**Empirical Investigation of Spatial Association Between Electrical Resistivity Values and Geotechnical Properties**

Mina Zamanian1; Asif Akhtar2; Mohsen Shahandashti3; and Sahadat Hossain4

1Graduate Research Assistant, Department of Civil Engineering, University of Texas at Arlington, 416 S. Yates St., Arlington, TX 76010 (corresponding author). email: [mina.zamanian@uta.edu](mailto:mina.zamanian@uta.edu)

2Graduate Research Assistant, Department of Civil Engineering, University of Texas at Arlington, 416 S. Yates St., Arlington, TX 76010. email: [mdasif.akhtar@mavs.uta.edu](mailto:mdasif.akhtar@mavs.uta.edu)

3Assistant Professor, Department of Civil Engineering, University of Texas at Arlington, 416 S. Yates St., Arlington, TX 76010. email: [mohsen@uta.edu](mailto:mohsen@uta.edu)

4Professor, Department of Civil Engineering, University of Texas at Arlington, 416 S. Yates St., Arlington, TX 76010. email: [hossain@uta.edu](mailto:hossain@uta.edu)

# Abstract

Evaluation of subsurface conditions and determination of geotechnical properties are essential for the design and construction of any project. Since conventional geotechnical site investigation methods are expensive, slow, and invasive, quantifying the geotechnical properties using a cost-effective, rapid, and non-invasive technique (i.e., geophysical methods) has gained attention in the recent decades. The ordinary least squares regression modeling has been widely used in the literature to quantify relationships between electrical resistivity values and geotechnical properties. However, spatial autocorrelation of data is ignored in these models, leading to biased and misleading inferences about the regression parameters. The objective of this study is to empirically investigate the spatial association between the electrical resistivity values and geotechnical properties using spatial regression models. The results show strong evidence of positive spatial autocorrelation between the OLS regression residuals, which affects the regression model's significance. It is also concluded that taking into account the spatial dependence of the observations in the standard regression model could improve the performance of the statistical models and lead to a better understanding of the effects of geotechnical properties on the variability of electrical resistivity values. These results are validated by creating spatial regression models that are capable of investigating the spatial effects on the variability of electrical resistivity values. The analyses were performed using 536 data points collected from laboratory soil physical property and electrical resistivity tests conducted on soil samples taken from different locations in Texas, US. Three spatial regression models, including Spatial Durbin Model (SDM), Spatial Lag model (SAR), and Spatial Error Model (SEM) were examined. The results show that the SEM model is the most appropriate compared to the SDM, SAR, and OLS to explain the spatial variability of electrical resistivity values based on the geotechnical properties variations. This research's findings contribute to the state of knowledge by investigating and quantifying the spatial relationship between the electrical resistivity values and geotechnical properties of clayey soils.

# Introduction

The accurate determination of geotechnical properties at a site is critical for any successful construction or development activity (Cosenza et al., 2006; Sudha et al., 2009). It is also essential to assess the stability of slopes along the roads and highways and identify the critical segments prone to failures (Shahandashti et al., 2020; Baral et al., 2021). Inadequate site investigation is the primary source of project delays and cost overruns in up to 50% of all infrastructure projects (Baynes, 2010). It may also lead to unexpected circumstances during construction or even lead to the failure of a project (Sirles 2006). Meanwhile, the conventional geotechnical site investigation methods are invasive, slow, expensive, and inherently limited in providing a continuous assessment of the subsurface (Kibria and Hossain, 2012; Siddiqui and Osman, 2012; Lin et al., 2017). Therefore, quantifying the geotechnical properties based on geophysical parameters obtained by non-invasive, rapid, and cost-effective geophysical investigation methods using statistical analysis has been of interest to many studies in recent decades (Cosenza et al., 2006; Kowalczyk et al., 2014). These methods could also provide a continuous assessment of the subsurface.

Among the geophysical methods, the electrical resistivity technique is widely used in the literature to characterize the geotechnical properties of clayey soils based on the electrical resistivity values. The soil electrical resistivity is a function of many factors such as moisture content, unit weight, porosity, pore fluid conductivity, degree of saturation, organic content, clay content, fabric structure, temperature, salinity, acidity, and soil compressibility (Rinaldi and Cuestas, 2002; Yang, 2002; Giao et al., 2003; Friedman, 2005; Samouëlian et al., 2005; Lapenna et al., 2005; Ekwue and Bartholomew, 2010; Kibria, 2014). Therefore, empirical and analytical studies have been conducted to develop statistical models using linear, power, and exponential regression functions to investigate the effects of different hydraulic and solid phase properties of clayey soils on the electrical resistivity. Among the hydraulic properties, soil moisture content has been identified as one of the major factors affecting the soil electrical resistivity (Friedman, 2005; Samouelian et al., 2005; Robinson et al., 2008a). Besson et al. (2010) also showed that 48% of the total variations of the electrical resistivity are attributed to the moisture content of the soil. The soil electrical resistivity decreases as the moisture content increases since the electrical current can be easily transmitted through the movement of ions in pore water (Siddiqui and Osman, 2012). The indirect relationship between the moisture content and electrical resistivity of clayey soils was also identified by Michot et al. (2003), Kibria and Hossian (2012), Siddiqui and Osman (2012), Abidin et al. (2013), and Rezaei et al. (2018). Abu-Hassanein et al. (1996) studied the effect of degree of saturation on the soil electrical resistivity. They concluded that an increase in the degree of clayey soil saturation leads to a decrease in the electrical resistivity values. Rinaldi and Cuestas (2002) showed that the void ratio (one of the controlling factors of moisture content and degree of saturation) has a significant effect on the variations of electrical resistivity. Abu-Hassanein et al. (1996) and Lin et al. (2016) indicated that the electrical resistivity of soils is influenced by their dry unit weights. They showed that there is an inverse relationship between the dry unit weight and electrical resistivity value. Nevertheless, the variability of electrical resistivity is less sensitive to the variations of unit weight than the moisture content, and it is almost negligible at the moisture contents above 30% (Kibria and Hossain, 2012). The effects of Atterberg limits on the variations of electrical resistivity were studied by Abu-Hassanein et al. (1996) and Long et al. (2012). They showed that the lower values of electrical resistivity are associated with the higher values of the plasticity index/liquid limit. It was also concluded that the percentage of fines (percent of soil finer than 75 microns) or percentage of clay (percent of soil finer than 2 microns) of soils impact the electrical resistivity of fine-grained soils.

Table 1 shows examples of empirical studies relating the electrical resistivity value to the geotechnical properties. The standard linear regression model has been widely used to explain the variability of electrical resistivity with the moisture content, plasticity index, void ratio, and porosity. The second-order regression models (quadratic regression models) are proposed by Kibria and Hossain (2012) and Lin et al. (2016) to study the effects of unit weight on the variations of electrical resistivity values. The power law and exponential regression functions have also been used to provide estimates for the unit weight, degree of saturation, and porosity using electrical resistivity values.

**Table 1.** Examples of empirical studies relating the electrical resistivity to the geotechnical properties

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Authors** | **Soil type** | **No. of data points** | **Correlation** | **Parameter values a** | **Coefficient of determination** |
| Goyal et al., 1996 |  |  | Linear, w-R | a = 500, b = -10 | 0.980 |
| Michot et al., 2003 | Loamy clay | 30-250 | Linear, w-R | a = 28.5 to 37.7,  b = -0.05 to 0.36 | 0.212 – 0.941 |
| Cosenza et al., 2006 | Sand and clay | 20 | Power law, R-w | a = 1.187, b = -2.444 | 0.821 |
| Kibria and Hossain, 2012 | Clay | 59 | Linear, R-w  Power law, R-Sr  Quadratic, R-γ | a = 119.26 to 328.03,  b = -1.094 to -1.351  a = 2.41 to 2.73,  b = -1.64 to -0.58  a = 0.095 to 0.7107,  b = -24.541 to 3.461,  c = 34.099 to 217.98 | 0.810 – 0.880  0.550 – 0.960  0.98 – 100.0 |
| Siddiqui and Osman, 2012 |  |  | Linear, w-ln(R)  Power law, γ – R | a = 0.644, b = -0.0451  a = 14.999, b = 0.0353 | 0.659  0.368 |
| Abidin et al., 2013 | Clayey silt | 25 | Power law, w-R | a = 121.88, b = -0.363  a = 109.98, b = -0.268 | 0.69 - 0.89 |
| Osman et al., 2014 | Clay | 16 | Power law, w-R | a = 81.12, b = -0.34 | 0.818 |
| Akinlabi and Adeyemi, 2014 |  |  | Linear, PI-R | a = 29.04, b = - 0.002 | 0.920 |
| Fallahsafari et al., 2010 | Clay | 25 | Exponential, w-R  Exponential, γ-R  Linear, e-ln(R)  Linear, n-ln(R) | a = 21.66, b = -0.19  a = 11426, b = 0.181  a = 0.702, b = -0.36  a = 0.415, b = -0.18 | 0.619  0.568  0.484  0.480 |
| Lin et al., 2016 | Marine clay |  | Power law, w-R  Exponential, PI-R  Linear, e-ln(R)  Quadratic, γ-R | a = 427.8, b = -1.13  a = 124.34, b = -0.239  a = 4.1663, b = -1.458  a = 0.16, b = -0.0166, c = 20.6 | 0.930  0.850  0.880  0.720 |
| Jusoh, and Osman, 2017 | Clay |  | Power law, w-R  Linear, PI-ln(R) | a = 123.93, b = −0.252  a = 29.793, b = -2.71 | 0.816  0.634 |
| Hazreek, et al., 2018 | Clayey silt | 25 | Power law, w-R | a = 110.68, b = -0.347 | 0.938 |
| Rezaei et al., 2018 |  | 15 | Power law, R-w | a = 2028.2, b = -1.496 | 0.68 |

Note: “R” denotes resistivity, “w” denotes moisture content, “γ” denotes unit weight, “PI” denotes plasticity index, “e” denotes void ratio, “n” denotes porosity, and “Sr” denotes degree of saturation.

a Coefficient of a, b, and c represent constant parameters in the linear (y=a+b.x), power law (y=a.xb), exponential (y=a.exp(b.x)), and quadratic (y=a.x2+b.x+c) regression functions.

Although most of these studies presented models with a relatively high goodness-of-fit, none of them have investigated the spatial association between the electrical resistivity values and geotechnical properties. The standard regression method is not the correct method for studying the spatial relationship between these variables. The presence of autocorrelated residuals in the standard regression model leads to wrong interpretations about the regression parameters and goodness-of-fit of the models. Spatial regression models consider the spatial dependence of the error terms to accurately determine the effects of a change in geotechnical properties on the variability of electrical resistivity values.

The objectives of this research are to (1) investigate whether there is a spatial association between the electrical resistivity and geotechnical properties and (2) determine the most appropriate spatial regression model to explain the variability of electrical resistivity values with the variations of geotechnical properties. The research approach to achieve the objectives of this paper is explained in the next section. The research approach is followed by presenting the empirical results. The conclusions of this research are discussed in the last section.

# Research Approach

The spatial relationship between the electrical resistivity and the geotechnical properties of clayey soils is investigated using spatial regression models to avoid misinterpretations resulting from ignoring the spatial effects in the standard regression model. The spatial regression approach is used in this paper for the first time to accurately determine the impact of geotechnical properties on the variability of electrical resistivity values of clayey soils while considering their spatial relationships. Figure 1 presents the flowchart of this research approach to collect data, investigate the spatial autocorrelation of the residuals, and develop spatial regression models. The research approach includes five main parts:

1. Collect soil samples from different locations and perform necessary laboratory tests to characterize the soil properties,
2. Investigate the spatial autocorrelation of the regression residuals using Moran’s I test statistic,
3. Perform the spatial regression analyses to study the relationships between the electrical resistivity and geotechnical properties by considering the spatial effects in the regression model using spatial weight matrices, and
4. Evaluate the performance of the spatial and standard regression models to determine a proper model to investigate the variability of electrical resistivity based on the variations of geotechnical properties using Lagrange Multiplier tests, Likelihood Ratio test, and diagnostic statistics of created models.



Figure 1. Flowchart of research approach to collect data, investigate the spatial autocorrelation, and develop spatial regression models

## Data Collection

Soil samples were collected from different locations across the state of Texas, US. Figure 2 shows the number of boreholes and obtained soil samples on the map of Texas. A total of 536 data points were collected from the laboratory physical property tests (e.g., moisture content, Atterberg limits, and specific gravity) and laboratory electrical resistivity tests on the obtained soil samples. The laboratory electrical resistivity tests were carried out at different moisture contents (in the range of 5 to 45%) and dry unit weights (in the range of 60 to 100 pcf) to investigate their effects on the variations of the electrical resistivity values. The laboratory electrical resistivity tests were conducted according to ASTM G187-05 standard test method using AGI SuperSting R8 instrument. To eliminate the variability of electrical resistivity measurements as a result of temperature variations, the measured resistivity values were corrected at a reference temperature of 15.5°C (60°F) using the following equation (ASTM G187-05):

(1)

where R15.5 is the corrected electrical resistivity at 15.5°C, RT is the measured electrical resistivity at the temperature T°C. Besides, Atterberg limits and specific gravity of soil samples were measured according to the ASTM D4318-17 and ASTM D854-14 standard test methods, respectively. The plasticity indices were observed in the range of 10.5 to 46.5, which are classified as low (CL) to high (CH) plasticity clayey soils according to the USCS (Unified Soil Classification System). The measured specific gravity values were used to determine the degree of saturation and void ratio for each data point.

Map

Description automatically generated

Figure 2. Number of boreholes and collected soil samples on clay map of Texas, US (Adapted from Olive et al. (1989))

## Spatial Autocorrelation

Spatial autocorrelation (also known as spatial dependence) is the degree of dependency among the similar/dissimilar neighboring observations and mainly emerges when the observations are collected from different locations in space. Linearity, homoscedasticity, independence, and normality are four critical assumptions associated with the linear regression model (Neter et al., 1996). Before making inferences regarding the model estimates, the model assumptions need to be checked by evaluating the residual plots and performing diagnostic tests such as Breusch-Pagan test for homoscedasticity, Shapiro-Wilk test for normality, and Moran’s I test for spatial autocorrelation. If any of the assumptions are violated, the OLS model is inappropriate and statistical inferences from the model are unreliable (Voss et al., 2006). In this paper, since the data were collected from different locations, the spatial association between the electrical resistivity values and geotechnical properties was investigated. Moran’s I test was used to examine the existence of an overall clustering in the OLS regression residuals. The Moran’s I test is represented as follow:

where n is the number of spatial units, x is the variable of interest, is the mean of x, and is an element of a spatial weight matrix. The spatial weight matrix (W) was used to identify the spatial structure of the observations. Each element of this matrix defines the dependency between two observations (Getis, 2009). The spatial weight matrix has different experimental forms based on the geometry of the spatial units, either by their boundaries or distances from each other (Anselin, 2005). The selection of a proper weight function is essential to achieve convincing results from spatial modeling, especially when the spatial autocorrelation is strong (Yan-guang, 2009; Elhorst, 2010). In this paper, the general distance between the locations of the collected soil samples was utilized to identify the neighboring structure of the observations and construct the spatial weights. The distance-based weight matrices are the most appropriate form for a data set with point locations (Anselin and Sergio, 2014). If “dij” denotes the distance between the location of i and j, and “d” indicates a threshold distance beyond which there is no direct spatial influence between the observations, the spatial weights of the corresponding weight matrix is constructed as follow (Chen, 2012):

which gives a binary matrix of 0 and 1. Typically, there is no unique approach to determine the threshold distance for identifying the neighboring locations (Walker et al., 2000; Anselin, 2005). The most widely used approaches are to assess the robustness of estimated spatial regression models and the magnitude of Moran’s I for a series of threshold distances. The distance at which the model shows the maximum log-likelihood value, highest Moran’s I value, highest pseudo-R-squared, and lowest residual standard error is determined as the appropriate threshold distance (Wang et al., 2007; Chi and Zhu, 2008; Stakhovych and Bijmolt, 2009; Elhorst, 2010). The other approach is to identify the threshold distance by creating a semi-variogram of the variables (Hession and Moore, 2011). The off-diagonal elements of the weight matrix with non-zero values denote the dependency of the neighboring observations. However, the diagonal elements of the weight matrix represent the self-influence of the observations that were excluded from the spatially lagged variables (i.e., diagonal elements of the weight matrix were set to zero). Then the weight matrix was standardized using a row-normalization approach in which all the weights in each row sum to unity ().

The null hypothesis of the Moran’s I test is that the regression residuals are randomly distributed in space. By rejecting the null hypothesis, it is concluded that there is evidence of spatially autocorrelated residuals (alternative hypothesis). Ignoring the presence of spatial dependence in the OLS model leads to underestimation or overestimation of actual variance in the case of positive and negative dependence, respectively, which consequently affects the significance of the model (Cressie, 1993; Schabenberger and Gotway, 2005). Moran’s I values range from -1 and +1, and its significance is evaluated using a P-value and z-score. The negative values represent the clustering between dissimilar values, while positive values represent the clustering between similar values. The zero value for Moran’s I implies that there is no spatial autocorrelation in the regression residuals, and the residuals are randomly distributed.

## Spatial Regression Models

The spatial dependence between the observations is accounted into a regression model using the spatial weight matrix by three methods; (1) inclusion of the effect of a change in the dependentvariable of one location on the dependent variable of a neighboring location (endogenous interaction effects), (2) inclusion of the effect of a change in the independent variables of one location on the dependent variable of a neighboring location (exogenous interaction effect), and (3) inclusion of the effect of dependency in the residuals in one location on a neighboring location (Calderon, 2009; Manski, 1993). In this study, three spatial regression models were examined: Spatial Durbin Model (SDM), Spatial Lag or Autoregressive Model (SAR), and Spatial Error Model (SEM). The SDM is a general model that includes both endogenous and exogenous interaction effects and has the form of:

(3)

where R is an (n×1) vector of observations on the electrical resistivity (dependent variable), X is an (n×k) matrix of observations on the geotechnical engineering parameters (independent variables), W is an (n×n) matrix of spatial weight, β is a (k×1) vector of regression parameters, ρ is a coefficient on the spatially lagged dependent variable, θ is a (k×1) vector of the spatially lagged independent variable, and ε is an (n×1) vector of independently and identically normally distributed errors. In this research, to avoid multicollinearity in the analyses, the geotechnical parameters (degree of saturation, liquid limit, and void ratio) with the lowest significant test statistics that have a high correlation with the other variables were removed from the model. The moisture content, dry unit weight, and plasticity index were selected as independent variables to explain the variability of electrical resistivity in the analyses. The SAR model only includes the endogenous interaction effects (θ=0 in equation 3) and expressed as:

(4)

where the variables are defined as the same for the SDM model. In contrast, the spatial dependence in the SEM is modeled only by the spatially lagged error terms and consider neither the exogenous nor endogenous interaction effects (θ=0 and ρ=0 in equation 3), which has the form of:

(5)

where 𝜆 is a spatial error lag coefficient. The Lagrange Multiplier (LM) tests were performed on the OLS residuals to decide whether the spatial lag (SAR) or spatial error model (SEM) is the most appropriate model for the analysis of the data (Anselin, 2005). There are four LM test statistics: standard LM-Error, standard LM-Lag, Robust LM-Error, and Robust LM-Lag. First, the standard LM tests are performed, and then the model with the significance test statistic is selected. If neither of the tests is significant, it indicates that the OLS model is more appropriate. However, if both standard LM tests are significant, which commonly happen in practice, the Robust forms of LM test are used, and the model with the (most) significance test statistic is selected as the most appropriate model (Anselin, 2005). Another approach is to start with the widely used model (i.e., SDM) if there is a global effect (Lesage, 2014). Then to further evaluate the goodness-of-fit of the nested models (when a complex model can be reduced to a simpler model by restricting certain parameters), the Likelihood Ratio (LR) test was utilized (Anselin, 2005). The null hypothesis of the test is that a complex model should be reduced to a simpler model by restricting some of the model parameters. By rejecting the null hypothesis, it is concluded that the complex model is more appropriate and should not be restricted to the simpler model (alternative hypothesis). Log-Likelihood (LIK), Akaike’s Information Criterion (AIC), and Bayesian Information Criterion (BIC) (also known as Schwarz Information Criterion (SIC)) were also used to compare the performance of the non-nested models (Yang and Fik, 2014; LeSage, 2008). The model with the highest LIK and lowest AIC or BIC was considered as the best model that fits the data.

# Research Results

## Results of OLS Analysis

The OLS regression model was fitted to the electrical resistivity data to test the performance of the model in defining a relationship between the geotechnical parameters and electrical resistivity values and check the model assumptions. Table 2 presents the results of the fitted standard regression model (OLS) before and after transforming the electrical resistivity values.

**Table 2.** Summary of results of OLS model before and after transforming the electrical resistivity values

|  |  |  |
| --- | --- | --- |
|  | OLS with no transforamtion | OLS using Box-Cox transformation |
| Intercept | 310.01\*\* | -0.611\*\* |
| Moisture Content | -2.989\*\* | 0.012\*\* |
| Dry Unit Weight | -18.73\*\* | 0.044\*\* |
| Plasticity Index | 0.656\* | 0.005\*\* |
| R-squared | 0.27a | 0.81a |
| Standard Error of Residual | 54.44 | 0.063 |
| LIK | -2901.10 | 726.27 |
| AIC | 5812.02 | -1442.5 |
| BIC | 5833.44 | -1421.1 |
| No. of Observations | 536 | 536 |

Notes: ‘\*’ indicates the significance at the 5% level and ‘\*\*’ indicates the significance at the 1% level

a Adjusted R-squared.

The results of the initial analysis with no transformation on the electrical resistivity values show that the moisture content and dry unit weight have significant inverse relationships with the electrical resistivity value. However, the plasticity index shows a significant direct relationship with the electrical resistivity value, which is inconsistent with the literature. Figure 4 illustrates the residual plots of the OLS regression model with no transformation on the electrical resistivity values. The presence of a funnel in the plot of residuals versus fitted values and skewness in the normal probability plot are indications of heteroskedasticity and non-normality of the error terms, respectively. Besides, the Breusch-Pagan test and Shapiro-Wilk test show that the assumptions of homoskedasticity and normality of linear regression model are not satisfied (rejection of null hypotheses at the 10% level of significance). Therefore, the electrical resistivity values were transformed using the Box-Cox transformation to stabilize the error variance and mitigate the problem of non-normality of the error terms.



**Figure 4.** Residual plots of OLS model with no transformation on the electrical resistivity values; (a) residuals versus fitted values and (b) normal probability plot.

The results of the OLS analysis with the transformed response variable (λ=-0.5) are presented in Table 2. The results show that the moisture content, dry unit weight, and plasticity index have significant inverse relationships with the electrical resistivity value, which is consistent with the literature. Figure 5 illustrates the residual plots of the OLS regression model using Box-Cox transformation on the electrical resistivity values. No clear pattern can be observed in the plot of residuals versus fitted values shown in Figure 5. The Breusch-Pagan test also shows that the assumption of the constant variance of error terms is satisfied after transformation at the 5% level of significance. The skewness of data in the normal probability plot is removed after transforming the electrical resistivity values, however, it can be observed that the residuals are less spread than the normal distribution (lighter-tailed). The Shapiro-Wilk test also confirms that the normality assumption is violated (rejection of null hypothesis at the 10% level of significance). Although the non-normality of error terms has remained even after transformation, no more transformations were used since the OLS model is relatively robust to non-normality in the absence of skewness (Neter et al., 1996).



Figure 5. Residual plots of OLS model using Box-Cox transformation on the electrical resistivity values; (a residuals versus fitted values and (b) normal probability plot.

The Moran’s I test provides strong evidence of positive spatial autocorrelation (Moran’s I=0.24 with P-value < 1%, z-value = 28.7) in the regression residuals, implying that the residuals are not independently distributed. Since the assumption of independence of the linear regression is violated, the OLS model might be an inappropriate approach to quantify the relationship between electrical resistivity and geotechnical properties. Besides, any statistical inferences regarding the coefficient estimates might be unreliable. Therefore, the spatial regression models were examined to account for the spatially autocorrelated residuals.

## Results of Spatial Regression Analysis

The simple and robust forms of Lagrange Multiplier tests (LM) were used on the OLS results to determine the most appropriate model for the analysis. Table 3 presents a summary of results of LM tests for the OLS residuals. Since both simple tests (LM error, LM lag) are highly significant and suggest using the spatial regression models, the robust form of LM error and LM lag tests were tested. The robust tests also show highly significant values for both SAR and SEM; however, it appears that the test statistic for the spatial error model (SEM) is more significant.

**Table 3.** Summary of results of Lagrange Multiplier tests for the OLS residuals

|  |  |
| --- | --- |
| **Test** | **Result** |
| LM error | 699.18\* |
| LM lag | 299.68\* |
| Robust LM error | 454.93\* |
| Robust LM lag | 55.426\* |

Notes: ‘\*’ indicates the significance at the 1% level.

Although it is concluded from the LM tests that the SEM is the most appropriate model, the SAR and SDM were also examined to compare the performance of different spatial regression models. The results of these spatial analyses with the transformed electrical resistivity values are summarized in Table 4.

**Table 4.** Summary of results of spatial regression models with the transformed electrical resistivity values

|  |  |  |  |
| --- | --- | --- | --- |
|  | SDM | SAR | SEM |
| Intercept | -0.455\* | -0.791\* | -0.573\* |
| Moisture Content | 0.012\* | 0.012\* | 0.012\* |
| Dry Unit Weight | 0.044\* | 0.045\* | 0.044\* |
| Plasticity Index | 0.004\* | 0.004\* | 0.003\* |
| Lag. Moisture Content | -0.007\* |  |  |
| Lag. Dry Unit Weight | -0.015\* |  |  |
| Lag. Plasticity Index | -0.001\* |  |  |
| ρ / 𝜆 Coefficient | 0.724 | 0.552 | 0.809 |
| R-squared | 0.86a | 0.85a | 0.86a |
| Standard Error of Residual | 0.054 | 0.055 | 0.054 |
| LIK | 801.44 | 794.03 | 798.29 |
| AIC | -1584.8 | -1576.1 | -1584.6 |
| BIC | -1546.3 | -1550.4 | -1558.9 |
| No. of Observations | 536 | 536 | 536 |

Notes: ‘\*’ indicates the significance at the 1% level.

a Pseudo-R-squared.

According to Table 4, the spatial lag coefficient of the SDM is positive, meaning that a change in the electrical resistivity of one location has positive effects on the electrical resistivity values of neighboring locations. These effects decay as moving towards higher-order neighbors. In other words, the variations of electrical resistivity values in one location influence the electrical resistivity of nearby locations more than further locations. The likelihood ratio test and Wald statistics show that the spatial lag coefficient of the SDM (ρ=0.724) is significant at the 1% level. Similarly, the spatial lag coefficient of the SAR model is positive and significantly different from zero at the 1% level (ρ=0.552). The SAR model presents positive but lower spillover effects in the neighboring locations rather than the SDM. The spatial error lag coefficient of the SEM is positive and significantly different from zero at the 1% level (𝜆=0.809). It shows the strength of spatial autocorrelation among the error terms meaning that the unexplained variabilities of the electrical resistivity values follow a systematic distribution in space.

The results of Lagrange Multiplier diagnostic tests for the spatial dependence show that the SEM and SAR models are removed the problem of spatially autocorrelated residuals at the 1% level of significance. However, the spatial autocorrelation has remained in the SDM residuals (null hypothesis is rejected at the 1% level of significance). Therefore, the performance of the SEM and SAR models are further assessed by evaluating the coefficients of geotechnical parameters, considering the results of the LR test, and evaluating the LIK, AIC, BIC statistics.

The signs and magnitudes of the SEM model parameters are similar to the standard regression model. They are also highly significant for the three geotechnical parameters (moisture content, dry unit weight, and plasticity index). Since the coefficients of the SAR model do not accurately explain the effects of geotechnical properties on the electrical resistivity, direct comparison of the regression parameters of the SAR model and standard regression model is inappropriate (LeSage and Dominguez, 2012). Therefore, the average direct, indirect, and total effects of a change in each of the three geotechnical parameters on the electrical resistivity were calculated for the SAR model and summarized in Table 5.

**Table 5.** Average effects of explanatory variables on the electrical resistivity values for the SAR model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Direct effect | Indirect effect | Total effect | P-value |
| Moisture Content | 0.012 | 0.014 | 0.026 | 0.000\* |
| Dry Unit Weight | 0.045 | 0.054 | 0.099 | 0.000\* |
| Plasticity Index | 0.004 | 0.005 | 0.009 | 0.000\* |

Notes: ‘\*’ indicates the significance at the 1% level.

According to Table 5, the corresponding direct effects of moisture content, dry unit weight, and plasticity index are smaller than the indirect effects, holding the same signs, which are associated with the transformation used on the electrical resistivity values. Similar to the standard regression model, the total effects of geotechnical properties on the electrical resistivity value are positive and highly significant at the 1% level. It again implies that an increase in the geotechnical properties has a decreasing effect on the electrical resistivity values. A noticeable difference is that the coefficients of the geotechnical parameters in the SAR model are shifted toward positive values compared to the standard regression model due to considering both direct and indirect effects. The coefficient variations imply that the variability of electrical resistivity is less influenced by the variation of geotechnical properties while considering the spatial effects.

Moreover, the likelihood ratio (LR) test was utilized to evaluate the goodness-of-fit of the nested models (i.e., SAR and OLS, or SEM and OLS). The test results show that the SAR and SEM models outperform the non-spatial regression model at the 1% level of significance, and they should not be restricted to a simpler model (i.e., OLS model). Comparing the LIK, AIC, BIC statistics, it appears that the SEM is a better fit to the electrical resisitivity data compared to the SAR. Therefore, according to the diagnostic tests and statistics, it is concluded that the spatial error model (SEM) is the best spatial model compared to the SDM and SAR models. Besides, it provides more accurate estimates of the regression parameters in comparison to the standard regression model due to considering the spatial effects in the analysis.

## Robustness of Spatial Regression Models Based on Threshold Distance

In this paper, different threshold distances (0.25, 0.5, 1, 2, 4, 6, 8, 10, 20, 30, 50, and 100 mi) were examined to construct spatial weight matrices to assess the robustness of spatial regression models and investigate the spatial autocorrelation in the regression residuals of electrical resistivity data (at shorter threshold distances than 0.25-mile, no neighbor was found for some locations). Table 6 presents the values of log-likelihood, pseudo-R-squared, residual standard error, and Moran’s I of OLS residual considering different threshold distances for the SEM and SAR. Although the log-likelihood of the SEM shows more variation than the SAR model, its value decreases as the threshold distance increases in both models. The log-likelihood has the highest value at 0.25-mile threshold distance in both models. Similarly, the value of Moran’s I decreases as the threshold distance increases and has the highest value at 0.25-mile threshold distance. The pseudo-R-squared and residual standard error have approximately constant values at different lag distances. Therefore, a threshold distance of 0.25-mile (0.4 km) was determined to construct the spatial weights and perform the spatial regression analyses on the electrical resistivity data based on the highest log-likelihood, highest Moran’s I, highest pseudo-R-squared, and lowest residual standard error.

**Table 6.** Variations of log-likelihood, pseudo-R-squared, residual standard error, and Moran’s I of OLS residual considering different threshold distances for the SEM and SAR

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Spatial Model** | **Threshold Distance**  **(mile)** | **Log-Liklihood** | **Pseudo-R2** | **Residual Standard Error** | **Moran’s I**  **OLS Residual** |
| SEM | 0.25 | 798.29 | 0.86 | 0.0538 | 0.243\* |
| 0.5 | 791.09 | 0.85 | 0.0547 | 0.232\* |
| 1 | 793.55 | 0.85 | 0.0546 | 0.225\* |
| 2 | 780.43 | 0.84 | 0.0560 | 0.176\* |
| 4 | 786.99 | 0.85 | 0.0553 | 0.195\* |
| 6 | 786.99 | 0.85 | 0.0553 | 0.195\* |
| 8 | 781.74 | 0.84 | 0.0559 | 0.177\* |
| 10 | 781.74 | 0.84 | 0.0559 | 0.177\* |
| 25 | 786.59 | 0.85 | 0.0554 | 0.191\* |
| 30 | 781.74 | 0.84 | 0.0559 | 0.177\* |
| 50 | 782.37 | 0.84 | 0.0560 | 0.177\* |
| 100 | 703.40 | 0.83 | 0.0586 | 0.159\* |
| SAR | 0.25 | 794.03 | 0.85 | 0.0547 | 0.243\* |
| 0.5 | 789.48 | 0.85 | 0.0553 | 0.232\* |
| 1 | 791.04 | 0.85 | 0.0551 | 0.225\* |
| 2 | 780.11 | 0.84 | 0.0563 | 0.176\* |
| 4 | 784.76 | 0.85 | 0.0558 | 0.195\* |
| 6 | 784.76 | 0.85 | 0.0558 | 0.195\* |
| 8 | 782.37 | 0.84 | 0.0560 | 0.177\* |
| 10 | 782.37 | 0.84 | 0.0560 | 0.177\* |
| 25 | 786.60 | 0.85 | 0.0556 | 0.191\* |
| 30 | 782.37 | 0.84 | 0.0560 | 0.177\* |
| 50 | 782.37 | 0.84 | 0.0560 | 0.177\* |
| 100 | 773.65 | 0.84 | 0.0570 | 0.159\* |

Notes: ‘\*’ indicates the significance at the 1% level.

# Conclusions

The presence of spatial autocorrelation in the OLS regression residuals could lead to wrong and unreliable interpretations about the effects of geotechnical properties on the electrical resistivity values. This wrong interpretation could result in a poor evaluation of the subsurface condition. The Moran’s I of the OLS regression residuals showed a highly significant value (I = 0.24), indicating that the regression residuals are spatially autocorrelated. Since the standard regression model could be inappropriate in the case of spatially autocorrelated residuals, the spatial regression models were examined. The spatial regression models were utilized in this paper for the first time to investigate the spatial association between the electrical resistivity values and influencing geotechnical properties of clayey soils (i.e., moisture content, dry unit weight, and plasticity index).

Using the likelihood ratio test for the nested models, it was found that the spatial regression models outperform the OLS model in explaining the variability of electrical resistivity with the variation of geotechnical properties. Unlike the SDM, no evidence of spatial autocorrelation was found in the SEM and SAR residuals. The Lagrange Multiplier tests showed that the SEM is a better fit to the electrical resistivity data compared to the SAR. Moreover, the LIK, AIC, and BIC proved that the SEM is the most appropriate model compared to spatial and non-spatial models. These findings indicate that the inclusion of spatial autocorrelation of residuals in the regression model could improve the performance of the regression model and lead to more accurate estimates for the effects of geotechnical properties on electrical resistivity values. This research's findings contribute to the state of knowledge by investigating and quantifying the spatial relationship between the electrical resistivity values and geotechnical properties of clayey soils. These findings help engineers to have a better understanding of the effects of geotechnical properties on the variability of electrical resistivity values to obtain more reliable evaluations of the subsurface characteristics using the electrical resistivity values.

# Data Availabaility

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.