



## Article

# Traffic Noise Modelling Using Land Use Regression Model Based on Machine Learning, Statistical Regression and GIS

Ahmed Abdulkareem Ahmed Adulaimi <sup>1</sup>, Biswajeet Pradhan <sup>1,2,\*</sup> , Subrata Chakraborty <sup>1</sup>   
and Abdullah Alamri <sup>3</sup>

<sup>1</sup> Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia; ahmed.a.adulaimi@student.uts.edu.au (A.A.A.); Subrata.Chakraborty@uts.edu.au (S.C.)

<sup>2</sup> Earth Observation Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, Bangi 43600 UKM, Selangor, Malaysia

<sup>3</sup> Department of Geology and Geophysics, College of Science, King Saud University, Riyadh 11451, Saudi Arabia; amsamri@ksu.edu.sa

\* Correspondence: biswajeet24@gmail.com or Biswajeet.Pradhan@uts.edu.au

**Abstract:** This study estimates the equivalent continuous sound pressure level ( $L_{eq}$ ) during peak daily periods ('rush hour') along the New Klang Valley Expressway (NKVE) in Shah Alam, Malaysia, using a land use regression (LUR) model based on machine learning, statistical regression, and geographical information systems (GIS). The research utilises two types of soft computing methods including machine learning (i.e., decision tree, random forest algorithms) and statistical regression (i.e., linear regression, support vector regression algorithms) to determine the best approach to create a prediction  $L_{eq}$  map at the NKVE in Shah Alam, Malaysia. The selection of the best algorithm is accomplished by considering correlation, correlation coefficient, mean-absolute-error, mean-square-error, root-mean-square-error, and mean absolute percentage error. Traffic noise level was monitored using three sound level meters (TES 52A), and a traffic tally was done to analyse the traffic flow. Wind speed was gauged using a wind speed meter. The study relied on a variety of noise predictors including wind speed, digital elevation model, land use type (specifically, if it was residential, industrial, or natural reserve), residential density, road type (expressway, primary, and secondary) and traffic noise average ( $L_{eq}$ ). The above parameters were fed as inputs into the LUR model. Additional noise influencing factors such as traffic lights, intersections, road toll gates, gas stations, and public transportation infrastructures (bus stop and bus line) are also considered in this study. The models utilised parameters derived from LiDAR (Light Detection and Ranging) data, and various GIS (Geographical Information Systems) layers were extracted to produce the prediction maps. The results highlighted the superior performances by the machine learning (random forest) models compared to the statistical regression-based models.

**Keywords:** traffic noise modelling; land use regression model; machine learning; GIS; LiDAR



**Citation:** Adulaimi, A.A.A.; Pradhan, B.; Chakraborty, S.; Alamri, A. Traffic Noise Modelling Using Land Use Regression Model Based on Machine Learning, Statistical Regression and GIS. *Energies* **2021**, *14*, 5095. <https://doi.org/10.3390/en14165095>

Academic Editor: Eduard Petlenkov

Received: 8 June 2021

Accepted: 13 August 2021

Published: 18 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Urban population is continuously exposed to traffic noise [1]. Road traffic is the most impacting noise source affecting human modern lifestyle [2]. The impacts of such noise on human health are well documented [3–6]. Adverse effects include cardiovascular health, insomnia and other sleep-related disorders, speech disorders, and psychological and physiological challenges [7–11]. The World Health Organization (WHO) published a study based on the degree of exposure to traffic noise experienced by population in European cities. Results show that 50% of people suffered exposure to traffic noise of greater than 55 dB [12]. Studies have also identified a strong correlation between traffic noise and the urban population [5,6].

The key factors affecting the traffic noise emission include elements such as type of tyres [13], engine types, and flow composition [14]. The acoustic impedance has been identified as another key factor in traffic noise emission [15,16]. Other than tyres, the acoustic performance depends on various pavement characteristics including pavement texture [17] and pavement age [18]. Understanding such noise characteristics helps in devising mitigating strategies, including pavement noise reduction with use of new technologies in design and materials [19], such as rubberized asphalt [20,21]. Understanding the noise factors helps in analysing the noise profile of a specific location along with the location characteristics.

The traffic noise variation is different based on the specific location characteristics. The common noise sources that are identified include road networks, rail networks, airports, construction sites, industrial zones, and human and social sources [22]. Having adjacent rail networks have high impact on the noise level of specific road networks and needs to be considered in noise modelling studies [23,24]. Road noise levels could be influenced by nearby airports and overhead air corridors as well [25,26]. Similarly, nearby port activities and ship movements may induce additional noise to road networks [27,28]. Hence, it is imperative these factors are considered in noise mapping studies of any area. However, the availability of traffic noise maps depends on several factors, including the size of the area, the type of input data, and the legal context [22,29]. Therefore, the key purpose of traffic noise map is identifying the noisy areas. Accordingly, developing traffic noise modelling for a specific area involves use of dedicated software, computing resources, and expert knowledge in order to generate reliable digital noise maps [30].

The application of land use regression (LUR) models is common in epidemiological research to examine air pollution exposure levels and in urban development research to examine levels of exposure to traffic noise [31–36]. LUR employs linear regression modelling techniques to create predictive data regarding noise levels or air pollution levels in a specific area by using predictor variables collected and analysed primarily through ArcGIS Software. Previous studies used LUR models for predicting the traffic noise. These models are challenging to generalize because of varying local conditions, such as the type of road networks in the area, variance in vehicle specifications, the type of land use area, meteorological conditions, and the size of the study area (e.g., city-wide/small scale) [35].

This study is based on the hypothesis that the LUR model can be used in traffic noise modelling with limited to field data collection. Thus, the main objective of this research is to identify the significance of key factors impacting traffic noise so that they can be used suitably in noise modelling. Moreover, the paper aims to perform a comparative analysis of several soft computing techniques so that the best one can be adopted with LUR model. Finally, it aims to demonstrate the appropriateness of LUR model for effective noise mapping where very limited field data is available.

This study aims to provide an objective summary of the key variables which exert a significant influence on how levels of traffic noise are defined as well as to analyse various soft computing techniques, commonly employed including decision trees (DT) [37], random forests (RF) [38], linear regression (LR) [39], and support vector regression (SVR) [40]. Based on the performance analysis of the chosen models, the most appropriate one is then selected to generate the predictive equivalent continuous sound pressure level (Leq) map for the designated area during peak traffic times. Model performances will be evaluated based on correlation ( $R$ ), correlation coefficient ( $R^2$ ), mean-absolute-error (MAE), mean-square-error (MSE), root-mean-square-error (RMSE), and mean absolute percentage error (MAPE). Finally, this paper will show which method is suitable for land use regression model to use for generating noise map for the study area by employing most efficient methods as well as the statistics indicators used to evaluate the models.

In the Section 2, a brief background of the literature is presented, followed by the explanation of the model development in Section 3. Next, results and discussion, including the main findings, are presented in Section 4. Finally, a short conclusion is presented in Section 5.



## 2. Related Work

Several traffic noise prediction models for cities have been proposed by previous studies based on the land use regression model (LUR). The LUR model based on the linear regression has been previously used for assessing and predicting such as traffic noise, air pollution, health, epidemiological studies, and others. Furthermore, LUR modelling can be scaled depending on the size of the city being examined; it has a high degree of accuracy and the capacity to manage complex variables and is less computationally expensive. The LUR model has been used successfully in North America [36], Africa [34], Asia [22], and Europe [29]. A study conducted by Aguilera et al. (2015) applied the LUR model to examine traffic noise in three different cities in Europe [29]. The data were recorded in a 20 min non-peak traffic period, and the input variables for the LUR model included roads, land use (industrial, residential), agricultural, forest, semi-natural, and population. The study developed LUR models based on linear regression and following the ESCAPE project in a large number of European study areas. Their study suggested that LUR modelling with accurate GIS source data can be a promising tool for noise mapping in epidemiological study [29]. Ragettli et al. [36] used LUR modelling with long-term noise measurements and land use characteristics to examine ambient levels of noise in Montreal, Canada. The study developed LUR models based on various transportation noise sources such as air, rail, and road, in order to predict the equivalent continuous sound pressure levels  $L_{eq24h}$ ,  $L_{night}$ , and  $L_{den}$  which was improved upon previous research conducted in the same area. In another study conducted in the Western Cape in South Africa, Sieber et al. [34] employed LUR modelling to assess the outdoor noise variability for adults living in informal settlements, which involved constant monitoring of outside noise levels during an entire week, and recorded data related to 134 homes in four different areas. The LUR model developed for the study considered noise sources such as transportation networks (air, rail, and road), local buildings, land use, and the community, in order to derive the daytime, evening and night-time values for the equivalent sound level. More recently Harouvi et al. (2018) utilized the LUR model with high-resolution transportation to estimate the noise in two periods of the day (rush hour and off-peak) at two cities in Israel, and it was discovered that using LUR supported by GIS approach provided good performance for estimation and mapping of noise pollution for environmental noise assessment [22].

In summary, as it can be seen the above, the previous research has employed the LUR model based on the linear regression with GIS through several variable descriptors for predicting the noise in each city area. The study by Harouvi et al. [22] used the existing predictor variables, whereas other studies utilized the potential predictor variables collected through GIS [22,34,36]. Notably, Harouvi et al. [22] further extended the analysis to include new variables derived from the volume of traffic and its position with respect to the city centre. In a somewhat different approach, Aguilera et al. [29] and Sieber et al. [25] used additional variables such as agricultural, forest, semi-natural, as well as vegetation as a normalized difference vegetation index (NDVI). Ragettli et al. [27] used two more variables, those of proximity of buffer areas to rivers, and the degree of density (low, high) of residential areas. The research discussed above has verified the successful application of the LUR model in cities in developed countries, as well as in undeveloped areas with unplanned settlements (informal settlements). Furthermore, the use of GIS software was found to be enhancing the predictor or independent variables used as input data for the LUR model.

The previous LUR models are based on linear regression. Thus, this study will attempt to develop LUR models based on machine learning (i.e., decision tree, random forest algorithms) and statistical regression (i.e., linear regression, support vector regression algorithms) using Python software with GIS for predicting traffic noise in a key expressway in Malaysia. In addition, the novelty of this study also lies in the use of additional predictor variables.

### 3. Methods

#### 3.1. Study Area

The study area is situated in Shah Alam, Malaysia. Shah Alam is home to the New Klang Valley Expressway (NKVE), a 35 km road linking Kuala Lumpur (Jalan Duta) with the commercial and rural areas of New Klang (Bukit Raja). The study area was divided a  $5 \times 5$  m grid cells with a centroid for each cell. The reason to choose this site because it contains a mix of various land usage including low/high density residential, industrial, and commercial areas and includes different types of road networks such as expressway, primary, and secondary roads. These diversities make the site suitable for studying the traffic noise for various conditions. Furthermore, noise maps of the area would help understand how the new expressway may impact residents who are concerned about the potential traffic impacts.

#### 3.2. Noise Data and Predictor Variables

The noise levels of traffic flow on Shah Alam roads were evaluated with three equipment of sound level meter (TES 52A). Additional data were collected from a wind speed meter and a traffic tally for reading wind speed and number of vehicles, respectively. The traffic noise measurements were generated randomly for various sites by using the ArcGIS sampling design tools across the city area [41,42]. On 11–12 February 2017, field data were collected at every 20 min at a height of 1.5 m during rush hours (6:30–8:00; 10:00–12:00; 18:00–20:00; 23:00–00:00). A total of 95 measurements were taken which were divided into 67 for training and 28 for testing as shown in Figure 1.

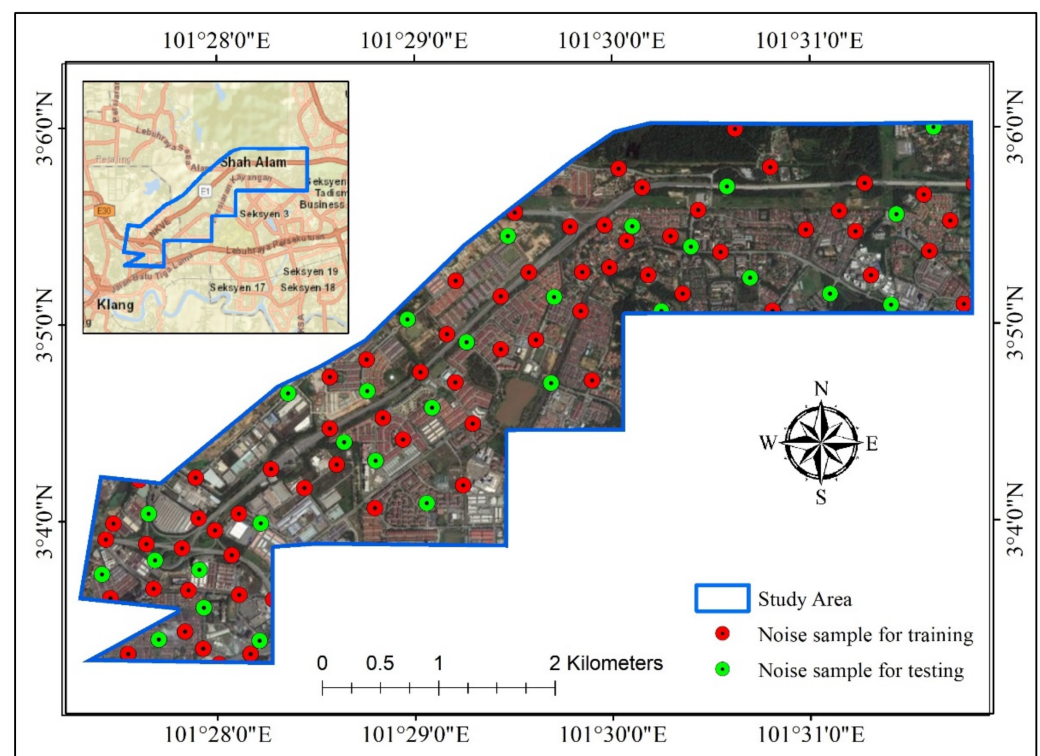


Figure 1. Shows the study area.

The LUR model used a wide selection of spatial predictors which might increase traffic noise [22,33,36]. The data used by the LUR model were to determine the traffic noise; these data are: area of residential low/high density, type of network road (expressway, primary, and secondary), land use (residential, industrial, and tree), digital surface model (DSM), wind speed (WS), and traffic noise average ( $L_{eq}$ ). Additional information on traffic jams, road intersections, traffic lights, road toll gate, public transport (bus stop and bus line),

and gas stations was also used. These details are summarised in Tables 1 and 2. As well as the predictor variables, the Figures below show the DSM raster (Figure 2); the area of residential high density, area of residential low density, industrial area, and trees area (Figure 3); the type of road network such as expressway, primary and secondary roads (Figure 4); the population raster (Figure 5); the road toll gate, gas station, traffic light, intersect, bus stop, and bus line (Figure 6); and wind speed (Figure 7). The raw sample from 1 data collection point at 4 different times of the day is shown in Figure 8.

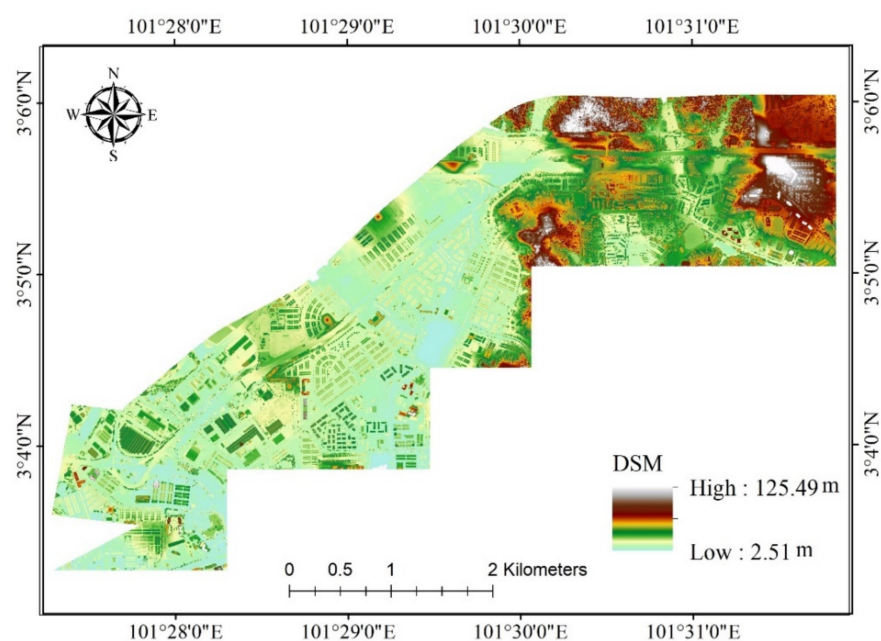


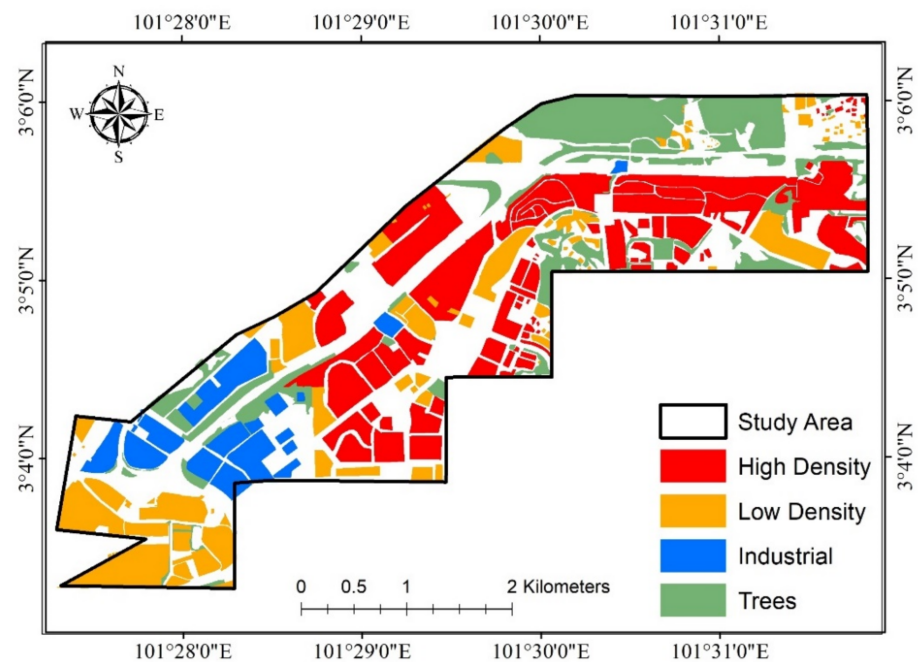
Figure 2. Digital surface raster (DSM) layer.

Table 1. Summary statistics of noise predictors.

Parameter (Noise Predictors)	Unit	Mean	Minimum	Maximum	Std. Dev.
Traffic volume (per 15 min)	Veh/hour	122	9	810	183.66
Distance from all type of roads	Meters	67.37	$2.11 \times 10^{-5}$	465.12	65.31
Distance from expressway	Meters	426.66	$1.74 \times 10^{-4}$	1638.76	334.54
Distance from primary road	Meters	468.26	$1.83 \times 10^{-3}$	1732.76	366.96
Distance from secondary road	Meters	97.90	$2.11 \times 10^{-5}$	483.40	88.58
Distance from area of residential high density	Meters	402.66	0	2826.72	615.76
Distance from area of residential low density	Meters	190.79	0	855.10	175.38
Distance from residential Area	Meters	94.60	0	855.10	159.93
Distance from industrial Area	Meters	705.13	0	2470.92	568.72
Distance from trees Area	Meters	157.48	0	947.91	168.87
DSM	Meters	19.25	2.51	125.49	16.27
WS	km/h	16.62	15.8	17.58	0.53
Distance from gas station	Meters	1183.00	0	2726.06	651.41
Distance from traffic lights	Meters	780.87	0.58	1975.83	432.05
Distance from intersect	Meters	203.25	0.14	909.12	159.71
Distance from road toll gate	Meters	1010.38	0	2239.69	513.18
Distance from bus stop	Meters	528.22	0.083	1736.24	334.74
Distance from bus line	Meters	214.94	$1.86 \times 10^{-4}$	946.33	161.00

**Table 2.** Results of assessing the contribution of noise predictors using the chi-square method.

Noise Predictor	Multiple R-Square	VIF
Traffic volume	0.64	2.78
All type of roads	0.25	1.33
Expressway	0.82	5.64
Primary road	0.89	9.15
Secondary road	0.37	1.59
Area of residential high density	0.82	5.49
Area of residential low density	0.63	2.71
Residential area	0.71	3.45
Industrial area	0.70	3.28
Trees area	0.49	1.96
DSM	0.53	2.14
WS	0.74	3.87
Gas station	0.75	4.06
Traffic lights	0.71	3.47
Intersect	0.61	2.56
Tool road	0.65	2.82
Bus stop	0.89	8.96
Bus line	0.72	3.52

**Figure 3.** Overall landcover of the area.

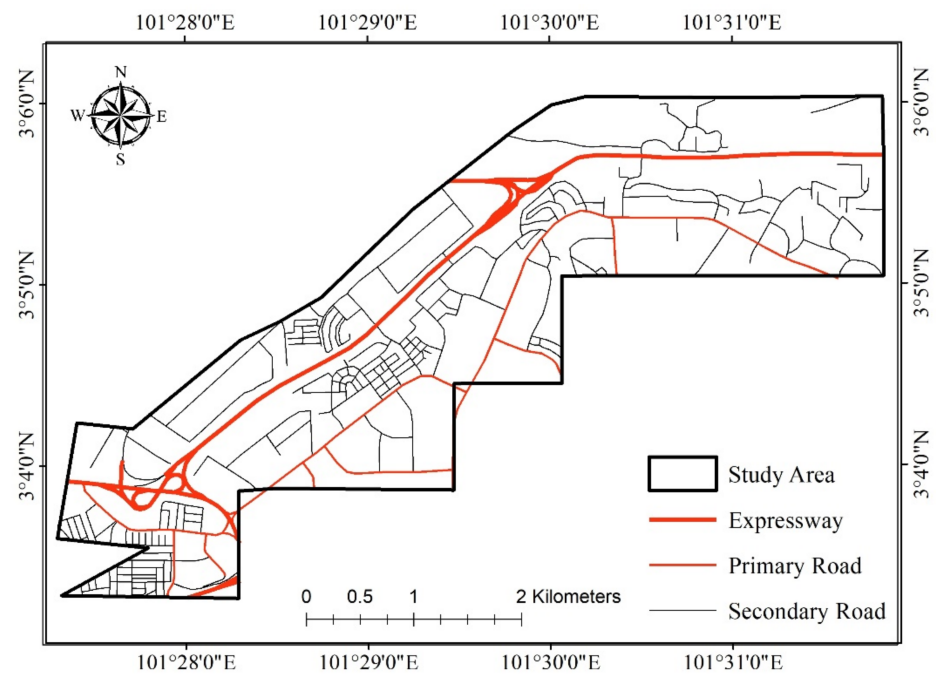


Figure 4. Type of road network such as expressway, primary, and secondary roads.

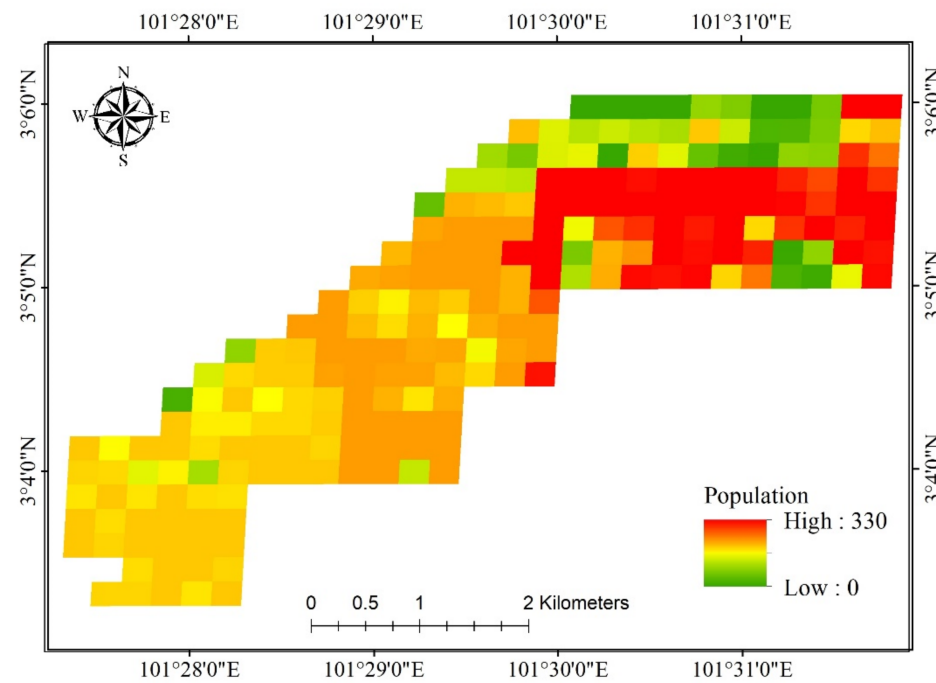


Figure 5. Population layer.



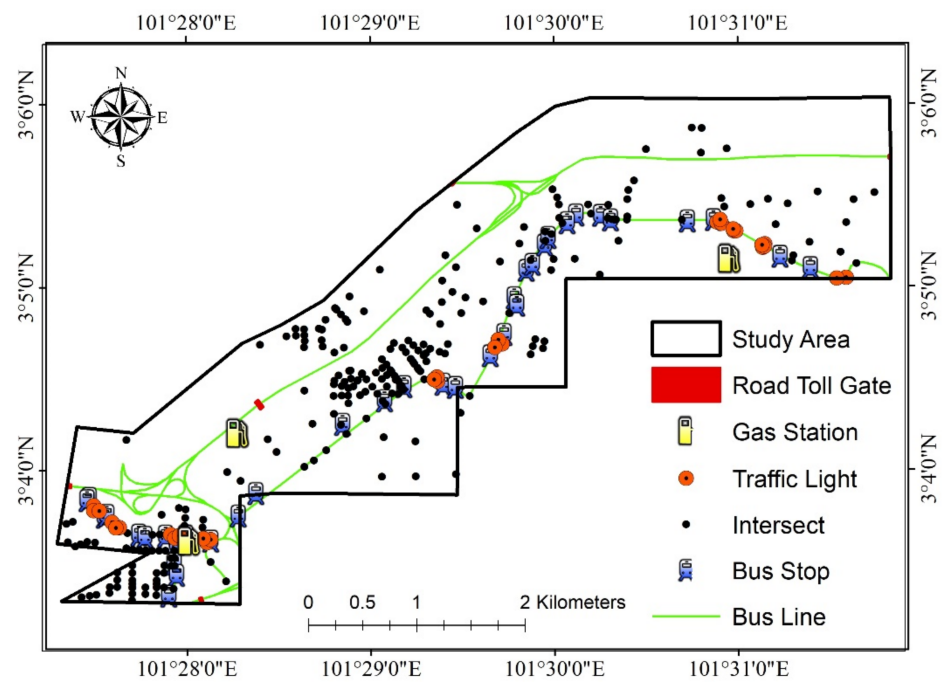


Figure 6. Road toll gate, gas station, traffic light, intersect, bus stop, and bus line.

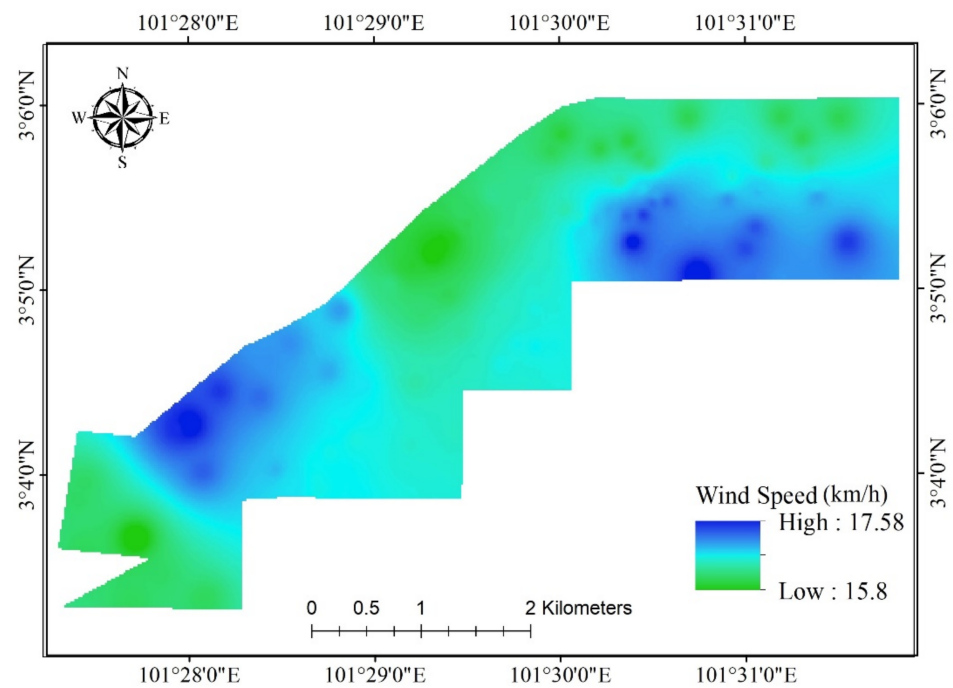
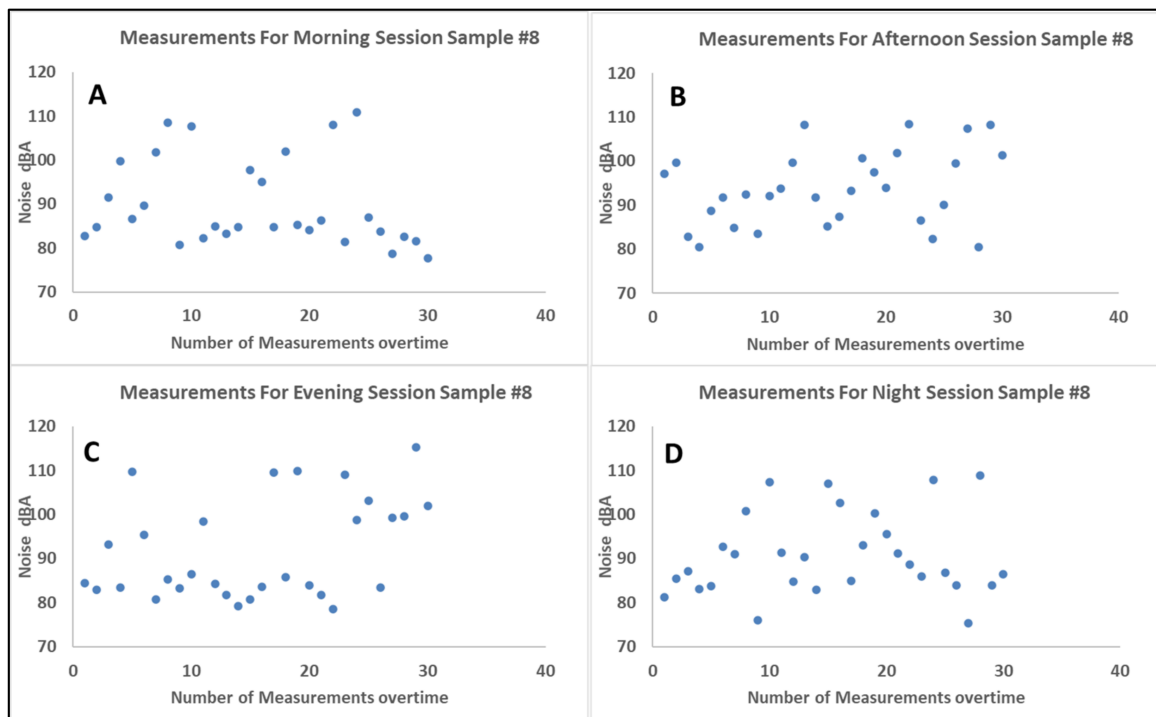


Figure 7. Wind speed layer.



**Figure 8.** Raw noise sample data taken during: (A) morning, (B) afternoon, (C) evening, and (D) night.

### 3.3. Data Pre-Processing

The correlation and variance inflation factor (VIF) [43,44] of the model's parameters were calculated using Equation (1), following which the researchers programmed the model using Python:

$$VIF = \frac{1}{1 - R^2} \quad (1)$$

where  $R^2$  = multiple correlation coefficients between a predictor of noise and predictors remaining.

In addition, the research also made use of the correlation-based feature selection (CFS) algorithm. This filter algorithm is a machine-learning algorithm that selects attributes according to existing concepts of correlation [45,46]. One of the advantages of the CFS algorithm is its ability to identify sub-groups within attributes that are unrelated within the group but do correlate with the targeted class of a study.

However, it is worth mentioning that the CFS algorithm omits attributes of a sub-group with a low correlation rate to the target class [45,46]. Accordingly, it is used to get rid of replicated attributes, assuming they will correlate in a different manner elsewhere. Aspects are therefore identified when the CFS algorithm predicts a group existing in an area that otherwise remains unidentified.

Equation (2) shows the feature sub-set assessment function of the CFS algorithm [46]:

$$Ms = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k - 1)\bar{r}_{ff}}} \quad (2)$$

$Ms$  relates to the heuristic 'merit' of a feature sub-set  $s$  containing  $k$  features;  
 $\bar{r}_{cf}$  is the mean of the feature-class correlation ( $f \ 2 \ s$ );  
 $\bar{r}_{ff}$  is the average of the feature–feature intercorrelation.

### 3.4. Land Use Regression Models

In this study, four different models were used as land use regression model such as RF, DT, LR, and SVR algorithms for the prediction of  $L_{eq}$ . A land use regression model (LUR model) is an algorithm often used for analysing pollution at any location, depending on the environmental characteristics of the surrounding [22,36]. Firstly, describe the methodology through these four soft computing methods to find the best model for predicting noise map of  $L_{eq}$  for Shah Alam, Malaysia. The overall methodology is shown in Figure 9. In December 2016, light detection and ranging (LiDAR) data were captured, along with the Worldview-3 satellite images. The data such as noise samples, traffic flow, and wind speed data were collected from the field during 11–12 February 2017. In addition, land use map was created using Worldview-3 image. The images were segmented using the multi-resolution segmentation algorithm. The parameters of the segmentation algorithm were selected through trial-and-error method. After segmentation, several features were extracted from Worldview-3 image. These features include spectral bands of the Worldview-3 image, textual features, and spatial features, such as shape index, rectangular fit, and length/width. These objects were then classified into three classes, including buildings, road network, and trees, using the support vector machine method (SVM).

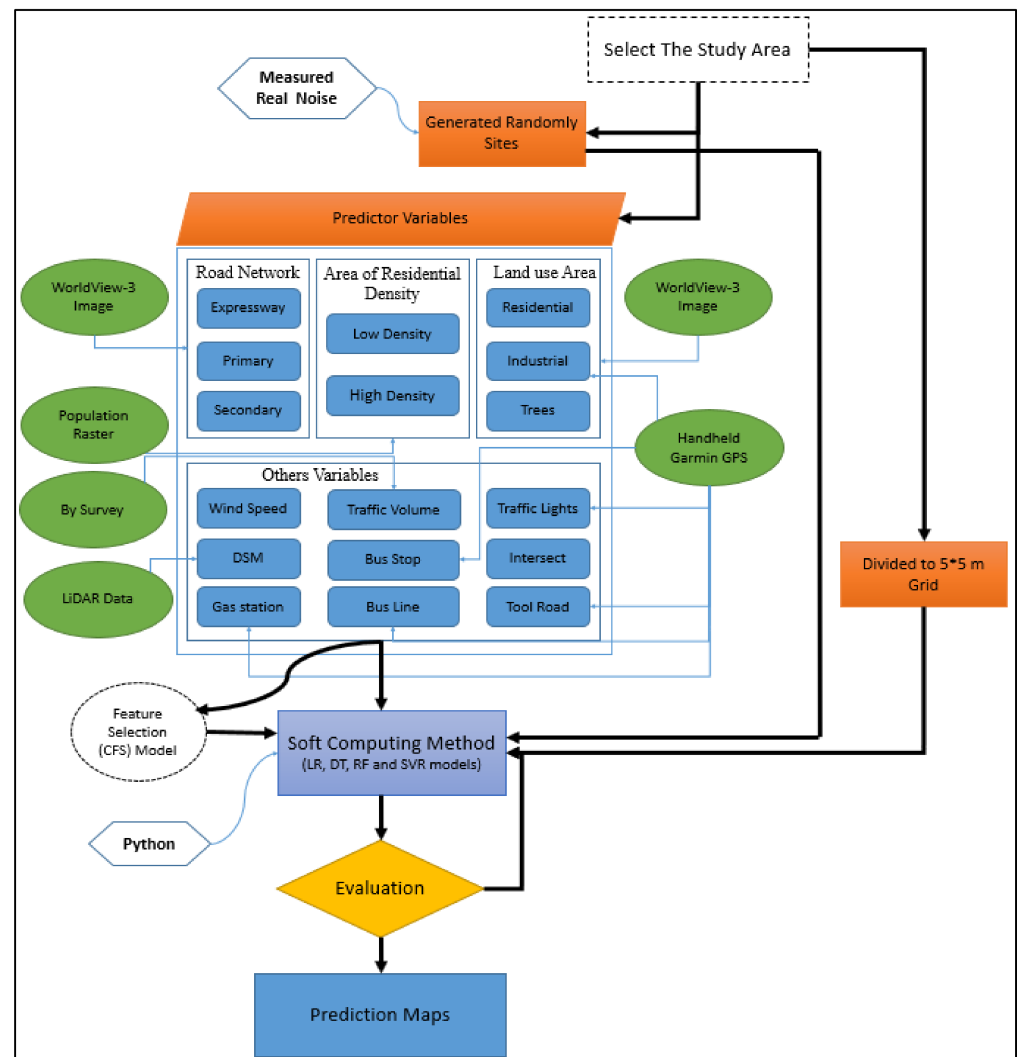


Figure 9. The overall methodology used in this study.

This algorithm requires several parameters, such as the kernel function and the penalty parameter for optimization. These parameters were then selected by grid search over

specific search domain. The analysis showed that the best combinations of segmentation parameters are as follows: scale = 25, shape = 0.3, and compactness = 0.8. The best SVM parameters were found to be rbf and  $c = 100$ . In addition, the population raster was utilized to extract the area of residential high and low density by using the spatial join tool (building with population raster) in GIS.

Next, the four different LUR model algorithms (DT, RF, LR, and SVR) were applied using Python software. Their effectiveness at predicting  $L_{eq}$  was decided by monitoring  $R$ ,  $R^2$ ,  $MAE$ ,  $MSE$ ,  $RMSE$ , and  $MAPE$ . And, on the other hand, combining models such as the correlation-based feature selection (CFS) with (DT, RF, LR, or SVR) models and also evaluating and validating to get the best model. Finally, the best model was identified through the validation of training and testing data to produce  $L_{eq}$  map by using GIS techniques for the study area.

### 3.5. Model Evaluation

The performance of the four models was ascertained by calculating six performance measures:  $R$ ,  $R^2$ ,  $MAE$ ,  $MSE$ ,  $RMSE$ , and  $MAPE$ , in the knowledge that this would give estimates of  $L_{eq}$ . These performance measures indicate the accuracy of model's predictions by comparing the actual parameter's value ( $a_i$ ), predicted value ( $b_i$ ), and number of sample data points ( $n$ ) and others such as an average of all observed values ( $\bar{a}$ ) and average of all predicted values ( $\bar{b}$ ) which could be useful when comparing different models. In each case, the independent variables entered into the four models were DSM, WS, type of land use (industrial, residential, and tree), type of road (expressway, primary, and secondary), and density of housing in area, plus the additional details on road intersections, road toll gate, traffic lights, gas stations, and public transportation (bus stop and bus line). The dependent variable was  $L_{eq}$ . (3) describes the correlation between the two data sets used to calculate linear relationship [10]. The value of  $R$  lies between  $-1$  and  $+1$ . Further, (4) was used to calculate the coefficient of determination, and the result value of  $R^2$  lies between  $-1$  and  $+1$  [10]. The  $MAE$  was calculated using (5). Determining  $MAE$  allowed the researchers to note the relationship (or difference) existing between two continuous variables. Equation (6) was used to calculate the  $MSE$ , which is the average squared difference between the estimated values [47–50]. Equation (7) is used to calculate the  $RMSE$  to evaluate the average performance of the model across different testing samples. Finally, Equation (8) is applied to calculate and find the  $MAPE$  which was shown the percentage and how accurate a forecasting method for prediction statistics [51].

$$R = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2 \sum_{i=1}^n (b_i - \bar{b})^2}} \quad (3)$$

$$R^2 = (R)^2 \quad (4)$$

$$MAE = \frac{\sum_{i=1}^n |b_i - a_i|}{n} \quad (5)$$

$$MSE = \frac{\sum_{i=1}^n (|b_i - a_i|)^2}{n} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (|b_i - a_i|)^2}{n}} \quad (7)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{a_i - b_i}{a_i} \right| \quad (8)$$

In this study, a 10-fold cross validation was used in order to examine the best model prediction for the testing data. This dataset was divided into 100 subsets, with 80 being used for training and the remaining 20 acting as the testing subsets. This process would then be

repeated with a different 20 subsets selected for testing and the original 20 reintegrated with the data set for training purposes. This process would be repeated 10 times, thus meaning that all the sub-sets of data were both tested and used for training.

## 4. Results and Discussion

### 4.1. Contribution of Noise Predictors

One of the findings of the study was that the traffic noise predictors actually contribute to the noise values gathered. As part of our statistical analysis, the chi-square method was used. Results revealed that the multi-collinearity of primary road and bus stop parameters amounted to 9.15 and 8.96, which was then used in the methodology because these noise predictors (primary road and bus stop) were important and crucial for predicting noise map.

The findings also showed how the traffic volume, road types (expressway, primary, and secondary), public transport, land use (industrial, residential, and tree), DSM, and WS all had a significant impact on prediction of noise levels. Table 2 describes this further.

### 4.2. Noise Prediction

The study applied four models, two each of machine learning and statistical regression. The number of data subsets totalled 95. This was divided into 67 for training and 28 for testing. Table 2 shows the 18 parameters used to predict the traffic noise ( $L_{eq}$ ). The first model is the LR algorithm.

The LR model fit for Shah Alam was calculated as given in Equation (9):

$$\begin{aligned} & (0.0204) \times \text{Traffic volume (per 15 min)} - (0.7139) \times \text{All type of roads} - (0.0085) \\ & \times \text{Expressway} - (0.0054) \times \text{Primary road} + (0.0067) \times \text{Secondary road} - \\ & (0.0005) \times \text{Area of residential high density} + (0.006) \times \text{Area of residential} \\ & \text{low density} - (0.0135) \times \text{Residential Area} - (0.0023) \times \text{Industrial Area} \quad (9) \\ & - (0.0041) \times \text{Trees Area} + (0.0008) \times \text{DSM} + (4.8595) \times \text{Wind speed} + (0.0007) \times \\ & \text{Gas station} - (0.0035) \times \text{Traffic lights} + (0.0106) \times \text{Intersect} - (0.0028) \times \text{Tool} \\ & \text{road} + (0.0108) \times \text{Bus stop} - (0.0164) \times \text{Bus line} + 45.5861 \end{aligned}$$

The second model was trained with the same parameters based on DT with the hyper-parameters such as min split = 20, max depth = 30, and min Bucket = 7. The same hyper-parameters were applied when using eleven parameters for training and testing.

The third model is the RF method for training and testing. In this case, the hyperparameters were: n estimators = 150; m try = 500; min split = 20; max depth = 30.

Finally, the fourth model, also known as the SVR model was trained and tested with the hyperparameters: kernel = rbf; gamma = sigmoid; tol = 3; decision function shape = ovo.

The SVR model fit for Shah Alam is given in Equation (10):

$$\begin{aligned} & (0.519) \times \text{Traffic volume (per 15 min)} - (0.0805) \times \text{All type of roads} - (0.1612) \\ & \times \text{Expressway} - (0.0571) \times \text{Primary road} + (0.0256) \times \text{Secondary road} \\ & - (0.0763) \times \text{Area of residential high density} + (0.1465) \times \text{Area of residential} \\ & \text{low density} - (0.2626) \times \text{Residential Area} - (0.1618) \times \text{Industrial Area} \quad (10) \\ & - (0.1304) \times \text{Trees Area} - (0.0713) \times \text{DSM} + (0.1042) \times \text{Wind speed} \\ & + (0.0738) \times \text{Gas station} - (0.2101) \times \text{Traffic lights} + (0.2613) \times \text{Intersect} \\ & - (0.1905) \times \text{Tool road} + (0.2157) \times \text{Bus stop} - (0.1546) \times \text{Bus line} + 0.6028 \end{aligned}$$

The findings for all four models showed that the difference between the predicted values was low (meaning the prediction errors for all models were small). From this, one could conclude that the RF algorithm was the most effective and successful. As shown in Table 3, the  $R$  value of RF model highest in training (0.95) and testing (0.93). Furthermore, for RMSE, training was 4.18 and testing 5.22. For MAE, training was 3.30 and 4.46. Table 4 describes this in more detail for all models.



**Table 3.** Results of predictions with LR, DT, RF, and SVM models with all parameters.

Method \ Evaluation	LR		DT		RF		SVM	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
<i>R</i>	0.93	0.91	0.92	0.91	0.95	0.93	0.94	0.92
<i>R</i> <sup>2</sup>	0.864	0.83	0.84	0.82	0.90	0.866	0.88	0.85
<i>MAE</i>	3.88	4.49	4.04	4.62	3.30	4.46	3.81	4.80
<i>MSE</i>	23.12	32.73	26.82	29.13	17.48	27.26	20.93	28.76
<i>RMSE</i>	4.81	5.72	5.18	5.40	4.18	5.22	4.58	5.36
<i>MAPE</i>	5.72	7.26	6.15	7.11	4.86	7.06	5.55	7.43

**Table 4.** Results of predictions with LR, DT, RF, and SVM models with eleven parameters.

Method \ Evaluation	LR		DT		RF		SVM	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
<i>R</i>	0.94	0.94	0.95	0.94	0.96	0.95	0.94	0.94
<i>R</i> <sup>2</sup>	0.88	0.88	0.90	0.88	0.92	0.90	0.89	0.88
<i>MAE</i>	3.66	4.46	3.36	4.26	2.99	3.86	3.64	4.30
<i>MSE</i>	20.25	22.91	17.99	23.35	13.99	19.96	19.10	23.95
<i>RMSE</i>	4.50	4.79	4.24	4.83	3.47	4.47	4.37	4.89
<i>MAPE</i>	5.45	6.28	5.05	6.42	4.37	5.94	5.35	6.65

The same models combined with the correlation-based feature selection model (CFS) were tested with the eleven parameters that had been identified by this model. In order to make the prediction process the best it could be, the CFS algorithm was used to identify the highly correlated parameters and to show which of the parameters were most effective when used to predict traffic noise for the study area.

From the CFS, it was found that the eleven parameters were the most effective for predicting traffic noise model, which was then trained and tested by the above four models. In the case of all four models, the hyperparameters were maintained the same.

However, the new equation regression of both the LR and SVR models (both using eleven parameters) registered a difference in the instances where eighteen parameters was used for the Shah Alam area.

The LR model for Shah Alam (trained and tested using eleven parameters) recorded the following results (Equation (11)):

$$\begin{aligned}
 &(0.0229) \times \text{Traffic volume (per 15 min)} - (8.5402) \times \text{All type of roads} - \\
 &(0.0092) \times \text{Expressway} + (0.0026) \times \text{Primary road} + (0.0013) \times \text{Area of} \\
 &\text{residential high density} + (0.0011) \times \text{Industrial Area} - (0.0027) \times \text{Trees} \\
 &\text{Area} - 0.0037 \times \text{DSM} + (3.9434) \times \text{Wind speed} + (0.0007) \times \text{Bus stop} \\
 &\quad - (0.0164) \times \text{Bus line} + 46.4032
 \end{aligned} \tag{11}$$

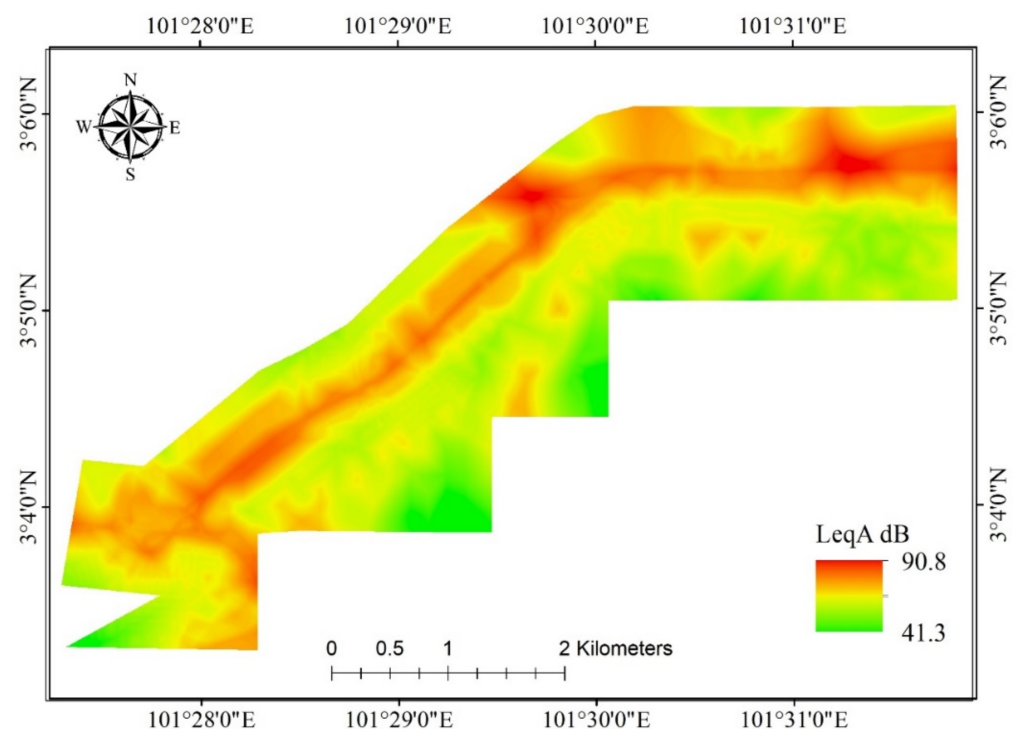
In turn, the SVR model (trained and tested using eleven parameters relevant to Shah Alam) recorded the following results (Equation (12)):

$$\begin{aligned}
 &(0.4495) \times \text{Traffic volume (per 15 min)} - (0.1553) \times \text{All type of roads} - \\
 &(0.3021) \times \text{Expressway} + (0.0955) \times \text{Primary road} - (0.0021) \times \text{Area of} \\
 &\text{residential high density} + (0.1113) \times \text{Industrial Area} - (0.11) \times \text{Trees} \\
 &\text{Area} - (0.1202) \times \text{DSM} + (0.1767) \times \text{Wind speed} - (0.0137) \times \text{Bus stop} \\
 &\quad - (0.2569) \times \text{Bus line} + 0.45
 \end{aligned} \tag{12}$$

Table 4 shows the results when filtered and reduced from eighteen parameters to eleven parameters based on the CFS method, which finds features that have a higher

correlation with the class but are uncorrelated with each other. Therefore, the highest correlated parameters were used for the prediction analysis, which resulted in improving the prediction accuracy. This is described in more detail in Table 4.

In this case, the results of all four models recorded improvements in the prediction accuracy and a decrease in the *MSE* and *RMSE* values, respectively. In addition, according to the results, it appeared the RF model was the most successful of the four models, even when the parameters were reduced from eighteen to eleven. By evaluating it in comparison to the other three models, the RF model repeatedly seemed to offer the most accurate and effective way of predicting noise for the Shah Alam area under consideration. Figure 10 shows the traffic noise prediction map for Shah Alam created using the RF model with 11 parameters.



**Figure 10.** The noise prediction map ( $L_{eq}$ ) on Shah Alam based on RF model with 11 parameters.

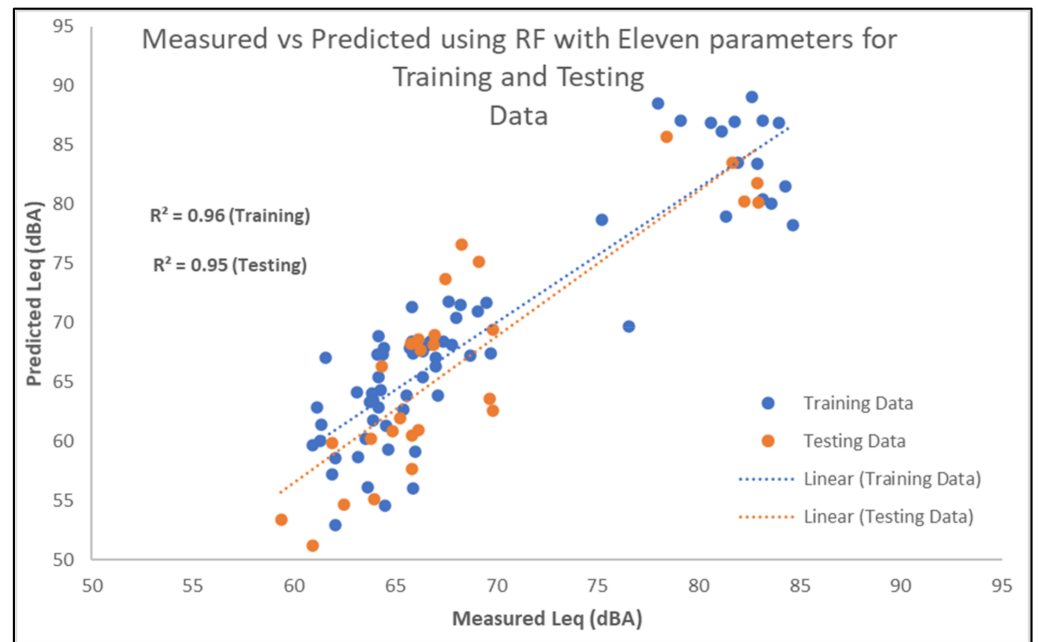
#### 4.3. Validation of Noise Prediction Maps

The validation of the machine learning and statistical regression models was performed by using the criteria of  $R$ ,  $R^2$ , *MAE*, *MSE*, *RMSE*, and *MAPE* methods. Tables 3 and 4 show the results for each model against six criteria.

The RF model still recorded better results even when the number of parameters were reduced from eighteen to eleven, which can be seen when considering, for example, the comparative *RMSE* values. When the parameters were set at eighteen, the *RMSE* for training was 4.18, while the *RMSE* for testing was 5.22. When it was changed to eleven parameters, the *RMSE* for training was 3.47 and for testing was 4.47. This identifiable trend continues with the *MSE* values (decreased with eleven noise predictors).

To further clarify how and why the RF model is the preferred algorithm, it is worth considering a few other attributes and potential applications for it. To start with, the RF model performed faster when used to process data sets of a large size, including multiple variables. The RF model is also capable of functioning and producing reliable results, even when some input values are missing. This is because it is an ensemble method. In turn, this makes the model able to create real-time predictions. Moreover, taken outside the research environment, the model would prove attractive to stakeholders requiring accurate predictive data for environmental reasons.

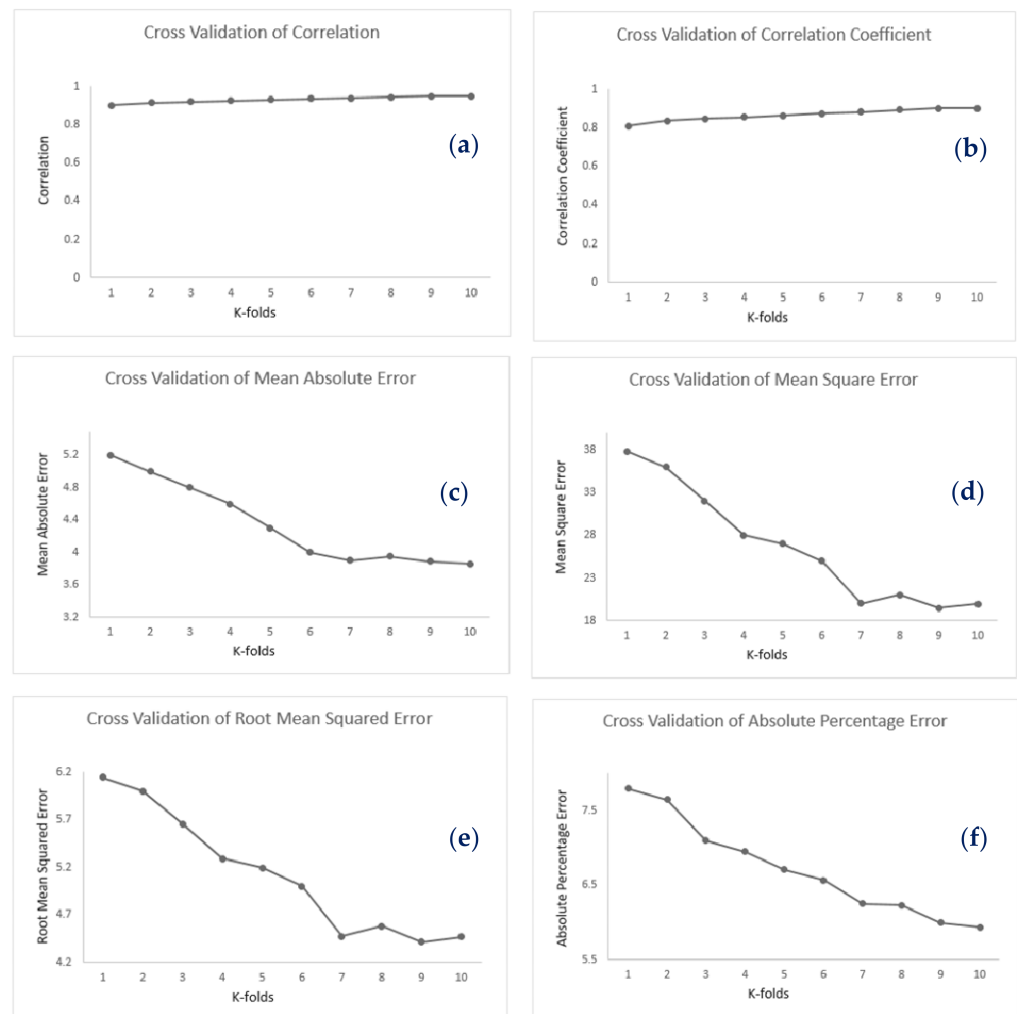
The success of the RF model at effectively and accurately predicting noise levels ( $L_{eq}$  dB) is described in Figure 11. These scatter plot graphs depict an imagined but plausible scenario that could occur based on the variables of the primary road, the bus stop, traffic volume, all type of roads, expressway, bus line, area of residential high density, industrial area, trees area, DSM, and WS. Predictions of the type shown would be of use to the environmental and town planning industries. For example, they could help predict the impact new infrastructure will have on the environment, or they could aide in designing and implementing programmes of traffic control, including the re-routing of vehicles and the creation of new roads.



**Figure 11.** Scatter plot of  $L_{eq}$  (measured vs predicted) using Random Forest for training and testing dataset with eleven parameters.

In addition to this, Figure 12 shows the data results when 10-fold cross-validation was performed on the RF model to see if it maintained its status as the best model for predicting noise levels.

Not only did the RF model stay performing well under the scrutiny of cross-validation, but it also proved itself to be stable. This is shown by how the six performance criteria stay regular across the six different iterations. Similarly, the observed and predicted  $L_{eq}$  figures for the testing data set were close in value. Figure 12 showed the values of  $L_{eq}$  predicted using RF fit well with the field data.



**Figure 12.** 10-fold cross-validation of (a)  $R$ , (b)  $R^2$ , (c)  $MAE$ , (d)  $MSE$ , (e)  $RMSE$ , and (f)  $MAPE$  in predicting  $L_{eq}$  using the testing dataset by RF method.

## 5. Conclusions

This research has evaluated the merits of four different soft computing models (machine learning and statistical regression) used to predict traffic noise levels at New Klang Valley Expressways in Shah Alam, Malaysia. In addition, it used six evaluation criteria for model performance assessment which would benefit researchers and practitioners. The successive stages of the research, including studying and changing the parameters, have been described in detail, along with information on the four different models and the research findings. The noise prediction models were developed, with  $L_{eq}$  as the output (dependent variable) and the following noise variables, primary road, bus stop, traffic volume, all types of roads, expressway, bus line, industrial area, trees area, DSM, and WS, as the independent variables. According to the performance criteria of  $R$ ,  $R^2$ ,  $MAE$ ,  $MSE$ ,  $RMSE$ , and  $MAPE$ , the results showed that the RF model was the most effective and reliable at predicting traffic noise levels. K-fold cross-validation further proved the stability of the RF model in making predictions. It is important to mention that proper validation of the results can only be done by evaluating them against long-term traffic data in the selected area. However, collecting long-term field data is both expensive and time-consuming. Therefore, this study considers both as limitation and out of scope. The data collection times were carefully selected to capture the general traffic trend in the selected area. The main objective was to demonstrate the capacity of LUR model as a potential tool for noise mapping. Results indicated that LUR model performed significantly

better than regression-based models despite using limited field data. Therefore, it can be ascertained that LUR models can be deployed for noise modelling when limited noise data is available.

The methodology introduced in this study can be extended and utilized to accommodate variation in number of variables to improve the prediction of traffic noise map. For future work, diverse traffic conditions with the inclusion of time as a variable can be explored by LUR based on the advanced artificial neural network and deep neural network models. Probably that may lead to increase the accuracy of prediction models like the ones discussed in this study. For urban areas like Shah Alam, the number of vehicles on city streets is the cause of high levels of pollution. In order to propose a useful model, this study combined the CFS, RF, and GIS models. The results of RF model then demonstrated the lowest *RMSE* 4.37 for testing data, and when all 18 parameters were used, the *RMSE* of testing was 5.22. The models were applied based on most of the parameters derived from LiDAR data, and GIS layers were extracted to produce noise prediction map. According to the prediction map, the highest values (high noise) were concentrated near expressway, whilst the lowest values (low noise) were distributed far away from expressway and primary road. However, both LUR statistical modelling and GIS techniques are important tools for planning and prediction maps. Ultimately, this study has proved that the machine learning model outperforms the regression model. The proposed models are an affordable and easy-to-use method for helping to monitor noise levels and could be useful for governmental and urban planning projects.

**Author Contributions:** Conceptualization, B.P.; resources, B.P.; methodology, B.P. and A.A.A.A.; software, B.P.; validation, A.A.A.A., B.P., S.C.; formal analysis, A.A.A.A.; investigation, A.A.A.A. and B.P.; data curation, A.A.A.A.; writing—A.A.A.A.; writing—review and editing, B.P., S.C. and A.A.; visualization, B.P., S.C. and A.A.; supervision, B.P.; project administration, B.P.; funding acquisition, B.P. and A.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Centre for Advanced Modelling & Geospatial Information Systems (CAMGIS), Faculty of Engineering & IT, University of Technology Sydney. Moreover, supported by the Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia, under Project RSP-2021/14. In addition, the second author, Biswajeet Pradhan, gratefully acknowledges the financial support from the UPM-PLUS industry project grant for collecting the various dataset used in this study. The APC was funded by Centre for Advanced Modelling & Geospatial Information Systems (CAMGIS), Faculty of Engineering & IT, University of Technology Sydney.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used in this study are available from the corresponding author.

**Acknowledgments:** The authors acknowledge and appreciate the provision of airborne laser scanning data (LiDAR), satellite images, and logistic support by the PLUS Berhad.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

Leq	Equivalent Continuous Sound Pressure
NKVE	New Klang Valley Expressway
LUR	Land Use Regression
GIS	Geographical Information Systems
LiDAR	Light Detection and Ranging
WHO	World Health Organization
DT	Decision Trees
RF	Random Forests



LR	Linear Regression
SVR	Support Vector Regression
R	Correlation
M	Correlation Coefficient
MAE	Mean Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
NDVI	Normalized Difference Vegetation Index
DSM	Digital Surface Model
WS	Wind Speed
VIF	Variance Inflation Factor
CFS	Correlation-Based Feature Selection
SVM	Support Vector Machine

## References

- Ahmed, A.A.; Pradhan, B. Vehicular traffic noise prediction and propagation modelling using neural networks and geospatial information system. *Environ. Monit. Assess.* **2019**, *191*, 1–17. [CrossRef]
- Ruiz-Padillo, A.; Ruiz, D.P.; Torija, A.J.; Ramos-Ridao, Á. Selection of suitable alternatives to reduce the environmental impact of road traffic noise using a fuzzy multi-criteria decision model. *Environ. Impact. Assess.* **2016**, *61*, 8–18. [CrossRef]
- Bluhm, G.; Nordling, E.; Berglind, N. Road traffic noise and annoyance—An increasing environmental health problem. *Noise Health* **2004**, *6*, 43–49.
- Moudon, A.V. Real noise from the urban environment: How ambient community noise affects health and what can be done about it. *Am. J. Prev. Med.* **2009**, *37*, 167–171. [CrossRef] [PubMed]
- Dzhambov, A.M.; Markevych, I.; Tilov, B.G.; Dimitrova, D.D. Residential greenspace might modify the effect of road traffic noise exposure on general mental health in students. *Urban For. Urban Green.* **2018**, *34*, 233–239. [CrossRef]
- Singh, D.; Kumari, N.; Sharma, P. A review of adverse effects of road traffic noise on human health. *Fluct. Noise Lett.* **2018**, *17*, 1830001. [CrossRef]
- Den Boer, L.C.; Schroten, A. Traffic noise reduction in Europe. *CE Delft* **2007**, *14*, 2057–2068.
- Steg, L.; Schuitema, G. Behavioural responses to transport pricing: A theoretical analysis. In *Threats from Car Traffic to the Quality of Urban Life*; Gärling, T., Steg, L., Eds.; Emerald Group Publishing Limited: Bingley, UK, 2007; pp. 347–366.
- Kawada, T. Noise and health—Sleep disturbance in adults. *J. Occup. Health* **2011**, *53*, 413–416. [CrossRef]
- Younes, I.; Shafiq, M.; Ghaffar, A.; Mehmood, S. *Spatial Patterns of Noise Pollution and Its Effects in Lahore City*; Anchor Academic Publishing: Humburg, Germany, 2007.
- Münzel, T.; Sørensen, M.; Schmidt, F.; Schmidt, E.; Steven, S.; Kröller-Schön, S.; Daiber, A. The adverse effects of environmental noise exposure on oxidative stress and cardiovascular risk. *Antioxid. Redox. Signal.* **2018**, *28*, 873–908. [CrossRef]
- Gan, W.Q.; McLean, K.; Brauer, M.; Chiarello, S.A.; Davies, H.W. Modeling population exposure to community noise and air pollution in a large metropolitan area. *Environ. Res.* **2012**, *116*, 11–16. [CrossRef]
- Licitra, G.; Teti, L.; Cerchiai, M.; Bianco, F. The influence of tyres on the use of the CPX method for evaluating the effectiveness of a noise mitigation action based on low-noise road surfaces. *Transp. Res. Part D Transp. Environ.* **2017**, *55*, 217–226. [CrossRef]
- Sandberg, U.; Ejsmont, J. *Tyre/Road Noise. Reference Book*; Infomex: Hard, Sweden, 2002; Available online: <https://trid.trb.org/view/730140> (accessed on 3 March 2017).
- Bianco, F.; Fredianelli, L.; Lo Castro, F.; Gagliardi, P.; Fidecaro, F.; Licitra, G. Stabilization of a pu sensor mounted on a vehicle for measuring the acoustic impedance of road surfaces. *Sensors* **2020**, *20*, 1239. [CrossRef]
- Praticò, F.G.; Fedele, R.; Pellicano, G. Monitoring Road Acoustic and Mechanical Performance. In *European Workshop on Structural Health Monitoring*; Rizzo, P., Milazzo, A., Eds.; Springer: Cham, Switzerland, 2020; Volume 127, pp. 594–602.
- Teti, L.; de León, G.; Del Pizzo, A.; Moro, A.; Bianco, F.; Fredianelli, L.; Licitra, G. Modelling the acoustic performance of newly laid low-noise pavements. *Constr. Build. Mater.* **2020**, *247*, 118509. [CrossRef]
- Del Pizzo, A.; Teti, L.; Moro, A.; Bianco, F.; Fredianelli, L.; Licitra, G. Influence of texture on tyre road noise spectra in rubberized pavements. *Appl. Acoust.* **2020**, *159*, 107080. [CrossRef]
- Praticò, F.G. On the dependence of acoustic performance on pavement characteristics. *Transp. Res. Part D Transp. Environ.* **2014**, *29*, 79–87. [CrossRef]
- Praticò, F.G.; Anfosso-Lédée, F. Trends and issues in mitigating traffic noise through quiet pavements. *Procedia-Soc. Behav. Sci.* **2012**, *53*, 203–212. [CrossRef]
- de León, G.; Del Pizzo, A.; Teti, L.; Moro, A.; Bianco, F.; Fredianelli, L.; Licitra, G. Evaluation of tyre/road noise and texture interaction on rubberised and conventional pavements using CPX and profiling measurements. *Road Mater. Pavement.* **2020**, *21*, 91–102. [CrossRef]
- Harouvi, O.; Ben-Elia, E.; Factor, R.; de Hoogh, K.; Kloog, I. Noise estimation model development using high-resolution transportation and land use regression. *J. Expo. Sci. Environ. Epidemiol.* **2018**, *28*, 559–567. [CrossRef]

23. Licitra, G.; Fredianelli, L.; Petri, D.; Vigotti, M.A. Annoyance evaluation due to overall railway noise and vibration in Pisa urban areas. *Sci. Total Environ.* **2016**, *568*, 1315–1325. [[CrossRef](#)]
24. Bunn, F.; Zannin, P.H.T. Assessment of railway noise in an urban setting. *Appl. Acoust.* **2016**, *104*, 16–23. [[CrossRef](#)]
25. Iglesias-Merchan, C.; Diaz-Balteiro, L.; Soliño, M. Transportation planning and quiet natural areas preservation: Aircraft overflights noise assessment in a National Park. *Transp. Res. Part D Transp. Environ.* **2015**, *41*, 1–12. [[CrossRef](#)]
26. Gagliardi, P.; Fredianelli, L.; Simonetti, D.; Licitra, G. ADS-B system as a useful tool for testing and redrawing noise management strategies at Pisa Airport. *Acta Acust. United Acust.* **2017**, *103*, 543–551. [[CrossRef](#)]
27. Fredianelli, L.; Bolognese, M.; Fidecaro, F.; Licitra, G. Classification of noise sources for port area noise mapping. *Environments* **2021**, *8*, 12. [[CrossRef](#)]
28. Fredianelli, L.; Nastasi, M.; Bernardini, M.; Fidecaro, F.; Licitra, G. Pass-by characterization of noise emitted by different categories of seagoing ships in ports. *Sustainability* **2020**, *12*, 1740. [[CrossRef](#)]
29. Aguilera, I.; Foraster, M.; Basagaña, X.; Corradi, E.; Deltell, A.; Morelli, X.; Künzli, N. Application of land use regression modelling to assess the spatial distribution of road traffic noise in three European cities. *J. Expo. Sci. Environ. Epidemiol.* **2015**, *25*, 97–105. [[CrossRef](#)] [[PubMed](#)]
30. Manvell, D.; van Banda, E.H. Good practice in the use of noise mapping software. *Appl. Acoust.* **2011**, *72*, 527–533. [[CrossRef](#)]
31. Basagana, X.; Rivera, M.; Aguilera, I.; Agis, D.; Bousa, L.; Elosua, R.; Kuenzli, N. Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmos. Environ.* **2012**, *54*, 634–642. [[CrossRef](#)]
32. Jerrett, M.; Arain, A.; Kanaroglou, P.; Beckerman, B.; Potoglou, D.; Sahuvaroglu, T.; Giovis, C. A review and evaluation of intraurban air pollution exposure models. *J. Expo. Sci. Environ. Epidemiol.* **2005**, *15*, 185–204. [[CrossRef](#)]
33. Xie, D.; Liu, Y.; Chen, J. Mapping urban environmental noise: A land use regression method. *Environ. Sci. Tech. Lib.* **2011**, *45*, 7358–7364. [[CrossRef](#)]
34. Sieber, C.; Ragetti, M.S.; Brink, M.; Toyib, O.; Baatjies, R.; Saucy, A.; Rössli, M. Land use regression modeling of outdoor noise exposure in informal settlements in Western Cape, South Africa. *Int. J. Environ. Res. Pub. Health* **2017**, *14*, 1262. [[CrossRef](#)]
35. Qing-Fang, M.; Yue-Hui, C.; Yu-Hua, P. Small-time scale network traffic prediction based on a local support vector machine regression model. *Chin. Phys. B* **2009**, *18*, 2194. [[CrossRef](#)]
36. Ragetti, M.S.; Goudreau, S.; Plante, C.; Fournier, M.; Hatzopoulou, M.; Perron, S.; Smargiassi, A. Statistical modeling of the spatial variability of environmental noise levels in Montreal, Canada, using noise measurements and land use characteristics. *J. Expo. Sci. Environ. Epidemiol.* **2016**, *26*, 597–605. [[CrossRef](#)]
37. Singh, D.; Nigam, S.P.; Agrawal, V.P.; Kumar, M. Vehicular traffic noise prediction using soft computing approach. *J. Environ. Manage.* **2016**, *183*, 59–66. [[CrossRef](#)] [[PubMed](#)]
38. Liu, Y.; Goudreau, S.; Oiamo, T.; Rainham, D.; Hatzopoulou, M.; Chen, H.; Smargiassi, A. Comparison of land use regression and random forests models on estimating noise levels in five Canadian cities. *Environ. Pollut.* **2020**, *256*, 113367. [[CrossRef](#)]
39. Garg, N.; Mangal, S.K.; Saini, P.K.; Dhiman, P.; Maji, S. Comparison of ANN and analytical models in traffic noise modeling and predictions. *Acoust. Aust.* **2015**, *43*, 179–189. [[CrossRef](#)]
40. Ahn, J.; Ko, E.; Kim, E.Y. Highway traffic flow prediction using support vector regression and Bayesian classifier. In Proceedings of the International Conference on Big Data and Smart Computing (BigComp), Hong Kong, China, 18–20 January 2016.
41. Crosetto, M.; Tarantola, S. Uncertainty and sensitivity analysis: Tools for GIS-based model implementation. *Int. J. Geogr. Inf. Sci.* **2001**, *15*, 415–437. [[CrossRef](#)]
42. Tao, S.; Manolopoulos, V.; Rodriguez Duenas, S.; Rusu, A. Real-time urban traffic state estimation with A-GPS mobile phones as probes. *J. Transp. Tech.* **2012**, *2*, 22–31. [[CrossRef](#)]
43. Hsieh, F.Y.; Bloch, D.A.; Larsen, M.D. A simple method of sample size calculation for linear and logistic regression. *Stat. Med.* **1998**, *17*, 1623–1634. [[CrossRef](#)]
44. Craney, T.A.; Surles, J.G. Model-dependent variance inflation factor cutoff values. *Qual. Eng.* **2002**, *14*, 391–403. [[CrossRef](#)]
45. Hall, M.A. Correlation-Based Feature Selection for Machine Learning. Ph.D. Thesis, The University of Waikato, Hamilton, New Zealand, 1999.
46. Azeez, O.S.; Pradhan, B.; Shafri, H.Z. Vehicular CO emission prediction using support vector regression model and GIS. *Sustainability* **2018**, *10*, 3434. [[CrossRef](#)]
47. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [[CrossRef](#)]
48. Draper, C.; Reichle, R.; de Jeu, R.; Naeimi, V.; Parinussa, R.; Wagner, W. Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. *Remote Sens. Environ.* **2013**, *137*, 288–298. [[CrossRef](#)]
49. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model. Dev.* **2014**, *7*, 1247–1250. [[CrossRef](#)]
50. Verlinden, B.; Dufloy, J.R.; Collin, P.; Cattrysse, D. Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study. *Int. J. Prod. Econ.* **2008**, *111*, 484–492. [[CrossRef](#)]
51. Atsalakis, G.S.; Valavanis, K.P. Surveying stock market forecasting techniques—Part II: Soft computing methods. *Expert Syst. Appl.* **2009**, *36*, 5932–5941. [[CrossRef](#)]