resistivity tomography (ERT) Electron. J. Geotech. Eng. 17 369–376
[6] Song W, Wang Y and Shao Z 2017 Categorical modeling on electrical anomaly of room-and-pillar coal mine fires and application for field electrical resistivity tomography, J. Appl. Geophys. 136 474–483

[7] De Donno G and Cardarelli E 2017 Tomographic inversion of time-domain resistivity and chargeability data for the investigation of landfills using a priori information. Waste Manage. 59 302–315

[8] Narayan S, Dusseault M B and Nobes D C Inversion techniques applied to resistivity inverse problems Inverse Problems 10 669

[9] Olayinka A I and Yaramanci U 2000 Use of block inversion in the 2-D interpretation of apparent resistivity data and its comparison with smooth inversion J. Appl. Geophys. 45 63–81

[10] Loke M H and Dahlin T 2002 A comparison of the Gauss-Newton and quasi-Newton methods in resistivity imaging inversion J. Appl. Geophys. 49 149–162

[11] Pain C C, Herwanger J V, Loke M H and Worthington C R E 2002 De Oliveira, Effective multidimensional resistivity inversion using finite-element techniques Geophys. J. Int. 151 710–728

[12] Gyulai Á, Ormos T and Dobroka M 2010 A quick 2-D geoelectric inversion method using series expansion J. Appl. Geophys. 72 232–241.

[13] Vachiratienchai C and Siripunvaraporn W 2013 An efficient inversion for two-dimensional direct current resistivity surveys based on the hybrid finite difference-finite element method Phys. Earth Planet. Inter. 215 1–11.

[14] Liu B, Nie L, Li S, Xu L, Liu Z, Song J and Lin C 2012 3D Electrical resistivity inversion tomography with spatial structural constraint, Yanshilixue Yu Gongcheng Xuebao/Chinese J. Rock Mech. Eng. 31 2258–2268.

[15] Singh U K, Tiwari R K and Singh S B 2010 Inversion of 2-D DC resistivity data using rapid optimization and minimal complexity neural network, Nonlinear Process. Geophys. 17 65–76

[16] Loke M H 2016 Tutorial: 2-D and 3-D Electrical Imaging Surveys, Revised Edition.

[17] Balarabe M, Abdullah K, Nawawi M N M and Khalil A E 2016 Monthly temporal-spatial variability and estimation of absorbing aerosol index using ground-based meteorological data in Nigeria Atmos. Climate Sci. 6 425–444

[18] Noori R, Hoshyariipour G and Ashrafi K 2010 Araabi, Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration Atmos. Environ. 44 476–482