Short-Term Traffic Flow Prediction for an Urban Highways using Time Series Forecasting Model

Dwijesh Shah[1], Paramsmit Sanghani[2] and Mrugendrasinh Rahevar[3]
[1] CSPIT, CHARUSAT University, Anand GJ 388421, India
[2] CSPIT, CHARUSAT University, Anand GJ 388421, India
[3] CSPIT, CHARUSAT University, Anand GJ 388421, India

Abstract:- Due to a significant increase in the number of automobiles, traffic congestion has become a serious issue in recent years. This paper discusses various techniques for forecasting traffic flow to resolve the issue of traffic congestion. To begin, we will demonstrate how time series can be applied in this field. Second, we will attempt to describe which time-series models will be most beneficial in resolving the most pressing issue. Following that, we'll compare the results obtained using various methods using accuracy parameters. Additionally, we observe that the ARIMA time series forecasting method is incapable of producing appropriate results due to the seasonality observed in the data. We discovered in this research paper that the SARIMA time series forecasting method produces more accurate results when forecasting traffic flow at 15-minute and 30-minute intervals. Additionally, we discovered that for short time intervals, i.e., one minute, FBProphet outperforms SARIMA.

Keywords:- *Traffic time-series, ARIMA, SARIMA, Facebook Prophet, Traffic Prediction.*

I. INTRODUCTION

Congested traffic is a significant threat to any country's growth. In developing countries such as India, the rapid growth of the population and motorized vehicles is the primary cause of traffic congestion. Traffic congestion costs commuters valuable work hours, which has a knock-on effect on the country's economy as a whole.

Accessibility and mobility are also hampered by traffic congestion. Congestion lengthens travel times and raises fuel prices, putting businesses and employees who distribute products and services at a disadvantage. Every day, millions of people all around the world are directly impacted by traffic congestion. Traffic congestion also causes severe air pollution and noise pollution, aggravating the entire environment. The cost of fuel, the expense of transportation, health-related issues, and environmental issues can all be attributed to traffic congestion. Qingyu et al. statically calculate external costs of traffic congestion pricing in cities. [1]

With a total length of 5.5 million kilometers, India boasts one of the world's greatest road networks. Roads in India transport more than 90% of the country's passenger traffic. This road system transports 64.5 percent of the country's commodities. [2] The figures offered show how crucial it is to fix this problem. As the connectivity between cities, towns, and villages has improved through time, road transportation has grown in popularity. With the increase in vehicle traffic, there is an increased need for effective and advanced traffic management solutions. Existing traffic management systems are based on real-time data and include pre-timed, vehicle-actuated, and adaptive traffic control systems.

The primary disadvantage of all of the solutions listed above is that they are all real-time, which means they are completely reliant on the cameras and traffic controllers installed at the crossroads. If any of the utilities fail to function properly due to a hardware failure or other malfunction, it will be unable to provide an effective solution for traffic management. As a result, a need exists for 'Predictive Analytics for Traffic Flow' that is based on historical data, i.e., a more advanced solution to the issue. The Government of India's initiative 'Smart City Missions' has also requested a resolution to this critical issue.

Time series are data and information that have been accumulated over time and can be used to forecast data in the near future. These data are gathered on a regular and detailed basis. The stock market, specifically in forecasting stock values, is one of the most extensively utilized uses of time series. Forecasting renewable energy is another time series application.

In numerous studies, time series are used to forecast traffic intensity and incoming traffic. Because traffic prediction is such an important aspect of ITS (Intelligent Transportation Systems), or Intelligent Transportation Systems. It uses time series to forecast data or information based on previously gathered data at regular intervals, with the goal of improving traffic conditions. The majority of Intelligent Transportation Systems (ITS) research and deployments rely heavily on the study of traffic time series.[3]

The remaining sections of this paper are organized as follows; Section 2 shows the literature review about various traffic flow forecasting methods proposed to resolve this issue. Section 3 introduces a time-series analysis of traffic data, and subsequently time-series techniques are used to define short-term traffic flow forecast models. Section 4 examines and discusses the predicting outcomes of various time-series models for short-term traffic flow. This study comes to a close with Section 5.

II. LITERATURE REVIEW

This section describes the literature on short-term traffic forecasting methods and the techniques presented in this paper.

Traffic flow, or the number of vehicles passing through a specific point in a given time period, is a point process, or a type of random process made up of a collection of isolated points accumulated over time. [4] When it comes to modelling such point processes, data-driven approaches that are based on, to find the stochasticity in observable data, statistical approaches are commonly used. [5] The two types of statistical techniques utilized to solve the problem of traffic flow prediction are non-parametric statistical techniques and parametric statistical techniques. [6] Nonparametric approaches such as nonparametric regression [7] and neural networks are examples of nonparametric techniques. [8-19] Linear and nonlinear regression, historical average methods [9], smoothing techniques [9,14,20], and autoregressive linear processes [6,10,14,20-29] are among the parametric techniques.

When compared to other accessible methodologies, time series analysis-based techniques such as the ARIMA (Auto Regressive Moving Average) have been proven to be one of the most precise methods for predicting traffic flow. [30] By dissecting trends and seasonal patterns and extending the pattern into the future, time series models attempt to find patterns in historical data. Because the traffic flow pattern demonstrates a strong seasonal pattern for specific hours, such as peak and off-peak traffic conditions, which repeats at roughly the same time every day. Models like SARIMA (Seasonal Auto Regressive Integrated Moving Average) are particularly useful for simulating traffic flow [6,26,27,29,30]. Many investigations have discovered that the SARIMA model outperforms models based on a random walk and simple ARIMA. [26,27,29,31] Smith et al. [32] showed that the best-performing non-parametric forecasting method is the k-Nearest Neighbor forecast model, SARIMA, a parametric forecasting method, failed to match its predictive performance.

The research done for flow prediction using SARIMA has one major flaw: they developed the model using huge datasets. Smith et al. [32], for example, used 45 days of data taken at 15-minute intervals to forecast traffic flow for the next day. Williams and Hoel [26] employed more than two months of traffic flow measurements, with over 60,000 traffic flow observations collected every three minutes. Stathopoulos and Karlaftis selected intervals that spanned a total of 106 days. [33] Ghoesh et al. [23] utilized 20 days of traffic flow data taken at the interval of 15 min. for model development. Furthermore, Mai et al. [30] utilized previous 26 days of traffic flow observations taken at the interval of 15 min. for fitting the SARIMA model.

Dong et al. [28] uses the previous two months of flow data aggregated to 5-minute intervals to train the ARIMA model for projecting traffic flow for the next day. Lippi et al. [29] used the SARIMA technique to create a model based on the preceding four months of traffic flow observations from loop detectors in nine California districts. Tan et al. [14] trained the ARIMA forecasting model using time-series data from traffic observations collected over several years. The use of such a large database for model building may limit its applicability in situations where data availability is an issue. At times, storing and maintaining historical databases can be a difficult task. As a result, it would be ideal if time series forecasting models for predicting traffic flow could be developed, which would require only limited input data for model development. FbProphet, [34] a Facebook-developed universal time series model released in 2017, is an open-source algorithm. A time-series prediction problem was solved using a curve fitting approach. When modelling data with piecewise trends and multi-cycle characteristics, FbProphet has a good effect; it's good for forecasting short-term traffic flow and has a lot of space for improvement. Kumar et. al. [35] have researched anticipating traffic flow using SARIMA and FbProphet.

The scope of the research is limited to applying various time-series forecasting methods to anticipate the traffic flow for the next upcoming day for various spans of intervals. This research is performed on the data obtained from traffic cameras located on urban highways at Gandhinagar, Gujarat.

III. TIMESERIES METHODS TO FORECAST SHORT-TERM TRAFFIC FLOW

A. *Time Series Analysis*

A time series is a collection of statistical observations that are organized chronologically. Every time series inherits various specific patterns that are Trend, Seasonality / Cyclic, Residual / Error, which can be shown by decomposition mechanism. The trend of time series is the long-term pattern. Seasonality can be described as repetitive patterns at particular fixed intervals. Cyclicity is represented as repetitive patterns at irregular intervals. Residual / Error is the unnecessary noise.

Time series can be expressed either additively or multiplicatively or the combination of both additive and multiplicative of trend, seasonality, and residual. For, our research we have used an additive model as the magnitude of seasonal data is independent of the magnitude of data. [36-37]

Additive: Observed = Trend + Seasonality + Cycle + Residual (1)

Time series decomposition of the traffic data used in the study can be represented as below, (see Fig.1.)

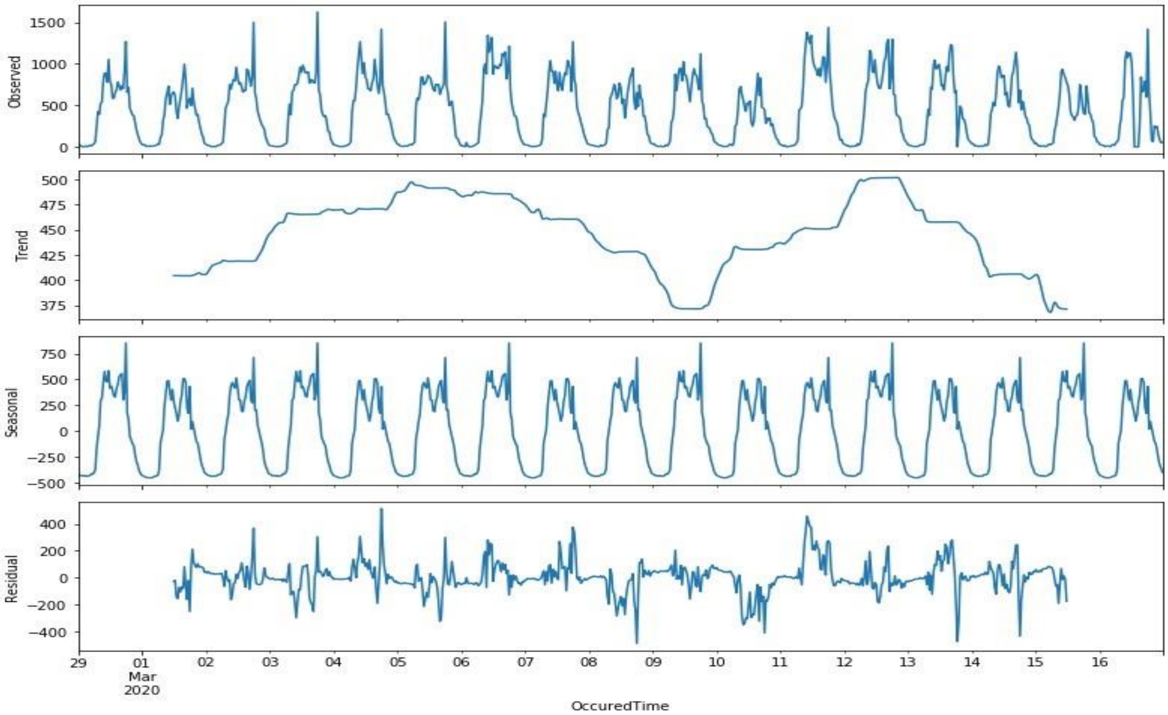

Fig. 1. Traffic Flow Time Series Decomposition

For this study, we anticipated traffic for the following day at various intervals of time, such as 5 minutes, 10 minutes, 15 minutes, 30 minutes, and 1 hour. As the forecasting is going to be done for short intervals of time it is known as short-term traffic forecasting..

Traffic time series data gathered from the camera consists of various fields as described below,

Table 1. Traffic dataset fields

OccurredTime	Time in UTC format at which data fetched from the camera.
OccuredDate	Date for which traffic data is collected.
Day	It represents the weekday name.
Event	It is the name of the junction where the camera is installed.
CameraIp	Unique identification of camera using Ip address.
Arm	It is represented as a number 1,2,3 and 4 for 4 cross-roads where the camera is installed.
VehicleType	Type of vehicle i.e., two-wheeler, three-wheeler, four-wheeler, and longvehicle.
VehicleCount	Count of each type of vehicle detected by the camera for each second.
Speed	Speed detected by the camera for each vehicle while crossing theroad.

As per the research, we have a scope limited to forecast only inflow traffic so, we can consider only OccuredTime and VehicleCount by ignoring other parameters as they have minimal effect on the result.

B. ARIMA Model

ARIMA stands for Auto Regressive Integrated Moving Average which consists of AR (Auto Regressive) and MA (Moving Average) models. AR of ARIMA consists of the weighted sum of lagged values of the time series while MA consists weighted sum of lagged predicted errors of the series and I (Integrated) contain the difference value of the time series.

ARIMA can be expressed as the ARIMA (p, d, q) model where, p is the order of AR, q is the order of MA and d is the corresponding degree of first differencing involved.

The ARIMA model has been widely and successfully applied to resolve different problems for univariate time-series data having trend and without seasonality. ARIMA models are applied to different systems such as social, economic, financial, industrial, agricultural, and many more systems to perform accurate forecasting.

In this paper, the ARIMA model is used with hyper tuned parameters of p, d, and q for the traffic flow forecasting of 1 hour time interval, and the result of forecasting can be shown as below (see Fig. 2.), but due to the limitation of the seasonality factor the model doesn't provide proper results.

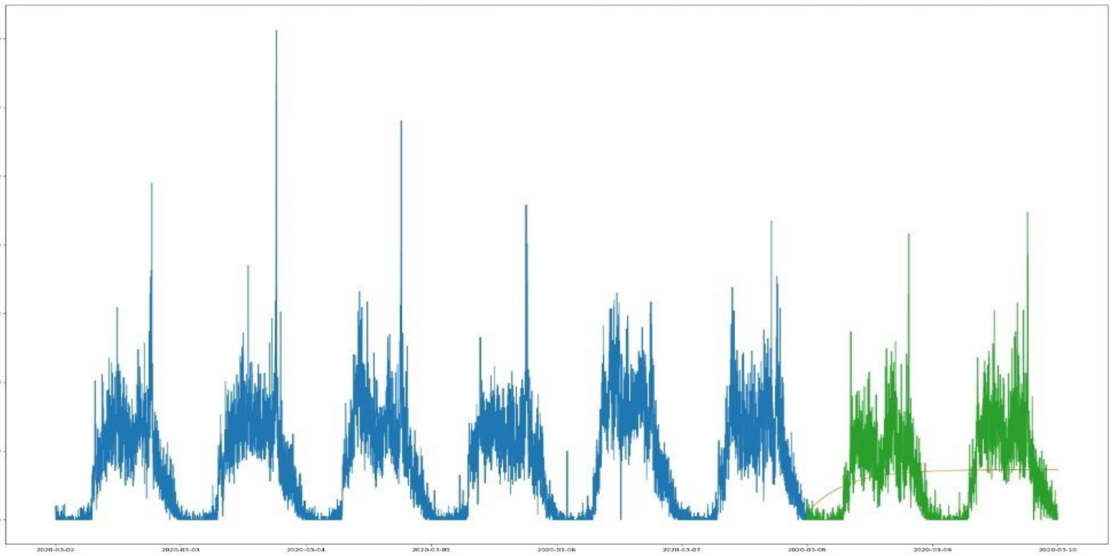

Fig. 2. ARIMA forecasting on traffic flow (1 minute)

C. SARIMA Model

SARIMA stands for Seasonal Auto Regressive Integrated Moving Average model which is the extended version of ARIMA with seasonality. As per the time series decomposition shown in the time series analysis, it is visible that traffic data consists of seasonality. For, traffic data seasonality can be seen for each day of the whole week as during morning time traffic is less, which reaches to peak level during office hours while it is a moderate level during noon time again in the evening traffic flow reaches at peak level while during night time traffic flow is nearly zero.

SARIMA model consists of parameters (P, D, Q, S) along with the (p, d, q). Parameters P, D, Q are similar to those of p, d, q of the ARIMA model. While S represents the seasonality of the SARIMA model i.e., 12 for yearly, 4 for quarterly. Likewise, for 30 minutes of the interval, there will be 2 data points at each hour and during 24 hours we will have 24 * 2 = 48 data points so, Seasonality will be 48 for 30 minutes of the time interval. Same way, if the data is taken for each minute, then during 24 hours seasonality will be 24 * 60 = 1440.

In this paper, the SARIMA model is used to forecast the traffic flow data for 15 minutes and 30 minutes intervals of the next day which can be described as below (see Fig. 3.and Fig. 4.),

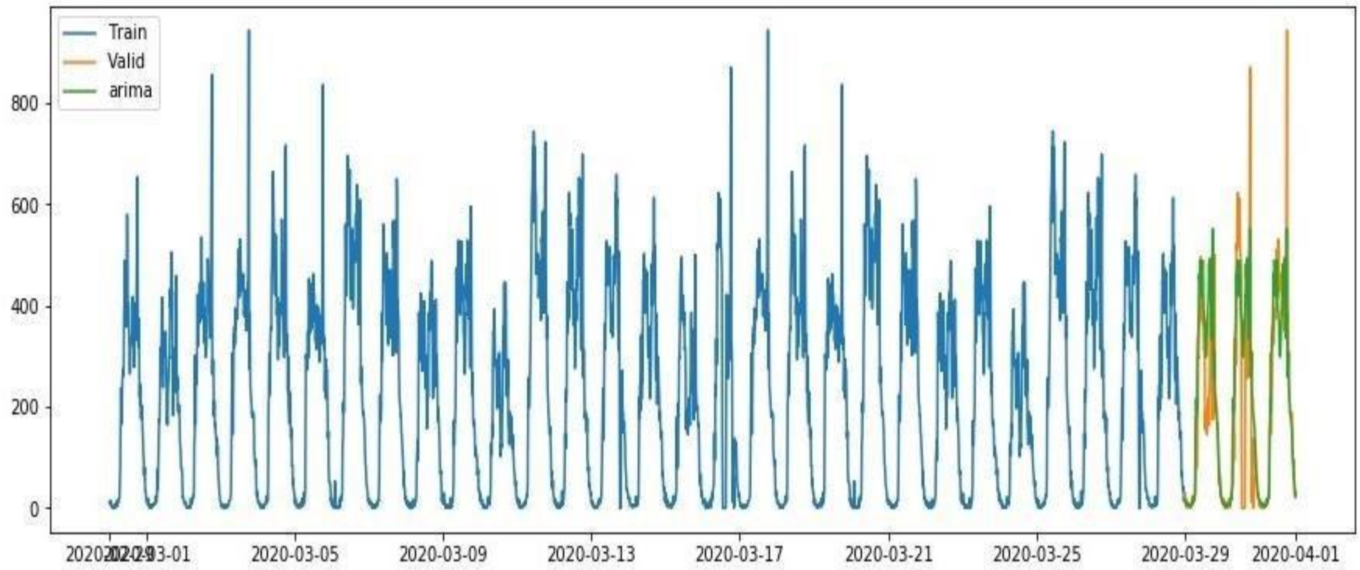

Fig. 3. SARIMA time series forecasting on traffic flow (15 minutes)

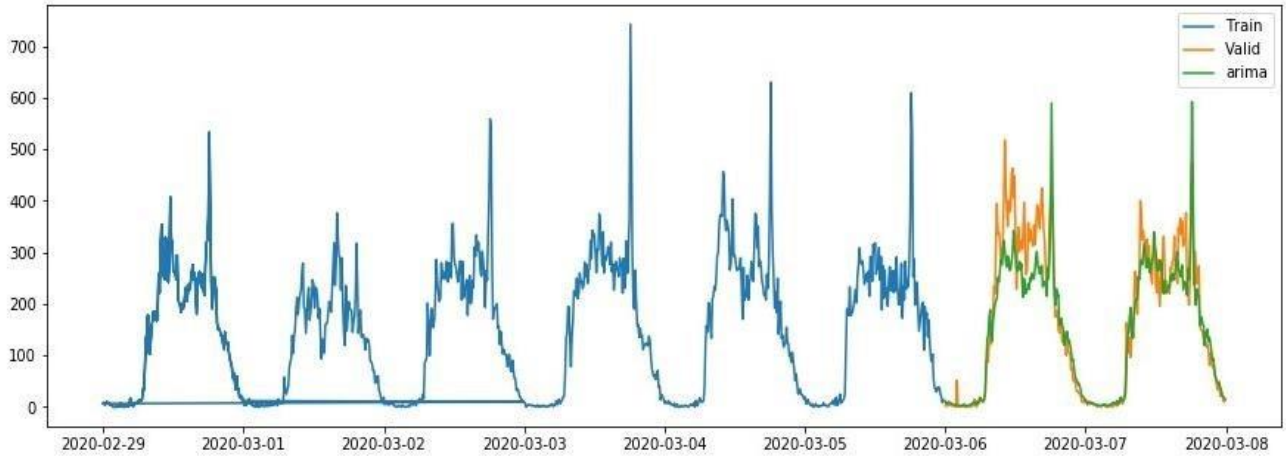

Fig. 4. SARIMA time series forecasting on traffic flow (30 minutes)

During the research, it is found that the SARIMA model doesn't work appropriately for short time-interval spans i.e., for 1 minute and 5 minutes due to the large computational time required by the model to provide forecasting. This limitation is overcome by the FbProphet model.

D. FbProphet Model

FbProphet stands for Facebook Prophet, which is an open-source time-series algorithm developed by the research team of Facebook. FbProphet accommodates seasonality with multiple periods. This method is also resilient to missing values which means the overall performance won't be affected by the missing values. This technique won't consider outliers while training the model on time series data which is one of the major points.

In this paper, FbProphet is used on traffic inflow data for short time intervals of data i.e., for 1 minute, to generate better and appropriate forecasting for the next day for every minute. The forecasting result of traffic flow for 1 minute of intervals for the day can be shown as below (see Fig. 5.),

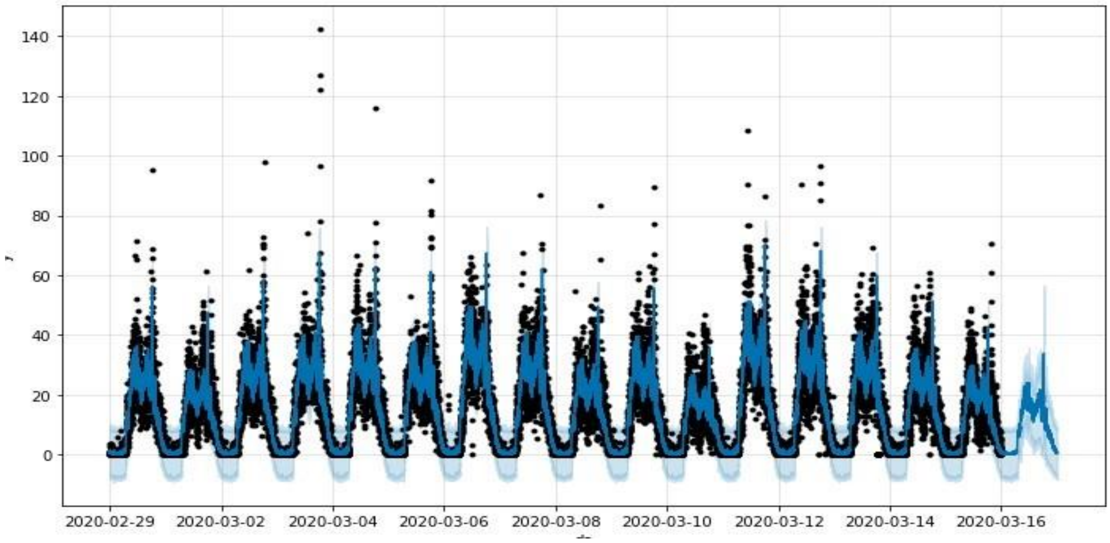

Fig. 5. FbProphet time series forecasting on traffic flow (1 minute)

IV. ANALYSIS AND IMPLICATIONS

In this research paper, we have obtained the results of forecasted traffic flow for the next upcoming day for various intervals of time. To choose an adequate model for a specific interval we have considered R2. R2 i.e., coefficient of determination is one of the accuracy measures which is used to get an accurate model.

Algorithm	$\mathbf{R^2}$
ARIMA	NA
SARIMA	-
FbProphet	0.65

Table 2. Result Analysis (1 Minute)

Algorithm	R^2
ARIMA	NA
SARIMA	0.70
FbProphet	-

Table 3. Result Analysis (15 Minutes)

Algorithm	R^2
ARIMA	NA
SARIMA	0.86
FbProphet	-

Table 4. Result Analysis (30 Minutes)

As per the above tables, it is visible that the ARIMA model fails to anticipate traffic flow for the upcoming next day as it doesn't consider the seasonality of data. While, SARIMA provides appropriate results up to some extent for 15 minutes and 30 minutes which is 70% and 86% respectively but fails for short time intervals (1 min., 5 min.) due to computational limitations. Furthermore, the FbProphet model provides nearly 65 % accurate results when forecasting of inflow traffic is performed for 1 minute for the next day.

V. CONCLUSION

People's demand for automobiles is growing in tandem with the economy's rapid growth. The primary cause of urban traffic congestion is an excess of automobiles. Traffic control and efficient vehicle routing planning can be performed by forecasting short-term traffic flow, thereby alleviating traffic congestion and reducing resource loss caused by traffic congestion.

In this paper, we revisit time series forecasting methods i.e., ARIMA, SARIMA, and FbProphet, and their real-world application to resolve one of the critical issues of urban cities of India. We show that accurate traffic time series modelling is essential for studies on this topic to produce relevant results. The main findings of this study are as follows, (1) Time series decomposition gives a better understanding of data to apply appropriate forecasting models. (2) By applying ARIMA we showed, how the seasonality factor affects the forecasting results. (3) We showed, SARIMA works better for a time interval of more than 15 minutes to provide appropriate prediction but fails to give proper results for a too small-time interval of data due to large computational limits. (4) We described the application of FbProphet to forecast traffic inflow for short period intervals of 1 minute for the next upcoming day.

Although this study proposes a traffic flow prediction method based on statistical approaches i.e., time series analysis, in the future the author will use other machine learning and deep learning approaches to find a better model to forecast more precisely. The findings of the research are limited to the studies only as they are not yet applied to the real world. The anticipation of the traffic flow is applied only on one side of the highway which will be implemented further for four crossroads to get appropriate insights to resolve the issue.

REFERENCES

[1]. Aftabuzzaman, M. (2007, September). Measuring traffic congestion-a critical review. In *30th Australasian transport research forum* (pp. 1-16).

[2]. CCI India Website, https://www.cciindia.org/highway.html (last accessed on 14th July 2021)

[3]. Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, *43*, (pp. 3-19).

[4]. Zhang, G. Y., Zhao, Q., Luo, Z. W., & Wei, H. (2009). Short-term traffic flow prediction with ACD and particle filter. In *ICCTP 2009: Critical Issues In Transportation Systems Planning, Development, and Management* (pp. 1-7).

[5]. Chien, S. I., Liu, X., & Ozbay, K. (2003). Predicting travel times for the South Jersey real-time motorist information system. *Transportation Research Record*, *1855*(1), (pp. 32-40).

[6]. Ghosh, B., Basu, B., & O'Mahony, M. (2007). Bayesian time-series model for short-term traffic flow forecasting. *Journal of transportation engineering*, *133*(3), (pp. 180-189).

[7]. Davis, G. A., & Nihan, N. L. (1991). Nonparametric regression and short-term freeway traffic forecasting. *Journal of Transportation Engineering*, *117*(2), (pp. 178-188).

[8]. Vythoulkas, P. (1993). Alternative approaches to short term traffic forecasting for use in driver information systems. *Transportation and traffic theory*, *12*, (pp. 485-506).

[9]. Smith, B. L., & Demetsky, M. J. (1994, October). Short-term traffic flow prediction models-a comparison of neural network and nonparametric regression approaches. In *Proceedings of IEEE international conference on systems, man and cybernetics* (Vol. 2, pp. 1706-1709).

[10]. Kirby, H. R., Watson, S. M., & Dougherty, M. S. (1997). Should we use neural networks or statistical models for short-term motorway traffic forecasting?. *International journal of forecasting*, *13*(1), (pp. 43-50).

[11]. Yin, H., Wong, S., Xu, J., & Wong, C. K. (2002). Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies*, *10*(2), (pp. 85-98).

[12]. Lee, D. H., Zheng, W. Z., & Shi, Q. X. (2004, January). Short-term freeway traffic flow prediction using a combined neural network model. In *Proceedings of the 84th Annual Transportation Research Board Meeting, Washington DC, USA.*

[13]. Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2005). Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C: Emerging Technologies*, *13*(3), (pp. 211-234).

[14]. Tan, M. C., Wong, S. C., Xu, J. M., Guan, Z. R., & Zhang, P. (2009). An aggregation approach to short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, *10*(1), (pp. 60-69).

[15]. Çetiner, B. G., Sari, M., & Borat, O. (2010). A neural network based traffic-flow prediction model. *Mathematical and Computational Applications*, *15*(2), (pp. 269-278).

[16]. Ogunwolu, L., Adedokun, O., Orimoloye, O., & Oke, S. A. (2011). A neuro-fuzzy approach to vehicular traffic flow prediction for a metropolis in a developing country.

[17]. Ge, Y., & Wang, G. (2011). Study of traffic flow short-time prediction based on wavelet neural network. In *Electrical engineering and control* (pp. 509-516). Springer, Berlin, Heidelberg.

[18]. Pamuła, T. (2012, October). Traffic flow analysis based on the real data using neural networks. In *International Conference on Transport Systems Telematics* (pp. 364-371). Springer, Berlin, Heidelberg.

[19]. Kumar, K., Parida, M., & Katiyar, V. K. (2015). Short term traffic flow prediction in heterogeneous condition using artificial neural network. *Transport*, *30*(4), (pp. 397-405).

[20]. Williams, B. M., Durvasula, P. K., & Brown, D. E. (1998). Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record*, *1644*(1), (pp. 132-141).

[21]. Ahmed, M. S., & Cook, A. R. (1979). Analysis of freeway traffic time-series data by using Box-Jenkins techniques (No. 722).

[22]. Levin, M., & Tsao, Y. D. (1980). On forecasting freeway occupancies and volumes (abridgment). *Transportation Research Record*, (773).

[23]. Nihan, N. L., & Holmesland, K. O. (1980). Use of the Box and Jenkins time series technique in traffic forecasting. *Transportation*, *9*(2), (pp. 125-143).

[24]. Hamed, M. M., Al-Masaeid, H. R., & Said, Z. M. B. (1995). Short-term prediction of traffic volume in urban arterials. *Journal of Transportation Engineering*, *121*(3), (pp. 249-254).

[25]. Lee, S., & Fambro, D. B. (1999). Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. *Transportation Research Record*, *1678*(1), (pp. 179-188).

[26]. Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of transportation engineering*, *129*(6), (pp. 664-672).

[27]. Ghosh, B., Basu, B., & O'Mahony, M. (2005, January). Time-series modelling for forecasting vehicular traffic flow in Dublin. In *84th Annual Meeting of the Transportation Research Board, Washington, DC.*

[28]. Dong, H., Jia, L., Sun, X., Li, C., & Qin, Y. (2009, August). Road traffic flow prediction with a time-oriented ARIMA model. In *2009 Fifth International Joint Conference on INC, IMS and IDC* (pp. 1649-1652).

[29]. Lippi, M., Bertini, M., & Frasconi, P. (2013). Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *IEEE Transactions on Intelligent Transportation Systems*, *14*(2), (pp. 871-882).

[30]. Mai, Tiep, Ghosh, Bidisha, Wilson, Simon (2012). Multivariate Short-Term Traffic Flow Forecasting Using Bayesian Vector Autoregressive Moving Average Model. *Transportation Research Board 91st Annual Meeting.*

[31]. Chung, E, Rosalion, N. (2001). Short Term Traffic Flow Prediction. *Australian Transport Research Forum (ATRF)*, 24th, 2001, Hobart, Tasmania, Australia

[32]. Smith, B. L., Williams, B. M., & Oswald, R. K. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, 10(4), (pp. 303-321).

[33]. Stathopoulos, A., & Karlaftis, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, 11(2), (pp. 121-135).

[34]. Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), (pp. 37-45)

[35]. Chikkakrishna, N. K., Hardik, C., Deepika, K., & Sparsha, N. (2019, December). Short-term traffic prediction using sarima and FbPROPHET. In *2019 IEEE 16th India Council International Conference (INDICON)* (pp. 1-4).

[36]. Machine Learning Mastery Blog, https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/ (last accessed on 14th July, 2021)

[37]. Minitab Support website, https://support.minitab.com/en-us/minitab/19/help-and-how-to/statistical-modeling/time-series/supporting-topics/time-series-models/additive-and-multiplicative-models/ (last accessed on 14th July, 2021)