FORECASTING ROAD-TRANSPORTATION CO_2 EMISSION IN BAHIA STATE CITIES, BRAZIL, THROUGH ARTIFICIAL NEURAL NETWORK WITH OPEN-ACCESS DATA AND WAVELET TRANSFORM

Marcos Nobuaki Kawai[1,a,b], PhD. Professor Erick Giovani Sperandio Nascimento[2,a ,b]

[a], Centro Universitário SENAI CIMATEC; Salvador; Bahia; Brazil;
[b] Computing Modeling and Industrial Technology department

Abstract: Brazil committed to the Paris Agreement, targeting to reduce by 37% 2005 emissions level in 2025. Focusing on road-transportation CO_2 emissions, there is no previous paper that applied ANN with wavelet transformation to predict it for Bahia state cities. This research aims to develop an artificial neural network capable to forecast it through available open-access data and analyze its efficiency. Multilayer perceptron (MLP) and long short-term memory (LSTM) algorithms were evaluated. Wavelet transformation resulted in better metrics; it was applied in the final model. LSTM resulted in better metrics and is capable to predict up to 6 months ahead with NMSE of 0.93 and R^2 of 1.01. Future research would increase the dataset variables, analyze other ANN techniques and develop scenarios based on variable's correlation.

Keywords: CO_2; road-transportation; artificial neural network; wavelet

PROJETANDO AS EMISSÕES DE CO_2 POR TRANSPORTE RODOVIÁRIO PELAS CIDADES DO ESTADO DA BAHIA, BRAZIL, ATRAVÉS DE REDE NEURAL ARTIFICIAL UTILIZANDO DADOS PÚBLICOS E TRANSFORMATA WAVELET

Resumo: O Brasil se comprometeu com o Acordo de Paris a reduzir em 37% os níveis de emissões de 2005 em 2025. Focando em emissões de CO_2 por transporte rodoviário, não há nenhum artigo prévio que aplicou ANN com a transformata wavelet para predizer isso para as cidades do estado da Bahia. Essa pesquisa visa desenvolver uma rede neural artificial capaz de predizer isso através de dados públicos e analisar sua eficiência. MLP e LSTM foram avaliados. A transformata Wavelet resultou em melhores métricas; foi aplicada no modelo final. A LSTM resultou em melhores métricas e é capaz de predizer até 6 meses à frente com um NMSE de 0.93 e R^2 de 1.01. Pesquisas futuras poderão aumentar as variáveis, analisar outras técnicas e desenvolver cenários baseados na correlação das variáveis.

Palavras-chave: CO_2, transporte rodoviário, rede neural artificial; wavelet

1. INTRODUCTION

Oceans have been absorbing carbon dioxide (CO_2); however, high concentrations would negatively impact water temperature. In addition, another impact originated from increase of CO_2 concentration over the seas is the increase of water acidity, which has a negative impact in marine animals [1]. In April 2021 worldwide atmospheric CO_2 concentration level was almost 50% higher than the beginning of the industrial age led by humans' activities such as the combustion of fossil fuels [2].

Global governments actions to reduce it has been implemented over the years. As part of these actions, Brazil committed to the Paris Agreement targeting to reduce by 37% 2005 emissions level in 2025 [3].

Focusing on road-transportation sector (combustion of fossil fuels) and CO_2, a prediction tool would help to understand if this target would be achieved in this sector and pollutant. Studies related to air pollutants demonstrated that an artificial neural network is an efficient forecasting tool [4].

Cities play an important role in the CO_2 emission level and there is a correlation with economic growth [5], however, data availability to analyze it is a concern.

No previous paper applied ANN with wavelet transformation to predict road-transportation CO_2 emissions with open-access data in Bahia.

In this context, this research aims to develop an artificial neural network algorithm capable to perform road-transportation CO_2 emission forecast based on Bahia state cities open-access data. Data availability and ANN algorithm potentiality to perform this forecast will be analyzed. It could help government authorities to plan countermeasures actions.

2. METHODOLOGY

2.1 Dataset

The dataset was built based on available open-access data. It was structured for all Bahia state cities in a monthly basis. The input variables were the fuel consumption by fuel type, automotive fleet volume by category, population, gross domestic product and CO_2 emissions (Table 1).

Table 1. Algorithm Input Variables

Variable	Unit	Variable	Unit	Variable	Unit
year	year	**Gasoline Consumption**	$1{,}0.10^6$ L[a]	**Number of Cars**	$1{,}0.10^6$ units[d]
month	month	**Diesel Consumption**	$1{,}0.10^6$ L[a]	**Number of Motorcycle**	$11{,}0.10^6$ units[d]
day	day	**Ethanol Consumption**	$1{,}0.10^6$ L[a]	**Number of Light Commercial**	$1{,}0.10^6$ units[d]
City Code	number	**Population**	$1{,}0.10^6$ inhabitants[b]	**Number of Truck**	$1{,}0.10^6$ units[d]
CO_2	$1{,}0.10^6$ Kg	**GDP**	$1{,}0.10^9$ Reais[c]	**Number of Bus**	$1{,}0.10^6$ units[d]

a: ANP [6] source; b: IBGE [7] source; b: IBGE [8] source, d: DENATRAN [9] source (registered vehicles)

CO_2 emissions were calculated (1) based on the top-down methodology which is an estimate considering the fuel demand in a specific geographic area, in a specific period of time and applying a fuel emission factor.

$$CO_2 = \sum_{f=1}^{F} Ef_f * D_f \quad (1)$$

In the equation 1, CO_2 is the total city emission in Kg in the specified month, *f* is the fuel type, *Ef* is the fuel type emission factor and *D* is the fuel type demand.

The fuel emission factor (Table 2) was based on CETESB 2019 São Paulo Road-Transportation Emissions Inventory [10].

Table 2. CETESB CO_2 Emission Factor (Ef) by Fuel (Kg/L)

Period	Gasoline	Anhydrous / Hydrate Ethanol	Diesel
2003-2018	2.212	1.526	2.603

With all variables combined, a threshold of 75% of minimum available data among the period was adopted. Cities that did not comply with it were removed from the analysis (Figures 1, 2, 3). The investigation period was restricted from 2003 to 2018 calendar years because of data availability.

ISSN: 2357-7592

Figure 1. Analyzed Bahia Cities

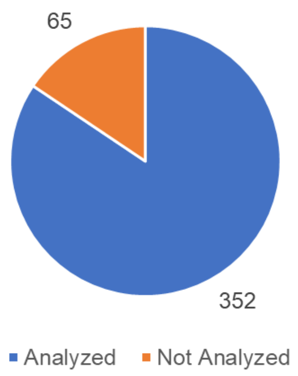


Figure 2. Analyzed Bahia Population

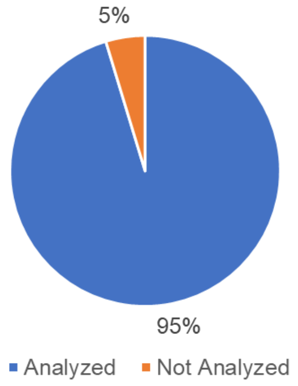


Figure 3. Not Analyzed Population (inhabitants)

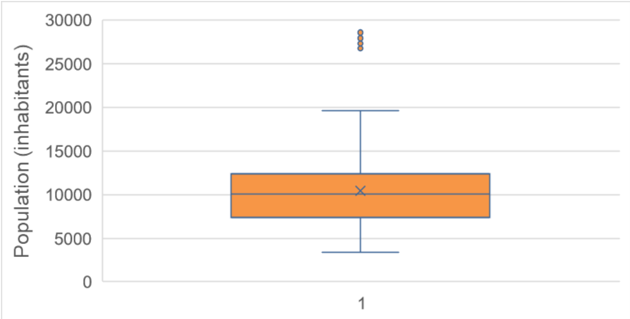


Figure 1 indicates that 65 cities out of 417 total Bahia cities did not have data to be analyzed. It corresponds to 5% (Figure 2) of total Bahia population and are cities with lower than 30000 inhabitants (Figure 3).

The dataset was split in training (49%), validation (21%), and test (30%) to analyze model's performance. Using the model that resulted in better metrics, a final run with a new dataset split, training (70%) and test (30%) was performed to get the final metrics.

2.2 Model Development

Discrete wavelet transformation [11] captures frequency and location information and may improve this ANN time series algorithm forecast metrics. Wavelet packed decomposition was applied in 5 levels in the dataset to evaluate its impact. The lowest root mean squared error (RMSE) was adopted to select the wavelet signal for each variable type (Table 3).

Table 3. Selected Discrete Wavelet Signal by Variable

Variable	Wavelet Signal	Variable	Wavelet Signal
CO_2	bior1.1	**Number of Cars**	bior1.1
Gasoline Consumption	bior1.1	**Number of Motorcycle**	bior1.1
Diesel Consumption	bior1.1	**Number of Light Commercial**	bior1.1
Ethanol Consumption	db21	**Number of Truck**	bior1.1
Population	bior1.1	**Number of Bus**	rbio2.2
GDP	bior1.1		

Two artificial neural network techniques were evaluated: Multilayer Perceptron (MLP) is a traditional feedforward ANN and Long Short-Term Memory (LSTM) is an artificial recurrent neural network. To define the model structure, up to 5 layers combination were evaluated. Three LSTM structure models were evaluated: Vanilla, Stacked and Bidirectional (Table 4).

Table 4. Layers Combination Evaluation

Model	MLP	LSTM (Vanilla, Stacked, Bidirectional)
Batch Normalization	X	X
Dropout	X	X
Hidden Layers	X	
Time Distributed		X
Weight Regularization	X	X

Bayesian Optimization [12] is a tool to support machine learning hyperparameters tuning. It was applied to support the algorithm's model structure definition and refine its learning hyperparameters.

2.2 Model Evaluation

The metrics to evaluate the algorithm's performance were the Root Mean Squared Error (RMSE), Normalized Root Mean Squared Error (NRMSE) and the coefficient of determination (R^2).

12 months prediction forecast was adopted to evaluate each model metrics month by month.

3. RESULTS AND DISCUSSION

MLP model layers that resulted in better metrics: 2 hidden layers.

LSTM model layers that resulted in better metrics: 1 LSTM (Vanilla), 1 dropout, 1 TimeDistributed.

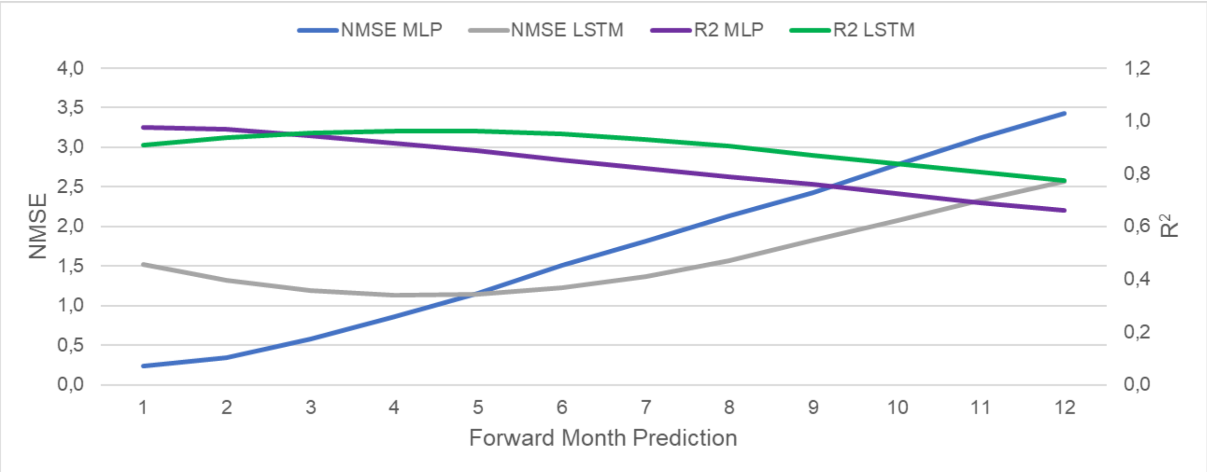
MLP and LSTM average 12 months forecast metrics are represented on (Table 5). It indicates that wavelet level 5 (WL5) resulted in the best metric result.

Table 5. MLP & LSTM - 12 months average forecast metrics

Type	RMSE MLP	NMSE MLP	R² MLP	RMSE LSTM	NMSE LSTM	R² LSTM
Without Wavelet	4,400	1,924	0,809	4,135	1,699	0,809
Wavelet Level 1	4,341	1,873	0,814	4,187	1,743	0,814
Wavelet Level 2	4,387	1,913	0,810	4,078	1,653	0,810
Wavelet Level 3	4,466	1,983	0,803	4,023	1,609	0,803
Wavelet Level 4	4,464	1,981	0,803	4,577	2,082	0,803
Wavelet Level 5	***4,133***	***1,697***	***0,831***	***4,020***	***1,606***	***0,831***

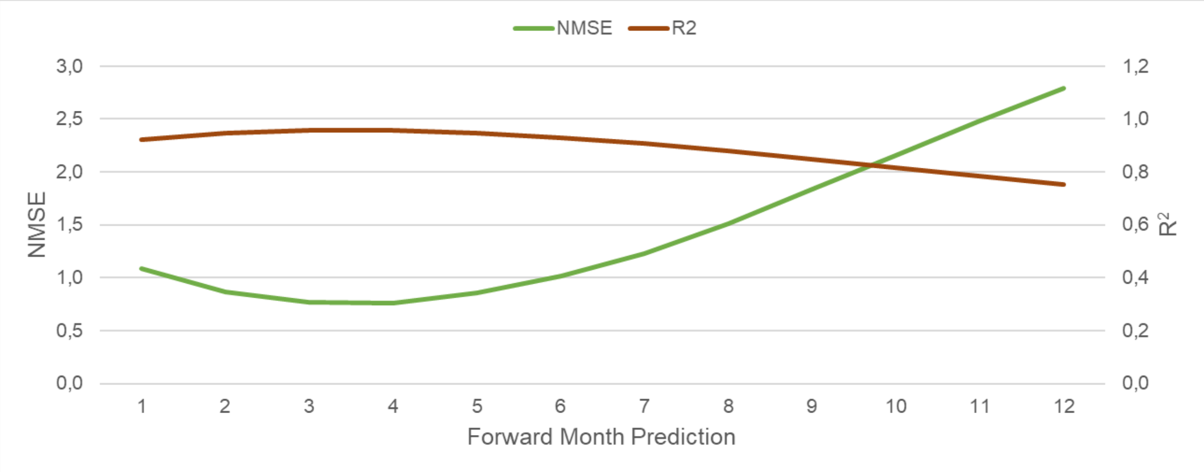
A combination of NMSE lower than 1 and R^2 around 1 should be pursued. Therefore, (Figure 4) indicates that both models could predict up to 5 months with the available dataset, nevertheless LSTM resulted in better metrics combination.

Figure 4. MLP & LSTM, WL5, Monthly Predictions - NMSE and R^2 Metrics



LSTM was adopted to perform a final run with an increased training dataset (70% training and 30% test). Metrics result are on (Figure 5).

Figure 5. LSTM, WL5, Monthly Predictions Metrics with Increased Training Dataset



It indicates that LSTM, with the available dataset and wavelet transformation level 5, is able to predict up to 6 months Bahia state cities road-transportation CO_2 emissions with an NMSE of 0.93 and R^2 of 1.01.

4. CONCLUSION

Bahia state cities open-access data can be used to build an ANN to forecast road-transportation CO_2 emissions limited to 84% of the total cities and to the period of 2003 to 2018 years.

Discrete wavelet transformation resulted on prediction metrics combination improvements. Therefore, the recommendation is to apply it.

LSTM Vanilla with TimeDistributed is capable to predict Bahia state cities road-transportation CO_2 emissions with a prediction window up to 6 months.

To improve the algorithm's forecast window capability, additional data might be considered such as fleet age and unemployment rate and/or increase the sample size (e.g.: region or country) as well as evaluate other ANN techniques.

This algorithm is not capable to predict CO_2 emissions up to 2025. Further studies might consider use ANN variables correlation to layout scenarios to 2025.

ISSN: 2357-7592

5. REFERENCES

[1] RIEBESELL, Ulf et al. Enhanced biological carbon consumption in a high CO 2 ocean. **Nature**, v. 450, n. 7169, p. 545-548, 2007.

[2] TANHUA, Toste et al. A vision for FAIR ocean data products. **Communications Earth & Environment**, v. 2, n. 1, p. 1-4, 2021.

[3] KASSNAYER, K.; NETO, H. J. F. A entrada em vigor do Acordo de Paris: o que muda para o Brasil. **Textos para Discussão. Núcleo de Estudos e Pesquisa da Consultoria Legislativa. Senado Federal**, v. 215, 2016.

[4] LIU, Bingchun et al. Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model. **Atmospheric Pollution Research**, p. 101144, 2021.

[5] LI, Ying; SUN, Yanwei. Modeling and predicting city-level CO 2 emissions using open access data and machine learning. **Environmental Science and Pollution Research**, v. 28, n. 15, p. 19260-19271, 2021.

[6] ANP - Agência Nacional do Petróleo, Gás Natural e biocombustíveis. **Arquivos – Dados abertos**. Rio de Janeiro. Available at: <https://www.gov.br/anp/pt-br/centrais-de-conteudo/dados-abertos/vendas-de-derivados-de-petroleo-e-biocombustiveis>. Accessed on: 17 Apr. 2021.

[7] IBGE, Censo. Instituto Brasileiro de Geografia e Estatística-IBGE. 2000. **Estimativas da População**. Available at: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=downloads>. Accessed on: 06 Apr. 2021.

[8] IBGE, Censo. Instituto Brasileiro de Geografia e Estatística-IBGE. 2000. **Arquivos – Sistema de Contas Nacionais Trimestrais**. Available at: <https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9300-contas-nacionais-trimestrais.html?=&t=resultados&utm_source=landing&utm_medium=explica&utm_campaign=pib#evolucao-taxa>. Accessed on: 15 May. 2021.

[9] DEPARTAMENTO NACIONAL DE TRÂNSITO (DENATRAN). Frota de veículos. 2017. **Estatísticas Frotas de Veículos**. Available at: <https://www.gov.br/infraestrutura/pt-br/assuntos/transito/conteudo-denatran/estatisticas-frota-de-veiculos-denatran>. Accessed on: 06 Apr. 2021.

[10] CETESB - Companhia Ambiental do Estado de São Paulo. **Relatório de Emissões Veiculares no Estado de São Paulo, 2019**. São Paulo, 2020. Available at: <https://cetesb.sp.gov.br/veicular/wp-content/uploads/sites/6/2020/11/Relatorio-Emissoes-Veiculares-no-Estado-de-Sao-Paulo-2019.pdf>. Accessed on: 08 Aug. 2021.

[11] ZUCATELLI, Pedro Junior et al. Nowcasting prediction of wind speed using computational intelligence and wavelet in Brazil. **International Journal for Computational Methods in Engineering Science and Mechanics**, v. 21, n. 6, p. 343-369, 2020.

[12] SNOEK, Jasper; LAROCHELLE, Hugo; ADAMS, Ryan P. Practical bayesian optimization of machine learning algorithms. **Advances in neural information processing systems**, v. 25, 2012.

ISSN: 2357-7592