

Forecast of traffic speeds with neural network LSTM encoder-decoder

Previsão de velocidades de tráfego com rede neural LSTM encoder-decoder

Douglas Zechin¹, Matheus Basso do Amaral², Helena Beatriz Bettella Cybis³

¹Federal University of Rio Grande do Sul, Rio Grande do Sul – Brazil, douglaszechin@gmail.com

²Federal University of Rio Grande do Sul, Rio Grande do Sul – Brazil, matheus.basso96@gmail.com

³Federal University of Rio Grande do Sul, Rio Grande do Sul – Brazil, helenabc@producao.ufrgs.br

Submitted:

17 de julho de 2021

Accepted for publication:

30 de agosto de 2022

Published:

14 de dezembro de 2022

Editor in charge:

Cira Souza Pitombo

Keywords:

Congestion.

Traffic forecasting.

Neural networks.

Palavras-chave:

Breakdown de tráfego.

Previsão de tráfego.

Redes neurais.

DOI:10.14295/transportes.v30i3.2660



ABSTRACT

This article proposes a speed prediction model for a highway segment in the city of Porto Alegre, which has daily traffic jams due to bottlenecks. We used traffic data and environmental variables, such as rainfall intensity, accidents and atypical events to make the forecasts. Then we proposed a neural network model with an encoder-decoder architecture and long short-term memory (LSTM) layers, which has the characteristic of establishing long-term relationships between the input variables, being relevant for applications in the Transportation area. As additional contributions, we evaluated the quality of forecasts for different prediction horizons and traffic regimes. We compared cumulative distribution functions (CDFs) generated using field and forecast data using a survival analysis method similar to the breakdown probability calculation. These CDFs represent the probability of a sudden speed drop due to the transition from the free-flow to the congested regime. The methodology presented a satisfactory performance based on both criteria, making good predictions even in critical traffic situations.

RESUMO

Este artigo tem como objetivo propor um modelo de previsão de velocidades para um trecho de rodovia na cidade de Porto Alegre, que apresenta congestionamentos diariamente por conta de gargalos. Para realizar as previsões foram utilizados dados de tráfego e variáveis ambientais, como intensidade de chuva, acidentes e eventos atípicos. Propôs-se então um modelo de rede neural com arquitetura encoder-decoder e camadas long short-term memory (LSTM), que possuem a característica de estabelecer relações de longa dependência temporal entre as variáveis de entrada, sendo pertinentes para aplicações na área de Transportes. Como contribuições adicionais, avaliou-se a qualidade das previsões para diferentes horizontes de previsão e regimes de tráfego, e comparou-se a capacidade e as curvas de probabilidade de breakdown calculadas com dados de campo e previstos. A metodologia apresentou desempenho satisfatório com base em ambos os critérios, sendo capaz de fazer boas previsões mesmo em situações críticas de tráfego.

1. INTRODUCTION

Traffic Engineering has received important contributions from recent technological advances in other areas, such as IoT (Internet of Things) and artificial intelligence. The intersection between these areas has led to the emergence of innovative fields of study, such as Smart Cities and autonomous vehicles, in addition to contributing to traditional areas, such as active traffic management (ATM), in which this study fits.

ATM has been around since the first half of the last century and traditionally proposes using simple algorithms and traffic and speed detectors to manage highway traffic operations. Although many traffic agencies still use these methods, ATM has received many contributions from data-driven approaches and seems to be increasingly merging with the concept of Smart Cities (Ma, Zhang and Ihler, 2020). An important feature made possible by more robust methods is improving traffic forecasts and anticipating undesired scenarios, such as congestion, accidents, and increased travel time.

In this paper, we propose using long short-term memory (LSTM) neural networks to perform speed predictions in the vicinity of a highway bottleneck located in the metropolitan region of Porto Alegre, Brazil. However, the proposed methodology aims to prioritize forecasts made close to the road capacity, which is the most critical moment for traffic management. The predictions consist of the expected average speed for the subsequent 5 time intervals of 5 minutes and are based on traffic data, precipitation, and other possibly relevant information such as the day of the week and detector malfunctions. We chose this approach because LSTMs can retain information by creating long-term dependencies, which generally results in better performance than parametric methods and standard neural networks for time series prediction.

Speed forecasting can lead to good results in terms of average error since traffic speed is mostly stable due to the existence of speed limits. However, a low average error can hide large forecast errors at critical times, such as during peak demand periods, where traffic characteristics change quickly. Other authors rarely address this problem, so we propose segregating the data into five sets with equivalent traffic characteristics and analyzing the model error for each one individually and for each forecast horizon.

We used Survival Analysis by the Kaplan-Meier method to confirm the quality of traffic forecasts during peak periods close to road capacity. In this case, survival is related to the maintenance of a non-congested regime, and death is associated with the beginning of the transition to a congested regime. We statistically tested the similarity of cumulative distribution functions (CDFs) constructed with field and predicted data. Although the region presents breakdowns daily, the measured phenomenon was not treated as a breakdown because the detectors are located upstream of the active bottleneck. In this way, the CDFs represent the probability of starting the transition from the free-flow regime to the congested regime.

Until the conclusion of this article, the evaluation of the quality of traffic forecasting methodologies from the comparison of survival curves made with the forecasts and with field data had not been used in other researches. However, we understand that this produces a solid comparison, as these methods are already well established among the traffic engineering community and allow for the calculation of road capacity. Therefore, in addition to a detailed discussion about the model's error, we propose evaluating its effectiveness from this approach.

2. LITERATURE REVIEW

The development and improvement of traffic forecasting methods are alternatives for improving traffic management on urban highways and arterials (Vlahogianni, Karlaftis and Golias, 2014). Precise short-term and real-time predictions can be used as input into ATM algorithms, contributing to more efficient and responsive traffic management (Gu *et al.*, 2019). Traffic predictions are often a specific application of parametric time series prediction methods such as naïve and ARIMA. Although these methods have greater physical interpretability and

their solution is usually simpler (Fu, Zhang and Li, 2017), the computational capacity and the great availability of currently existing data allow the use of more robust models, such as neural networks.

Due to the dynamic nature of demand, non-linear non-parametric models tend to be better suited to capture traffic's spatial and temporal evolution to make good speed predictions. Recurrent Neural Networks (RNNs) adapt well to this type of problem, as they are a type of neural network capable of processing temporal sequences. However, there are different subtypes of RNNs with different purposes, and one of the best suited to this study is the LSTM. LSTMs can retain relationships with long temporal dependence, which is crucial for correctly interpreting traffic seasonality.

Hochreiter (1997) proposed the LSTM architecture with the main objective of modeling long dependencies, which is not possible with standard RNNs. Short-term traffic predictions can be defined as estimating the state of traffic for a close time in the future (Gu *et al.*, 2019). For this reason, accuracy and precision are essential aspects that must be considered. LSTM is a great candidate as it captures the non-linearity of traffic dynamics in an effective way across using memory blocks and thus has a superior capacity for predicting time series with long time dependencies (Ma *et al.*, 2015).

The ease of access to high-level neural network programming tools has enabled rapid assimilation of new techniques for specific applications (Chollet, 2018; Géron, 2019). Because of this, the use of LSTM neural networks has gained space for solving traffic problems, which are highly time-dependent and have multiple variables that are related in a complex way. Fu *et al.* (2017) showed that LSTM and GRU neural networks (Gated Recurrent Units) have similar performance for traffic flow prediction and perform better when compared to the ARIMA method. Laptev *et al.* (2017) proposed an application of an LSTM neural network with an encoder-decoder structure to forecast the travel demand of an urban private transport company and capable of making predictions with high quality. A comparison between FFN (Feed Forward Network), CNN (Convolutional Neural Network), and LSTM was made by Asplund (2019), who obtained better results using the LSTM neural network to predict traffic conditions using public transport traffic information as input data. As stated by Vlahogianni *et al.* (2014), the interest of researchers has shifted towards more responsive prediction methods and models for non-recurring traffic conditions through the development of prediction systems with high algorithmic complexity. Furthermore, do Amaral (2020) compared the quality of velocity predictions in the same locality using different predictive models and concluded that an LSTM neural network produced better predictions than traditional methods such as linear regression, ARIMA, and regular neural networks.

In this article, therefore, we propose making speed predictions in a segment of a suburban highway where breakdowns are observed daily due to the existence of a bottleneck. To make these predictions, we used environmental and traffic data collected with inductive loops upstream of the bottleneck. We chose as the model a LSTM neural network with encoder-decoder architecture to increase the predictive capabilities of LSTM neural networks pointed out in other studies. We assessed the quality of the predictions by comparing the error of the forecasts in traffic situations with similar characteristics and testing whether the CDF calculated with the predictions is equivalent to that calculated with field data.

3. METHODOLOGY

This article proposes using an LSTM neural network to make speed predictions using traffic data from a point on a Brazilian highway. Information on precipitation, road accidents, and atypical events were concatenated with traffic data and then grouped by lane at regular intervals to generate the input variables that feed the neural network. As input and output variables, we defined how much time in the past and the future the proposed network would consider to make predictions. After training the neural network, we evaluated the results for different regions of the fundamental diagram and compared them with the CDF obtained through the field data.

3.1. Study site

The study region comprises a section of the BR-209 highway in Porto Alegre, RS, selected due to the high traffic volumes in the morning peak period. The breakdown phenomenon occurs regularly on weekdays due to this great demand, bottlenecks in the approaches, and the lifting of the mobile span of the Guaíba Bridge downstream of the data detection location (Caleffi *et al.*, 2016; Caleffi, 2018; Zechin, Caleffi and Cybis, 2020), as shown in Figure 1.

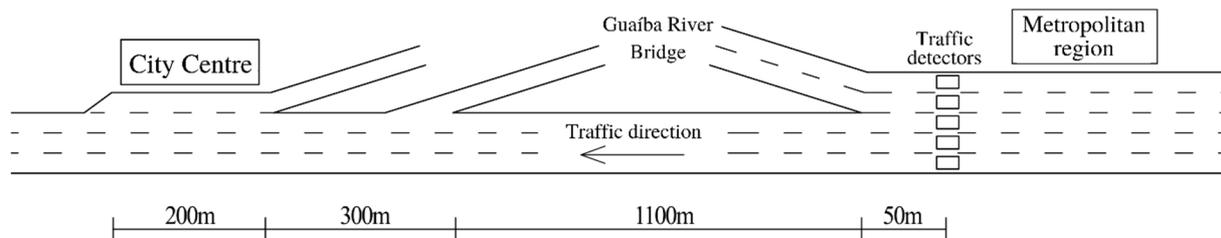


Figure 1. Study region

3.2. Traffic and environmental data

The data used in this article were made available by the company Triunfo Concepa, the concessionaire that operated the stretch of the highway. These data were collected using inductive loops located approximately 50 meters upstream of a fork that connects the road to the Guaíba Bridge. The data consists of two years (2016 and 2017) of disaggregated traffic counts with information on the instant of each vehicle's passage, speed, and lane. We only used data from the three lanes on the left since the others do not present congestion and connect the road to the bridge. We discarded data from days when the detectors malfunctioned, weekends, and days with accidents within a 5 km radius of the detectors, resulting in a useful sample of 263 days.

We also used environmental data to provide the network with as much useful information as possible. We obtained rainfall data from a rain gauge 500 m away from the inductive loops from the Cemaden (National Center for Monitoring and Alerts for Natural Disasters) online portal. We treated it as a continuous variable since rainfall intensity was provided at intervals of up to 10 minutes. We replicated the rainfall intensity calculated for a given instant to the previous data aggregation intervals used in the study until the time when another measurement was reported. This methodology is compatible with the data aggregation methodology used by Cemaden. In addition to rainfall data, we used the day of the week and bridge lifts as dummy variables.

In this region, the breakdown phenomenon occurs daily around 7:30am with no important exceptions. Because of this, we defined 4am to 11am as a suitable period for the analysis based on the speed profile of the highway. This covers the development of demand in the early morning, congestion, and the recovery of the free flow regime.

3.3. Generation of inputs and outputs

LSTM neural networks require data spaced in regular intervals to make adequate predictions, so we aggregated the data at 5 min intervals. Then, from the aggregated data, we created the variables volume, standard deviation of speed, average speed, minimum speed, median speed, and maximum speed per lane. We consolidated environmental variables and traffic variables, and continuous variables were normalized.

We defined the neural network inputs as 12 intervals in the past (60 min), each comprised of the previously created variables. For the outputs, we defined a forecast horizon of 25 min, corresponding to 5 intervals of 5 min, and the predicted variable was the average speed of the road. The first 80% of the data, in chronological order, was used for training and the remaining for testing. We did so to bring the study closer to an actual application, where past data would be used to predict unknown future events.

3.4. LSTM neural network with encoder-decoder architecture

Although neural networks with cells of the LSTM type have a remarkable ability to predict time series, relying on the ability to retain long-term information, there are network architectures that allow predictions to be even more accurate. In this work, we propose using the encoder-decoder architecture, as shown in Figure 2, which has shown promising results in applications in the transport area (Laptev *et al.*, 2017). This architecture interprets the information in two stages: the encoder processes the data, and the decoder computes the model outputs.

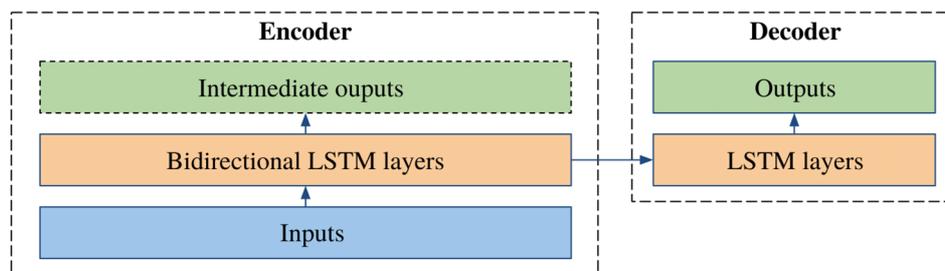


Figure 2. Encoder-decoder architecture with bidirectional LSTM layers

We inserted the input data into the neural network through the encoder. It passes through the bidirectional intermediate layers (Schuster and Paliwal, 1997), which make abstractions using LSTM cells. The computed information then follows two paths: (i) it is passed to a layer that generates intermediate outputs with the exclusive objective of increasing the assertiveness and stability of the model, and (ii) it is passed to the decoder, where it passes through intermediate layers before generating the outputs that are actually used as a forecast.

4. RESULTS

We created the proposed neural network model using the Keras (Chollet, 2015) and Tensorflow libraries in Python. As it is a relatively small neural network, it was possible to carry out the

training on the Google Colab cloud computing service, which has 12 GB of RAM memory and an NVIDIA Tesla P100 graphics card.

Neural network models have many parameters that can be adjusted to obtain better predictions. These parameters include the number of intermediate layers, number of neurons in each layer, activation functions, loss functions, optimization algorithms, and regularization algorithms. Although default values are used for general purposes, some parameters must necessarily be adjusted. These adjustments, in turn, can be made by trial and error or using some structured methodology. In this study, we used the hyperband technique (Li *et al.*, 2018), which has proven more time efficient and accurate than other techniques, such as grid search and random search. We also used the mean absolute error (MAE) of the decoder predictions as the objective function to be optimized. The optimization of the network hyperparameters took about 2 h. We present the optimized parameters and the respective optimal values in Table 1.

Table 1 – Optimized Parameters

Parameter	Tested values	Optimum value
Bidirectional LSTM layers of the encoder	0 – 3 bidirectional LSTM + 1 LSTM	2
LSTM layers of the decoder	1 - 5	1
Bidirectional LSTM layer neurons	32 - 512	512
LSTM layer neurons	32 - 512	256
Loss function	Mean square error; absolute mean error; percent average absolute error	Mean square error
Optimizer	Adam; RMSprop ; adagrad ; adadelta	RMSprop
Dropout	0.1 - 0.4	0.15

In addition to these parameters, we used a variable learning rate as a function of the number of training epochs of the neural network. It started with a learning rate of 10^{-3} and was divided by 10 every 20 training epochs.

Then we retrained the optimal model found with the hyperband technique for 60 epochs to achieve complete convergence. We used the model with the lowest MAE in the test portion for the following stages of the study since the use of many epochs can lead to overfitting (Chollet, 2015; Gal and Ghahramani, 2016).

4.1. Forecasts evaluation

The evaluation of the quality of traffic predictions on highways and arterials is not trivial since it does not have uniform characteristics in time and space. Traffic on these roads is usually classified as free flow or congested, and traffic behavior in each of these situations is entirely different and requires different and specific strategies. The transitions between these regimes also present peculiarities and are of particular interest for traffic management since they are linked to the operational capacity of the roads.

With this in mind, we proposed segregating the data into analysis regions with similar traffic characteristics from the flow-speed diagram. In this way, the error can be compared by analysis region and forecast horizon, as shown in Figure 3. We created the proposed regions empirically according to the following criteria: (R1) free flow; (R2) drop in speed due to proximity to capacity; (R3) transition to the congested state; (R4) congestion; and (R5) free flow recovery.

The MAE of the forecasts was 5.40 km/h globally. However, we observed that the error differs in order of magnitude when comparing different traffic regions and forecast horizons:

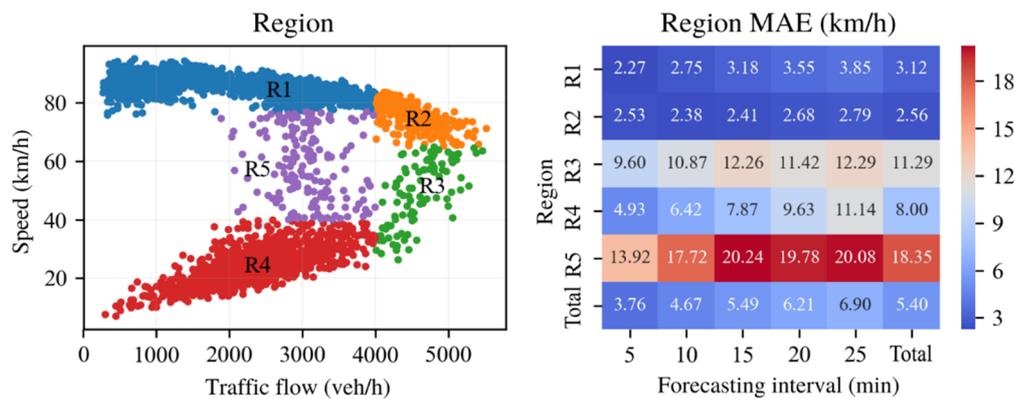


Figure 3. Regions of analysis and MAE by region and forecast horizon

- R1: In this region, vehicles travel at speeds limited by the legal limits of the road. Because of this, the MAE is expected to be low, basically resulting from different individual desired speed choices (Galvan, Zechin and Cybis, 2019). A low MAE was achieved by the proposed model, with little error increase even for the maximum forecast horizon;
- R2: this was the region where the model made the most accurate predictions, which is interesting since it precedes the beginning of the transition to the congested state. In this region, there is greater speed homogeneity resulting from the increase in traffic flow. However, the speed profile does not follow a stable pattern like the R1 region. Good predictions, especially for longer horizons, indicate that the model is capable of predicting the onset of congestion;
- R3: This region refers to the transition from the free flow to the congested state. In this region, a sudden drop in speed is observed, and the calculated average velocity depends significantly on the instant within the aggregation interval (5min in this study) in which this phenomenon occurred. Because of this, there is great speed variability in this region, and it is natural that larger errors are observed proportionally to the size of the chosen data aggregation interval. Thus, in this region, it is expected that the model is able to capture the rapid downward trend even with larger errors than in the other regions. Based on this, we understand that the errors found are compatible with expectations;
- R4: vehicles travel in a stop-and-go motion in this region, and the speed variability is more significant. This happens mainly because data was collected with inductive loops, which measure the instantaneous speed of vehicles. The model errors are smaller for shorter prediction horizons and are close to the errors measured in the R1 region, but increase for larger horizons. There is less interest in obtaining highly accurate predictions in this region since the possibilities of acting on traffic are lower during congestion due to the high density and low speed of vehicles;
- R5: Although predictions during congestion are not very interesting, the possibility of predicting free flow recovery is interesting, and this is done in the R5 region. However, this is the region where the model incurred the most significant errors. The probability that the congestion will end depends on the volume upstream of the bottleneck approaches decreasing, which cannot be measured with just one detector, especially during congestion. Because of this, we expected forecasts in this region to be reactive, respond to measured velocity variations, and have a low anticipation capacity. We stated that this occurred, since the error is high and increases as the forecast horizons increase.

To support the interpretations above, we propose evaluating how the error behaves as a function of time. Figure 4 shows the speed profile used in the test portion of the neural network along with the predictions made, the error of each prediction, and the volume used as weight during training. To evaluate the quality of predictions in the future, we present the predictions made for the first (5 min) and fifth (25 min) predicted interval. As the test portion is large, we show a sample of 200 predicted sequences, where some important phenomena can be observed.

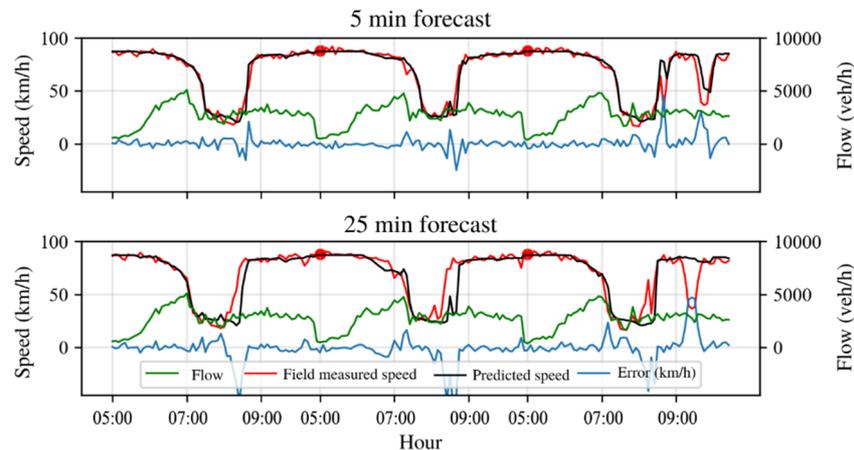


Figure 4. Speed predictions over time. The volumes are scaled to correspond to the vertical axis. Red signals indicate the start of a new morning.

Although distant in time, we observe that the predictions made for 5 and 25 min in the future are similar in terms of error and have good adherence to the speeds measured in the field. The error is noticeably smaller in the regions close to the transition from the free flow to the congested regime since the volume was used as weight during the training process, increasing the relative importance of these intervals. This is a highly desired effect since good forecasts close to capacity are necessary to anticipate the beginning of the transition to the congested regime. In free-flow moments, speed variability is greater since the volume is low, and most vehicles travel unimpeded. It is interesting to note that the model converged to linear predictions in these situations since the main trend is stable and the weights are smaller because they are proportional to the volume. In the congested regime, both forecast horizons have larger and similar errors due to the speed fluctuations that occur during the stop-and-go motion. The biggest difference between the predicted intervals happens in the transition from the congested to the free flow regime; in this case, the speed forecasts seem to react to changes on the road without anticipation of speed recovery. This is clear by looking at the delay between forecasts and field measurements, which is even more significant in the 25 min forecast. As expected, the predictions regarding the recovery of free flow are more erratic than the others since they are highly dependent on the flow of vehicles upstream of the bottleneck under analysis. As this information does not exist in this study, it is natural that the observed error is greater.

4.2. Validation using the predictions to calculate CDFs

The analyses indicate that the proposed model performs satisfactorily for the speed prediction task, especially in regions of particular interest for active traffic management.

Model validation was performed by calculating and statistically comparing CDFs constructed with field and predicted data. These curves were estimated using the breakdown probability calculation methodology suggested by Brilon et al. (2005) to provide robustness to the model validation (Han and Ahn, 2018). Then we statistically tested the hypothesis that the CDFs generated with measured and predicted speeds in the field are different. In this study, we did not use the term breakdown to refer to the measured phenomenon due to the unfavorable position of the detectors. However, the survival analysis does not make this distinction. It is sufficient that we compute the observed phenomenon the same way using predicted and field data for the statistical analysis to be valid.

The methodology for breakdown probability calculation used by Brilon et al. (2005) is widely recognized for its effectiveness and simplicity, having also been used in several studies that followed it (Andrade and Setti, 2014; Elefteriadou et al., 2011, 2014). The original methodology defines a speed threshold, so that the interval preceding a drop in speed that exceeds this limit is considered a breakdown. This interval is censored (received a 1 marker) and the intervals preceding it receive a 0 marker. We discarded intervals following the breakdown. Then we sorted the markers and their respective volumes from the entire database by volume and applied them to the non-parametric Kaplan-Meier model (Kaplan and Meier, 1958) to generate breakdown probability curves as a function of volume. In this study, we considered that the beginning of the transition to the congested regime is analogous to the breakdown phenomenon treated in these studies. We adapted the methodology by Brilon et al. (2005), adding as a criterion for identifying a censored interval the need for 2 consecutive intervals to be below the established speed threshold. We did so to reduce the likelihood of identifying false positives.

Although the breakdown probability curve provides a stochastic view of the road's capacity, traffic managers tend to prefer to use a deterministic value for it. Shojaat et al. (2016) proposed the sustainable flow index (SFI) to meet this demand without giving up the information offered by the probability distribution. This metric originates from the concept of risk, defined as the multiplication of the probability of an adverse event occurring and the damage caused by it. In the context of traffic engineering, and more precisely of the occurrence of a breakdown, the SFI represents the volume that transits through a road and is calculated by the product between the volume and the complementary probability of the occurrence of a breakdown.

The capacity, therefore, is obtained by maximizing the SFI. As an example, the SFI curves, the CDFs made with the predictions, and the speeds measured in the field are shown in Figure 5. We used the last predicted interval (25 min in the future) and a speed threshold of 65 km/h.

Then we investigated the quality of the predictions applied to this methodology by varying the threshold velocity to identify the highest threshold velocity that (i) generates statistically identical probability curves and (ii) produces similar capabilities. The hypothesis that the generated curves are identical was tested by fitting the Cox survival model (Cox, 1972) to the volume data, from the binary marker of early transition to the congested regime calculated previously and an accessory variable that indicates whether the data refers to a prediction or a field measurement. We tested the significance of the accessory variable in the model through the likelihood ratio test, so that p-values greater than an assigned acceptance limit $\alpha = 0.05$ do not allow rejecting the null hypothesis that the curves are identical with 95% confidence, which is desirable in this study. Figure 6 shows the p-values obtained in comparing the curves generated for different speed thresholds and each forecast horizon. Note that we only created 4 curves, since we considered making two predictions lower than the established speed

threshold an identification criterion for the beginning of the transition to the congested regime. We also present the calculated capacities.

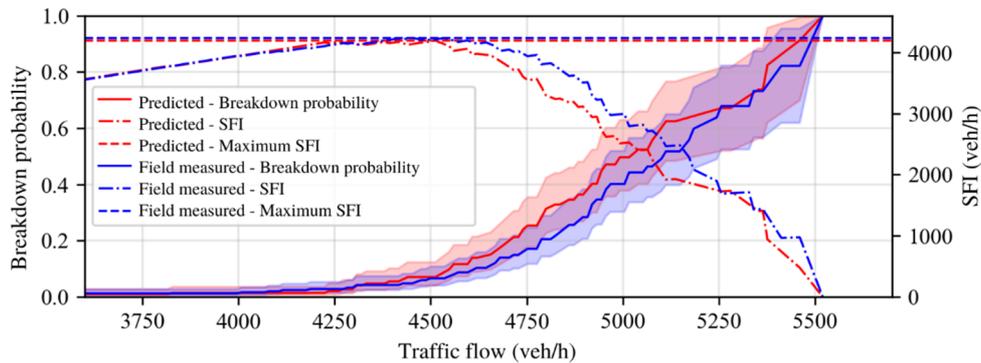


Figure 5. CDFs of the beginning of the transition to the congested regime and SFI for speed threshold = 65 km/h with 25 min prediction data

Speeds greater than 70 km/h generally produce p-values below the limit α , where the null hypothesis that the distributions are identical is rejected. However, we note that the threshold speed from which the p-values become greater than this threshold decreases as the forecast horizon increases. We understand that this occurs because the forecasts are more imprecise the longer the forecast horizon, and the assertiveness of the forecasts increases when there are clearer signs that the speed drop has started and lower speeds are measured.

Visual inspection in Figure 6 shows a convergence between the calculated capacity values for values close to 65 km/h, where there is a maximum absolute difference below 200 veh/h. As we observe convergence between capacities for this speed threshold and all p-values are greater than 0.05, we understand that the neural network well represents both the beginning of the transition to the congested regime and capacity.

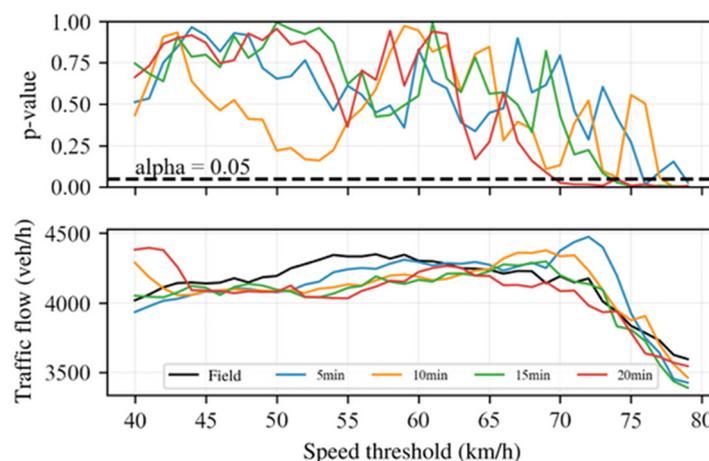


Figure 6. p-value of the accessory variable in the Cox survival model and capacity for different limit speeds

In this application, the speed threshold of 65 km/h could be suggested to characterize the beginning of the transition to the congested regime from speeds predicted by the real-time model in practical applications. However, it is noteworthy that this value is suggested based on the data of this specific case study, so that the ideal speed threshold may differ in other locations due to geometric and behavioral specificities and peculiarities in the demand profile.

5. CONCLUSIONS

In this article, we proposed using an LSTM neural network with encoder-decoder architecture to perform speed predictions of a road segment where breakdowns are observed daily due to a bottleneck. We used rainfall and traffic data collected with inductive loops, including road accidents and lifting information from the mobile span of a bridge, to aggregate as much information relevant to the neural network as possible. We evaluated the forecast results for different traffic states to detail the model's quality. We also validated the results by applying predictions in the calculation of CDFs that represent the probability of the beginning of the transition to the congested regime.

With an MAE of 5.40 km/h, the forecast errors obtained in the regions of greatest interest showed satisfactory results for all predicted intervals, but it is noted that the error increases with the forecast horizon. The use of volumes as a sample weight allowed the reduction of prediction errors in situations where traffic is close to capacity. Because of this, we observed convergence between the probability curves calculated with field and predicted data, indicating that the model can also make good predictions at critical moments for traffic.

Practical applications of the proposed methodology must consider the peculiarities of the used data. The hyperparameters found during the neural network optimization process may differ depending on factors such as the amount of data, the number of variables created, data aggregation, and traffic characteristics in the studied region. The suitability of the methodology for the chosen region can also be verified through the generation and statistical comparison of CDFs.

We suggest for future work using data from detectors located closer to the bottlenecks, so that the breakdown characterization can be performed with greater precision, and to assess whether the location of the detectors significantly influences the results. The use of data from multiple sections of the segment, especially upstream, would also be interesting, as it would allow the model to consider the local traffic state and the volume of vehicles that will pass through the section in the future. Traffic has a stochastic nature, so the probabilistic prediction of speeds may be a more appropriate tool (Fortunato et al., 2017; Kendall and Gal, 2017). Making predictions using adaptations of LSTM neural networks compatible with disaggregated traffic data can also contribute to maximizing the use of information (Neil, Pfeiffer and Liu, 2016). Neural network models are often considered black-box models. However, recent advances indicate ways to create visualizations for humans (Arras et al., 2019). Crossing traffic data with other databases can add even more information to the network, such as the use of traffic images, Bluetooth data, telephony data, and integrations with mobility applications, as well as a previous study of the significance of the variables, to reduce the number of variables used and provide only relevant information. Other models of neural networks, such as transformers networks, also seem promising for solving traffic problems (Wu et al., 2020) but still demand more studies.

REFERENCES

- do Amaral, M.B. (2020) *Previsão de velocidades de tráfego com rede neural LSTM*. Bachelor thesis. Departamento de Engenharia Civil: Universidade Federal do Rio Grande do Sul. Porto Alegre, Brazil.
- Andrade, G.R. and Setti, J.R. (2014) Speed-Flow Relationship and Capacity for Expressways in Brazil, in *Innovative Applications of the Highway Capacity Manual 2010*. DOI: <http://onlinepubs.trb.org/onlinepubs/circulars/ec190.pdf>.
- Arras, L. et al. (2019) Explaining and Interpreting LSTMs, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11700 LNCS(2019), pp. 211–238. DOI: 10.1007/978-3-030-28954-6_11.

- Brilon, W., Geisfeldt, J. and Regler, M. (2005) Reliability of Freeway Traffic Flow: A stochastic Concept of Capacity, *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, (July), pp. 125–144.
- Caleffi, F. et al. (2016) Influência das condições climáticas e de acidentes na caracterização do comportamento do tráfego em rodovias, *Transportes*, 24(4), p. 57. DOI: 10.14295/transportes.v24i4.1104.
- Caleffi, F. (2018) *Proposição de um método de harmonização da velocidade baseado em modelo de previsão de conflitos veiculares*. PhD Thesis. Departamento de Engenharia de Produção e Transportes: Universidade Federal do Rio Grande do Sul. Porto Alegre, Brazil.
- Chollet, F. (2018) *Deep Learning with Python*. 1st edn. Shelten Island, NY: Manning Publications.
- Eleftheriadou, L. et al. (2011) Proactive ramp management under the threat of freeway-flow breakdown, *Procedia - Social and Behavioral Sciences*, 16, pp. 4–14. DOI: 10.1016/j.sbspro.2011.04.424.
- Eleftheriadou, L. et al. (2014) Enhancing ramp metering algorithms with the use of probability of breakdown models, *Journal of Transportation Engineering*, 140(4), pp. 1–9. DOI: 10.1061/(ASCE)TE.1943-5436.0000653.
- Fortunato, M., Blundell, C. and Vinyals, O. (2017) Bayesian Recurrent Neural Networks, pp. 1–14. DOI: 10.48550/arXiv.1704.02798.
- Fu, R., Zhang, Z. and Li, L. (2017) Using LSTM and GRU neural network methods for traffic flow prediction, *Proceedings - 2016 31st Youth Academic Annual Conference of Chinese Association of Automation, YAC 2016*, pp. 324–328. DOI: 10.1109/YAC.2016.7804912.
- Gal, Y. and Ghahramani, Z. (2016) A theoretically grounded application of dropout in recurrent neural networks, *Advances in Neural Information Processing Systems*, pp. 1027–1035. DOI: 10.48550/arXiv.1512.05287.
- Galvan, Y.T., Zechin, D. and Cybis, H.B.B. (2019) Utilização do método Kaplan-Meier na estimativa das distribuições de velocidade desejada, *Anais do 33º Congresso de Pesquisa e Ensino em Transportes - 2019*. Rio de Janeiro: Associação Nacional de Pesquisa e Ensino em Transportes.
- Géron, A. (2019) *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. 2nd edn. Sebastopol: O'Reilly Media
- Gu, Y. et al. (2019) Short-term prediction of lane-level traffic speeds: A fusion deep learning model, *Transportation Research Part C: Emerging Technologies*, 106(July), pp. 1–16. DOI: 10.1016/j.trc.2019.07.003.
- Han, Y. and Ahn, S. (2018) Stochastic modeling of breakdown at freeway merge bottleneck and traffic control method using connected automated vehicle, *Transportation Research Part B: Methodological*, 107, pp. 146–166. DOI: 10.1016/j.trb.2017.11.007.
- Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory, *Neural Computation*, 9(8), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
- Iung, B. (2013) Cœur et grosseesse, *EMC - Traité de médecine AKOS*, 8(2), pp. 1–4. DOI: 10.1016/s1634-6939(13)59289-1.
- Kendall, A. and Gal, Y. (2017) What uncertainties do we need in Bayesian deep learning for computer vision?, *Advances in Neural Information Processing Systems*, 2017-December(Nips), pp. 5575–5585. DOI: 10.48550/arXiv.1703.04977.
- Laptev, N. et al. (2017) Time-series Extreme Event Forecasting with Neural Networks at Uber, *International Conference on Machine Learning - Time Series Workshop*, pp. 1–5. Available at: <http://www.cs.columbia.edu/~lierranli/publications/TSW2017_paper.pdf>. Accessed at: 12/08/2022.
- Li, L. et al. (2018) Hyperband: A novel bandit-based approach to hyperparameter optimization, *Journal of Machine Learning Research*, 18, pp. 1–52. DOI: 10.48550/arXiv.1603.06560.
- Li, Y., Wang, N. and Carroll, R.J. (2010) Generalized functional linear models with semiparametric single-index interactions, *Journal of the American Statistical Association*, 105(490), pp. 621–633. DOI: 10.1198/jasa.2010.tm09313.
- Ma, X. et al. (2015) Long short-term memory neural network for traffic speed prediction using remote microwave sensor data, *Transportation Research Part C: Emerging Technologies*, 54, pp. 187–197. DOI: 10.1016/j.trc.2015.03.014.
- Ma, Y., Zhang, Z. and Ihler, A. (2020) Multi-Lane Short-Term Traffic Forecasting with Convolutional LSTM Network, *IEEE Access*, 8, pp. 34629–34643. DOI: 10.1109/ACCESS.2020.2974575.
- Neil, D., Pfeiffer, M. and Liu, S.C. (2016) Phased LSTM: Accelerating recurrent network training for long or event-based sequences, *Advances in Neural Information Processing Systems*, (Nips), pp. 3889–3897. DOI: 10.48550/arXiv.1610.09513.
- Pujol, B.S., Asplund, M. and Serra, J.C. (2019) *Machine learning for early detection of traffic congestion using public transport traffic data*. Bachelor Thesis. Department of Computer and Information Science Machine: Linköping University. Linköping, Sweden.
- Vlahogianni, E.I., Karlaftis, M.G. and 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. DOI: 10.1016/j.trc.2014.01.005.
- Zechin, D., Caleffi, F. and Cybis, H.B.B. (2020) Influence of Rain on Highway Breakdown Probability, *Transportation Research Record*, 2674(8), pp. 687–695. DOI: 10.1177/0361198120919754.