PM10 POLLUTANT PREDICTION WITH N-BEATS AND N-HITS: AN ANALYSIS OF THE CAPACITY OF NEURAL MODELS IN ESTIMATING THE AIR QUALITY INDEX

GJV MEIRA FILHO[1]

[1]Maringá State University, Chemical Engineering Department
Contact e-mail: pg406359@uem.br

ABSTRACT – Forecasting air quality indicators, such as PM10, is critical for public health and environmental management, particularly within chemical engineering, where emissions control, process safety, and environmental compliance are central concerns. Anticipating PM10 levels supports informed decision-making in industrial operations, urban planning, and regulatory policy, helping to minimize pollutant exposure and optimize mitigation strategies. This study evaluates the performance of two neural forecasting architectures – N-BEATS and N-HiTS – for predicting PM10 levels at the Tietê Riverbank in São Paulo, Brazil, using daily AQI data from 2016 to 2022. The models were optimized using Tree-Structured Parzen Estimators across various hyperparameters, including stack configuration, transformation methods, and kernel parameters. Performance was assessed across 30-, 60-, and 90-day horizons using five metrics: MAE, MAPE, MSE, sMAPE, and MAX. Results show that both N-BEATS and N-HiTS consistently ranked among the top models for MAPE, demonstrating superior adaptability and generalization compared to traditional statistical methods. While they did not outperform in all metrics – particularly in capturing peak values and complex fluctuations – their stability and interpretability suggest strong potential for broader application in environmental forecasting. These findings support neural models as competitive, scalable alternatives for air quality prediction, with advantages in long-term performance and model tuning flexibility.

1 INTRODUCTION

1.1 Particulate Matter

The Air Quality Index (AQI) is a globally recognized metric used to indicate ambient air pollution levels. It is divided into six distinct categories, designed to convey the potential health risks associated with specific pollutant concentrations. The classification for particulate matter with a diameter greater than 10 µm (PM$_{10}$) is presented in Table 1. In categories one and two, the vast majority of the population is unlikely to experience significant health effects. From category three onward, adverse symptoms become increasingly common (Baumbach, 1996).

Particulate matter originates from a wide range of sources, with industrial combustion being one of the primary contributors. These particles, which are on average ten times smaller than a grain of sand, may consist of ash, metal oxides, fuel residues, soot, and other substances. Among natural sources, volcanic activity is considered the most significant, while industrial processes are the leading anthropogenic source. However, attention must also be given to the rising levels of particulate matter resulting from deforestation, desertification, and the potential atmospheric consequences of a nuclear winter (Baumbach, 1996).

Tabel 1: Air quality index (AQI) classification and PM_{10} concentration ranges.

Index Values	Levels of Concern	Concentration / (μg/m^3)
0 to 50	Good	0 to 54
51 to 100	Moderate	55 to 154
101 to 150	Unhealthy for Sensitive Groups	155 to 254
151 to 200	Unhealthy	255 to 354
201 to 300	Very Unhealthy	355 to 424
Above 301	Hazardous	Above 425

The World Health Organization (WHO) has classified all particulate matter (PM) as a Group 1 carcinogen, capable of penetrating the respiratory system and contributing to severe cardiovascular and respiratory diseases (Kim, 2017). In line with this, a study by Bosco et al. (2005) demonstrated that a petrochemical plant in Gela, Italy, significantly contributed to elevated PM levels, particularly concerning trace metals such as Mo, Ni, V, Pd, Se, and As.

1.2 N-BEATS Model

Developed by Oreshkin *et al*. (2020), the Neural Basis Expansion Analysis for Time Series (N-BEATS) is a neural network architecture specifically designed for time series forecasting. The model receives T temporal observations as input and produces forecasts for H future time steps. Its architecture consists of a sequence of stacks, where the output of each stack contributes additively to the final forecast. Additionally, the "filtered output" of each stack s is used as the input (lookback window) for the subsequent stack, enabling iterative refinement of the prediction.

Within each stack, a series of blocks ℓ composed of four fully connected layers compute the forward and backward expansion coefficients θ. These coefficients are then passed to the basis layer g, which uses them to generate both the backcast (reconstruction of the input, with index b) and the forecast (prediction of future values, with index f). The g functions can take on various interpretable forms, such as a trend form – exhibiting linear behavior; a seasonal form – with periodic trigonometric behavior; or a generic form – representing a black-box model with no explicit interpretability (Oreshkin *et al*., 2020).

1.3 N-HiTS Model

The N-HiTS model was introduced by Challu *et al*. (2022) as a more efficient and general-purpose neural forecasting architecture built upon the foundation of N-BEATS. Its key innovations include the incorporation of a Max Pooling layer before each block's fully connected neural network, with a kernel size of k, and a hierarchical non-linear interpolation mechanism at the output stage to assign values to each coefficient.

Through multi-rate signal sampling and hierarchical interpolation, the model can capture various frequency components within the complex structure of a time series, enhancing performance by allowing each stack to specialize in a distinct frequency – corresponding to a known cyclical component – while maintaining control over the number of parameters. The residuals from each stack are subsequently subtracted from the backcast input of the next stack. The N-HiTS model significantly reduces computational complexity as the forecasting horizon increases, especially when compared to the interpretable and general versions of N-BEATS (N-BEATS-I and N-BEATS-G, respectively) (Challu *et al.*, 2022).

2 METHODOLOGY

Data from the Air Quality Historical Data Platform – available at AQICN (2024) – was used to access daily AQI PM_{10} data from April 1st, 2016, up to December 31st, 2022, at the Tietê Riverbank (Remedy Bridge), São Paulo, Brazil. All missing data was filled using a Bernstein basis piecewise polynomial to promote continuity and smoothness at the breakpoints.

The programming language used was Python 3.12.0, along with the following libraries: numpy==2.0.0, pandas==2.2.3, torch==2.5.1+cu124, neuralforecast==1.7.6, statsforecast==2.0.0, and optuna==4.1.0. Reproducibility is ensured by setting the random seed to 1 for all stochastic processes. The four selected forecasting horizons were 30, 60, and 90 days, which were chosen to capture monthly patterns throughout the period.

For the weekly time series, distinct time windows and forecasting horizons were defined, with all four selected horizons utilizing a total of 2466 historical observations for both training and testing. To evaluate the accuracy and compare the forecasting methods, the following metrics were used: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Maximum Error (MAX), and Symmetric Mean Absolute Percentage Error (sMAPE). For comparison, 11 distinct statistical methods were evaluated on the same forecasting periods and dataset, including AutoCES, AutoETS, and AutoTheta, along with their respective main variations.

For the neural forecasters, the best-performing models underwent hyperparameter optimization. The optimized parameters included input size (integers between 90 and 1100), number of stacks (ranging from 2 to 7 for N-BEATS and 3 to 7 for N-HiTS), maximum number of training epochs (10 to 700), number of blocks per stack (1 to 7), and the input data preprocessing method (none, standardization, Box-Cox transformation, or Min-Max normalization).

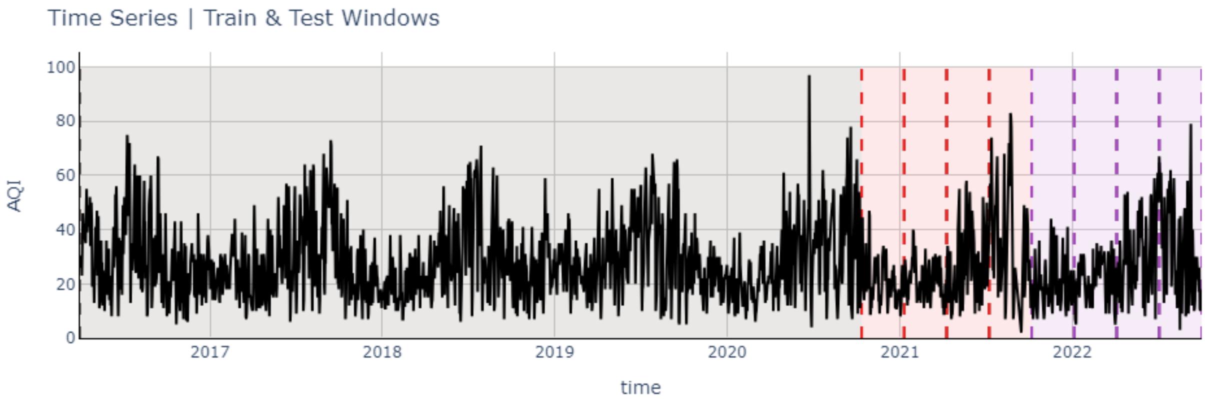
Additionally, for N-HiTS, the kernel size was optimized using permutations of the set [1, 2, 3] for the final three stacks, while earlier stacks did not use kernels, and the down sampling used a permutation of [1, 7, 90, 180, 365] applied only for the first three stacks, while the remainder used unitary frequency. Other hyperparameters were fixed according to the recommendations by Challu et al. (2022), including the kernel pooling method (MaxPool1D), activation function (ReLU), and interpolation method (Linear). Moreover, N-BEATS exclusive parameters optimized was the interpretability, which corresponded to the selection of two out of three g functions – trend, seasonal, or generic [T, S, G] – for the first two stacks.

The optimization was performed over 200 trials for each forecast horizon and neural model. The objective function primarily minimized the symmetric Mean Absolute Percentage Error (sMAPE) – due to its scale-independence and balanced treatment of over and under-predictions – with all trial data recorded for subsequent analysis.

To ensure the validity of the tests, a 4-fold cross-validation was used, ranging from the most recent data point to four times the value of the forecasting horizon, with no overlap. To further prevent overfitting during optimization, an additional 4-fold cross-validation was performed with no temporal overlap, extending up to eight times the value of the forecasting horizon, stopping just before the data used in the test set. This ensured that no test data was used in the training process. A representation of this is shown in Figure 1.

The optimization employed Tree-Structured Parzen Estimators (TPE) to enhance the sampling strategy and facilitate efficient hyperparameter tuning. The optimal combination of hyperparameters was subsequently applied in the final test evaluation, enabling a comparative analysis of the performance metrics.

Figure 1: AQI for PM10 Time Series with Red Indicating the Optimization K-Folds for Cross-Validation and Purple Representing the Test Sets.



3 RESULTS AND DISCUSSION

Based on the metrics, the performance of the optimized models was evaluated across 4 distinct k-folds for cross-validation. Table 2 presents the results of both neural forecasters, considering all horizons and metrics, ranked such as 1 represents the best model and 13 represents the worst. Both models performed well in terms of MAPE across all studied horizons, with N-BEATS in the top position and N-HiTS in second place for the 30 and 90 horizons, and in fourth place for the 60 horizon.

Tabel 2: Relative ranking of performance among 13 forecasting models based on the average of each metric for 4 k-fold cross-validations in the test set.

Method	H	smape	mae	mape	mse	max
NHITS	30	5	6	2	5	6
	60	5	5	4	2	3
	90	5	5	2	6	5
NBEATS	30	6	5	**1**	3	7
	60	6	6	**1**	6	6
	90	6	6	**1**	5	6

Although the models outperformed the statistical methods in terms of MAPE, the other metrics indicate potential shortcomings in prediction. The MAX's (metric that helps evaluate a model's ability to handle extreme pollution events) intermediate position in the ranking demonstrates that the models were not highly sensitive to capturing peaks or even outliers, which could be attributed to the high complexity of the series' patterns. The sMAPE and MAE in intermediate positions suggest that the neural models achieved a certain degree of accuracy but could still not fully capture both seasonal and more complex patterns during testing.

The hyperparameters of the optimized models are presented in Table 3 for both models studied and their respective horizons. The optimization of the N-HiTS model selected large kernels as well as downsampling values, indicating the model's effort to capture the influence of strong autocorrelation in future data. Similarly, all horizons of N-BEATS included a stack with interpretable activation functions, which similarly suggests that interpretability was important for both models in improving their results.

Tabel 3: Final hyperparameter configuration for deployed NHITS and NBEATS Models.

Method	H	input size	n stacks	n blocks	max steps	transform	kernel size	down sampling	Interpret.	Trial
NHITS	30	570	6	5	92	minmax	[2, 3, 2]	[90, 1, 90]	–	165
	60	288	3	5	35	–	[2, 2, 3]	[1, 365, 365]	–	78
	90	224	3	2	389	–	[3, 2, 3]	[90, 180, 90]	–	171
NBEATS	30	527	5	1	155	minmax	–	–	[S, G]	2
	60	189	3	2	61	–	–	–	[S, G]	174
	90	162	2	5	101	boxcox	–	–	[G, T]	143

A deeper investigation into each trial of the optimizer reveals that, although sMAPE was minimized in fewer iterations, the MAX reached its lowest value in the final iterations. This strongly indicates that additional trials could yield better results for other metrics. This conclusion is strongly supported by the optimization of the 30 horizon in N-BEATS, where the best sMAPE value was randomly found in the second trial, but the best MAX values were observed in the final iterations.

Tabel 4: Performance among neural forecasters (white), best (green) and worst (red) statistical forecasting models based on the average of each metric for 4 k-fold cross-validations in the test set.

Method	H	smape	mae	mape	mse	max
NHITS	30	44.834	9.875	56.368	156.762	26.179
	60	44.140	11.948	65.913	208.412	34.111
	90	41.385	10.713	55.688	197.568	33.510
NBEATS	30	45.001	9.715	55.299	146.006	26.371
	60	47.201	12.577	61.950	264.730	37.261
	90	43.470	10.891	54.387	195.129	34.314
AutoCES-S	30	40.168	9.384	59.781	144.524	25.880
	60	41.525	11.365	62.820	201.952	34.199
	90	39.527	10.183	58.052	168.639	30.897
AutoCES-Z	30	39.897	9.551	58.730	159.578	25.186
	60	41.653	11.668	66.290	224.424	36.015
	90	40.574	10.588	61.097	188.723	32.533
AutoETS-ZAA	30	39.066	9.117	57.018	141.132	24.659
	60	41.740	11.532	63.799	211.610	32.654
	90	40.175	10.336	56.824	178.095	29.999
AutoCES-P	30	99.652	15.006	75.595	338.913	34.005
	60	103.907	19.052	77.785	527.176	47.721
	90	131.246	21.032	88.360	602.003	52.274
AutoETS-ZAZ	30	80.651	34.105	221.477	1766.127	69.068
	60	104.816	72.891	399.056	7308.747	146.891
	90	120.006	101.143	560.424	14088.676	206.469
AutoETS-ZAN	30	80.620	34.081	221.337	1764.167	69.017
	60	105.924	74.397	404.529	7536.920	149.449
	90	121.118	103.600	569.761	14701.104	210.885

Table 4 presents the average metrics obtained after four conclusive k-fold cross-validation rounds for both neural forecasters, as well as for the three best and three worst statistical methods. For sMAPE – the optimized metric – the lowest standard deviations for horizons of 30, 60, and 90 are 10.72, 4.45, and 6.52, respectively, all of which correspond to the neural forecasters. This indicates that, although the neural forecasters did not outperform statistical methods in terms of the optimized metric only, their performance is still comparable to that of the best models. This is not the case for the worst-performing methods, whose errors are disproportionately higher than those of the neural forecasters.

The best-performing statistical methods proved suitable for forecasting multiple horizons of the PM10 time series. However, this may not be held for other pollutants or forecast horizons. In contrast, N-HiTS and N-BEATS – being adaptable neural models – have the potential to deliver consistently strong results across time series with varying characteristics, while remaining competitively aligned with the top-performing statistical methods. Further analysis involving other types of autoregressive structures with different properties is necessary to confirm this generalizability.

4 CONCLUSION

Overall, the findings demonstrate that both N-HiTS and N-BEATS provide robust forecasting performance across all evaluated horizons, particularly concerning MAPE, where they consistently outperform traditional statistical methods. While they do not lead across all evaluation metrics – such as sMAPE and MAX – their adaptability to varying temporal structures and the complexity of the PM10 time series suggest strong generalization capabilities.

The stability observed in cross-validation results, along with interpretable and well-optimized configurations, underscores the models' suitability for environmental forecasting tasks. These results support the continued investigation of neural forecasting architectures as competitive and scalable alternatives to classical methods in air quality prediction. Therefore, neural methods can be effectively applied in the context of PM10 prediction, as they yield strong results and are competitively comparable to the best-performing statistical methods, while consistently outperforming average and weaker approaches.

5 REFERENCES

AQICN. *Historical Air Quality Data*. Available at: https://aqicn.org/historical/. Accessed on: 27 dec. 2024.

BAUMBACH G, *Air Quality Control: Formation and Sources, Dispersion, Characteristics and Impact of Air Pollutants - Measuring Methods, Techniques for Reduction of Emissions and Regulations for Air Quality Control*. Berlin; Heidelberg: Springer-Verlag, 1996.

CHALLU C. et al. N-HiTS: Neural Hierarchical Interpolation for Time Series Forecasting. 30 jan. 2022. *arXiv*. https://arxiv.org/abs/2201.12886

KIM J, Assessment and estimation of particulate matter formation potential and respiratory effects from air emission matters in industrial sectors and cities/regions. *Journal of Korean Society of Environmental Engineers*, [S.I.], v. 39, p. 220-228, 2017.

OREShKIN BN, et al. N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting. 2019. *arXiv*. https://arxiv.org/abs/1905.10437