Journal of Traffic and Transportation Management
Volume 3, No. 1 (2021) pp. 17-24

Machine Learning and statistic predictive modeling for road traffic flow

Nadia Slimani[a]*, Ilham Slimani[b], Nawal Sbiti[a], Mustapha Amghar[a]

[a] *Computer Systems and Productivity Team, EMI, Mohammed V University, Rabat, Morocco.*
[b] *Laboratory SmartICT, ENSAO, Mohammed I University, Oujda, Morocco.*

Abstract

Traffic forecasting is a research topic debated by several researchers affiliated to a range of disciplines. It is becoming increasingly important given the growth of motorized vehicles on the one hand, and the scarcity of lands for new transportation infrastructure on the other. Indeed, in the context of smart cities and with the uninterrupted increase of the number of vehicles, road congestion is taking up an important place in research. In this context, the ability to provide highly accurate traffic forecasts is of fundamental importance to manage traffic, especially in the context of smart cities. This work is in line with this perspective and aims to solve this problem. The proposed methodology plans to forecast day-by-day traffic stream using three different models: the Multilayer Perceptron of Artificial Neural Networks (ANN), the Seasonal Autoregressive Integrated Moving Average (SARIMA) and the Support Machine Regression (SMOreg). Using those three models, the forecast is realized based on a history of real traffic data recorded on a road section over 42 months. Besides, a recognized traffic manager in Morocco provides this dataset; the performance is then tested based on predefined criteria. From the experiment results, it is clear that the proposed ANN model achieves highest prediction accuracy with the lowest absolute relative error of 0.57%.

Keywords: *Road traffic forecasting, artificial neural networks, MLP, statistical forecasting, SMOreg, SARIMA.*

1. Introduction

With the frequent congestions experienced in urban and inter-urban perimeters, the advent of new data innovations in the transportation field, called Intelligent Transportation Systems (ITS), presents a major opportunity to mitigate this scourge, it allows a better knowledge of the road traffic especially in the context of connected and smart cities concepts. Traffic flow forecasting is an important part of the intelligent transportation system. By providing real-time traffic data information for traffic drivers, it is more efficient to select the optimal route and reduce the loss of time and costs due to traffic congestion.

During the recent three decades, this issue was concentrated by several authors and with multiple approaches under several perspectives. Those works can be grouped under concentrating on procedures of traffic modeling or contemplating the performance of different prediction models including parametric and non-parametric techniques or contrasting the viability of those models by looking at the forecast horizon (short or long-term) and the context of study (urban or non-urban areas).

The literature offers a variety of approaches to traffic forecasting, which remains the key success factor for good traffic management. This research is in line with this perspective and is based on real daily traffic stream recorded by a recognized traffic manager in Morocco in order to predict future traffic data using three forecasting models:

- Artificial neural network (ANN) as non-parametric model [1];
- Seasonal Autoregressive Integrated Moving Average (SARIMA) as a classic parametric model [2] ;
- SMOreg as a second non-parametric model [3].

In order to verify the effectiveness of the proposed model, two groups of datasets and different models are studied in the experiment. The performance of these models will then be compared between different simulations according to established criteria.

The current paper is sorted out as follows: The first section is dedicated to the literature review including the study's context and a review of various strategies used to tackle the traffic forecasting. The subsequent part is dedicated to the proposed methodology by depicting in subtleties the dataset and the model of forecasting. Finally, in the third part, numerical experimentations are introduced, with the acquired results and the evaluation of the best forecast model dependent on various criteria alongside the conclusions.

* Corresponding author. Tel.: +212 661 456 185
Fax: +212 537 600 015; E-mail:slimani.nadia@gmail.com

DOI: 10.5383/JTTM.03.01.003

2. Literature review

2.1. Context and problem description

Every country in the world is experiencing a rapid development of motorized transport, which is expected to continue on an upward trend. However, the development of road infrastructure has not kept pace, which implies a reduction in space available per vehicle on the roads and several bottlenecks [4] which can be categorized in two types [5]: recurrent congestions with a specific periodicity and stochastic caused by external factors: accidents, road works, climate, social and cultural events (matches, religious celebrations, etc…)

Traffic flow prediction is undeniably a key component of traffic management and ITS (Intelligent Transportation System). In past few decades, various traffic flow prediction models have been created to aid traffic the board and control for improving transportation efficiency going from course direction and vehicle steering to flag coordination. The evolution of traffic flow can be considered as a temporal and a spatial process. The traffic flow expectation issue can be expressed as follows:
Let X_i^t denote the observed traffic flow quantity during the t time interval at the i observation location in a transportation network. Given a sequence $\{X_i^t\}$ of observed traffic flow data, with $i = 1,2, \ldots, m$ and $t=1,2,\ldots,T$.
The problem is to predict the traffic flow at time interval $(t+\Delta)$ for some prediction horizon Δ.
As an endeavor to take care of this pragmatic issue, the current paper means to anticipate every day traffic stream utilizing three diverse determining methods including neural network, SARIMA and SMOreg and utilizing a real dataset of traffic that was recorded in Morocco to test the performance of each model.

2.2. An overview of traffic forecasting models

Throughout the literature, the problem of traffic modelling has been widely addressed from several perspectives. The dominating forecasting models can be segmented and summarized into three clusters as follows:

2.2.1. Parametric models:

The parametric models are built from statistical assumptions and exploit data related to time series defined by the problem being addressed. These models are assimilated to a box to which we provide input information (historical data and variables) and return output information (previsions) contained in variables; the model will provide forecast results depending on these variables and relationships induced by the choice of model.

Mainly, the research concerned the application of Box-Jenkins time series models to the dynamic system of single point traffic flow forecasting. This work has addressed most of the parametric model concerns for traffic condition data by establishing a theoretical foundation for using seasonal ARIMA forecast models.

Those statistical approaches mainly include Auto–regressive integrated moving average (ARIMA) [6] and K–nearest neighbor (KNN) [7] methods.
Several researches with ARIMA model was widely used to forecast several aspects of traffic, such as [8] where Williams et al. propose SARIMA model to forecast urban freeway traffic flow and Wang et Liu [9] used ARIMA to forecast velocity of vehicles. Lee et Fambro [10] adopted ARIMA for freeway context.
Tong et al. [11] proposed SARIMA model to forecast highway traffic volume, the numerical application demonstrate that the accuracy was better than grouped regression model, variable seasonal index forecasting model and seasonal regression model. A. Emami et al. [12] and [13] presented a low-cost mean for short-term flow prediction based on the connected vehicle data, based on Kalman filter equations.

2.2.2. Non-parametric models:

The non-parametric models have the ability to learn from the past data to predict future values of the outputs. They offer the advantage of innate ability to learn even with non-linearity and complexity of parameters, in addition these models are not forcing the data to follow a specific model that would not describe all possible situations. Widely used models are gravitating around artificial intelligence.

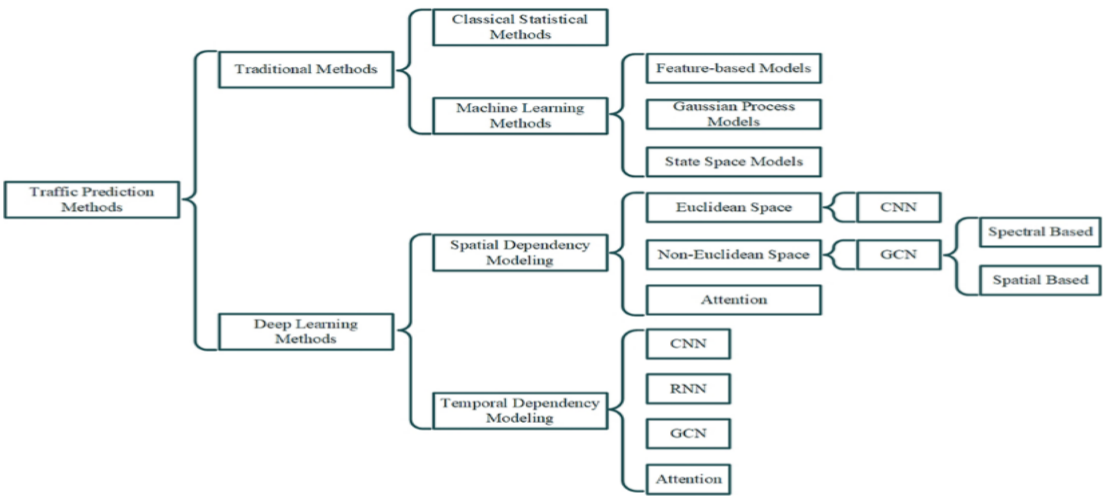

Fig. 1. Categories of traffic prediction models

Artificial intelligence and Support vector machine (SVM) [14] has been deployed in wide-spectrum domains. Qu et al. [15] using realistic data and contextual factor data to forecast daily traffic flow. Zhuo et al. [16] presented a new Long Short–Term Neural Network (LSTM) with Deep Neural Networks (DNN) to improve the accuracy of the prediction.

2.2.3. Hybrid models:

The hybrid models are based on a synergetic combination of models to improve the emulating and prediction performances. T.Ma, et al. [17] developed a novel hybrid approach NN-ARIMA for network-wide traffic forecast where a neural network employed to capture network-scale co-movement pattern of all traffic flows, and ARIMA is used to further extract location-specific traffic features in the residual time series out of Neural Network. In [18] employed a novel hybrid model by combining a Spatiotemporal Graph Convolutional Network and a Gated Recurrent Units Neural Network for short-term traffic speed forecasting The experimental results show that the accuracy of prediction was improved.

The figure 1 presents another way to present the different categories of traffic forecasting models.

3. Experiments with research procedure

3.1. Aim and contribution

After presenting a literature review of traffic forecasting methods, this section is dedicated to presenting the proposed methodology as well as analyzing and comparing the obtained results. Indeed, this research aims to explore three models to forecast daily traffic volume (based on parametric and non-parametric models), then an evaluation of each model is made according to the chosen criteria.

3.2. Data description

The data used to evaluate the performance of the proposed model was collected from a heavily used toll station in Morocco's highway. The station S140 is located between the most important economic and political poles of Morocco (see Figure 2 below), it carries the highest traffic of highways in Morocco. It is an open system toll station, i.e. motorists pass through only one toll station, where they pay a single amount, independent of the number of kilometers traveled on the highway.

The dataset contains traffic stream data recorded inside every hour during years 2015, 2016, 2017 and 2018 and fragmented by characterization of the kind of vehicle (lightweight vehicle: class 1, truck: class 2, vehicle with at least three axles: class 3). Examples of recorded traffic are presented in the following Table 1.

Fig. 2. Location of the toll station S140 used in the study.

The dataset contains 30792 recordings of hourly traffic flow. This dataset is divided into 42 months of traffic information; it is separated into two collections: training and testing set.
The training set is detailed utilizing daily traffic information of the initial three years including 2015, 2016 and 2017. The test set contains day-by-day traffic information of the initial half year of 2018. The following figures (see Figure 3 and Figure 4) illustrate the dataset used, as the real traffic recorded in the station S140 during 36 months.

Table 1. Dataset example.

Day	Hour	Class 1	Class 2	Class 3	Hourly average
06/04/2015	09:00	1367	128	71	522
25/08/2016	17:00	1342	149	25	505
30/12/2017	16:00	1981	185	132	766
08/06/2018	01:00	389	116	97	200

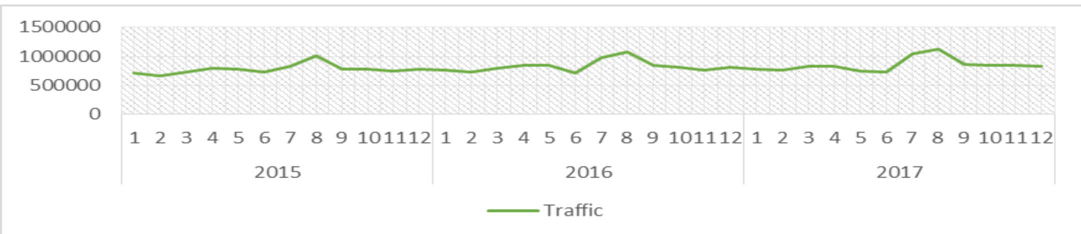

Fig. 3. Daily traffic recorded in the station S140 during 2015, 2016 and 2017.

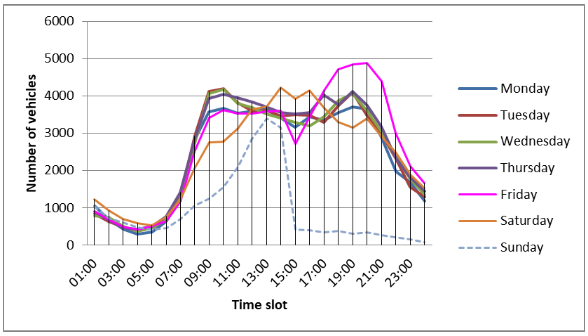

Fig. 4. Hourly traffic flow by weekday from station S140 in 2017.

In the following sections, the principles of the three used models in this study and the obtained results are explained.

3.3. Multilayer perceptron

An artificial neural network is designed to simulate the human brain analyzing and processes information. It solves problems that could be complex to implement in statistical tools. The ANN contains different type of neurons, each one have a functional property in the structure: pick up the signal from the upstream neurons, generate information from those signals and transfer the signal to the downstream neurons.
The Multilayer Perceptron (MLP) is a class of feedforward artificial neural network (Figure 5). Learning occurs in the perceptron by changing connection weights after each piece of data is processed, based on the amount of error in the output compared to the expected result [19]. This is an example of supervised learning, and it is carried out through backpropagation algorithm, a generalization of the least mean squares algorithm in the linear perceptron [20]. Back-propagation is the practice of fine-tuning the weights of a neural net based on the error rate obtained in the previous iterations, making the model reliable.

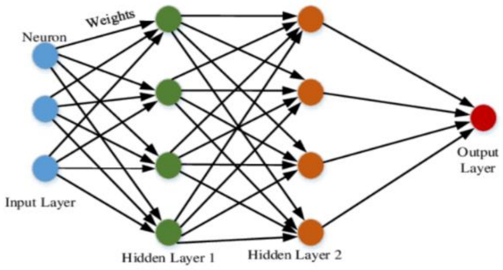

Fig. 5. Architecture of Back-propagation Multi-Layer neural network.

Our proposed neural network contains an input layer provided by 18 values. The output represents the expected traffic flow of the day in the next order. The input values are explained bellow:

- 8 calendar information as follows:

Feature 1	Working day (Yes / No)
Feature 2	Weekend (Yes / No)
Feature 3	National holiday (Yes / No)
Feature 4	Religious holiday (Yes / No)
Feature 5	School holiday (Yes / No)
Feature 6	Ramadan (Yes / No)
Feature 7	Strike (Yes / No)
Feature 8	Chronological order of the day (for example first day of a religious celebration, or last day of school holiday)

- 10 previous daily traffic flows as follows:

Feature 9	Traffic of the previous day (day-1)
Feature 10	Traffic observed 2 days before (day-2)
Feature 11	Traffic observed 3days before (day-3)
Feature 12	Traffic observed 4 days before (day-4)
Feature 13	Traffic observed 5 days before (day-5)
Feature 14	Traffic observed 6 days before (day-6)
Feature 15	Traffic observed one week before (day-7)
Feature 16	Traffic observed 2 weeks before (day-14)
Feature 17	Traffic observed 3 weeks before (day-21)
Feature 18	Traffic observed one year before (day-365)

The next step consists to scaling data with the Min-Max normalization using this operation:

$$x_i \rightarrow \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

Where:

x_i The traffic flow of day i to be normalized

x_{min} The minimal traffic flow recorded in the learning database

x_{max} The maximal traffic flow recorded in the learning database

Several artificial neural networks structures were tested for traffic flow prediction, including Adaline, NoProp and Multi-Layer Perceptron.

The database used for the history was split into two parts: the daily traffic of three years (i.e. 85% of the data) for training, and the daily traffic of the following six months (i.e. 15% of the remaining data) for testing. Therefore, 1096 and 181 patterns are respectively employed in training and testing phases. Moreover, to find the adequate network's topology offering the best traffic forecasts in several attempts are applied by changing the number of hidden layers and the number of neurons in those hidden layers.

The best traffic flow predictions during our experimentations are obtained with a Multi-Layer Perceptron of three hidden layers, the first hidden layer is composed of five neurons, the second is composed of eight neurons and the third one is composed of two neurons. The transfer function is Sigmoid and

Backpropagation as learning algorithm with the existence of the Bias neuron.

Table 2. Monthly traffic forecasting results using the best MLP topology (hidden layers 5-8-2)

Month	Actual traffic	Forecasted traffic	Error
Jan-18	855.016	833.918	11.098
Feb-18	773.178	765.904	7.274
Mar-18	838.361	833.108	5.253
Apr-18	934.568	893.573	3.995
May-18	835.100	837.943	-2.843
Jun-18	827.913	832837	-4.924

The best MLP result is obtained with a total Mean Square Error (MSE) of 0.00927 in the train set (see Figure 6) and 0.01321 in the test set.

A presentation of the traffic prediction results is presented in the following table (see Tables 2 and 3). For more details about this study, please refer to [1].

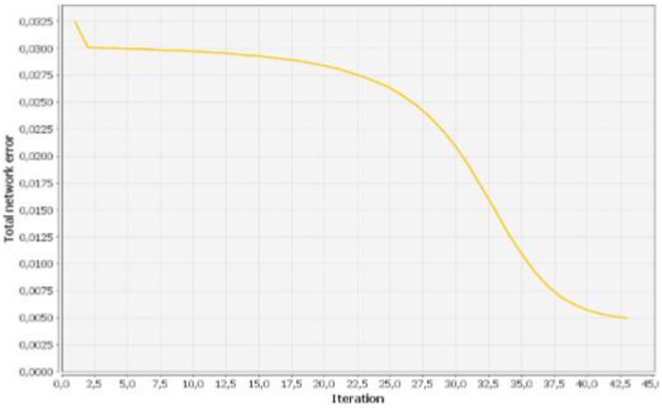

Fig. 6. Graph of the total MSE error of the network: training set

Table 3. An overview of daily traffic forecasting results using the best MLP topology (hidden layers 5-8-2)

Day	Actual traffic	Forecasted traffic	Difference
08/01/2018	23147	22902	245
17/01/2018	27739	28176	-437
28/02/2018	27471	28347	-876
24/05/2018	25080	24587	493

3.4. Seasonal Autoregressive Integrated Moving Average

Autoregressive Integrated Moving Average (ARIMA) is one of the most used techniques for mathematical forecasts modeling. SARIMA designed the seasonal property of periodic variations that characterizes the studied time series.

This model is suitable for linear predictions and is based on modelling the time series in the past with parameters to find out then using those same parameters and built model in order to predict future time series values.

The term ARIMA model includes many variants. In the model ARIMA(p,d,q), the parameter p denotes the order of the auto-regressive part, the parameter q the order of the moving average part, and d the number of differentiation steps.

The main steps of establishing the seasonal ARIMA model [11] are explained in our previous work [21]. The modeling process was realized by the software Excel by downloading the function NUMXL. The first step is to determine the stationarity of the input data series via the autocorrelation function (ACF). In our study, the time series of traffic flow is not stationary. ARIMA models require the input data to have a constant mean, variance, and autocorrelation through time. Therefore, the input data series were transformed into a stationary model through a differencing process. The number of non-seasonal differences (d) was set to 1 to ensure the stationarity of the data. Following this, parameter estimation in the ARIMA models was performed, and then traffic flow was forecasted using the SARIMA models.

SARIMA model were first trained using the data in the training set (From January 2015 to December 2017), and then tested using the testing set (From January 2018 to June 2018). Several models was tested, and the best forecasting performance was SARIMA (1,0,1) (1,0,1)12 whose results are showed in table 4 bellow:

Table 4. Traffic forecasting results using SARIMA

Month	Actual traffic	Forecasted traffic	Error
Jan-18	855.016	754.851	100.165
Feb-18	773.178	740.389	32.789
Mar-18	838.361	747.625	90.736
Apr-18	934.568	784.813	149.755
May-18	835.100	826.381	8.719
Jun-18	827.913	841.929	-14.016

The SARIMA model has not been included in the comparison below since the data studied shows periodic monthly variations.

3.5. Support Vector Regression

The Support Vector Regression model (SMOreg) is inspired from the Support Vector Machine (SVM) algorithm dedicated to solve only classification problems [22]. Indeed, Support Vector Machine can be utilized as a regression technique keeping up all the features characterizing the algorithm including the maximal margin concept. With just a minor contrast, the SVR algorithm utilizes indistinguishable standards from the SVM algorithm for classification. To solve a regression problem, a margin (epsilon) is set to approximate the SVM. “Basically, what SVR does is to find a regression function that fits well the training instances by minimizing the prediction error, this error is user defined and forms tube around the regression function, discarding all the data points that are outside the margin. Besides minimizing the error, the algorithm maximizes the flatness of the regression function for generalization purpose. The larger the tube (bigger error deviation) the flatter the function” [23].

The SVR function can be used in linear and non-linear problems. Indeed, in a non-linear problem, Kernel function, Polynomial or Gaussian Radial Basis are used to allow a linear fit of the training data among the support vectors and the attribute instances.
Using the SMOreg algorithm in Weka with the function RegOptimizer in order to learn automatically most of the algorithm's parameters. Road traffic predictions are obtained using the following Scheme:
SMOreg -C 1.0 -N 0 -I "RegSMOImproved -T 0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K "PolyKernel -E 1.0 -C 250007"

The improved version is chosen to better performance (I). The C is the complexity parameter responsible for the fitting level. The rest of the parameter are set by default, N parameter is set to normalize the variable, 0.001 for the stopping criterion (T), by default the variant of the algorithm is 1 of 2, the epsilon is the error deviation threshold and is set to 1 (p). The epsilon for the loss-function is also set to 1 (L), 1 random seed (w). Finally, the nonlinear function is Polykernel, it is employed in most of SVM algorithms.
With the already described data, the used dataset is composed of all historical attributes, as explained earlier, and formatted as an arff file: "RoadTrafficDataset.arff". An overview of the obtained prediction results is displayed in the following table (see Table 5):

Table 5. An overview of traffic predictions using SMOreg

Day	Actual traffic	Forecasted traffic	Error
15/06/2018	22395	29175	6780
21/06/2018	29973	28787	1186
22/06/2018	31653	29926	1727
23/06/2018	29718	28658	1060
27/06/2018	32991	28464	4527
28/06/2018	33409	29410	3999
29/06/2018	35602	30531	5071
30/06/2018	34025	29243	4782

3.6. Results: comparison and discussion

The experimental results confirm that the proposed model with artificial neural network gives the best forecasting results. Indeed, an example comparing the absolute error of the proposed models MLP, SARIMA and SMOreg is given in the next table (see Table 6) where the minimum absolute error (4924) is obtained by using the Multilayer Perceptron model [16].

Table 6. Absolute error comparison: Example June 2018

Performance criteria	Total traffic	Absolute error	Relative error
Actual traffic	827913	-	-
MLP	832837	4924	0,57%
SARIMA	841929	14016	1,69%
SMOreg	807560	20353	2,45%

The MLP recorded the best forecasting performance with 0.57% absolute error. For a better results' evaluation, other performance measures are taken into account to compare the performance of MLP and SMOreg forecasts including RMSE (Root mean squared error), MAE (Mean absolute error) and MSE (Mean squared error).
They are defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - \hat{f}_i|$$

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|f_i - \hat{f}_i|}{f_i}$$

$$\mathrm{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}\left(|f_i - \hat{f}_i|\right)^2\right]^{\frac{1}{2}}$$

where f_i is the observed traffic flow, and $\hat{f}_i$ is the predicted traffic flow.

After performing we obtained the best architecture for different prediction tasks, which are compared in Table 7. Therefore, Multilayer Perceptron model can be used to produce traffic forecasts with a good accuracy. These performances are comparable with those obtained by Wu et al. [24] using a hybrid deep learning model .

Table 7. Traffic forecasting: comparison of performance criteria (1-step-ahead)

Performance criteria	SMOreg	SMOreg
RMSE	0,6436	0,1605
MAE	0,5545	0,0939
MSE	0,4142	0,0258

An illustration comparing daily traffic predictions, by zooming in on the results for the month of June, is presented in the next figure (see Figure 7).

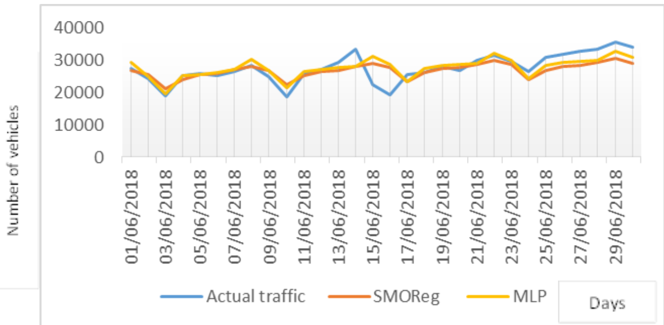

Fig. 7. Daily forecasting results: Actual traffic vs. MLP forecasts vs. SMOreg forecasts

4. Conclusion and future lines

Prediction is definitely a key component of road traffic control. Although traffic prediction has made great progress in recent years, there are still many challenges to be investigated.

Considering that, this paper tackles the use of various methods to forecast traffic based on real Moroccan dataset of station S140 "Bouznika" located between Rabat and Casablanca, the most important economic and political poles in Morocco.
The proposed approach consists of comparing the predictive performances of three different methods, namely the MLP as a neural networks structure, SARIMA as a mathematical modeling method and the SMOreg inspired from the Support Vector Machine algorithm. The results clearly show that the best predictions are obtained by using the MLP technique with an absolute relative error of 0.57%.

However, to increase the model's performance, it would we will be interested to consider the following aspects in our future work:

- Include external conditions (weather, road maintenance work, adjacent points of interest…) that are non-recurrent, that makes difficulties to obtain data, thus, it involves a small simple size and learning more difficult than under normal traffic conditions. However, sufficient data is usually a prerequisite for deep learning methods. A solution is to use transfer learning techniques to perform deep spatio-temporal prediction tasks.
- Multi-source information: generally bringing together a variety of detectors (cameras, passage loop, and traffic collection station) and a multitude of data sources (social networks, satellite views, etc.)
- Real-time forecasting: The reason for ongoing traffic forecast is to lead information preparing and traffic condition appraisal in a brief timeframe. The challenge is significant, in order to allow a prediction from a dynamic incoming data flow.
- Benchmarking and comparing the obtained results with other methods like Long Short Term Memory a (LSTM), recurrent neural networks and deep learning approach.

References

[1] N. Slimani, I. Slimani, N. Sbiti, et M. Amghar, Traffic forecasting in Morocco using artificial neural networks Traffic forecasting in Morocco using artificial neural networks », *Procedia Computer Science*, vol. 151, nº 2018, p. 471-476, 2019, doi: 10.1016/j.procs.2019.04.064. https://doi.org/10.1016/j.procs.2019.04.064

[2] N. Slimani, I. Slimani, M. Amghar, et N. Sbiti, « Comparative analysis of traffic prediction models based on a real case study ». INTIS 2019.

[3] N. Slimani, I. Slimani, M. Amghar, et N. Sbiti, « Road traffic forecasting using a real data set in Morocco », *Procedia Computer Science*, vol. 177, p. 128-135, 2020, doi: 10.1016/j.procs.2020.10.020. https://doi.org/10.1016/j.procs.2020.10.020

[4] T. Nagatani, « The physics of traffic jams », *Rep. Prog. Phys.*, vol. 65, nº 9, p. 1331-1386, sept. 2002, doi: 10.1088/0034-4885/65/9/203. https://doi.org/10.1088/0034-4885/65/9/203

[5] M. Shahgholian et D. Gharavian, « Advanced Traffic Management Systems : An Overview and A Development Strategy », p. 1-39.

[6] D. Pavlyuk, « Short-term Traffic Forecasting Using Multivariate Autoregressive Models », *Procedia Engineering*, vol. 178, p. 57-66, 2017, doi: 10.1016/j.proeng.2017.01.062. https://doi.org/10.1016/j.proeng.2017.01.062

[7] Z. Zheng et D. Su, « Short-term traffic volume forecasting: A k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm », *Transportation Research Part C: Emerging Technologies*, vol. 43, p. 143-157, juin 2014, doi: 10.1016/j.trc.2014.02.009. https://doi.org/10.1016/j.trc.2014.02.009

[8] B. M. Williams et L. A. Hoel, « Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results », *Journal of Transportation Engineering*, vol. 129, nº 6, p. 664-672, nov. 2003, doi: 10.1061/(ASCE)0733-947X(2003)129:6(664). https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664)

[9] J. Wang et Y. Liu, « Mean Velocity Prediction Information Feedback Strategy in Two-Route Systems under ATIS », *Advances in Mechanical Engineering*, vol. 7, nº 2, p. 640416, févr. 2015, doi: 10.1155/2014/640416. https://doi.org/10.1155/2014/640416

[10] S. Lee et D. B. Fambro, « Application of Subset Autoregressive Integrated Moving Average Model for Short-Term Freeway Traffic Volume Forecasting », *Transportation Research Record*, vol. 1678, nº 1, p. 179-188, janv. 1999, doi: 10.3141/1678-22. https://doi.org/10.3141/1678-22

[11] M. Tong et H. Xue, « Highway Traffic Volume Forecasting Based on Seasonal ARIMA Model », *Journal of Highway and Transportation Research and Development (English Edition)*, vol. 3, nº 2, p. 109-112, 2008, doi: 10.1061/JHTRCQ.0000255. https://doi.org/10.1061/JHTRCQ.0000255

[12] A. Emami, M. Sarvi, et S. Asadi Bagloee, « Using Kalman filter algorithm for short-term traffic flow prediction in a connected vehicle environment », *J. Mod. Transport.*, vol. 27, nº 3, p. 222-232, sept. 2019, doi: 10.1007/s40534-019-0193-2. https://doi.org/10.1007/s40534-019-0193-2

[13] D. Xu, Y. Wang, L. Jia, Y. Qin, et H. Dong, « Real-time road traffic state prediction based on ARIMA and Kalman filter », *Frontiers Inf Technol Electronic Eng*, vol. 18, nº 2, p. 287-302, févr. 2017, doi: 10.1631/FITEE.1500381. https://doi.org/10.1631/FITEE.1500381

[14] P. Li, Y. Peng, P. Jiang, et Q. Dong, « A support vector machine based semiparametric mixture cure model », *Comput Stat*, vol. 35, nº 3, p. 931-945, sept. 2020, doi: 10.1007/s00180-019-00931-w. https://doi.org/10.1007/s00180-019-00931-w

[15] L. Qu, W. Li, W. Li, D. Ma, et Y. Wang, « Daily long-term traffic flow forecasting based on a deep neural network », *Expert Systems with Applications*, vol. 121, p. 304-312, mai 2019, doi: 10.1016/j.eswa.2018.12.031. https://doi.org/10.1016/j.eswa.2018.12.031

[16] Q. Zhuo, Q. Li, H. Yan, et Y. Qi, « Long short-term memory neural network for network traffic prediction », in *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Nanjing, nov. 2017, p. 1-6, doi: 10.1109/ISKE.2017.8258815. https://doi.org/10.1109/ISKE.2017.8258815

[17] T. Ma, C. Antoniou, et T. Toledo, « Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast », *Transportation Research Part C: Emerging Technologies*, vol. 111, p. 352-372, févr. 2020, doi: 10.1016/j.trc.2019.12.022. https://doi.org/10.1016/j.trc.2019.12.022

[18] M. Jiang, W. Chen, et X. Li, « S-GCN-GRU-NN: A novel hybrid model by combining a Spatiotemporal Graph Convolutional

Network and a Gated Recurrent Units Neural Network for short-term traffic speed forecasting », *J. of Data, Inf. and Manag.*, janv. 2021, doi: 10.1007/s42488-020-00037-9. https://doi.org/10.1007/s42488-020-00037-9

[19] S. Zhang et K. Lin, « Short-Term Tra ffi c Flow Forecasting Based on Data-Driven Model », p. 1-17, 2020. https://doi.org/10.3390/math8020152

[20] D. Pavlyuk, « Short-Term Traffic Forecasting Using Multivariate Autoregressive Models », *Procedia Engineering*, vol. 178, p. 57-66, 2017, doi: 10.1016/j.proeng.2017.01.062. https://doi.org/10.1016/j.proeng.2017.01.062

[21] C. Proceedings, « 2019 8 », 2019.

[22] A. J. Smola et B. S. C. H. Olkopf, « A tutorial on support vector regression ∗ ¨ », p. 199-222, 2004. https://doi.org/10.1023/B:STCO.0000035301.49549.88

[23] H. Ricardo, *Forecasting Tourism Demand for Lisbon's Region. A Data Mining Approach,*. Munich: GRIN Verlag, 2017.

[24] Y. Wu, H. Tan, L. Qin, B. Ran, et Z. Jiang, « A hybrid deep learning based traffic flow prediction method and its understanding », *Transportation Research Part C: Emerging Technologies*, vol. 90, p. 166-180, mai 2018, doi: 10.1016/j.trc.2018.03.001. https://doi.org/10.1016/j.trc.2018.03.001