Application of Multiple Linear Regression Technique to Predict Noise Pollution Levels and Their Spatial Patterns in the Tarkwa Mining Community of Ghana

Peter Ekow Baffoe¹, *, Alfred Allen Duker²

¹Department of Geomatic Engineering, Faculty of Mineral Resources Technology, University of Mines and Technology, Tarkwa, Ghana
²Department of Geomatic Engineering, School of Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Email address: pbaffoe@umat.edu.gh (P. E. Baffoe)
*Corresponding author

To cite this article:

Received: April 30, 2019; Accepted: June 4, 2019; Published: June 20, 2019

Abstract: Predicting and preventing intraurban noise levels in our communities are very challenging for urban planning, epidemiological studies and environmental management, especially in the developing world. Most existing noise-predicting models are limited in providing changes in noise levels during intraurban development and the corresponding noise pollution. In this study, noise levels were measured at 50 purpose-designed monitoring stations and then a land-use regression model was developed for the intraurban noise prediction applying the multiple linear regression (MLR) technique. The measured and the predicted noise levels were compared. These were further compared with noise estimates from a standard noise model, Lyons Empirical model. The results from the developed MLR model did not show any significant differences in the patterns as compared with those of the Lyons Empirical model. The model performance indicators showed a standard deviation of 1.585, high correlation (R) of 0.98, \(R^2\) of 0.961 and RMSE of 1.569. The resulting maps showed a heterogeneous distribution of the noise pollution levels in the community. This confirms the usefulness of the method for assessing the spatial pattern of noise pollution in a community. This makes it a useful tool for urban planning, epidemiological studies and environmental management.

Keywords: Noise Level, Noise Pollution, Noise Models, Land Use Regression Models

1. Introduction

Intraurban noise pollution is prevalent in our cities, especially in the mining communities where many factors contribute to noise pollution and distribution. Increase in urban noise pollution brings about associated health problems including hearing impairment, sleep disturbances, interference with spoken communication, cardiovascular problems, and disturbances in mental health, impaired task performance, negative social behavior and annoyance reactions [1]. In line with that, many researchers have studied the problem of noise pollution in many cities throughout the world [2-9]. Other studies have also confirmed that noise pollution is a threat to the health and well-being of humans [1, 10-12]. With the continuous trend of population growth, urbanisation and its associated varied and mobile sources of noise, noise pollution will increase in magnitude and severity. This therefore, calls for more research in this area of noise pollution. However, comprehensive noise exposure assessment techniques are not available.

Most noise predicting models, especially the existing Lyons Empirical model which is being applied for noise prediction in a similar community, are limited in predicting long term noise levels due to the fact that they are limited to only one variable namely, traffic noise. In view of this land-use regression (LUR) models, which could be used for a wide range of variables, are currently being used for assessing the exposure to air pollution [13]. This method uses least-squares regression modelling techniques to predict air pollution levels based on the monitoring data accessed at the
purpose-designed monitoring stations (PMS) and predictor variables collected mainly through Geographic Information Systems [13]. LUR models are very easy to adopt since they have quality performance in detecting environmental air pollutions in the urban areas, as their empirical structure requires the use of standardized approaches.

In the field of noise exposure, distribution and prediction, application of LUR modelling has been least explored. It was first applied in north-east China, where the technique was used in two different sites using 101 PML for model development and 101 PML for model validation. The model performance explained 83.2% variability of the noise pollution levels and was successfully used at three different scales [14]. The second application of the LUR was in three different European cities to explain variability of the intraurban noise pollution of the cities. The model performance was good with adjusted R² range of 0.66-0.87 and also 0.70-0.89 in both applications. The short-term noise measurements gave a correlation of 0.62-0.78 with noise estimates from the standard noise models when compared with it [13].

In line with the reported merits of LUR in literature, the aim therefore of this current study is to develop a generic LUR model for predicting noise pollution levels in urban areas, especially in the mining communities using MLR method. The results are potential for urban planning, epidemiological studies and environmental management.

2. Resources and Methods Used

2.1. Study Area

Tarkwa Mining Community (TMC) is an area on the south-western part of Ghana and is within the Tarkwa Nsuaem Municipality. The study area is geographically located between latitudes 4° 00’ 00” N and 5° 00’ 00” N and longitudes 1° 45’ 00” W and 2° 00’ 00” W, and is about 89 km north of Takoradi, the capital of the Western Region of Ghana.

It is an old mining town which is well noted for the mining of minerals such as gold and manganese. Goldfields Ghana Limited, Anglo-gold Ashanti, and Ghana Manganese Company are some of the large scale mining companies found in the TMC. There are also numerous allied mining companies located in the study area. Several small scale mining activities are also going on in the TMC. Over the past few years TMC has seen infrastructural developments including road constructions, building of health posts, education, industries, banking, hospitality services and private business development [15]. Figure 1 shows a map of the Tarkwa Mining Community.
2.2. Measurements of Field Data

The geo-spatial locations of the purpose-designed monitoring stations (PMS) in the TMC were surveyed using Garmin GPS 60CSx handheld Global Positioning System (GPS) of 2 m accuracy. A calibrated Larson Davis’s SoundTrack LxT Sound Level Meter was used to measure the noise levels in the study area. The outdoor noise levels were measured from August 2014 to January 2015. The measurements of the PMS were taken at street level and were also determined with the aid of the city digital map.

To avoid noise reflections, the noise-level meter was set on a tripod at about 1.5 m above the ground level and separated from the noise sources by at least 1.5 m. This decision was made in connection with what has been reported and accepted in the literature. For example, in [16] used 1.5 m above ground level and 1.22-1.52 m from the source of the noise. The tolerance of the calibrated Larson Davis’s Sound Track LxT trademark device is ±0.6 dBA. A-weighted instantaneous sound pressure levels were recorded three times daily at the selected positions in the study area. The total number of the points used for the modelling was 50.

2.3. Predicting Noise Pollution Levels Using MLR Approach

The noise predictive model for forecasting noise levels in the study area was developed by following the normal procedures for developing land-use regression model, using the multiple linear regression method. The noise level was defined as the dependent variable and the areas of the various land-uses within the study area were defined as independent variables.

Thus, the general equation consisted of five independent variables namely land-use, traffic intensity, road network, distance to the main road, and population density. Due to the heterogeneous nature of the independent variables, they were converted into a homogeneous one that is adaptable in the multiple linear regression equation. This was done using the Analytic Hierarchy Process (AHP) of the Multi-Criteria Decision Analysis technique, which is used to solve complex multi-criteria decisions. The equations formulated from AHP were solved using matrices.

2.3.1. Application of the Analytic Hierarchy Process

A mathematical expression was developed to provide weights for each criterion as a function of its rank. In computing the vector of criteria weights, a pairwise comparison matrix (A) was formulated for each pair of independent variables. The matrix A is an mxm matrix, where m is the number of evaluation criteria that was considered. Each entry, \( a_{jk} \) of matrix A represented the importance of the jth criterion compared to the kth criterion. The following conditions were set as defined in AHP procedures: if \( a_{jk} > 1 \), then the jth criterion is more important than the kth criterion and vice versa. If two criteria have the same importance, then the entry \( a_{jk} = 1 \). The entries, \( a_{jk} \) and \( a_{kj} \) satisfy the following constraint in Equation 1:

\[
\sum_{k=1}^{m} a_{jk} = 1 \quad \text{for} \quad j \neq k
\]

In this study, the importance of each factor criterion over the other was determined and quantified using the scale of pairwise comparison as developed by Saaty in the AHP [17].

The comparisons were quantified on a scale of 1 to 9. The quantity 1 represents two factors of equal importance. The quantity 9 also denotes a factor with extreme importance over the other [18]. Based on literature as well as experts’ opinion, the judgements for the independent variables were formulated. Each independent variable has five alternatives and five decision criteria. Each of the alternatives was evaluated in terms of the decision criteria and the relative importance (or weight) of each criterion.

The results are thus represented in normalized pairwise comparison matrices in proceeding Equations 2 and 3.

\[
\bar{a}_{jk} = \frac{a_{jk}}{\sum_{i=1}^{m} a_{ik}} \quad (2)
\]

After creating the normalized pairwise comparison matrices for the independent variables, the criteria weight vector, \( w \) is built by averaging the entries on each row of \( A_{cm} \) as shown in Equation 3.

\[
w_j = \frac{\sum_{j=1}^{m} \bar{a}_{ij}}{m} \quad (3)
\]

The whole processes were then summarized thus; the judgment tables was represented by 5 x 5 matrices and then squared to obtain an eigenvector. The result was then normalised by summing the eigenvector and dividing each value of the eigenvector by the sum. The weights for the individual factors were obtained after the normalisation process. The process was then repeated a number of times until the weights assigned to each factor were consistent. A consistency ratio of 0.02 was achieved which is less than the maximum allowable ratio of 0.10.

2.3.2. Applying the Multiple Linear Regression Approach

The prediction model used to estimate noise pollution levels in the study area was developed using multiple linear regression approach, with matrix notation and analyses using the Statistical Package for the Social Sciences (SPSS) and MATLAB. From the multiple linear regression (MLR) model the dependent variable is related to five independent variables. The general multiple linear regression expression for \( k \) variables is as given in Equation 4.

\[
y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + e_i, \quad i = 1, 2, \ldots, n \quad (4)
\]

In Equation 4, \( y \) is the dependent variable which is the noise pollution levels in the study area, \( \beta_1, \beta_2 \) and \( \beta_k \) are coefficients of the regression equation and \( \beta_0 \) is the intercept that is the value of \( y \) when all \( x \)s are zero.
In order to solve for the regression coefficient in the MLR model, as in Equation (4), the method of least squares is used. These coefficients (see Equation 4), illustrate the unrelated contributions of each independent variable towards predicting the dependent variable. It is important to note that, the computations used in finding the regression coefficients (βi, i = 1, 2, 3, ..., k), residual sum of square (SSE), regression sum of squares (SSR), etc is complex therefore, the multiple regression model in terms of observations were written using matrix notation.

This is because using matrix allows for a more compact framework in terms of vectors representing the observations, levels of regressor variables, regression coefficients, and random errors. Therefore, Equation 4 was represented in a compact as in Equation 5:

\[ Y = X\beta + \epsilon \]  \hspace{1cm} (5)

Where, Y is noise pollution levels, X represents the independent variables; ε is residuals; and A is the value of Y when all Xs are zero. After formulating the matrix equations, matrix commands in Excel were used to solve the equations and the results compared with that from SPSS software. Finally using MATLAB, the values for the coefficient of regression β were calculated and the prediction equation was presented as in Equation 6.

\[ Y = h_0 + h_1x_1 + h_2x_2 + h_3x_3 + h_4x_4 + h_5x_5 + \epsilon \]  \hspace{1cm} (6)

Where, Y is dependent variable, what is being modelled or predicted, x is explanatory variables, variables that influence or help explain the dependent variable, b is Coefficients, values computed by the regression equations, reflecting the relationship and strength of each explanatory variable to the dependent variable; ε is residuals, the portion of the dependent variable that isn’t explained by the model; the model under and over predictions.

The least square estimator of the coefficients of the regression equation (β) is given by Equation 7.

\[ \beta = (X^T X)^{-1} (X^T Y) \]  \hspace{1cm} (7)

Since X is not usually a squared matrix, it is multiplied by the transpose of X, that is X'X and the inverse of (X’X) calculated. Hence the estimator β is calculated thus using Equation 7.

The data was populated using the MATLAB software and the predicted noise levels as projected for the future were then used to develop the spatial distribution of the estimated noise levels.

2.3.3. Applying the Lyons Empirical Model

In order to confirm the developed MLR noise prediction model, the Lyons Empirical prediction model was compared to it. The Lyons Empirical Model is a mathematical model developed and applied in an area of similar conditions as that of the TMC by Lyons in [19]. The long term estimates of noise pollution levels by this model were achieved by using Equation 8.

\[ L = 10\log V - 15\log D + 30 \log S + 10 \log \tanh (1.19 \times 10^{-3}) \times VD/S + 29 \]  \hspace{1cm} (8)

indicated ‘Sites’ along the major road (as explained earlier) and the computed values of noise level descriptors for the various locations in the study area.

<table>
<thead>
<tr>
<th>Sites</th>
<th>L_{Aeq}</th>
<th>L_{Leq}</th>
<th>L_{TNI}</th>
<th>L_{Aeq}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65</td>
<td>68</td>
<td>58</td>
<td>68</td>
</tr>
<tr>
<td>2</td>
<td>78</td>
<td>81</td>
<td>73</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>84</td>
<td>88</td>
<td>74</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>84</td>
<td>86</td>
<td>75</td>
<td>89</td>
</tr>
<tr>
<td>5</td>
<td>75</td>
<td>77</td>
<td>67</td>
<td>77</td>
</tr>
<tr>
<td>6</td>
<td>86</td>
<td>88</td>
<td>79</td>
<td>85</td>
</tr>
<tr>
<td>7</td>
<td>79</td>
<td>81</td>
<td>72</td>
<td>78</td>
</tr>
<tr>
<td>8</td>
<td>88</td>
<td>92</td>
<td>76</td>
<td>110</td>
</tr>
<tr>
<td>9</td>
<td>85</td>
<td>91</td>
<td>80</td>
<td>94</td>
</tr>
<tr>
<td>10</td>
<td>86</td>
<td>91</td>
<td>79</td>
<td>97</td>
</tr>
<tr>
<td>11</td>
<td>89</td>
<td>92</td>
<td>81</td>
<td>95</td>
</tr>
<tr>
<td>12</td>
<td>90</td>
<td>94</td>
<td>82</td>
<td>100</td>
</tr>
<tr>
<td>13</td>
<td>91</td>
<td>96</td>
<td>85</td>
<td>99</td>
</tr>
<tr>
<td>14</td>
<td>98</td>
<td>100</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>15</td>
<td>96</td>
<td>99</td>
<td>86</td>
<td>108</td>
</tr>
<tr>
<td>16</td>
<td>94</td>
<td>96</td>
<td>81</td>
<td>111</td>
</tr>
<tr>
<td>17</td>
<td>83</td>
<td>85</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td>18</td>
<td>81</td>
<td>85</td>
<td>74</td>
<td>88</td>
</tr>
<tr>
<td>19</td>
<td>84</td>
<td>87</td>
<td>72</td>
<td>102</td>
</tr>
<tr>
<td>20</td>
<td>85</td>
<td>90</td>
<td>71</td>
<td>117</td>
</tr>
<tr>
<td>21</td>
<td>75</td>
<td>81</td>
<td>69</td>
<td>87</td>
</tr>
<tr>
<td>22</td>
<td>76</td>
<td>79</td>
<td>68</td>
<td>82</td>
</tr>
</tbody>
</table>

3. Results and Discussion

3.1. Results of the Measured Noise Levels

The results of the measured noise levels are presented in Table 1. These are presented with locations of the PMS
### 3.2. Results from the Predictive Model

The results of the MLR model developed for predicting the noise levels of the study area are presented. The essence was to apply Landuse variables (POP, Traffic, Roadnet, Distance) as input in the MLR technique to predict the level of noise within the TMC. This regression analysis allows modelling, examining, and exploring of spatial relationships. This helps to better understand the factors behind observed spatial patterns, and to predict outcomes based on that understanding.

The MLR noise prediction model developed based on the AHP principle is presented in Equation 9. The resulting correlation coefficient between the variables, summary coefficients of the multiple linear regression and others from the SPSS statistical software are presented in Table 2. The model performance based on statistical indicators is presented in Table 3.

#### Table 2. Coefficients of the Regression Equation.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>68.088</td>
<td>5.473E-44</td>
<td>65.907</td>
</tr>
<tr>
<td>POP</td>
<td>112.602</td>
<td>1.38E-25</td>
<td>102.610</td>
</tr>
<tr>
<td>ROADNET</td>
<td>36.570</td>
<td>4.699E-21</td>
<td>32.324</td>
</tr>
<tr>
<td>LANDUSE</td>
<td>14.146</td>
<td>6.624E-08</td>
<td>9.763</td>
</tr>
<tr>
<td>DIST</td>
<td>45.503</td>
<td>5.22E-19</td>
<td>39.506</td>
</tr>
</tbody>
</table>

The predicted noise levels from the MLR model is presented in Figure 2. Figure 2 shows the trend of predicted and measured values of noise level during day-night in the study.

The MLR model developed was run in the MATLAB environment and the results were used to plot the spatial distribution of the noise levels of the area. The spatial distribution of the estimated noise pollution levels from the monitoring stations in the study area forecasted are presented in Figure 3 demonstrating that GIS could be used for noise mapping.

A standard model, Lyon’s Empirical model, was used to confirm the developed MLR model. The development was also used to map the TMC and the results are presented in Figure 4. The predicted noise levels were between 65.94 dB (A) being minimum level and 110.42 dB (A) being the maximum level.

#### Table 3. Model Summary.

<table>
<thead>
<tr>
<th>Mathematical Model</th>
<th>Performance Indicators</th>
<th>RME</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Maximum</th>
<th>Minimum</th>
<th>$R^2$</th>
<th>$R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR</td>
<td></td>
<td>1.569</td>
<td>2.462</td>
<td>1.585</td>
<td>3.718</td>
<td>-3.756</td>
<td>0.961</td>
<td>0.98</td>
</tr>
</tbody>
</table>

RME = root mean square error, $R^2$ = coefficient of determination, $R$ = correlation coefficient.
Figure 3. Noise Map for the Predicted Noise Levels.
Figure 4. Noise Map Created from Lyons’ Empirical Noise Prediction Mode.
3.3. Discussion

The equivalent noise levels measured in the study area are presented in Table 1. The Table 1 shows the daily average values of noise descriptors for the monitoring stations along the major Takoradi-Tarkwa road. The monitoring stations were designated with numbers 1 to 50.

From Table 1, location 36 recorded the highest values of $L_{Aeq}$ (98 dBA), location 37 is the next station that recorded level of $L_{Aeq}$ (97 dBA), location 1 recorded the lowest $L_{Aeq}$ of (65 dBA), and the station was followed by location 50. It was noted that monitoring locations with high noise descriptors were found in areas of high noise activities and vice versa. Among factors responsible for the differences in noise levels in the monitoring stations surveyed include location site, presence of intrusive noise, traffic volume, and commercial activities.

This further confirm the assertion made by [20] that noise pollution generated from ambient noise levels in a given area depend on a number of specific variables including road traffic characteristics; especially traffic volume, vehicle horn, rolling stock and tires. Some researchers have demonstrated in literature that the urban conditions of a given area are also important factor influencing the noise levels. There is a variation in the noise levels with the period of the day and the nature of the location.

Considering the regulations from Environmental Protection Agency (EPA) standards, in which equivalent sound level above 65 dBA is judged as high and could put the population at high risk, only 9 locations, representing 18%, of the 50 monitoring stations surveyed, can be classified as normally acceptable. The rest of the stations can be classified as clearly unacceptable. It is demonstrated in literature that living in areas of high noise pollution put that population in the area at risk of numerous health effects of noise pollution including, psychological, sleep and behavioural disorder as in [2] and [1]. Therefore, in comparing the values in Table 1, it could be observed that there is a danger (i.e., adverse health impact) in staying in such communities. These may results in fluctuations in the costs of house pricing in the area.

The design and implementation of noise prediction model using MLR technique yielded very good results. This will help forecast long-term variability of urban noise levels in the Tarkwa mining communities. The developed equation has brought to bear for the first time that MLR could be used to formulate noise prediction models. The results obtained, as shown in Figure 2, indicate that it is possible to develop a LUR model using the MLR technique with independent variables and that the results could be used for noise mapping as illustrated in Figure 3. The MLR produced results compared well with the study conducted by [13], which confirmed that the noise prediction model could be used in epidemiological studies.

It should be noted that Pearson correlation analysis at 0.05 significance level (two-tailed) was performed in a step wise manner on the input and output data. The essence was to select the most suitable input parameters applicable for developing the MLR models. The results revealed that the relationship existing between POP, Traffic, Roadnet, Landuse, Distance and measured noise level are statistically significant with $p \leq 0.05$. The interpretation is that the measured noise level provides enough convincing evidence to reject the null hypothesis and accept the alternative hypothesis that “the population correlation coefficient is different from zero” that is $\rho \neq 0$. This can be confirmed from Table 2. Therefore, in the MLR formulation, this study selected POP, Traffic, Roadnet, Landuse, and Distance as input parameters.

The performance indicators of this model, as observed in Table 3 showed RMSE, standard deviation, $R^2$ and $R$ values of 1.569, 1.585, 0.961 and 0.98 respectively. On the basis of quantitative evidence (Table 3), the obtained RMSE value signifies that the MLR model predictions are in consonance with the observed data. Hence, the RMSE result indicates a good statistical estimation measure of the residuals generated by the MLR model. In continuance of that, the standard deviation value shows the extent of precision of the predicted MLR results. The $R^2$ value here indicates the tolerability of the MLR prediction values. Thus, 96.1% changes in the measured noise levels are explained by the variation in the predicted outcomes from the MLR model. The $R$ findings, on the other hand, show the strength and direction of linear dependency existing between the observed noise levels and the predicted outputs from the MLR model. The inference made in line with the maximum and minimum values (Table 3) is that when the MLR model is applied within the study area to predict noise level a maximum error of 3.718 and minimum -3.756 could be achieved. Therefore, in this study, it is obvious that the MLR structure appears to perform satisfactorily. Figure 2 intuitively confirm this assertion where it is evident that the MLR predicted values are closely related to the observed noise levels. Therefore, it can be concluded that the statistical indicators presented in Table 3 are all within the acceptable limits.

The noise prediction model developed in this research is exclusively different from the current LUR models, since in this model a variety of independent variables are applied both in the formation of the MLR equations and application. It is also different from other basic prediction models by the consideration of land-use and other relevant variables. The results from this modelling process show that, with accurate data, noise prediction models are now promising tools, as demonstrated, for noise exposure assessment with potential applications in urban planning, epidemiological studies and environmental management, particularly in areas where noise predictions models or noise maps from competent authorities are not available.

4. Conclusion

A noise prediction model has been designed and implemented, the first time of its kind in the West African
sub-region, using MLR method. The statistical indicators used to verify the viability of the developed MLR model showed a RMSE of 1.569, standard deviation of 1.585, correlation coefficient (R) of 0.98 and R² of 0.961. These statistical findings have revealed that the developed LUR based on the MLR could be used for noise prediction of an area with an accuracy of 98%. This assertion is based on the R of 0.98 obtained in this study. The use of R as a model assessor is in line with the study of [6] and [5] who also used the same approach for the model adequacy. Furthermore, results obtained from the MLR model is in consonance with the standard noise model (Lyons Empirical model). This was noticed from the visual observation from the noise maps produced in Figures 3 and 4.

The developed MLR model had also demonstrated its use in mapping intrarural noise in relation to urban land-use as changes occur. This is useful for urban planning and environmental noise management. It could also be applied predict intrarural noise changes with time and epidemiological studies as well as decision-making tool. It was observed from the study that the more the monitoring stations, the better the model performance. Therefore, the model performs better for large scale area like the study area and vice versa.

References


