

Article

Hourly Particulate Matter (PM₁₀) Concentration Forecast in Germany Using Extreme Gradient Boosting

Stefan Wallek ^{1,2,*} , Marcel Langner ^{1,2} , Sebastian Schubert ³ , Raphael Franke ⁴  and Tobias Sauter ² 

¹ German Environment Agency, Wörlitzer Platz 1, 06844 Dessau-Roßlau, Germany

² Geography Department, Faculty of Mathematics and Natural Sciences, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

³ Urban Ecosystem Science, Institute of Ecology, Faculty VI—Planning Building Environment, Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany

⁴ School of Business and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

* Correspondence: stefan.wallek@uba.de

Abstract: Air pollution remains a significant issue, particularly in urban areas. This study explored the prediction of hourly point-based PM₁₀ concentrations using the XGBoost algorithm to assimilate them into a geostatistical land use regression model for spatially and temporally high-resolution prediction maps. The model configuration and training incorporated meteorological data, station metadata, and time variables based on statistical values and expert knowledge. Hourly measurements from approximately 400 stations from 2009 to 2017 were used for training. The selected model performed with a mean absolute error (MAE) of 6.88 µg m⁻³, root mean squared error (RMSE) of 9.95 µg m⁻³, and an R² of 0.65, with variations depending on the siting type and surrounding area. The model achieved a high accuracy of 98.54% and a precision of 73.96% in predicting exceedances of the current EU-limit value for the daily mean of 50 µg m⁻³. Despite identified limitations, the model can effectively predict hourly values for assimilation into a geostatistical land use regression model.

Keywords: air pollution; machine learning; XGBoost; COSMO-REA6



Citation: Wallek, S.; Langner, M.; Schubert, S.; Franke, R.; Sauter, T. Hourly Particulate Matter (PM₁₀) Concentration Forecast in Germany Using Extreme Gradient Boosting. *Atmosphere* **2024**, *15*, 525. <https://doi.org/10.3390/atmos15050525>

Academic Editors: Elena Hristova, Manousos Ioannis Manousakas, Anikó Angyal and Maria Gini

Received: 28 March 2024

Revised: 19 April 2024

Accepted: 22 April 2024

Published: 25 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Air pollution poses one of the most significant environmental risks to human health, contributing to a myriad of adverse health effects, such as stroke, heart disease, lung cancer, and respiratory diseases, including asthma. Despite efforts to mitigate its impact, air pollution remains a pressing global concern. In 2019, a staggering 99% of the world's population resided in areas where air quality fell below the guidelines set by the World Health Organization (WHO). Tragically, ambient air pollution was responsible for an estimated 4.2 million premature deaths worldwide that year, with a disproportionately high burden observed in low- and middle-income countries, particularly within the WHO South-East Asia and Western Pacific Regions [1].

In 2021 in the European Union, 253,000 deaths were attributable to exposure to particulate matter (PM_{2.5}) above WHO's guideline level of 5 µg m⁻³, 52,000 deaths were attributable to exposure to NO₂ concentrations above WHO's guideline level of 10 µg m⁻³, and 22,000 deaths were attributable to short-term exposure to O₃ concentrations above 70 µg m⁻³. Particulate matter (PM) is of significant interest because of its widespread presence in the atmosphere and its detrimental effects on human health and the environment [2].

Addressing this conjuncture requires comprehensive policies and investments aimed at reducing key sources of outdoor air pollution, such as transportation, energy production, industrial processes, and waste management. Moreover, promoting access to clean household energy sources can significantly alleviate air pollution in certain regions [1]. Legal frameworks and international agreements, including WHO Air Quality Guidelines [3], the

United Nations Economic Commission for Europe (UNECE) Convention on Long-range Transboundary Air Pollution (CLRTAP) [4], including its protocols, and the EU Ambient Air Quality Directive [5], which is to be revised at present, play a crucial role in guiding these efforts. Recognising the heightened vulnerability of certain populations, such as children, elderly persons, and those with chronic illnesses, underscores the urgency of implementing effective air quality management strategies.

Moreover, accurate forecasting of particulate matter (PM) concentrations on urban, local, and regional scales is essential for informing public health interventions and mitigating exposure risks. While existing forecasting models, such as the Copernicus Atmosphere Monitoring Service (CAMS), provide valuable insights, they often lack the necessary spatial and temporal resolution required for precise predictions, particularly in urban areas where PM concentrations tend to be highest.

Pappa and Kioutsioukis [6] assessed the accuracy of PM forecasts generated by CAMS at a local scale, comparing them against actual in situ measurements gathered over a two-year period from a network of monitoring stations situated in an urban coastal Mediterranean city in Greece. Their evaluation focused on forecasting $PM_{2.5}$ and PM_{10} concentrations over four consecutive days at intervals of 6 h at individual monitoring stations. The findings revealed that CAMS forecasts tend to underestimate $PM_{2.5}$ and PM_{10} concentrations by a factor of two during the winter season, suggesting a deficiency in capturing anthropogenic particulate emissions like those from wood-burning activities. Conversely, an overestimation of concentrations was observed during other seasons.

Bailey et al. [7] utilised the Copernicus Atmospheric Monitoring Service (CAMS) reanalysis data as a critical component in estimating $PM_{2.5}$ levels for both city and national scales, as required by Sustainable Development Goal (SDG) Indicator 11.6.2. Leveraging CAMS data, which incorporates in situ and remote sensing information at a resolution of 0.1° , their approach provided a robust framework for assessing air quality across Europe. By integrating this comprehensive dataset into their methodology, they aimed to enhance the granularity and accuracy of $PM_{2.5}$ estimations, addressing the limitations posed by sparse monitoring networks.

A huge variety of techniques and algorithms exist to predict PM concentration. Choubin et al. [8] introduced a hybrid model that combines air mass trajectory analysis and wavelet transformation to enhance the accuracy of artificial neural network (ANN) forecasts for daily average concentrations of $PM_{2.5}$ two days in advance. Developed using data from 13 air pollution monitoring stations in the Jing-Jin-Ji area, China (Beijing, Tianjin, and Hebei province), the model leverages air mass trajectories to identify distinct transport corridors for “dirty” and “clean” air. By decomposing the original time series of $PM_{2.5}$ concentrations using wavelet transformation and integrating meteorological forecast variables, the model achieves a significant reduction in root mean squared error (RMSE) of up to 40%, with a detection rate for high $PM_{2.5}$ days reaching an average of 90%.

The study from Kowalski et al. [9] noted the significant impact of air pollution in Poland. Focusing on PM_{10} concentration caused by adverse weather conditions and human activities, the study aimed to evaluate the efficacy of modern neural networks in predicting PM_{10} levels for the hours of the subsequent day. The model is based on data from the Polish sensor network Airly composed of 2458 stations. Employing machine learning algorithms, including convolutional and deep learning neural networks, the research demonstrated the effectiveness of a proposed convergent neural network model in providing detailed air quality forecasts for the next 24 h.

Czerniecki et al. [10] addressed the persistent issue of air pollution in European urban areas, particularly highlighting the impact of elevated PM levels on premature deaths, predominantly due to heart disease and stroke. With Poland being identified as one of the most polluted countries in Europe, especially during winter months, the study emphasised the need for accurate PM forecasting alongside municipal mitigation efforts. By analysing 10 years of hourly winter PM_{10} and $PM_{2.5}$ concentrations from 11 urban air quality monitoring stations across four major Polish agglomerations, the research assessed the feasibility

of short-term PM forecasting using machine learning (ML) techniques. Among the tested ML models, Extreme Gradient Boosting (XGBoost) emerged as the most effective, followed by random forests and neural networks, while stepwise regression exhibited the lowest performance. These findings underscore the significant potential of ML in short-term air quality prediction.

Furthermore, the work of Park et al. [11] demonstrated strong performance in capturing spatial contrasts and temporal variability in PM₁₀ concentrations using ML techniques and propounded that these models offer reliable PM₁₀ concentration values for pollution management, prevention, and mitigation. For future improvements, they suggested the inclusion of additional variables related to spatial and seasonal characteristics to enhance model accuracy.

Gilik et al. [12] trained models based on hybrid deep learning architecture to predict concentrations of different pollutants with publicly available data in the cities of Barcelona, Spain, Kocaeli, Turkey, and Istanbul, Turkey. They also observed an effect of meteorological conditions on the prediction. However, the study acknowledged several limitations. Firstly, there was a significant number of missing or poor-quality data in the publicly available sources for the selected cities. Consequently, the training dataset for the model was constrained by the scarcity of usable data collected from all sensors within the cities. This resulted in small samples used as input for the model, making it challenging for the model to extract meaningful relationships from the data. The authors additionally indicated that the transferability of local models of individual cities to other cities is not easily guaranteed and does not make sense if they are too far apart.

While previous studies have explored machine learning (ML) models for predicting PM concentrations, Feng et al. [13] stated that spatial hazard modelling remains limited. Their study addressed this gap by developing new ML models for predicting PM₁₀ hazard in the Barcelona province of Spain. Using data from 75 stations, healthy and unhealthy locations were identified, and ML models were calibrated and validated, achieving accuracy and precision of >87% and >86%, respectively. Spatial hazard maps generated by the models highlighted high-risk areas primarily situated in the middle of the Barcelona province rather than in the metropolitan area.

For the high-resolution spatial–temporal distribution of point information on PM₁₀ concentrations, we have developed a geostatistical model using land-use regression [14]. This study aims to investigate the potential of augmenting this model with predicted hourly PM₁₀ concentrations using the XGBoost approach to create a spatially and temporally high-resolution prediction model. This is intended to depict pollution situations in urban areas more realistically than can be achieved by regional models such as CAMS. If more precise and comprehensive information on future pollution situations is available at a high resolution, especially high-risk groups can benefit, as they can immediately and effectively reduce their exposure to air pollution through appropriate planning of their activities. In situations relevant to the general population, city administrations can issue warnings or take measures to proactively mitigate the severity of pollution, thereby reducing the disease burden on the population as a whole. This work builds upon previous findings, particularly regarding the use of meteorological predictors and the utilisation of machine learning algorithms for modelling PM concentrations, and extends them.

2. Materials and Methods

2.1. Study Area

The study area encompasses the Federal Republic of Germany, offering a diverse range of geographical and climatic conditions that influence air pollution dynamics. From the mountainous terrain of the Bavarian Alps to the coastal plains along the North Sea and Baltic Sea, Germany's geography is characterised by a mix of landscapes, including forests, agricultural areas, urban centres, and industrial zones. The climate is influenced by its location in central Europe, with distinct seasonal variations and weather patterns. Coastal areas experience maritime influences, while inland regions are subject to continental

climatic conditions. These climatic variations, coupled with topographical features, such as valleys, hills, and plateaus, create microclimates that can significantly impact air pollutant dispersion and transport. Germany's dense population centres, including major cities, such as Berlin, Munich, Hamburg, Frankfurt, and the Ruhr area, are hubs of economic activity and transportation networks. Industrial activities, vehicular emissions, and residential heating contribute to localised air pollution hotspots, particularly in urban areas with high population densities and traffic congestion. With a robust network of air quality monitoring stations, Germany offers rich and extensive datasets for conducting detailed air pollution studies in a broader context. This includes data on particulate matter concentrations as well as meteorological parameters crucial for understanding pollutant dispersion patterns. Figure 1 shows a terrain map of Germany and used stations categorised by siting type.

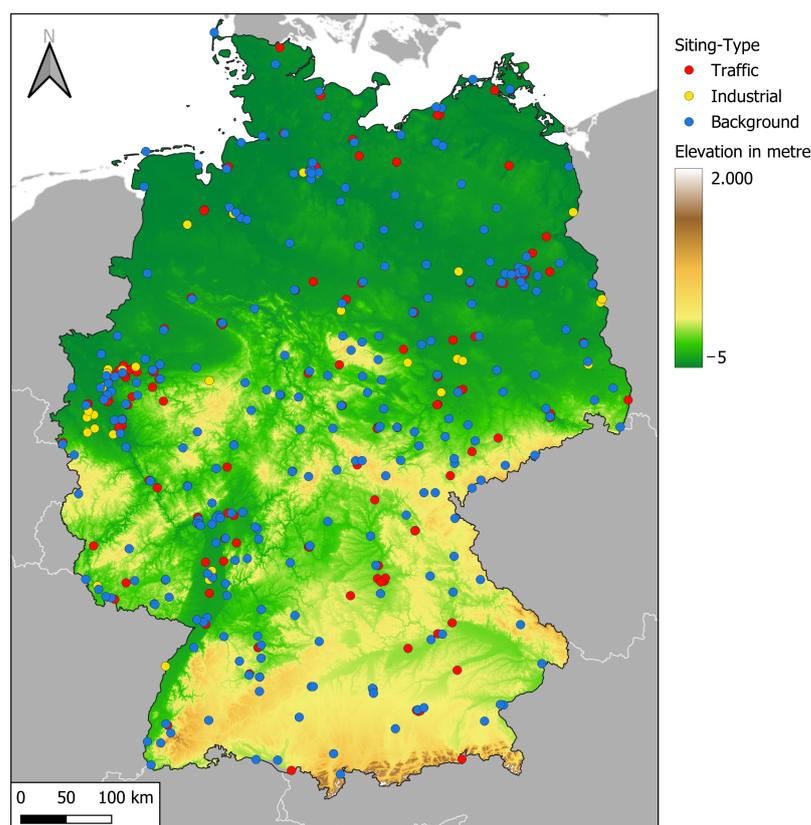


Figure 1. Terrain map of Germany and used stations categorised by siting type. Background stations may overlap with traffic or industrial stations, especially in urban areas. Data sources: State and federal air quality monitoring networks, Federal Agency for Cartography and Geodesy, OpenStreetMap Contributors.

2.2. Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) [15] is a powerful and widely-used machine learning algorithm known for its efficiency and effectiveness in predictive modelling tasks. It belongs to the family of ensemble learning methods, specifically boosting algorithms, which combine multiple weak learners (typically decision trees) to create a strong learner capable of making accurate predictions. XGBoost builds upon the traditional gradient boosting framework by introducing several enhancements to improve performance and scalability. One key innovation is the integration of a regularisation term into the objective function, which helps prevent overfitting and improves generalisation of unseen data. Additionally, XGBoost employs a highly optimised implementation, leveraging parallel processing and distributed computing to achieve faster training and inference times. The algorithm's success can be attributed to its ability to handle diverse data types, including numerical, categorical, and missing values, without the need for extensive preprocessing.

Moreover, XGBoost offers flexibility in model tuning through a wide range of hyperparameters, allowing users to fine-tune the algorithm to suit specific datasets and prediction tasks in their domain. XGBoost has been widely adopted across various domains, including finance, healthcare, marketing, and environmental science, for tasks such as regression, classification, and ranking. Its robust performance, ease of use, and interpretability make it a preferred choice for both academic research and real-world applications.

2.3. Dataset and Model Configuration

Hourly data of PM₁₀ concentrations (capped at the 99.99% percentile) from state and federal air quality monitoring stations were used as target variables. The measurements meet the requirements specified for compliance with the obligations arising from the EU Ambient Air Quality Directive [5]. Hourly measurements from 2009 to 2018 from a total of 429 stations were used in this study. However, the number and distribution of active stations are subject to change because of continuous adjustments to the networks. Table 1 shows the summary statistics for measured PM₁₀ concentrations and the number of active stations per year.

Table 1. Summary statistics for measured PM₁₀ concentrations and the number of active stations per year. Min, max, mean, and standard deviation (SD) are in $\mu\text{g m}^{-3}$.

Year	Min	Max	Mean	SD	Stations
2009	0	1062.85	22.09	17.31	347
2010	0	16,142.50	23.06	25.71	356
2011	0	1966.63	23.08	18.87	370
2012	0	3617.21	19.93	16.41	357
2013	0	1409.69	20.26	15.22	355
2014	0	3014.80	20.23	15.85	345
2015	0	3407.47	18.89	15.13	347
2016	0	2737.70	17.65	13.96	345
2017	0	1860.08	17.50	14.91	349
2018	0	1330.84	18.69	13.50	360

Additionally, meteorological variables from COSMO-REA6 [16–19], along with time variables (day of the week, day of the year), and station metadata (latitude, longitude, altitude) served as predictor variables. The COSMO-REA6 dataset is a high-resolution reanalysis system developed based on the numeric weather prediction model COSMO and provides detailed atmospheric simulations over Continental Europe at a resolution of 0.055° (6 km). Incorporating observational data assimilation through COSMO's nudging scheme and specialised modules for snow, sea surface temperature, and soil moisture analysis, it utilises ERA-Interim data for lateral boundary conditions. Covering the period from 1995 to August 2019, COSMO-REA6 serves as a valuable resource for regional weather forecasting, climate modelling, and research, facilitating the study of local climate patterns, the assessment of climate change impacts, and support for various applications, such as agriculture and disaster management. The full model output of COSMO-REA6 comprises a set of 150 variables. We focused on the 35 available 2D parameters and chose 9 of them to use as features to build the model.

The selection of variables was carried out iteratively and was based, in the first step, on the correlation with the target variable and the other variables. We selected variables with strong correlation coefficients with the target variables. Some terms could be grouped together, which themselves exhibited high multicollinearity, such as the group of radiation terms, which were divided into 6 variables depending on direction and wavelength. If variables had a high degree of multicollinearity and were part of a group such as radiation, a term from this group was selected as a representative. In the second step, the variables were selected based on expert knowledge. The selected variables served as proxies for various physical processes and interactions in emission, transmission, and ambient concentration.

In addition to variables such as planetary boundary layer height or wind speed and direction, which characterise the mixing potential and magnitude of exchange processes within the troposphere, the specific surface humidity was also included to better describe the process of particle resuspension. This process is of particular importance in urban areas [20]. The hourly PM₁₀ concentrations were aggregated to daily arithmetic averages and added as predictor variables for each 1, 2, and 3 days in the past. The dataset spanned from 2009 to 2018, with the training data covering the period from 2009 to 2017 and the test data from 2018. The year 2018 was selected as the test dataset because it was the most recent year fully covered by the COSMO-REA6 dataset and was also interesting because of its meteorology.

Various aggregated models were trained using R [21] in combination with the caret package [22]. Different sets of variables were used, including all selected variables as described above, a subset of them, and only the 5 most important variables. The hyperparameters of the XGBoost algorithm were optimised using tune grids to find suitable values for training the final model. Model selection was based on the lowest root mean squared error (RMSE) resulting from the internal 5-fold cross-validation performed during the training process with a training/test ratio of 75/25.

3. Results

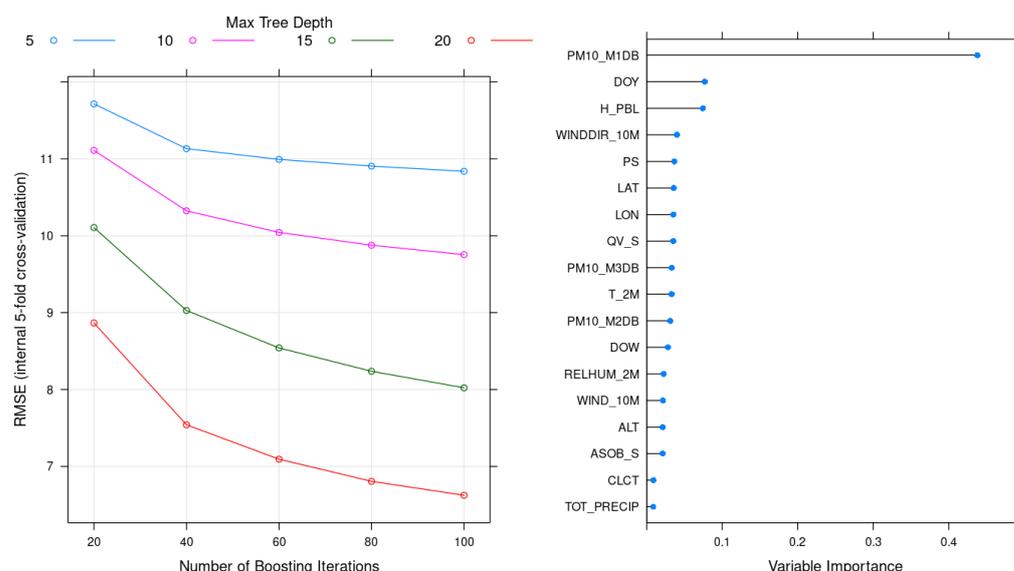
The selection criterion for the final model was the lowest RMSE value of the internal 5-fold cross-validation. The selected model was configured with the following hyperparameters: the number of iterations (nrounds) was configured to 100 boosting rounds to balance model complexity and training time; the maximum tree depth (max_depth) was set to 20 to control the depth of each decision tree; the learning rate (eta) was chosen as 0.1 to moderate the step size during optimisations for smoother convergence; the gamma value was set to zero to enforce minimum loss reduction for further node partitioning; the fraction of features to be sampled for each tree (colsample_bytree) was maintained at one; to impose minimum instance weight requirements in child nodes, the hyperparameter min_child_weight was set to one; and the subsample value was set to one, indicating the fraction of training data samples used for each boosting iteration. These parameter values were selected to strike a balance between model complexity and generalisation performance.

The internal 5-fold cross-validation resulted in an MAE of 4.32 µg m⁻³, an RMSE of 6.62 µg m⁻³, and an R² of 0.82. Figure 2 depicts the variation of RMSE as a function of maximum tree depth and the number of boosting iterations (left), as well as the importance of each predictor variable (right). RMSE values decrease with increasing tree depth and boosting iterations. Table 2 provides an overview of used predictor variables with their name, unit, and importance.

The most important independent variable by far is the mean PM₁₀ concentration from the previous day. Mean PM₁₀ concentrations from 2 or 3 days prior follow later and with less importance. The second most important variable of the model is the day of the year, which serves as a proxy for both seasonal variation and associated meteorological conditions, as well as anthropogenic activities related to these seasons. The third most important variable is the height of the planetary boundary layer, the first variable from the COSMO-REA6 dataset, serving as a crucial proxy for potential dynamic exchange processes within the troposphere. Subsequent to this are variables directly related to it, such as wind direction and air pressure. Interrupted by metadata on station latitude and longitude, the variable of the specific surface humidity, serving as a proxy for resuspension, follows. Meteorological variables, such as cloud cover and total precipitation, have the least importance. In all preliminary modelling attempts, the station type variable was explicitly included but consistently demonstrated the least importance; thus, it was not further considered in subsequent runs.

Table 2. Overview of used predictor variables with their name, unit, and importance.

Variable	Name	Unit	Importance
PM10_M1DB	Mean PM ₁₀ concentration 1 day before	$\mu\text{g m}^{-3}$	0.438
DOY	Day of year	day	0.077
H_PBL	Height of planetary boundary layer	m	0.074
WINDDIR_10M	Wind direction 10 m above ground	degree	0.040
PS	Surface pressure	Pa	0.037
LAT	Geographical latitude	degree	0.036
LON	Geographical longitude	degree	0.035
QV_S	Surface specific humidity	kg kg^{-1}	0.035
PM10_M3DB	Mean PM ₁₀ concentration 2 days before	$\mu\text{g m}^{-3}$	0.033
T_2M	Air temperature 2 m above ground	K	0.033
PM10_M2DB	Mean PM ₁₀ concentration 3 days before	$\mu\text{g m}^{-3}$	0.031
DOW	Day of week	day	0.028
RELHUM_2M	Relative humidity 2 m above ground	%	0.022
WIND_10M	Wind speed 10 m above ground	m s^{-1}	0.021
ALT	Height above mean sea level	m	0.021
ASOB_S	Avg. surface net downward shortwave radiation	W m^{-2}	0.021
CLCT	Total cloud cover	%	0.009
TOT_PRECIP	Total precipitation	kg m^{-2}	0.009

**Figure 2.** RMSE values of the internal 5-fold cross-validation depending on the hyperparameter maximum tree depth and number of boosting iterations (**left**) and the variable importance of the final model with the lowest RMSE (**right**.)

Using this model, predictions were made for the test set comprising hourly values for the year 2018. The selected model performed with an MAE of $6.88 \mu\text{g m}^{-3}$, an RMSE of $9.95 \mu\text{g m}^{-3}$, and an R^2 of 0.65. The mean bias value (predicted minus observed) of $-0.86 \mu\text{g m}^{-3}$ indicates a slight underestimation. The bias median is $-1.73 \mu\text{g m}^{-3}$ with a standard deviation of $9.35 \mu\text{g m}^{-3}$.

In addition, Figures 3 and 4 show Q–Q plots with different value ranges. Both show the line of best fit in magenta as well as the 95% and 99% percentiles of measured values as blue dotted and dashed lines, respectively. Figure 3 displays the entire range of values and illustrates that the model underestimates values above the 95% percentile, with the underestimation increasing significantly as values rise, reaching deviations of up to $100 \mu\text{g m}^{-3}$. Figure 4 focuses on the value range up to $70 \mu\text{g m}^{-3}$ and demonstrates that overestimation begins at values above approximately $25 \mu\text{g m}^{-3}$. Conversely, in the range from 0 to $25 \mu\text{g m}^{-3}$, there is a systematic overestimation of values. Around $25 \mu\text{g m}^{-3}$, the

points align closely with the line of best fit. Figure 5 provides a histogram of measured PM_{10} values corresponding to the Q–Q plot in Figure 4.

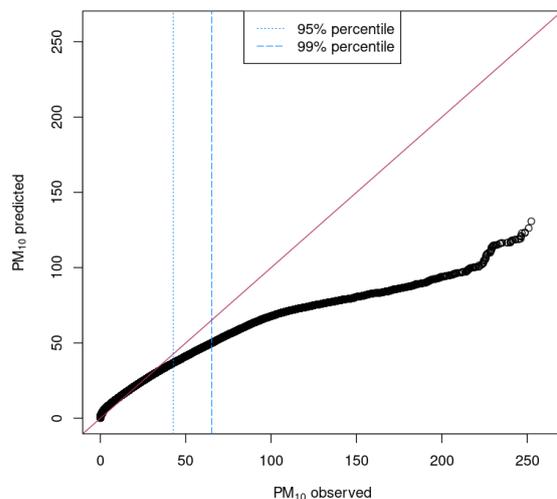


Figure 3. Q–Q plot of predicted and observed PM_{10} values in $\mu\text{g m}^{-3}$ of the test dataset (2018) containing the full value range with the line of best fit (magenta) as well as the 95% and 99% percentiles of measured values as blue dotted and dashed lines, respectively.

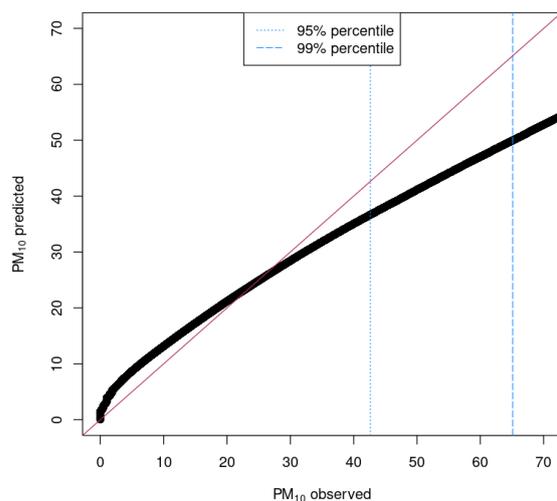


Figure 4. Q–Q plot of predicted and observed PM_{10} values in $\mu\text{g m}^{-3}$ of the test dataset (2018) containing only the 99% percentile of values with the line of best fit (magenta) as well as the 95% and 99% percentiles of measured values as blue dotted and dashed lines, respectively.

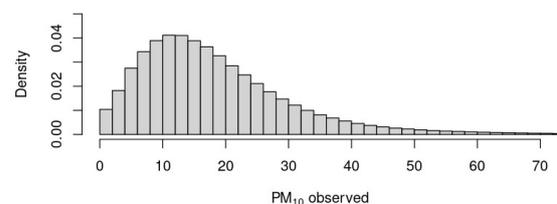


Figure 5. Histogram of observed PM_{10} values in $\mu\text{g m}^{-3}$ of the test dataset (2018) corresponding to the Q–Q plot in Figure 4.

Figure 6 displays the observed PM_{10} concentrations in 2018 and exhibits their temporal variations, as reflected in the model performance metric RMSE shown in Figure 7. The year 2018 was characterised by notably dry conditions, featuring numerous stable high-pressure

weather systems, which led to reduced heights of the planetary boundary layer shown in Figure 8. Additionally, in 2018, several instances occurred where dust originating from the Sahara was transported to Germany.

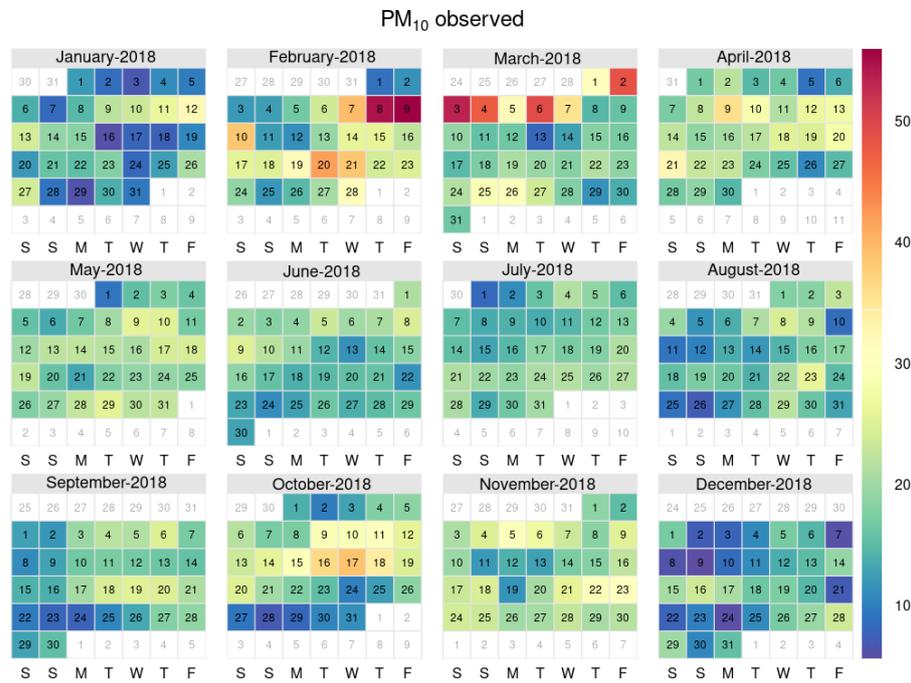


Figure 6. Calendar plot of observed PM₁₀ concentrations in $\mu\text{g m}^{-3}$ of the test dataset (2018).

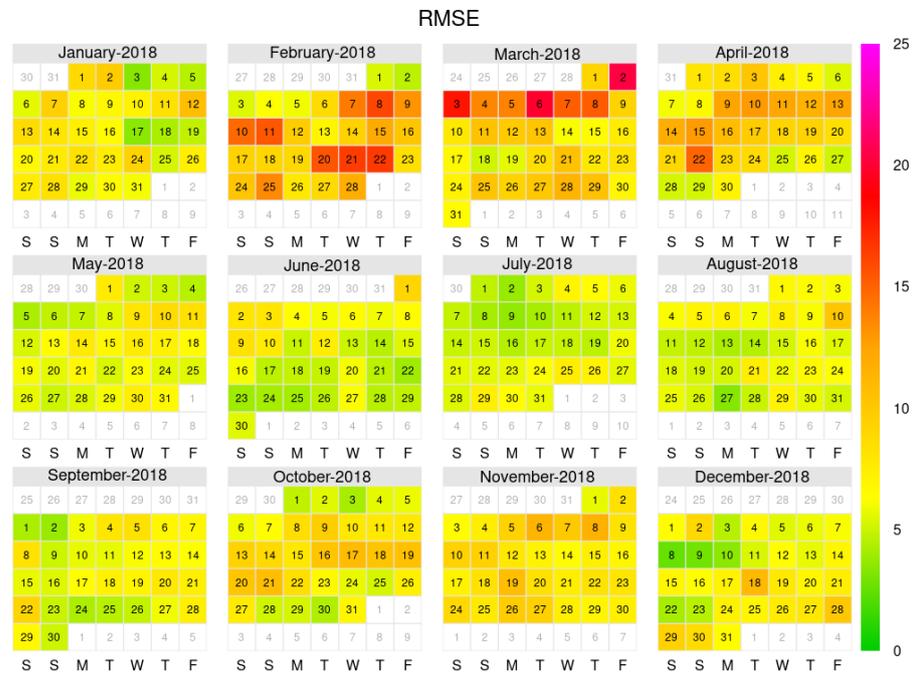


Figure 7. Calendar plot of RMSE values in $\mu\text{g m}^{-3}$ from the validation process of the test dataset (2018).

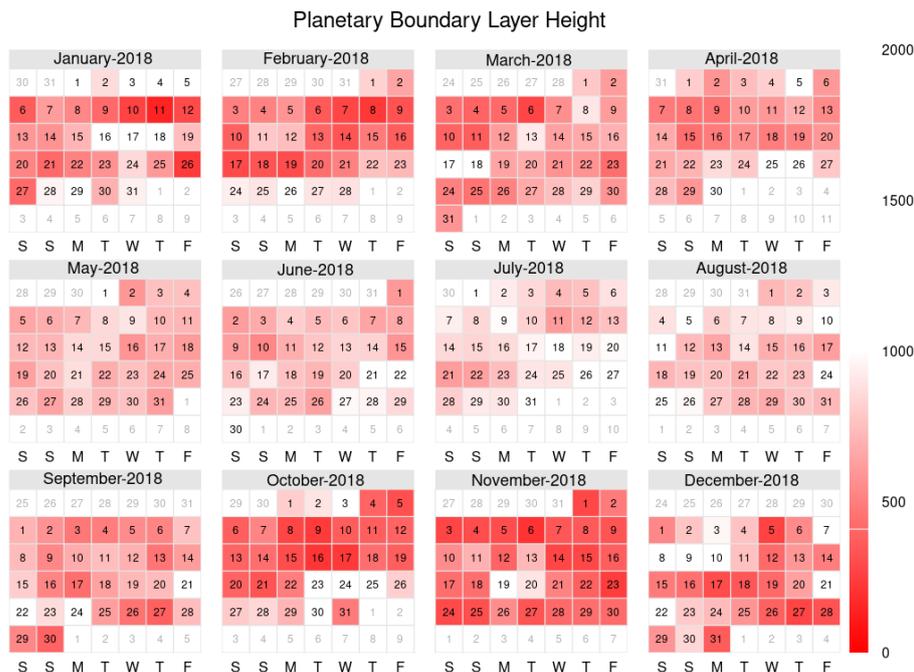


Figure 8. Calendar plot of the planetary boundary layer height in metres of the test dataset (2018).

Excluding days with a Sahara Dust Index (SDI) [23] greater than 0.4, along with the days immediately preceding and following (due to the singular observation station in Bavaria, Hohenpeißenberg, rendering it non-representative for the entire study area), results in a slight improvement of the error metrics MAE and RMSE, as shown in Table 3.

Table 3. Metrics of the model evaluation with training (2009–2017) and test (2018) datasets including and excluding days (± 1) with Sahara Dust Index (SDI) > 0.4. All metrics except R^2 are in $\mu\text{g m}^{-3}$.

SDI > 0.4	MAE	RMSE	R^2	Bias (Mean)	Bias (Median)	Bias (SD)
included	6.88	9.95	0.65	−0.86	−1.73	9.35
excluded	6.61	9.53	0.65	−1.03	−1.84	9.03

The occurrence of several multi-day events posed a challenge, with the model exhibiting underestimation at the onset and overestimation towards the conclusion of such events. These trends are also evident in the calendar plots and apply to the above-mentioned Sahara dust events, which are predicted with a time lag.

Model performance also varies in terms of the classification of the siting type and the surrounding area of the station. The respective values of the evaluation process are shown in Table 4. As anticipated, model performance is poorer at traffic stations because of the inherently greater magnitude and variability in values. The best evaluation metrics are those from rural background stations, and the worst model performance is observed at industrial stations in urban areas (based on RMSE).

For the month of March 2018, which experienced the highest pollution levels, a separate analysis was conducted for each exemplarily chosen station representing the most frequent combinations of station type and environment. The hourly values were aggregated into rolling 3 h means for better visualisation. Figures 9–12 depict the measured values (black) compared to the modelled values (red).

Table 4. Evaluation metrics by pairs of siting type and classification of the surrounding area of the station according to Annex III of the directive 2008/50/EC on ambient air quality and cleaner air for Europe [5]. MAE and RMSE are in $\mu\text{g m}^{-3}$.

Siting Type	Area	MAE	RMSE	R ²
Background	Rural	5.83	8.49	0.65
Background	Suburban	6.53	9.45	0.63
Background	Urban	6.82	9.73	0.63
Industrial	Rural	8.23	12.10	0.58
Industrial	Suburban	6.87	9.70	0.66
Industrial	Urban	8.19	11.60	0.58
Traffic	Rural	6.66	9.89	0.63
Traffic	Suburban	7.64	11.10	0.61
Traffic	Urban	7.53	10.90	0.64

Representing urban industry, the station in Warstein (DENW181) was selected. The station is situated on a paved area at the eastern edge of the town. Approximately 6 m west lies a two-lane road primarily used as access to quarries. The quarries begin about 400 m southeast of the station and extend south. A federal highway is approximately 450 m from the station. Modelled values often exceed the mostly moderate measured values, exhibiting slight peaks, although the trend of measured values remains relatively constant. Individual events with particularly high concentrations cannot be accurately modelled. The R² value between the measured and modelled values is 0.58.

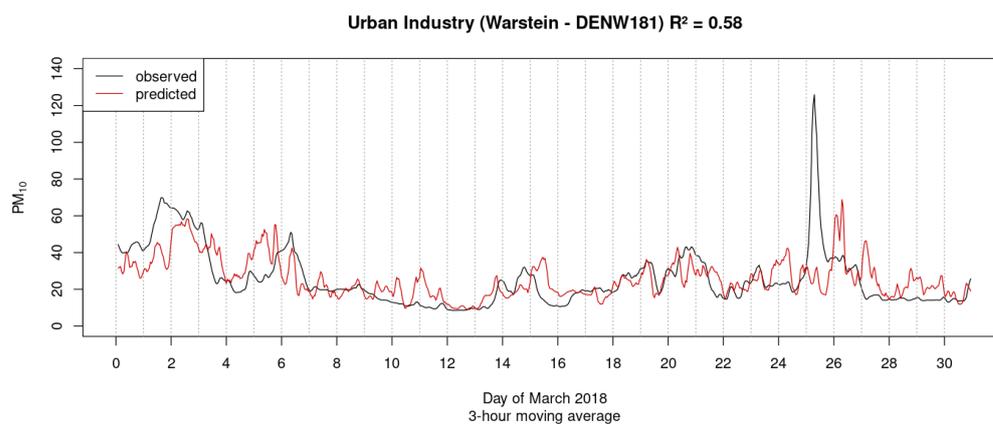


Figure 9. Comparison of observed and predicted PM₁₀ values in $\mu\text{g m}^{-3}$ during March 2018 for Warstein as an example of the siting type urban industry.

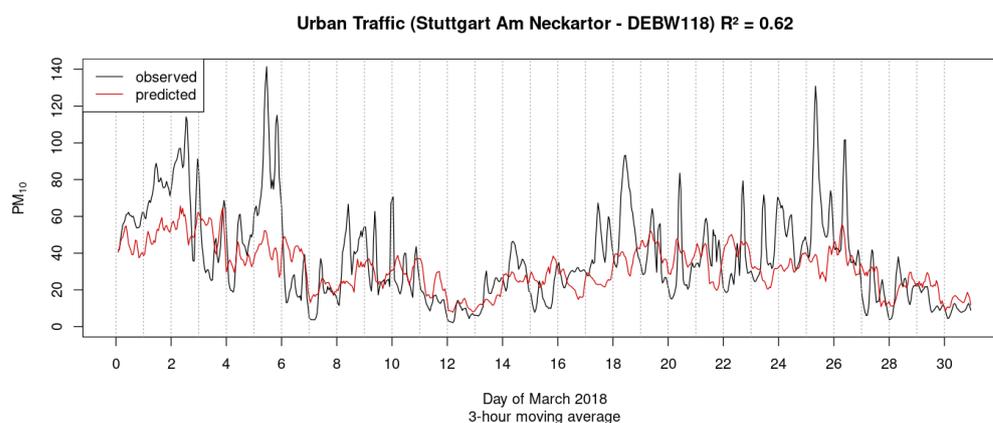


Figure 10. Comparison of observed and predicted PM₁₀ values in $\mu\text{g m}^{-3}$ during March 2018 for Stuttgart Am Neckartor as an example of the siting type urban traffic.

Similarly, for urban traffic stations, the selected station Stuttgart Am Neckartor (DEBW118) exhibits comparable behaviour. Situated in the capital city of Baden-Württemberg, the station is located on a five-lane (plus one bus lane) road with heavy traffic flow between dense urban development and a green area. Peaks occur more frequently here and are also not accurately reflected by the model. Apart from the peaks, the model follows the general trend of concentration levels throughout the month with an R^2 of 0.62.

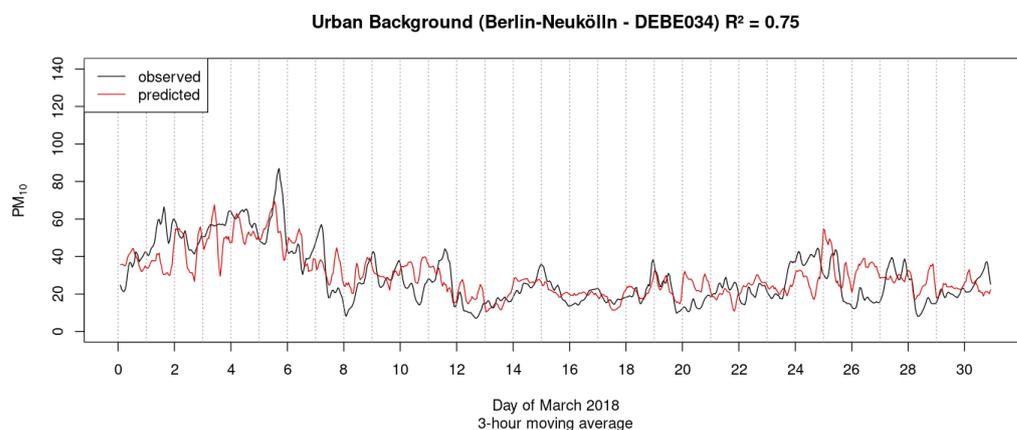


Figure 11. Comparison of observed and predicted PM_{10} values in $\mu g m^{-3}$ during March 2018 for Berlin-Neukölln as an example of the siting type urban background.

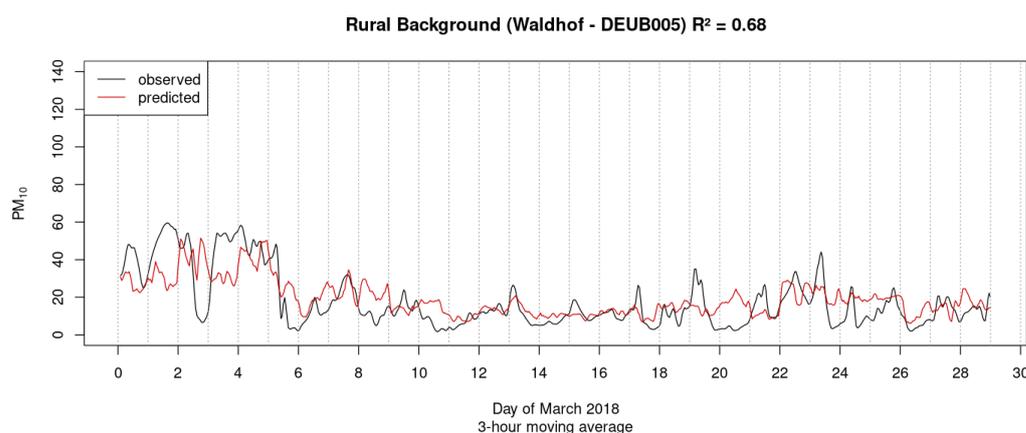


Figure 12. Comparison of observed and predicted PM_{10} values in $\mu g m^{-3}$ during March 2018 for Waldhof as an example of the siting type rural background.

This observation is also evident when considering values for the urban background station Berlin-Neukölln (DEBE034). With few exceptions, the line of modelled values closely follows the line of measured values throughout March 2018, reaching a correlation of $R^2 = 0.75$. The measurement station is located in a densely populated residential area in the city centre with moderate traffic flow near a daycare centre.

For the category of rural background, the station Waldhof (DEUB005) was selected. It is located in the eastern part of the Lüneburg Heath on Lower Saxony territory. The nearest settlement is located approximately 3 km to the west. Here, concentrations fluctuate relatively strongly at the beginning of the month, and the model values do not follow those of the measurements. As the month progresses, both values remain mostly low and close to each other. In the last third of the month, the variability of measured concentrations increases again, and the model tends to overestimate them. The correlation of measured and predicted values for this station in March 2018 reaches an R^2 of 0.68.

The exemplary examination of a selection of representatives for most station-type combinations illustrates the various typical trends and variations for each station type. In most cases, the model could capture these characteristic trends well, achieving correlation

coefficients of up to 0.75. However, the peaks that commonly occur at traffic stations could not be captured by the model, resulting in lower coefficients of determination.

The model was not specifically trained to detect exceedances of the currently applicable daily EU-limit value for PM_{10} of $50 \mu g m^{-3}$. However, for better comparison with similar models, the model's ability to fulfil this task was still investigated. For this purpose, daily mean concentrations were calculated from the measured and modelled hourly concentration values for all stations in 2018. The model achieved a high accuracy of 98.54%, indicating its overall effectiveness in making correct predictions. However, a closer examination of precision and recall metrics unveiled areas for improvement. While precision was relatively high at 73.96%, suggesting that the model's positive predictions were generally reliable, the recall value was lower at 27.81%. This indicates that the model might be missing a significant portion of instances where PM_{10} concentrations exceed the limit value.

4. Discussion

4.1. Sampling

Model training began with a 70/30 train/test split of the entire dataset spanning from 2009 to 2018. The results obtained from this initial setup closely mirrored the outcomes of the internal 5-fold cross-validation performed during the XGBoost training process. However, it is important to note that such a configuration lacks the characteristics of a true forecast model; rather, it operates more as a gap-filling model. In an effort to imbue the model with a more pronounced forecast character, subsequent training sessions utilised data from 2009 to 2017 for training, while the model's performance was evaluated against data from 2018. This approach yielded the results presented above. However, it is worth mentioning that a rolling-point-forecast model would have been even more optimal in this scenario. It is crucial to acknowledge that samples in the internal 5-fold cross-validation of the model training were not independent, primarily because of the temporal component. The interdependence between samples introduced a certain level of autocorrelation, particularly concerning time-based variables. Ignoring temporal dependencies in time-series data during XGBoost modelling and hyperparameter tuning could lead to violations of assumptions, data leakage, and misleading feature importance. This consideration is essential for understanding the limitations of the model and has to be kept in mind when it comes to validation and application.

4.2. Model Architecture

The model architecture presents opportunities for enhancement through ensemble techniques employing multiple algorithms, which have the potential to increase overall model performance. Additionally, the implementation of one aggregated model for each siting type could mitigate unwanted interference, particularly in cases where stations of different types are in close vicinity. This effect is partly attenuated when coupled with the land use regression model from our previous research [14], leveraging functions for de-trending and re-trending based on emission/land-use coefficients. Notably, in the context of station-type consideration, urban backgrounds, such as Berlin, exhibit promising performance, likely attributable to the dense network of air monitoring stations in the vicinity. Moreover, the selected meteorological variables contribute significantly to the model, warranting their pre-selection based on expert knowledge. Despite this, metadata such as station latitude and longitude also hold significance, presumed to reflect spatial patterns within the study area.

The importance of the station-type variable ranked last in all preliminary modelling attempts and was subsequently excluded. This could be justified by the fact that the mean concentration of the previous day already implicitly contained this information, as distributions of the measurements follow characteristic patterns depending on the siting type. For example, mean values and measures of dispersion for values from an urban traffic station significantly differ from those of a station in a rural background. The mean PM_{10}

concentration from the preceding day emerges as the most crucial variable, with subsequent days' concentrations showing diminished importance, possibly because autocorrelation decreases the more time elapses between the points. However, there is potential for refinement in the treatment of this variable; while the model currently considers mean PM_{10} concentration from one, two, or three days prior, optimisation can involve adjusting the timeframe to better align with predictive accuracy. In the worst case, this value is 23 h apart from the value to be predicted.

The unexpectedly low importance of the variables for cloud cover and precipitation can stem from several factors. Firstly, these variables may be inadequately modelled in the weather model and, hence, in the reanalysis dataset. Additionally, their relatively coarse spatial resolution is a potential limitation. Furthermore, their high correlation with other variables, such as humidity and surface moisture, which are also used and considered more important, can contribute to their diminished importance.

The variable for the hour of the day is also not included in the final model, as it proved to be insignificant in previous iterations. While this may initially seem surprising, it is supported by the weaker diurnal variation compared to the annual cycle observed in PM_{10} levels. This can also be interpreted as a result of implemented emission reduction measures. Thus, exhaust emissions in Germany have been declining since the mid-1990s and have even fallen below road traffic emissions from abrasion (tyres, brake pads, road surface) since 2015. Unlike gaseous air pollutants, such as NO_2 , local sources of PM, especially the coarse fraction PM_{10} , contribute less to the overall concentration. Statistically, the low importance of the variable can also be attributed to some meteorological parameters included in the model, such as air temperature, which typically exhibit a relatively strong diurnal pattern.

4.3. Limitations and Improvement Suggestions

The model faces limitations in predicting very high values, primarily because of the mathematical configuration of the algorithm and statistical constraints. These constraints stem from the infrequency of occurrence of such high values, resulting in insufficient data points for effective learning. While attempts were made to address this issue through bias correction using quantile mapping, these efforts proved challenging as only a few values were affected, leading to a limitation in the narrower sense. In evaluating the model's performance, considerations of accuracy, precision, and recall for limit exceedances are essential. It was observed that in cases of incorrectly identified limit exceedances, measured values mostly hovered just above the threshold, while modelled values of unrecorded exceedances ranged from 30 with increasing density up to the limit of 50, as shown in Figure 13. In some cases, even with a significant exceedance of the daily limit value, this exceedance was not accurately captured by the model. The study of Feng et al. [13] achieved values for accuracy and precision of >87% and >86%, respectively. Our model's accuracy of 98.54% also lies above 87%, but the precision of 73.96% is close to but still below 86%. The reason for this could be that this model was not explicitly trained for this task but rather aims to predict an hourly mean value using regression. It may be possible to improve the model for this purpose by training a classification model for daily mean values instead.

Moreover, the model was unable to accurately capture situations originating entirely outside the study area, such as Sahara dust episodes or long-range transboundary transport of air pollution originating from sources like heating and coal power plants, e.g., in Poland or the Czech Republic. Those impacts could be partially covered by taking into account the wind direction, which varies depending on the season of the year, as shown in Figure 14. Pültz et al. [24] investigated the source attribution of particulate matter in Berlin and found that about one-third of the foreign shares can be attributed to Germany's neighbouring countries Poland and the Czech Republic. However, these contributions can differ significantly during episodes. A potential avenue for improvement involves integrating complementary models like the Copernicus Atmosphere Monitoring Service (CAMS) to learn from errors between models, offering a straightforward yet effective approach. Furthermore, the in-

clusion of variables from land use regression, traffic, building density, green area, or noise alongside meteorological and time variables can enhance predictive capabilities.

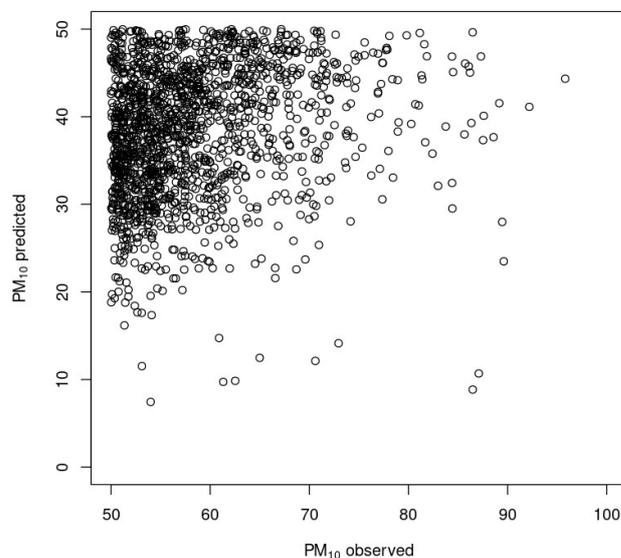


Figure 13. Observed and modelled daily mean PM₁₀ concentrations in $\mu\text{g m}^{-3}$ of the test dataset (2018) for instances where an exceedance of the EU-limit value of $50 \mu\text{g m}^{-3}$ occurred but was not accurately predicted. Cases where no exceedance of the limit value of $50 \mu\text{g m}^{-3}$ occurred are not shown in this graphic.

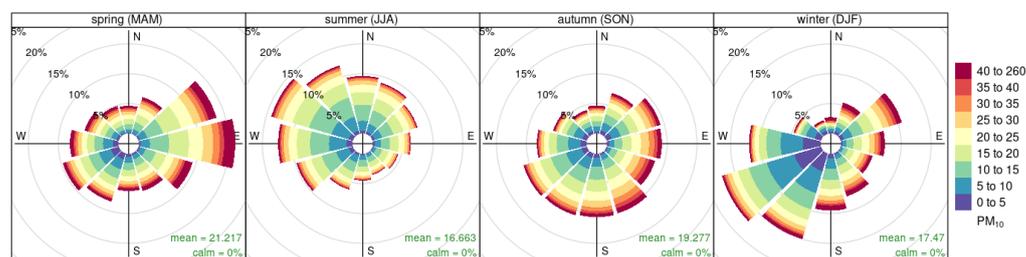


Figure 14. Observed PM₁₀ concentrations in $\mu\text{g m}^{-3}$ with frequency of counts (%) by wind directions (10 m above ground) and seasons of the test data set (2018).

The model demonstrates effectiveness in predicting particulate matter concentrations and may be adaptable for finer fractions of particulate matter. However, caution must be exercised when adapting the model for gaseous contaminants, as their behaviour differs significantly. Additionally, the training and execution of the model are highly cost-, data-, and energy-efficient compared to chemical transport models, aligning with the principles of Green IT and meeting certain environmental requirements.

4.4. Outlook and Further Research Directions

The methodology outlined in Section 2.3 is optimised for maximising predictive accuracy. However, compared to a classical statistical model not tailored to enhance our comprehension of feature influences on particulate matter concentration, it may lack transparency in attributing specific contributions to results. If a deeper understanding of feature effects was the objective, employing an experimental design enabling statistical tests alongside a more interpretable statistical approach, such as a generalised additive model, would be advisable.

Nevertheless, understanding feature importance remains crucial for comprehending the inner workings of the model and assessing whether our knowledge of influences on the response variable aligns with model attributions. The feature importance illustrated in

Figure 2 is quantified using gain, as originally proposed by Breiman [25], which evaluates the data homogeneity of child nodes compared to their parent node in a decision tree. However, despite its widespread use, this method can exhibit inconsistency, as demonstrated by Lundberg et al. [26]. For instance, a feature's reliance within the model may increase even as its importance decreases. This discrepancy arises because early splits in decision trees, being more crucial, tend to be weighted higher, while gain favours later splits, reflecting a bias inherent in the greedy construction of decision trees. This theoretical limitation persists in tree ensembles like XGBoost.

Theoretically superior methods for measuring feature importance include permutation importance and SHAP (SHapley Additive exPlanations) importance, as defined by Lundberg and Lee [27]. These methods offer consistency and align closely with human intuition regarding the significance of features [26]. The SHAP importance for the trained XGBoost model described in Section 3 is visualised in Figure 15. While the feature ranking closely resembles the gain-based feature importance depicted in Figure 2, notable differences arise: the contribution of the previous day's value appears to have been overestimated in the gain-based importance, while the influence of variables such as the height of the planetary boundary layer and temperature is more pronounced.

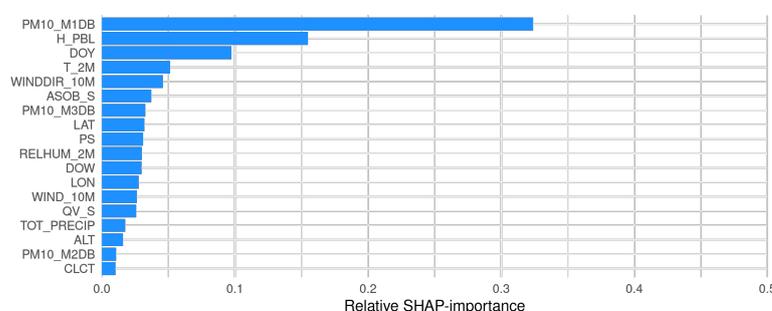


Figure 15. The relative feature importance of the trained XGBoost model as defined by the normalised mean of absolute SHAP values.

Furthermore, SHAP values, the basis for SHAP importance, offer deeper insights by providing feature contributions for each observation. These values, being directional, allow visualisation of not only the magnitude but also the direction of feature contributions; for instance, small values in a feature may correspond to increased values in the target. Additionally, we can visualise the impact of a single feature through dependence plots or for a single observation using waterfall plots, elucidating how each feature contributes to the prediction.

A promising extension is transitioning from point forecasts to probabilistic forecasts, increasingly popular in weather forecasting, as discussed in works such as Gneiting and Katzfuss [28] and Scheuerer and Hamill [29]. This shift offers several benefits, including deeper insights into differences between stations or over time, the quantification of uncertainty through confidence intervals, and the ability to generate varying point forecasts without retraining by retrieving percentiles and expectiles from the distribution. A straightforward approach involves selecting a suitable distribution and forecasting its parameters, as demonstrated in the XGBoostLSS algorithm by März and Kneib [30], which builds upon the standard XGBoost library. Alternatively, more flexible options, such as transformation forests, as explored by Schlosser et al. [31], offer probabilistic forecasting without pre-defining a distribution.

5. Conclusions

Our present study highlights the feasibility of a point-based prediction of PM₁₀ concentration across a large area encompassing numerous stations, leveraging meteorological variables, station metadata, and time variables. The integration of these factors sheds light on their intricate interactions, revealing the high potential for utilising XGBoost algorithms

in predicting PM₁₀ levels. Overall, the performance of the model is deemed satisfactory, notwithstanding the challenges posed by differing station types and associated biases. This model has the capability to provide accurate point-based prediction values for assimilation into the geostatistical land-use regression model presented in the study conducted by Wallek et al. [14], which demonstrated the efficacy of spatial interpolation of point-based PM₁₀ concentrations using a geostatistical land-use regression model with open data. This approach proves capable of delivering satisfactory results for large areas, achieving high spatial (100 m × 100 m) and temporal (hourly) resolutions simultaneously. Utilising the resulting spatially predicted concentrations, it can provide crucial information on inter-urban and regional transport of particulate matter to significantly contribute to improved health outcomes. It is worth noting that the geostatistical model's functions for de-trending and re-trending, coupled with land-use emission coefficients, likely mitigate above-mentioned biases. However, the coupling with geostatistical models may introduce error propagation, necessitating careful consideration when transitioning to an operational model. Despite these challenges, the model's cost-effectiveness, data efficiency, spatio-temporal resolution, and energy efficiency compared to Chemical Transport Models (CTM) make it a promising tool for a broad spectrum of users and stakeholders, ranging from individual citizens to governmental entities at various levels, as well as scientists.

Author Contributions: Conceptualisation, S.W., M.L. and T.S.; methodology, S.W., M.L., S.S. and T.S.; software, S.W. and S.S.; validation, S.W., M.L. and T.S.; formal analysis, S.W. and R.F.; investigation, S.W.; resources, S.S.; data curation, S.W. and S.S.; writing—original draft preparation, S.W.; writing—review and editing, S.W., M.L., S.S., R.F. and T.S.; visualisation, S.W.; supervision, M.L. and T.S.; project administration, S.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to their ownership to the states and the federal government. Access to these datasets necessitates a formal request.

Acknowledgments: We express our profound gratitude to Anselm Arndt for his unwavering and invaluable assistance throughout the technical implementation and execution of model runs on the computing cluster, including prompt resolution of associated issues. Special thanks to Stefan Feigenspan for his valuable guidance and feedback on methodology and evaluation. Our sincere appreciation also goes to our colleagues at the German Meteorological Service for generously providing their data. We are equally grateful to Milan Flach and Andreas Kerschbaumer for their enlightening discussions and fruitful scientific exchanges, particularly in the domains of sampling and variable selection.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. World Health Organization. Ambient (Outdoor) Air Quality and Health. 2021. Available online: [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health) (accessed on 7 February 2024).
2. European Environment Agency. Harm to Human Health from Air Pollution. 2023. Available online: https://www.eea.europa.eu/ds_resolveuid/29d273f7a5ce447cbd588b300a8eab8d (accessed on 7 February 2024).
3. World Health Organization. *WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*; World Health Organization: Geneva, Switzerland, 2021.
4. United Nations Economic Commission for Europe. Convention on Long-Range Transboundary Air Pollution. 1979. Available online: <https://unece.org/sites/default/files/2021-05/1979%20CLRTAP.e.pdf> (accessed on 7 February 2024).
5. European Union. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. *Off. J. Eur. Union* **2008**, *29*, 169–212.
6. Pappa, A.; Kioutsoukakis, I. Forecasting Particulate Pollution in an Urban Area: From Copernicus to Sub-Km Scale. *Atmosphere* **2021**, *12*, 881. [CrossRef]

7. Bailey, J.; Ramacher, M.O.P.; Speyer, O.; Athanasopoulou, E.; Karl, M.; Gerasopoulos, E. Localizing SDG 11.6.2 via Earth Observation, Modelling Applications, and Harmonised City Definitions: Policy Implications on Addressing Air Pollution. *Remote Sens.* **2023**, *15*, 1082. [[CrossRef](#)]
8. Choubin, B.; Abdolshahnejad, M.; Moradi, E.; Querol, X.; Mosavi, A.; Shamshirband, S.; Ghamisi, P. Artificial neural networks forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation. *Atmos. Environ.* **2015**, *107*, 118–128. [[CrossRef](#)]
9. Kowalski, P.A.; Sapała, K.; Warchałowski, W. PM₁₀ forecasting through applying convolution neural network techniques. In *Air Pollution Studies*; WIT Press: Southampton, UK, 2020; p. 47.
10. Czernecki, B.; Marosz, M.; Jędruszkiewicz, J. Assessment of Machine Learning Algorithms in Short-term Forecasting of PM₁₀ and PM_{2.5} Concentrations in Selected Polish Agglomerations. *Aerosol Air Qual. Res.* **2021**, *21*, 200586. [[CrossRef](#)]
11. Park, S.; Son, S.; Bae, J.; Lee, D.; Kim, J.J.; Kim, J. Robust Spatiotemporal Estimation of PM Concentrations Using Boosting-Based Ensemble Models. *Sustainability* **2021**, *13*, 13782. [[CrossRef](#)]
12. Gilik, A.; Ogrenci, A.S.; Ozmen, A. Air quality prediction using CNN+LSTM-based hybrid deep learning architecture. *Environ. Sci. Pollut. Res.* **2022**, *29*, 11920–11938. [[CrossRef](#)] [[PubMed](#)]
13. Feng, X.; Li, Q.; Zhu, Y.; Hou, J.; Jin, L.; Wang, J. Spatial hazard assessment of the PM₁₀ using machine learning models in Barcelona, Spain. *Sci. Total. Environ.* **2020**, *701*, 134474. [[CrossRef](#)] [[PubMed](#)]
14. Wallek, S.; Langner, M.; Schubert, S.; Schneider, C. Modelling Hourly Particulate Matter (PM₁₀) Concentrations at High Spatial Resolution in Germany Using Land Use Regression and Open Data. *Atmosphere* **2022**, *13*, 1282. [[CrossRef](#)]
15. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
16. Hans-Ertel-Centre for Weather Research—Climate Monitoring and Diagnostics (Universities Bonn and Cologne); German Meteorological Service (DWD). COSMO-REA6 Reanalysis Data. 2021. Available online: <https://reanalysis.meteo.uni-bonn.de/?COSMO-REA6> (accessed on 8 August 2021).
17. Bollmeyer, C.; Keller, J.D.; Ohlwein, C.; Wahl, S.; Crewell, S.; Friederichs, P.; Hense, A.; Keune, J.; Kneifel, S.; Pscheidt, I.; et al. Towards a high-resolution regional reanalysis for the European CORDEX domain. *Q. J. R. Meteorol. Soc.* **2015**, *141*, 1–15. [[CrossRef](#)]
18. Wahl, S.; Bollmeyer, C.; Crewell, S.; Figura, C.; Friederichs, P.; Hense, A.; Keller, J.; Ohlwein, C. A novel convective-scale regional reanalyses COSMO-REA2: Improving the representation of precipitation. *Meteorol. Z.* **2016**, *26*, 345–361. [[CrossRef](#)]
19. Frank, C.W.; Wahl, S.; Keller, J.D.; Pospichal, B.; Hense, A.; Crewell, S. Bias correction of a novel European reanalysis data set for solar energy applications. *Sol. Energy* **2018**, *164*, 12–24. [[CrossRef](#)]
20. Banari, A.; Hertel, D.; Schlink, U.; Hampel, U.; Lecrivain, G. Simulation of particle resuspension by wind in an urban system. *Environ. Fluid Mech.* **2023**, *23*, 41–63. [[CrossRef](#)]
21. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2021.
22. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
23. Flentje, H.; Briel, B.; Beck, C.; Collaud Coen, M.; Fricke, M.; Cyrys, J.; Gu, J.; Pitz, M.; Thomas, W. Identification and monitoring of Saharan dust: An inventory representative for south Germany since 1997. *Atmos. Environ.* **2015**, *109*, 87–96. [[CrossRef](#)]
24. Pültz, J.; Banzhaf, S.; Thürkow, M.; Kranenburg, R.; Schaap, M. Source attribution of particulate matter in Berlin. *Atmos. Environ.* **2023**, *292*, 119416. [[CrossRef](#)]
25. Breiman, L. *Classification and Regression Trees*, 1st ed.; Routledge: London, UK, 1984. [[CrossRef](#)]
26. Lundberg, S.M.; Erion, G.; Lee, S.I. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv* **2018**, arXiv:1802.03888.
27. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
28. Gneiting, T.; Katzfuss, M. Probabilistic Forecasting. *Annu. Rev. Stat. Its Appl.* **2014**, *1*, 125–151. [[CrossRef](#)]
29. Scheuerer, M.; Hamill, T.M. Statistical Postprocessing of Ensemble Precipitation Forecasts by Fitting Censored, Shifted Gamma Distributions. *Mon. Weather. Rev.* **2015**, *143*, 4578–4596. [[CrossRef](#)]
30. März, A.; Kneib, T. Distributional Gradient Boosting Machines. *arXiv* **2022**, arXiv:2204.00778.
31. Schlosser, L.; Hothorn, T.; Stauffer, R.; Zeileis, A. Distributional regression forests for probabilistic precipitation forecasting in complex terrain. *Ann. Appl. Stat.* **2019**, *13*, 1564–1589. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.