Journal of Scientific & Industrial Research
Vol. 81, April 2022, pp. 408-415

Time Series Traffic Flow Prediction with Hyper-Parameter Optimized ARIMA Models for Intelligent Transportation System

Praveen Kumar B[1]* and Hariharan K[2]

[1]Department of Mechatronics Engineering, [2]Department of Electronics & Communication Engineering, Thiagarajar College of Engineering, Madurai 625 015, India

Received 06 June 2021; Revised 23 March 2022; Accepted 24 March 2022

Intelligent Transportation System (ITS) has become the need of the day to manage heavy traffic problems due to the exponential growth of road transportation. This is also very much essential for building the smart cities and to improve the comfort of the vehicle drivers. The electric and autonomous vehicles are going to be the future transport systems for which we need an intelligent traffic management system. This requires a lot of growth in infrastructure. The integration of technologies such as Sensors, Internet of Things (IoT), Cloud Computing, etc. has to be done for this. The traffic prediction is one of thekey requirement for establishing the ITS. In this paper we present our study on ARIMA model with optimized hyper-parameter using grid search technique for traffic flow predictions. The model validation is done on the whole day traffic flow, morning and evening peak time traffic flow datasets. The prediction results show good performance metrics with RMSE of 8.953, 11.007 and 11.837 for those three datasets.

Keywords: ARIMA, Forecast, Grid search, Road transport, Time series prediction

Introduction

With the rapid proliferation of vehicles, the cities in India are facing complex traffic management problems.[1] The transportation industry is focusing on the smart city applications in the recent years to increase the comfort of city-dwellers. This majorly includes the Intelligent Transportation System (ITS) applications such as traffic forecasting, travel time estimation, route estimation, accident zone prediction, autonomous transportations, and vehicle to infrastructure communication, etc.[2–4] The vehicle delay also affects the traffic.[5] Among them, the accurate traffic flow prediction is the most essential component of the Advanced Traffic Management Systems (ATMSs) for reducing the traffic jams and other unprecedented incidents on the road.[6] Along with increased quality of life, ITS and ATMSs also provide safety to the valuable lives. By the integration of latest innovative technologies, it is possible to make an efficient, systemized and intelligent transportation system.

With the advent of sensors, a wide variety of traffic sensors such as inductive loops, ultrasonic sensor, magnetometer, infrared sensor, microwave RADAR, image processors, Global Positioning System (GPS), RFID etc.[7] are available for transportation application. Such sensors are the sources of traffic related data such as traffic flow, speed, road occupancy, traffic video streams, vehicle position, vehicle ID and traffic movement direction.

The communication and network infrastructure is well established with the technological advancement. Also the cloud utility services opened up to a large storage capability on pay-for-use basis. These sensors are connected to the communication infrastructure through the gateway devices. Internet of Things implementation is done by connecting the sensor to the network infrastructure through gateway devices using communication protocols such as WiFi, WiFiMax, Bluetooth, Zigbee, 6LoWPAN.[8] Thus, the petabytes of data is collected and stored in the IoT cloud platform as a big data, which is used for further analytics and useful predictions.[9]

Most of these traffic data are available in the form of time series data. The traffic data thus collected from those sensors are applied to the data driven machine learning models for analysis and predicting different traffic parameters.

The time series forecasting is used to predict the near future accurately given the past historical data. The short term forecasting is mostly applied for time

*Author for Correspondence
E-mail: bpkmech@tce.edu

series forecasting problems such as the stock prices, weather prediction, electricity demand forecasting, etc. Similarly, the time series traffic flow forecast can be done with statistical, classical and modern machine learning techniques.

The statistical machine learning techniques include Auto Regression (AR), Moving Average (MA), Auto Regressive Moving Average (ARMA), Auto Regressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA) models, etc.[10] The classical machine learning techniques such as Support Vector Machine (SVM), Naïve Bayesian, K- means, Hidden Markov Model, Decision trees are used for traffic predictions.[11] The modern machine learning techniquesinclude deep neural networks, Recurrent Neural Networks (RNN), Long Short Term Model (LSTM), Gated Recurrent Unit (GRU), etc.[12] Also the multivariate time series forecasting models are implemented with the machine learning models considering multiple factors for improving prediction accuracy.[13]

In this proposed study, the ARIMA, one of the popular and powerful machine learning model with optimized hyper parameters is used to forecast the traffic flow. The ARIMA model has less computational overhead and also quite faster in real time traffic forecasting scenarios. Since traffic jam problems are very high during the morning and evening peak hours of the day in most of the cities, the forecasting is performed for different interval of time such as whole day traffic flow, morning peak time traffic flow and evening peak time traffic flow. The performance metrics are then compared.

Literature Review

Box-Jenkins autoregressive integrated moving average (ARIMA) model, one of the most successful statistical techniques is popular among the researchers for time series predictions.

Ho *et al.* use ARIMA model for predicting the performance of chiller in a Heating, ventilation, and air conditioning (HVAC) system. They use a hybrid ARIMA – Regression model by which they could achieve improved prediction accuracy for the coefficient of chiller performance system.[14]

Gocheva-Ilieva *et al.* present a SARIMA based statistical approach for forecasting the short term air pollution in urban areas considering various pollutants in the environment. They use Yeo–Johnson power transformation for data pre-processing and Bayesian information Criterion (BIC) for ARIMA model selection.[15]

Erdem *et al.* study the wind speed and its direction forecasting using ARMA model for improvingthe efficiency of wind turbine operation during energy harvesting. They also employ Vector Auto Regression (VAR) models for forecasting the tuple of wind attributes.[16]

Vu *et al.* propose a method to forecast the short-term electricity demand using auto-regressive based time varying model. The coefficients of the auto-regressive model is updated at a pre-set interval of time which improves the relationship between present and past values and thus enhances the prediction accuracy.[17] Contreras *et al.* also use ARIMA model for forecasting the short term electricity prices.[18]

The researchers also use this ARIMA based model for forecasting the road traffic related parameter such as traffic flow and traffic density. Alzyout *et al.* propose an ARIMA based prediction model for predicting the vehicles' future GPS location with good accuracy and less execution time. The proposed framework automatically selects the suitable ARIMA model considering dynamic data and the prediction deadline.[19]

Guo *et al.* present an ARIMA method for forecasting the short term hybrid electric vehicle velocity and the road gradient for enhancing the energy management and optimizing the power train control during horizons.[20]

Ma *et al.* propose a hybrid statistical and machine learning model for short term traffic forecast. They develop a hybrid model NN-ARIMA with multi-layer perceptron neural network and ARIMA for capturing the network wide traffic flows and traffic features. Then the Multidimensional Support Vector Regression (MSVR) model is used toverify the ARIMA prediction.[21]

Williams *et al.* present a vehicular flow forecast model using seasonal ARIMA process. They use Wold decomposition theorem and seasonal differencing for stationarity transformation.[22]

Ding *et al.* predict the traffic volume with Space Time Auto Regressive Integrated Moving Average (STARIMA). This model combines the historical traffic data and the spatial features of an urban road network.[23]

For optimized selection of ARIMA model parameter, researchers use grid search methods. Azari *et al.* choose optimum ARIMA model parameters p, d and q using grid search technique for bitcoin price prediction with three years dataset.[24] In another work by Azari *et al.* on cellular traffic prediction and classification with ARIMA model, they use the grid search method for selecting the parameters for the ARIMA model.[25]

Even though there are many sophisticated machine learning techniques available for the time series prediction, this work focuses on ARIMA based statistical model which is one of the traditional and also efficient technique for traffic flow predictions. Also we propose to use the grid search method for the optimal hyper-parameter selection for the ARIMA based traffic flow prediction model.

Proposed Methodology

ARIMA is a statistical technique, widely used for the time series forecasting applications. This includes Autoregressive (AR), Integration (I) and Moving Average (MA) concepts.

Auto Regressive AR (p) Models

In auto regressive model, the future value prediction of time series data is done with the past historical values. The auto regressive model depends on the aspect that the value at time step t in a time series data can be represented as a function of past t-p values.

This can be well illustrated with Eq. (1) where the present value in AR (p) model is computed with the past p values. The q is the order of the moving average model.

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots . + \Phi_p y_{t-p} + e_t \quad \ldots(1)$$

where, the value y_t is the value at time t. $y_{t-1}, y_{t-2}, \ldots y_{t-p}$ are the lag values till p time steps. $\Phi_0, \Phi_1, \ldots, \Phi_p$ are the coefficients of autoregressive function, e_t is the time varying error term.

Moving Average MA (q) Models

The moving average model does the future value prediction of time series with past forecast errors as show in Eq. (2) for an MA model. q is the order of moving average model.

$$y_t = \mu + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots . + \theta_p e_{t-q} + e_t \quad \ldots(2)$$

where, $\theta_1, \theta_2, \ldots . \theta_q$ are the coefficients of moving average function.

Integration (I)

The integration makes the time series data to be stationary by determining the differences between the raw data with its past lagged data.

ARIMA (p, d, q) Models

The ARIMA (p, d, q) model combines the auto regression, moving average and integration methods. This shall be represented as shown in the following equation.[26]

$$(1 - \Phi_1 B - \Phi_2 B^2 + \cdots . - \Phi_p B^p)(1 - B)^d y_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots . + \theta_q B^q)\, w_t \quad \ldots(3)$$

w_t is the white noise process with zero mean and variance σ^2. The back shift operator B is given by, $B^k y_t = y_{t-k}$. The ARIMA model is used for the time series traffic flow prediction because of its reduced computational complexity and it is a proven technique for the time series predictions.

The process of our methodology for the ARMIA based traffic flow prediction is shown in the Fig. 1.

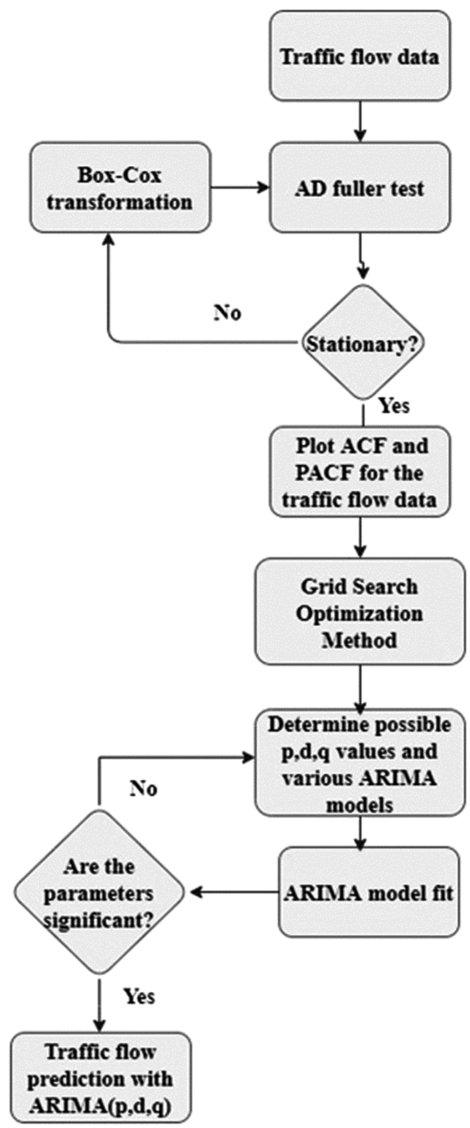

Fig.1 — Proposed ARIMA based traffic flow prediction

Augmented Dickey Fuller Test (ADF Test)

The ADF test is first done on the traffic flow dataset. The ADF test is used to verify the stationarity of the time series data. The null hypothesis is set as the traffic flow data is not stationary. Then the ADF test is performed and the p-value is estimated. When the p-value is greater than the significant value ($p>0.05$), the null hypothesis is accepted and the traffic flow data is considered as non-stationary. The transformations are then applied to the data to make it stationary with Box-cox transformation. When the p-value is less than significant value ($p<0.05$), the null hypothesis is rejected and the traffic flow data is considered to be stationary.

Once the ADF test is passed and the time series dataset is confirmed to be stationary, the autocorrelation coefficients are calculated and ACF and PACF plots are made to have an idea of selecting the ARIMA model.

ACF and PACF

The Autocorrelation Function (ACF) and partial autocorrelation function (PACF) help in choosing the parameters such as order p and q for the AR and MA models. The ACF describes the complete correlation of a present value at time stept of a time series data and its lag values at t-k. The ACF coefficients are computed using the Eq. (4). This coefficient is plotted as a bar chart called as the ACF plot.

$$r_k = \frac{\sum_{t=k+1}^{N}(y_t-\bar{y})(y_{t-k}-\bar{y})}{\sum_{t=1}^{N}(y_t-\bar{y})^2} \quad ...(4)$$

The PACF describes the partial correlation, which is the amount of correlation of residual with the lag values. It does not give the complete correlation.

Further to find an optimized hyper-parameters for the ARIMA model, we use grid search method with minimum Root Mean Square Error (RMSE). Once optimum hyper-parameters are identified, the model is built and once the parameters of the model is satisfactory we do the traffic predictions with the chosen ARIMA model.

Data Set Description

For our study, the PeMS traffic flow data set is chosen from a particular sensor station of California Transportation Agencies (CalTrans) Performance Measurement.[27] One week whole day traffic flow data, morning peak time (9.00 AM to 11.00 AM) and evening peak time(5.00 PM to 07.00PM) traffic flow data sets are considered in this study. The whole day traffic flow data is plotted in the Fig. 2, the morning peak traffic flow is plotted in Fig. 3 and the evening peak traffic flow data is plotted in Fig. 4

Time Series Components Decomposition

The time series data may have the trend T_t, seasonality S_t and noise R_t which can be decomposed with additive $Y_{at}$and multiplicative Y_{mt} models as shown in Eqs (5) & (6).

$$Y_{at} = T_t + S_t + R_t \quad ...(5)$$

$$Y_{mt} = T_t \text{ x } S_t \text{ x } R_t \quad ...(6)$$

The decomposition gives better understanding of the data. The decomposed data using additive decomposition model is shown in Fig. 5. The decomposition clearly depicts that the trend of the traffic flow repeats on each day. On each day, the

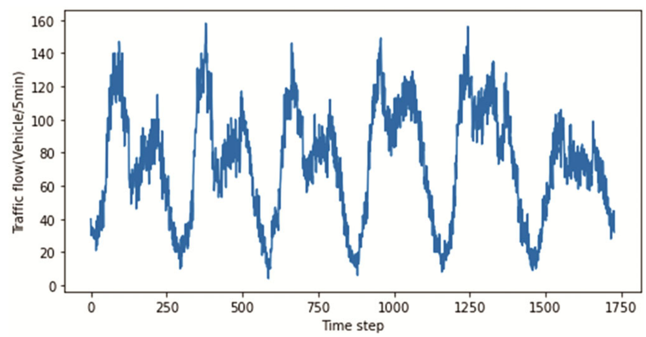

Fig. 2 — A week long whole day traffic flow

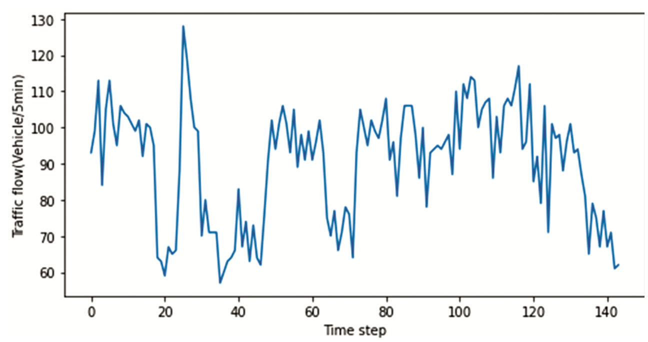

Fig. 3— Morning peak time traffic flow (9.00 A.M to 11.00 A.M)

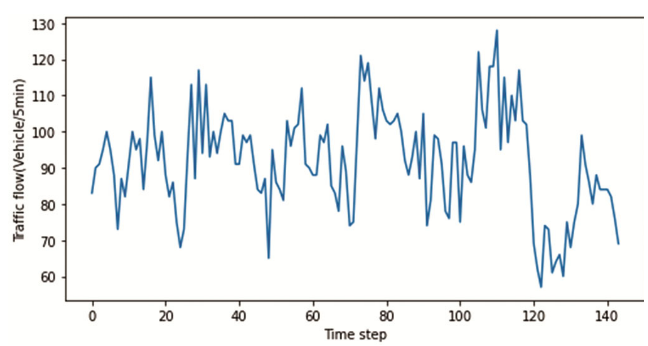

Fig. 4 — Evening peak time traffic flow (5.00 P.M to 7.00 P.M)

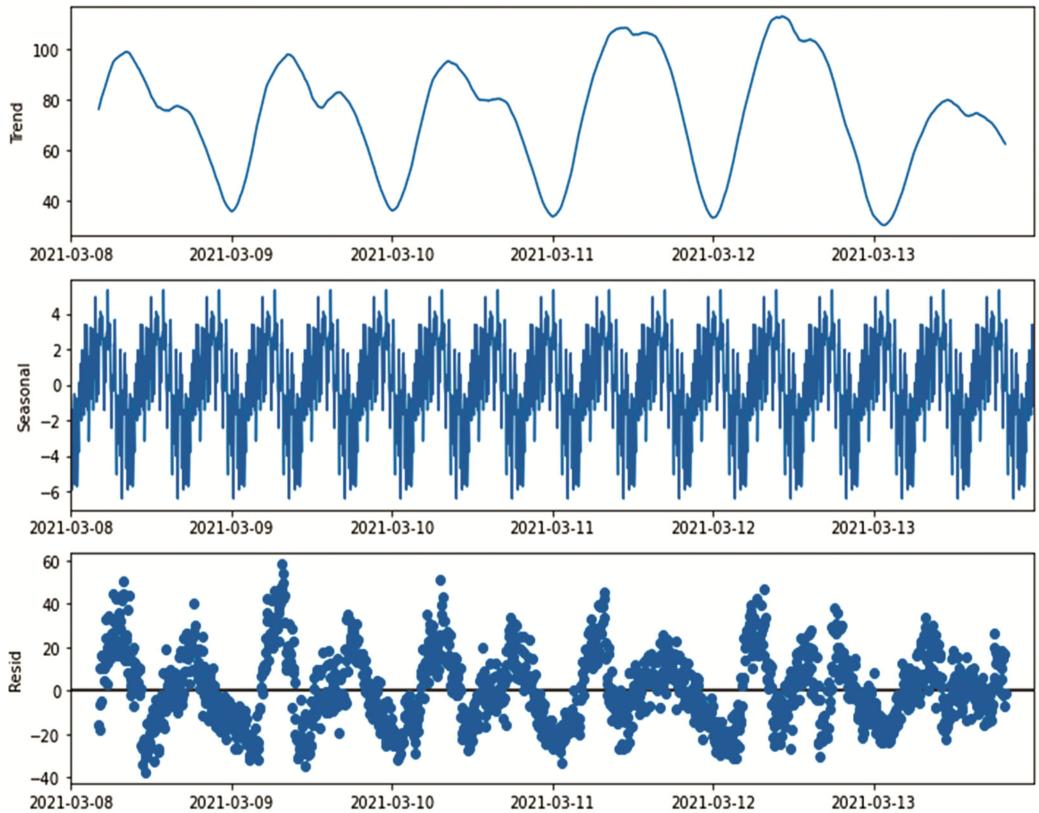

Fig. 5 — Additive decomposition of a week traffic flow data

flow is very less initially, then increases to the peak during the day and then again decreases in the night.

The seasonality also shows that the cycle repeats every day. Also the noise in the time series is plotted as residue.

Results and Discussion

The ARIMA model for traffic flow prediction is implemented with python 3.7 programming language in Jupyter Notebook. Initially, the autocorrelation and partial autocorrelation are computed for the three different datasets. The ACF and PACF plots are shown in Figs 6 & 7 for the whole day data set.

The ACF and PACF plot shows that the partial autocorrelation reduces to significant level as the number of lags is increased. This clearly indicated that the data is not stationary. To make the data stationary, differencing need to be performed.

With ACF and PACF plot alone it is difficult to choose appropriate model for our prediction. For choosing optimized values for p, d and q, the grid search optimization is required. Thus, the hyper-parameter tuning is performed with this grid search method.

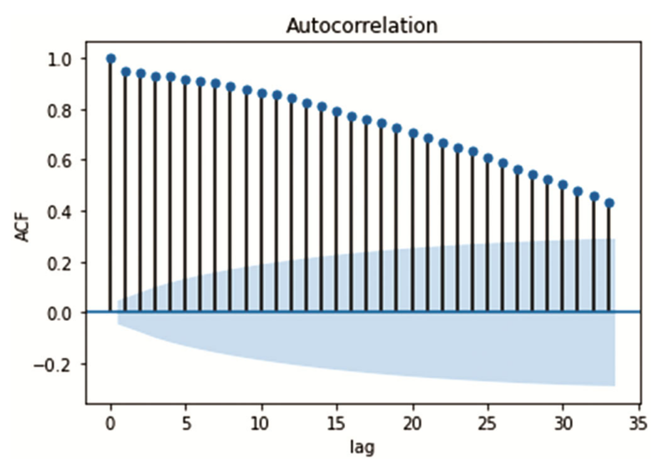

Fig. 6 — Autocorrelation of the traffic flow data

The grid search optimization is applied to the whole day as well the morning and evening peak traffic datasets. The models evaluation is done with following set of (p,d,q) values: p [1,2,3,4] , q [0,1,2] and d[0,1,2]. For different combination of model parameters, the results of Root Mean Square Error (RMSE) is shown in Table 1.

From the grid search method we shall infer that the best model for the whole day traffic flow data is

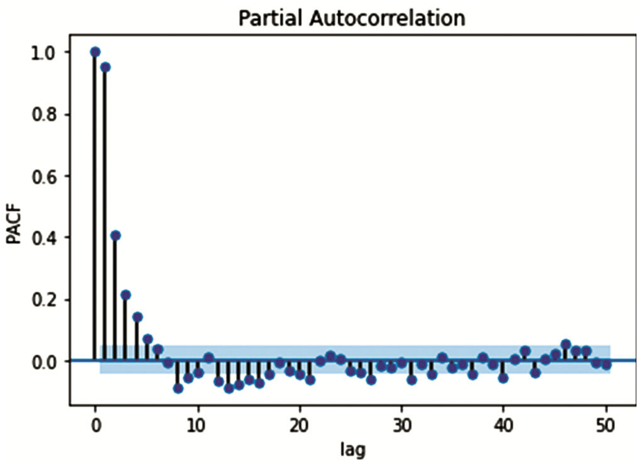

Fig. 7 — Partial autocorrelation of the traffic flow data

ARIMA (4,0,2) with RSME of 8.953. The best model for morning peak time and evening peak time traffic flow data are ARIMA (4,1,0) with RMSE of 11.007 and ARIMA (4,1,0) with RMSE of 11.837 respectively. Hence, the ARIMA (4,0,2) based prediction modelis chosen for training and validation of the whole day traffic flow dataset. Similarly, the ARIMA (4,1,0) based prediction model is chosen for training and validation of peak hours traffic flow dataset.

The data set is split into training and test data sets for the whole day, morning and evening peak traffic data sets. The ARIMA models identified by the grid search parameter optimization method are validated with these data sets. The prediction result is plotted in Fig. 8 in comparison with the ground truth of the original test data of whole day traffic flow dataset.

Similarly, the prediction result and the original test data for morning peak time traffic flow are plotted in the Fig. 9. The prediction done with ARIMA (4,1,0) is closer to the actual traffic flow data.

In the same way, the prediction result and the original test data for evening peak time traffic flow are plotted in the Fig. 10. The prediction of traffic flow using ARIMA (4,1,0) model is closer to the ground truth.

The model performance is evaluated with the metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The measure of standard deviation of the prediction errors is RMSE which is the difference between the predicted traffic flow (TF) and true value ($\widehat{TF}$). The absolute difference between the predicted and true values is MAE. Mean percentage of the absolute errors of the prediction is MAPE. These metrics are computed with Eqs (7) – (9).

Table 1 — RMSE values for various ARIMA models from grid search method

ARIMA Model	RMSE Whole day	RMSE Morning peak	RMSE Evening peak
ARIMA(1,0,0)	10.551	12.139	12.683
ARIMA(1,0,1)	9.057	11.350	12.227
ARIMA(1,0,2)	9.058	11.058	12.437
ARIMA(1,1,0)	9.478	11.057	11.867
ARIMA(1,1,1)	9.073	11.160	12.423
ARIMA(1,1,2)	9.021	NA	NA
ARIMA(1,2,0)	13.693	15.039	15.763
ARIMA(1,2,1)	NA	11.178	NA
ARIMA(1,2,2)	NA	11.331	NA
ARIMA(2,0,0)	9.434	11.106	12.186
ARIMA(2,0,1)	9.058	11.223	12.211
ARIMA(2,0,2)	8.973	11.030	NA
ARIMA(2,1,0)	9.177	11.113	11.902
ARIMA(2,1,1)	9.072	11.153	12.270
ARIMA(2,1,2)	9.008	11.297	NA
ARIMA(2,2,0)	11.915	13.908	15.133
ARIMA(2,2,1)	NA	11.227	NA
ARIMA(2,2,2)	NA	11.224	NA
ARIMA(3,0,0)	9.152	11.126	12.352
ARIMA(3,0,1)	9.057	11.206	12.270
ARIMA(3,0,2)	8.958	11.595	12.147
ARIMA(3,1,0)	9.113	11.112	11.679
ARIMA(3,1,1)	9.083	11.219	11.988
ARIMA(3,1,2)	8.992	12.186	NA
ARIMA(3,2,0)	11.084	12.359	13.279
ARIMA(3,2,1)	NA	11.227	11.806
ARIMA(3,2,2)	NA	11.341	12.420
ARIMA(4,0,0)	9.093	11.207	12.187
ARIMA(4,0,1)	9.086	11.358	12.183
ARIMA(4,0,2)	8.953	11.866	NA
ARIMA(4,1,0)	9.086	11.007	11.837
ARIMA(4,1,1)	NA	11.208	12.163
ARIMA(4,1,2)	9.103	11.061	12.494
ARIMA(4,2,0)	10.567	12.266	13.017
ARIMA(4,2,1)	NA	11.117	11.948
ARIMA(4,2,2)	NA	11.120	12.168

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(TF_i - \widehat{TF_i})^2} \quad \ldots(7)$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|TF_i - \widehat{TF_i}| \quad \ldots(8)$$

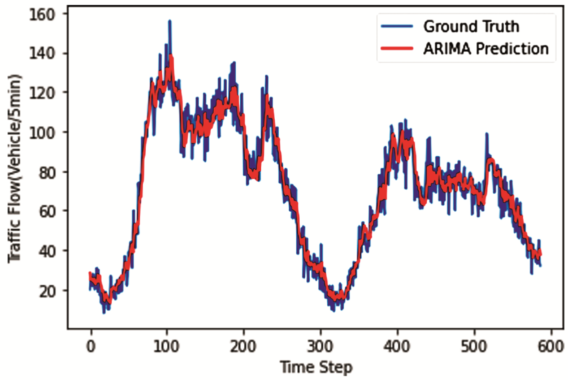

Fig. 8 — ARIMA (4,0,2) traffic flow prediction for whole day traffic flow dataset

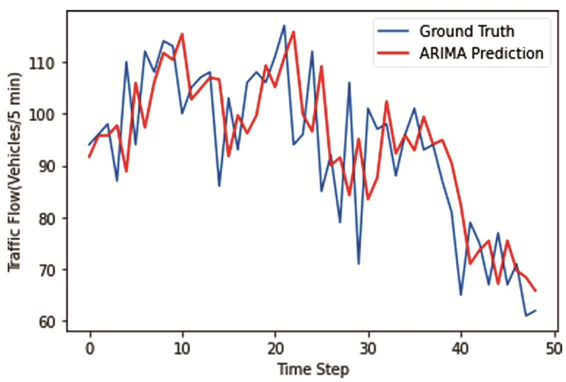

Fig. 9 — ARIMA (4,1,0) traffic flow prediction for morning peak time traffic flow dataset

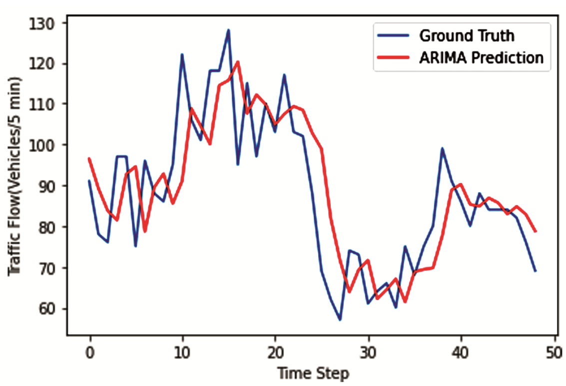

Fig. 10 — ARIMA (4,1,0) traffic flow prediction for evening peak time traffic flow dataset

$$\text{MAPE}=\frac{100}{N}\sum_{i=1}^{N}\frac{(|TF_i-\widehat{TF_i}|)}{|TF_i|} \quad \ldots(9)$$

The RMSE, MAE and MAPE for the prediction results are shown in the Table 2.

The prediction results indicate that the chosen models show comparatively better performance in traffic flow prediction. The RMSE on the prediction of test datasets are 8.953, 11.007 and 11.837 for the three datasets- whole day, morning peak time and evening peak time traffic flow respectively. Similarly the MAE are 7.008, 8.620 and 9.137 for the prediction results. And the MAPE values are 16.564%, 18.367% and 22.785%.

Table 2 — Performance metrics for the ARIMA based traffic flow prediction

Datasets	RMSE	MAE	MAPE (%)
Whole day ARIMA (4,0,2)	8.953	7.008	16.564
Morning peak time, ARIMA(4,1,0)	11.007	8.620	18.367
Evening peak time, ARIMA(4,1,0)	11.837	9.137	22.785

Conclusions

The traffic prediction is one the most essential part of the Intelligent Transportation System. The ARIMA based statistical method is used in the study, since it is easier to implement and it is a well-established technique for time series predictions. Since the traffic flow is peak during morning and evening times, the prediction is done for the whole day, morning and evening time traffic flow datasets separately and the results are compared. The optimal hyper-parameters of the ARIMA model are chosen with the grid search method. The traffic flow prediction is done with ARIMA (4,0,2) model for the whole day dataset and ARIMA (4,1,0) for morning and evening traffic flow datasets. These two models show good performance metrics with RMSE of 8.953, 11.007 and 11.837 for the three datasets respectively.

References

1 Xie P, Li T, Liu J, Du S, Yang X & Zhang J, Urban flow prediction from spatiotemporal data using machine learning: A survey, *Inf Fusion*, **59** (2020) 1–12.

2 Lana I, Del Ser J, Velez M & Vlahogianni E I, Road traffic forecasting: Recent advances and new challenges, *IEEE Intell Transp Syst Mag*, **10** (2018) 93–109.

3 George S & Santra A K, Traffic prediction using multifaceted techniques : A survey, *Wireless Pers Commun*, **115** (2020) 1047–1106

4 Singh K & Singh Tomar D D, Architecture, enabling technologies, security and privacy, and applications of internet of things: A survey, *Proc Int Conf on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, **4** (2019) 642–646.

5 Murat Y S, Modeling vehicle delays at signalized junctions : Artificial neural networks approach, *Adv traffic Manag Syst*, **65** (2006) 558–564.

6 Zhang Z, Li M, Lin X, Wang Y & He F, Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies, *Transp Res Part C Emerg Technol*, **105** (2019) 297–322.

7 Nagy A M & Simon V, Survey on traffic prediction in smart cities, *Pervasive Mob Comput*, **50** (2018) 148–163.

8 Sadowski S & Spachos P, Wireless technologies for smart agricultural monitoring using internet of things devices with

energy harvesting capabilities, *Comput Electron Agric*, **172** (2020) 1–8.

9 Celesti A, Galletta A, Carnevale L, Fazio M, Lay-Ekuakille A & Villari M, An IoT cloud system for traffic monitoring and vehicular accidents prevention based on mobile sensor data processing, *IEEE Sens J*, **18** (2018) 4795–4802.

10 Karlaftis M G & Vlahogianni E I, Statistical methods versus neural networks in transportation research: Differences, similarities and some insights, *Transp Res Part C Emerg Technol*, **19** (2011) 387–399.

11 Zhou F, Yang Q, Zhang K, Trajcevski G, Zhong T & Khokhar A, Reinforced spatiotemporal attentive graph neural networks for traffic forecasting, *IEEE Internet Things J*, **7**(2020) 6414–6428.

12 Muhammad A N, Aseere A M, Chiroma H, Shah H, Gital A Y & Hashem I A T, Deep learning application in smart cities: recent development, taxonomy, challenges and research prospects, *Neural Comput Appl*, **33** (2021) 2973–3009.

13 Praveen Kumar B & Hariharan K, Multivariate time series traffic forecast with long short term memory based deep learning model, *Proc IEEE Int Conf Power, Instrum Control Comput*, **1** (2020) 1–5.

14 Ho W T & Yu F W, Predicting chiller system performance using ARIMA-regression models, *J Build Eng*, **33** (2021) 1–9.

15 Gocheva-Ilieva S G, Ivanov A V, Voynikova D S & Boyadzhiev D T, Time series analysis and forecasting for air pollution in small urban area: An SARIMA and factor analysis approach, *Stoch Environ Res Risk Assess*, **28** (2014) 1045–1060, 2014.

16 Erdem E & Shi J, ARMA based approaches for forecasting the tuple of wind speed and direction, *Appl Energy*, **88** (2011) 1405–1414.

17 Vu D H, Muttaqi K M, Agalgaonkar A P & Bouzerdoum A, Short-term electricity demand forecasting using autoregressive based time varying model incorporating representative data adjustment, *Appl Energy*, **205** (2017) 790–801.

18 Contreras J, Espínola R, Nogales F J & Conejo A J, ARIMA models to predict next-day electricity prices, *IEEE Trans Power Syst*, **18** (2003) 1014–1020.

19 Alzyout M S & Alsmirat M A, Performance of design options of automated ARIMA model construction for dynamic vehicle GPS location prediction, *Simul Model Pract Theory*, **104** (2020) 102148.

20 Guo J, He H & Sun C, ARIMA-based road gradient and vehicle velocity prediction for hybrid electric vehicle energy management, *IEEE Trans Veh Technol*, **68** (2019) 5309–5320.

21 Ma T, Antoniou C & Toledo T, Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast, *Transp Res Part C Emerg Technol*, **111** (2020) 352–372.

22 Williams B M & Hoel L A, Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results, *J Transp Eng*, **129** (2003) 664–672.

23 Ding Q, Wang X, Zhang X & Sun Z, Forecasting traffic volume with space-time ARIMA model, *Adv Mater Res*, **156** (2011) 979–983.

24 Azari A, Bitcoin price prediction : An ARIMA approach, *Soc Inf Networks*, **1** (2019) 1–3.

25 Azari A, Papapetrou P, Denic S & Peters G, Comparative evaluation of LSTM and ARIMA, *Lect Notes Comput Sci*, **11828** (2019) 1–16.

26 Ding C, Duan J, Zhang Y, Wu X & Yu G, Using an ARIMA-GARCH Modeling approach to improve subway short-term ridership forecasting accounting for dynamic volatility, *IEEE Trans Intell Transp Syst*, **19** (2018) 1054–1064.

27 http:// pems.dot.ca.gov/