

MARINGÁ STATE UNIVERSITY  
TECHNOLOGY CENTER  
UNDERGRADUATE COURSE IN CHEMICAL ENGINEERING

Gustavo Junqueira Valias Meira Filho  
João Marcos Scarparo

**EVALUATING NEURAL FORECASTERS (N-BEATS AND N-HITS) FOR AIR  
QUALITY INDEX PREDICTION**

Maringá - PR  
2024

MARINGÁ STATE UNIVERSITY  
TECHNOLOGY CENTER  
UNDERGRADUATE COURSE IN CHEMICAL ENGINEERING

Gustavo Junqueira Valias Meira Filho  
João Marcos Scarparo

**EVALUATING NEURAL FORECASTERS (N-BEATS AND N-HITS) FOR AIR  
QUALITY INDEX PREDICTION**

Completion of Undergraduate Coursework presented to the  
Undergraduate Course in Chemical Engineering at the State  
University of Maringá as part of the requirements for  
obtaining the title of Chemical Engineer.

Advisor: Dr. Leandro Vitor Pavão

Maringá - PR  
2024

Gustavo Junqueira Valias Meira Filho

João Marcos Scarparo

**EVALUATING NEURAL FORECASTERS (N-BEATS AND N-HITS) FOR AIR  
QUALITY INDEX PREDICTION**

Graduation course completion work presented to the Undergraduate Course in Chemical Engineering at the State University of Maringá as part of the requirements for obtaining the title of Chemical Engineer.

Approved in: \_\_\_\_ de \_\_\_\_ de \_\_\_\_.

**EXAMINATION BOARD**

---

Paulo Cesar Ossani - CCE/UEM

---

Carlos Henrique Vassoler - DEQ/UEM

---

Leandro Vitor Pavão – DEQ/UEM (advisor)

## ABSTRACT

Air Quality Index (AQI) is a crucial metric that standardizes various atmospheric pollutants into a single scale, reflecting the potential health risks associated with air pollution. Accurate forecasting of AQI is essential for public health and environmental management. Recently, machine learning and deep learning techniques have gained prominence in time series forecasting due to their ability to generalize across datasets and capture intricate temporal patterns beyond the capabilities of traditional statistical methods, which primarily rely on trends, seasonality, and residual analysis. Among these advanced techniques, N-BEATS and N-HiTS are state-of-the-art neural forecasting models that show significant promise for AQI prediction. This study evaluates their performance against classical statistical methods using forecasting metrics and analyzing their sensitivity to outliers. The findings indicate that air pollutants exhibit highly complex temporal behaviors, and neural forecasters often performed competitively with, or even outperformed, statistical approaches in many scenarios. However, a notable challenge with neural forecasting models is the opaque nature of their hyperparameter optimization processes, which can complicate their interpretation and fine-tuning. This requirement for extensive optimization introduces inefficiencies in terms of time and resources. Despite these challenges, neural forecasting models represent a significant innovation in time series analysis, offering a promising avenue for improving AQI prediction. Future work should focus on refining these models, exploring novel combinations, and enhancing their accuracy and efficiency to better serve the needs of AQI forecasting and broader time series applications.

Keywords: N-BEATS; N-HiTS; Neural forecasting; Air quality index; Time series.

## **RESUMO**

O Índice de Qualidade do Ar (AQI) é uma métrica crucial que padroniza diversos poluentes atmosféricos em uma única escala, refletindo os potenciais riscos à saúde associados à poluição do ar. A previsão precisa do AQI é essencial para a saúde pública e a gestão ambiental. Recentemente, técnicas de aprendizado de máquina e aprendizado profundo têm ganhado destaque na previsão de séries temporais devido à sua capacidade de generalizar entre conjuntos de dados e capturar padrões temporais intrincados, além das capacidades dos métodos estatísticos tradicionais, que se baseiam principalmente em tendências, sazonalidade e análise de resíduos. Entre essas técnicas avançadas, os modelos de previsão neural N-BEATS e N-HiTS são os mais avançados e mostram um potencial significativo para a previsão do AQI. Este estudo avalia seu desempenho em comparação com métodos estatísticos clássicos, utilizando métricas de previsão e analisando sua sensibilidade a outliers. Os resultados indicam que os poluentes atmosféricos apresentam comportamentos temporais altamente complexos, e os preditores neurais frequentemente tiveram um desempenho competitivo ou até superaram as abordagens estatísticas em muitos cenários. No entanto, um desafio notável com os modelos de previsão neural é a natureza opaca dos processos de otimização de hiperparâmetros, o que pode complicar sua interpretação e ajuste fino. Essa necessidade de otimização extensa introduz ineficiências em termos de tempo e recursos. Apesar desses desafios, os modelos de previsão neural representam uma inovação significativa na análise de séries temporais, oferecendo um caminho promissor para melhorar a previsão do AQI. Trabalhos futuros devem se concentrar no refinamento desses modelos, na exploração de combinações inovadoras e no aprimoramento de sua precisão e eficiência para melhor atender às necessidades da previsão do AQI e de aplicações mais amplas de séries temporais.

**Palavras-chave:** N-BEATS; N-HiTS; Previsão neural; Índice de qualidade do ar; Séries temporais.

## SUMMARY

<b>1 INTRODUCTION</b>	<b>6</b>
<b>2 THEORETICAL BACKGROUND</b>	<b>7</b>
2.1 Air Pollution	7
2.1.1 Air Quality Index (AQI)	10
2.1.2 O <sub>3</sub>	11
2.1.3 NO <sub>x</sub>	13
2.1.4 CO	18
2.1.5 Particulate Matter	19
2.2 Forecasting	20
2.2.1 Time Series Properties	20
2.2.2 State-of-the-Art	22
2.2.3 Difficulties	25
2.3 Neural Forecasters	26
2.3.1 Introduction to Neural Networks	26
2.3.2 Applied Neural Networks in Forecasting	28
2.3.3 N-BEATS	30
2.3.4 N-HiTS	33
<b>3 METHODOLOGY</b>	<b>37</b>
<b>4 RESULTS AND DISCUSSION</b>	<b>41</b>
4.1 Overview of Forecasting Performance Across Methods	41
4.2 Pollutant-Specific Forecasting Performance	42
4.3 Horizon Effects and Impact of Hyperparameters on Neural Forecasters	44
4.4 Metric-Specific Analysis	47
4.5 Neural vs. Statistical Forecasting Capacity	50
4.6 Implications for Air Quality Forecasting	51
<b>5 CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK</b>	<b>53</b>
<b>REFERENCES</b>	<b>54</b>
<b>APPENDIX A – STATISTICAL METHODS</b>	<b>56</b>
A.1 ARIMA	56
A.2 Error Trend Seasonality	56
A.3 Theta	58
A.4 Complex Exponential Smoothing	59
<b>APPENDIX B – FANOVA</b>	<b>61</b>
<b>APPENDIX C – RESULTS TABLES</b>	<b>62</b>
<b>APPENDIX D – TIME SERIES PROPERTIES</b>	<b>67</b>
<b>APPENDIX E – HYPERPARAMETERS IMPORTANCE</b>	<b>73</b>

## 1 INTRODUCTION

The Air Quality Index (AQI) is an extremely relevant measure for society as it provides a standardized scale for various atmospheric pollutants, alerting the population to the potential harm these pollutants may cause at any given time. Consequently, it is in the public interest to predict the damage caused by pollutants in advance, fostering awareness and social preparedness, particularly among public agencies and healthcare professionals, to mitigate the consequences of air pollution and to highlight the detrimental effects each pollutant can have on human health and the common good. While many studies have addressed this topic, significant gaps remain, particularly in the Brazilian context.

All pollutants exhibit time series with highly complex behavioral patterns that are challenging to capture using traditional statistical or mathematical models. Although some methods yield satisfactory results, they often fall short of adequately modeling these behaviors. As an investigative alternative, machine learning and deep learning methods are increasingly being explored to recognize implicit and non-obvious mathematical patterns within these time series.

Significant progress has been made in the development of neural predictors. However, in 2020 and 2022, two specific deep learning models for forecasting, N-BEATS and N-HiTS, were introduced. These models feature architectures capable of abstracting complex autoregressive temporal patterns and capturing their behavioral trends, offering a technologically innovative alternative in this research area. While deep learning methods were traditionally regarded as black-box models, these specialized neural forecasters are equipped to interpret key time-series characteristics such as seasonality, trends, and autoregressive patterns, thereby providing enhanced interpretability and sensitivity to the unique attributes of temporal data, including cyclicalities.

Incorporating these methods into AQI prediction represents a significant advancement in both statistical and environmental knowledge, paving the way for future studies to further refine and expand the application of neural forecasters in this critical area.

## 2 THEORETICAL BACKGROUND

### 2.1 AIR POLLUTION

All substances that alter the composition of air are considered atmospheric pollutants. Air is primarily composed of nitrogen (approximately 78.10%) and oxygen (approximately 20.93%), with other compounds present in smaller quantities. These pollutants can have natural origins, such as volcanic activity, storms, wildfires, pollen, and others. However, there is a significant anthropogenic (human-induced) contribution, which has been present since the discovery and use of fire.

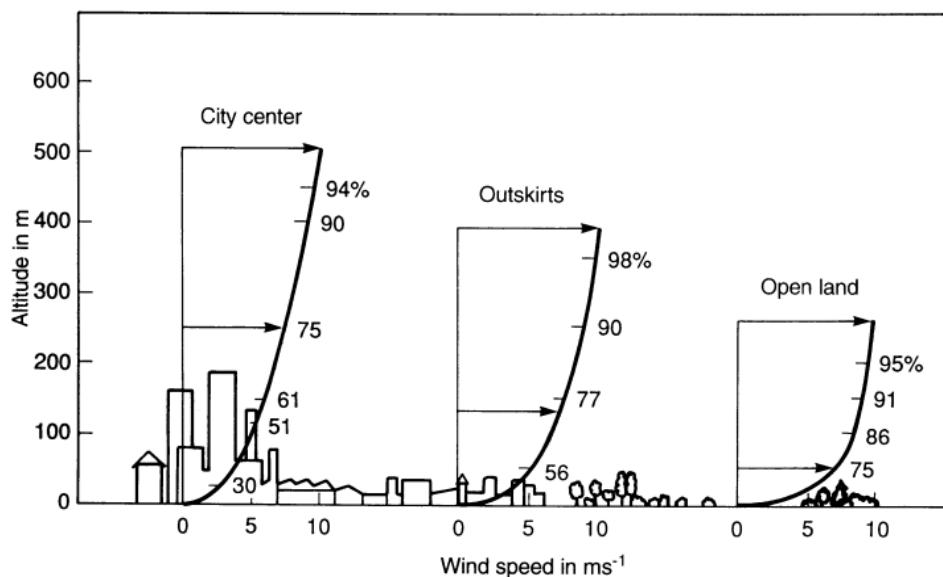
The air pollutants can be classified as primary and secondary. The first one refers to pollutants emitted directly from a specific source to the atmosphere. Its original form is already harmful to human health and the environment, and doesn't need a chemical transformation for this. The main ones are sulfur dioxide ( $\text{SO}_2$ ), carbon monoxide (CO), Large Particulate Matter ( $\text{PM}_{10}$ ), Fine Particulate Matter ( $\text{PM}_{2.5}$ ) and nitrogen oxides ( $\text{NO}_x$ ). The secondary pollutants are the ones that are formed through reactions between primary pollutants and other compounds. Some examples of this class are: tropospheric ozone ( $\text{O}_3$ ), ammonium nitrate ( $\text{NH}_4\text{NO}_3$ ), ammonium sulfate ( $(\text{NH}_4)_2\text{SO}_4$ ) and photochemical smog.

Natural pollution encompasses a range of environmental phenomena with implications that can be categorized into three distinct levels: local, global, and interactive. Locally, natural pollution significantly impacts air quality, leading to adverse effects on respiratory and cardiovascular health in nearby populations. On a global scale, events such as volcanic eruptions can induce substantial changes in the climate system, influencing global temperatures and disrupting the hydrological cycle. Furthermore, natural pollution often interacts with anthropogenic emissions, exacerbating issues such as photochemical smog and intensifying the broader impacts of human-induced pollution. Although natural sources of pollution cannot be controlled, understanding their mechanisms and patterns is crucial. This knowledge is essential for interpreting test results in environmental studies and developing strategies to mitigate the compounded effects of natural and anthropogenic pollutants.

Another important point to be considered in air pollution analyzes is the geographic influence. It plays a crucial role in the dispersion, concentration and impact of air pollution. Factors such as relief, climate, location and natural characteristics of a region (like vegetation and proximity to bodies of water) can intensify or mitigate the effects of pollutants in the atmosphere.

For example, cities like Los Angeles (USA) and Santiago (Chile) often face problems with the concentration of pollutants due to the mountainous terrain. It happens because regions surrounded by mountains tend to trap pollutants, especially during thermal inversion phenomena, where a layer of cold, heavier air is trapped beneath a layer of hot air. On the other hand, lowland regions generally allow greater dispersion of pollutants, but the absence of natural barriers can result in the transport of pollution over long distances.

In addition to relief, other geographic aspects that influence air pollution are wind and rain. Winds play an essential role in the dispersion of pollutants. In areas with strong, consistent winds, pollutants are more easily diluted. However, winds can transport pollutants to distant regions, affecting air quality in areas that do not have local emission sources. As for rain, it helps to "wash" the atmosphere, removing solid particles and soluble gases, such as sulfur dioxide ( $\text{SO}_2$ ). On the other hand, the lack of rain in arid areas contributes to the accumulation of suspended particles in the air. Another aspect strongly linked to rain is humidity. High humidity levels can intensify the formation of secondary pollutants, such as photochemical smog and secondary aerosols. Still dealing with climate and time, there is the issue of temperature and thermal inversion. Under normal conditions, hot air rises and disperses pollutants. During a thermal inversion, cold air is trapped close to the ground, making it difficult to disperse and worsening pollution.



**Figure 2.1** - Relationship between urbanization and wind speed.

Source: BAUMBACH (1996)

Regarding vegetation, forests can act as natural filters, absorbing CO<sub>2</sub> and suspended particles. However, they can also emit volatile organic compounds (VOCs), which contribute to the formation of tropospheric ozone. On the other hand, the lack of vegetation increases vulnerability to the transport and concentration of particles by wind, especially in arid or degraded regions. In relation to the proximity of bodies of water, coastal regions often experience sea breezes that can disperse pollutants out to sea. However, these areas can also be affected by emissions from ships, port activities and salt aerosols. Furthermore, areas close to bodies of water may have specific microclimates that influence the dispersion of pollutants.

Another important aspect to consider is the location of cities. Cities at high altitudes may face less pollutant dispersion due to reduced air density at high altitudes. Cities in tropical regions tend to have greater dispersion of pollutants due to strong thermal convection. In cities in polar regions, emissions can remain concentrated for prolonged periods due to the absence of convection and atmospheric stability.

There are still some more specific factors, such as the transport and deposition of pollutants. They can travel thousands of kilometers due to winds, affecting areas geographically distant from the emission sources. An example is sandstorms from the Sahara that reach South America or volcanic particles that circulate globally.

Not only these, but air pollution is also a major environmental risk to human and animal health, with effects that vary depending on pollutant type, concentration, and exposure duration.

In humans, the respiratory system is commonly affected, with pollutants such as PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, and O<sub>3</sub> causing bronchitis, asthma, and long-term risks like COPD and lung cancer. The cardiovascular system is also at risk, as CO and PM<sub>2.5</sub> can lead to arrhythmias, hypertension, heart attacks, and strokes. The nervous system is vulnerable to heavy metals and ultrafine particles, which are linked to cognitive deficits, neurodegenerative diseases, and developmental issues in children. Reproductive health can be impaired by pollutants like PM<sub>2.5</sub> and NO<sub>2</sub>, leading to fertility issues, miscarriages, and low birth weight. Additionally, prolonged exposure to carcinogens such as benzene and PAHs raises cancer risks, while immune system suppression increases susceptibility to infections. Vulnerable groups include children, the elderly, pregnant women, individuals with chronic diseases, and low-income populations with higher exposure to pollutants.

In animals, air pollution induces similar respiratory issues, including bronchitis and reduced lung capacity. Heavy metals and neurotoxic pollutants impair behavior, reproduction, and nervous system function, while ozone damages respiratory tissues and hinders survival behaviors like hunting. Prolonged exposure to pollutants such as dioxins and PCBs leads to hormonal disruptions, infertility, and genetic defects. Contamination of food chains through bioaccumulation compounds these impacts, posing a broad threat to ecosystems.

### 2.1.1 AIR QUALITY INDEX (AQI)

The Air Quality Index (AQI) serves as a standardized measure of air quality based on the concentration of pollutants. It is an indicator developed by governmental agencies to communicate the current level of air pollution in a specific region and to predict potential future pollution levels. The AQI is divided into six categories. The first category indicates minimal atmospheric pollution. The second category poses a low risk for individuals who are exceptionally sensitive to pollutants. In the third category, symptoms may begin to manifest in at-risk groups. The fourth category signals potential impacts on the respiratory and cardiovascular systems of healthy individuals, with symptoms becoming more severe in the fifth category. Finally, in the sixth category, significant health issues and pronounced symptoms are almost certain to occur (WANG, 2022).

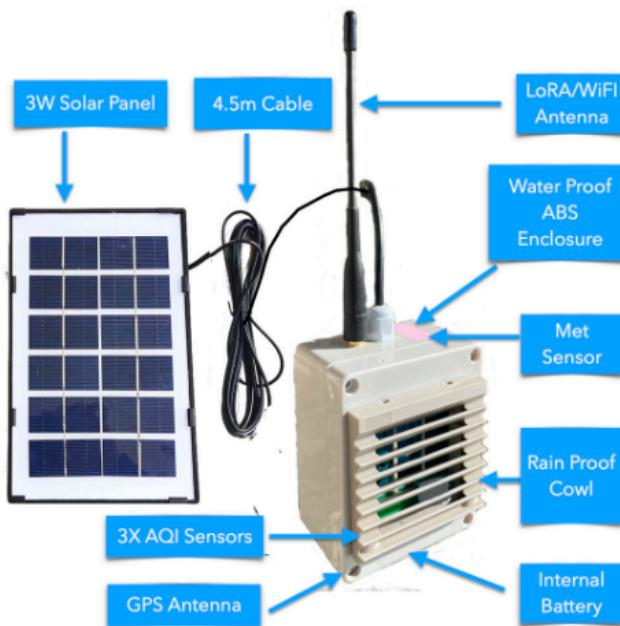
AQI	Air Pollution Level	Health Implications	Cautionary Statement (for PM2.5)
0 - 50	Good	Air quality is considered satisfactory, and air pollution poses little or no risk	None
51 - 100	Moderate	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.	Active children and adults, and people with respiratory disease, such as asthma, should limit prolonged outdoor exertion.
101-150	Unhealthy for Sensitive Groups	Members of sensitive groups may experience health effects. The general public is not likely to be affected.	Active children and adults, and people with respiratory disease, such as asthma, should limit prolonged outdoor exertion.
151-200	Unhealthy	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects	Active children and adults, and people with respiratory disease, such as asthma, should avoid prolonged outdoor exertion; everyone else, especially children, should limit prolonged outdoor exertion
201-300	Very Unhealthy	Health warnings of emergency conditions. The entire population is more likely to be affected.	Active children and adults, and people with respiratory disease, such as asthma, should avoid all outdoor exertion; everyone else, especially children, should limit outdoor exertion.
300+	Hazardous	Health alert: everyone may experience more serious health effects	Everyone should avoid all outdoor exertion

**Figure 2.2 - Air Quality Index categories table.**

Source: AQICN (2024)

The AQI for each pollutant is determined based on its atmospheric concentration over a specific time period, as measured by monitoring stations, with each pollutant having its own calculation method for the index. Additionally, many countries have their own standardized AQI system; this study adopts the U.S. AQI methodology.

Pollutant measurements can be conducted using various types of monitors. The monitors employed by AQICN are of the GAIA type. An example is the GAIA A18, which has a detection cycle ranging from 30 seconds to 5 minutes. It is extremely lightweight and compact, featuring multiple protective mechanisms such as a solar panel for energy recharge and waterproof cables. These monitors are typically placed in locations shielded from direct wind, rain, and dust sources. They are installed at a height of over 1,5 meters and at least 10 meters away from the nearest road – a distance that increases based on the average daily vehicle traffic, potentially reaching up to 250 meters.



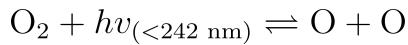
**Figure 2.3 - GAIA A18 monitoring station.**

Source: AQICN (2024)

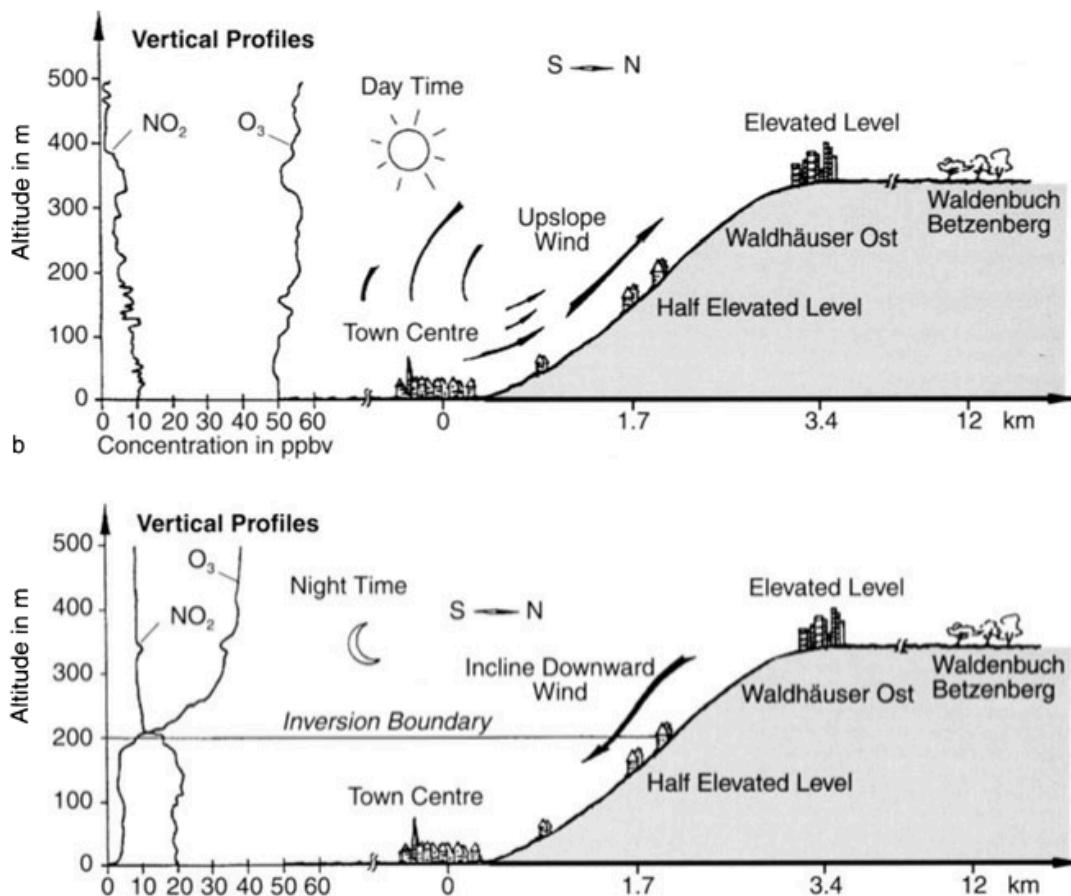
### 2.1.2 O<sub>3</sub>

Ozone is formed when molecular oxygen is exposed to UV light of short wavelengths at very high altitudes, leading to the creation of the stratospheric ozone layer. A harmful implication of anthropogenic air pollution is its impact on this process. Since ozone formation depends heavily on UV radiation and it readily decomposes into NO<sub>2</sub>, high concentrations of

$O_3$  at lower altitudes during the day are reversed into  $NO_2$  at night. Considering M a third molecule like  $N_2$  absorbing excess energy:



Ground-level ozone is also produced when sunlight interacts with molecules of volatile organic compounds (VOCs) and nitrogen oxides, which are emitted by vehicles, industries, and various other sectors. During winter, temperature inversions occur, trapping VOCs in denser air near the surface and preventing them from escaping to higher altitudes. This confinement facilitates the chemical reactions that produce significant amounts of ozone (RAVINDIRAN, 2023).



**Figure 2.4 -  $NO_2$  and  $O_3$  vertical profiles during day and night over Tübingen, Germany.**

Source: BAUMBACH (1996)

Ozone plays a crucial role in global warming. The climate depends on the interplay between the atmosphere, hydrosphere, and biosphere, along with the radiation balance across various systems. The stratospheric ozone layer filters UV rays, making life on Earth possible.

A sustained and intermittent increase in UV radiation reaching the Earth's surface contributes to global warming, driven by the destruction of the ozone layer and the accumulation of greenhouse gases. This leads to uninhabitable regions, desert expansion, flooding, and inundation of land due to the melting of polar ice caps.

The depletion of the ozone layer is closely linked to chlorofluorocarbons (CFCs) – highly stable compounds with long lifespans – and nitrogen oxides ( $\text{NO}_x$ ). This depletion allows greater penetration of UV-B and UV-C radiation, to which animals, plants, and humans are highly sensitive.

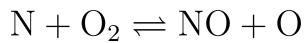
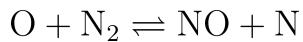
Conspicuously, ozone has a highly detrimental effect on certain plants (such as wheat, beans, potatoes, vines, tomatoes, and pine), causing visible damage to foliage, including a reddish-purple coloration and necrosis. It is absorbed through the stomata and induces oxidative destruction of surface lipoproteins, cell membranes, metabolites, and other cellular components.

Although  $\text{O}_3$  is a blue gas in large quantities, it is typically transparent with a distinct odor, characterized by being a strong oxidant. It causes severe irritation in the respiratory tract and penetrates the lungs due to its low affinity for water. Upon entering the lungs, it oxidizes proteins, enzymes, lipids, and many essential biocompounds, leading to chromosomal breakage, increased susceptibility to infections, coughing irritation, and eye discomfort.

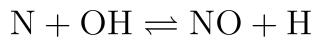
### 2.1.3 $\text{NO}_x$

Nitrogen oxides ( $\text{NO}_x$ ) originate from high-temperature combustion through the oxidation of atmospheric nitrogen and nitrogen present in the fuel. Initially, nitric oxide ( $\text{NO}$ ) is formed, and nitrogen dioxide ( $\text{NO}_2$ ) is subsequently synthesized in the post-combustion phase, where higher concentrations of  $\text{O}_2$  are present in the exhaust gases and atmosphere. Consequently,  $\text{NO}_2$  is frequently observed in processes with excess oxygen. There are three main pathways for NO formation.

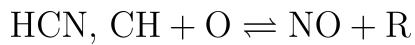
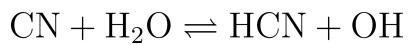
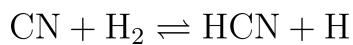
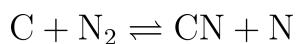
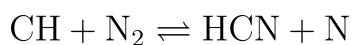
At temperatures above  $1300^\circ\text{C}$ , the concentration of free oxygen ( $\text{O}$ ) increases due to the dissociation of  $\text{O}_2$ , leading to higher concentrations of NO and, subsequently,  $\text{NO}_2$ . This is commonly referred to as the thermal  $\text{NO}_2$  formation mechanism. Under conditions of oxygen excess, the following reactions dominate:



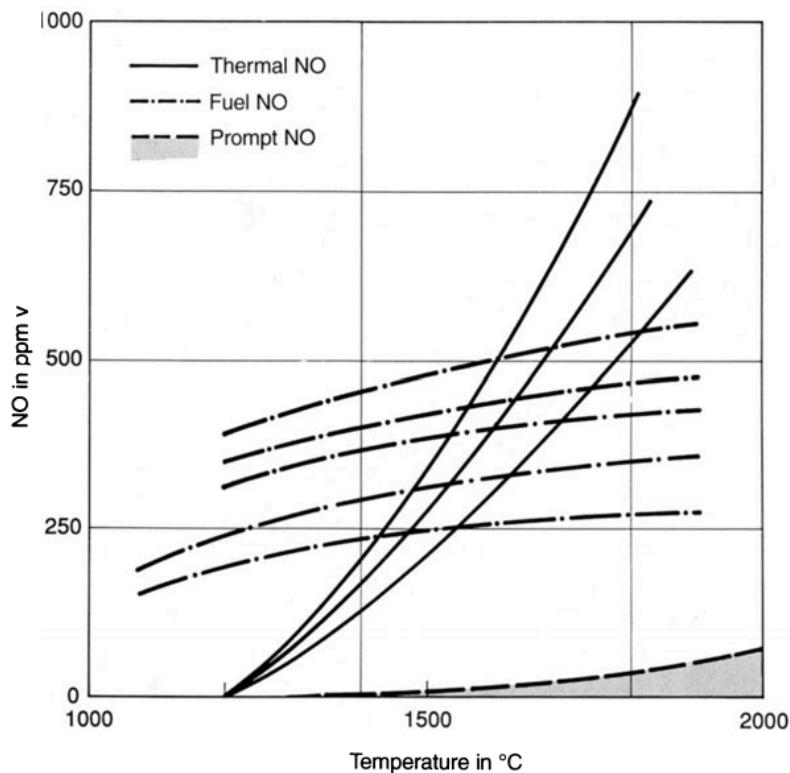
In fuel-rich regions, the following processes occur:



In regions with low oxygen availability, NO can be formed from fuel radicals, such as CH, reacting with molecular oxygen. This mechanism is known as "prompt NO." The following reactions are understood to contribute to the formation of NO and organic residues (R):



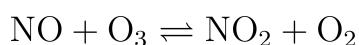
Furthermore, fuel NO originates from nitrogenous organic compounds present in certain fuels, such as oils, including amines, amides, nitrobenzenes, pyridines, among others. The oxidation of this nitrogen occurs even at low temperatures and increases under excess air conditions.



**Figure 2.5 - NO formation based on combustion temperature considering excess of O<sub>2</sub>.**

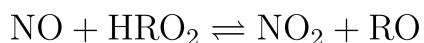
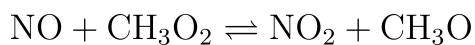
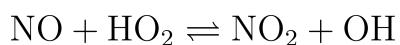
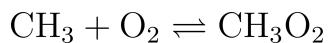
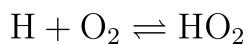
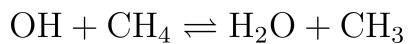
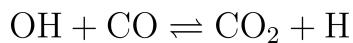
Source: BAUMBACH (1996)

Nitrogen oxides act in the atmosphere as both reactants and inhibitors, with nitric oxide (NO) and nitrogen dioxide (NO<sub>2</sub>) being the most significant. In principle, there are two main reactions by which NO<sub>2</sub> is formed from NO in the atmosphere. The first involves oxidation by atmospheric O<sub>2</sub> in a termolecular reaction, which proceeds at a slow rate under low NO concentrations. The second, faster reaction involves oxidation by ozone (O<sub>3</sub>) in a bimolecular process. In urban areas with high vehicular traffic, there is an increased synthesis of ozone, which, when combined with warmer periods of the day, leads to higher NO<sub>2</sub> formation. Another significant concern, in addition to the formation of NO<sub>2</sub>, is the consequent synthesis of O<sub>3</sub> through reversible reactions.

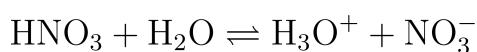


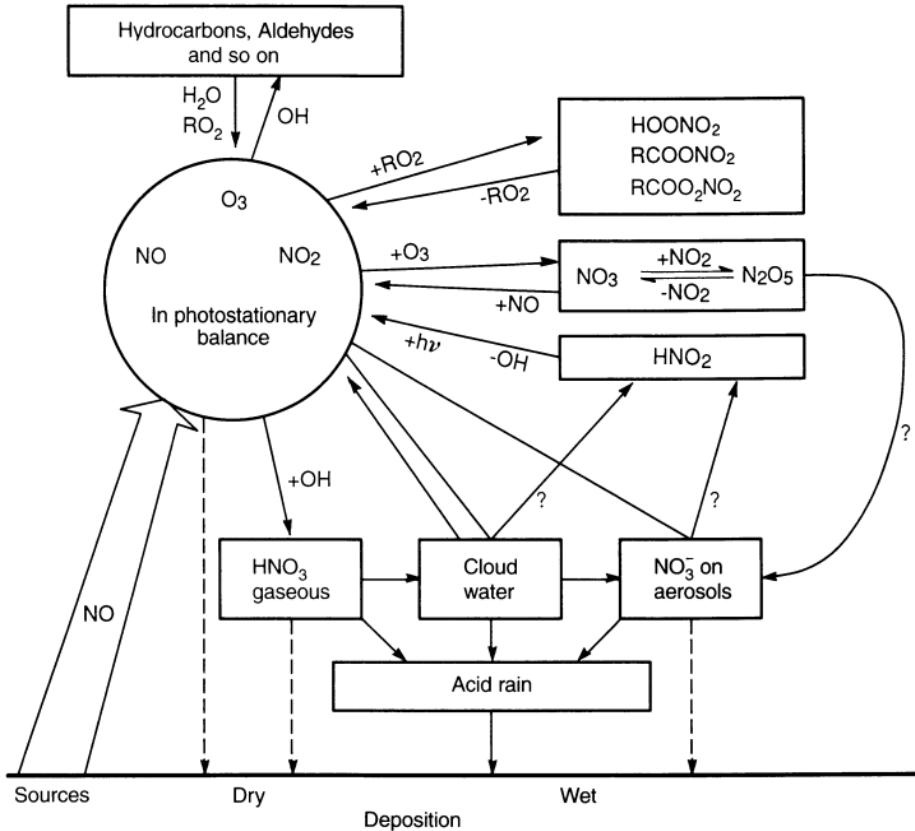
Hydrocarbons, whether of anthropogenic or natural origin, that contain a hydroxyl (OH) group contribute to the formation of peroxy groups, which tend to oxidize NO into NO<sub>2</sub>. Hydroxyl radicals are particularly more concentrated in the presence of solar radiation. When the NO<sub>2</sub>/NO ratio is high, there is increased O<sub>3</sub> formation, as well as a higher likelihood of

$\text{RO}_2$  radicals reacting with  $\text{NO}_2$ , leading to the formation of peroxides such as peroxyacetyl nitrate (PAN), a well-known component of the photochemical smog observed in Los Angeles.



The combination of  $\text{NO}_2$  and  $\text{O}_3$  also facilitates the formation of  $\text{NO}_3$  and  $\text{N}_2\text{O}_5$  during nighttime, although their concentrations are much lower compared to other pollutants. During the day,  $\text{NO}_3$  is rapidly reverted. A major concern lies in the reaction of  $\text{NO}_2$  with OH radicals and a third-body collision partner (M), leading to the formation of nitric acid ( $\text{HNO}_3$ ), a highly stable and water-soluble compound. Consequently,  $\text{HNO}_3$  is leached from the atmosphere and deposited onto the soil. The oxidation of  $\text{NO}_2$  into  $\text{HNO}_3$  in the gas phase is, therefore, environmentally significant, as it acidifies water and contributes approximately 30% to its acidification, alongside sulfates.





**Figure 2.6 - Diagram of NO<sub>x</sub> reactions in the atmosphere.**

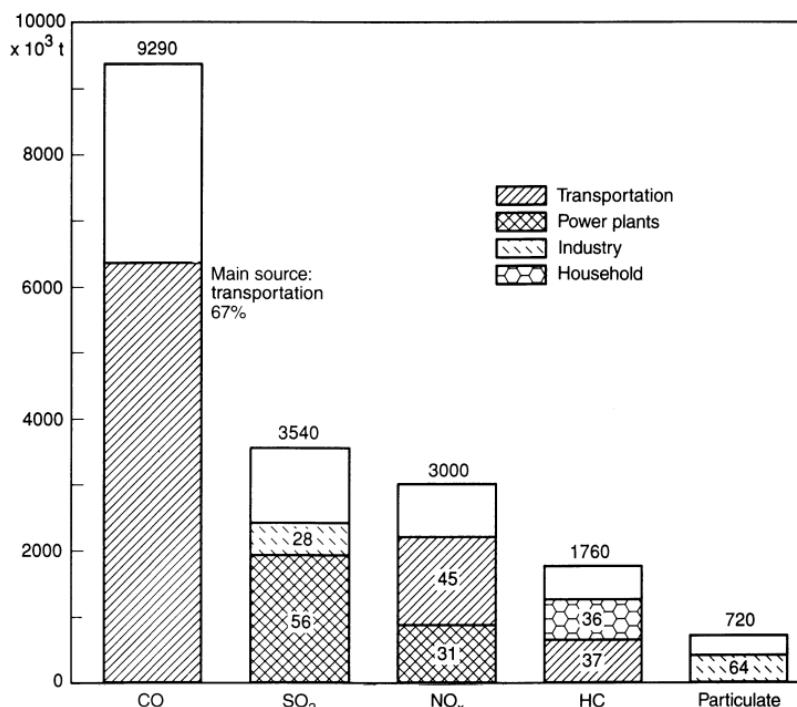
Source: BAUMBACH (1996)

Peroxyacetyl nitrate (PAN) is responsible for inhibiting lipid synthesis in plants, enzyme activity, and causing premature aging. It leads to a bronzing coloration of the leaves and discoloration of the pulp. It primarily affects crops such as beans, lettuce, and species of pine. Moreover, NO and NO<sub>2</sub> primarily affect pine and spruce trees, causing brown and reddish discoloration. They are absorbed through the stomata and lead to plant acidification, disruption of internal functions, and the potential formation of nitrosamines.

Forbye, NO<sub>2</sub> is an irritant gas for the human respiratory system, dissolving in the mucous membranes and increasing the susceptibility of the respiratory tract to germs. Similarly, NO causes irritation to these membranes and leads to the formation of methemoglobin in the blood, rendering cells unable to transport O<sub>2</sub>.

## 2.1.4 CO

Like soot and hydrocarbons, carbon monoxide (CO) is a byproduct generated during the incomplete combustion of carbon into carbon dioxide ( $\text{CO}_2$ ). Its oxidation requires a minimum ignition temperature of approximately  $715\text{ }^{\circ}\text{C}$ , along with sufficient residence time for the reaction to proceed effectively. If these conditions are not met, CO will escape as a byproduct. This phenomenon is particularly prevalent in scenarios where there is an insufficient supply of reactive oxygen, thereby compromising the efficiency of the combustion process.



**Figure 2.7** - Annual emissions of air pollutants in Federal States of Germany 1982.

Source: BAUMBACH (1996)

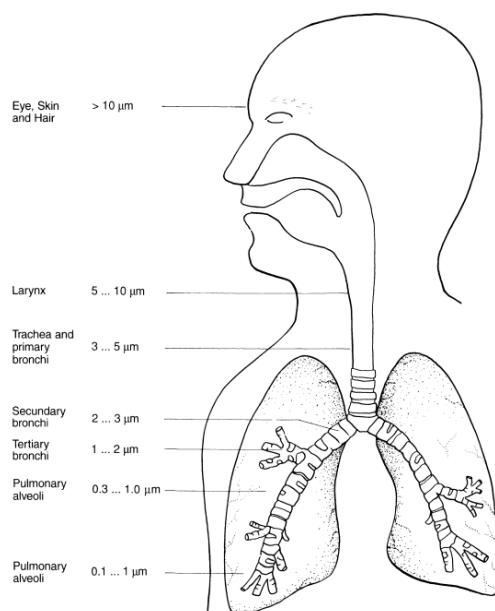
CO is highly dangerous to human health due to its strong affinity for hemoglobin (the blood pigment), surpassing that of  $\text{O}_2$ . This results in the formation of carboxyhemoglobin ( $\text{COHb}$ ), which obstructs the transport of  $\text{O}_2$  in the blood, severely affecting the central nervous system and cardiovascular system, and potentially leading to death.

### 2.1.5 PARTICULATE MATTER

Particulate matter (PM) refers to extremely small solid particles originating from various sources, primarily combustion. These particles may derive from the fuel and can consist of soot, high-molecular-weight condensed hydrocarbons, ash, metal oxides, and carbon particles.

Earth's temperature decreases when radiation is reduced by atmospheric pollution, with particulate matter playing a key role by partially reflecting and absorbing this radiation. The most significant source of aerosols is volcanic activity; however, anthropogenic influence is primarily driven by industrial areas. Withal, desert formation and deforestation contribute to the increase in particulate matter over time. The so-called nuclear winter, caused by atomic bombs, is also a major concern due to its potential for severe and widespread climate disruption.

The respiratory tract is the primary pathway through which particulate matter affects the human body. Depending on their aerodynamic properties, particles are absorbed and eliminated in different regions. Larger particles are generally retained in the nasopharyngeal region, while smaller particles can be deposited in the lungs. Numerous polycyclic aromatic hydrocarbons (PAHs) result from incomplete fossil fuel combustion and are frequently absorbed by particulate matter, including soot. When PAHs enter the respiratory tract, they are considered carcinogenic, with benzo[a]pyrene being a prominent example.



**Figure 2.8 - Particle deposition in dependence of the mean diameter of the particle.**

Source: BAUMBACH (1996)

## 2.2 FORECASTING

Forecasting is a mathematical technique for accurately predicting the future by analyzing historical data and identifying patterns. Its main purpose is to provide informative predictions of future events, which helps with decision-making, supply control, risk management, and more (HYNDMAN, 2021).

In air quality control, forecasting can play a crucial role in predicting pollution levels, enabling timely interventions and more effective management strategies. By forecasting future air quality, authorities can issue warnings, optimize traffic flow, regulate industrial activities, and implement other preventive measures to mitigate the impact of pollutants on public health and the environment (CASTELLI, 2020).

A good forecast depends on what is the final goal established by the scientist. Also, it is important to comprehend the horizon's extension, therefore the amount of time in the future that will be forecasted, to determine if the forecast is short, medium, or long-term (HYNDMAN, 2021). Distinct approaches might lead to different necessities and forecasting methods.

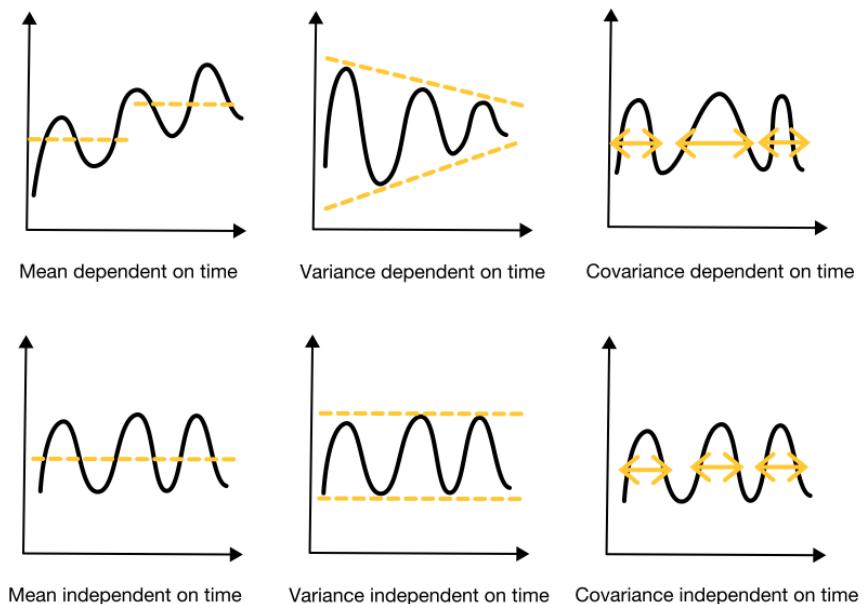
Classical regression models and techniques could be more efficient in forecasting time series data due to the time dependence inherent in such data, known as autocorrelation. This property occurs when the values in a time series are correlated with their past values, violating the assumption of independence required by traditional regression models. This dependency structure can lead to biased estimates and reduced predictive accuracy, making specialized time series models more appropriate for capturing temporal dynamics (HYNDMAN, 2021).

### 2.2.1 TIME SERIES PROPERTIES

Time series data often exhibit underlying patterns that, when identified, significantly enhance the accuracy of forecasting models. Decomposing a time series into its core components – trend, seasonality, and residuals – provides a comprehensive understanding of historical data behavior (HYNDMAN, 2021). This decomposition not only elucidates the systematic variations over time but also enables researchers and forecasters to identify periods or regions where specific values are most probable. By leveraging these components, forecasting models can better capture and predict future fluctuations, thereby improving predictions' overall quality and reliability.

While time series decomposition offers valuable insights into the behavior of historical data, additional features and properties can further enhance forecasting accuracy. One such impactful property is stationarity, which refers to a time series whose statistical properties – mean, variance, and autocovariance – remain constant over time. Stationary time series are particularly important in many forecasting models, as they often lead to more stable and reliable predictions (HYNDMAN, 2021). Therefore, trend, complex seasonality, variance, and noise are examples of characteristics that generate non-stationarity in a dataset.

Differencing the data one or more times is a common technique employed in forecasting methods such as ARIMA (AutoRegressive Integrated Moving Average) to achieve higher accuracy – presented in Appendix A. This process transforms a non-stationary time series into a stationary one by removing trends and reducing variability, thereby satisfying the stationarity assumption required by many statistical models. By stabilizing the mean and variance over time, differencing enables ARIMA models to more effectively capture underlying patterns and generate more reliable forecasts (NIELSEN, 2019).



**Figure 2.9 -** Graphic explanation of stationarity.

Source: MONIGATTI (2024)

Stationarity tests typically revolve around hypotheses concerning the presence or absence of a unit root in the time series, which indicates whether the series is non-stationary. Commonly used tests for stationarity include the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, the Augmented Dickey-Fuller (ADF) test, and the Phillips-Perron (PP) test. The KPSS test evaluates the null hypothesis that the series is stationary, whereas the ADF and PP tests

assess the null hypothesis of a unit root, implying non-stationarity. These tests provide critical insights into the characteristics of the data and guide the appropriate transformations for reliable forecasting (NIELSEN, 2019).

### 2.2.2 STATE-OF-THE-ART

Forecasting has its roots in statistical methods that emerged in the early 20th century. Initially, time series analysis and statistical modeling were the primary techniques for predicting future outcomes. Early models, such as moving averages and exponential smoothing, were employed to capture basic patterns in the data. As statistical theory advanced, more sophisticated models, including the Autoregressive Integrated Moving Average (ARIMA) and the Box-Jenkins methodology, gained prominence. These methods focused on understanding and modeling the underlying time-dependent structure of data through components such as trend, seasonality, and noise. Despite their utility, traditional statistical methods often struggled with non-linear relationships and high-dimensional data, which limited their ability to accurately forecast complex, real-world phenomena (ZHANG, 2024).

With the rise of machine learning (ML) in the late 20th and early 21st centuries, forecasting techniques evolved to incorporate algorithms capable of learning from data without explicit programming for every potential pattern. Machine learning models, particularly regression trees, support vector machines, and neural networks, introduced a new level of flexibility in handling both linear and non-linear relationships (ZHANG, 2024).

These models are particularly adept at identifying intricate patterns in large datasets, providing more accurate and robust forecasts. For example, recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) were designed to better capture temporal dependencies in sequential data, addressing many of the limitations of earlier statistical methods. As ML algorithms became more computationally efficient, they facilitated forecasting in domains ranging from finance to energy consumption and weather prediction (ZHANG, 2024).

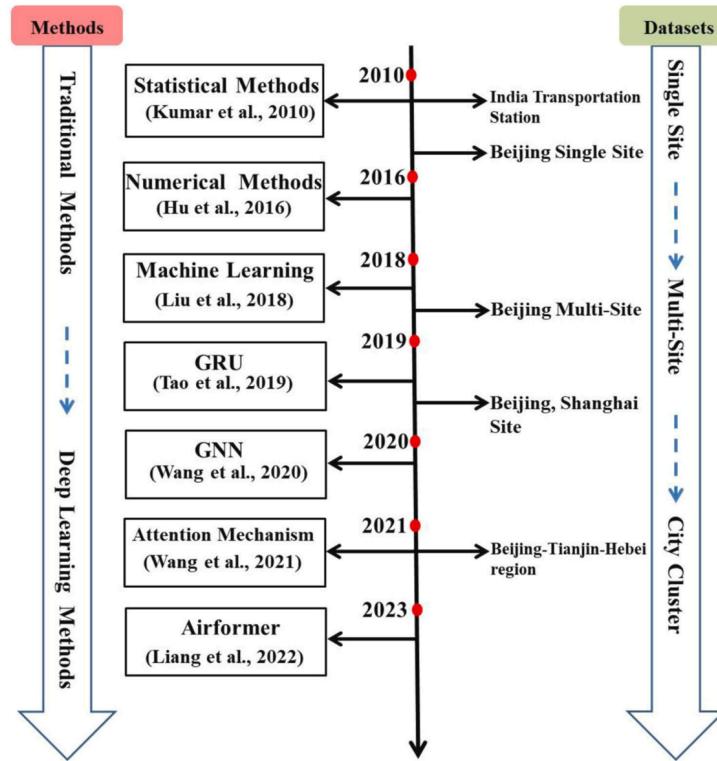
Today, forecasting has become a blend of traditional statistical methods and cutting-edge machine learning techniques, further enhanced by advances in big data and artificial intelligence (AI). Modern forecasting leverages a combination of time series models and deep learning techniques to make highly accurate predictions, even in the presence of

complex, noisy, or missing data. Techniques like hybrid models, which integrate classical statistical methods with deep learning algorithms, are increasingly popular. Additionally, the availability of real-time data from IoT devices, social media, and other digital sources has transformed forecasting into a dynamic, adaptive process. Automated machine learning (AutoML) frameworks and the rise of explainable AI (XAI) have made these advanced methods more accessible (ZHANG, 2024).

When forecasting the Air Quality Index, predictions primarily rely on numerical modeling, statistical methods, and traditional machine learning, using climate and spatial information to aid in the modeling. Among the numerical predictors are the Community Multiscale Air Quality Modeling System (CMAQ), the Comprehensive Air Quality Model with Extensions (CAQME), and the Nested Air Quality Prediction Modeling System (NAQPMS) (ZHANG, 2024).

Statistical methods were constrained by the assumption of stationarity and linear dependency, including approaches such as autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), and multivariable linear regression (MLR). Meanwhile, machine learning methods were based on Support Vector Regression (SVR) and Random Forest Regression (RFR) (ZHANG, 2024).

Hu et al. achieved promising predictions based on CMAQ for PM<sub>2.5</sub> and O<sub>3</sub> in an annual experiment conducted in China in 2016. In 2010, Kumar et al. measured air quality in Delhi, India, using ARMA and ARIMA models. In Taiwan in 2009, Cheng et al. proposed a moving average model with Order Weight Average (OWA) that effectively calculated ozone concentrations. Other notable studies include Lei et al., who employed Classification and Regression Trees (CART) and Multiple Regression (MR) in Macau; Nieto et al., who utilized Support Vector Regression (SVR) in Spain with high nonlinear accuracies; and Kumar et al., who combined ARIMA with Principal Component Regression (PCR) in Delhi. These studies demonstrate the dependence of forecasting accuracy on the comprehensiveness of the employed models and their various limitations, in addition to the high computational demand and requirements (ZHANG, 2024).



**Figure 2.10** - Evolution process of air quality prediction methods and relevant datasets.

Source: ZHANG (2024)

The current landscape of air quality index (AQI) forecasting heavily relies on deep learning techniques, particularly those that address temporal dependencies in data. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have become widely used for modeling time-series data due to their ability to capture long-range dependencies. Moreover, hybrid architectures that combine space-temporal data with Convolutional Neural Networks (CNNs) allow for the extraction of spatial features, which are then processed in memory-specialized cells, further enhancing forecasting accuracy (ZHANG, 2024).

More recently, complex Transformer-based models have been applied to AQI forecasting due to their superior performance in handling sequential data and capturing long-range dependencies. These models offer advanced capabilities in understanding and predicting air quality patterns with greater precision (ZHANG, 2024).

### 2.2.3 DIFFICULTIES

The difficulties in accurate forecasting arise from several factors, including data complexity, model selection, and external influences. One of the primary challenges is the huge variability and noise present in many datasets, which can obscure underlying patterns and make accurate predictions difficult. Additionally, selecting the appropriate model to capture the specific characteristics of the data requires expertise and careful tuning.

Besides this, the quality and availability of historical data play a crucial role in determining the reliability of forecasts, as missing or inconsistent data can skew results. Lastly, many forecasting models, particularly complex machine learning models, are computationally intensive and require significant resources for training and implementation.

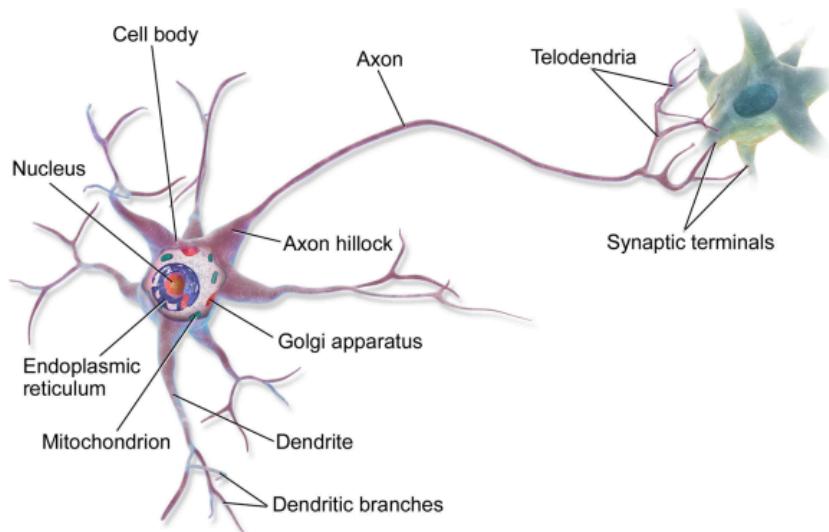
Specifically in the context of AQI forecasting, historical air quality data are often scattered across various sources, including monitoring stations, satellites, and aircraft. This fragmentation can introduce significant variance and noise into the data due to spatial and temporal heterogeneity. The differences in data collection methods, geographical coverage, and time intervals between measurements make it challenging to integrate and analyze the data cohesively, which can reduce the accuracy and reliability of forecasts (ZHANG, 2024).

## 2.3 NEURAL FORECASTERS

### 2.3.1 INTRODUCTION TO NEURAL NETWORKS

In recent years, deep learning has gained increasing importance due to its ability to process and analyze large volumes of data with unprecedented accuracy. This makes it a cornerstone in fields such as computer vision, natural language processing, and predictive analytics.

The foundation of deep learning lies in Artificial Neural Networks (ANNs), which are computational models inspired by the structure and functioning of human neurons. These networks process information by passing data through layers of interconnected nodes, mimicking how biological neurons transmit signals (GÉRON, 2021).



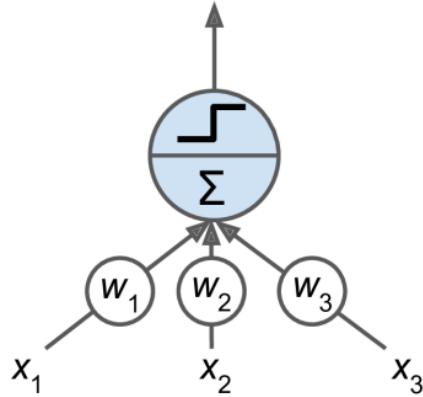
**Figure 2.11** - Biological neuron representation.

Source: GÉRON (2019)

Dendrites receive electrical signals from the axons of other neurons, processing these signals and propagating them as electrical impulses along the axon. These impulses are then transmitted to the axon terminals, or telodendria, which relay the information to other neurons or target cells through synaptic connections. This process enables the transmission and propagation of information across neural networks, forming the basis of neural communication.

Inspired by biological neural networks, ANNs receive data as input and propagate these values through artificial processing units known as perceptrons. Each perceptron processes the incoming signals by applying a weighted sum, followed by an activation function determining whether the signal should be passed to the next layer. These perceptrons

work collectively to identify patterns, make predictions, or classify data, enabling the network to learn complex representations from large datasets (GÉRON, 2021).



**Figure 2.12 - Architecture of a single perceptron.**

Source: GÉRON (2019)

Mathematically, the perceptron computes a weighted sum of all its received inputs, denoted as  $\mathbf{x}$ , by applying their respective weights,  $\mathbf{w}$ . This is expressed as:

$$z = \mathbf{x}^T \mathbf{w} \quad (1)$$

where  $\mathbf{x}^T$  represents the transpose of the input vector, and  $\mathbf{w}$  is the weight vector. The result,  $z$ , is then passed through an activation function to determine the perceptron's output. The activation function is given by

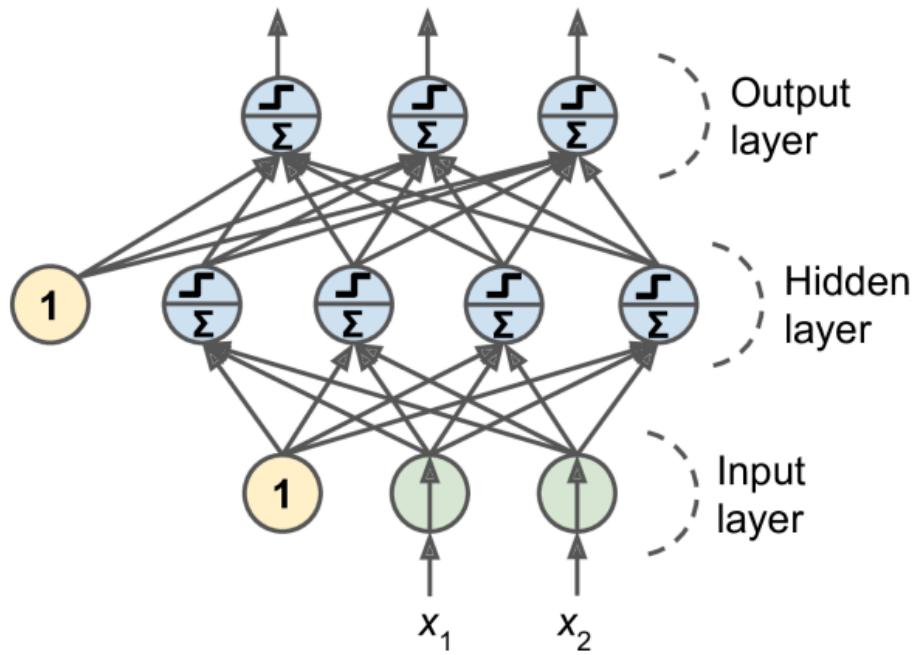
$$h_{\mathbf{w}}(\mathbf{x}) = \phi(z) \quad (2)$$

where  $\phi$  is the specific activation function. The function can take various forms, such as the step function, sigmoid, ReLU (Rectified Linear Unit), or others, depending on the problem being solved. It introduces non-linearity to the model, enabling it to learn more complex patterns and relationships within the data (GÉRON, 2021).

When working with a fully connected layer of perceptrons, the input data matrix  $\mathbf{X}$  is multiplied by the weight matrix  $\mathbf{W}$ , and optionally summed with a bias vector  $\mathbf{b}$ , before being passed through the activation function. This operation can be expressed as:

$$h_{\mathbf{W}, \mathbf{b}}(\mathbf{X}) = \phi(\mathbf{X}\mathbf{W}^T + \mathbf{b}) \quad (3)$$

where  $\phi$  is the activation function,  $\mathbf{X}\mathbf{W}^T$  represents the weighted sum of inputs, and  $\mathbf{b}$  is the bias vector added to each of the output neurons. The result is then passed through the activation function to produce the output of the layer (GÉRON, 2021).



**Figure 2.13** - Architecture of a multilayer perceptron.

Source: GÉRON (2019)

After the forward pass, the artificial neural network (ANN) adjusts its parameters (weights and biases) to better fit the data using a process called backpropagation, combined with automatic differentiation. Backpropagation computes the gradient of the loss function concerning each weight and bias by applying the chain rule of calculus, allowing the network to update its parameters iteratively in the direction that minimizes the error. This process is typically performed using optimization algorithms like gradient descent (GÉRON, 2021).

### 2.3.2 APPLIED NEURAL NETWORKS IN FORECASTING

Artificial Neural Networks (ANNs) operate by utilizing a collection of training data to train the model. This data typically consists of input features paired with their respective labels or target values, which the network learns to predict through an iterative optimization process. During training, the network adjusts its parameters (weights and biases) using algorithms such as backpropagation and gradient descent to minimize the error between the predicted outputs and the actual labels. This process enables the model to generalize patterns from the training data, allowing it to make accurate predictions on unseen data.

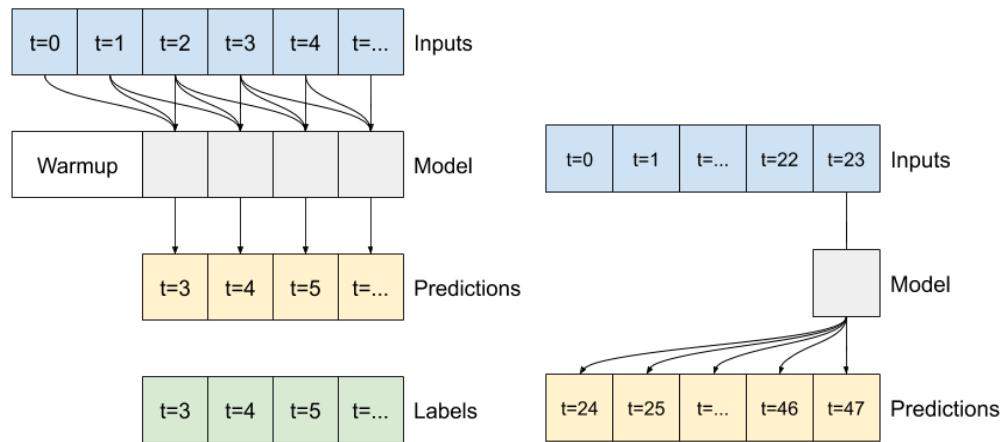
For time series data, Artificial Neural Networks are trained similarly to other tasks. However, in this context, the training process involves using a fixed range  $L$  of the historical data  $T$ , where  $L < T$ . This subset of historical data acts as the input window for the model.

The labels for each training instance correspond to the values in the subsequent time steps that the model is tasked with predicting, commonly referred to as the forecast horizon.

Mathematically, if the historical time series is represented as  $\{y_1, y_2, \dots, y_T\}$ , the training pairs consist of:

- Input:  $\{y_{t-L}, y_{t-L+1}, \dots, y_t\}$ , where  $t$  denotes the current time step.
- Labels:  $\{y_{t+1}, y_{t+2}, \dots, y_{t+H}\}$ , where  $H$  represents the forecast horizon.

This approach enables the model to learn temporal dependencies within the time series, leveraging patterns in the historical window  $L$  to accurately predict future values  $H$ .

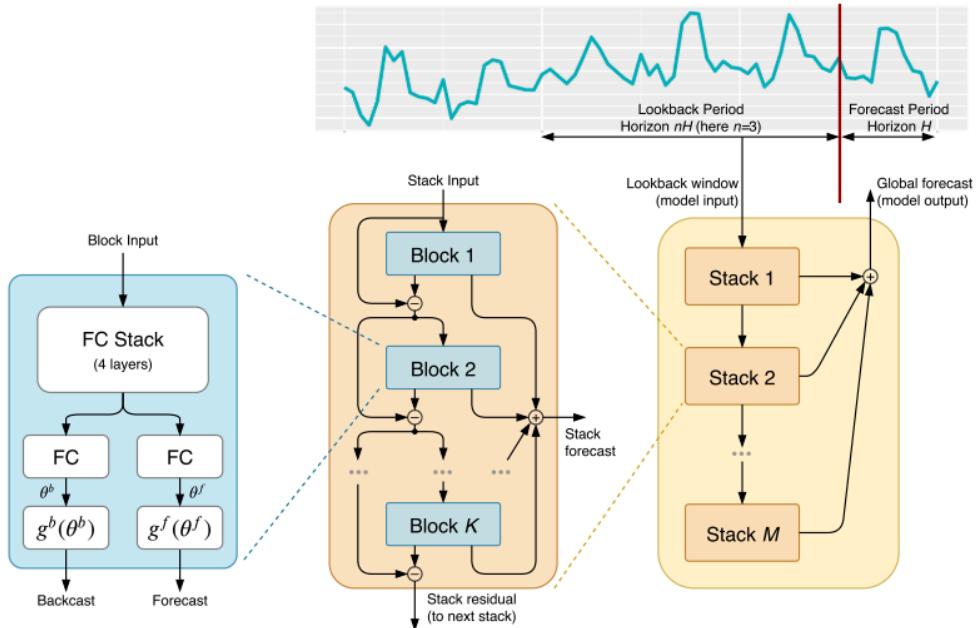


**Figure 2.14** - Graphical representation of Deep Learning modeling for time series forecasting.

Source: TENSORFLOW (2024)

### 2.3.3 N-BEATS

N-BEATS (Neural Basis Expansion Analysis for Time Series) is a univariate neural forecasting model designed to predict future values of a time series. Given a time series consisting of  $T$  temporal observations, the model generates forecasts for a horizon of  $H$  future time steps (ORESHKIN, 2020). Its architecture can be seen in Fig 2.15.



**Figure 2.15** - Graphical representation of N-HiTS architecture.

Source: ORESHKIN (2020)

The lookback period  $x$ , representing the input data, is fed into the first stack  $s$  of the model. This stack generates a forecast  $\hat{y}$ , which will later be aggregated with the forecasts produced by subsequent stacks. Additionally, the backcast output  $\hat{x}$  from the first stack serves as the input for the next stack in the sequence (ORESHKIN, 2020).

Within each stack, there are multiple blocks  $\ell$  responsible for generating forecasts  $\hat{y}_\ell$  (which are aggregated to form the stack's final forecast) and backcasts  $\hat{x}_\ell$ . The backcast output is subtracted from the input data to create the input for the subsequent block. Each block comprises four fully connected layers that compute the forward  $\theta_\ell^f$  and backward  $\theta_\ell^b$  predictors of expansion coefficients. These coefficients are then passed to a basis layer  $g_\ell$ , which utilizes them to produce the corresponding backcast and forecast (ORESHKIN, 2020).

$$\mathbf{h}_{\ell,1} = \text{FC}_{\ell,1}(\mathbf{x}_\ell) = \text{ReLU}(\mathbf{W}_{\ell,1}\mathbf{x}_\ell + \mathbf{b}_{\ell,1}) \quad (4)$$

$$\mathbf{h}_{\ell,2} = \text{FC}_{\ell,2}(\mathbf{x}_\ell) = \text{ReLU}(\mathbf{W}_{\ell,2}\mathbf{x}_\ell + \mathbf{b}_{\ell,2}) \quad (5)$$

$$\mathbf{h}_{\ell,3} = \text{FC}_{\ell,3}(\mathbf{x}_\ell) = \text{ReLU}(\mathbf{W}_{\ell,3}\mathbf{x}_\ell + \mathbf{b}_{\ell,3}) \quad (6)$$

$$\mathbf{h}_{\ell,4} = \text{FC}_{\ell,4}(\mathbf{x}_\ell) = \text{ReLU}(\mathbf{W}_{\ell,4}\mathbf{x}_\ell + \mathbf{b}_{\ell,4})) \quad (7)$$

$$\theta_\ell^b = \mathbf{W}_\ell^b \mathbf{h}_{\ell,4} = \text{Linear}_\ell^b(\mathbf{h}_{\ell,4}) \quad (8)$$

$$\theta_\ell^f = \mathbf{W}_\ell^f \mathbf{h}_{\ell,4} = \text{Linear}_\ell^f(\mathbf{h}_{\ell,4}) \quad (9)$$

The primary function of this architecture is to optimize the precision of the expansion coefficient predictors for forecasting, thereby improving the accuracy of time series predictions. The secondary function is to perform a backcast operation, which eliminates irrelevant information for future prediction, effectively creating a "filtered" input for the next block (ORESHKIN, 2020). Forecasts and backcasts are generated as follows:

$$\hat{\mathbf{y}}_\ell = g_\ell^f(\theta_\ell^f) = \sum_{i=1}^{\dim(\theta_\ell^f)} \theta_{\ell,i}^f \mathbf{v}_i^f \quad (10)$$

$$\hat{\mathbf{x}}_\ell = g_\ell^b(\theta_\ell^b) = \sum_{i=1}^{\dim(\theta_\ell^b)} \theta_{\ell,i}^b \mathbf{v}_i^b \quad (11)$$

Here, the vectors  $\mathbf{v}_i^f$  and  $\mathbf{v}_i^b$  represent the basis vectors for the forecast and backcast, respectively, for the  $i$ -th element of the coefficients. The role of the functions  $g$  is to provide the sets of basis vectors  $\{\mathbf{v}_i^f\}_{i=1}^{\dim(\theta_\ell^f)}$  and  $\{\mathbf{v}_i^b\}_{i=1}^{\dim(\theta_\ell^b)}$ , ensuring that their outputs vary appropriately according to the predictor coefficients. These functions  $g$  can either be fully learned by the model or predefined based on prior knowledge of the problem (ORESHKIN, 2020).

The architecture employs two residual branches: one for the backcast, using subtraction across stacks  $\mathbf{x}_\ell = \mathbf{x}_{\ell-1} - \hat{\mathbf{x}}_{\ell-1}$ , and one for forecast, aggregating outputs via summation  $\hat{\mathbf{y}} = \sum_\ell \hat{\mathbf{y}}_\ell$ . This structure facilitates efficient backpropagation while enabling hierarchical temporal decomposition. Alternative formulations of  $g$  functions further enhance the flexibility and significance of each stack (ORESHKIN, 2020).

The N-BEATS architecture predominantly assumes that the  $\mathcal{G}$  functions are linear projections, forming the basis of its generic framework. However, two interpretable mechanisms have been proposed to enhance the model's ability to decompose time series data into trend and seasonality components, allowing for greater interpretability and improved forecasting performance (ORESHKIN, 2020).

Trends' behavior can be effectively represented using low-degree polynomials. For a polynomial of degree  $p$ , the trend forecast at stack  $\ell$  is modeled as:

$$\hat{\mathbf{y}}_{s,\ell} = \sum_{i=0}^p \theta_{s,\ell,i}^f \cdot t^i \quad (12)$$

Here, the normalized time vector is defined as  $\mathbf{t} = [0, 1, 2, \dots, H-1]^T/H$ , and the polynomial basis matrix  $\mathbf{T}$  is constructed as:

$$\mathbf{T} = [\mathbf{1}, \mathbf{t}, \mathbf{t}^2, \dots, \mathbf{t}^p] \quad (13)$$

Using this formulation, the trend component forecast is expressed as:

$$\hat{\mathbf{y}}_{s,\ell}^{\text{tr}} = \mathbf{T} \theta_{s,\ell}^f \quad (14)$$

Typically, values of  $p = 2$  or  $p = 3$  are chosen to adequately capture the smooth, low-frequency nature of trends in the data (ORESHKIN, 2020).

Seasonality reflects regular, recurring fluctuations in the time series, characterized by periodic patterns where  $y_t = y_{t-\Delta}$ , with  $\Delta$  representing the seasonal period. To model these periodic variations, a Fourier series representation is employed. The seasonality component at stack  $\ell$  is given by:

$$\hat{\mathbf{y}}_{s,\ell} = \sum_{i=0}^{\lfloor H/2-1 \rfloor} \theta_{s,\ell,i}^f \cos(2\pi i t) + \theta_{s,\ell,i+\lfloor H/2 \rfloor}^f \sin(2\pi i t) \quad (15)$$

The corresponding Fourier basis matrix  $\mathbf{S}$  is defined as:

$$\mathbf{S} = [\mathbf{1}, \cos(2\pi \mathbf{t}), \dots, \cos(2\pi \lfloor H/2 - 1 \rfloor \mathbf{t}), \sin(2\pi \mathbf{t}), \dots, \sin(2\pi \lfloor H/2 - 1 \rfloor \mathbf{t})] \quad (16)$$

Using this matrix, the seasonality component forecast can be expressed as:

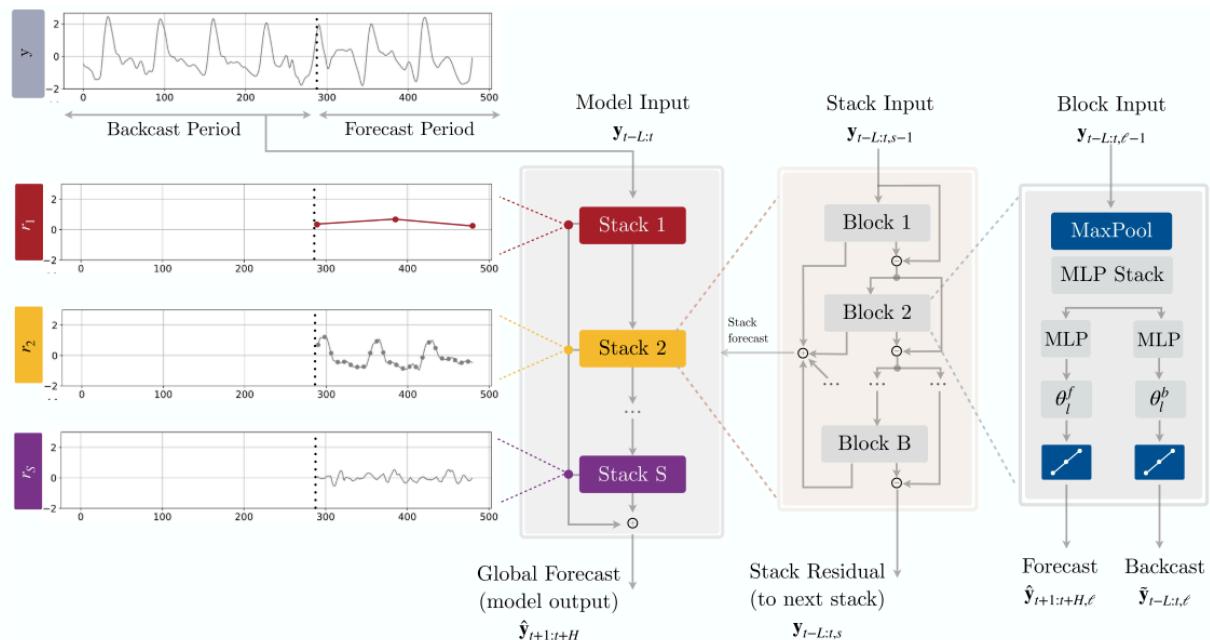
$$\hat{\mathbf{y}}_{s,\ell}^{\text{seas}} = \mathbf{S} \theta_{s,\ell}^f \quad (17)$$

The combination of trend and seasonality components allows for a structured decomposition of the time series. The trend component removes non-cyclic variations,

thereby "cleaning" the data, while the seasonality component isolates periodic fluctuations. This hierarchical approach enables the model to provide interpretable forecasts, with each stack contributing distinctly to the overall decomposition. By varying the equations used for the  $g$  functions, the model can further adapt to the specific characteristics of the data, enhancing its flexibility and effectiveness (ORESHKIN, 2020).

### 2.3.4 N-HiTS

Building on the foundational principles of N-BEATS for time series forecasting, N-HiTS (Neural Hierarchical Interpolation for Time Series) introduces an enhanced approach by leveraging database backcasts and frequency filter structures to generate forecasts. This architecture aims to provide a more comprehensive and flexible framework for capturing temporal dynamics and refining predictions through hierarchical interpolation (CHALLU, 2020). Its architecture can be seen in Fig 2.16.



**Figure 2.16 - Graphical representation of N-HiTS architecture.**

Source: CHALLU (2022)

The main differences between this algorithm and N-BEATS are the multi-rate signal sampling and the hierarchical interpolation. Before each block  $\ell$  in the N-HiTS architecture, a MaxPooling layer with kernel size  $k_\ell$  is applied to the input data. This operation enables the block to analyze components of the input at a specific scale. Larger kernel sizes result in a greater reduction of lower frequencies (smaller temporal scales), thereby forcing the fully

connected (dense) layers to focus on capturing higher-frequency components. In other words, the multiple layers can capture signals corresponding to different scales within the input data. The ability of a block to capture higher temporal scales is a key factor in its effectiveness for making long-term forecasts (CHALLU, 2020).

Additionally, the MaxPooling operation helps to reduce the dimensionality (length) of the input to the multilayer perceptron (MLP), which not only decreases memory usage but also mitigates the risks of overfitting (CHALLU, 2020).

$$\mathbf{y}_{t-L:t,\ell}^{(p)} = \text{MaxPool}(\mathbf{y}_{t-L:t,\ell}, k_\ell) \quad (18)$$

Following the sub-sampling step, each block  $\ell$  performs a non-linear regression on the input by learning the interpolation coefficients  $\theta_\ell^f$  (for the forward forecast) and  $\theta_\ell^b$  (for the backward backcast) using an MLP. These coefficients are learned from the hidden layer vector  $\mathbf{h}_\ell \in \mathbb{R}^{N_h}$ , which is obtained through the MLP's forward pass. The hidden vector  $\mathbf{h}_\ell$  is then linearly projected to yield the forward and backward coefficients (CHALLU, 2020).

The forward pass through the block is as follows:

$$\mathbf{h}_\ell = \text{MLP}_\ell \left( \mathbf{y}_{t-L:t,\ell}^{(p)} \right) \quad (19)$$

The forward and backward coefficients are computed as:

$$\theta_\ell^f = \text{Linear}^f(\mathbf{h}_\ell) \quad (20)$$

$$\theta_\ell^b = \text{Linear}^b(\mathbf{h}_\ell) \quad (21)$$

These learned coefficients are subsequently used in the hierarchical interpolation process to synthesize the block's outputs. Specifically, the backcast  $\hat{\mathbf{x}}_{t-L:t,\ell}$  and the forecast  $\hat{\mathbf{y}}_{t-L:t,\ell}$  are generated as follows:

$$\hat{\mathbf{x}}_{t-L:t,\ell}, \hat{\mathbf{y}}_{t-L:t,\ell} = \text{Hierarchical Interpolation}(\theta_\ell^b, \theta_\ell^f) \quad (22)$$

By applying this non-linear regression and hierarchical interpolation framework, the model captures complex temporal patterns and produces accurate forecasts and backcasts, enabling improved performance in time series prediction (CHALLU, 2020).

In most multi-horizon forecasting models, the cardinality of the neural network corresponds to the dimension of the forecast horizon, leading to a rapid increase in computational demands. To mitigate this issue, N-HiTS employs temporal interpolation,

where the dimensionality of the interpolation coefficients is expressed in terms of an expressivity ratio  $r_\ell$ , which controls the number of parameters per unit of output time:

$$|\theta_\ell^f| = \lceil r_\ell H \rceil \quad (23)$$

This expressivity ratio balances computational efficiency while ensuring the model captures sufficient temporal patterns. To recover the original sample and forecast the entire horizon, an interpolation function  $g$  is applied:

$$\hat{y}_{\tau,\ell} = g(\tau, \theta_\ell^f), \quad \forall \tau \in \{t+1, \dots, t+H\} \quad (24)$$

$$\hat{x}_{\tau,\ell} = g(\tau, \theta_\ell^b), \quad \forall \tau \in \{t-L, \dots, t\} \quad (25)$$

The interpolation mechanism is designed to distribute the expressivity ratios across the blocks in a manner synchronized with the multi-rate sampling. Blocks closer to the input have smaller  $r_\ell$  values and larger kernel sizes  $k_\ell$ , resulting in lower-granularity signals produced by more aggressive interpolation. These blocks are also subjected to more aggressive sub-sampling, which produces smoothed signals. The forecast for the entire horizon,  $\hat{\mathbf{y}}_{t+1:t+H}$ , is derived by summing the outputs interpolated at different temporal scales (CHALLU, 2020).

The model utilizes exponentially increasing expressivity ratios to handle a broad range of frequencies while controlling the number of parameters. Alternatively, each stack may specialize in a known cyclical component of the time series (e.g., daily, weekly) by using corresponding frequencies (CHALLU, 2020).

Finally, the residuals from the backcast are subtracted from the input of the next level, amplifying the capture of relevant signals while filtering out components that have already been addressed:

$$\hat{\mathbf{y}}_{t+1:t+H} = \sum_{\ell=1}^L \hat{\mathbf{y}}_{t+1:t+H,\ell} \quad (26)$$

$$\hat{\mathbf{y}}_{t-L:t,\ell+1} = \mathbf{y}_{t-L:t,\ell} - \hat{\mathbf{x}}_{t-L:t,\ell} \quad (27)$$

By organizing the model architecture in this manner, N-HiTS effectively balances computational efficiency with the capacity to capture complex temporal patterns over long horizons. The smoothness of the projection functions, coupled with hierarchical interpolation, allows for forecasts over an arbitrarily large horizon (CHALLU, 2020).

Regarding computational performance in practical scenarios, the input to the model is typically defined as  $L = \mathcal{O}(H)$ , where the model operates to predict a forecast horizon  $H$  based on this input  $L$ . Let  $B$  represent the number of blocks in the architecture, and  $N_h$  denote the number of fully connected layers in each block. For the N-BEATS-G model, the computational effort increases linearly concerning the forecast horizon and the number of blocks, resulting in a computational complexity of  $\mathcal{O}(HB)$  for the forecast and  $\mathcal{O}(HN_h)$  for the number of layer operations (CHALLU, 2020).

In contrast, the N-HiTS model introduces a significant optimization by specializing each stack for a specific frequency based on the expressivity ratio. This specialization effectively reduces the number of parameters required in each layer. The expressivity ratio grows exponentially across the layers, where  $r_\ell = r^l$ , resulting in a geometric scaling of the spatial complexity of the model. Specifically, the computational complexity of the N-HiTS architecture can be expressed as:

$$\mathcal{O} \left( \sum_{l=0}^B Hr^l \right) = \mathcal{O} \left( H \frac{1 - r^B}{1 - r} \right) \quad (28)$$

This formulation demonstrates that the N-HiTS model scales more efficiently with the forecast horizon  $H$  than N-BEATS-G, particularly as the number of blocks  $B$  increases. By leveraging exponential growth in the expressivity ratio and specializing each stack for different temporal frequencies, N-HiTS significantly reduces the overall computational burden while maintaining the flexibility to capture complex patterns in time series data (CHALLU, 2020).

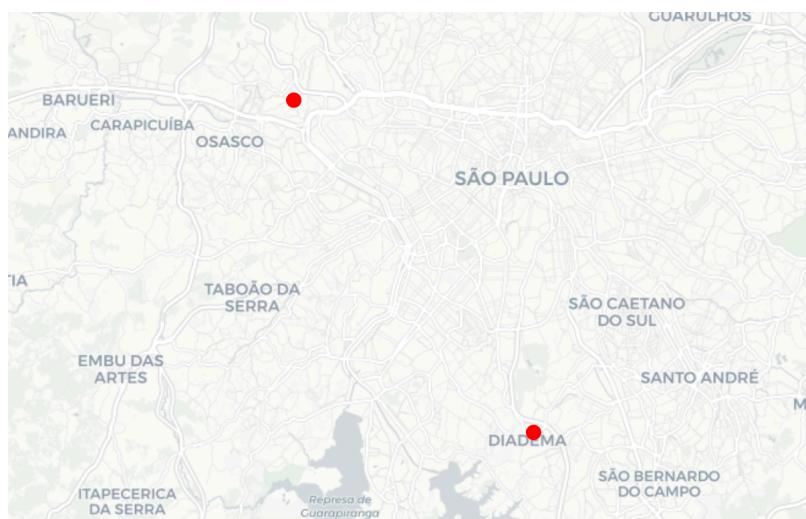
MODEL	TIME	MEMORY
LSTM	$\mathcal{O}(H)$	$O(H)$
ESRNN	$\mathcal{O}(H)$	$O(H)$
TCN	$\mathcal{O}(H)$	$O(H)$
Transformer	$\mathcal{O}(H^2)$	$O(H^2)$
Reformer	$\mathcal{O}(H \log H)$	$O(H \log H)$
Informer	$\mathcal{O}(H \log H)$	$O(H \log H)$
Autoformer	$\mathcal{O}(H \log H)$	$O(H \log H)$
LogTrans	$\mathcal{O}(H \log H)$	$O(H^2)$
N-BEATSi	$\mathcal{O}(H^2 B)$	$O(H^2 B)$
N-BEATSG	$\mathcal{O}(HB)$	$O(HB)$
N-HiTS	$\mathcal{O}(H(1 - r^B)/(1 - r))$	$O(H(1 - r^B)/(1 - r))$

**Figure 2.17** - Big  $\mathcal{O}$  evaluation for many Deep Learning forecasting methods.

Source: CHALLU (2020)

### 3 METHODOLOGY

The data used in this study were obtained from the Air Quality Historical Data Platform, available at AQICN Data Platform (AQICN, 2024), which provides daily measurements of various pollutants from multiple cities worldwide, accessible free of charge. All collected data span from April 1, 2016, to December 31, 2022, and were converted to weekly data by calculating the weekly averages. The air quality indices for PM<sub>10</sub>, PM<sub>2.5</sub>, CO, and NO<sub>2</sub> were sourced from an air quality monitor located at the Tietê Riverbank, at Ponte dos Remédios, situated between São Paulo and Osasco, SP. The data for O<sub>3</sub> were obtained from the municipal government building (Paço Municipal) in Guarulhos, SP.



**Figure 3.1** - Partial map of São Paulo indicating the locations of monitoring devices.

Source: The Authors

Missing data in all time series were interpolated using a piecewise polynomial in the Bernstein basis, which ensures compatibility with the specified values and derivatives at breakpoints. This method constructs a series of polynomial segments, each defined over a specific interval, and the polynomials are expressed in Bernstein form. The Bernstein basis allows for the interpolation to preserve continuity and smoothness at the breakpoints, ensuring that both the values and their derivatives match at the segment boundaries, resulting in a smooth and accurate approximation of the missing data.

The primary packages used in this study were Python 3.12.0, 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. The specification of package versions is crucial for ensuring the reproducibility and consistency of results, as updates or changes in package versions may

introduce variations in functionality or behavior that could affect the outcome of the analysis. All stochastic processes were executed with a fixed seed value of 1 to guarantee the reproducibility of results across different runs.

For each pollutant time series, training and testing were conducted over different forecasting horizons. The horizons considered were 12 weeks, 26 weeks, 52 weeks, and 78 weeks, using a fixed set of 353 historical data points. The statistical forecasting methods utilized the entire training dataset for each horizon, while the neural forecasters had their input size optimized for each respective horizon.

**Table 3.1** - Forecast horizon details and data availability.

Forecast Horizon (Weeks)	Time Equivalence	Available Historical Data	Proportional Data Availability
$h = 12$	3 months	341	$28 \cdot h$
$h = 26$	1 semester	327	$12 \cdot h$
$h = 52$	1 year	301	$5, 5 \cdot h$
$h = 78$	1.5 years	275	$3, 5 \cdot h$

Source: The Authors

The metrics used to compare the performance of the models across all forecasting horizons were MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), MSE (Mean Squared Error), MAX (Maximum Error), and sMAPE (Symmetric Mean Absolute Percentage Error).

Each of these metrics provides a different perspective on model performance: MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. MAPE expresses the errors as a percentage of the actual values, providing a normalized measure of accuracy. MSE penalizes larger errors more heavily due to the squaring of residuals, making it sensitive to outliers. MAX quantifies the largest error in the forecast, highlighting extreme prediction failures. sMAPE, unlike MAPE, is symmetric, which prevents the metric from being biased towards over- or under-predictions, offering a more balanced assessment of accuracy.

$$\text{MAE} = \frac{1}{H} \sum_{i=1}^H |y_i - \hat{y}_i| \quad (29)$$

$$\text{MAPE} = \frac{100}{H} \sum_{i=1}^H \frac{|y_{T+i} - \hat{y}_{T+i}|}{|y_{T+i}|} \quad (30)$$

$$\text{MSE} = \frac{1}{H} \sum_{\tau=t}^{t+H} (y_\tau - \hat{y}_\tau)^2 \quad (31)$$

$$\text{MAX} = \max_{i=1}^H |y_i - \hat{y}_i| \quad (32)$$

$$\text{sMAPE} = \frac{200}{H} \sum_{i=1}^H \frac{|y_{T+i} - \hat{y}_{T+i}|}{|y_{T+i}| + |\hat{y}_{T+i}|} \quad (33)$$

Nineteen distinct statistical models from the statsforecast==2.0.0 package were utilized, including AutoARIMA, AutoCES, AutoETS, and AutoTheta. These models employ statistical and optimization techniques, such as the Akaike Information Criterion (AIC) and cross-validation, to identify optimal parameter values, ensuring accurate and efficient forecasting by selecting the best-fitting model configuration for the given data.

The hyperparameters of the N-HiTS and N-BEATS models were optimized individually for each forecasting horizon and pollutant. The optimization process for both models considered the input size (integer values between 4 and 156), the number of stacks (integer values between 3 and 7 for N-HiTS, and 2 and 7 for N-BEATS), the number of blocks per stack (integer values between 1 and 7), the maximum number of epochs (integer values between 10 and 700), and the type of preprocessing applied to the input data (options included {No-transformation, Standardization, Box-Cox transformation, or Min-Max Normalization}).

Additionally, for the N-HiTS model, the kernel size was also optimized. This involved a permutation of values [1, 2, 3], applied only to the last three stacks, while earlier stacks, if present, had no kernel value. The downsampling frequency was another optimized parameter, with permutations of values [1, 4, 12, 52], applied only to the first three stacks, while the remaining stacks used a unitary frequency. Certain other hyperparameters were fixed based on recommendations from the literature, such as the kernel pooling method (MaxPool1D), the activation mechanism of perceptrons (ReLU), and the interpolation method (linear).

For the N-BEATS model, an additional hyperparameter optimized was interpretability. This involved selecting two out of three options – Trend, Seasonality, or Generic – which defined the type of  $g$ -function used by the stack. Any remaining stacks, if present, were configured as Generic. Like N-HiTS, all perceptrons in N-BEATS used ReLU as the activation function.

The optimization process was conducted over 100 trials for each forecasting horizon and pollutant. In each iteration, the neural forecasters were trained and tested on the datasets consisting of the first 341 and 337 weekly historical data points, respectively, for better optimization in a cross-validation scenario. The objective function aimed to minimize both the sMAPE and MAE metrics, prioritizing sMAPE over MAE. All iterations were recorded in a data frame, and the results were stored in a pickle file for subsequent analysis.

In the Optuna framework, each trial corresponds to a single objective function evaluation with a specific set of hyperparameter values. Optuna uses advanced search algorithms, such as Tree-structured Parzen Estimators (TPE), to sample hyperparameter combinations intelligently. These algorithms prioritize promising regions of the hyperparameter space based on the performance of previous trials, balancing exploration of new areas and exploitation of known good configurations. This iterative process allows Optuna to converge on an optimal or near-optimal set of hyperparameters efficiently (OZAKI, 2020).

Finally, the best trial's hyperparameters were applied to the final neural forecasting model used to predict the most recent data points in the dataset for each pollutant and forecasting horizon. This allowed for directly comparing the resulting performance metrics with those obtained from the statistical forecasting methods.

## 4 RESULTS AND DISCUSSION

### 4.1 OVERVIEW OF FORECASTING PERFORMANCE ACROSS METHODS

Tables 4.1 and 4.2 summarize the results obtained for N-HiTS and N-BEATS, respectively. They include the optimized hyperparameters used by each model for the specific pollutant and forecasting horizon, as well as their ranking position for each metric (for instance, a rank of 1 indicates the best performance among all methods, while a rank of 21 represents the worst performance out of the 21 methods).

In Appendix C, all tables are presented in detail, sorted by sMAPE, and include the values for all metrics: sMAPE, MAE, MAX, MAPE, and MSE. For all metrics, lower values indicate lower forecasting errors. It is also evident that a method can perform best for one metric while not performing as well for another, as each metric captures specific aspects of the model's performance.

The neural forecasters, particularly N-BEATS, demonstrated very satisfactory results when applied to the time series of O<sub>3</sub>, performing well for both shorter and longer forecasting horizons. However, for the other pollutants, their performance varied significantly, ranging from very good results to some of the worst observed outcomes.

**Table 4.1** - Optimized Hyperparameters and Rankings for N-HiTS Across All Studied Pollutants and Horizons.

N-HiTS													
info		hyperparameters							ranking				
pollutant	h	input	epochs	stacks	blocks	downsample	kernel	scalar	smape	mae	max	mape	mse
co	12	109	70	3	3	[52, 12, 52]	[3, 1, 1]	standard	7	6	8	10	6
pm10	12	134	456	3	1	[12, 12, 4]	[1, 3, 1]	boxcox	11	11	13	11	12
pm25	12	141	166	5	5	[12, 1, 1]	[2, 2, 2]	boxcox	11	11	12	13	13
o3	12	6	167	7	4	[12, 52, 12]	[1, 1, 1]	standard	2	2	21	2	2
no2	12	117	85	7	5	[12, 4, 52]	[1, 2, 3]	minmax	6	6	5	5	5
co	26	83	99	3	3	[12, 1, 1]	[1, 3, 1]	None	4	4	4	6	4
pm10	26	45	140	4	2	[4, 1, 52]	[2, 3, 1]	boxcox	10	10	13	3	8
pm25	26	109	92	3	6	[1, 12, 12]	[1, 2, 3]	None	1	1	12	1	1
o3	26	94	310	3	6	[12, 12, 12]	[1, 1, 2]	boxcox	1	1	1	1	1
no2	26	107	486	3	6	[52, 12, 52]	[3, 2, 3]	standard	14	14	14	14	14
co	52	41	393	3	7	[12, 52, 4]	[3, 1, 1]	None	1	2	5	1	3
pm10	52	89	397	4	5	[1, 12, 12]	[2, 1, 1]	standard	13	10	13	7	13
pm25	52	110	156	6	6	[4, 12, 52]	[1, 3, 1]	standard	13	11	12	13	12
o3	52	28	73	4	1	[52, 1, 12]	[2, 2, 2]	boxcox	15	15	13	15	14
no2	52	81	336	7	7	[52, 4, 1]	[2, 1, 2]	None	10	8	10	11	10
co	78	93	614	6	1	[4, 12, 12]	[1, 2, 3]	standard	17	17	10	11	17
pm10	78	75	685	6	4	[12, 4, 1]	[3, 2, 2]	boxcox	12	12	12	13	12
pm25	78	59	681	3	6	[52, 12, 12]	[2, 3, 2]	minmax	9	9	1	4	1
o3	78	28	546	6	2	[1, 1, 1]	[3, 2, 2]	standard	2	2	3	8	3
no2	78	69	75	7	2	[1, 1, 4]	[3, 2, 3]	boxcox	19	19	13	19	19

Source: The Authors

**Table 4.2 - Optimized Hyperparameters and Rankings for N-BEATS Across All Studied Pollutants and Horizons.**

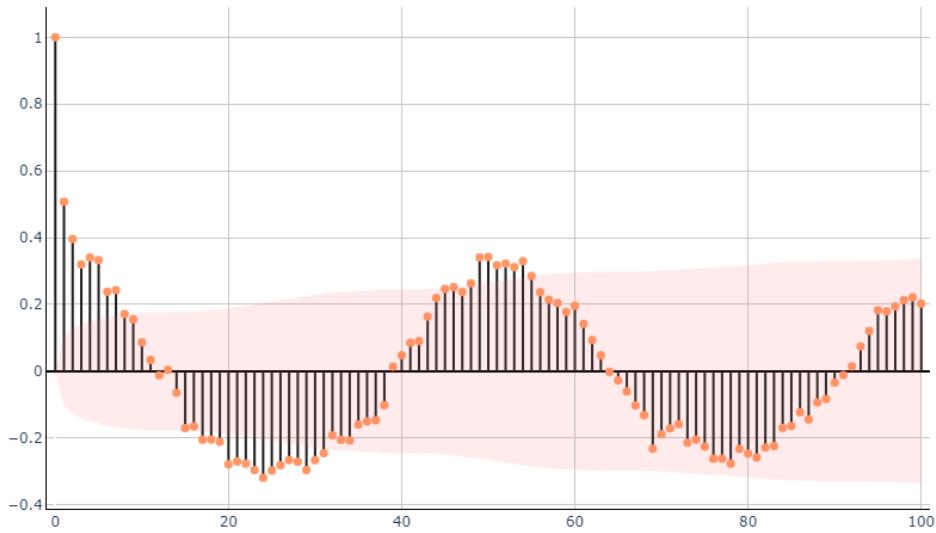
N-BEATS												
info		hyperparameters						ranking				
pollutant	h	input	epochs	stacks	blocks	interpretability	scalar	smape	mae	max	mape	mse
co	12	66	587	5	3	[identity, trend]	boxcox	5	4	21	<b>1</b>	13
pm10	12	40	88	5	5	[seasonality, seasonality]	boxcox	14	14	12	13	13
pm25	12	81	614	2	1	[seasonality, trend]	boxcox	21	21	19	21	21
o3	12	12	586	7	4	[identity, identity]	None	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
no2	12	144	290	4	2	[trend, seasonality]	minmax	<b>1</b>	<b>1</b>	21	6	8
co	26	42	198	7	3	[trend, identity]	boxcox	14	12	19	7	15
pm10	26	50	456	2	2	[seasonality, trend]	boxcox	13	13	15	13	14
pm25	26	62	155	2	1	[trend, identity]	None	2	2	14	2	3
o3	26	32	97	2	3	[trend, trend]	None	3	3	3	9	3
no2	26	114	39	3	5	[identity, identity]	minmax	<b>1</b>	<b>1</b>	2	<b>1</b>	<b>1</b>
co	52	135	605	2	5	[identity, seasonality]	boxcox	2	<b>1</b>	<b>1</b>	13	<b>1</b>
pm10	52	127	134	5	4	[identity, seasonality]	standard	10	7	12	8	8
pm25	52	116	398	6	4	[identity, identity]	None	9	6	13	6	13
o3	52	24	38	2	2	[trend, identity]	minmax	5	5	2	5	<b>1</b>
no2	52	85	255	3	2	[trend, trend]	boxcox	14	12	7	14	12
co	78	104	15	3	5	[trend, identity]	standard	21	19	15	18	18
pm10	78	51	380	7	3	[trend, trend]	minmax	13	13	19	14	13
pm25	78	53	689	2	4	[identity, seasonality]	boxcox	10	10	7	12	10
o3	78	4	95	3	2	[trend, trend]	minmax	<b>1</b>	<b>1</b>	<b>1</b>	7	<b>1</b>
no2	78	114	539	4	5	[trend, identity]	minmax	16	18	15	18	16

Source: The Authors

#### 4.2 POLLUTANT-SPECIFIC FORECASTING PERFORMANCE

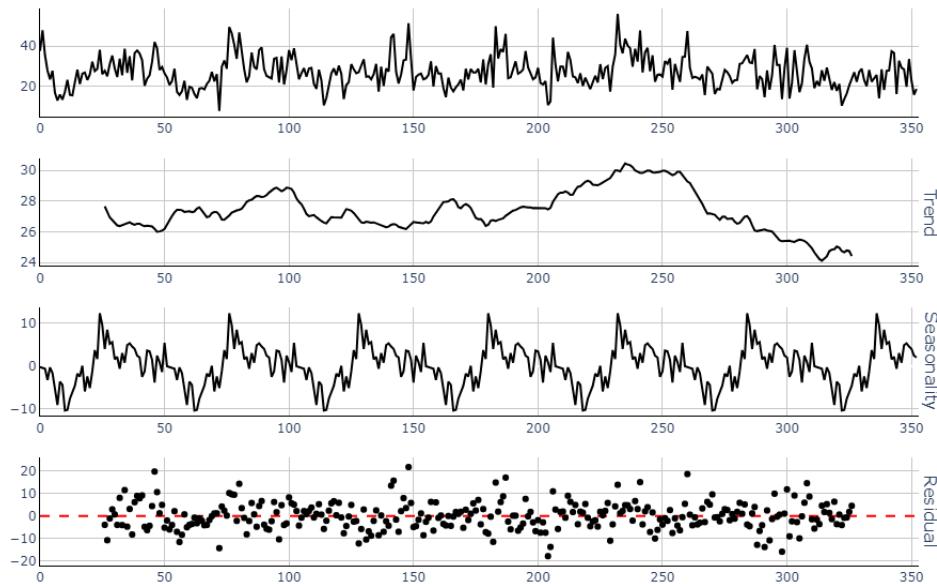
Appendix D presents the relevant time series plots for each studied time series. It can be observed that all the series exhibit highly complex seasonal decomposition, where the trend cannot be easily defined in polynomial terms. This complexity poses challenges for models attempting to recognize a regressive pattern. A well-defined trend would typically allow for the explanation of long-term movements in the data, such as persistent increases or decreases over time, and would facilitate predictions of the general direction of the series, independent of short-term fluctuations or seasonal variations.

Furthermore, it is evident that the series exhibit seasonality, as demonstrated by the autocorrelation functions, which reveal periodic and cyclical behavior, increasing confidence in the seasonal patterns as the lags progress. Not only this, but the seasonal decomposition indicates that seasonality is informative for annual periods, additively and multiplicatively, making it a crucial component for forecasting models. However, the seasonal patterns remain highly complex and challenging to capture accurately, particularly for more interpretative methods that rely on simpler or more transparent assumptions about the data structure.

Autocorrelation Function (ACF) - PM<sub>10</sub>**Figure 4.1** - Autocorrelation function (ACF) of PM<sub>10</sub>.

Source: The Authors

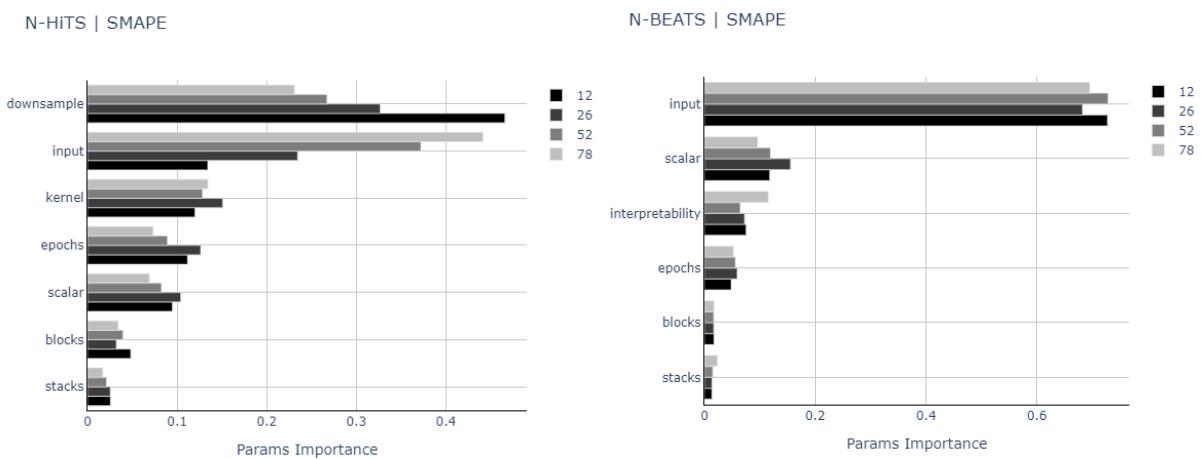
A significant issue, however, is the residual quantity left after the decomposition of the series. The residuals can exhibit a multiplicative influence greater than 2, highlighting models' considerable difficulty in recognizing these complex behaviors. This suggests that the remaining noise or irregularities after decomposition still contain substantial patterns that are challenging for forecasting models to capture accurately.

Seasonal Decomposition - O<sub>3</sub>**Figure 4.2** - Seasonal decomposition of O<sub>3</sub>.

Source: The Authors

### 4.3 HORIZON EFFECTS AND IMPACT OF HYPERPARAMETERS ON NEURAL FORECASTERS

It is well-known that the forecasting horizon influences the effectiveness of temporal predictions. In general, there is a consensus that long-term forecasts are consistently more difficult to predict accurately than shorter horizons. This occurs because, as the forecast horizon increases, the uncertainty in the model's predictions also increases. Over longer periods, more factors can change, such as seasonal patterns, external influences, and trends that are harder to capture. Additionally, the accumulation of errors over time makes it more challenging to maintain accuracy, as small deviations early in the forecast can compound, leading to larger discrepancies in longer-term predictions.



**Figure 4.3 -** fANOVA results for sMAPE hyperparameter importance in N-HiTS and N-BEATS models.

Source: The Authors

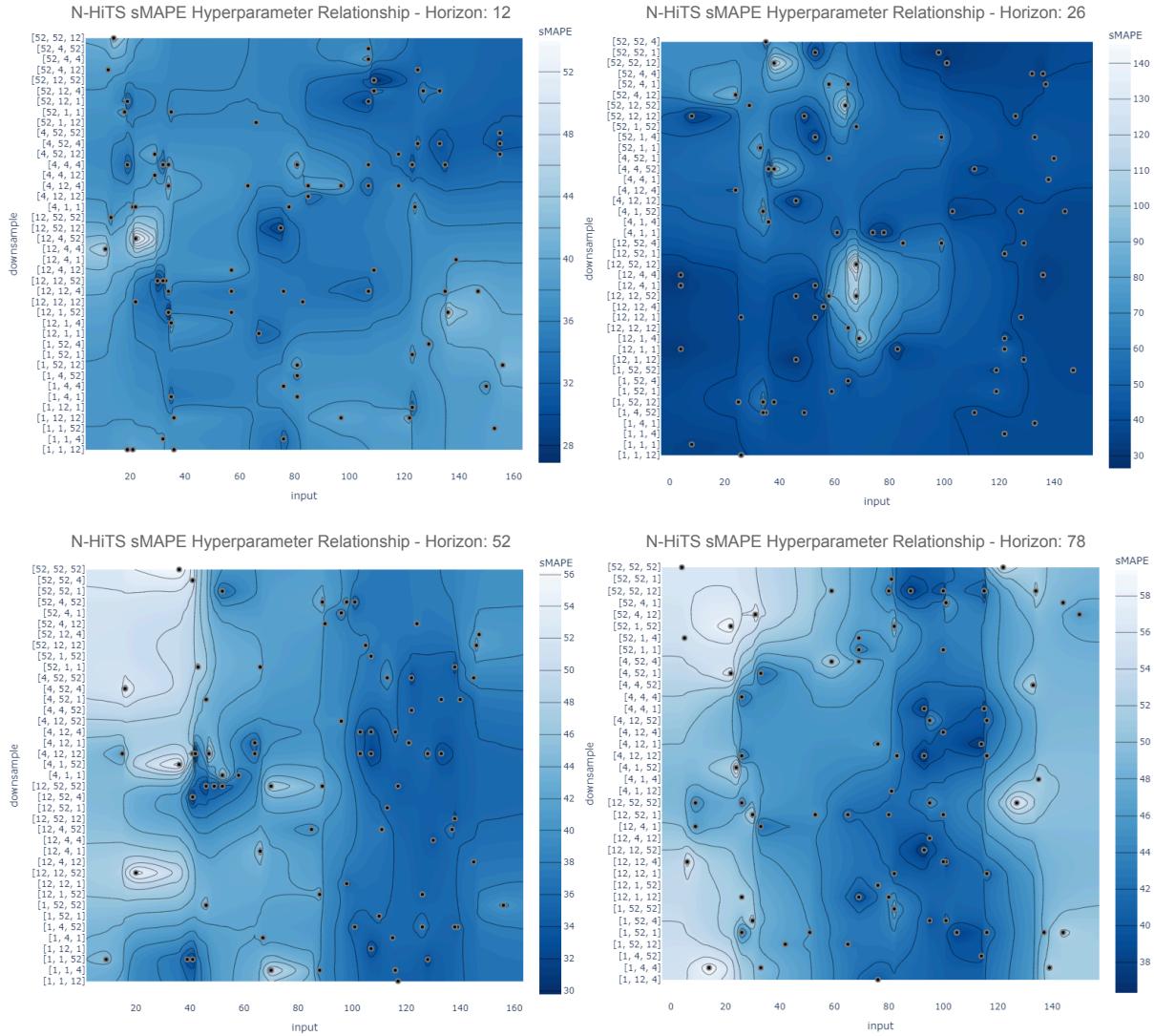
Figure 4.3 and Appendix E demonstrate that for N-BEATS, regardless of the forecasting horizon, the hyperparameters exhibit a well-defined hierarchy of importance according to the fANOVA test. Specifically, an accurate result is approximately 70% dependent on the size of the model's input data. Following that, the type of scaling transformation, interpretability, and the number of training epochs significantly influence the model's efficiency. Finally, the number of blocks and stacks becomes less relevant, together accounting for less than 5% of the importance in the forecast.

Contradictorily, the adjustment of N-HiTS appears to be highly dependent on the forecasting horizon. While the type of scaling transformation, the number of blocks, and the number of stacks are consistently presented as the least relevant hyperparameters, the downsample rate and input size fluctuate significantly depending on the horizon. Intuitively,

as the forecasting horizon increases, the relevance of the input size becomes considerably greater, with downsampling becoming more significant for shorter horizons.

This can be intuitively explained by two factors. The first factor relates to pattern recognition, where capturing patterns from the current moment has a greater impact on future predictions as the horizon lengthens. The second factor is linked to the amount of available data, as with longer forecasting horizons, the training data becomes increasingly scarce, reaching a maximum of 3.5 times the horizon length, compared to the availability of up to 28 times the horizon length for a 12-week forecast.

From the contour plots in Figure 4.4, it is evident that as the forecasting horizon increases, it becomes increasingly difficult to find local minima (which is noticeable by the increase in white areas in the plots). A higher number of trials would be beneficial to capture even more non-linear patterns. However, it is already clear that there are minimum regions strongly associated with the input size and downsampling, with the optimal input size range becoming more clearly defined as the horizon lengthens in N-HiTS.



**Figure 4.4** - Contour plot showing the hyperparameter relationship between input size and downsampling frequency for N-HiTS CO forecasting at 12, 26, 52, and 78-week horizons.

Source: The Authors

The high complexity of patterns present in time series data necessitates the optimization of model hyperparameters. For example, the hyperparameters `{input_size: 28, n_stacks: 6, n_blocks: 2, max_steps: 546, local_scalar_type: standard, n_pool_kernel_size: [3, 2, 2], n_freq_downsample: [1, 1, 1]}` resulted in an MAE approximately 4 times lower than those with the hyperparameters `{input_size: 155, n_stacks: 3, n_blocks: 6, max_steps: 527, local_scalar_type: boxcox, n_pool_kernel_size: [1, 1, 2], n_freq_downsample: [1, 1, 2]}` for O<sub>3</sub> in a 78-week forecast horizon. Counterintuitively, a smaller input size and the absence of downsampling produced better results than the opposite, highlighting that these hyperparameters would be difficult to define without an optimizer. Another example is the input size of 4 for forecasting 78 weeks in the N-BEATS model for O<sub>3</sub>. This configuration

achieved the best performance across all metrics – except for MAPE – while using two trend stacks.

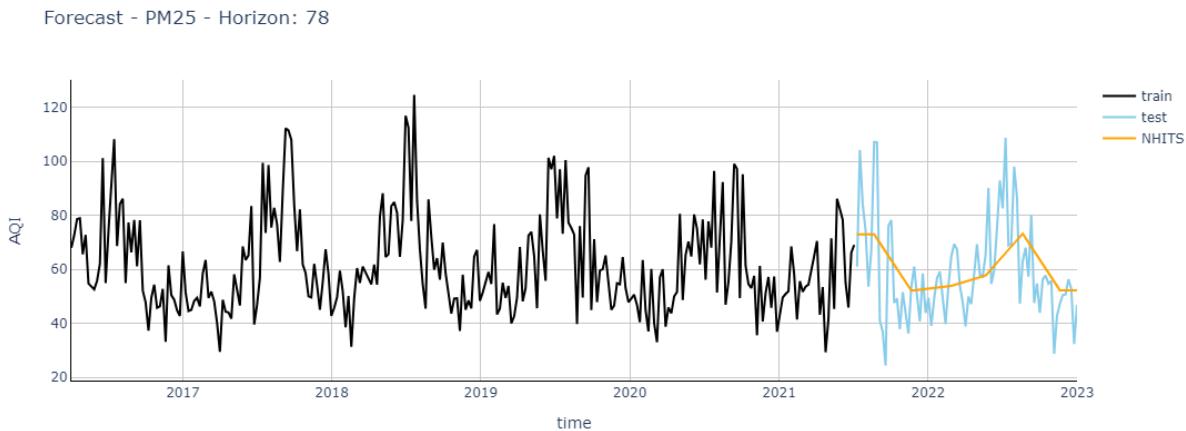
However, for both methods, the forecast accuracy does not show a noticeable decrease in effectiveness as the forecasting horizon increases, demonstrating that they can be used effectively for both short and long horizons.

#### 4.4 METRIC-SPECIFIC ANALYSIS

The sMAPE metric was used as the primary optimization criterion for the model because it is a metric that allows for direct comparison between different models, expressed in percentage terms. It is symmetric, meaning that errors in a series with high AQI values are weighted equally to errors in a series with low AQI values. This makes it intuitive, and less sensitive to outliers (which are present throughout the study), providing a more balanced and meaningful evaluation of model performance.

However, the metrics collected for each model serve different purposes. MAE (Mean Absolute Error) quantifies the average magnitude of errors in a set of forecasts, giving insight into the overall prediction accuracy without considering direction. MSE (Mean Squared Error) emphasizes larger errors by penalizing larger discrepancies more heavily, making it useful for identifying models that may have significant outliers. Max measures the largest forecast error, highlighting the worst-case scenario in the predictions. MAPE (Mean Absolute Percentage Error) offers a percentage-based error metric, useful for comparing models across different datasets or time scales, though it can be sensitive to small values in the data. Each metric provides a different perspective on model performance, helping to evaluate various aspects of forecasting quality.

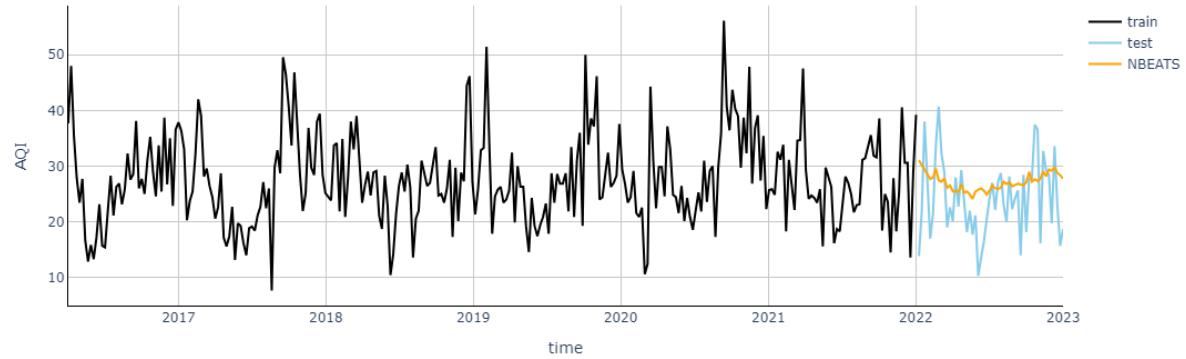
In cases where the goal is to minimize errors from outliers and extreme values, it is noticeable that models with greater smoothing can yield optimal results. For example, N-HiTS for forecasting PM<sub>2.5</sub> at 78 weeks was intermediate for sMAPE and MAE but achieved the best results for MAX and MSE. A similar trend was observed with N-BEATS for O<sub>3</sub> at 52 weeks, although it also performed well in other metrics. It is also observed that N-HiTS has a greater power of smoothing than the other models, and this is the leading quality that allows it to be a good long-horizon forecaster in good conditions.



**Figure 4.5** - N-HiTS time series forecasting for PM<sub>2.5</sub> in a 78-week prediction horizon.

Source: The Authors

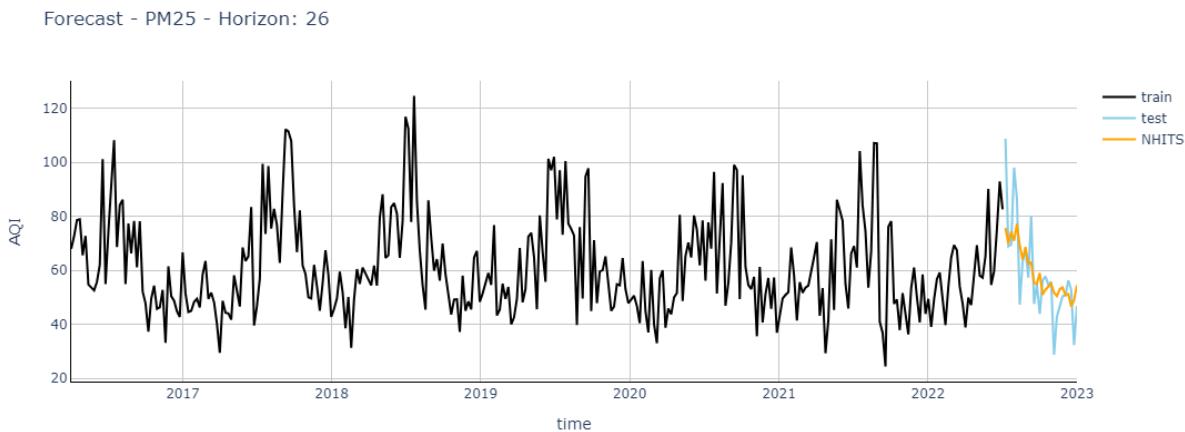
Forecast - O3 - Horizon: 52



**Figure 4.6** - N-BEATS time series forecasting for O<sub>3</sub> in a 52-week prediction horizon.

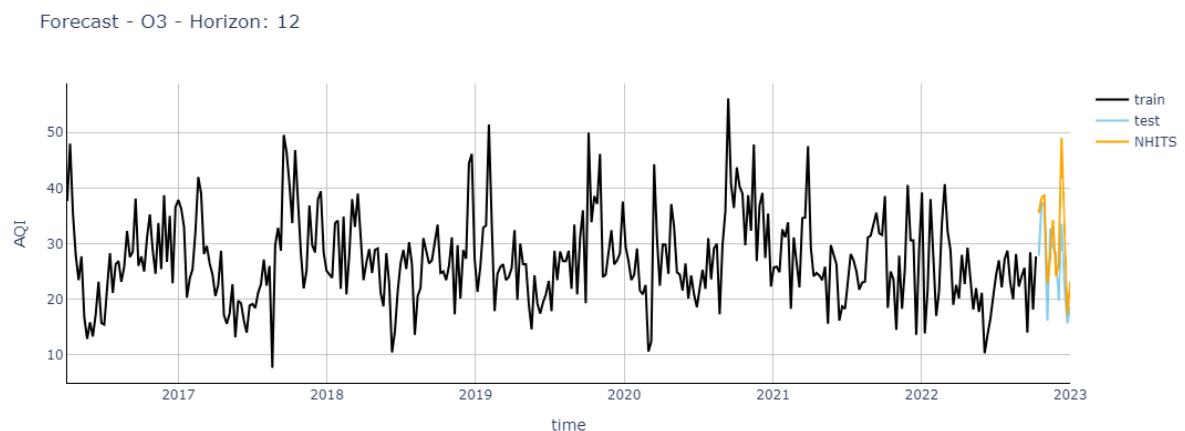
Source: The Authors

On the other hand, models that focus more on point-by-point accuracy tend to have more fluctuations and are more susceptible to larger errors, particularly with MAX. However, when they do succeed, they outperform smoothed models, as their predictions capture individual data points more precisely. This is the case for models such as N-HiTS for PM<sub>2.5</sub> at a 26-week horizon and O<sub>3</sub> at a 12-week horizon, and especially for N-BEATS models like NO<sub>2</sub> at a 26-week horizon and CO at a 52-week horizon.



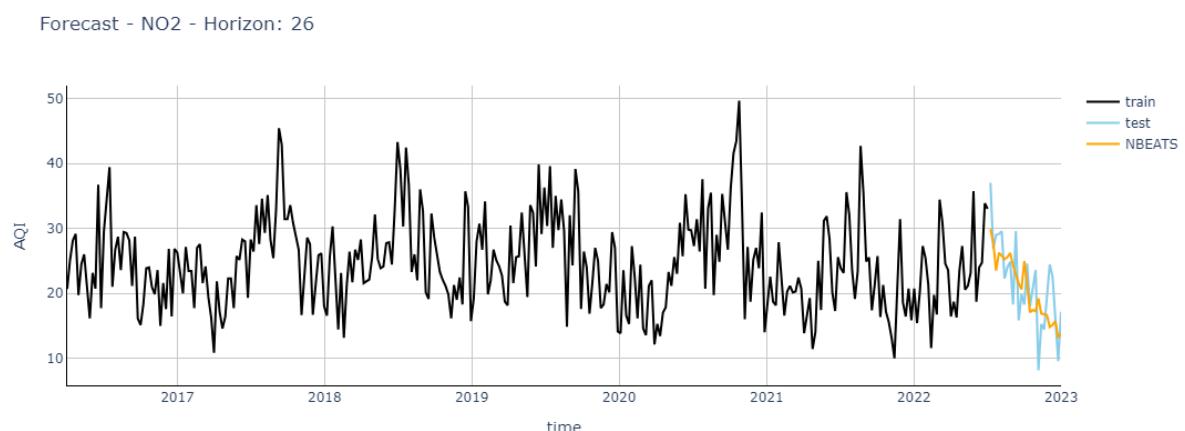
**Figure 4.7** - N-HiTS time series forecasting for PM<sub>2.5</sub> in a 26-week prediction horizon.

Source: The Authors



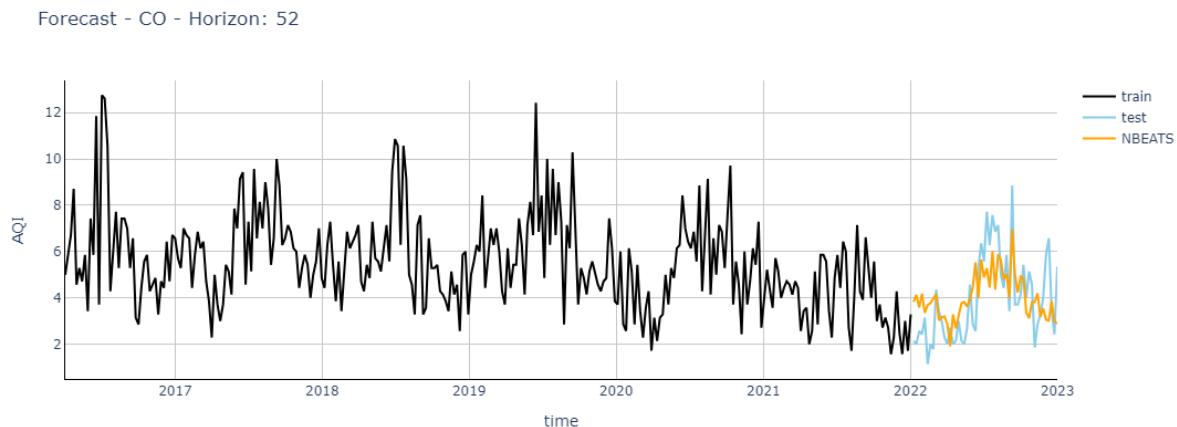
**Figure 4.8** - N-HiTS time series forecasting for O<sub>3</sub> in a 12-week prediction horizon.

Source: The Authors



**Figure 4.9** - N-BEATS time series forecasting for NO<sub>2</sub> in a 26-week prediction horizon.

Source: The Authors



**Figure 4.10** - N-BEATS time series forecasting for CO in a 52-week prediction horizon.

Source: The Authors

Therefore, during model optimization, it is important to focus on the desired metric to ensure that the forecasters meet their specific objectives. Alternatively, a more extensive and intensive optimization process can be performed to optimize multiple metrics simultaneously.

#### 4.5 NEURAL VS. STATISTICAL FORECASTING CAPACITY

In general, neural forecasters show great potential but still face significant challenges. One of the main challenges is the modeling time. Although neural forecasters are designed to be generalizable and interpretable, their hyperparameters still require extensive optimization to achieve impressive results and become competitive. While computational tools now allow for timely optimization, statistical methods are still incomparably faster than neural forecasters.

However, the ability to use auto differentiation in dense networks enables the capture of complex patterns that are often not captured by trend and seasonality models, making it a significant emulous advantage for neural forecasters – although many forecasts still benefit more from models with these specific features. This explains the growing interest in developing hybrid models that combine neural and statistical forecasters.

It is still difficult to predict the cases in which statistical models outperform neural models. What becomes evident is that, with properly optimized hyperparameters, neural models are competitive and can achieve highly beneficial results – especially since no single statistical model was found to perform equally well across all cases. Nevertheless, pure Deep

Learning models have proven to be promising and should be the focus of further research and technological investment.

#### 4.6 IMPLICATIONS FOR AIR QUALITY FORECASTING

Many models showed extremely satisfactory results in predicting the AQI for various pollutants, even ranking as the best model among all those studied. This was particularly the case for both models in the O<sub>3</sub> series, the PM<sub>2.5</sub> series with a 26-week horizon, the CO series with a 52-week horizon, and for N-BEATS in the NO<sub>2</sub> series with medium-term horizons.

These results highlight the potential of neural forecasters in not only improving the accuracy of air quality predictions but also in offering relevant insights that can inform public health responses. The ability to forecast air pollution levels with such precision allows for proactive measures, such as issuing warnings to vulnerable populations or implementing policies to mitigate exposure. For example, individuals with respiratory conditions, the elderly, and children can benefit from early warnings about elevated pollution levels, helping them take precautions and avoid health risks.

Moreover, these predictions can also be crucial for economic sectors that are sensitive to air quality, such as agriculture, transportation, and energy production. By anticipating pollution spikes, businesses can adjust operations or prepare for potential disruptions. The integration of neural forecasting models into real-time air quality monitoring systems could, therefore, contribute significantly to both public health and economic resilience, underscoring the importance of continued development and optimization of these models.

In the São Paulo region, PM<sub>2.5</sub> is the pollutant with the highest AQI, reflecting its significant impact on air quality and public health. The presence of fine particulate matter is closely linked to various sources of pollution, including vehicular emissions, industrial activities, and even natural factors like dust.

Carbon monoxide (CO) has the lowest AQI in the São Paulo region, primarily due to its relatively lower concentrations in comparison to other pollutants like PM<sub>2.5</sub>, O<sub>3</sub>, and NO<sub>2</sub>. CO is primarily produced by incomplete combustion processes, such as those from vehicles, industrial activities, and residential heating. However, in urban areas like São Paulo, while CO levels can still be significant, they are generally lower than particulate matter or ozone due to various factors.

One key reason for the lower AQI of CO is its shorter atmospheric lifetime. Unlike particulate matter, which can remain in the air for extended periods, CO typically disperses quickly and is diluted by wind and other meteorological factors. This makes CO levels more transient, with lower concentrations observed over time, especially in areas with good air circulation.

While CO still plays a role in air quality, its relatively lower AQI compared to other pollutants reflects its lower concentration and the successful mitigation of its sources in urban areas. However, continuous monitoring remains essential to ensure that CO levels stay within safe limits, particularly in densely populated areas with high traffic or industrial activity.

## 5 CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

In summary, the study demonstrated that neural forecasters are strong contenders for forecasting complex time series and can be an excellent tool for predicting the Air Quality Index (AQI). Additionally, it provided a deeper understanding of their hyperparameters and the significant efficiency improvements that can be achieved with proper optimization. Furthermore, a greater understanding of the issues caused by atmospheric pollutants was gained.

However, there remains a substantial opportunity to expand upon the present work. First, collecting a larger historical dataset and a greater diversity of pollutants would allow for the evaluation of models on additional datasets and enable the assessment of longer forecasting horizons with more historical data. This could potentially lead to a reinterpretation of hyperparameters, particularly for N-HiTS, as the broader context and more data may offer new insights into model performance.

Second, expanding the comparison to include more models from different classes, such as other machine learning models or even more general deep learning approaches, could provide valuable insights into how these alternative models perform on air quality forecasting tasks. This would help identify the most effective methods for capturing complex patterns in environmental data. Not only this, but a more in depth feature engineering and exogenous variables could make the model achieve better results.

Lastly, pursuing dynamic optimization for hyperparameters could offer greater flexibility in model adaptation, allowing for more precise tuning in real time or as new data becomes available. This would enhance the models' ability to generalize across varying conditions and improve their robustness in predicting AQI across different periods and pollutant types.

Overall, these extensions could further enhance the understanding of neural forecasters in the context of air quality prediction and lead to more robust, flexible, and scalable forecasting systems.

## REFERENCES

- AQICN. **Historical Air Quality Data.** Available at: <https://aqicn.org/historical/>. Accessed on: 27 dez. 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. (Environmental Engineering). ISBN 978-3-642-79003-4.
- BINBUSAYYIS, A. et al. **A Deep Learning Approach for Prediction of Air Quality Index in Smart City.** Discover Sustainability, v. 5, n. 1, 1 dez. 2024.
- BRUM, A. N. et al. **Impact of Temperature Increase on Air Pollutants - A Case Study in a Small City in Southern Brazil.** Case Studies in Chemical and Environmental Engineering, v. 9, 1 jun. 2024.
- CASTELLI, M. et al. **A Machine Learning Approach to Predict Air Quality in California.** Complexity, v. 2020, 4 ago. 2020.
- CHALLU, C. et al. **N-HiTS: Neural Hierarchical Interpolation for Time Series Forecasting.** 30 jan. 2022.
- FIORUCCI, J. A. et al. **Models for Optimising the Theta Method and Their Relationship to State Space Models.** International Journal of Forecasting, v. 32, n. 4, p. 1151–1161, 1 out. 2016.
- GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.** 2nd ed. Sebastopol: O'Reilly Media, Inc., 2019. ISBN 9781492032649.
- HUTTER, F.; HOOS, H.; LEYTON-BROWN, K. **An efficient approach for assessing hyperparameter importance.** In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 31., 2014, Beijing, China. JMLR: W&CP, v. 32, 2014.
- HYNDMAN, R.J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice.** 3rd ed. Melbourne: OTexts, 2021. Available at: <https://otexts.com/fpp3>. Accessed on: 27 dez. 2024.

MONIGATTI, L. **Stationarity in Time Series: A Comprehensive Guide.** Available at: <https://towardsdatascience.com/stationarity-in-time-series-a-comprehensive-guide-8beabe20d68>. Accessed on: 27 dez. 2024.

NIELSEN, A. **Practical Time Series Analysis: Prediction with Statistics and Machine Learning.** 1st ed. Sebastopol: O'Reilly Media, 2019. 497 p. ISBN 978-1492041658.

ORESHKIN, B. N. et al. **N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting.** 24 maio 2019.

OZAKI, Y.; TANIGAKI, Y.; WATANABE, S.; ONISHI, M. **Multiobjective Tree-structured Parzen Estimator for Computationally Expensive Optimization Problems.** In: GENETIC AND EVOLUTIONARY COMPUTATION CONFERENCE, GECCO '20, 2020, Cancún, Mexico. New York: ACM, 2020. Available at: <https://doi.org/10.1145/3377930.3389817>. Accessed on: 13 jan. 2025.

RAVINDIRAN, G. et al. **Air Quality Prediction by Machine Learning Models: A Predictive Study on the Indian Coastal City of Visakhapatnam.** Chemosphere, v. 338, 1 out. 2023.

SUN, Z. **Comparison of Trend Forecast Using ARIMA and ETS Models for S&P500 Close Price.** ACM International Conference Proceeding Series. Anais...Association for Computing Machinery, 9 out. 2020.

SVETUNKOV, I.; KOURENTZES, N.; ORD, J. K. **Complex Exponential Smoothing.** Naval Research Logistics, v. 69, n. 8, p. 1108–1123, 1 dez. 2022.

TENSORFLOW. **Time Series Forecasting with TensorFlow.** Available at: [https://www.tensorflow.org/tutorials/structured\\_data/time\\_series](https://www.tensorflow.org/tutorials/structured_data/time_series). Accessed on: 27 Dec. 2024.

WANG, J. et al. **An Air Quality Index Prediction Model Based on CNN-ILSTM.** Scientific Reports, v. 12, n. 1, 1 dez. 2022.

ZHANG, Z. et al. **A Systematic Survey of Air Quality Prediction Based on Deep Learning.** Alexandria Engineering JournalElsevier B.V., , 1 abr. 2024.

## APPENDIX A – STATISTICAL METHODS

This appendix provides a detailed explanation of the calculations and methodologies for all the statistical methods used throughout the study.

### A.1 ARIMA

The Autoregressive Integrated Moving Average (ARIMA) model is a cornerstone in time series analysis and forecasting. It combines three key aspects: autoregression (AR), differencing (integration, I), and moving average (MA). The AR component captures the relationship between an observation and a number of its lagged values, the I component involves differencing the data to achieve stationarity, and the MA component models the relationship between an observation and residual errors from a moving average model applied to lagged observations. The model is formally denoted as ARIMA(p, d, q), where p represents the order of the autoregressive part, d the degree of first differencing involved, and q the order of the moving average part (HYNDMAN, 2021).

The ARIMA modeling process begins with ensuring that the time series is stationary, meaning its statistical properties such as mean and variance are constant over time. If the data exhibit trends or seasonality, differencing procedures are applied to eliminate these effects, corresponding to the 'Integrated' part of the model. Once stationarity is achieved, the autoregressive and moving average parameters (p and q) are identified typically using autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. These plots help in selecting the appropriate number of lags for the AR and MA components by highlighting the correlations at different lags (HYNDMAN, 2021).

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (34)$$

### A.2 ERROR TREND SEASONALITY

The Exponential Smoothing State Space Model (ETS) represents a sophisticated approach to time series forecasting that elegantly integrates error, trend, and seasonal components within a unified statistical framework. This model extends traditional exponential smoothing methods by formulating them in a state space context, which not only provides a solid theoretical foundation but also facilitates the generation of accurate point forecasts and coherent prediction intervals (SUN, 2020).

In the ETS framework, a time series is decomposed into its fundamental elements: the level, which captures the underlying baseline of the series; the trend, which accounts for systematic upward or downward movements over time; and the seasonality, which reflects regular, repeating patterns at fixed intervals. Each of these components can interact additively or multiplicatively, allowing the model to adapt to various types of time series behaviors. Additive interactions are suitable when the magnitude of the components remains relatively constant over time, while multiplicative interactions are appropriate when the components scale with the level of the series (SUN, 2020).

Mathematically, for an additive ETS model with additive errors, the equations can be articulated as follows: the observed value is the sum of the previous level, trend, and seasonal components plus an error term. The level component updates based on the previous level and trend, adjusted by the level smoothing parameter and the error term. Similarly, the trend component evolves from the previous trend, modified by the trend smoothing parameter and the error term. The seasonal component adjusts based on the seasonal smoothing parameter and the error term. This structure allows the model to capture the dynamics of the time series comprehensively (SUN, 2020).

The ETS (Error-Trend-Seasonal) model has a clear taxonomy based on the forms of its components. When it assumes no special properties (N), it reduces to simple exponential smoothing. Alternatively, it can incorporate additive properties (A), damped additive trends ( $A_d$ ), multiplicative properties (M), multiplicative damped trends ( $M_d$ ), or optimized components (Z) (HYNDMAN, 2021).

Different equations are formulated for each of these models. The variable  $\ell$  represents the level component,  $b$  the trend component, and  $s$  the seasonal component. The parameter  $\Phi$  represents the damping factor, and all Greek letters denote estimated parameters. Figure A.1 shows the major possible combinations (HYNDMAN, 2021).

## ADDITIVE ERROR MODELS

Trend		Seasonal		
N		A		M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$		$y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$		$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$
<b>A<sub>d</sub></b>	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$		$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$

## MULTIPLICATIVE ERROR MODELS

Trend		Seasonal		
N		A		M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha \varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$		$y_t = \ell_{t-1} s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha \varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma \varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$		$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma \varepsilon_t)$
<b>A<sub>d</sub></b>	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$		$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma \varepsilon_t)$

**Figure A.1** - Equations for distinct combinations for the ETS model.

Source: (HYNDMAN, 2021)

## A.3 THETA

The Theta method's fundamental principle involves decomposing the original time series into two or more separate components, called theta lines, which are modified versions of the original data with different local curvatures (FIORUCCI, 2016).

In its basic form, the Theta method applies a coefficient  $\theta$  to the second differences of the data, effectively modifying the local curvature of the time series. The most common implementation, known as the standard Theta method, uses two theta lines: one with  $\theta = 0$ , which represents a linear regression line, and another with  $\theta = 2$ , which doubles the local curvatures of the original series. These theta lines are then extrapolated separately using

different forecasting techniques, and their forecasts are combined with equal weights to produce the final forecast (FIORUCCI, 2016).

Considering  $Y_1, \dots, Y_n$  the original deseasonalized time series and  $\nabla$  the difference operator, the Theta Method, is given by the following equations (FIORUCCI, 2016).

$$\nabla^2 Z_t(\theta) = \theta \nabla^2 Y_t, \quad t = 3, \dots, n \quad (35)$$

$$Z_t(\theta) = \theta Y_t + (1 + \theta)(A_n + B_n t), \quad t = 1, \dots, n \quad (36)$$

$$A_n = \frac{1}{n} \sum_{t=1}^n Y_t - \frac{n+1}{2} B_n \quad (37)$$

$$B_n = \frac{6}{n^2 - 1} \left( \frac{2}{n} \sum_{t=1}^n t \cdot T_t - \frac{1+n}{n} \sum_{t=1}^n Y_t \right) \quad (38)$$

#### A.4 COMPLEX EXPONENTIAL SMOOTHING

The conventional exponential decay of weights in time series models may not adequately capture real-world phenomena, where historical observations might exhibit more complex patterns of influence, including harmonic declines where certain past observations carry greater weight than more recent ones. To accommodate such complex weight distributions, we extend the traditional framework by introducing complex variables into the formulation (SVETUNKOV, 2022).

The transformation begins by replacing the real-valued observation  $y_{t-j}$  with its complex counterpart  $y_{t-j} + i e_{t-j}$ , where  $e_t$  represents the model's error term and  $i$  denotes the imaginary unit (defined by  $i^2 = -1$ ). This complex representation enables the simultaneous incorporation of both actual values and their associated errors in the forecasting mechanism. Furthermore, we introduce complexity into the smoothing parameter by replacing the real-valued  $\alpha$  with the complex parameter  $\alpha_0 + i \alpha_i$ , and extending unity to  $1 + i$  to facilitate harmonic weight decay (SVETUNKOV, 2022).

The interaction of these complex components generates weight distributions that can manifest various temporal patterns, including but not limited to exponential decay, oscillatory behavior, and harmonic progression. Given that the product of complex numbers yields another complex number, we represent the forecast term  $\hat{y}_{t-j}$  as  $\hat{y}_{t-j} + i \hat{e}_{t-j}$ , where  $\hat{e}_{t-j}$

serves as an error term proxy (SVETUNKOV, 2022). The resulting Complex Exponential Smoothing (CES) model can be expressed mathematically as:

$$\hat{y}_t + i\hat{e}_t = (\alpha_0 + i\alpha_1) \sum_{j=1}^{t-1} (1 + i - (\alpha_0 + i\alpha_1))^{j-1} (y_{t-j} + ie_{t-j}) \quad (39)$$

This formulation provides a more flexible and comprehensive framework for capturing intricate temporal dependencies in time series data, allowing for more nuanced modeling of real-world phenomena where the influence of historical observations follows complex patterns (SVETUNKOV, 2022).

## APPENDIX B – FANOVA

Complex deep learning models are highly dependent on their numerous hyperparameters, as a minor change can differentiate a conventional model from a state-of-the-art one. Optimization becomes essential because grid search is not scalable. Problems with low dimensionality and numeric hyperparameters are well-suited to Bayesian optimizers. Heuristic optimization methods have also become increasingly efficient for combinatorial hyperparameter optimization (HUTTER, 2014).

Although optimizers can produce excellent results, they do not inherently provide interpretability for the optimization process, creating a need for such insight. Functional analysis of variance (functional ANOVA) decomposes the variance  $V$  of a black-box function  $f : \Theta_1 \times \dots \times \Theta_n \rightarrow \mathbb{R}$  into additive components  $V_U$  (HUTTER, 2014).

When applying this new method to quantify the importance of hyperparameters in machine learning algorithms and combinatorial optimization procedures, following Hutter et al. (2011), we consider a setting slightly more general than black-box function optimization: given an algorithm A with configuration space  $\Theta$ , a set of training scenarios  $\pi_1, \dots, \pi_k$ , and a performance metric  $m(\theta, \pi)$  that captures A's performance with hyperparameter configuration  $\theta \in \Theta$  on scenario  $\pi$ , the goal is to find a configuration  $\theta \in \Theta$  that minimizes  $m$  over  $\pi_1, \dots, \pi_k$ , i.e., that minimizes the function

$$f(\theta) := \sum_{i=1}^k m(\theta, \pi_i)$$

(HUTTER, 2014, p. 2).

The fANOVA algorithm begins by collecting a set of function samples at different hyperparameter configurations and constructs a regression tree model to approximate the function of interest. Using this model, the algorithm decomposes the total variance into additive components for each hyperparameter, estimating the portion of the total variance explained by each term. This estimation is derived from summing the model's conditional responses, thereby isolating the effect of each individual hyperparameter (HUTTER, 2014).

$$V = \sum_{U \subset N} V_U, \quad \text{where } V_U = \frac{1}{|\Theta_U|} \int \hat{f}_U(\theta_U)^2 d\theta_U \quad (40)$$

## APPENDIX C – RESULTS TABLES

This appendix contains all the results for the metrics of the 19 statistical models, as well as N-HiTS and N-BEATS, for each pollutant and across all forecasting horizons sorted by sMAPE.

**Table C.1** - Metrics results for all 21 methods used in CO pollutant horizons.

CO   12 Weeks Forecast						CO   26 Weeks Forecast					
method	smape	mae	max	mape	mse	method	smape	mae	max	mape	mse
AutoETS-ZNA	24.89825	1.08606	2.19688	30.34291	1.62110	AutoETS-ZNA	21.36007	1.03179	2.20909	24.50068	1.60446
AutoTheta-Multi	24.92643	1.08936	2.43284	27.70377	1.73129	AutoCES-S	21.79861	1.04752	2.19877	25.49829	1.53900
AutoETS-ZNM	25.58223	1.10997	2.01239	31.45860	1.66209	AutoETS-ZNM	21.92618	1.06031	2.08637	25.18853	1.62310
AutoCES-S	25.99763	1.13074	2.21889	33.01316	1.67015	NHTS	<b>27.79316</b>	<b>1.38275</b>	<b>3.26325</b>	<b>30.55455</b>	<b>2.61440</b>
<b>NBEATS</b>	<b>26.56401</b>	<b>1.12545</b>	<b>3.52487</b>	<b>27.44780</b>	<b>2.71418</b>	AutoCES-N	28.27537	1.41958	4.43133	29.86320	3.14328
AutoTheta-Add	27.08290	1.18473	2.46277	28.35081	1.95795	AutoETS-ZZZ	28.91501	1.45238	4.07414	33.71978	3.05047
<b>NHITS</b>	<b>27.20602</b>	<b>1.16666</b>	<b>2.76189</b>	<b>31.67121</b>	<b>1.96824</b>	AutoETS-ZZN	28.91501	1.45238	4.07414	33.71978	3.05047
AutoCES-Z	27.33860	1.19544	2.52749	28.25192	2.01730	AutoETS-ZNZ	28.91501	1.45238	4.07414	33.71978	3.05047
AutoCES-P	27.33860	1.19544	2.52749	28.25192	2.01730	AutoARIMA	28.92279	1.45238	4.14431	33.22512	3.08302
AutoCES-N	29.63030	1.30465	2.93478	42.50113	2.30758	AutoTheta-Multi	33.42988	1.61161	4.04285	28.65411	4.02968
AutoETS-ZNZ	30.22121	1.34102	3.04686	44.72782	2.50598	AutoETS-ZAZ	33.98782	1.77354	4.03155	46.78798	4.32116
AutoETS-ZNN	30.22121	1.34102	3.04686	44.72782	2.50598	AutoETS-ZAN	34.04843	1.77765	4.04007	46.91889	4.33774
AutoARIMA	30.28516	1.34485	3.05836	44.91081	2.52434	<b>NBEATS</b>	<b>38.13497</b>	<b>1.73641</b>	<b>4.83805</b>	<b>31.88950</b>	<b>4.59865</b>
AutoETS-ZZZ	32.44155	1.48383	3.33963	51.43317	3.31401	AutoTheta-Add	41.60929	1.80218	3.93368	33.93770	4.59754
AutoETS-ZAZ	32.44155	1.48383	3.33963	51.43317	3.31401	AutoCES-Z	44.32542	1.88062	3.92155	35.61063	4.96474
AutoETS-ZZN	32.44554	1.48408	3.34034	51.44290	3.31527	AutoCES-P	44.32542	1.88062	3.92155	35.61063	4.96474
AutoETS-ZAN	32.44554	1.48408	3.34034	51.44290	3.31527	AutoETS-ZZM	59.79300	2.44817	5.33726	45.93693	7.79392
AutoETS-ZZM	32.88189	1.38798	3.24534	29.63046	3.01348	AutoETS-ZAM	59.79300	2.44817	5.33726	45.93693	7.79392
AutoETS-ZAM	32.88189	1.38798	3.24534	29.63046	3.01348	AutoETS-ZZA	67.45234	2.42319	4.65696	48.98937	7.48489
AutoETS-ZZA	42.30859	1.67448	3.47417	35.84860	4.05856	AutoETS-ZAA	67.45234	2.42319	4.65696	48.98937	7.48489
AutoETS-ZAA	42.30859	1.67448	3.47417	35.84860	4.05856						
CO   52 Weeks Forecast						CO   78 Weeks Forecast					
method	smape	mae	max	mape	mse	method	smape	mae	max	mape	mse
<b>NHITS</b>	<b>32.46016</b>	<b>1.29389</b>	<b>4.22526</b>	<b>33.18276</b>	<b>2.84438</b>	AutoTheta-Multi	32.27121	1.23658	3.69492	40.73419	2.29409
<b>NBEATS</b>	<b>32.72267</b>	<b>1.25873</b>	<b>3.60952</b>	<b>39.60075</b>	<b>2.26351</b>	AutoTheta-Add	33.22976	1.25365	4.24196	37.69623	2.48414
AutoTheta-Multi	33.37884	1.32545	4.62213	33.19242	2.86070	AutoETS-ZZM	35.14494	1.36102	4.38891	42.46259	2.84304
AutoTheta-Add	34.28817	1.29758	3.72869	35.15670	2.60405	AutoETS-ZAM	35.14494	1.36102	4.38891	42.46259	2.84304
AutoCES-Z	36.93523	1.37189	3.95345	35.90394	2.96532	AutoETS-ZZA	35.25966	1.37799	4.60221	46.91123	3.01706
AutoCES-P	36.93523	1.37189	3.95345	35.90394	2.96532	AutoETS-ZAA	35.25966	1.37799	4.60221	46.91123	3.01706
AutoETS-ZNA	40.58155	1.67728	5.30735	62.90774	4.08942	AutoCES-Z	37.01976	1.34663	4.30429	37.07075	2.87971
AutoETS-ZNM	41.01251	1.68830	5.26967	63.23621	4.02158	AutoCES-P	37.01976	1.34663	4.30429	37.07075	2.87971
AutoCES-S	41.69359	1.73764	5.68852	65.05120	4.32632	AutoCES-N	39.73874	1.54585	5.28604	49.33033	3.43038
AutoETS-ZZM	41.73753	1.59399	4.98591	40.73940	3.96101	AutoARIMA	41.78898	1.66236	4.23493	60.02543	3.61476
AutoETS-ZAM	41.73753	1.59399	4.98591	40.73940	3.96101	AutoETS-ZAZ	42.20639	1.68456	4.18389	61.16711	3.69175
AutoARIMA	41.98032	1.64306	6.19458	36.07173	5.01564	AutoETS-ZAN	42.23366	1.68603	4.18048	61.24231	3.69703
AutoETS-ZZZ	42.06347	1.64568	6.20138	36.07783	5.03326	AutoETS-ZZZ	42.25036	1.68692	4.17859	61.28854	3.70016
AutoETS-ZZM	42.06347	1.64568	6.20138	36.07783	5.03326	AutoETS-ZZM	42.25036	1.68692	4.17859	61.28854	3.70016
AutoETS-ZN2	42.06347	1.64568	6.20138	36.07783	5.03326	AutoETS-ZN2	42.25036	1.68692	4.17859	61.28854	3.70016
AutoETS-ZNN	42.06347	1.64568	6.20138	36.07783	5.03326	<b>NHITS</b>	<b>42.27536</b>	<b>1.72012</b>	<b>4.25614</b>	<b>60.59577</b>	<b>4.19717</b>
AutoETS-ZAZ	42.08602	1.64637	6.20285	36.08718	5.03717	AutoCES-S	44.17297	1.87275	5.73468	70.99617	5.01375
AutoETS-ZAN	42.10027	1.64681	6.20391	36.09028	5.03997	AutoETS-ZNM	44.82230	1.90702	6.07158	71.85905	5.08255
AutoETS-ZZA	45.38226	1.59273	4.24290	42.86022	3.94029	AutoETS-ZNA	44.83121	1.92052	6.05220	72.10977	5.22471
AutoETS-ZAA	45.38226	1.59273	4.24290	42.86022	3.94029	<b>NBEATS</b>	<b>44.93846</b>	<b>1.87298</b>	<b>4.42813</b>	<b>69.89092</b>	<b>4.60611</b>
AutoCES-N	50.41089	1.88058	6.65153	40.15671	6.34885						

Source: The Authors

**Table C.2** - Metrics results for all 21 methods used in NO<sub>2</sub> pollutant horizons.

NO <sub>2</sub>   12 Weeks Forecast						NO <sub>2</sub>   26 Weeks Forecast					
method	smape	mae	max	mape	mse	method	smape	mae	max	mape	mse
NBEATS	19.35326	3.68582	17.23598	30.57020	34.38209	NBEATS	20.04710	4.10677	10.41262	21.93479	25.61800
AutoTheta-Add	20.61730	3.71682	7.61952	24.08457	20.14518	AutoETS-ZAA	21.00792	4.51694	11.45969	26.75229	30.13842
AutoTheta-Multi	20.91795	3.73501	7.79041	26.53022	20.21888	AutoETS-ZAM	21.21413	4.51154	10.10165	26.88303	29.48900
AutoCES-Z	20.99396	3.78193	8.24851	23.52069	20.93008	AutoCES-Z	22.66273	4.98101	12.59554	29.40711	35.56542
AutoCES-P	20.99396	3.78193	8.24851	23.52069	20.93008	AutoCES-P	22.66273	4.98101	12.59554	29.40711	35.56542
<b>NHITS</b>	<b>23.12973</b>	<b>4.11706</b>	<b>9.70491</b>	<b>30.44112</b>	<b>24.99521</b>	AutoCES-S	23.16904	5.00271	12.68700	30.91803	36.00578
AutoETS-ZAA	23.32937	4.33420	10.72149	32.01612	28.42982	AutoETS-ZZM	24.54479	5.40832	11.86625	32.25806	39.69178
AutoETS-ZAM	25.36642	4.81609	9.92098	35.44764	33.10183	AutoETS-ZNM	24.54479	5.40832	11.86625	32.25806	39.69178
AutoCES-N	26.43192	4.90963	13.12811	40.06296	37.76772	AutoETS-ZZA	24.62280	5.46598	13.19592	32.66507	41.91195
AutoETS-ZZA	27.37900	5.31300	12.46172	39.66012	40.91443	AutoETS-ZNA	24.62280	5.46598	13.19592	32.66507	41.91195
AutoETS-ZNA	27.37900	5.31300	12.46172	39.66012	40.91443	AutoTheta-Add	24.99956	5.57633	13.58927	33.21944	43.41843
AutoCES-S	27.94570	5.41147	12.56811	41.61006	44.31287	AutoTheta-Multi	25.18528	5.63185	13.39186	33.50899	44.01221
AutoETS-ZZZ	28.29415	5.38109	13.86138	44.22638	46.13554	AutoARIMA	25.42046	5.49716	15.28473	35.60053	45.71868
AutoETS-ZZN	28.29415	5.38109	13.86138	44.22638	46.13554	<b>NHITS</b>	<b>28.56887</b>	<b>6.42333</b>	<b>16.85192</b>	<b>41.27299</b>	<b>60.53444</b>
AutoETS-ZNZ	28.29415	5.38109	13.86138	44.22638	46.13554	AutoCES-N	30.16430	6.83465	18.64014	44.70729	69.39517
AutoETS-ZNN	28.29415	5.38109	13.86138	44.22638	46.13554	AutoETS-ZZZ	33.57766	7.88653	20.24566	51.27665	90.32945
AutoETS-ZZM	29.15440	5.78123	11.73997	42.40025	46.32840	AutoETS-ZZN	33.57766	7.88653	20.24566	51.27665	90.32945
AutoETS-ZNM	29.15440	5.78123	11.73997	42.40025	46.32840	AutoETS-ZZP	33.57766	7.88653	20.24566	51.27665	90.32945
AutoETS-ZAZ	30.56770	5.98752	14.56301	49.07181	57.10476	AutoETS-ZNN	33.57766	7.88653	20.24566	51.27665	90.32945
AutoETS-ZAN	30.59013	5.99363	14.56887	49.11765	57.21061	AutoETS-ZAZ	39.28238	9.82336	23.10341	63.21875	136.86635
AutoARIMA	31.13595	6.12611	14.08968	49.37144	57.25235	AutoETS-ZAN	39.56542	9.92332	23.23101	63.79800	139.22958
NO <sub>2</sub>   52 Weeks Forecast						NO <sub>2</sub>   78 Weeks Forecast					
method	smape	mae	max	mape	mse	method	smape	mae	max	mape	mse
AutoETS-ZAA	20.93117	4.52018	17.53883	20.27477	33.67513	AutoTheta-Multi	22.02803	4.85587	23.08251	22.93882	40.47842
AutoETS-ZAM	21.05063	4.59769	16.86638	20.47409	34.27567	AutoTheta-Add	23.56354	5.10969	24.25762	23.26739	45.09265
AutoCES-Z	21.08129	4.54697	17.60069	20.70771	34.03772	AutoETS-ZZZ	24.09333	5.41525	19.33011	28.74479	46.97736
AutoCES-P	21.08129	4.54697	17.60069	20.70771	34.03772	AutoETS-ZZN	24.09333	5.41525	19.33011	28.74479	46.97736
AutoETS-ZZA	21.95212	4.91348	14.39983	26.45200	36.45116	AutoETS-ZNP	24.09333	5.41525	19.33011	28.74479	46.97736
AutoETS-ZNA	21.95212	4.91348	14.39983	26.45200	36.45116	AutoETS-ZNN	24.09333	5.41525	19.33011	28.74479	46.97736
AutoCES-S	21.96164	4.85540	13.10249	26.52861	33.84463	AutoETS-ZAM	24.24642	5.52829	17.49471	28.84775	45.64446
AutoETS-ZZM	22.38096	4.98860	14.09895	26.58487	35.30120	AutoETS-ZAA	24.24775	5.57136	19.21023	29.20855	47.41714
AutoETS-ZNM	22.38096	4.98860	14.09895	26.58487	35.30120	AutoCES-N	24.56966	5.47483	20.15323	24.75500	51.67958
<b>NHITS</b>	<b>22.49183</b>	<b>4.96158</b>	<b>16.72252</b>	<b>26.17384</b>	<b>36.81941</b>	AutoCES-S	24.73888	5.65784	17.08933	30.34873	47.79460
AutoARIMA	23.60548	5.27928	14.13214	29.73588	39.60622	AutoETS-ZZA	24.96507	5.76815	19.03308	30.59917	50.42073
AutoETS-ZAN	24.44304	5.47442	15.98521	28.87449	48.33409	AutoETS-ZNA	24.96507	5.76815	19.03308	30.59917	50.42073
AutoETS-ZAZ	24.51614	5.49597	16.13210	29.08962	48.60420	AutoETS-ZZM	25.27043	5.82696	16.81146	30.91771	50.14621
<b>NBEATS</b>	<b>24.95149</b>	<b>5.37297</b>	<b>15.31680</b>	<b>26.50524</b>	<b>44.90472</b>	AutoETS-ZNM	25.27043	5.82696	16.81146	30.91771	50.14621
AutoETS-ZZZ	25.55010	5.63369	17.82405	25.54027	51.80762	AutoARIMA	26.03093	5.92942	19.27662	32.56704	53.57444
AutoETS-ZZN	25.55010	5.63369	17.82405	25.54027	51.80762	<b>NBEATS</b>	<b>26.89697</b>	<b>6.18421</b>	<b>21.66166</b>	<b>33.13167</b>	<b>58.98654</b>
AutoETS-ZN0	25.55010	5.63369	17.82405	25.54027	51.80762	AutoCES-Z	29.42606	6.09228	24.85552	25.86171	60.80146
AutoETS-ZNN	25.55010	5.63369	17.82405	25.54027	51.80762	AutoCES-P	29.42606	6.09228	24.85552	25.86171	60.80146
AutoTheta-Multi	25.65044	5.65364	17.90241	25.53516	52.20805	<b>NHITS</b>	<b>29.43881</b>	<b>6.91954</b>	<b>19.44538</b>	<b>37.53492</b>	<b>68.64623</b>
AutoTheta-Add	25.65044	5.65364	17.90241	25.53516	52.20805	AutoETS-ZAN	34.28847	8.45036	24.68815	48.40221	101.96103
AutoCES-N	31.46113	6.73515	20.31454	28.02930	71.55946	AutoETS-ZAZ	34.41639	8.49331	24.78499	48.64655	102.92312

Source: The Authors

**Table C.3 - Metrics results for all 21 methods used in O<sub>3</sub> pollutant horizons.**

O <sub>3</sub>   12 Weeks Forecast						O <sub>3</sub>   26 Weeks Forecast					
method	smape	mae	max	mape	mse	method	smape	mae	max	mape	mse
NBEATS	17.29154	<b>4.76248</b>	7.63589	19.73952	28.23547	NHITS	19.32135	4.94071	<b>12.07812</b>	23.32120	<b>38.74121</b>
<b>NHITS</b>	<b>19.15312</b>	<b>5.85018</b>	<b>16.20169</b>	<b>23.60264</b>	<b>50.58697</b>	AutoARIMA	20.56468	5.29631	13.57660	24.98395	44.10201
AutoETS-ZAM	23.63952	6.60557	12.48650	30.46827	57.89326	<b>NBEATS</b>	<b>21.11663</b>	<b>5.39052</b>	<b>14.02995</b>	<b>26.05770</b>	<b>46.15378</b>
AutoTheta-Multi	24.05676	6.67765	13.43097	26.64872	58.00644	AutoTheta-Add	22.80393	6.09573	20.03322	28.74767	61.57701
AutoTheta-Add	24.10671	6.70269	14.41162	25.79861	60.36447	AutoCES-Z	23.18484	6.24174	20.60400	29.42838	64.57313
AutoETS-ZZM	24.21108	6.80787	13.61418	32.14811	64.97988	AutoCES-P	23.18484	6.24174	20.60400	29.42838	64.57313
AutoETS-ZNM	24.21108	6.80787	13.61418	32.14811	64.97988	AutoTheta-Add	23.65413	6.40203	21.26677	30.41771	67.76903
AutoCES-Z	24.55952	6.84736	15.57986	24.88485	66.80967	AutoETS-ZAM	23.92935	6.54107	20.75464	31.73251	71.37319
AutoCES-P	24.55952	6.84736	15.57986	24.88485	66.80967	AutoETS-ZZM	24.20874	6.63768	20.68457	32.43886	74.46341
AutoETS-ZAA	24.73739	6.97433	12.95824	31.90357	62.05319	AutoCES-S	24.42225	6.70399	20.01144	32.63786	74.91115
AutoARIMA	24.97330	6.95700	11.62340	30.03541	63.74514	AutoETS-ZZA	24.72706	6.79843	21.84945	32.96854	76.13074
AutoETS-ZZA	25.04787	7.09358	13.70535	33.05060	67.00217	AutoETS-ZZA	24.99475	6.89179	21.98166	33.56529	78.55783
AutoETS-ZNA	25.04787	7.09358	13.70535	33.05060	67.00217	AutoETS-ZNA	24.99475	6.89179	21.98166	33.56529	78.55783
AutoETS-ZNZ	25.65355	7.11115	13.71451	27.33670	63.79457	AutoETS-ZNN	26.63485	6.63136	17.55178	24.30733	65.39729
AutoETS-ZNN	25.65355	7.11115	13.71451	27.33670	63.79457	AutoETS-ZZZ	27.01586	6.71302	17.70345	24.53268	67.03789
AutoETS-ZZZ	25.70525	7.12495	13.79689	27.30038	64.27308	AutoETS-ZAZ	27.01586	6.71302	17.70345	24.53268	67.03789
AutoETS-ZZN	25.70525	7.12495	13.79689	27.30038	64.27308	AutoETS-ZAZ	27.01586	6.71302	17.70345	24.53268	67.03789
AutoETS-ZAZ	25.70525	7.12495	13.79689	27.30038	64.27308	AutoETS-ZAN	27.01586	6.71302	17.70345	24.53268	67.03789
AutoETS-ZAN	25.70525	7.12495	13.79689	27.30038	64.27308	AutoCES-N	32.66135	7.85959	19.99194	27.93882	92.89032
AutoCES-S	26.42963	7.56407	14.72714	35.25810	76.27480						
O <sub>3</sub>   52 Weeks Forecast						O <sub>3</sub>   78 Weeks Forecast					
method	smape	mae	max	mape	mse	method	smape	mae	max	mape	mse
AutoTheta-Multi	22.35501	5.63858	21.74843	26.64931	53.78683	<b>NBEATS</b>	<b>23.00007</b>	<b>5.82421</b>	<b>16.53123</b>	<b>28.03761</b>	<b>51.16080</b>
AutoTheta-Add	22.38963	5.63857	21.68314	25.87532	54.18467	<b>NHITS</b>	<b>23.06942</b>	<b>5.86367</b>	<b>18.51774</b>	<b>28.69973</b>	<b>55.14087</b>
AutoCES-Z	22.71863	5.65106	19.87966	25.12109	51.99078	AutoCES-S	23.57174	6.19990	21.68484	30.16933	66.91301
AutoCES-P	22.71863	5.65106	19.87966	25.12109	51.99078	AutoARIMA	23.60720	5.95052	17.31146	29.21786	53.87865
<b>NBEATS</b>	<b>23.35874</b>	<b>5.71842</b>	<b>17.12701</b>	<b>29.54159</b>	<b>50.74597</b>	AutoETS-ZZM	23.66459	6.24806	22.08258	30.39245	68.42871
AutoETS-ZZM	24.49414	6.29526	21.22560	30.89119	63.71468	AutoETS-ZNM	23.66459	6.24806	22.08258	30.39245	68.42871
AutoETS-ZNM	24.49414	6.29526	21.22560	30.89119	63.71468	AutoETS-ZZA	23.67089	6.23665	22.82374	30.26147	67.23888
AutoCES-S	24.55848	6.29281	20.26770	30.83833	63.94614	AutoETS-ZNA	23.67089	6.23665	22.82374	30.26147	67.23888
AutoETS-ZZA	24.62147	6.32487	22.42444	30.94088	64.39331	AutoCES-Z	23.71711	6.29249	24.40842	30.52833	69.31174
AutoETS-ZNA	24.62147	6.32487	22.42444	30.94088	64.39331	AutoCES-P	23.71711	6.29249	24.40842	30.52833	69.31174
AutoCES-ZNA	24.76737	6.08375	16.33794	29.54554	54.23059	AutoTheta-Add	25.23994	6.79620	25.95421	33.35729	79.15600
AutoETS-ZAA	24.98854	6.44046	22.98906	31.73879	66.56634	AutoETS-ZZZ	25.92667	6.43380	19.36190	24.73622	65.03163
AutoETS-ZAM	25.08003	6.48427	21.51663	32.13580	67.78069	AutoETS-ZZN	25.92667	6.43380	19.36190	24.73622	65.03163
AutoARIMA	25.71634	6.38775	17.32753	33.04855	60.96466	AutoETS-ZAZ	25.92667	6.43380	19.36190	24.73622	65.03163
<b>NHITS</b>	<b>26.02657</b>	<b>6.53235</b>	<b>20.91330</b>	<b>34.64532</b>	<b>67.14850</b>	AutoETS-ZAN	26.13368	6.47984	19.46706	24.82981	65.88655
AutoETS-ZZZ	29.75578	7.64156	19.89988	40.08219	81.14071	AutoETS-ZNZ	26.13368	6.47984	19.46706	24.82981	65.88655
AutoETS-ZZN	29.75578	7.64156	19.89988	40.08219	81.14071	AutoTheta-Add	26.21569	7.16606	27.28038	35.15442	87.90174
AutoETS-ZAZ	29.75578	7.64156	19.89988	40.08219	81.14071	AutoETS-ZAM	30.76077	8.75347	28.96440	42.77002	121.40966
AutoETS-ZAN	29.75578	7.64156	19.89988	40.08219	81.14071	AutoETS-ZAA	31.74684	8.98997	28.90678	44.39977	123.88805
AutoETS-ZNZ	30.59444	7.91460	20.29397	41.51078	86.00519	AutoCES-N	45.12270	9.94947	24.81012	35.95522	139.85382
AutoETS-ZNN	30.59444	7.91460	20.29397	41.51078	86.00519						

Source: The Authors

**Table C.4** - Metrics results for all 21 methods used in PM<sub>10</sub> pollutant horizons.

PM <sub>10</sub>   12 Weeks Forecast						PM <sub>10</sub>   26 Weeks Forecast					
method	smape	mae	max	mape	mse	method	smape	mae	max	mape	mse
AutoETS-ZAA	12.08282	2.33090	7.85979	15.51141	11.96371	AutoETS-ZAA	19.19188	5.30872	14.16923	23.21640	48.85419
AutoCES-Z	12.44258	2.40267	8.44559	16.26149	13.10412	AutoETS-ZAM	19.55163	5.36528	15.32325	23.55221	50.19438
AutoCES-P	12.44258	2.40267	8.44559	16.26149	13.10412	AutoCES-S	19.80832	5.31160	12.98603	24.72474	45.66286
AutoETS-ZAM	12.68651	2.44994	7.79319	16.27909	13.19058	AutoCES-Z	19.94849	5.53785	14.27091	24.58250	52.11677
AutoTheta-Multi	12.89419	2.52445	7.73717	16.01092	12.90712	AutoETS-ZZM	20.68464	5.69113	15.48426	26.31501	57.16580
AutoETS-ZZM	14.06534	2.76747	8.80028	18.73892	15.80509	AutoETS-ZNM	20.68464	5.69113	15.48426	26.31501	57.16580
AutoETS-ZNM	14.06534	2.76747	8.80028	18.73892	15.80509	AutoETS-ZZA	21.00112	5.77272	14.94734	26.36294	54.13158
AutoETS-ZZA	14.40857	2.84798	9.22936	19.28716	16.05698	AutoETS-ZNA	21.00112	5.77272	14.94734	26.36294	54.13158
AutoETS-ZNA	14.40857	2.84798	9.22936	19.28716	16.05698	<b>NHITS</b>	<b>22.23798</b>	<b>5.91001</b>	<b>19.45783</b>	<b>24.11750</b>	<b>56.49301</b>
AutoTheta-Add	14.58514	2.88534	7.76309	16.82531	13.11535	AutoTheta-Multi	22.60462	6.30407	15.98750	28.79857	62.19486
<b>NHITS</b>	<b>15.13263</b>	<b>3.02690</b>	<b>12.34681</b>	<b>20.65426</b>	<b>20.21119</b>	AutoTheta-Add	22.96791	6.36793	16.20504	29.33961	62.28904
AutoCES-S	15.42759	3.09212	9.78535	20.93884	18.82381	<b>NBEATS</b>	<b>24.18492</b>	<b>6.50554</b>	<b>26.03581</b>	<b>29.50578</b>	<b>84.80744</b>
AutoCES-N	18.10880	3.74724	13.10807	25.85020	26.72093	AutoARIMA	28.87501	7.80292	17.36611	38.31814	82.57601
<b>NBEATS</b>	<b>19.42658</b>	<b>3.92753</b>	<b>11.55366</b>	<b>25.35711</b>	<b>23.82812</b>	AutoCES-N	39.13789	11.29826	24.00380	58.10868	159.65321
AutoETS-ZZZ	21.86323	4.72749	14.12694	31.94257	37.77315	AutoETS-ZZZ	46.25028	14.25245	28.45989	73.92078	254.60880
AutoETS-ZZN	21.86323	4.72749	14.12694	31.94257	37.77315	AutoETS-ZZN	46.25028	14.25245	28.45989	73.92078	254.60880
AutoETS-ZNZ	21.86323	4.72749	14.12694	31.94257	37.77315	AutoETS-ZNN	46.25028	14.25245	28.45989	73.92078	254.60880
AutoETS-ZNN	21.86323	4.72749	14.12694	31.94257	37.77315	AutoETS-ZNN	46.25028	14.25245	28.45989	73.92078	254.60880
AutoARIMA	24.79517	5.53956	14.25935	36.34815	47.08373	AutoETS-ZAN	52.47175	17.27504	32.93581	90.01370	379.82276
AutoETS-ZAN	25.94166	5.88171	15.07360	38.70046	53.46165	AutoETS-ZAZ	52.52493	17.30228	32.97259	90.15323	380.98502
AutoETS-ZAZ	26.01122	5.90198	15.09344	38.81591	53.73987						
PM <sub>10</sub>   52 Weeks Forecast						PM <sub>10</sub>   78 Weeks Forecast					
method	smape	mae	max	mape	mse	method	smape	mae	max	mape	mse
AutoETS-ZAA	18.13249	4.85589	16.09169	20.02276	42.25891	AutoETS-ZAA	21.71142	5.97638	34.92364	26.42801	76.94611
AutoTheta-Multi	18.21126	4.87571	16.83621	19.84534	44.90825	AutoCES-Z	21.78226	5.97598	34.94582	26.35522	75.70440
AutoCES-P	18.39206	4.90429	14.77497	20.40625	41.91913	AutoCES-P	21.78226	5.97598	34.94582	26.35522	75.70440
AutoCES-Z	18.54263	4.93791	14.72630	20.85480	41.84812	<b>AutoTheta-Multi</b>	<b>21.85139</b>	<b>5.95594</b>	<b>37.79048</b>	<b>24.37863</b>	<b>77.12380</b>
AutoTheta-Add	18.77686	5.00194	17.38565	20.12429	45.66619	AutoETS-ZAM	22.09252	6.03524	31.45628	26.63088	72.33348
AutoETS-ZAM	18.97990	5.05266	15.05179	21.05321	43.20567	AutoETS-ZZA	22.74089	6.23853	34.19825	28.45613	79.75946
AutoETS-ZZA	19.47881	5.18246	17.65907	22.71603	44.19734	AutoETS-ZNA	22.74089	6.23853	34.19825	28.45613	79.75946
AutoETS-ZNA	19.47881	5.18246	17.65907	22.71603	44.19734	AutoCES-S	22.75612	6.18244	31.82724	27.94608	72.63332
AutoCES-S	19.86167	5.22929	14.30606	23.15747	42.62577	AutoETS-ZZM	22.88616	6.26100	31.43450	28.48942	75.60109
<b>NBEATS</b>	<b>19.97608</b>	<b>5.16013</b>	<b>17.94013</b>	<b>21.56901</b>	<b>44.25038</b>	AutoETS-ZNM	22.88616	6.26100	31.43450	28.48942	75.60109
AutoETS-ZZM	20.01770	5.32369	14.92899	23.44348	46.11952	AutoTheta-Add	23.58452	6.29051	38.88745	25.48672	81.22875
AutoETS-ZNM	20.01770	5.32369	14.92899	23.44348	46.11952	<b>NHITS</b>	<b>25.94896</b>	<b>6.97708</b>	<b>33.19374</b>	<b>30.19141</b>	<b>88.40331</b>
<b>NHITS</b>	<b>20.16401</b>	<b>5.22101</b>	<b>18.71591</b>	<b>21.54627</b>	<b>46.15025</b>	<b>NBEATS</b>	<b>26.59539</b>	<b>7.21587</b>	<b>35.64838</b>	<b>32.35436</b>	<b>96.80831</b>
AutoARIMA	26.05906	7.00151	25.00677	30.42067	77.46796	AutoCES-N	29.44282	7.91962	31.88632	29.64521	115.91146
AutoETS-ZAN	27.60869	7.47552	28.20909	30.52420	95.42637	AutoARIMA	30.14023	8.17187	33.34851	36.02719	103.34123
AutoETS-ZAZ	27.65506	7.49086	28.15291	30.64850	95.45961	AutoETS-ZZZ	32.47527	8.84711	30.53559	39.64929	111.32747
AutoETS-ZZZ	28.55719	7.57596	32.86738	25.83243	119.24451	AutoETS-ZZN	32.47527	8.84711	30.53559	39.64929	111.32747
AutoETS-ZZN	28.55719	7.57596	32.86738	25.83243	119.24451	AutoETS-ZNZ	32.47527	8.84711	30.53559	39.64929	111.32747
AutoETS-ZNN	28.55719	7.57596	32.86738	25.83243	119.24451	AutoETS-ZAN	32.47527	8.84711	30.53559	39.64929	111.32747
AutoCES-N	44.03798	10.52815	37.14154	35.33412	186.52827	AutoETS-ZAZ	45.72045	13.85360	32.94445	67.91357	249.46425
						AutoETS-ZAN	45.78092	13.88046	33.00284	68.04825	250.40024

Source: The Authors

**Table C.5** - Metrics results for all 21 methods used in PM<sub>2.5</sub> pollutant horizons.

PM <sub>2.5</sub>   12 Weeks Forecast						PM <sub>2.5</sub>   26 Weeks Forecast					
method	smape	mae	max	mape	mse	method	smape	mae	max	mape	mse
AutoETS-ZAA	10.11535	4.76016	14.42661	12.26340	42.57126	NHITS	14.63253	8.89594	32.79903	16.37173	155.16133
AutoETS-ZZA	10.37780	4.87270	15.94328	13.04160	49.10307	NBEATS	15.26479	9.33524	37.89914	17.14728	181.84932
AutoETS-ZNA	10.37780	4.87270	15.94328	13.04160	49.10307	AutoCES-S	16.18900	10.25305	25.45771	18.89918	180.42911
AutoETS-ZZM	10.61776	5.00365	16.04866	13.26546	51.24380	AutoETS-ZAA	16.19274	10.45176	30.13033	18.53366	200.48677
AutoETS-ZNM	10.61776	5.00365	16.04866	13.26546	51.24380	AutoETS-ZAM	16.89630	10.81111	29.66513	19.32569	205.70999
AutoETS-ZAM	10.90024	5.17125	14.55874	13.13209	46.67707	AutoCES-Z	16.96506	11.02500	28.38437	19.96703	215.56757
AutoCES-S	11.33413	5.39295	17.72600	14.44027	60.80736	AutoCES-P	16.96506	11.02500	28.38437	19.96703	215.56757
AutoTheta-Multi	11.41406	5.47995	13.38595	13.26542	46.66477	AutoETS-ZZA	17.05040	11.01999	29.40132	20.04062	212.03106
AutoTheta-Add	12.88096	6.27463	10.98842	14.19399	50.13133	AutoETS-ZNA	17.05040	11.01999	29.40132	20.04062	212.03106
AutoCES-N	13.56578	6.63845	24.78097	18.28864	98.98733	AutoETS-ZZM	17.44199	11.31118	32.09352	20.53511	221.34016
<b>NHITS</b>	<b>14.51509</b>	<b>7.27727</b>	<b>23.21763</b>	<b>18.50082</b>	<b>99.32117</b>	AutoETS-ZNM	17.44199	11.31118	32.09352	20.53511	221.34016
AutoCES-Z	14.90310	7.30849	11.23516	15.79707	61.88717	AutoTheta-Add	17.89178	11.60800	29.78513	21.25530	230.40716
AutoCES-P	14.90310	7.30849	11.23516	15.79707	61.88717	AutoTheta-Multi	18.13063	11.79266	29.83728	21.59738	235.48153
AutoETS-ZZZ	15.88873	8.03921	26.83635	21.93769	136.00685	AutoARIMA	22.27193	13.80845	34.37141	26.97277	269.37256
AutoETS-ZZN	15.88873	8.03921	26.83635	21.93769	136.00685	AutoCES-N	30.60860	19.77987	42.90058	40.33712	475.74187
AutoETS-ZNZ	15.88873	8.03921	26.83635	21.93769	136.00685	AutoETS-ZZZ	35.19718	23.52266	48.54728	48.44745	668.80250
AutoETS-ZNN	15.88873	8.03921	26.83635	21.93769	136.00685	AutoETS-ZN	35.19718	23.52266	48.54728	48.44745	668.80250
AutoETS-ZAZ	18.72772	9.83001	28.25151	26.02001	175.57600	AutoETS-ZNN	35.19718	23.52266	48.54728	48.44745	668.80250
AutoETS-ZAN	18.76059	9.85121	28.26345	26.06732	176.06295	AutoETS-ZAN	39.80118	27.65555	54.62784	57.36414	928.79518
AutoARIMA	18.92501	9.96111	26.77341	26.12363	175.26510	AutoETS-ZAZ	39.90641	27.75433	54.76691	57.57276	935.39476
<b>NBEATS</b>	<b>19.84084</b>	<b>10.54560</b>	<b>27.83623</b>	<b>27.38343</b>	<b>185.32766</b>						
PM <sub>2.5</sub>   52 Weeks Forecast						PM <sub>2.5</sub>   78 Weeks Forecast					
method	smape	mae	max	mape	mse	method	smape	mae	max	mape	mse
AutoETS-ZAA	15.09911	9.42521	33.65948	16.11659	160.64256	AutoETS-ZAM	18.64072	11.61933	59.28899	22.25933	281.12515
AutoETS-ZZA	15.11434	9.48997	36.52129	16.76730	165.59933	AutoETS-ZAA	18.70968	11.78952	61.84132	22.52747	293.19199
AutoETS-ZNA	15.11434	9.48997	36.52129	16.76730	165.59933	AutoCES-S	18.83142	11.74763	54.72548	22.69138	275.25595
AutoETS-ZAM	15.26206	9.42826	30.67827	16.18626	159.76828	AutoETS-ZZM	18.85288	11.81429	61.12144	23.12171	288.71884
AutoETS-ZZM	15.55795	9.68507	32.11376	17.16788	167.73780	AutoETS-ZNM	18.85288	11.81429	61.12144	23.12171	288.71884
AutoETS-ZNM	15.55795	9.68507	32.11376	17.16788	167.73780	AutoETS-ZZA	19.05671	12.01275	62.74781	23.38332	298.91449
AutoTheta-Multi	15.55902	9.67022	36.15954	16.42138	170.05883	AutoETS-ZNA	19.05671	12.01275	62.74781	23.38332	298.91449
AutoCES-S	15.58476	9.59944	25.33848	17.14909	155.87159	AutoTheta-Multi	19.59725	12.06239	57.62363	22.05715	293.06937
<b>NBEATS</b>	<b>16.06791</b>	<b>9.60111</b>	<b>45.48816</b>	<b>16.84733</b>	<b>187.68926</b>	<b>NHITS</b>	<b>19.88790</b>	<b>12.07724</b>	<b>43.54848</b>	<b>22.60644</b>	<b>255.25784</b>
AutoTheta-Add	16.26146	10.05148	37.27603	16.98788	174.79502	<b>NBEATS</b>	<b>21.90096</b>	<b>13.41827</b>	<b>47.53751</b>	<b>26.65842</b>	<b>298.99184</b>
AutoCES-Z	17.08354	10.30621	34.95526	17.05219	172.07074	AutoTheta-Add	22.61452	13.47416	56.86066	24.07421	323.01434
AutoCES-P	17.08354	10.30621	34.95526	17.05219	172.07074	AutoCES-N	23.23054	14.04435	56.73702	24.09480	356.54554
<b>NHITS</b>	<b>17.10547</b>	<b>10.19212</b>	<b>42.10567</b>	<b>17.67195</b>	<b>183.40929</b>	AutoARIMA	23.73424	14.43169	47.27400	27.42442	335.94981
AutoARIMA	19.16038	11.69497	47.68776	21.07835	239.73551	AutoETS-ZZZ	25.64626	15.65182	44.53463	30.27304	359.50467
AutoETS-ZAZ	20.72591	12.65539	53.21272	21.31983	294.51289	AutoETS-ZZN	25.64626	15.65182	44.53463	30.27304	359.50467
AutoETS-ZAN	20.74759	12.67026	53.13745	21.37295	294.33070	AutoETS-ZNZ	25.64626	15.65182	44.53463	30.27304	359.50467
AutoETS-ZZZ	22.43242	13.49483	59.62471	20.58922	361.33737	AutoETS-ZNN	25.64626	15.65182	44.53463	30.27304	359.50467
AutoETS-ZZN	22.43242	13.49483	59.62471	20.58922	361.33737	AutoCES-Z	30.26911	16.68018	59.05152	28.70238	431.77935
AutoETS-ZNZ	22.43242	13.49483	59.62471	20.58922	361.33737	AutoCES-P	30.26911	16.68018	59.05152	28.70238	431.77935
AutoETS-ZAN	22.43242	13.49483	59.62471	20.58922	361.33737	AutoETS-ZAN	36.46622	24.06790	58.12659	49.03550	738.57426
AutoCES-N	33.17526	18.51783	66.95080	28.02105	558.15177	AutoETS-ZAZ	36.59581	24.18208	58.38882	49.27595	745.55905

Source: The Authors

## APPENDIX D – TIME SERIES PROPERTIES

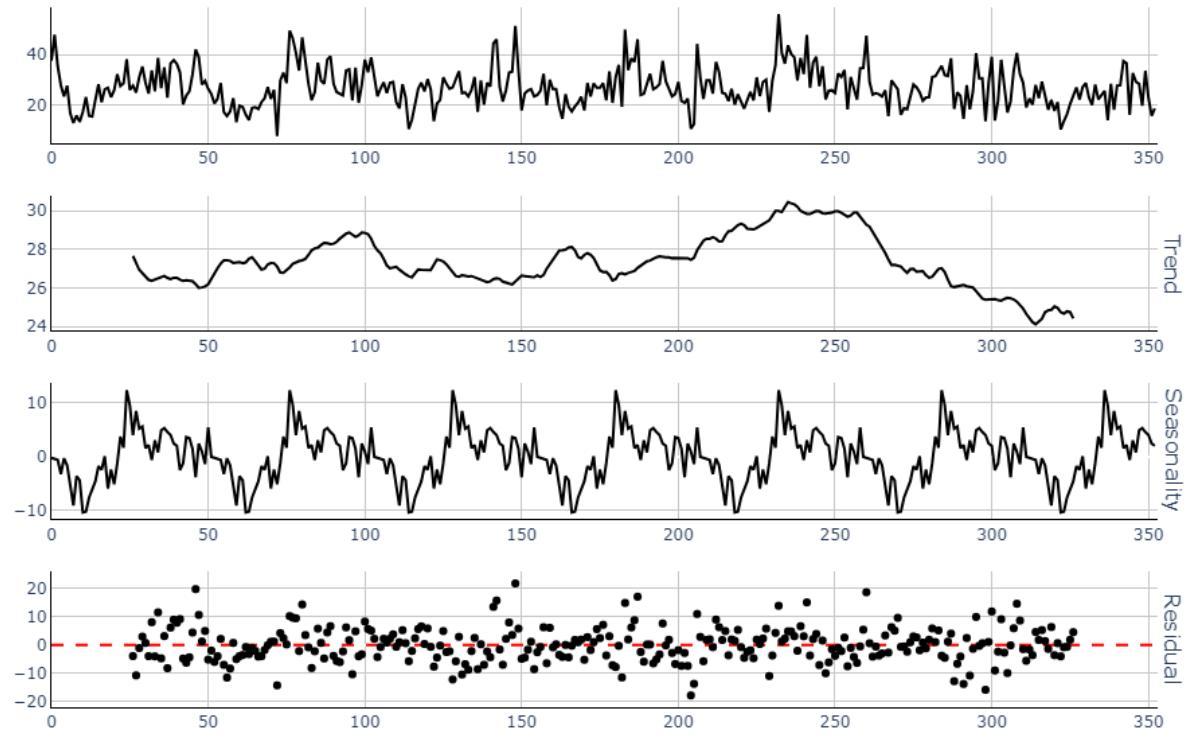
This appendix provides graphical information on each pollutant's weekly AQI time series. To achieve this, the components are subjected to (a) additive seasonal decomposition, which breaks down the series into trend, seasonality, and residual components.

The decomposition is performed using techniques such as classical decomposition or STL (Seasonal-Trend decomposition using LOESS), where each component is extracted iteratively, enabling a detailed analysis of the underlying structure of the time series. This approach aids in understanding the distinct characteristics of the AQI for each pollutant, providing valuable insights into patterns and irregularities in the data.

Additionally, it presents (b) the autocorrelation plot (ACF), which measures the correlation of the time series with its own lagged values, helping identify repeating patterns, such as seasonality and it also assesses how observations at different time points are related. Moreover, it shows (c) the partial autocorrelation (PACF) plot, which isolates the direct relationship between the time series and its lagged values by removing the influence of intermediate lags. It is particularly useful for determining the appropriate number of lag terms to include in autoregressive models, as it highlights the number of lags that have a significant direct impact on the series.

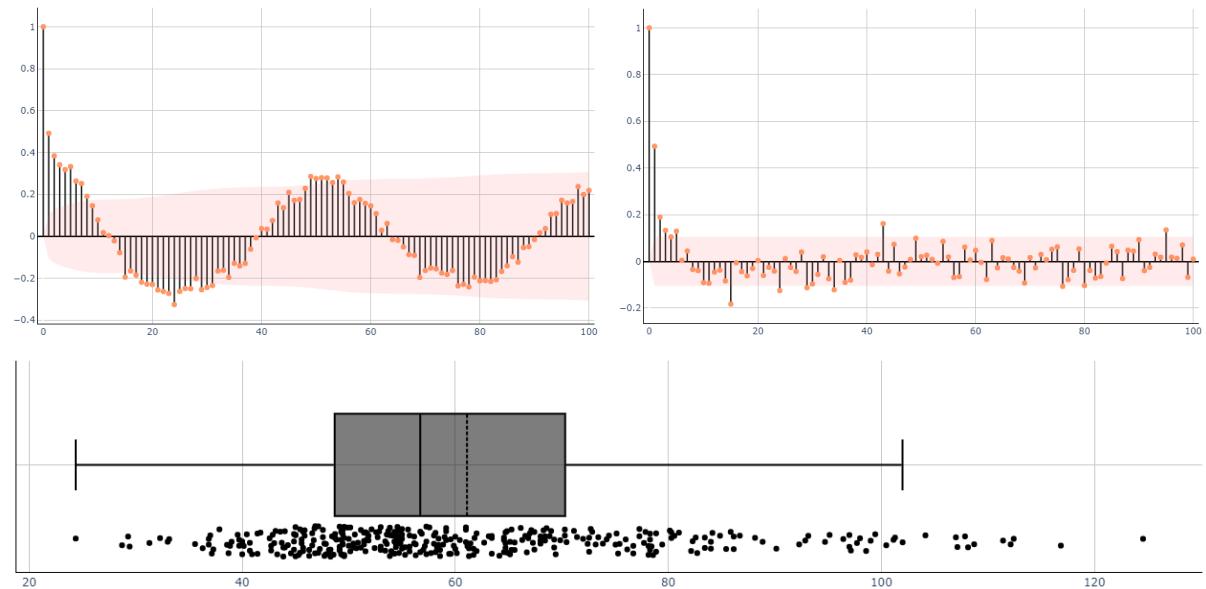
Finally, a (d) boxplot of the time series values for each pollutant is presented. This plot allows for the visualization of key frequentist statistical properties, providing insights into the distribution, variability, and potential outliers within the data. It serves as a valuable tool for understanding the overall behavior of the series and for identifying patterns or anomalies that may influence the forecasting process.

### Seasonal Decomposition - O3



Autocorrelation Function (ACF) - PM25

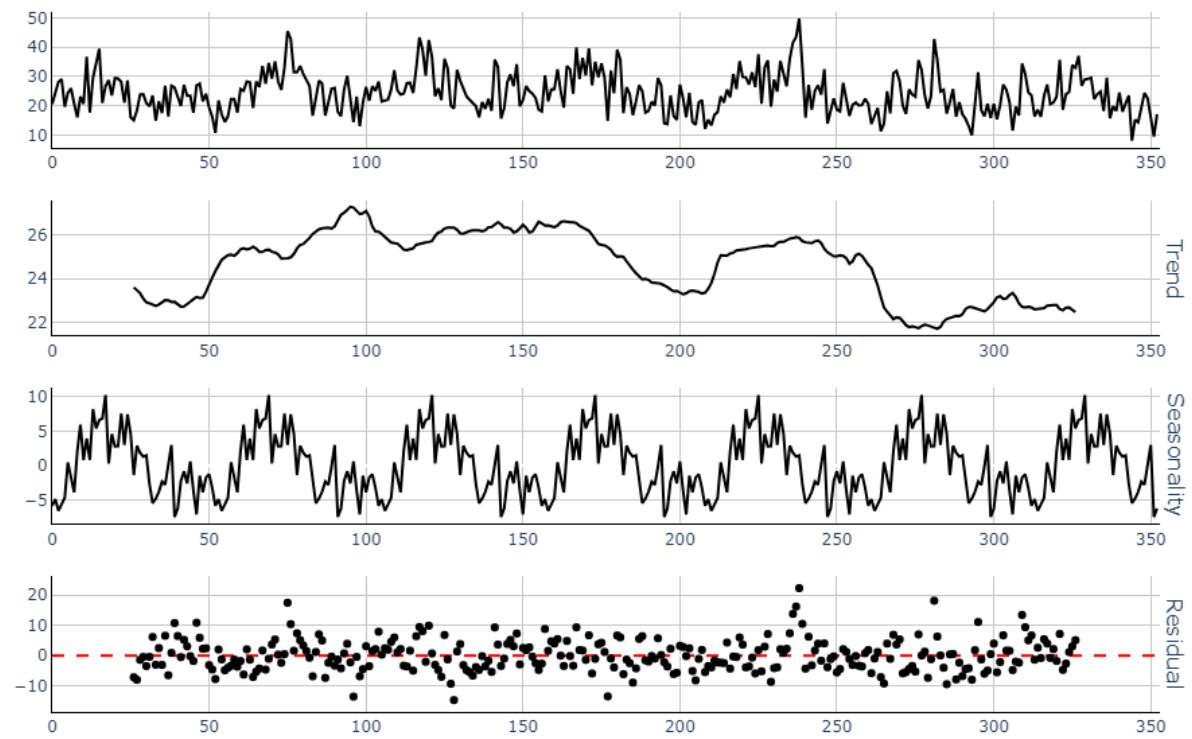
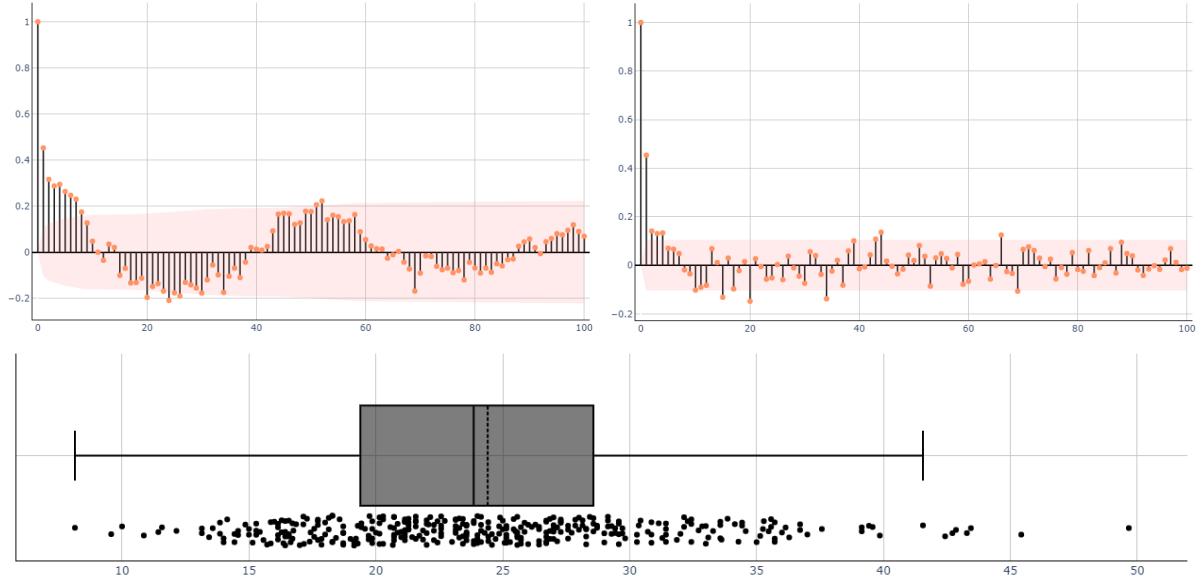
Partial Autocorrelation Function (PACF) - PM25



**Figure D.1** - O<sub>3</sub> time series properties. (a) seasonal decomposition, (b) autocorrelation function, (c) partial autocorrelation function, (d) boxplot dispersion.

Source: The Authors

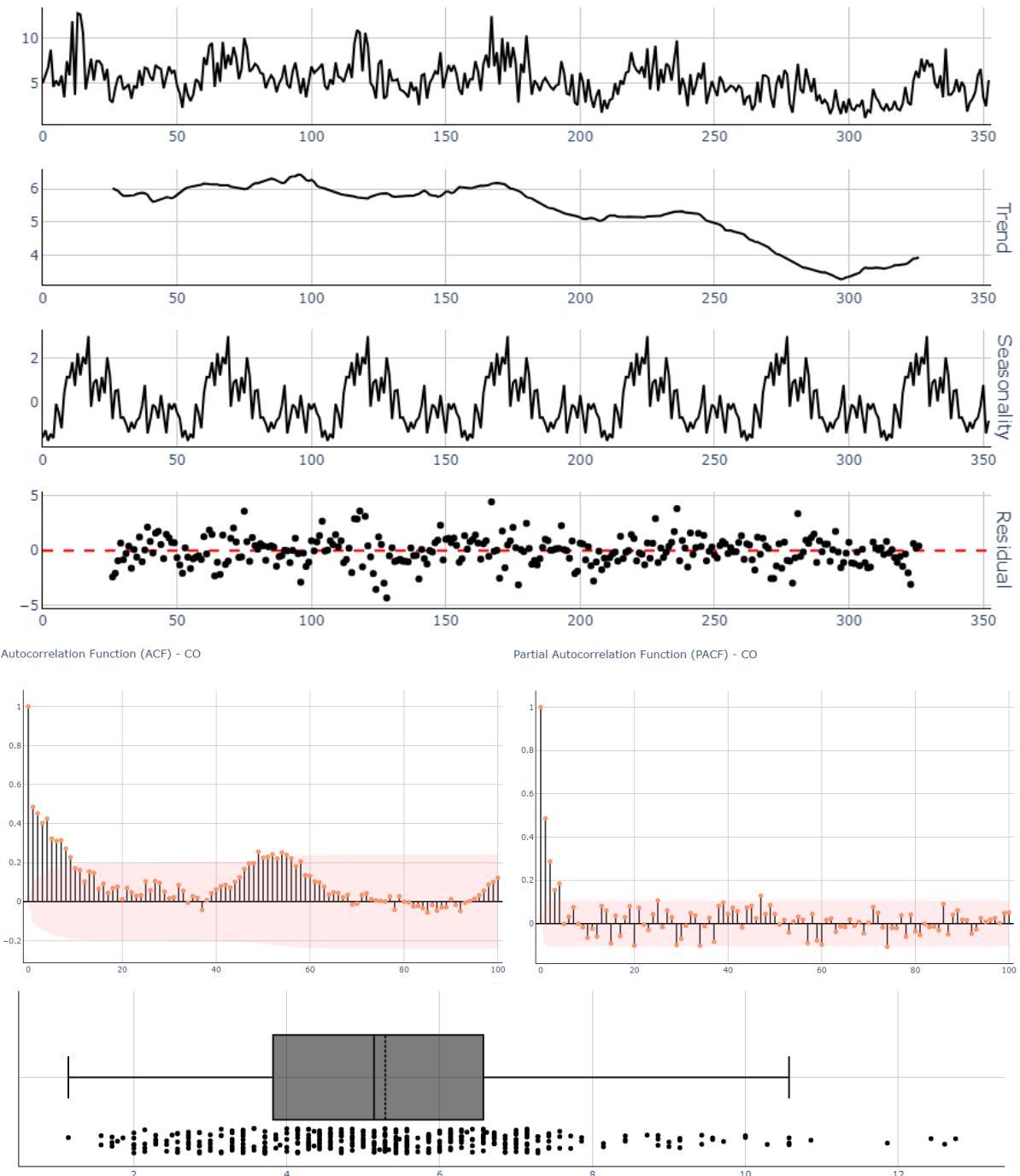
### Seasonal Decomposition - NO<sub>2</sub>

Autocorrelation Function (ACF) - NO<sub>2</sub>Partial Autocorrelation Function (PACF) - NO<sub>2</sub>

**Figure D.2** - NO<sub>2</sub> time series properties. (a) seasonal decomposition, (b) autocorrelation function, (c) partial autocorrelation function, (d) boxplot dispersion.

Source: The Authors

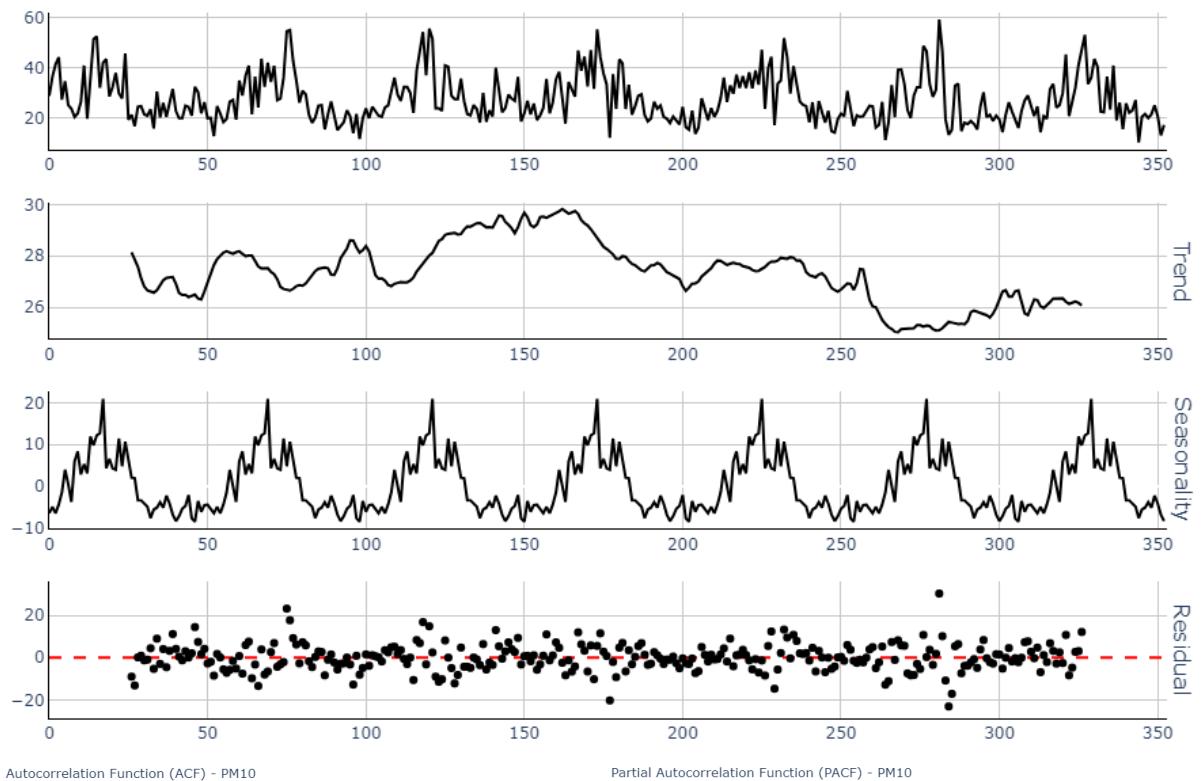
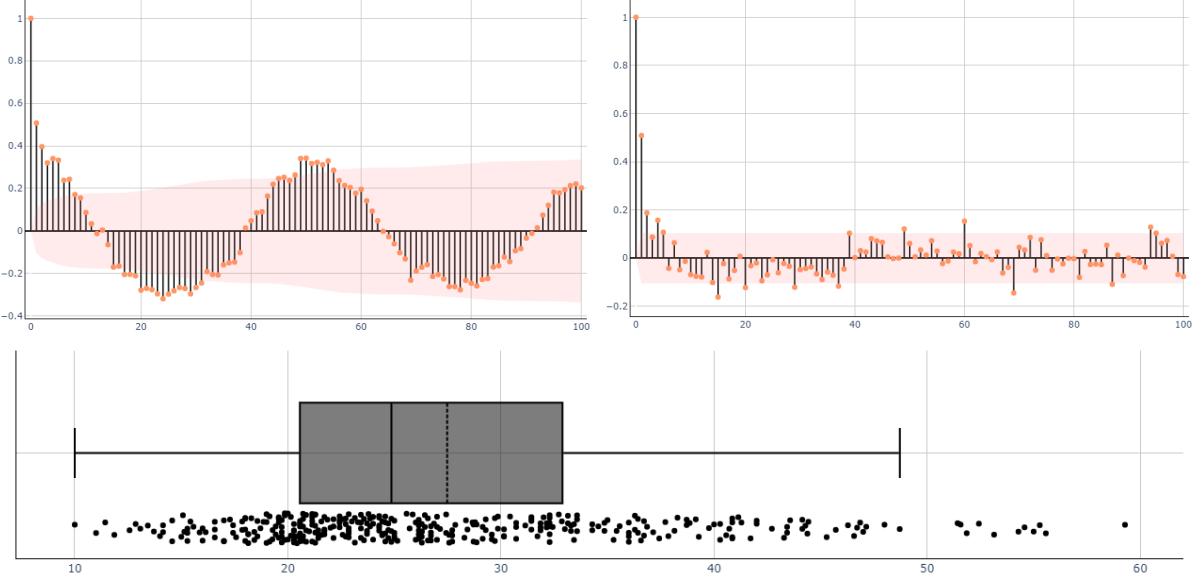
### Seasonal Decomposition - CO



**Figure D.3** - CO time series properties. (a) seasonal decomposition, (b) autocorrelation function, (c) partial autocorrelation function, (d) boxplot dispersion.

Source: The Authors

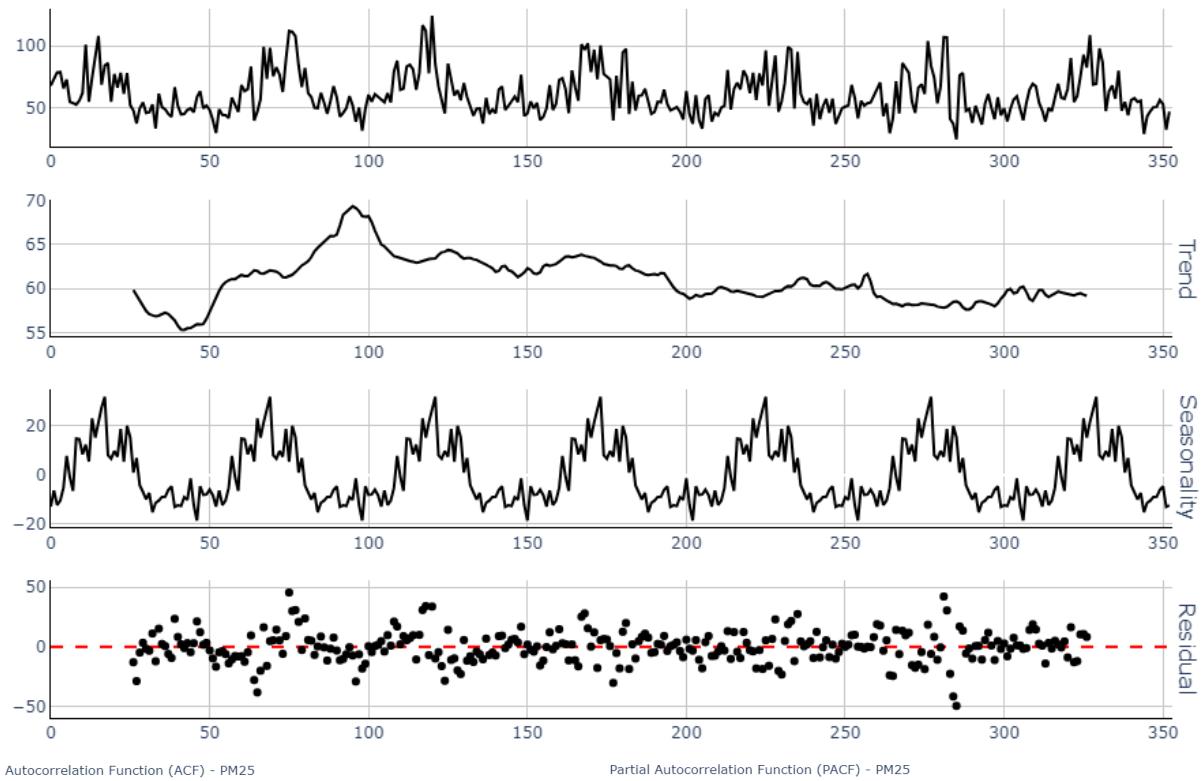
### Seasonal Decomposition - PM<sub>10</sub>

Autocorrelation Function (ACF) - PM<sub>10</sub>Partial Autocorrelation Function (PACF) - PM<sub>10</sub>

**Figure D.4** - PM<sub>10</sub> time series properties. (a) seasonal decomposition, (b) autocorrelation function, (c) partial autocorrelation function, (d) boxplot dispersion.

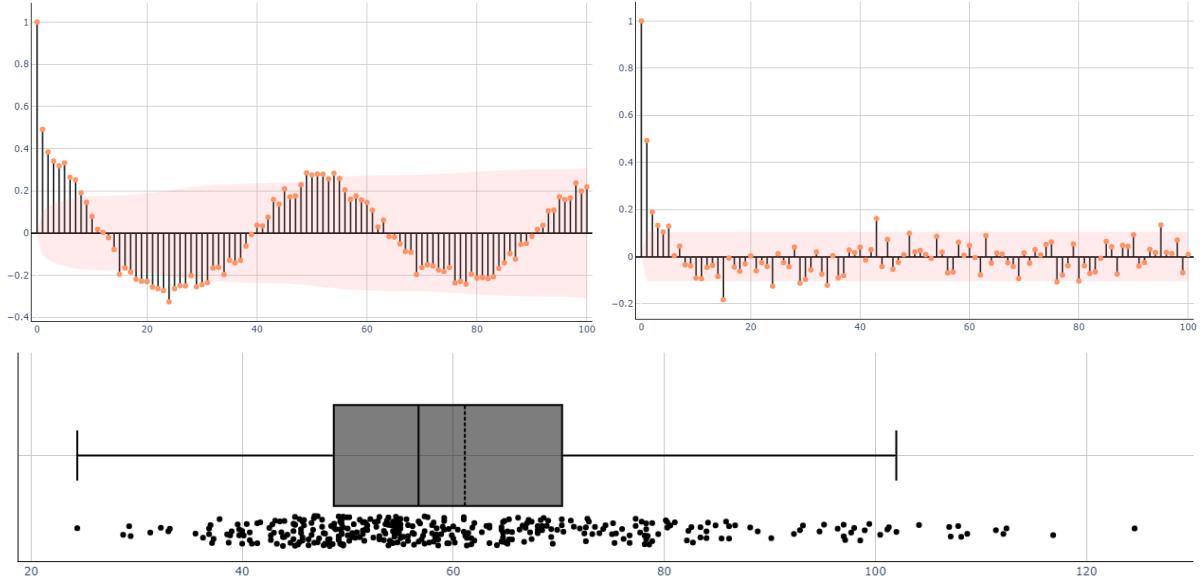
Source: The Authors

### Seasonal Decomposition - PM25



Autocorrelation Function (ACF) - PM25

Partial Autocorrelation Function (PACF) - PM25

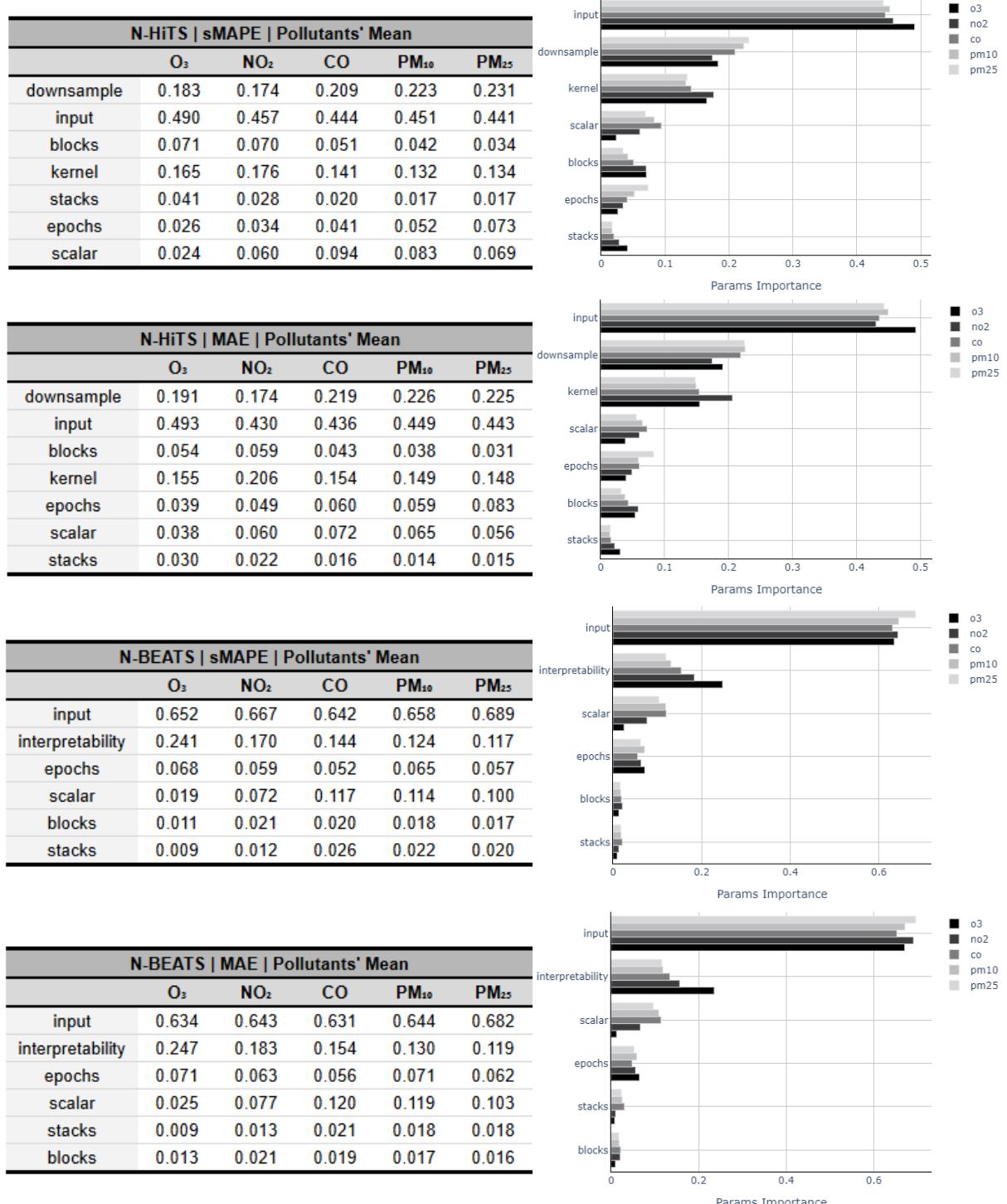


**Figure D.5** - PM<sub>2.5</sub> time series properties. (a) seasonal decomposition, (b) autocorrelation function, (c) partial autocorrelation function, (d) boxplot dispersion.

Source: The Authors

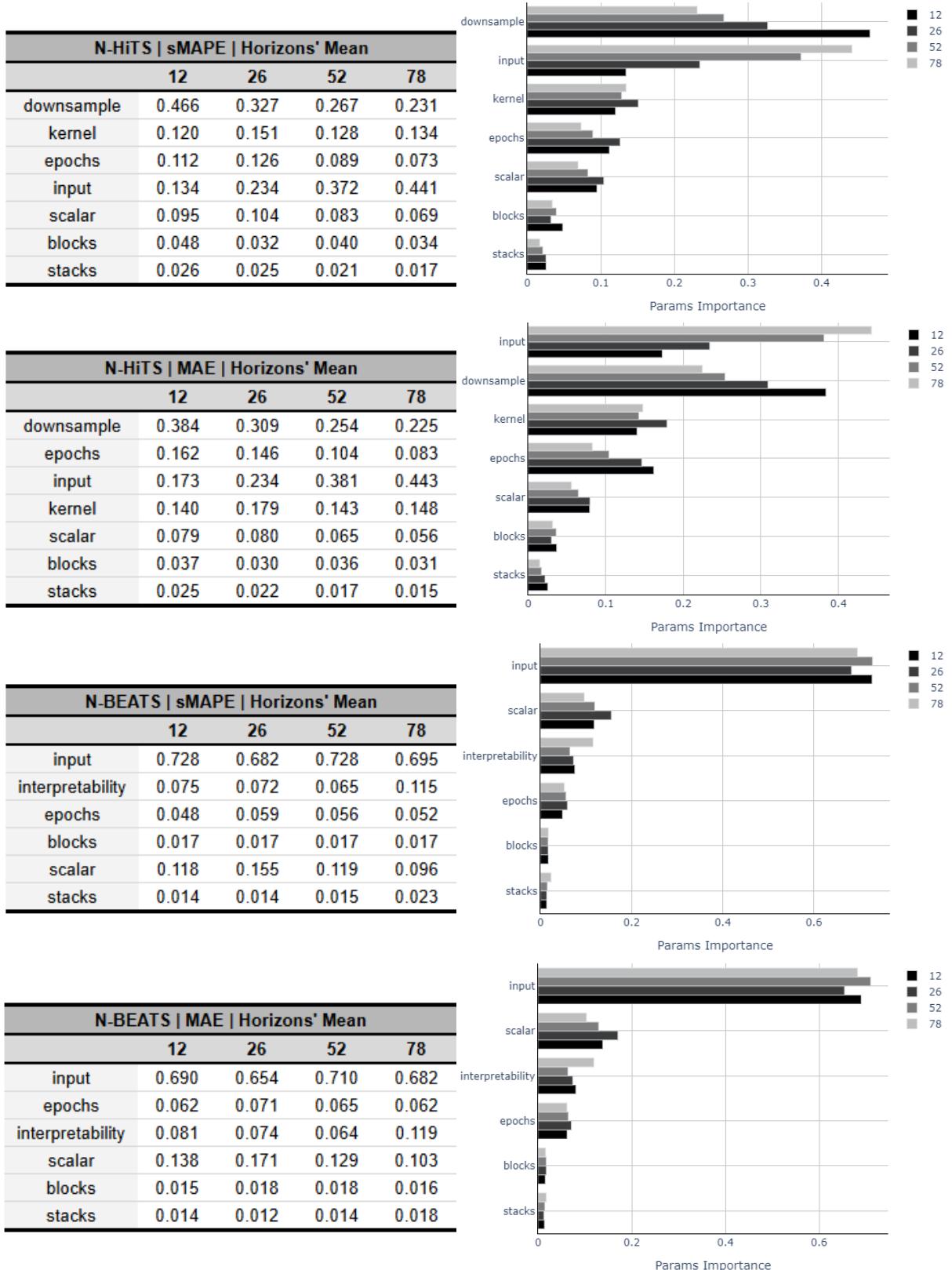
## APPENDIX E – HYPERPARAMETERS IMPORTANCE

This appendix presents the results regarding the importance of hyperparameters based on the fANOVA test, considering the average results across pollutants and forecasting horizons, along with their respective graphical representations.



**Figure E.1** - Parameters importance for neural forecasters based on pollutants.

Source: The Authors



**Figure E.2** - Parameters importance for neural forecasters based on horizons.

Source: The Authors