

Automated prediction of air pollution conditions in environment monitoring systems

Dawid Białka dawid.bialka@gmail.com,
Małgorzata Zajęcka¹[0009-0009-7914-2653] mzajECKa@agh.edu.pl,
Ada Brzoza-Zajęcka¹[0000-0002-8340-5454] abrzoza@agh.edu.pl, and
Tomasz Pelech-Pilichowski¹[0000-0003-2212-7806] tomek@agh.edu.pl

AGH University of Science and Technology, Poland**

Abstract. This paper aims to explore the problem of air pollution forecasting, especially the particulate matter (PM) concentration in the air. Other quantities such as air temperature, atmospheric pressure, and relative humidity are also considered. Moreover, a large part of the discussion in this paper can be extended and applied to a variety of other quantities which are stored and expressed as data series. The goal is to evaluate different time series forecasting models on a selected air pollution data set. The proposed model is compared with other implemented state-of-the-art methods in order to validate whether it could be a reliable pick for air pollution forecasting problem.

Keywords: Air Pollution · Sensor Networks · Severe environmental conditions · Forecasting Models · Time series · Internet of Things · Energy crisis

1 Introduction

Forecasting severe environmental conditions is very useful to protect the health and well-being of citizens. In general, forecasting atmospheric phenomena is rather complicated and requires very complex numerical models. On the other hand environmental pollution is related to human activity, which in some situations may be easier to forecast with the help of modern algorithms, techniques of data analysis, and artificial intelligence methods. In the first section we present the motivation and goal of our research and the related work. The rest of the paper is organized as follows: section 2 presents the proposed solution and in section 3 we discuss the obtained results. Section 4 concludes the paper.

1.1 Motivation and goal

The increased air pollution leads to many diseases and premature deaths [32, 20]. The smallest particles, like PM2.5, are the most threatening. The situation

** Acknowledgements: The research presented in this paper was partially financed by the funds of Polish Ministry of Education and Science assigned to AGH University.

is even worse in developing countries, which do not pay much attention to that issue [3].

The environment monitoring stations typically provide only real-time air quality information, which cannot be used to notify people in advance. Therefore, the best way to prevent people from entering areas with high concentration of air pollutants is to forecast the trend of air pollution and alert citizens before the hazardous situation occurs. However, it is challenging to forecast air pollution because it is affected by complex factors, such as air pollution accumulation, meteorology, traffic flow and industrial emissions. It is difficult to obtain sufficient data to model each factor. Therefore, there is a need to develop better models, which yields more accurate results.

Main objective of the research is to explore the problem of air pollution forecasting, especially the particulate matter (PM) concentration in the air. Other quantities such as air temperature, atmospheric pressure, and relative humidity are also considered. The goal is to evaluate different time series forecasting models on a selected air pollution data set. The proposed deep Echo State Network (ESN) model is compared with other implemented state-of-the-art methods in order to validate whether it could be a reliable pick for air pollution forecasting problem. Moreover, a large part of the discussion in this work can be extended and applied even further, to a variety of other quantities which are stored and expressed as data series.

1.2 Literature review

In case of air pollution data, there are basically three types of time series forecast models that are being used [4]: numerical models, statistical models, and machine learning models.

Numerical models Numerical models use a range of equations and mathematical functions to describe and simulate the physical processes that contribute to air pollution. The Atmospheric Dispersion Modelling System (ADMS) [17, 26] employs a three-dimensional Gaussian model to estimate the dispersion of pollutants. The California Puff Model (CALPUFF) [23] is a three-dimensional Lagrangian puff dispersion and transport model which allows to simulate the discrete and transform processes of matter emitted from a source. It can be combined with other models to produce better results [1, 27]. The Comprehensive Air Quality Model (CMAQ) [4] is a comprehensive air quality modeling system that consists of multiple processors and chemical-transport models.

Statistical models Statistical analysis models are a common way of performing time series forecasting [21]. The Autoregressive Integrated Moving Average (ARIMA) [25, 16, 2] and The Seasonal Autoregressive Integrated Moving Average [18, 22] (SARIMA) are models that are widely used for time series forecasting. The VARMA [33, 9, 10] builds on the principles of VAR [6, 8] models by incorporating moving average terms. Forecasting air pollution by applying VARMA

model yielded better results than ARIMA model [10]. Prophet [28, 24, 31] is an open-source time series forecasting model. It is designed to handle the complexities of time series data such as seasonality, trends and holidays.

Machine learning models Machine learning models are more difficult to create and are more complex but they are able to provide a very accurate and repeatable predictions [15]. Multilayer Perceptrons (MLP) [7] were used to construct a high accuracy model for predicting air pollution in London. MLP and LSTM (Long Short-Term Memory) were combined for hourly prediction of PM10 in Lima [5]. Deep spatial-temporal ensemble model coupling many stacked LSTM models were used by Wang et. al [12] for air quality prediction. Alléon, Antoine, et al. [13] created a large-scale air quality forecasting model using the Convolutional LSTM (ConvLSTM) network called PlumeNet. Xu, Xinghan et al. [30] utilized the Echo State Network (ESN) and improved particle swarm optimization (IPSO) to forecast PM2.5 concentrations in Beijing. However, their analysis did not incorporate meteorological data or account for spatial dependencies in their models.

2 The proposed solution

In this chapter improvements to the typical approach of time series forecasting are shown along with proposed models architectures. Based on their popularity in related work the selected models are: state-of-the-art models Long-Short Term Memory (LSTM), Multilayer Perceptron (MLP), and the Echo State Network (ESN) model, which this paper aims to evaluate.

A typical approach to time series forecasting is to make models that are trained on historical data obtained from a single source to forecast future values. In the case of air pollution forecasting, that approach can be improved in two ways. First, we propose that air pollution time series forecasting can be enhanced by incorporating weather forecasts as an additional factor in the predictive models. Second, we intend to consider that the air pollution is not limited to a single location but the pollution from one area can impact pollution levels in other areas. For instance, a strong wind can carry emissions from a different station, thereby influencing the air pollution levels at the first station. Therefore, it is also necessary to account for the spatial dependencies and interactions between different locations when forecasting air pollution values.

2.1 Proposed architecture of the models

Every model is tuned for each of four scenarios:

1. Base - model is trained only on historical air pollution.
2. Meteorological data - model is trained on historical air pollution and meteorological data.

3. Weather forecast improvement - model is trained using historical air pollution and meteorological data, as well as incorporating additional information from weather forecasts.
4. Spatial dependencies improvement - model undergoes training using a comprehensive data set that includes historical air pollution data, meteorological information, weather forecasts and additionally data from other stations.

Figure 1 illustrates the common workflow for the entire forecasting process, which was used for all of the proposed models to forecast PM2.5 values. The workflow consists of following steps: *Collect* - air pollution and meteorological data along with weather forecast data is collected, based on which the models will be trained; *Data fusion* - air pollution and meteorological data for one station, for which the forecasts will be made is fused with air pollution data from other stations and meteorological data forecast to provide the model with more context; *Data cleaning, transformation and normalization* - the process of preparing the data to be fed into the models. This include operations like removing outliers, interpolating missing data, normalizing the data and changing the data format to one that the model will accept and understand as an input; *Model training, evaluation and tuning* - feeding the training and test data, parameters tuning; *Forecasting based on input data* - validating the model by providing the input data and comparing the obtained forecasts with ground truth values.

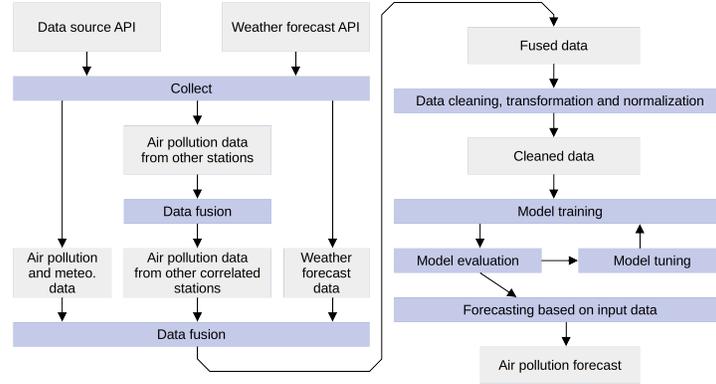


Fig. 1: High level workflow of the whole forecasting process.

ESN architecture which comprises an input node, reservoir node and readout node. In the first two scenarios, we propose a straightforward architecture, which includes a single reservoir node and a readout node. The reservoir node is connected to both the input and readout node. In the following two more complex scenarios, we introduced a deep ESN architecture shown in Figure 2. It incorporates three interconnected reservoirs, each connected to the input and readout node.

The *LSTM architecture* which is common for every scenario and consists of five consecutive layers: LSTM layer, dropout layer, LSTM layer, dense layer and reshape layer. Parameters and hyperparameters differ for the scenarios and are explained in the section 2.2 in more depth.

The *MLP architecture* which is also common for every scenario and comprises of five consecutive layers: dense layer, dropout layer, two dense layers and reshape layer. Parameters and hyperparameters vary across the scenarios.

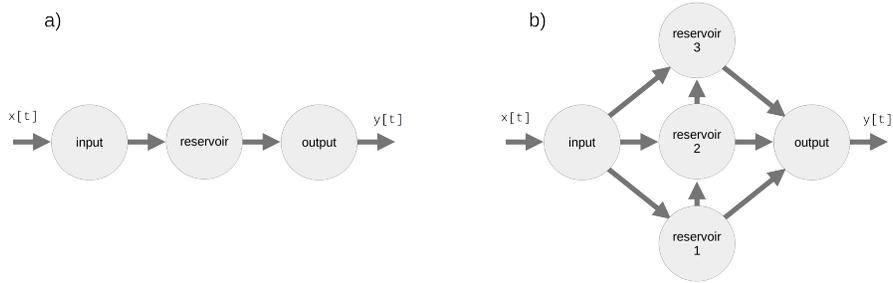


Fig. 2: The proposed deep ESN architecture with one reservoir node (a) and three reservoir nodes (b) and with one readout node.

2.2 A sample implementation

This section outlines the practical implementation of the aforementioned methodology visualised in Figure 1 which authors used on specific data. This resulted in the development of models capable of forecasting future air pollution values.

Data collection The AIRLY¹ service was chosen as a source of air pollution data. It is an organization that monitors and gathers air pollution and meteorological data across Europe, including Poland and Krakow city, North America, and South-East Asia. The data collection process utilizes a network of sensors distributed in various locations. These measurements are subsequently transmitted to a central server and stored in a database.

A Python script was developed to collect the data by making requests for the previous 24 hours' worth of data from 194 stations in and around Krakow. The data used in this study spans from 11th March 2022 to 16th March 2023, providing a substantial timeframe for analysis. The whole data set contains 1635934 single measurements from all stations.

Data description Each measurement in the data set comprises the hourly average values of the following variables: date time of the start and end of the measurement [yyyy - mm - dd hh : mm : ss], PM1, PM10, PM2.5 [$\mu\text{g}/\text{m}^3$],

¹ <https://airly.org/>

temperature [$^{\circ}C$], pressure [hPa], humidity [%], wind speed [m/s] and wind bearing [$azimuth$]. There is also information about station id, longitude and latitude.

The further analysis primarily focuses on a specific station with ID 1026, located close to the city center of Kraków, Poland. The measurements from that station have no outliers and have a low number of missing values. Total of 8760 samples are available for that station and a statistical description of the measurements from that station are shown in Table 1.

Variable	Unit	Min	Max	Mean	Standard deviation
PM1	$\mu g/m^3$	0.03	66.27	12.32	8.92
PM10	$\mu g/m^3$	0.31	129.08	24.48	19.01
PM2.5	$\mu g/m^3$	0.15	101.96	18.12	14.02
Temperature	$^{\circ}C$	-16.35	34.58	10.55	8.49
Pressure	hPa	988.35	1041.88	1015.75	8.66
Humidity	%	28.93	100	72.12	11.83
Wind speed	m/s	0.26	40.71	10.17	6.25
Wind bearing	$azimuth$	0	359.98	193.62	94.43

Table 1: Statistical description of data obtained from station with ID 1026.

Data cleansing The FBEWMA (forwards-backwards exponential weighted moving average) was employed for outlier detection and removal [11]. The method was appropriately adjusted and consists of the following steps:

- Removing measurements that exceeded the maximum or minimum value determined based on domain knowledge.
- Calculating FBEWMA for the whole time interval and determining the threshold value used to remove the measurements.
- Calculating the distance between FBEWMA and measurement value and removing measurements that exceed the determined threshold.

Out of the total 8760 samples, 25 of them were found to be invalid due to null values, and there were 115 missing measurements (gaps) in the data set [19]. To address these gaps and create a continuous set of measurements, rows with all values set to null apart from date and time were inserted into the data set for all missing points.

Data exploration No apparent trend was evident for any of the features. The data comes from human activities that disrupt natural processes and therefore we do not check nor assume a normal distribution of samples. However, daily seasonality was visible in all features except for pressure. Additionally, both PM and temperature values exhibit yearly seasonality. One interesting observation can be spotted in Figure 3. As the temperature and wind speed value rise, the PM values evidently decrease. Those two features are strong candidates for being included in the training set.

Seasonality Analysis Fast Fourier Transform (FFT) was utilized to extract important frequencies from the PM2.5 time series. The peak frequencies were

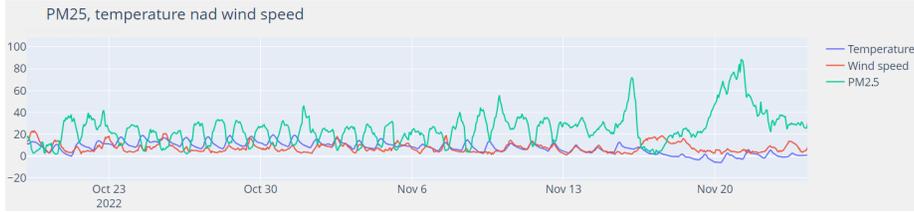


Fig. 3: PM2.5, temperature and wind speed on the same chart for data obtained from station with ID 1026.

detected near one year and one day, which corresponds to the information obtained from visual examination during data exploration. Subsequently, seasonal decomposition was conducted using the the day frequency.

Stationary assessment Two tests were performed to assess stationarity: the ADF test for difference stationarity and the KPSS test for trend stationarity. In order for the time series to be both difference and trend stationary, the null hypothesis should be rejected in the ADF test and not rejected in the KPSS test. The null hypothesis is rejected at a significance level of less than 5% when the test statistic value is smaller than the critical value. Table 2 shows that the null hypothesis is rejected in the ADF test for all variables, while the null hypothesis is not rejected in the KPSS test for all variables except for pressure.

Variable	ADF		KPSS	
	Test statistic	5% critical value	Test statistic	5% critical value
PM1	-7.93	-2.86	2.21	0.463
PM10	-7.86	-2.86	1.95	0.463
PM2.5	-7.71	-2.86	2.48	0.463
Temperature	-3.16	-2.86	7.16	0.463
Pressure	-7.25	-2.86	0.08	0.463
Humidity	-9.81	-2.86	0.54	0.463
Wx	-12.19	-2.86	1.7	0.463
Wy	-8.9	-2.86	0.74	0.463

Table 2: Results of ADF and KPSS tests.

Correlation analysis The autocorrelation analysis of the PM2.5 feature revealed that current points are significantly influenced by the previous 12 points, with a correlation strength exceeding 50% . Based on this information, the model is initialized to consider the last 12 previous points as a starting value for model learning. Furthermore, the autocorrelation analysis unveils a daily seasonality in the data, as evidenced by peak correlation values at the 24th, 48th, 72nd and 96th points, which furthermore confirms that assumption obtained from previous methods.

The Pearson correlation between PM features and rest of the features for data obtained from station with ID 1026 is shown in Figure 4. The correlation was the strongest for temperature and for wind speed, indicating a strong relationship. The correlation is relatively lower for humidity and pressure, and the wind bearing shows the lowest correlation. The correlation value is almost 1 between all PM features, thus indicating a very significant dependency.

The spatial correlation between PM2.5 for data obtained from station with ID 1026 and PM2.5 for data obtained from other stations decreases linearly as the distance increases.

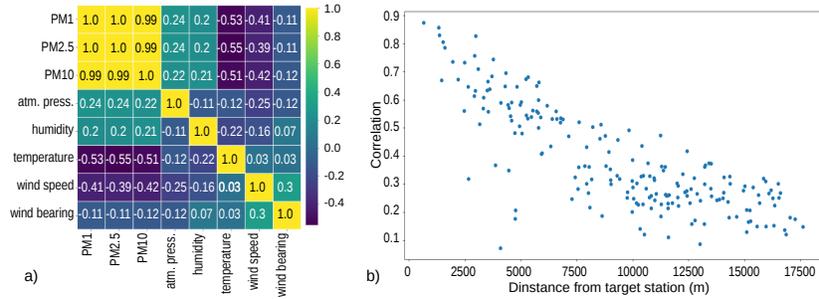


Fig. 4: a) Pearson correlation matrix of features obtained from station ID 1026. b) Pearson correlation of PM2.5 data between station ID 1026 and other stations.

Data normalization For both the LSTM and MLP models, the data was normalized by min-max scaling to a range between 0 and 1. The data for ESN was scaled using z-score scaling to improve performance.

Feature engineering and selection The following new features were introduced: x and y wind speed vector components, day and year signal, and weekend flag. Based on the analysis, we chose the features included in the training set for every testing scenarios.

- First scenario - PM2.5, day signal, year signal, weekend flag
- Second scenario - PM2.5, day signal, year signal, weekend flag, temperature, pressure, wind speed x vector component and wind speed y vector component
- Third scenario - PM2.5, day signal, year signal, weekend flag, temperature, pressure, wind speed x vector component and wind speed y vector component. Forecast values of temperature, pressure, wind speed x vector component and wind speed y vector component
- Fourth scenario - PM2.5, day signal, year signal, weekend flag, temperature, pressure, wind speed x vector component, wind speed y vector component, PM2.5 of correlated stations. Forecast values of temperature, pressure, wind speed x vector component and wind speed y vector component. Only the stations with the correlation coefficient greater than 0.6 and with low amount of invalid data were chosen.

Forecasting models creation and optimization The forecast is made 6 hours ahead for PM2.5 value for the target station with ID 1026 utilizing multi-step forecasting approach. The data was split into training, validation and test set in standard 80% : 10% : 10% ratio. The hyperparameters were tuned by utilizing random search and manual tuning. Models are presented with the best achieved configuration. We implemented the models in Python, utilizing the TensorFlow and ReservoirPy libraries.

Forecasting models testing and final evaluation In the final step of the process, the models were tested using the data from the test set. This testing phase was conducted for each model and each scenario. The performance of the models was evaluated and compared using two commonly used metrics: Mean Squared Error (MSE) [29] and R-squared score [14].

3 Results

This section is an overview of the final performance results for all models across all testing scenarios.

3.1 First scenario: base scenario

In this scenario, all models are trained on 12 previous hours of PM2.5 value, day signal, year signal and weekend flag to forecast future 6 hours of PM2.5 value. For every hour of forecast, the MSE values, R-squared scores, and the percentages of forecasted value points for which the distance from the ground truth is less than 10% of maximum value of the test set are shown in Table 3.

Model	1h	2h	3h	4h	5h	6h
	MSE					
ESN	11.9	34.35	56.34	76.99	98.27	119.13
LSTM	14.18	37.36	62.26	87.9	113.37	136.3
MLP	14.71	34.37	56.05	78.32	100.02	119.02
	R-squared					
ESN	0.96	0.88	0.82	0.75	0.67	0.59
LSTM	0.95	0.86	0.77	0.66	0.55	0.45
MLP	0.94	0.87	0.78	0.68	0.59	0.52
	Percentage of points					
ESN	97.58	89.4	82.03	75.35	70.62	66.94
LSTM	96.27	88.56	81.09	72.35	66.51	58.46
MLP	95.92	89.15	81.56	75.14	71.06	68.84

Table 3: MSE values highlighted in red are the smallest for given hour. R-squared values and percentage values highlighted in red are the greatest for given hour.

In the base scenario the ESN model demonstrates superior performance compared to other models. It achieves the highest R-squared scores and percentage of points close to ground truth for all forecast hours. It also achieves the lowest MSE for all forecast hours, except for the third hour.

3.2 Second scenario: meteorological data scenario

The models in this scenario forecast 6 future hours of PM2.5 value based on 12 past hours of PM2.5 value, day signal, year signal, weekend flag, temperature, pressure, wind speed x vector component and wind speed y vector component.

Model	1h	2h	3h	4h	5h	6h
	MSE					
ESN	15.42	31.64	46.48	62.33	73.99	83.28
LSTM	20.07	34.37	46.24	59.77	72.06	83.18
MLP	17.53	32.76	49.26	66.24	79.34	92.32
	R-squared					
ESN	0.95	0.89	0.85	0.78	0.74	0.70
LSTM	0.93	0.86	0.82	0.77	0.72	0.69
MLP	0.94	0.87	0.8	0.73	0.68	0.63
	Percentage of points					
ESN	95.50	89.98	84.36	76.72	73.37	70.63
LSTM	95.57	90.09	86.35	80.23	77.65	73.98
MLP	95.79	90.08	83.43	78.92	73.91	71.81

Table 4: MSE values highlighted in red are the smallest for given hour. R-squared values and percentage values highlighted in red are the greatest for given hour.

The MSE values, R-squared scores, and the percentages of forecasted value points for which the distance from the ground truth is less than 10% of maximum value of the test set are shown in Table 4.

For the second scenario the LSTM model yields the best results. It has the highest percentage of forecast points close to the ground truth, except for the first hour. Additionally, the MSE is the lowest for the last four forecast hours when using the LSTM model. Only the R-squared score is greater for all forecast hours for the ESN.

3.3 Third scenario: weather forecast improvement scenario

In the third scenario, the models are trained to forecast the future values of PM2.5 for the next 6 hours based on the previous training set extended with weather forecast data. Features used for training include the 12 previous hours of PM2.5 values, day signal, year signal, weekend flag value, temperature, pressure, wind speed x vector component and wind speed y vector component as well as the 6 future points of temperature, pressure, wind speed x vector component, and wind speed y vector component.

The MSE values, R-squared scores, and the percentages of forecasted value points sufficiently close to the the ground truth are shown in Table 5.

For the third scenario the LSTM model again yields the best results. It has the highest percentage of forecast points for which the distance between the ground truth and forecast value is less than 10% of the maximum for all forecast hours. The MSE is the lowest for the last three forecast hours when using the LSTM model. Only the R-squared score is greater for almost all forecast hours for the ESN model.

3.4 Fourth scenario: spatial dependencies improvement scenario

In the last scenario PM2.5 data from other correlated stations was added to the training set. Thus the features used for training include the 12 previous hours of PM2.5 values, PM2.5 values obtained from 11 correlated stations, day

Model	1h	2h	3h	4h	5h	6h
	MSE					
ESN	16.62	34.23	50.05	60.53	72.79	82.01
LSTM	22.31	34.94	47.69	56.42	65.9	74.23
MLP	16.74	32.91	50.51	63.38	78.89	90.02
	R-squared					
ESN	0.95	0.89	0.84	0.79	0.75	0.71
LSTM	0.92	0.84	0.80	0.78	0.74	0.72
MLP	0.91	0.83	0.76	0.70	0.66	0.60
	Percentage of points					
ESN	95.73	88.74	81.56	79.72	75.34	72.93
LSTM	95.80	90.43	85.41	80.51	77.82	74.09
MLP	93.23	87.04	81.68	79.65	75.73	73.75

Table 5: MSE values of forecasts for the third scenario for every hour. MSE values highlighted in red are the smallest for given hour.

signal, year signal, weekend flag, temperature, pressure, wind speed x vector component and wind speed y vector component, as well as the 6 future points of temperature, pressure, wind speed x vector component, and wind speed y vector component.

The MSE values, R-squared scores, and the percentages of forecasted value points sufficiently close to the the ground truth are shown in Table 6.

Model	1h	2h	3h	4h	5h	6h
	MSE					
ESN	19.04	36.38	53.46	65.22	77.34	90.22
LSTM	19.96	37.38	49.64	61.29	73.07	78.90
MLP	20.83	41.71	54.95	68.65	83.81	93.60
	R-squared					
ESN	0.95	0.87	0.79	0.72	0.69	0.66
LSTM	0.92	0.86	0.81	0.76	0.70	0.66
MLP	0.92	0.81	0.72	0.66	0.62	0.58
	Percentage of points					
ESN	95.65	87.56	76.95	72.39	67.70	64.63
LSTM	95.12	86.55	82.39	76.42	73.21	70.84
MLP	93.51	86.63	78.06	75.14	74.03	72.62

Table 6: MSE values of forecasts for the fourth scenario for every hour. MSE values highlighted in red are the smallest for given hour.

For this scenario, overall, the LSTM model again yields the best results. It achieves the highest R-squared scores for the last four forecast hours and the highest percentage of forecast points where the distance between the ground truth and forecast value is sufficiently small. Furthermore, the LSTM model exhibits the lowest mean squared error (MSE) for the last four forecast hours.

3.5 Comparison

From the analysis of the tables 3, 4, 5 and 6, several observations can be made. For the last three forecast hours the third scenario stands out with the LSTM

model achieving the lowest MSE and the highest percentage of forecast points within the desired range. On the other hand, the ESN model performs the best in terms of R-squared scores for the same hours.

However, for the first three hours, the metrics are better for the first three scenarios. This suggests that incorporating weather forecast data into the training set improves the model’s ability to forecast long-term values (last three hours), but it may have a negative impact on forecasting the first three hours.

Another interesting finding is that for none of the forecast hours or metrics the fourth scenario performs the best. This could be due to the simplicity of the models in handling both temporal and spatial dependencies, as well as the large number of features in this scenario.

Overall, these results highlight the importance of considering the specific requirements and characteristics of the forecasting task when selecting and designing the appropriate model and training set.

3.6 Forecasting visualization

The visualizations of forecasts of PM2.5 values for the third scenario three hours in the future compared to ground truth values from validation set are presented in Figure 5 for visual overview purposes. Every point on the graph is a single forecast of PM2.5 value for three hours in the future based on 12 points of input data (each point consisting of the features of third scenario described in 2.2 Section) compared to the PM2.5 ground truth value.

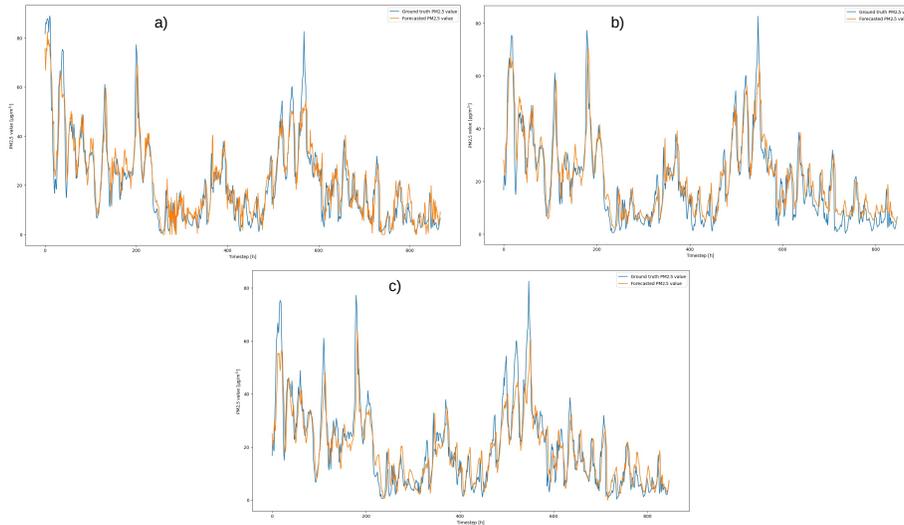


Fig. 5: Forecasts of PM2.5 values for three hours in the future obtained from a) ESN, b) LSTM, c) MLP model.

4 Summary and future work

4.1 Designed models and methods

The ESN model performed very well in terms of speed, outpacing both the LSTM and MLP models while maintaining relatively satisfactory performance. The ESN model was almost 60 times faster in fitting the data compared to the LSTM model and 5 times faster than MLP model for the third and most favorable scenario. However, when considering the forecast evaluation metrics, the LSTM model achieved the best results, while the MLP model exhibited the poorest performance.

Training the models solely on historical PM_{2.5} data yielded satisfactory results for the short-term forecasts. However, for long-term forecasts, the performance deteriorated significantly. By incorporating additional meteorological data in the second scenario the long-term forecasts improved.

Furthermore, the inclusion of weather forecast data further enhanced the long-term forecasting capabilities. However, incorporating spatial dependencies actually led to a decrease in the performance of all models. This suggests that the designed models might be too simplistic to effectively capture and utilize spatial and temporal features in the data simultaneously.

The obtained results from the models can be regarded as highly satisfactory, as clearly demonstrated by the visual comparison between the forecasted values and the ground truth. The visualizations validate the effectiveness of the models in forecasting the target variable (PM_{2.5} values) and highlight their potential for practical applications.

Overall, the ESN model presents a promising alternative to state-of-the-art methods by significantly reducing the fitting process time compared to other models, while achieving a comparable forecast performance. The results are specific to a particular geographical region, period of time, and measurement conditions. Therefore it is very difficult to perform a qualitative comparison between our results and presented by authors from different areas.

4.2 Improvements

To further enhance the modeling process, it is advisable to develop more complex models capable of effectively capturing both temporal and spatial dependencies within the data. One approach is to design a more sophisticated models with more layers and longer training time. Another approach to achieve this is by implementing ensemble learning, where multiple models are combined to leverage their individual strengths. In this context, one model can be specifically designed to capture temporal dependencies, while another model can focus on spatial dependencies.

By integrating these two models through ensemble learning techniques, such as model averaging or stacking, it becomes possible to leverage the collective intelligence of both models and potentially achieve superior performance. This approach takes advantage of the complementary nature of temporal and spatial

information, allowing for a more comprehensive understanding of the underlying patterns and dynamics in the data.

In addition to model complexity, allocating more computational resources can enable more extensive model tuning. This involves exploring a wider range of hyperparameter configurations and optimizing the model settings to achieve the best possible performance.

In future work, there are several areas to explore for enhancing the models and their applications. The presented approach can be extended and evaluated for other quantities apart from PM_{2.5}, such as NO, NO₂, SO₂ or O₃. Another aspect is to investigate alternative methods for selecting stations, such as employing Granger causality analysis. This might help identify influential stations better than Pearson Correlation. Additionally, expanding the dataset by collecting real-time weather data alongside air pollution and meteorological data, thus making the training data set larger can potentially lead to better forecasts.

References

1. Abdul-Wahab, S., Sappurd, A., Al-Damkhi, A.: Application of California Puff (CALPUFF) model: a case study for Oman. *Clean Technologies and Environmental Policy* (4 2011)
2. Abhilash, M.S.K., Thakur, A., Gupta, D., Sreevidya, B., Tiwari, S., Trivedi, M.C., Mishra, K.K.: Time series analysis of air pollution in Bengaluru using ARIMA model. Springer Singapore (3 2018)
3. Amegah, A., Jaakkola, J.: Household air pollution and the sustainable development goals. *Bull World Health Organ* (3 2016)
4. Bai, L., Wang, J., Ma, X., Lu, H.: Air Pollution Forecasts: An Overview. *International Journal of Environmental Research and Public Health* (4 2018)
5. Cordova, C.H., Portocarrero, M.N.L., Salas, R., Torres, R., Rodrigues, P.C., López-Gonzales, J.L.: Air quality assessment and pollution forecasting using artificial neural networks in metropolitan lima-peru. *Scientific Reports* **11**(1), 24232 (2021)
6. Davis, R.A., Zang, P., Zheng, T.: Sparse Vector Autoregressive Modeling. *Journal of Computational and Graphical Statistics* (2016)
7. Fionn, M.: Multilayer perceptrons for classification and regression. *Neurocomputing* (1991)
8. Gholamzadeh, F., Bourbour, S.: Air pollution forecasting for Tehran city using Vector Auto Regression. 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS) (12 2020)
9. Guo, H., Liu, X., Sun, Z.: Multivariate time series prediction using a hybridization of VARMA models and Bayesian networks. *Journal of Applied Statistics* (2016)
10. Hajar, H., Benjamin, H.: Multivariate time series modelling for urban air quality. *Urban Climate* (4 2021)
11. Hodge, V., Austin, J.: A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review* **22** (2004)
12. Junshan, W., Guojie, S.: A Deep Spatial-Temporal Ensemble Model for Air Quality Prediction. *Neurocomputing* (2018)
13. Junshan, W., Guojie, S.: PlumeNet: Large-scale air quality forecasting using a convolutional LSTM network [Online, access date: 26.03.2023]. arXiv:2006.09204, <https://arxiv.org/abs/2006.09204> (2020)

14. Lewis-Beck, M.S., Skalaban, A.: The R-Squared: Some Straight Talk. *Political Analysis* (1990)
15. Masini, R.P., Medeiros, M.C., Mendes, E.F.: Machine learning advances for time series forecasting. *Journal of economic surveys* (2023)
16. Mondal, P., Shit, L., Goswami, S.: Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications* (04 2014)
17. Nasstrom, J., Sugiyama, G., Leone, J., Ermak, D.: A real-time atmospheric dispersion modeling system. Tech. rep., Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States) (1999)
18. Peng, C., Aichen, N., Duanyang, L., Wei, J., Bin, M.: Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing. *IOP Conference Series: Materials Science and Engineering* (2018)
19. Pratama, I., Permanasari, A.E., Ardiyanto, I., Indrayani, R.: A review of missing values handling methods on time-series data. *2016 International Conference on Information Technology Systems and Innovation (ICITSI)* (2016)
20. Radim, S., Blanka, B., Miroslav, D., Michaela, M.D., Helena, L., Alena, M., Pavel, R., Andrea, R., Jana, S., Vlasta, S., Jan, T., Hana, V.: Health impact of air pollution to children. *Interntional Journal of Hygiene and Environmental Health* (8 2013)
21. Richard, W., Marcus, O.: Judgemental and statistical time series forecasting: a review of the literature. *International Journal of Forecasting* (1996)
22. Samal, K.K.R., Babu, K.S., Das, S.K., Acharaya, A.: Time series based air pollution forecasting using sarima and prophet model pp. 80–85 (2019)
23. Scire, J.S., Strimaitis, D.G., Yamartino, R.J., et al.: A user's guide for the calpuff dispersion model. *Earth Tech, Inc* **521**, 1–521 (2000)
24. Setianingrum, A.H., Anggraini, N., Ikram, M.F.D.: Prophet model performance analysis for jakarta air quality forecasting pp. 1–7 (2022)
25. Shumway, R.H., Stoffer, D.S.: Arima models. *time series analysis and its applications* (2017)
26. Snoun, H., Kanfoudi, H., Christoudias, T., Chahed, J.: One-way coupling the Weather Research and Forecasting Model with ADMS for fine-scale air pollution assessment. *Colloque scientifique national* (2 2017)
27. Tartakovsky, D., Broday, D.M., Stern, E.: Evaluation of AERMOD and CALPUFF for predicting ambient concentrations of total suspended particulate matter (TSP) emissions from a quarry in complex terrain. *Environmental Pollution* (7 2013)
28. Taylor, S.J., Letham, B.: Forecasting at scale. *The American Statistician* **72**(1), 37–45 (2018)
29. Wallach, D., Goffinet, B.: Mean squared error of prediction as a criterion for evaluating and comparing system models. *Ecological Modelling* (1989)
30. Xu, X., Ren, W.: Application of a Hybrid Model Based on Echo State Network and Improved Particle Swarm Optimization in PM_{2.5} Concentration Forecasting: A Case Study of Beijing, China. *Sustainability* (2019)
31. Ye, Z.: Air pollutants prediction in Shenzhen based on ARIMA and Prophet method. *E3S Web of Conferences* (1 2019)
32. Yin, F., Jinhua, C., Jun, S., Han, S.: Spatial Effects of Air Pollution on Public Health in China. *Environ Resource Econ* (5 2018)
33. Yunus, K., Chen, P., Thiringer, T.: Modelling spatially and temporally correlated wind speed time series over a large geographical area using VARMA. *IET Renewable Power Generation* (2017)