Applications of AutoRegressive Integrated Moving Average (ARIMA) approach in time-series prediction of traffic noise pollution

N. Garg[a)], K. Soni[b)], T.K. Saxena[b)] and S. Maji[c)]

(Received: 13 September 2014; Revised: 20 April 2015; Accepted: 22 April 2015)

The paper analyzes the long-term noise monitoring data using the AutoRegressive Integrated Moving Average (ARIMA) modeling technique. Box–Jenkins ARIMA approach has been adapted to simulate the daily mean L_{Day} (06–22 h) and L_{Night} (22–06 h) in A- and C-weightings in conjunction with single-noise metrics, day–night average sound level (DNL) for a period of 6 months. The autocorrelation function (ACF) and partial autocorrelation function (PACF) have been obtained to find suitable orders of autoregressive (*p*) and moving average (*q*) parameters for ARIMA (*p*, *d*, *q*) models so developed for all the single-noise metrics. The ARIMA models, namely, ARIMA(0,0,14), ARIMA(0,1,1), ARIMA(7,0,0), ARIMA(1,0,0) and ARIMA(0,1,14), have been developed as the most suitable for simulating and forecasting the daily mean L_{Day} dBA, L_{Night} dBA, L_{Day} dBC, L_{Night} dBC, and day–night average sound level (DNL) respectively. The performance of the model so developed is ascertained using the statistical tests. The work reveals that the ARIMA approach is reliable for time-series modeling of traffic noise levels. **© *2015 Institute of Noise Control Engineering.***

Primary subject classification: 52.3; Secondary subject classification: 56.1

1 INTRODUCTION

Traffic noise has accentuated to an intolerable level in recent years in Delhi city owing to the alarming increase in vehicular density. With rapid urbanization and economic advancements, vehicular density has considerably increased, causing a serious noise and air pollution in Delhi. Delhi has been suffering from adverse traffic conditions for the past few years with rapid urbanization and population explosion. It is thus imperative to monitor the accentuated ambient noise levels for devising methods for combating the radiated vehicular noise and also planning of guidelines to be implemented for the new projects. A calibrated noise model with highest accuracy is thus essential for town planners and architects. There has been a great emphasis on noise and its ill effects in developing nations and each nation has now come up with a validated road traffic noise model[1,2]. The European Directive 2002/49/EC lays emphasis on environmental noise and related problems and obligates the member states to provide access to the information on noise pollution[3]. Numerous noise mapping studies have been reported in different countries[4–9]. However, as of yet, no study has been reported on the time-series prediction of traffic noise levels.

There has been no standard traffic noise model followed in Indian conditions, although many studies have been reported in recent years in various parts of the country[10–15]. The Central Pollution Control Board (CPCB), Delhi has taken many initiatives and carried out numerous studies for monitoring the ambient noise levels at hot spots in Delhi and implementation of suitable measures to be adopted for noise mitigation. The introduction of mass rapid transit systems (Delhi Metro) and subsequent expansion has, however, provided a major relief from high vehicular density plying on Delhi roads, although the operation of metro trains can cause a cumulative increase in ambient A-weighted noise levels of a maximum of 2 to 3 dB in medium and high traffic density areas[16]. The growing awareness of community towards noise pollution has led to serious concerns and implementation of strict measures for mitigation of noise as well as systematic planning beforehand for new projects. Long-term noise monitoring has been reported in only a few studies in India due to lack of infrastructural support[17–22]. CPCB initiatives in this regard in setting up of noise monitoring stations in various parts of the country are an effective approach for analyzing the ambient noise levels and planning for suitable measures for

[a)] CSIR—National Physical Laboratory, New Delhi – 110 012, INDIA; email: ngarg@nplindia.org.
[b)] CSIR—National Physical Laboratory, New Delhi – 110 012, INDIA.
[c)] G. B. Pant Govt. Engg. College, New Delhi – 110 020, INDIA.

traffic noise abatement. The noise data gathered in systematic monitoring of ambient noise under the National Ambient Noise Monitoring Network (NANMN) would help the government better implement the recently amended Noise Pollution (Regulation and Control) Rules, according to which the use of construction machines, musical instruments, bursting of noise-emitting firecrackers and horns beyond permissible limits at nights in residential areas have been made punishable offenses[23]. A continuous long-term noise monitoring of ambient noise levels is cumbersome and quite expensive and in many cases the larger resources involved are not justified in terms of better accuracy than that achievable by cheaper and feasible temporal samplings such as that reported by Brambilla et al.[24]. Gaja et al.[25] summarizes 5 years of continuous noise measurements carried out in Valencia, Spain and recommends that a random day strategy for sampling gives a more accurate representation than consecutive day strategy. Recently, Morillas and Gajardo[26] investigations show that it needs to take measurements for 9 days spread randomly throughout the year for obtaining an estimate of day–evening–night levels, L_{den} with a probability of success within 95% confidence interval. The number of sampling days is recommended between 30 and 35 for a probability of 95%. A recent study conducted by the authors for the noise data gathered under pilot NANMN project recommended the random two month strategy, whereby an error of ± 2 dBA is achieved with a probability higher than 90%[27]. The time-series analysis of ambient noise shall be thus instrumental in forecasting the future noise levels in addition to the continuous noise monitoring attributed to the stochastic nature of traffic noise. It can also serve as a suitable substitute to continuous long-term noise monitoring provided the predicted data matches well with the actual measurement data. There have been very few studies reported so far[28–31] on the application of this approach to noise modeling. Kumar and Jain[28] analyzed the short-term noise levels measured at 10 s intervals in the vicinity of a busy road carrying vehicular traffic using the ARIMA approach. DeVor et al.[29] used ARMA (AutoRegressive Moving Average) model to assess the level of autocorrelation in the data via the Dynamic Data System approach to time series analysis. For reliable estimation of the mean level within a ± 5 dB range, it was recommended that the sample size in the range of 20–50 consecutive daily averages would be required. Schomer et al.[30] simulations demonstrate that nonconsecutive sampling strategies reduce the overall sampling requirements for non stationary data. The exhaustive literature review reveals that ARIMA methodology has not been implemented so far for long-term noise monitoring, although it has been extensively used in air and water pollution predictions[32–36]. The present work extends Kumar and Jain[34] utilizing the ARIMA approach for time-series predictions and forecasting of traffic noise levels. The stationary R^2, R^2, root mean squared error, mean absolute percentage error, normalized Bayesian information criterion, Ljung–Box analysis were used to ascertain the validity and suitability of the developed ARIMA model.

2 METHODOLOGY

The noise monitoring data analyzed in the present study are reported from a CPCB noise monitoring station situated in Delhi that comes under the commercial zone. The noise monitoring terminal is a standalone operating remote terminal sound-level meter consisting of a high quality microphone connected to an advanced acoustic signal-processing unit connected to an advanced high resolution data logger. The noise data are acquired locally, archived and communicated to a central station through an integrated GPRS modem[37]. The noise data considered in the present work is 181 days continuous noise monitoring data taken from September 2013 to February, 2014[38] at the CPCB noise monitoring station situated in the commercial area in Delhi.

2.1 L_{Day} (06–22 h) and L_{Night} (22–06 h) and Day–Night Average Sound Level (DNL)

The value of L_{Day} and L_{Night} is calculated as:

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

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

where n is the number of days or nights included in the long-term and $L_{\text{Day},i}$ and $L_{\text{Night},i}$ are the ith corresponding A-weighted equivalent level for the considered period. In present case, $n = 181$, while according to the 2002/49/EC requirements $n = 365$. The day-time means from 6.00 a.m. to 10.00 p.m., while the night time means from 10.00 p.m. to 6.00 a.m.[23]. It may be noted that the ambient air quality standards w.r.t. noise are prescribed by CPCB in terms of L_{Day} (and L_{Night}), while single-noise metrics, day–night average sound level (DNL) calculated using equation (3) is used to know the sound exposure on people/residents due to aircrafts and for land use planning around airports[39]. The present study considers DNL as an additional representative noise metric for the assessment of overall average sound levels and for ARIMA modeling.

$$DNL = 10\ \log\left[\frac{1}{24}\left(16\times 10^{\left(\frac{L_{Day}}{10}\right)} + 8\times 10^{\left(\frac{L_{Night}+10}{10}\right)}\right)\right]. \quad (3)$$

The standard deviation of the DNL is given by:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(L_{DNL}^{(i)} - \overline{L_{DNL}}\right)^2}{n-1}}, \quad (4)$$

where $\overline{L_{DNL}}$ is average of values of L_{DNL}^{i} for 181 days.

2.2 ARIMA Modeling Approach

The ARIMA approach was first popularized by Box and Jenkins, and ARIMA models are often referred to as Box–Jenkins models. This process is classified as linear models that is capable in presenting both stationary and non-stationary time-series. This approach has been extensively reported in many studies pertaining to time-series forecasting in various fields[40] although it suffers from the limitation in presuming linear form of the model[41]. AutoRegressive Moving Average (ARMA) model is based on a combination of two processes, AR (AutoRegressive) and MA (Moving Average). AR (p) model denotes that the current value of x_i of a stationary series can be explained as a linear function of p past values, $x_{t-1}, x_{t-2}\ldots x_{t-p}$ where p determines the number of steps into the past needed to forecast the current value.

The mathematical formulation of the ARIMA (p, d, q) model using lag polynomials for a time-series data Y_t where t is an integer and Y_t is real numbers[42,43]:

$$\left(1-\sum_{i=1}^{p}\varphi_i L^i\right)(1-L)^d Y_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t, \quad (5)$$

where p, d and q are integers greater than or equal to zero and refer to the order of autoregressive, integrated and moving average parts of the model respectively. The integer d controls the level of differencing. When $d = 0$, then it reduces to an ARMA (p, q) model. ARIMA model is a generalization of an ARMA model to include the case of non-stationary as well. In Eqn. (5), L is the lag operator, φ_i is the parameters of the autoregressive part of the model, θ_i is the parameters of the moving average part and ε_t is error terms. In time-series analysis, an event occurring at time $t + k$ ($k > 0$) is said to lag behind an event occurring at time t, the extent of the lag being k. The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean. In practice, most economic or business time-series can be modeled with rather modest numbers of terms, p and q, in the form of an AutoRegressive (AR), a Moving Average (MA), or an ARMA model. In order to achieve parsimony, the forecaster's task is to identify the smallest numbers of terms, p and q, to include within the model and still satisfactorily forecast the series[44]. An ARIMA (p, 0, q) or ARMA (p, q) is a model for time-series that depends on p past values of itself and q past random error terms e_t. This method has a form of Eqn. (6) as[45]:

$$\begin{aligned} Y_t = \theta_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots\ldots + \varphi_p Y_{t-p} \\ + \mu - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots\ldots - \theta_q e_{t-q} + e_t \end{aligned}, \quad (6)$$

where θ_1, θ_2........ θ_q are the finite weights, μ is the mean of series, Y_t is the forecasted output, Y_{t-p} is observation at time $t - p$ and $\varphi_1, \varphi_{2,}$.......... φ_p is a set of finite parameters determined by linear regression and e_t is an error associated with the regression.

The ARIMA modeling involves three stages: identification stage, estimation stage and diagnostic checking stage. In identification stage, it is ensured that the time-series is sufficiently stationary, i.e. free from trend and seasonality[41]. The estimation stage involves the estimation of parameters p and q of autoregressive moving

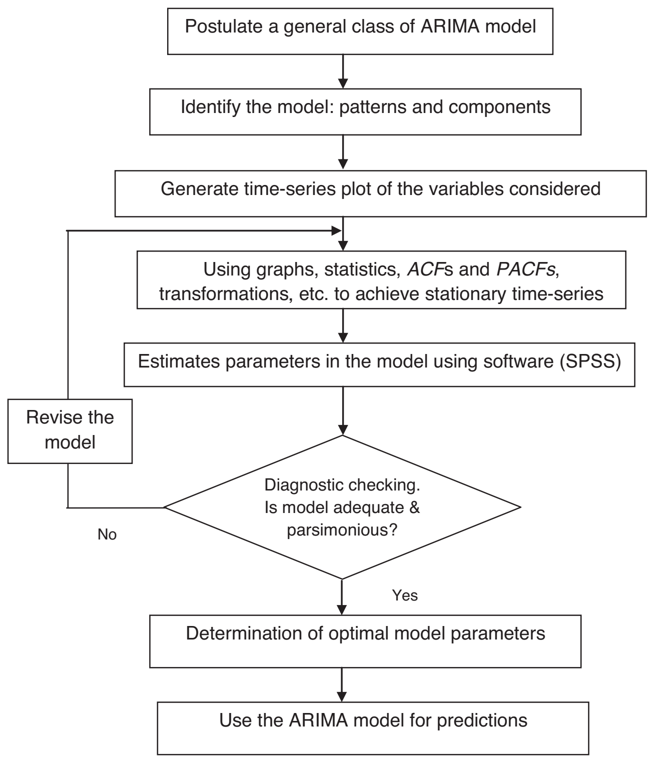

Fig. 1—Flow chart depicting the ARIMA methodology used.

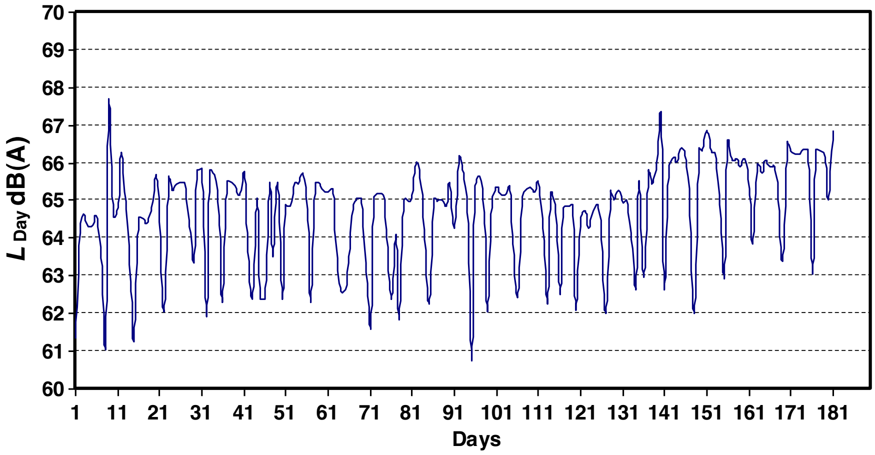

Fig. 2—Time sequence plot of L_{Day} (in dBA) for September, 2013 to February, 2014.

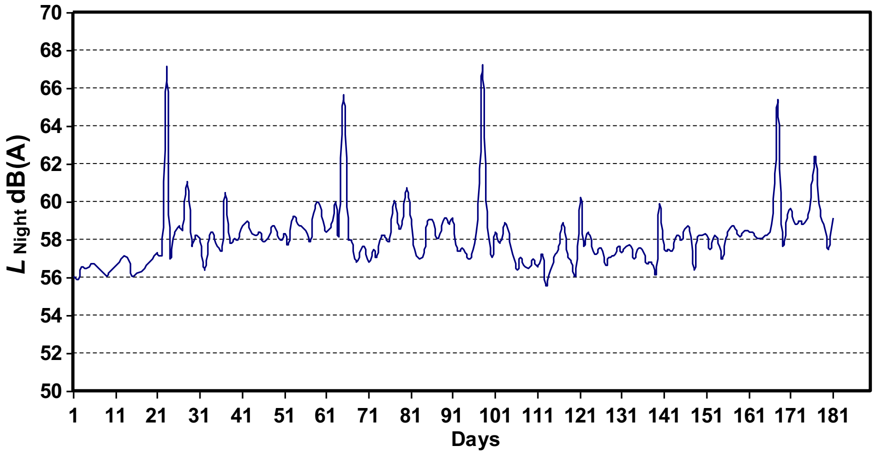

Fig. 3—Time sequence plot of L_{Night} (in dBA) for September, 2013 to February, 2014.

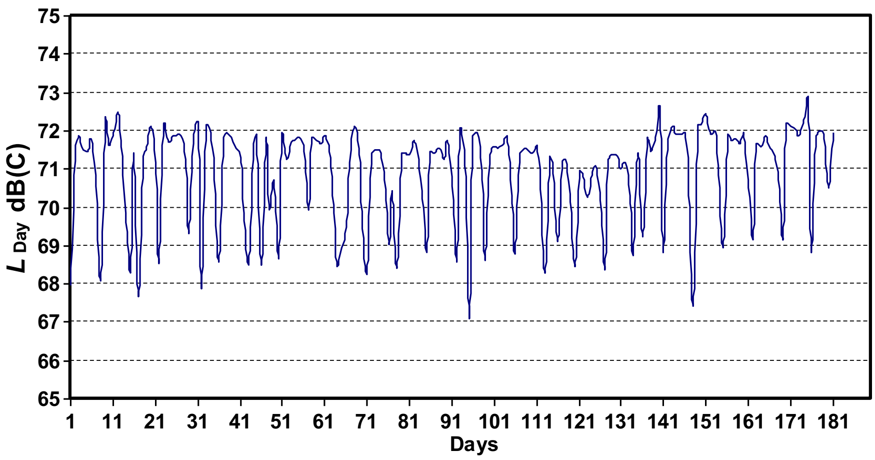

Fig. 4—Time sequence plot of L_{Day} (in dBC) for September, 2013 to February, 2014.

average terms. The diagnostic stage ascertains whether the developed ARIMA model fits well with the input time-series data or not. Once it is ensured that the difference between the predicted and actual observations is sufficiently small, then the model can be utilized for forecasting.

Firstly, for the identification of the suitable model, it is necessary to determine whether the time-series is

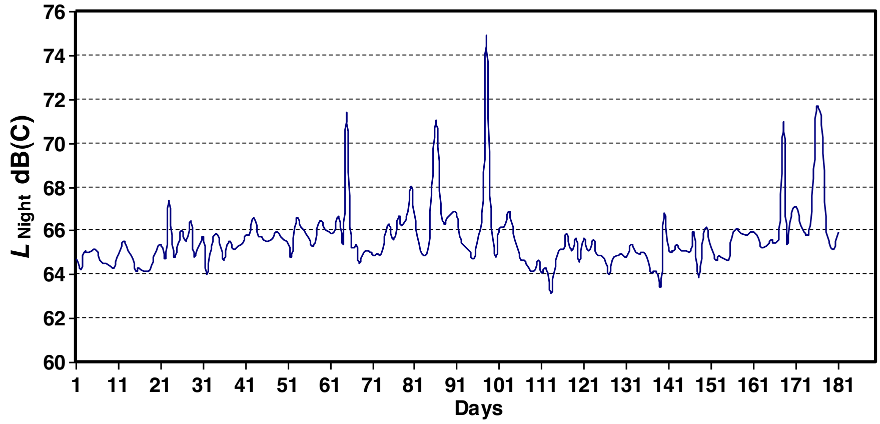

Fig. 5—Time sequence plot of L_{Night} (in dBC) for September, 2013 to February, 2014.

stationary or not. For non-stationary time-series data, a suitable differencing process is employed till the seasonality in the data disappears. Both AutoCorrelation Functions (ACFs) and Partial AutoCorrelation Functions (PACFs) are commonly used in evaluating a time-series variable's dependency on its past. The autoregressive and moving average terms of the stationary time-series are determined by examining the patterns of the graphs of ACFs and PACFs. Autocorrelation is the correlation between time-series and the same time-series lag, while partial autocorrelations are the correlation coefficients between the basic time-series and the same time series lag, but with the elimination of the influence of the members in between[33]. The number of AR and MA terms can be identified by looking at the AutoCorrelation Function (ACF), which is a bar chart of time-series of coefficients of correlation of time-series and lags of itself and Partial AutoCorrelation Function (PACF), which represents the plot of partial coefficients of correlation of time-series and lags of itself[46].

In the present study, the time-series has been analyzed using the SPSS software (Statistical Package for the Social Sciences), Version 17. Figure 1 shows the flow chart depicting the ARIMA methodology used in present study. The software package includes an expert modeler that automatically identifies and estimates the best-fitting ARIMA model for one or more dependent variable series, thus eliminating the need to identify an appropriate model through trial and error[47].

3 RESULTS AND DISCUSSION

3.1 Variations of L_{Day}, L_{Night} and Day–Night Average Sound Level (DNL)

Figure 2 shows the time sequence plot of L_{Day} in dBA for 6 months from September, 2013 to February, 2014. The average value of L_{Day} is calculated to be 64.9 $\pm$ 1.4 dBA. The monthly variation ranges from 64.4 dBA in November, 2013 to 65.30 dBA in January, 2014. The monthly variation of L_{Night} ranges from 57.7 dBA in January, 2014 to 59.1 dBA in February, 2014. The average value of L_{Night} is calculated to be 58.6 $\pm$ 1.7 dBA. Figure 3 shows the time sequence plot of L_{Night} in dBA for 6 months from September 2013 to

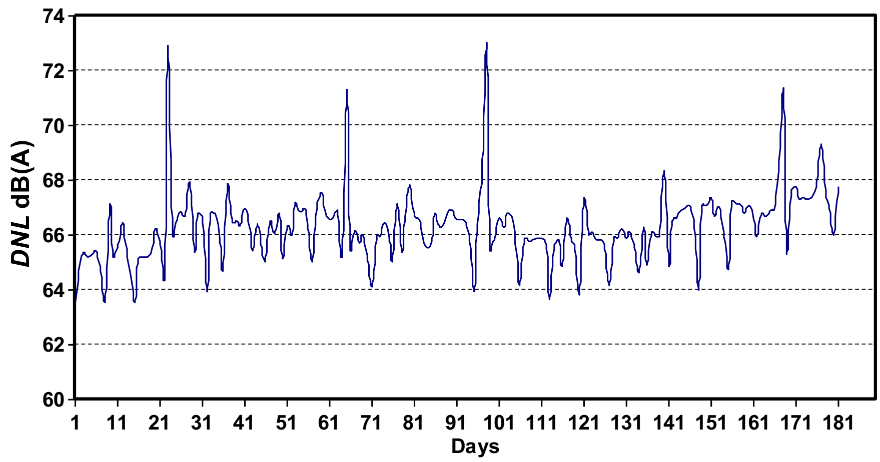

Fig. 6—Time sequence plot of day–night average sound level (DNL) (in dBA) for September, 2013 to February, 2014.

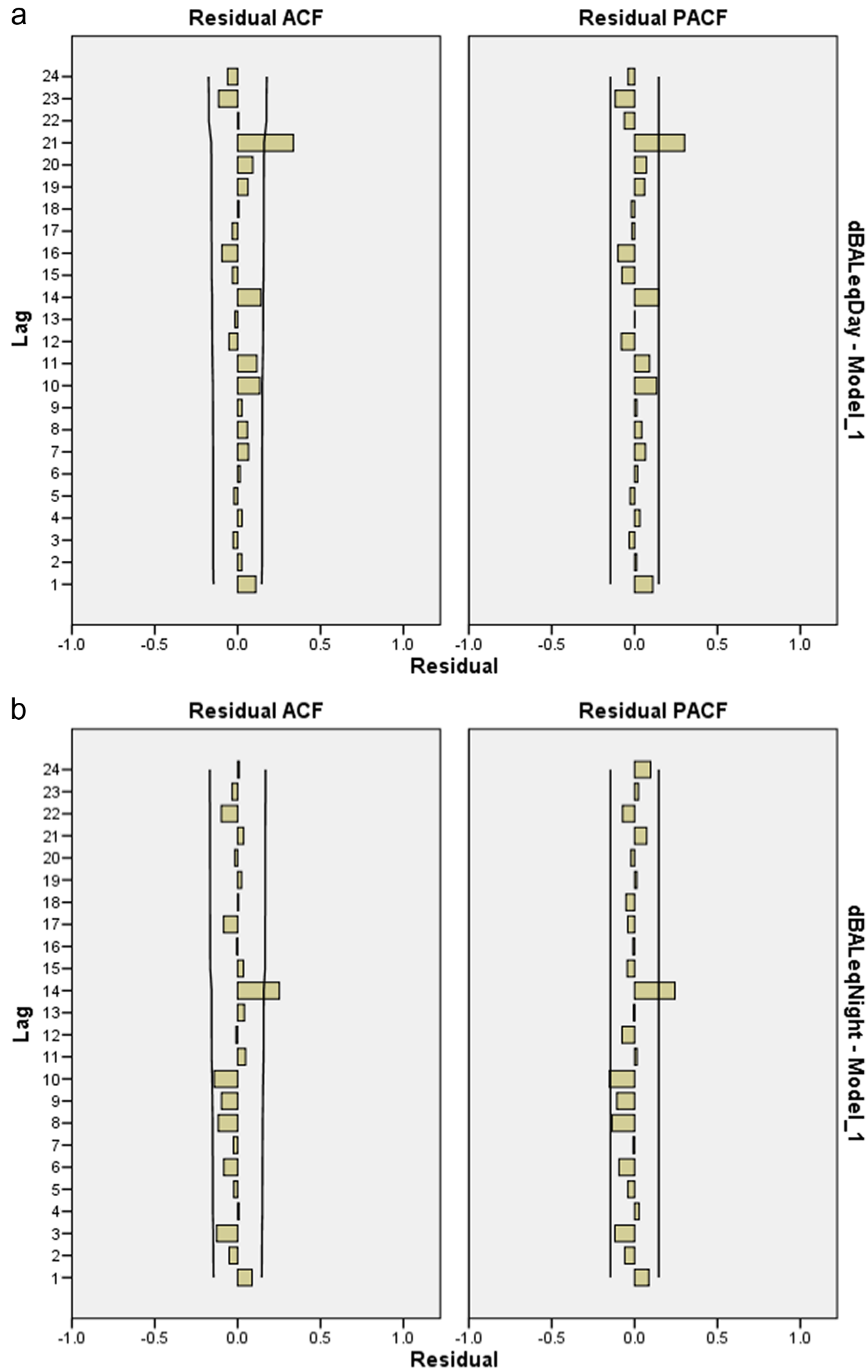

Fig. 7—(a) to (e). Autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) residuals for L_{Day} (in dBA), L_{Night} (in dBA), L_{Day} (in dBC), L_{Night} (in dBC) and day–night average sound levels (DNLs) (in dBA). The vertical lines indicate the upper and lower confidence limits.

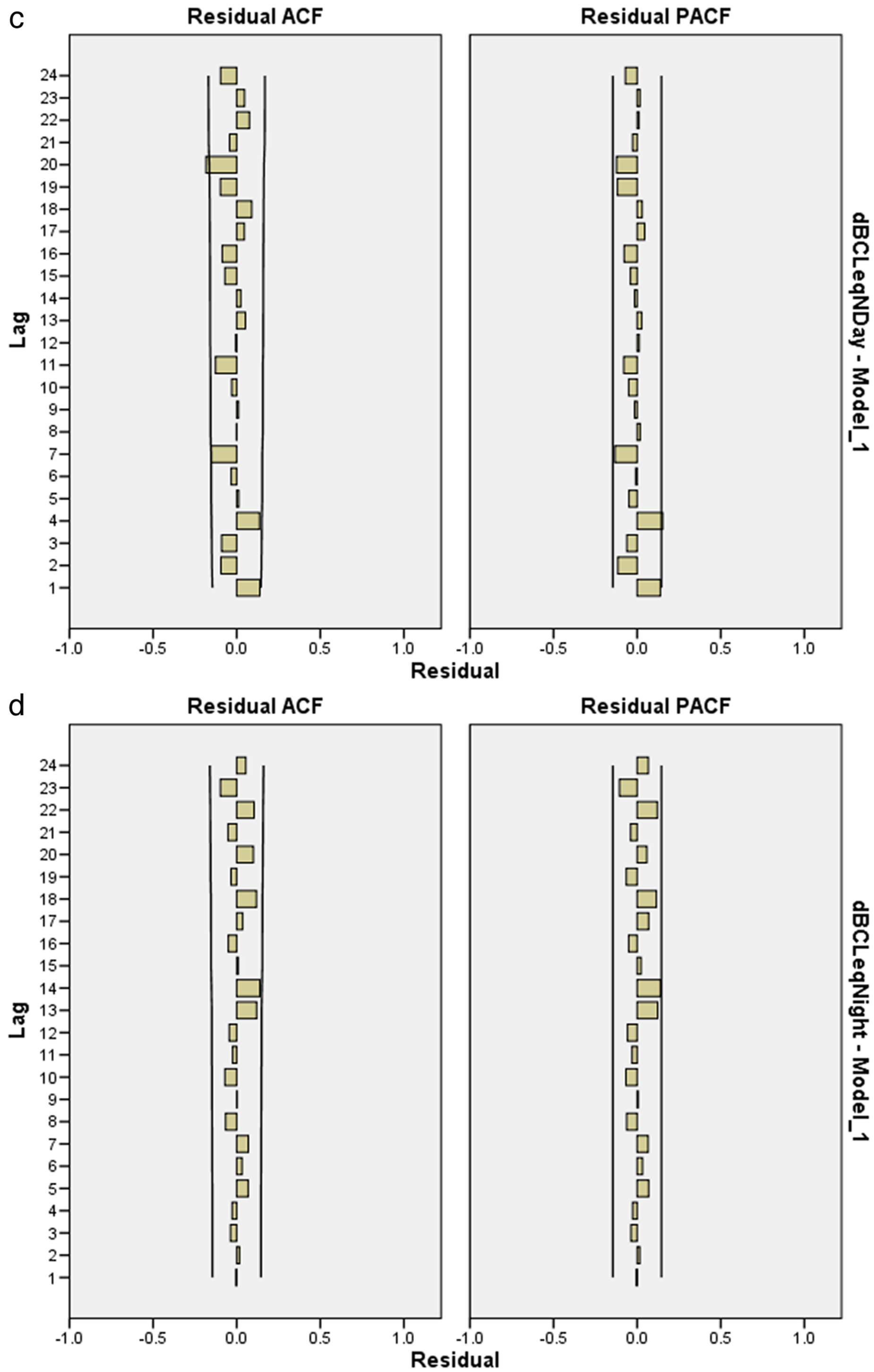

Fig. 7—Continued.

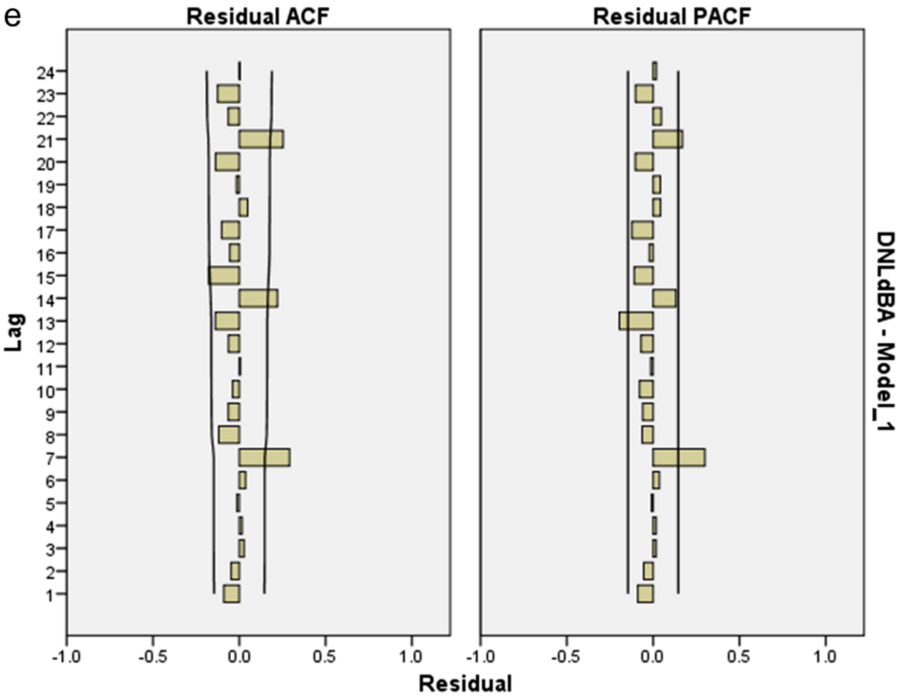

Fig. 7—Continued.

February, 2014. Figures 4 and 5 show the time sequence plot of L_{Day} and L_{Night} in dBC for a six month period. The average value of L_{Day} is observed to be 71.03 ± 1.3 dBC and L_{Night} is 65.93 ± 1.5 dBC. Figure 6 shows the time sequence plot of DNL value for a six month period. The value ranges from 66.2 dBA in September, 2013 to 67.0 dBA in February, 2014, whereby an average level of 66.5 ± 1.3 dBA is calculated. Thus, the objective of the present work is to conduct a time-series analysis via ARIMA so as to generate an ARIMA model that is most suitable for simulating and forecasting the L_{Day} and L_{Night} levels for the concerned sample site. However, the comparison of these noise levels with the recommended noise standards[38] is beyond the scope of present study.

3.2 Time-Series Analysis via ARIMA

Figures 7(a) to (e) illustrate the ACF and PACF for the daily mean time-series data of L_{Day} and L_{Night} in dBA, L_{Day} and L_{Night} in dBC and DNL. The two vertical lines in the ACF and PACF plots designate the 95% confidence intervals for the estimated autocorrelation and partial autocorrelation coefficients. The y axis of the ACF and PACF plots indicates the lag at which the autocorrelation is computed; the x axis indicates the value of the correlation (between −1 and 1). Large or frequent excursions from the bounds suggest the requirement of a model for the explanation of the dependence. Both positive and negative correlations have been observed as shown in Figs. 7(a) to (e). A positive correlation indicates that large current values correspond to large values at the specified lag and a negative correlation indicates that large current values correspond to small values at the specified lag. The confidence limits are provided to display when ACF or PACF are significantly different from zero, suggesting that the lags having values outside these limits should be considered to have significant correlation. Visual inspection of Fig. 7 shows significant deviations from

Table 1—Model statistics.

Parameter	RMSE	MAPE	Normalized BIC	Model fit statistics		Ljung–Box Q			AIC	Model type
				Stationary R^2	R^2	Statistics	df	Sig.		
L_{Day} dBA	0.859	0.982	0.096	0.641	0.641	27.658	16	0.035	−0.23	ARIMA(0,0,14)
L_{Night} dBA	0.719	0.920	−0.257	0.878	0.830	29.797	17	0.028	−0.73	ARIMA(0,1,1)
L_{Day} dB (C)	0.682	0.638	−0.422	0.736	0.736	23.553	17	0.132	−0.75	ARIMA(7,0,0)
L_{Night} dBC	0.574	0.629	−0.681	0.855	0.855	15.029	17	0.593	−1.18	ARIMA(1,0,0)
DNL dBA	0.909	1.034	0.066	0.746	0.552	46.056	13	0.000	0.11	ARIMA(0,1,14)

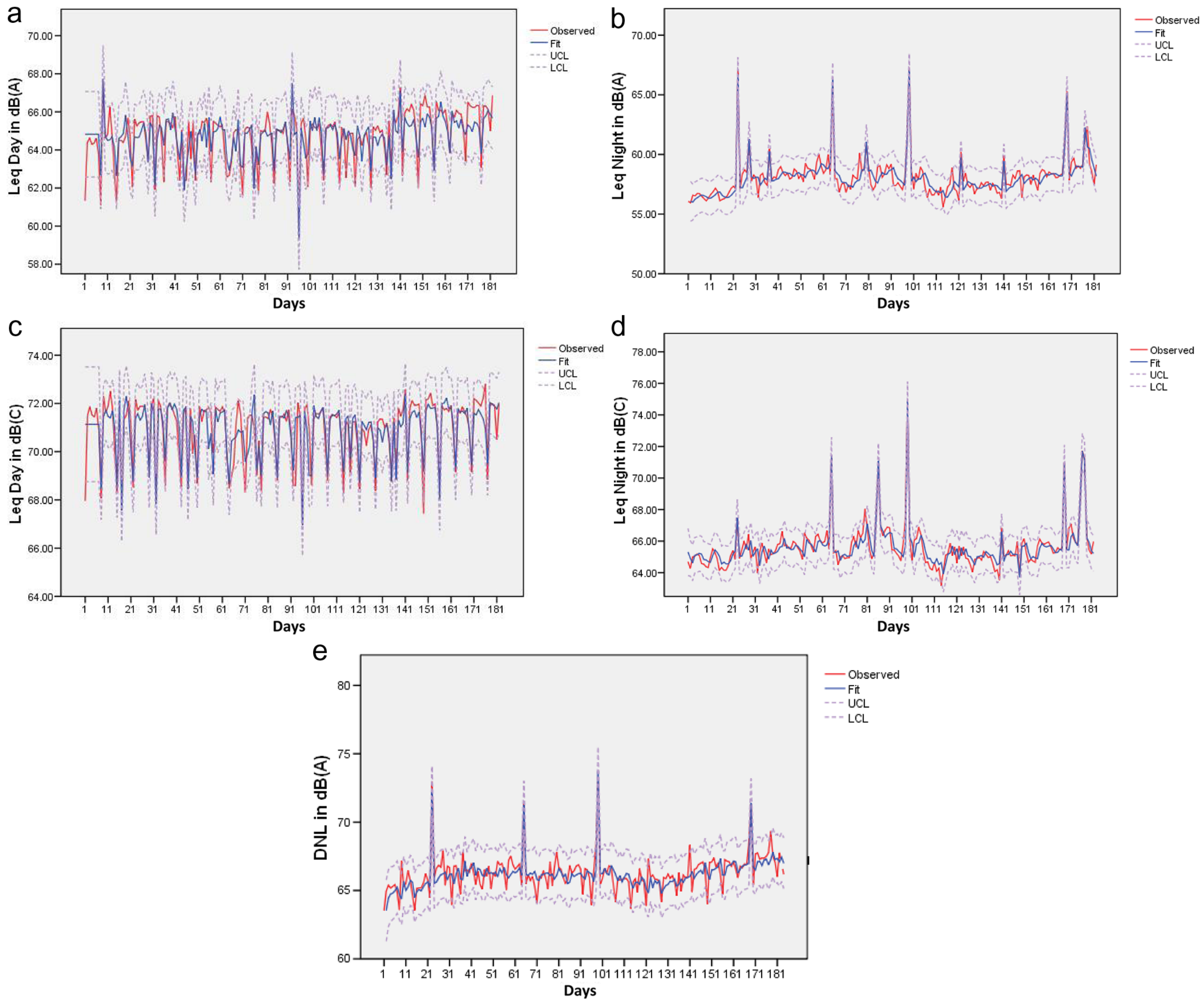

Fig. 8—(a) to (e). Comparison of ARIMA model simulations (blue line) and observations (red line) of L_{eq} Day or L_{Day} (in dBA) L_{eq} Night or L_{Night} (in dBA), L_{eq} Day or L_{Day} (in dBC), L_{eq} Night or L_{Night} (in dBC) and day–night average sound level (DNL) (in dBA) data.

zero. The interpretation suggests that the all ACF and PACF values are correlated to each other during successive days. The model results of the single-noise metrics viz., L_{Day}, L_{Night} and DNL are shown in Figs. 8(a) to (e). The ARIMA models, namely, ARIMA(0,0,14), ARIMA (0,1,1), ARIMA(7,0,0), ARIMA(1,0,0) and ARIMA (0,1,14), have been observed as the most suitable for simulating and forecasting the daily mean L_{Day} dBA, L_{Night} dBA, L_{Day} dBC, L_{Night} dBC and day–night average sound level (DNL) respectively. In order to choose the best ARIMA (p, d, q) model for each time-series, Akaike Information Criterion (AIC) was applied in the model selection procedure. For a fitted ARIMA time series of length n, the AIC is defined as[48]:

$$AIC = \ln\left(\hat{\sigma}^2_{p,q}\right) + 2(p+q)/n, \quad (7)$$

where $\hat{\sigma}^2_{p,q}$ is the residual error variance from the fitted model. AIC is an objective measure that balances the model fit and complexity. When comparing the fitted model, the basic idea is that the model with smallest AIC value is chosen[49]. Table 1 represents the AIC values corresponding to the parsimonious models developed for each of these sites. For the single-noise metrics, L_{Night} dBA and DNL, the time-series is found to be non-stationary. Thus, after the first order of differencing[41], the transformed series is stationary and observed to be governed by a moving average process of order 1 for L_{Night} (in dBA) and that of 14 for DNL (in dBA). It may be noted here that an autoregressive order of 1 specifies that the value of series one time period in the past is used to predict the current value, while the moving-average order of 14 specifies that deviations from the mean value of the series from each

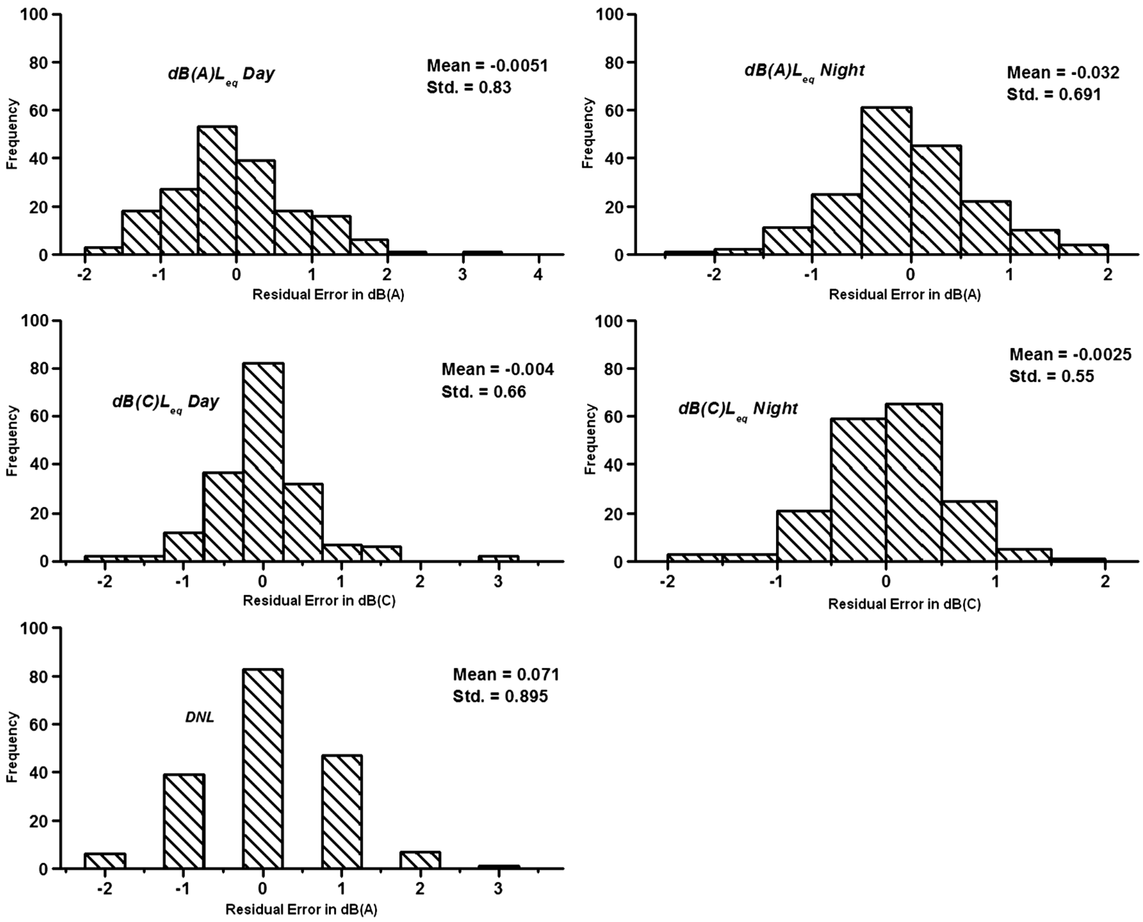

Fig. 9—Standardized residual analysis of ARIMA model for L_{Day} (in dBA), L_{Night} (in dBA), L_{Day} (in dBC), L_{Night} (in dBC) and day–night average sound level (DNL) (in dBA). X-axis denotes the residual error in dBA or dBC, and Y-axis denotes the frequency of occurrence.

of the last 14 time periods be considered when predicting the current value of the series.

These models are parsimonious among all the other possibilities, especially when no suitable model can be found with the available methodology[42,50]. These results also suggest that the model forecasted values are following the observed trend quite well. Consequently, the model gives satisfactory results and can be used as a reliable predictive tool. Several measures of accuracy were applied to ascertain the performance of the ARIMA models so developed such as stationary R^2, R^2, root mean square error (RMSE), mean absolute percentage error (MAPE) and normalized BIC (Bayesian information criterion). Table 1 represents the statistical analysis of ARIMA models developed in terms of stationary R^2, R^2, RMSE, MAPE, normalized BIC, and Ljung–Box for all the single-noise metrics. The Ljung–Box statistic, also known as the modified Box–Pierce statistic, provides an indication of whether the model is correctly specified. Using Ljung–Box model, value of statistics lies between 15 and 46, significance level varies from 0.00 to 0.59 for all the single-noise metrics, which confirms that the model is correctly specified[33]. The stationary R^2 (Table 1) is a measure that compares the stationary part of the model to a simple mean model. This measure is preferable to ordinary R^2 when there is a trend or pattern. Similarly R^2 is an estimate of the proportion of the total variation in the series that is explained by the model. The higher value of R^2 indicate a perfect prediction over the mean[36]. Normalized BIC is a general measure of the overall fit of a model and has been widely used for model identification in time-series and linear regression analysis. Negative values of BIC for L_{Night} (in dBA), L_{Night} (in dBC) and L_{Day} (in dBC) indicate a higher accuracy in the model so developed. The performance and accuracy of the model is evaluated (Table 1) in three stages: predictive capability (R^2), precision (RMSE) and goodness-of-fit (BIC). Lower values of the BIC, RMSE and high value of R^2 were preferable (Table 1). The low RMSE indicates that the dependent series is closest with model predicted levels and thus the

Table 2—Average values predicted by developed ARIMA models for various single-noise metrics (LCL denotes the lower confidence limit and UCL denotes the upper confidence limit).

Single-noise metrics		Month						
		Sept. 2013	Oct. 2013	Nov. 2013	Dec. 2013	Jan. 2014	Feb. 2014	Next 15 days, March 2014
L_{Day} dBA	Observed	64.6	64.6	64.3	64.5	65.1	65.6	65.8
	Predicted	64.7	64.7	64.6	64.7	64.8	65.1	65.0
	LCL	62.3	62.3	62.3	62.4	62.4	62.8	62.6
	UCL	67.1	67.1	67.0	67.1	67.1	67.5	67.4
L_{Night} dBA	Observed	57.5	58.4	58.6	57.9	57.6	59.1	58.6
	Predicted	57.1	58.3	57.0	58.1	57.5	58.7	59.3
	LCL	53.9	55.2	55.4	54.9	54.4	55.5	55.9
	UCL	60.3	61.5	61.8	61.3	60.7	61.8	62.6
L_{Day} dBC	Observed	71.0	71.0	70.6	70.7	71.0	71.3	71.4
	Predicted	70.9	71.0	70.9	70.9	70.9	71.1	71.1
	LCL	68.7	68.9	68.8	68.8	68.8	69.0	68.8
	UCL	73.1	73.1	73.0	73.0	73.0	73.1	73.4
L_{Night} dBC	Observed	65.0	65.6	66.3	65.5	65.0	66.4	65.8
	Predicted	65.4	65.6	65.9	65.6	65.4	65.9	65.6
	LCL	62.8	63.0	63.3	63.0	62.8	63.3	62.8
	UCL	68.0	68.2	68.5	68.2	68.0	68.6	68.5
DNL dBA	Observed	65.9	66.3	66.3	66.1	66.1	67.2	67.0
	Predicted	65.7	66.2	66.4	66.3	65.9	66.8	67.3
	LCL	63.1	63.7	63.8	63.7	63.4	64.3	64.7
	UCL	68.2	68.8	68.9	68.8	68.5	69.4	70.0

predictive model is useful at 95% confidence limits[35]. On the basis of the above discussion, it can be concluded that the model is performing satisfactory for all the single-noise metrics. The relative success of statistical models in reproducing the measured time-series can also be measured in terms of residuals of error. The frequency distributions of the residuals of the ARIMA models for all single-noise metrics are presented in Fig. 9. The study of residuals is very essential in deciding the appropriateness of the statistical model. The histogram distribution pattern (Fig. 9) displays that the residuals were, in general, distributed equally around zero approaching the Gaussian distribution, which again validates the suitability of the statistical models developed in the present study.

The ARIMA models so developed are thus utilized for forecasting the single-noise metrics for next 15 days for the month of March 2014 for the sample site in Delhi. Table 2 shows the monthly average values predicted by the developed model. The ARIMA forecasting provides results in three different options, which are i) the 95% upper confidence limit (UCL), ii) the 95% lower confidence limit (LCL) and, iii) the predicted values. However, for a larger database longer than a year (n = 181 in the present case), the ARIMA model can satisfactorily predict and forecast noise levels for longer durations e.g. for 1 month. It can be observed that the maximum error between the predicted and measured values is observed to be 0.8 dB for L_{Day} (in dBA), which confirms the validity and reliability of the developed ARIMA models.

4 CONCLUSIONS

The study focused on a statistical analysis of 181 days continuous noise data using well known ARIMA approach, covering the period from September, 2013 to February, 2014 in Delhi. In this respect, daily mean L_{Day} and L_{Night} in A- and C-weightings in conjunction with noise metrics, day–night average sound level (DNL) were used. The ARIMA models, namely, ARIMA(0,0,14), ARIMA(0,1,1), ARIMA(7,0,0), ARIMA (1,0,0) and ARIMA(0,1,14), have been developed as the most suitable for simulating and forecasting the daily mean L_{Day} dBA, L_{Night} dBA, L_{Day} dBC, L_{Night} dBC, and day–night average sound level (DNL) respectively. The validation and suitability of the developed ARIMA models are ascertained at various stages in the paper. The observed and predicted values have been found to match well. The statistical parameters stationary R^2, R^2, root mean square error (RMSE), mean absolute percentage error (MAPE) and normalized BIC

(Bayesian Information Criterion) were used to test the validity and applicability of the developed ARIMA models indicating that the models fit reasonably well with the observed data series. The Ljung–Box analysis and standardized residual error analysis further confirm the suitability and validity of the ARIMA models developed for different single-noise metrics. Furthermore, the results of the present study suggest that it is possible to predict the noise levels and single-noise metrics using statistical analysis of the present and historical time-series data sets obtained from continuous long-term noise monitoring. The ARIMA methodology demonstrated in the present work can thus serve as a suitable substitute to the long-term noise monitoring and is thus indispensable for saving costs and time incurred on continuous noise monitoring. However, the ARIMA models so developed for single-noise metrics are adequate for the particular site in the commercial area of Delhi and cannot be generalized for the other sites as well. It is obvious that for the other sites, the ARIMA model can be developed afresh with different *p*, *d*, *q* values as dependent upon the measured time-series data of noise levels. The present study thus provides a basis for further investigations on the applicability of the ARIMA methodology for serving as a prognostic model for predictions of L_{den} and L_{Aeq} in urban cities. Also, it is recommended to further investigate the suitability and reliability of ARIMA modeling for analyzing the time-series data of longer durations ranging from one to five years for long-term traffic noise predictions.

5 ACKNOWLEDGMENTS

The authors thank Kulwinder Singh Parmar, Faculty, Sant Baba Bhag Singh University, Jalandhar, India for his help and support. Authors also thank CPCB, New Delhi officials for information about the noise monitoring database in CPCB website, http://www.cpcbnoise.com.

6 REFERENCES

1. C. Steele, "A critical review of some traffic noise prediction models", *Applied Acoustics* **62**, 271–287, (2001).
2. N. Garg and S. Maji, "A critical review of principal traffic noise models: strategies and implications", *Environmental Impact Assessment Review*, **46**, 68–81, (2014).
3. European Noise Directive, "Assessment and Management of Environmental Noise", 2002/49/EU, (2002).
4. E.A. Björk, "Community noise in different seasons in Kuopio, Finland", *Applied Acoustics*, **42**, 137–150, (1994).
5. M. Ausejo, M. Recuero, C. Asensio, I. Pavón and J.M. López, "Study of precision, deviations and uncertainty in the design of the strategic noise map of the macrocenter of the city of Buenos Aires, Argentina", *Environmental Modeling and Assessment* **15**, 125–135, (2010).
6. G. La Torre, U. Moscato, F. La Torre, P. Ballini, S. Marchi and W. Ricciardi, "Environmental noise exposure and population health: a cross-sectional study in the Province of Rome", *Journal of Public Health* **15**, 339–344, (2007).
7. K-T. Tsai, M-D. Lin and Y-H. Chen, "Noise mapping in urban environments: a Taiwan study", *Applied Acoustics*, **70**, 964–972, (2009).
8. E. Murphy, E.A. King and H.J. Rice, "Estimating human exposure to transport noise in central Dublin, Ireland", *Environment International*, **35**, 298–302, (2009).
9. P.H.T. Zannin, F.B. Diniz and W.A. Barbosa, "Environmental noise pollution in the city of Curitiba, Brazil", *Applied Acoustics*, **63**, 351–358, (2002).
10. P. Rao and S. Rao, "Prediction of L_{AT} traffic noise levels in the city of Visakhapatnam, India", *Applied Acoustics*, **34**, 101–110, (1991).
11. R.S. Nirjar, S.S. Jain, M. Parida, V.S. Katiyar and N. Mittal, "A study of transport related noise pollution in Delhi", *Journal of the Institute of Engineers (I) – Environment*, **84**, 6–15, (2002).
12. H.N. Rajakumara and R.M.M. Gowda, "Road traffic noise pollution model under interrupted traffic flow condition", *Environmental Monitoring Assessment*, **14**, 251–257, (2009).
13. S. Agarwal and S.L. Swami, "Comprehensive approach for the development of traffic noise prediction model for Jaipur City", *Environmental Monitoring Assessment*, **172**, 113–120, (2011).
14. R. Kalaiselvi and A. Ramachandraiah, "A model for traffic noise prediction in heterogeneous traffic conditions", *International Journal of Current Research*, **4**, 180–184, (2012).
15. A. Sharma, G.L. Bodhe and G. Schimak, "Development of a traffic noise prediction model for an urban environment", *Noise and Health*, **16**, 63–67, (2014).
16. N. Garg, O. Sharma and S. Maji, "Noise impact assessment of mass rapid transit systems in Delhi city", *Indian Journal of Pure and Applied Physics*, **49**, 257–262, (2011).
17. B.B. Singh and V.K. Jain, "A comparative study of noise levels in some residential, industrial and commercial areas of Delhi", *Environmental Monitoring and Assessment*, **35**, 1–11, (1995).
18. D. Chakraborty, S.C. Santra, A. Mukherjee, B. Roy and P. Das, "Status of road traffic noise in Calcutta metropolis, India", *J Acoust. Soc. Am.*, **101**, 943–949, (1997).
19. S.K. Chakrabarty and D. Banerjee, "Monthly variation in night time noise levels at residential areas of Asansol city (India)", *J. Environ. Sci. Engg.*, **48**, 39–44, (2006).
20. K. Kumar, V.K. Jain and D.N. Rao, "A predictive model of noise for Delhi", *J. Acoust. Soc Am.*, **103**, 1677–1679, (1998).
21. S. Anyogita, A. Prakas and V.K. Jain, "A study of noise in CNG driven modes of transport in Delhi", *Applied Acoustics*, **65**, 195–201, (2004).
22. P. Mandal, M. Prakash and J.K. Bassin, "Impact of Diwali celebrations on urban air and noise quality in Delhi city, India", *Environmental Monitoring and Assessment*, **184**, 209–215, (2009).
23. "Pollution, Assessment, Monitoring and Survey", CPCB, India. http://www.cpcb.nic.in/divisionsofheadoffice/pams/namp.pdf
24. G. Brambilla, F.L. Castro, A. Cerniglia and P. Verardi, "Accuracy of temporal samplings of environmental noise to estimate the long-term L_{den} value", *InterNoise07*, (2007).
25. E. Gaja, A. Gimenez, S. Sancho and A. Reig, "Sampling techniques for the estimation of the annual equivalent noise level under urban traffic conditions", *Applied Acoustics*, **64**, 43–53, (2003).
26. J.M.B. Morillas and C.P. Gajardo, "Uncertainty evaluation of continuous noise sampling", *Applied Acoustics*, **75**, 27–36, (2014).
27. N. Garg, T.K. Saxena and S. Maji, "Long-term versus short-term noise monitoring: strategies and implications in India", *Noise Control Engr. J.*, **63**(1), 26–35, (2015).
28. K. Kumar and V.K. Jain, "Autoregressive integrated moving averages (ARIMA) modeling of a traffic noise time series", *Applied Acoustics*, **58**, 283–294, (1999).

29. R.E. DeVor, P.D. Schomer, W.A. Kline and R.D. Neathamer, "Development of temporal sampling strategies for monitoring noise", *J. Acoust. Soc. Am.*, **66**, 763–771, (1979).
30. P.D. Schomer, R.E. De Vor and W.A. Kline, Sampling strategies for monitoring noise in the vicinity of airports", *J. Acoust. Soc. Am.*, **73**, 2041–2050, (1983).
31. G.M. Machimbarrena, V. A. Turino, S. J. González, R. J. I. Sánchez and B. M. C. Valseros, "The seasonal nature of urban noise: a case of study", *InterNoise97*, (1997).
32. P. Goyal, A.T. Chan and N. Jaiswal, "Statistical models for the prediction of respirable suspended particulate matter in urban cities", *Atmospheric Environment.* **40**, 2068–2077, (2006).
33. K. Soni, S. Kapoor, K.S. Parmar and D.G. Kaskaoutis, "Statistical analysis of aerosols over the Indo-Gangetic plains and southern slopes of the Himalayas using ARIMA model based on long-term MODIS observations", *Atmospheric Research*, **14**, 174–192, (2004).
34. U. Kumar and V.K. Jain, "ARIMA forecasting of ambient air pollutants (O_3, NO, NO_2 and CO)", *Stoch. Environ. Res. Risk Assessment*, **24**, 751–760, (2010).
35. K.S. Parmar and R. Bharadwaj, "Water quality management using statistical analysis and time series prediction model", *Applied Water Science*, **4**, 425–434, (2014).
36. K. Soni, K.S. Parmar and S. Kapoor, "Time series model prediction and trend variability of aerosol optical depth over coal mines in India", *Journal of Environmental Science and Pollution Research*, **22**(5), 3652–3671, (2015).
37. "National Ambient Noise Monitoring Network", Information Brochure, Central Pollution Control Board, India, (2011).
38. Central Pollution Control Board, Ministry of Environment and Forests, Govt. of India, http://www.cpcbnoise.com
39. "Requirement and procedure for monitoring ambient noise level due to aircrafts", Central Pollution Control Board, (2008), http://cpcb.nic.in/divisionsofheadoffice/pci2/package_Airport%20Noise.pdf
40. A.B. Schimdt, *Quantitative Finance for Physicists*, Elsevier, Academic Press, USA, Chapter 5, 43–58, (2005).
41. G.P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", *Neurocomputing*, **50**, 159–175, (2003).
42. R. Adhikari and R.K. Agrawal, *An Introductory Study on Time Series Modeling and Forecasting*, Lambert Academic Publishing, Germany, (2013). http://arxiv.org/ftp/arxiv/papers/1302/1302.6613.pdf
43. X.H. Cai, "Time series analysis of air pollution CO in California South Coast area with seasonal ARIMA model and VAR model", Master of Science Thesis, University of California, (2008).
44. Dick Stanford Economics, "Time series forecasting tools", http://www.dickstanford.com/EFT/h8.htm
45. L.A. Díaz-Robles, J.C. Ortega, F.S. Fu, G.D. Reed, J.C. Chow, J.G. Watson and J.A. Moncada-Herrera, "A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: the case of Temuco, Chile", *Atmospheric Environment*, **42**, 8331–8340, (2008).
46. G.E.P. Box, G.M. Jenkins and G.C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th Edition, John Wiley & Sons; Inc. U.K., (2008).
47. SPSS Forecasting, Version 17 Manual. http://www.docs.is.ed.ac.uk/skills/documents/3663/SPSSForecasting17.0.pdf
48. Y. Zhai, "Time series forecasting competition among three sophisticated paradigms", Master of Science Dissertation, University of North Carolina, (2005). http://libres.uncg.edu/ir/uncw/f/zhai2005-2.pdf.
49. Y. Sakamoto, M. Ishiguro and G. Kitagawa, *Akaike Information Criterion Statistics*, D. Reidel Publishing Company, (1986).
50. A.I. Mc Leod and Parsimony, "Model Adequacy and Periodic Correlation in Time Series Forecasting", *The International Statistical Review*, **61**, 387–393, (1993).