Modelling of Ambient Noise Levels in Urban Environment

S. K. Tiwari, L. A. Kumaraswamidhas, and N. Garg

Abstract The study conducts time-series approach for analysing one year noise monitoring data. Support vector machine (SVM) technique is used as a modelling technique for time-series approach. The noise data is trained using tenfold cross-validation to get optimum hyperparameters (γ, ε, C). The performance and accuracy of model are determined by statistical parameters like MSE, RMSE, MAPE in %, R^2. The paper predicts an error of ± 2 dB(A) with the implementation of support vector machine (SVM).

Keywords Noise monitoring · Support vector machine · Ambient noise levels · Tenfold cross-validation · Day noise level · Night noise level

1 Introduction

Noise pollution has significantly increased especially in the urban areas in Indian scenario. With advancement of vehicles and urbanization in cities, there has been a quick addition in traffic volume. Regardless of the way that transportation is an essential part of urban society, its superiority is obscured by its negativity. Inappropriate placement of vehicles at different locations nearby roads is one of the major cause of traffic jam. Some studies affirm that noise pollution has adverse influence on human health [1, 2]. It comprises slant stress impact, sleeping disturbances which clearly cause 'prompt effect' on mental and physical perspective. The Central Pollution Control Board has directed noise levels for different zones, i.e. silence, industrial, commercial, residential zones, and carried many studies for noise monitoring in Indian scenario [3]. Garg et al. [4] discussed the pilot project on the establishment of National Ambient Noise Monitoring Network (*NANMN*) at 35 locations across

S. K. Tiwari (✉) · L. A. Kumaraswamidhas
Indian Institute of Technology (ISM), Dhanbad 826001, India
e-mail: shashikanttiwari61@gmail.com

N. Garg
CSIR-National Physical Laboratory, New Delhi 110012, India

the seven major cities of the country. The European Environmental Noise Directive 2002/49/EC [5] gives direction for noise mapping that include future plans with financial information and cost-effective assessment. European Directive permits to survey and to look at, inside EU Member States, noise exposure data, particularly for the future execution steps, when noise maps should be drawn up with the basic evaluation strategies.

There are various techniques used for noise assessment and monitoring. Some uses long-term noise monitoring strategy, while Garg et al. [6] emphasized on short-term noise monitoring strategy as a reliable strategy within an accuracy of ± 2 dB(A). The high costs of installing and maintaining permanent networks are primarily the main reason for analysing the suitability of short-term strategies to ascertain whether they can provide a suitable and reliable alternative or not as compared to the long-term noise monitoring. There are some illustrations whereby extensive networks have been installed [7]. Morillas and Gajardo [8] evaluated 90% probability interval for random 9 days data to measure L_{den}. Hence, there is a need of alternative approach to predict and forecast ambient noise level by using time-series approach. DeVor et al. [9] used autoregressive moving average (*ARMA*) and Garg et al. [10] used autoregressive integrated moving average (*ARIMA*) model to predict and forecast noise level in time-series analysis. This study explores SVM technique for forecasting and determining the accuracy and performance of the predicted noise level.

2 Methodology

2.1 Ambient Noise Level Calculation

The present study is helpful to opt an optimized strategy for forecasting with the help of a time-series method (*SVM*) in Indian scenario using the 3-year database with minimum error.

The value of A-weighted day noise level (L_{day}) and night noise level (L_{night}) is calculated:

$$L_{\text{day},n} = 10 \log \left[\frac{1}{n} \sum_{i=1}^{n} 10^{0.1(l_{\text{day},i})} \right] \tag{1}$$

$$L_{\text{night},n} = 10 \log \left[\frac{1}{n} \sum_{i=1}^{n} 10^{0.1(l_{\text{night},i})} \right] \tag{2}$$

where n denotes the numbers of days and nights in long-term noise monitoring strategy. The error is calculated as difference of observed noise level with the predicted noise level for the commercial site for a particular period of data.

2.2 Support Vector Machine

The idea of support vector machine (*SVM*) is mapping a nonlinear dataset. The approach focuses to solve a regression using linear function. The hyperplane is also known as classifier separates classes to get an optimal solution. The data is spitted into training and testing data. Suppose X_i represents input data ($i = 1, ..., n$) where n is the number of training data points. The function of hyperplane is as [11]:

$$Y(x) = w^T X_i + b \tag{3}$$

where w is the orientation and b is the position of hyperplane classifying the training data into two classes. C_1 is the positive class, and C_2 is the negative class. The main focus of SVM is to find a new classifier.

$$Y(x_1) = w^T x_i + b > 0 \tag{4}$$

$x_1 \square C_1$, if x_1 lies on the positive side of the hyperplane.

$$Y(x_2) = w^T x_i + b < 0 \tag{5}$$

$x_2 \square C_2$, if x_1 lies on the positive side of the hyperplane.

2.3 Nonlinearity in Data

For a nonlinear separable data, SVM can be a optimized time-series technique to have a good solution. The w and b are determined by two mathematical formulations of a regularized function ($R(c)$) in *SVM* [11–13].

$$R(c) = \frac{1}{2}\|w\|^2 + \frac{c}{n} \sum_{i=1}^{n} L_s(d_i, y_i) \tag{6}$$

$$L_s(d_i, y_i) = \begin{cases} |d_i - y_i| - \varepsilon & 0 \\ |d_i - y_i| & \text{otherwise} \end{cases} \tag{7}$$

$\frac{c}{n} \sum_{i=1}^{n} L_s(d_i, y_i)$ is empirical error.

The dot product of two input $\psi(x)$ and $\psi(y)$ vectors is kernel function that should affirm Mercer's condition. Mainly, four kernels are utilized in support vector machine (SVM) modelling which are as follows [12]:

$$\text{Linear Kernel: } K(x, y) = x^T y \tag{8}$$

$$\text{Radial Basis Function Kernel: } K(x, y) = \exp(-\gamma/x - y^2), \quad \gamma > 0 \qquad (9)$$

$$\text{Polynomial Kernel: } K(x, y) = (\gamma x^{\mathrm{T}} y + r)^d, \quad \gamma > 0 \qquad (10)$$

$$\text{Sigmoid Kernel: } K(x, y) = \tanh(\gamma x^{\mathrm{T}} y + r) \qquad (11)$$

Here, the kernel parameters are d, r and γ. Kernel parameters have important significance in the performance of support vector machine model. The complexity of best parameter is controlled by the kernel functions. These kernel functions increase the accuracy of the model as compared to the other NN models with the help of other hyperparameters. The main purpose of kernel parameters is to convert a complex data, i.e. in the form of lower-dimensional space to higher-dimensional space. Here for the research work, radial basis function is used because of its higher accuracy in comparison to other kernel functions.

3   Results and Discussion

Figure 1a shows variation plot of L_{day} in dB(A) for 365 days for a residential site in Delhi. Maximum value of L_{day} is 72 dB(A), and minimum value is 60 dB(A). While Fig. 1b shows variation plot of L_{night} in dB(A) for 365 days for a residential site in Delhi. Maximum value of L_{night} is 72 dB(A), and minimum value is 60 dB(A). The noise values are taken in A-weighting because in most of the industrial application, noise levels are taken in the form of A-weighting only in comparison to C-weighting. The C-weighting noise values are used for very tuned and fined noise, while A-weighting noise levels are used for human audible range.

The kernel applied in the study is radial basis function (RBF) kernel. It has better performance in comparison to the others kernel functions. In RBF kernel, three hyperparameters have been used to analyse the performance of SVM model. These three hyperparameters are: Gamma (γ), Epsilon (ε) and Cost (C). The first stage is to find an optimized parametric combination of these hyperparameters. Hit-and-trial approach was attempted to get the optimum value of hyperparameters. The parametric combination (γ, ε, C) of optimized hyperparameters is (2^5, 0.4, 2^5) for both day and night. Figure 2a, b shows the plot of predicted day and night ambient noise level in comparison to observed noise levels.

The input data is divided into testing and training data. 90% of the input data is taken as training data, and 10% of the data is used as testing data. Mean square error (MSE), root mean square error ($RMSE$), mean average percentage error ($MAPE$ in%), coefficient of determination (R^2) are the parameters that ascertain the efficiency of the model.

Table 1 shows the statistics performance of training data for both day and night noise levels. The maximum error is 4.54 dB(A) for day and 5.37 dB(A) for night, but the MSE and RMSE error lies within ± 2 dB(A) which is sought of reliable accuracy.
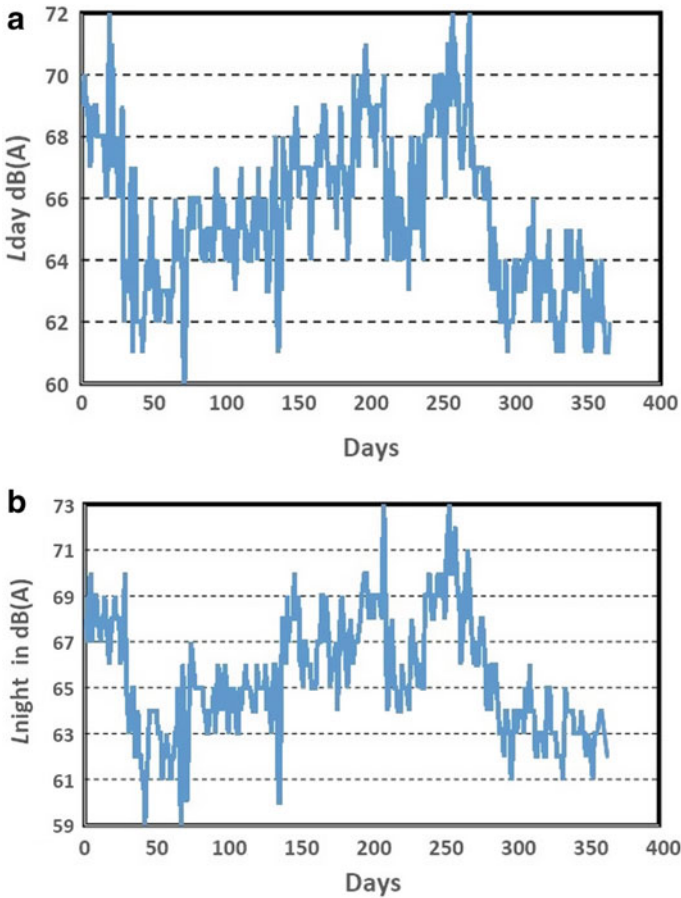
Fig. 1 **a** Time sequence plot of L_{day} in dB(A) for 365 days for a residential site in Delhi. **b** Time sequence plot of L_{night} in dB(A) for 365 days for a residential site in Delhi

The probability of the training data determined from input data can be taken as per the study analysis. There is no particular approach to determine the probability like in the present study, the training data is taken 90%. The determination of coefficient is 0.6 for day and 0.4 for night which implies better performance of the classifier.

To test the model, the testing data is taken 10% of the input data. The SVM model predicts an error of ± 2 dB(A) for testing data as well. The determination of coefficient is 0.6 for day and 0.5 for night which implies better performance of the classifier of predicted testing data (Table 2). Hence, it can be observed that the support vector machine is a good approach for forecasting of ambient noise level within an accuracy of ± 2 dB(A). Figure 3a, b shows the standardized residual analysis of *SVM* model for L_{day} and L_{night} dB(A) for both training and testing data.
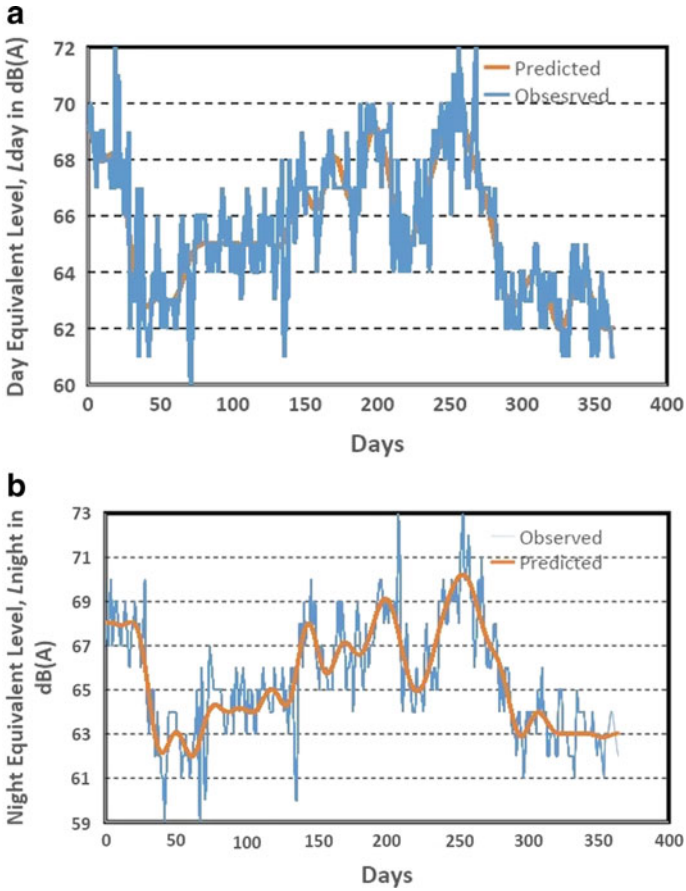
Fig. 2 a Comparison of measured (blue line) and predicted (red line) values of L_{day} in dB(A).
b Comparison of measured (blue line) and predicted (red line) values of L_{night} in dB(A)

Table 1 *SVM* model statistics for training data

Statistical parameter	L_{day} dB(A)	L_{night} dB(A)
MSE in dB(A)^2	1.640	1.664
RMSE in dB(A)	1.281	1.290
MAPE in %	1.47	1.43
Maximum error in dB(A)	4.54	5.37
Minimum error in dB(A)	−4.37	−5.96
R^2	0.6	0.4

Table 2 *SVM* model statistics for testing data

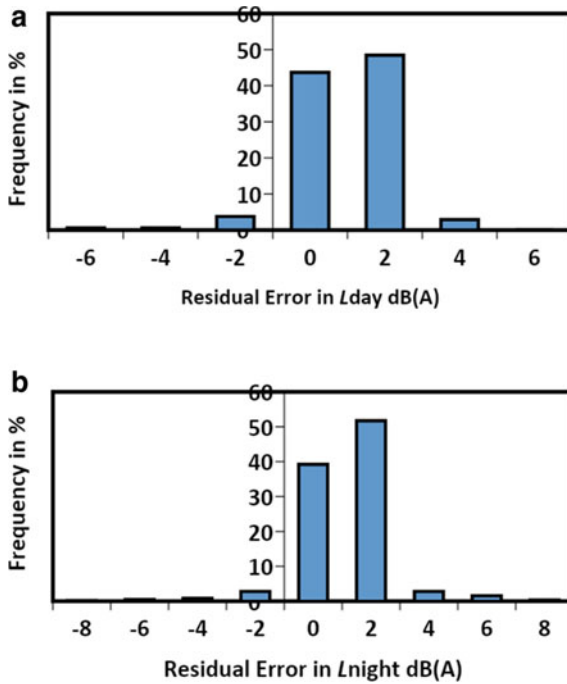| Statistical parameter	L_{day} dB(A)	L_{night} dB(A)
MSE in dB(A)^2	2.316	1.695
RMSE in dB(A)	1.522	1.302
MAPE in %	1.45	1.82
Maximum error in dB(A)	2.47	1.97
Minimum error in dB(A)	−3.20	−3.09
R^2	0.6	0.5

Fig. 3 **a** Standardized residual analysis of *SVM* model for L_{day} dB(A) for both training and testing data. **b** Standardized residual analysis of *SVM* model for L_{night} dB(A) for both training and testing data

4 Conclusion

In the study, *SVM* is used as a time-series modelling technique for the statistical analysis of one-year noise monitoring data set. *SVM* is an outperforming technique that can profoundly predict the ambient noise levels L_{day} and L_{night}. The parametric combination for both day and night is $(2^5, 0.4, 2^5)$. Meanwhile, this is the best set of hyperparameter for classifier which represents a similar trend as the observed pattern.

The application of SVM has rarely been used in the determination of ambient noise level. The result shows that this model can be used as a better fitting model for predicting and forecasting noise levels. The work also emphasizes on the use of RBF kernel to analyse *SVM* model. The performance of model is determined by the statistical parameters like *MSE*, *RMSE*, *MAPE* in % and R^2.

References

1. Ising H, Kruppa B (2004) Health effects caused by noise: evidence in the literature from the past 25 years. Noise Health 6(22):5
2. World Health Organization (2011) Burden of disease from environmental noise. WHO, Geneva
3. Central Pollution Control Board, Annual report, 2011–2012. https://cpcb.nic.in/upload/Annual Reports/AnnualReport%2043AR%202011-12%20English.pdf [access on 02.03.2016]
4. Garg N, Sinha AK, Gandhi V, Bhardwaj RM, Akolkar AB (2016) A pilot study on the establishment of national ambient noise monitoring network across the major cities of India. Appl Acoust 103:20–29
5. European Noise Directive (2002). Assessment and management of environmental noise. 2002/49/EU, Official Journal of European Communities. DIRECTIVE 2002/49/EC of the European Parliament and of the council of 25 June 2002 relating to the assessment and management of environmental noise
6. Garg N, Saxena TK, Maji S (2015) Long-term versus short-term noise monitoring: strategies and implications in India. Noise Control Eng J 63(1):26–35(10)
7. Czyzewski A, Kotus J, Szczodrak M (2012) On-line urban acoustic noise monitoring system. Noise Control Eng J 60(1):69–84
8. Morillas JB, Gajardo CP (2014) Uncertainty evaluation of continuous noise sampling. Appl Acoust 75:27–36
9. DeVor RE, Schomer PD, Kline WA, Neathamer RD (1979) Development of temporal sampling strategies for monitoring noise. J Acoust Soc Am 66(3):763–771
10. Garg N, Soni K, Saxena TK, Maji S (2015) Applications of autoregressive integrated moving average (ARIMA) approach in time-series prediction of traffic noise pollution. Noise Control Eng J 63(2):182–194
11. Samsudin R, Shabri A, Saad P (2010) A comparison of time-series forecasting using support vector machine and artificial neural network model. J Appl Sci 10(11):950–958
12. Ding Y, Song X, Zen Y (2008) Forecasting financial condition of Chinese listed companies based on support vector machine. Expert Syst Appl 34(4):3081–3089
13. Benítez-Peña S, Blanquero R, Carrizosa E, Ramírez-Cobo P (2019) Cost-sensitive feature selection for support vector machines. Comput Oper Res 106:169–178