

Article

Machine-Learning Proxy of Air-Quality Variability Associated with Intercontinental Bridge Traffic over the Bosphorus in Istanbul

Enes Birinci ^{1,*} , Qusay Al Dulaimi ² , Ola Surakhi ³ , Hüseyin Ozdemir ⁴, Ali Deniz ¹ and Tareq Hussein ^{2,5,*} 

¹ Department of Climate Science and Meteorological Engineering, Istanbul Technical University, 34469 Maslak, Istanbul, Türkiye; denizali@itu.edu.tr

² Institute for Atmospheric and Earth System Research (INAR/Physics), Faculty of Science, University of Helsinki, FI-00014 Helsinki, Finland; qusay.aldulaimi@helsinki.fi

³ Faculty of Information Technology, Data Science and Artificial Intelligence Department, American University of Madaba, Madaba 11821, Jordan; o.surakhi@aum.edu.jo

⁴ Eurasia Institute of Earth Sciences, Climate and Marine Sciences Department, Istanbul Technical University, 34469 Maslak, Istanbul, Türkiye; ozdemirhuseyin@itu.edu.tr

⁵ Environmental and Atmospheric Research Laboratory (EARL), Department of Physics, School of Science, University of Jordan, Amman 11942, Jordan

* Correspondence: birinci16@itu.edu.tr (E.B.); tareq.hussein@helsinki.fi (T.H.)

Abstract

Traffic-based air quality proxy models that perform well in compact urban settings have not been evaluated in megacity environments where complex topography, sea–land breeze circulation, and strong meteorological modulation are expected to dominate pollutant transport. We evaluate this transferability for the Bosphorus, an intercontinental corridor between Asia and Europe, using a 17,520-h panel for 2022–2023 that combines hourly traffic counts from six traffic sensors on the 15 Temmuz and Fatih Sultan Mehmet bridges with the closest air quality and meteorological observations. A traffic-proxy baseline (Prophet decomposition with a 9th-degree polynomial for hourly traffic flow) fails uniformly across all bridge × pollutant combinations, yielding negative coefficients of determination. A meteorology-aware machine-learning pipeline (LightGBM, XGBoost, and a bidirectional LSTM-Attention model) trained on traffic, meteorology, temporal indicators, precipitation lags, and short autoregressive terms recovers consistently positive predictive skill for NO₂, NO, NO_x, PM₁₀, PM_{2.5}, and the derived coarse-particle channel across both bridges. The mean of the twelve per-combination hold-out R² values is ≈0.77 (range 0.58–0.86), yielding a 47–63% reduction in RMSE relative to the baseline; the three model families are practically equivalent in overall performance. SHAP ranks temporal autocorrelation, seasonal contrast, wind direction, and heavy vehicles (rather than total traffic) as the most influential predictors, whereas precipitation-delay characteristics carry diagnostic weight for particle channels. Distribution-free 95% prediction intervals are provided through split-conformal calibration. These results indicate that atmospheric persistence dominates predictive skill, with meteorology and traffic as secondary but systematic modulators.

Keywords: air pollution prediction; machine learning; LightGBM; LSTM-Attention; SHAP; Istanbul



Academic Editor: Klaus Schäfer

Received: 15 September 2026

Revised: 13 October 2026

Accepted: 15 October 2026

Published: 18 October 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Road traffic is among the most significant sources of urban air pollution, with emissions of nitrogen oxides (NO_x), particles with a diameter lower than 10 μm (PM₁₀) [1],

and particles with a diameter lower than $2.5\ \mu\text{m}$ ($\text{PM}_{2.5}$) [2] posing a significant risk to human health in areas with high traffic density. Recently, the World Health Organization (WHO) estimated that 4.2 million premature deaths occurred globally in 2016 [3]. Furthermore, recent large-scale systematic reviews have reconfirmed that long-term exposure to traffic-related air pollution is associated with a wide spectrum of diseases ranging from cardiovascular diseases to childhood asthma [4]. Therefore, air quality forecasting and exposure assessment have become increasingly critical agenda items.

The literature uses various approaches to estimate and analyze air pollution. Land Use Regression (LUR) models [1,5] explain long-term average concentrations using geographical predictors, while Chemistry-Transport models (CMAQ, WRF-Chem), which incorporate physical and chemical processes, generate hourly spatial data [6].

Over the last decade, machine learning (ML) methods have proven successful for high-resolution, station-based air quality modeling [7,8]. Gradient boosting [9], random forest-based regression, and deep learning approaches [7,10] have demonstrated particularly strong performance in atmospheric pollutant prediction studies [11–13]. Studies linking traffic volume to air quality concentrations have increased. A traffic-based air quality proxy model was proposed for Helsinki [14], combining the Prophet time series model [15] with 9th-degree polynomial fitting to compare NO_2 and PM_{10} hourly time series with city-scale total traffic flow. In recent years, ML-based air quality prediction studies have intensified in various urban environments. Hybrid neural network frameworks combined with time series decomposition have been used in air quality index (AQI) estimation in major cities in China [16], demonstrating the value of decomposition-based preprocessing in highly variable urban pollution series. Also, Random Forest and Gradient Boosting models achieved strong predictive performance in $\text{PM}_{2.5}$ prediction [17]. Another study compared CNN, ANN, and classical ML models for the city of Jaipur, reporting that CNN performed best on a university campus [18]. LightGBM and ensemble stacking models were applied for NO_2 estimation for the London Marylebone Road high-traffic corridor [19], and traffic-induced NO_2 and $\text{PM}_{2.5}$ were decomposed around Dublin train station with XGBoost-based low-cost sensor calibration [20]. ML-based air pollutant prediction has also been studied in Istanbul. ML-based air quality prediction has rapidly become widespread. Studies focusing on Istanbul report high success rates across architectures and pollutants: $\text{PM}_{2.5}$ prediction for the next hour with stacked LSTM, approximately 1 km NO_2 mapping by combining Sentinel-5P TROPOMI and GEOS-CF with meteorological and anthropogenic features, and winter PM_{10} prediction at seasonally classified Istanbul stations, and SHAP consistently identifies population density, road length, and meteorological factors as dominant predictors [21–23]. On a national scale, $\text{PM}_{2.5}$ trends across Türkiye for 1998–2019 were mapped using satellite-based aerosol optical depth and GEOS-Chem model integration, and a significant upward trend was identified in Bursa, Kütahya, Ankara, and Istanbul [24].

The transferability of traffic-based proxy models from compact urban settings to megacity environments remains unexplored. Helsinki, where the proxy was developed, has a population of approximately 0.7 million, low topographic relief, and weak coastal modulation. Istanbul, with a population of approximately 16 million, presents a contrasting setting in which all road traffic between Europe and Asia is channeled through two suspension bridges carrying approximately 420,000 vehicles per day, and in which complex topography, sea and land breeze circulation, and winter temperature inversions modulate pollutant dispersion on hourly time scales. The timing of these air circulation patterns relative to traffic demand is of critical importance. Boundary layer observations in Istanbul indicate that the layer is frequently shallow and stable during the night and early morning hours; that mixing heights remain below a few hundred meters during

pollution episodes [25] and that persistent winter inversions under blocking conditions can cause PM₁₀ levels to spike to several hundred $\mu\text{g}/\text{m}^3$ within hours [26]. Consequently, while peak morning traffic (07:00–09:00 Local Time) coincides with the phase of the diurnal cycle when pollutant dispersion capacity is at its lowest, midday traffic is diluted by the onset of the sea breeze and the convective expansion of the mixing layer; conversely, peak evening traffic occurs as the sea breeze weakens and the layer subsides. Ultimately, the same hourly traffic volume can result in markedly different concentrations at receptor locations depending on the specific phase of the diurnal air circulation, an interaction that cannot be captured by an indicator (proxy) based solely on traffic data. This sensitivity of near-surface concentrations to local circulation is not specific to Istanbul. At Lake Baikal, surface pollutant concentrations increase under stable stratification and low-level jets, when the vertical turbulent heat flux becomes negative [27], and breeze circulation together with the thermal contrast between the water surface and the surrounding land suppresses convective mixing and retains aerosol near the surface [28]. Whether a traffic proxy remains valid under these conditions, and what additional information is needed if it does not, matters for applying proxy-based monitoring in rapidly growing megacities. In this study, air quality forecasts for the period 2022–2023 on the two intercontinental bridges of the Bosphorus (15 Temmuz and Fatih Sultan Mehmet) were obtained by integrating RTMS sub-2-min traffic count data (sensors 25, 60, 268, 302, 303, 417), Istanbul Metropolitan Municipality hourly air quality measurements (PM₁₀, PM_{2.5}, NO₂, NO_x), and TSMS hourly meteorological observations (temperature, relative humidity, wind speed, wind direction, precipitation). Previous studies have generally focused on a single target pollutant [29–31] or on a single model family [32,33]. Although traffic and meteorological predictors have been combined before, including for roadside corridors in London [19] and Dublin [20] and for Istanbul stations [21–23], these studies did not benchmark against a physically motivated traffic-only baseline, did not quantify the separate contributions of traffic, meteorology and persistence, and did not test whether a proxy developed in one city transfers to another.

The present study addresses these three gaps directly. We geographically matched air quality stations with the nearest meteorological stations. The Helsinki Prophet + 9th-degree polynomial baseline [14] is replicated one-to-one to generate a directly comparable skill score; a feature-group ablation separates the contributions of traffic, meteorology and atmospheric persistence. Cross-bridge experiments test spatial transfer between the two receptors; distribution-free 95% uncertainty bands are provided with split conformal prediction, and LightGBM/XGBoost gradient boosting models are compared with the LSTM-Attention sequence model on the same hold-out. Modeling six pollutants in a single pipeline enables comparison of the primary (NO), secondary (NO₂), and non-combustion (PM_{coarse}) channels under identical data, features, and validation, which is not possible across separate single-pollutant studies. We included split-conformal prediction intervals because point forecasts alone are insufficient for regulatory or exposure assessment applications. Separating the predictor groups also makes it possible to examine whether the dominant role of meteorological factors reported in earlier Istanbul studies is confirmed or needs to be refined once persistence is taken into account. Accordingly, the main research objectives of the present study can be summarized under five headings: (i) simultaneous modeling of six pollutants (PM₁₀, PM_{2.5}, PM_{coarse}, NO₂, NO_x, NO) instead of a single target, (ii) testing whether the compact-city traffic proxy transfers to the Bosphorus, where the two bridges constitute the sole intercontinental road link, (iii) combining SHAP-based interpretability with split conformal uncertainty quantization under the same framework, (iv) comparing two tree-based gradient boosting models (LightGBM, XGBoost) with the LSTM-Attention sequence model in the Bosphorus air basin, and (v) quantifying the sepa-

rate contributions of traffic, meteorology and persistence, and thereby testing whether the previously reported dominance of meteorological factors holds.

2. Materials and Methods

2.1. Study Area and Data

The study area encompasses the two main suspension bridges connecting the Bosphorus on an intercontinental scale, the Maslak and Beşiktaş air quality monitoring stations (AQMS), and the ITU-Maslak and Şişli meteorological stations (Figure 1). The 15 Temmuz Bridge (41.0475° N, 29.0337° E; total span 1560 m), located to the south, has a daily traffic volume of approximately 190,000 vehicles, while the Fatih Sultan Mehmet (FSM) Bridge (41.0914° N, 29.0610° E; total span 1510 m), located approximately 5 km to the north, is the main load-bearing bridge of the corridor with approximately 230,000 vehicles per day. All distances given in this study are great-circle distances from the stated station coordinates to the mid-span coordinates of the bridges. The topographical contrast between the two bridges is significant for local atmospheric mixing: the 15 Temmuz Corridor runs near sea level, while the FSM Corridor passes through higher, hilly areas.

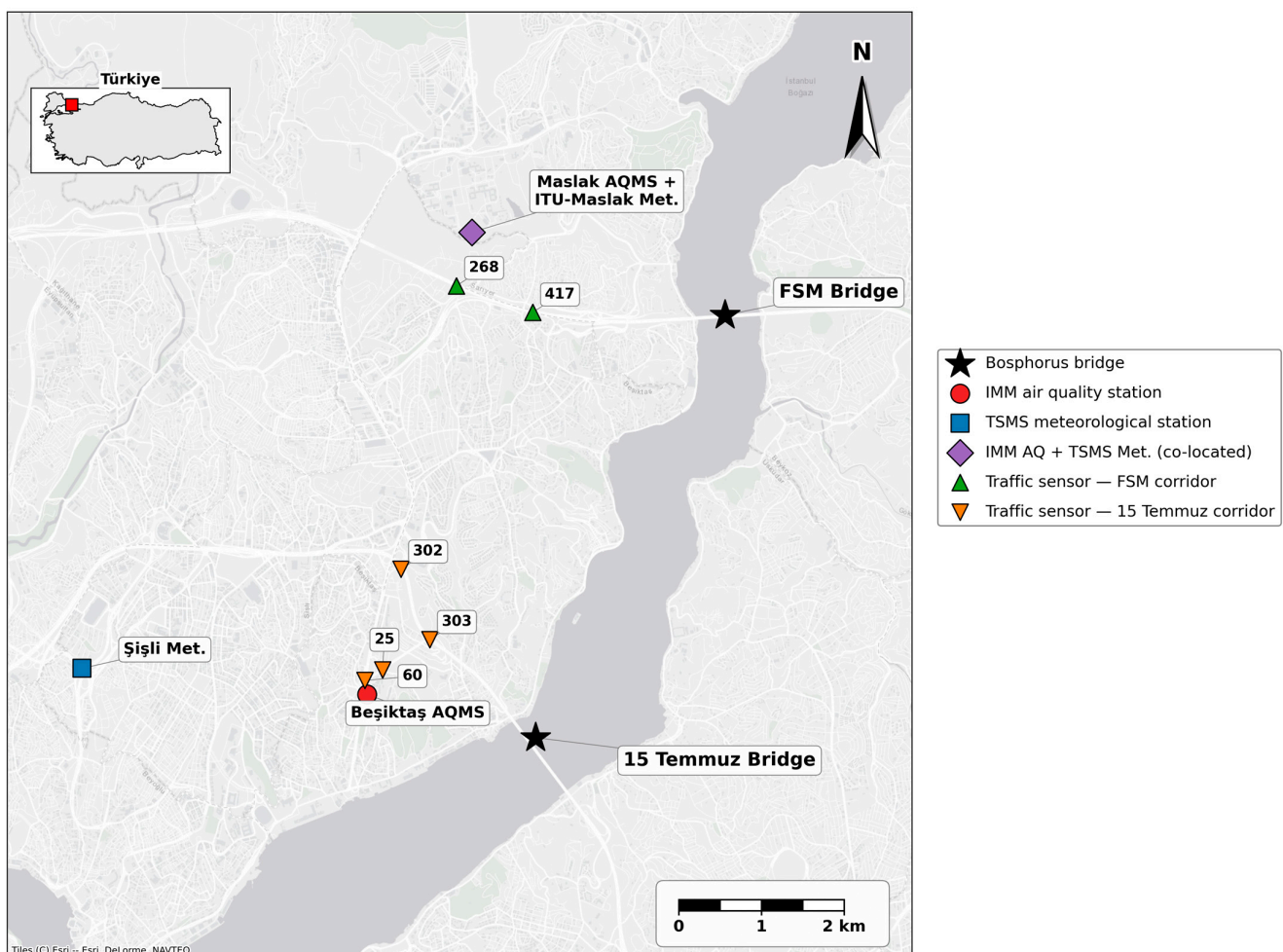


Figure 1. Location of bridges, AQMS, and meteorological stations in İstanbul (The red square in the inset marks the study area).

Air quality observations were compiled from two fixed monitoring stations operated by the Istanbul Metropolitan Municipality (IMM). The Beşiktaş AQMS (41.0520° N, 29.0094° E) is the closest station to the 15 Temmuz Bridge and is located about 2.1 km from the mid-span of the bridge and within approximately 100 m of the bridge approach road, in

a relatively dense urban environment. The Maslak AQMS (41.1000° N, 29.0245° E), located approximately 3.2 km from the mid-span of the FSM Bridge, represents urban background conditions. Meteorological data were obtained from two stations of the Turkish State Meteorological Service (TSMS): Şişli station (41.0547° N, 28.9683° E) and ITU Maslak station (41.0999° N, 29.0251° E). The Şişli station was paired with the Beşiktaş AQMS based on geographical proximity and microclimate similarity, while the İTÜ Maslak station was paired with the Maslak AQMS. Şişli is located approximately 3.4 km southwest of the Beşiktaş AQMS, in the dense Şişli-Beşiktaş urban area, at a similar elevation. Şişli is the closest operational TSMS station to Beşiktaş AQMS. No other TSMS stations are in the Beşiktaş-Şişli district. ITU Maslak is located approximately 50 m from the Maslak AQMS. All four stations lie at similar elevations (Beşiktaş AQMS 85 m, Maslak AQMS 70 m, Şişli 60 m, ITU Maslak 75 m above sea level (Asl)), and air quality inlets are 4 m above ground. The Beşiktaş AQMS is classified by IMM as a traffic station (6 m from the nearest road) and the Maslak AQMS as an urban background station. The decks of both bridges are approximately 64 m above sea level at mid-span. We obtained hourly traffic flow from six radar sensors within the General Directorate of Highways (GDH) Remote Traffic Management System (RTMS). Of the two sensors selected for the FSM corridor, sensor number 268 (TEM–Levent, two-way, 4 + 4 lanes) represents the main artery at the bridge approach, and sensor number 417 (TEM–FSM before the toll booth area, in its modern two-way configuration during the 2019–2023 period, 6 lanes) represents the flow in the direction of travel before the toll booths. For the 15 Temmuz Bridge corridor, four sensor locations were used: 25 (D100–Yıldız Access), 60 (D100–15 Temmuz European entrance), 302 (D100–Zincirlikuyu), and 303 (15 Temmuz–Yıldız Access). Sensor 25 is included in the model as an auxiliary (control) feature because it measures lateral flow toward Beşiktaş, not directly toward the bridge. In contrast, sensors 60, 302, and 303 are officially marked as “Bosphorus Bridge” with the “Outbound” direction label, and these three sensors directly indicate the flow towards the bridge. The modeling period is set as 1 January 2022–31 December 2023; it was run on an hourly grid with $n = 17,520$ h. The window’s upper limit is restricted to the last date of the RTMS data (31 December 2023). This was chosen to fully represent seasonal patterns and to statistically model weekday/weekend contrasts.

RTMS traffic data is collected with timestamps at approximately 2-min intervals for each sensor, and each record includes average speed (km/h), total number of vehicles, and long/short vehicle classification (≥ 7 m and < 7 m classes, respectively) per lane. Direction: Extract direction labels from sensor metadata files. IBB AQMS measures PM_{10} , $PM_{2.5}$, NO_2 , and NO_x concentrations ($\mu g/m^3$) on an hourly basis. At IMM stations, NO , NO_2 and NO_x are measured by chemiluminescence analysers (EN 14211; detection limit ≈ 0.4 ppb) and PM_{10} and $PM_{2.5}$ by beta-attenuation monitors (EN 12341, EN 14907), with calibration and maintenance performed by the IMM Air Quality Monitoring Centre. Meteorological data come from TSMS automated weather observing systems (AWOSs). Traffic counts are recorded by side-fire microwave radar detectors (RTMSs); as an accuracy check, daily counts at sensor 60 agreed within approximately 5% with toll-gate reference counts, and hourly counts at sensors 60 and 303 correlated at $r = 0.956$. However, on-site meteorological variables were completely missing at the Beşiktaş station and 28% missing at the Maslak station during 2022–2023. The primary purpose of AQMS is to measure air pollutants, but some stations in Istanbul also measure meteorological data. No meteorological measurements were taken at the Beşiktaş AQMS during the study period, whereas the Maslak AQMS had 28% data availability. Therefore, the meteorological stations closest to the AQMSs were selected. Şişli and ITU-Maslak stations provide hourly temperature, relative humidity, wind speed, wind direction, and precipitation measurements. To maintain feature symmetry between the two stations, only five common variables were used. NO and

coarse PM (PM_{coarse}) were derived as $NO = NO_x - NO_2$ and $PM_{coarse} = PM_{10} - PM_{2.5}$, with negative values (a small fraction, arising from independent calibration offsets of the two analyzers and therefore physically impossible) clipped at zero. Table 1 presents the technical specifications and missing-data status of all source data. The table includes the number of channels, the percentage of missing data in the expected 17,520-h grid, mean $\pm$ standard deviation, quartiles (Q1, median, Q3), minimum-maximum, skewness, and kurtosis values for each variable.

Table 1. Descriptive statistics of hourly air quality and meteorological observations at the four monitoring stations.

Variable	Unit	<i>n</i>	Missing (%)	Mean	SD	Min	Q1	Median	Q3	Max	Skewness	Kurtosis
Air quality ($\mu\text{g}/\text{m}^3$)												
Beşiktaş AQMS												
NO ₂		17,174	1.97	44.5	19.8	2.8	30.3	41.9	54.3	188	1.18	2.66
NO _x		17,176	1.96	138.5	99.5	5.6	65.6	100.4	188.5	777.2	1.38	1.6
PM ₁₀		17,257	1.5	37.4	23.4	0.4	21.4	32	47.8	253.1	1.64	4.6
PM _{2.5}		17,278	1.38	19.7	11.9	0.8	11.4	17.1	25.1	143.8	1.71	6.13
Maslak AQMS												
NO ₂		14,029	19.93	19.3	14.3	0.5	9.2	15.6	25.2	116.6	1.68	3.73
NO _x		13,973	20.25	53.9	50.1	1.2	20.6	40.6	70.7	777.8	3.38	21.67
PM ₁₀		16,995	3	31.7	24.3	0.1	14.4	26.1	41.8	246.3	1.83	5.91
PM _{2.5}		17,001	2.96	15.7	11.7	0.1	7.3	12.8	21.3	112.7	1.55	3.56
Meteorology												
Sisli TSMS—paired with Besiktas AQMS												
Temperature	°C	17,340	1.03	16.4	7.6	−3.4	10.4	16.3	22.4	39.2	−0.01	−0.79
Relative humidity	%	17,350	0.97	78.3	17.5	20	66	81	95	99	−0.64	−0.48
Wind speed	m/s	17,403	0.67	2.05	0.88	0	1.4	2	2.6	7.2	0.49	0.34
Wind direction *	°	17,403	0.67	16.9	87	-	-	-	-	-	-	-
Precipitation	mm	17,281	1.36	0.09	0.58	0	0	0	0	22.5	16.74	435.98
Maslak TSMS—paired with Maslak AQMS												
Temperature	°C	15,928	9.09	15.8	7.6	−3.6	9.8	15.8	22	40.4	−0.02	−0.77
Relative humidity	%	15,929	9.08	74.2	16.8	17	62	76	88	100	−0.48	−0.46
Wind speed	m/s	12,963	26.01	2.35	1.84	0	1.1	2.1	3.2	37.5	3.27	34.95
Wind direction *	°	12,963	26.01	351.8	78.6	-	-	-	-	-	-	-
Precipitation	mm	16,129	7.94	0.11	0.96	0	0	0	0	67.6	36.18	2026.19

* For wind direction, the circular mean and circular standard deviation are given; the other statistics are not meaningful for a circular variable.

2.2. Data Preprocessing and Quality Control

Two-minute sensor recordings from traffic data were converted to hourly values and reindexed on a full grid of 17,520 h. The total number of vehicles arriving and departing, as well as the number of long vehicles, was totaled within each hour, and average speeds were calculated as vehicle-weighted averages. Values exceeding physical limits in air quality and meteorological measurements were marked as NaN: NO₂ < 0 or >1500 $\mu\text{g}/\text{m}^3$; PM₁₀ < 0 or >1500 $\mu\text{g}/\text{m}^3$; temperature < −25 °C or >50 °C; relative humidity < 0% or >100%; wind speed < 0 or >60 m/s. Repeated occurrences of the same value for ≥ 24 h were recorded as a stuck-value flag in the QC report; automatic deletion was not performed. In the second stage of quality control, winsorization based on 0.5 and 99.5 percentiles was applied for extreme values of the target pollutants [34]. Winsorization was applied before kNN-based imputation. This sorting prevents outliers from contaminating the neighborhood calculus and ensures that the marginal distribution of the imputed series retains its original distribution. This process significantly affected the Maslak NO_x. Skewness decreased from 3.38 to 2.20, over-kurtosis from 21.67 to 7.23, and the maximum value decreased from 777.8 $\mu\text{g}/\text{m}^3$ to 300.9 $\mu\text{g}/\text{m}^3$. Winsorization at the 0.5/99.5 percentiles was applied to stabilize the log-transformed training target

against isolated spikes, not to classify them as measurement artifacts. We acknowledge that some of the capped values may correspond to genuine short-lived pollution episodes (e.g., Saharan dust intrusions or winter inversion events). Because approximately 0.5% of hourly values per channel are affected, the impact on overall skill metrics is marginal; however, the models are consequently not designed to reproduce extreme peaks, and peak underestimation should be expected during exceptional episodes.

2.3. Missing Data Imputation

For the remaining missing values, a data-specific filling method was applied. For air pollution data, we filled missing values using kNN-based assignment ($k = 5$, distance-weighted Euclidean metric). Neighbourhood calculation was defined in the multivariate feature space across concurrent channels (other pollutants, meteorology, and temporal indicators) other than the target pollutant; this approach preserves intervariable dependencies and does not smooth outliers, unlike linear or seasonal mean-based assignment methods [35]. The k value was chosen as 5 based on sensitivity analyses performed for hourly air quality series in the range of 3–10 and the best performance reported in recent years [8]. In this study, with $k = 5$, the central measures (mean, median) and dispersion shape (Q1, Q3) of the assigned dispersion agreed with the original dispersion within 1%; this result confirms that the missing-data filling step preserves the signal structure. For meteorological temperature, relative humidity, and wind speed, forward/backward filling was applied for edge hours following linear interpolation along the time axis [30]. For traffic vehicle numbers, a three-factor seasonal average filling of hour $\times$ day of the week $\times$ month was used; if the matching group was empty, hour $\times$ day of the week and then hour-only backfills were applied. Linear interpolation was used for traffic speeds. For target pollutants, k -nearest neighbor (kNN, $k = 5$, distance-weighted) filling was preferred. The feature domain was constructed from time of day, day of the week, month, and four meteorological channels (T, RH, WS, precipitation) and normalized with StandardScaler. A three-stage dispersion comparison (raw data, post-quality control, and post-quality control + imputation) is presented in Table S4. All preprocessing, quality control, imputation, baseline, machine-learning, and SHAP analyses were performed in Python 3.11 within Google Colab notebooks (Google LLC, Mountain View, CA, USA); the LSTM-Attention model was trained on an NVIDIA Tesla T4 GPU (NVIDIA Corporation, Santa Clara, CA, USA) under PyTorch 2.11.0. To numerically assess the uncertainty of kNN imputation, an artificial masking experiment was performed on a random subsample of the full observed portion of each target pollutant series over 4380 h. For 5%, 10%, and 20% masking ratios, randomly selected observation values were withheld, reimputed with the same kNN procedure ($k = 5$, distance-weighted), and compared with the withheld original values (Table S2). The experiment was repeated on three different random seeds; aggregated RMSE, MAE, and R^2 values are reported as mean $\pm$ standard deviation in Table S1 (Supplementary Material). Imputation recovered 65% to 78% of the variance ($R^2 = 0.65\text{--}0.78$) across eight bridge $\times$ pollutant channels; distortion between 5% and 20% masking levels was limited to $\Delta R^2 \leq 0.03$, and RMSE remained within a small fraction of each channel's own standard deviation. These results confirm that the kNN-based imputation step preserves both the marginal distribution of the original series and the short-term autocorrelation structure.

2.4. Feature Engineering

A total of 22 features were derived across five categories from the filled hourly data to form the common input set for both baseline and machine learning models. Nine variables were used for traffic: hourly incoming and outgoing vehicle counts for each bridge, subtotals for long (≥ 7 m) and short (< 7 m) vehicle classes, vehicle-weighted average speed,

and the two-way directional total. For meteorological input, hourly temperature, relative humidity, wind speed, wind direction, and precipitation were directly used from the paired TSMS station. Wind direction was entered in degrees. In a sensitivity test, wind speed and direction were replaced by the zonal (u) and meridional (v) wind components; hold-out R^2 changed by -0.02 to $+0.01$ across the twelve combinations (mean -0.001), so the original representation was retained. To maintain temporal cyclicity, the time of day and day of the year were coded as sine-cosine pairs; the day of the week and weekend were also added to the panel. To transfer the memory of the target series to the model, autoregressive lags lag-1, lag-24, and lag-168 were derived for each pollutant. To capture the delayed wet-scavenging effect of precipitation, precipitation delays of 1, 2, 3, and 6 h were added, and missing values in this series were filled with zeros to indicate “no rain.” The modeling target set included four directly measured channels (NO_2 , NO_x , PM_{10} , $\text{PM}_{2.5}$) as well as two derivative channels defined as $\text{NO} = \text{NO}_x - \text{NO}_2$ and $\text{PM}_{\text{coarse}} = \text{PM}_{10} - \text{PM}_{2.5}$; small negative values resulting from instrument noise were clipped to zero. All features were normalized with a StandardScaler fitted only on the training set after train/test separation, and the same transformation parameters were applied to the test set to prevent data leakage. The baseline model uses only total bidirectional traffic; LightGBM and XGBoost use all 22 features; the LSTM-Attention model uses the same traffic, meteorological, precipitation-lag, and temporal features, replacing the three autoregressive lag terms with the log-transformed target history within the 24-h input sequence. To quantify the marginal contribution of traffic predictors, a feature-group ablation was conducted with LightGBM under an 8-fold quarterly block cross-validation covering all seasons of 2022–2023. Four configurations were compared: the full 22-feature model, the full model without the nine traffic features, and both configurations without the three autoregressive target lags. The no-lag variants isolate the exogenous signal by removing persistence, which otherwise absorbs most of the hourly variance.

2.5. Baseline Model

The Prophet + 9th-degree polynomial framework of [14] was applied. The hourly total bidirectional traffic flow for each bridge (the two-way directional total for 15 Temmuz, the two-way directional total for FSM) was input into the Prophet time series forecasting model [15]. It was structured with daily, weekly, and annual seasonality components and a multiplicative mode. The hour-of-day average of the Prophet model output was calculated (24 points), and this 24-point series was fitted with a 9th-degree polynomial by least-squares regression. The same hour-of-day average was also calculated for each of the target pollutants. Both the traffic polynomial and the pollutant 24-point series were compared using min-max 0–1 normalization. The baseline is therefore not a forecasting model in the conventional sense: it uses traffic flow as its only input, contains no meteorological or autoregressive predictors, and evaluates whether the mean diurnal shape of traffic reproduces the mean diurnal shape of each pollutant. The multiplicative seasonality mode means daily, weekly, and annual components scale with the trend level rather than being added to it, which suits traffic counts whose diurnal amplitude grows with overall volume.

2.6. Machine Learning Algorithms

2.6.1. LightGBM and XGBoost

LightGBM [36] and XGBoost [37] regression models were trained for each bridge $\times$ pollutant combination; a total of 36 model fits were made (6 pollutants $\times$ 2 bridges $\times$ 3 models: LightGBM, XGBoost, LSTM-Attention). The 168-h delay corresponds to a full week (7×24 h), and its weekly periodicity was specifically chosen to capture the systematic weekday/weekend differentiation and associated pollutant cycles in heavy

vehicle mobility and suburban traffic; shorter delays (1 h and 24 h) cannot achieve this. The target variable was modeled using the $\log_{1p}(y) = \log(1 + y)$ transformation, and predictions were scaled back to the original scale by the inverse transform $\exp(\cdot) - 1$. This transformation stabilizes heavy-tailed pollutant distributions, increasing the model's robustness to extreme values. The training/test split was performed using a time series-based forward-chaining approach [38]. 87.5% of the total 17,520 h of data was used as the training set. The remaining 12.5% is reserved as a hold-out test set. The hold-out window is fixed as the last quartile of the data. This selection spans 1 October to 31 December 2023. This selection reflects three considerations: first, to assess the model's ability to generalize to the most challenging atmospheric conditions; second, to prevent seasonal leakage between the training and test sets (the training set spans two cold seasons while the hold-out constitutes an independent third one); and third, to evaluate performance during the most policy-critical period for air-quality management. The two-year training window aligns with the World Meteorological Organization's climatological practice principles based on full seasonal-cycle observation for climatic representation [39] and is compatible for baseline comparison, as it is of the same order of magnitude as the time horizon used in the Helsinki replication by [14]. All hyperparameter values of the models and the fixed randomness seed are reported in Supplementary Materials Table S1.

2.6.2. LSTM-Attention

A sequence model consisting of a combination of a two-layer bidirectional Long Short-Term Memory [40] network and self-attention pooling [41] was trained. The network input was a feature sequence from the last 24 h, constructed by spreading 20 features derived from traffic, meteorology, temporal indicators, and a log-transformed target variable across 24 time steps. The network architecture consists of a scalar self-attention layer followed by a two-layer bidirectional LSTM block with a hidden dimension of 64, then a dense layer of 32 units, and finally a regression output that produces a single scalar value. To reduce the risk of overfitting, a dropout of 0.2 was applied to all layers. During the training phase, the Adam optimizer (learning rate 1×10^{-3}), mean squared error (MSE) loss, a mini-batch size of 64, and a maximum of 40 epochs were used. For the early stopping criterion, the last 10% of the training set was reserved as a validation set, and training was terminated if no improvement was observed over 6 epochs. Table S3 reports the complete layer-by-layer architecture and training configuration of the LSTM-Attention model.

2.7. Feature Importance Analysis

The SHapley Additive exPlanations (SHAP) framework was used for model explainability [42]. The TreeExplainer algorithm proposed by [43] for tree-based models was preferred in this study for all LightGBM and XGBoost models because it can accurately calculate Shapley values in polynomial time using the internal state of the tree structure. For each bridge $\times$ pollutant combination, the model was evaluated on a hold-out test set, and feature-based Shapley values were derived using the same set as a background reference. The SHAP method has recently become widely used to interpret urban air quality prediction models [23]. SHAP values quantify each feature's contribution to the model's prediction and should be interpreted as measures of model reliance rather than as causal effects.

2.8. Quantification of Uncertainty: Split Conformal Prediction

Split conformal prediction was applied to express model prediction uncertainties with distribution-free reliability [44–46]. 80% of the training set was allocated for model fitting, and the remaining 20% for calibration. The LightGBM model was fitted on the log-transformed target, absolute log-residuals were calculated in the calibration set, and the

critical value q was taken as the quantile $(1 - \alpha + 1/n_{\text{cal}})$. $\alpha = 0.05$ (95% nominal coverage) was targeted. $\pm q$ symmetric intervals were constructed in the log-space around the test set predictions, then translated to the original scale with the inverse transform $\exp(\cdot) - 1$ and capped with a lower limit of 0 (negative pollutant concentrations are physically impossible). The empirical coverage ratio was calculated as the ratio of the condition $y_i \in [L_i, U_i]$ on the test set.

2.9. Validation Strategy and Performance Metrics

All models were evaluated on an independent hold-out test set using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) [47]. The combined use of these three metrics simultaneously quantifies the variance explained by the model (R^2), the level of total error sensitivity to large deviations RMSE, and the MAE, consistent with the triple assessment approach widely adopted in air quality forecasting studies. NRMSE and relative MAE (both normalized by the mean observed concentration) are also reported. All metrics are computed separately for each of the twelve bridge $\times$ pollutant combinations on its own hold-out set. Summary values reported in the text (e.g., mean $R^2 \approx 0.77$) are the unweighted arithmetic means of these twelve per-combination values; no pooled evaluation across combinations is performed, since the pollutants differ in concentration scale. The Supplementary Materials Section S2 provides mathematical definitions of the performance metrics (Equations (S1)–(S4)). All models operate at hourly resolution on the 17,520-h panel. The training/test split is chronological: January 2022 to September 2023 (15,312 h, 87.5%) for training and October to December 2023 (2208 h, 12.5%) as the hold-out test set, with no shuffling. The same split is used for all three model families and for the baseline comparison.

3. Results and Discussion

3.1. Helsinki Baseline Replication

Table S5 shows the baseline results for 10 different combinations. In all ten bridge $\times$ pollutant combinations, the coefficient of determination took a negative value ($R^2 = -0.53$ to -2.71); in other words, baseline estimates produce worse predictions than the average observation. NO channel exhibited the most negative R^2 (-2.71 at 15 Temmuz), reflecting the very short atmospheric lifetime of NO ($\text{NO} + \text{O}_3 \rightarrow \text{NO}_2$ on a timescale of minutes), which precludes a meaningful linear association with the hourly traffic curve. The worst result was obtained on the FSM Bridge PM_{10} ($R^2 = -2.11$). Figure 2 compares the baseline R^2 values for the ten studied areas with the $R^2 \approx 0.5$ reference line reported for Helsinki [14]. This finding shows that the simple traffic-proxy assumption, valid at the Helsinki street-artery scale, cannot be directly transferred to a complex megacity air basin like the Bosphorus. This reflects atmospheric boundary-layer dynamics: an anti-correlation between traffic and PM arises because convective mixing becomes dominant and dilutes pollutant concentrations when traffic density peaks during daylight hours [48]. This behavior is particularly evident at the Maslak urban background station. The Helsinki framework's reliance solely on hour-of-day seasonality fails to capture this meteorological modulation. The fact that the traffic-based proxy structure, successful at the Helsinki scale, produces a negative R^2 in the Bosphorus panel is not an unexpected result from an atmospheric science perspective. In an urban environment with a relatively flat topography and lower vehicle density like Helsinki, the traffic- NO_2 relationship remains almost linear and directly modulated; However, in a megacity like Istanbul, with its complex topography, sea-land breeze oscillations, and intense winter inversions, the modulation of traffic emissions by atmospheric mixing is inevitable [6,49]. In particular, at urban background sites such as Maslak, increased convective-layer thickness during daylight hours dilutes local

emissions, creating a systematic anti-correlation between traffic and PM. Because the pure traffic-proxy structure does not include this directional information, the model's prediction direction can often oppose the observed direction.

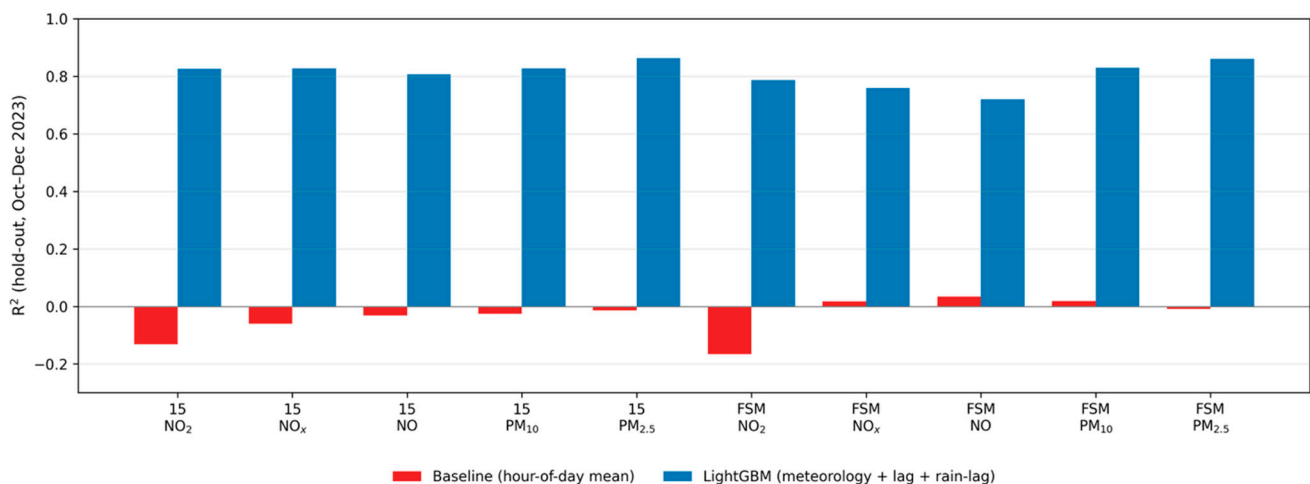


Figure 2. R^2 performance of simple traffic proxy model (Prophet + 9th degree polynomial) for 10 bridge $\times$ pollutant combinations.

3.2. Diurnal, Weekday/Weekend, and Seasonal Cycles

Descriptive time-series analysis of the panel generated over the Bosphorus visually reveals why the baseline model failed and which signals the ML model should utilize. The scaled diurnal cycles in Figure 3 overlay the Prophet + 9th-degree polynomial fit of hourly traffic (black) on the six pollutant traces at the two bridges (Figure 3a, 15 Temmuz; Figure 3b, FSM). The traffic fit rises sharply between 06:00 and 10:00 and then remains on a high plateau until ~22:00, whereas the pollutant series follow markedly different timing. NO₂, NO_x, and NO channels show distinct peaks during the morning (07:00–09:00) and evening (17:00–22:00) commuter windows but drop to minima at midday despite the traffic plateau, while PM₁₀ and PM_{2.5} remain comparatively low through midday owing to the increased atmospheric mixing height. Similar dual-peak commuter patterns with no midday rise have recently been documented in other urban interpretable ML-based air quality studies [50]. Low-cost sensor deployments have likewise reported rush-hour-dominated diurnal NO₂ patterns in Lagos megacity traffic corridors [51]. This pollutant-specific profile differentiation, shown in Figure 3 by the systematic divergence between the smoothed traffic curve and each pollutant trace, explains why a uniform traffic proxy model cannot succeed for all six pollutants simultaneously. The NO trace (purple) shows the sharpest morning peak at both bridges, consistent with NO's short atmospheric residence time, which leads to high concentrations only near the emission source. The derived PM_{coarse} closely tracks the PM₁₀ profile while isolating the non-combustion emission component (road dust, brake/tire wear, road surface wear). Non-exhaust traffic sources have been reported to contribute up to approximately 55% of total traffic-related PM₁₀ by mass, supporting the use of the coarse fraction as a robust tracer of these mechanical processes [52]. A weekday-weekend comparison (Figure S1) shows a 20–35% decrease in NO_x concentrations on weekends due to reduced heavy-vehicle mobility. This effect is more pronounced in the 15 Temmuz corridor than in the FSM, consistent with the stronger weekday-weekend contrast in heavy-vehicle counts on the bridge-bound sensors of this corridor. In the monthly seasonal plot (Figure S2), significant increases are observed across all pollutants during the November–February period. This pattern reflects the combined effects of heating-season emissions and winter inversions [21,30].

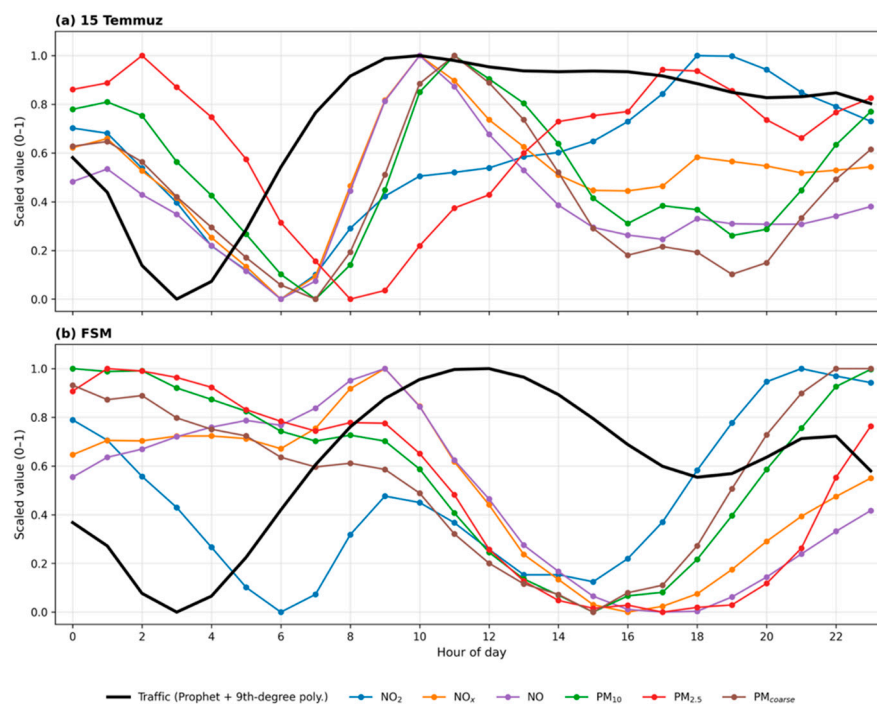


Figure 3. Scaled (0–1 min–max) diurnal profiles of hourly mean concentrations of six pollutants (NO_2 , NO_x , NO , PM_{10} , $\text{PM}_{2.5}$, $\text{PM}_{\text{coarse}}$) and the Prophet + 9th-degree polynomial fit of hourly traffic counts at (a) 15 Temmuz and (b) FSM bridges, 2022–2023.

3.3. Meteorology-Aware Machine Learning Performance

LightGBM and XGBoost models, enriched with meteorological characteristics, lag values, and temporal indicators, produced positive and high coefficients of determination across all twelve bridge $\times$ pollutant combinations in the hold-out test set (Table 2). R^2 values for LightGBM ranged from 0.58 to 0.86, with the highest success rate in the 15 Temmuz $\text{PM}_{2.5}$ channel ($R^2 = 0.86$) and the lowest performance in the $\text{PM}_{\text{coarse}}$ channel at FSM ($R^2 = 0.58$). This reflects the dominance of non-combustion sources (resuspension, brake/tire wear) in the coarse fraction, which traffic counts do not directly capture. XGBoost values differed from LightGBM by less than 0.04; the two libraries are practically equivalent. Recently, hybrid LightGBM-based modeling frameworks combined with SHAP have been applied to NO_2 and $\text{PM}_{2.5}$ at regional scales, and topographic and meteorological factors have also been found to be significant [53]. Comparable ML-based $\text{PM}_{2.5}$ studies in densely populated polluted basins have reported similar R^2 ranges (0.79–0.82) when meteorological predictors such as boundary-layer height and wind components are incorporated [54]. Observation-prediction scatter plots for the twelve combinations are presented in Figure 4 (LightGBM) and Figure S3 (XGBoost) on log-log axes. RMSE decreased by 47–63% relative to the baseline; the greatest improvement was for 15 Temmuz $\text{PM}_{2.5}$, and the least for FSM NO . Relative errors (Table 2) are lowest for NO_2 and $\text{PM}_{2.5}$ at Beşiktaş (NRMSE 15–22%) and highest for $\text{PM}_{\text{coarse}}$ and the FSM nitrogen oxides, where the low mean concentration at the background site and a few high-concentration episodes inflate RMSE more than MAE. Normalized by the hold-out mean, RMSE corresponds to 15–33% for the gas-phase channels at Beşiktaş and 22–31% for PM_{10} and $\text{PM}_{2.5}$ at both sites, but rises to 49–54% for $\text{PM}_{\text{coarse}}$ and 33–87% for the FSM nitrogen oxides (Table 2). The latter reflects the low mean NO_x load at the background site combined with a few high-concentration episodes, which inflate RMSE more than MAE (rMAE 23–43%). The RMSE-based skill score was positive across all combinations. Table 3 shows that removing traffic features from the full model changes cross-validated R^2 by only +0.00 to +0.02, because the lag-1 autoregressive term already encodes the recent emission history. When the autoregressive lags are removed,

however, traffic features contribute a consistent positive increment in all twelve bridge $\times$ pollutant combinations ($\Delta R^2 = +0.01$ to $+0.24$). The increment scales with the primary character of the pollutant—largest for NO ($+0.12$ at 15 Temmuz, $+0.24$ at FSM), intermediate for NO₂ and NO_x, and smallest for the particle channels—consistent with the short atmospheric lifetime of NO as a fresh-exhaust tracer. The contribution is roughly twice as large at Maslak as at Beşiktaş: the Beşiktaş AQMS lies within 100 m of the bridge approach road, so the traffic load is nearly saturated and largely captured by the lag terms, whereas at Maslak the FSM/TEM corridor signal arrives more intermittently and retains independent predictive value. These results indicate that a bridge-traffic signal is detectable at both receptor sites, but is small relative to atmospheric persistence and meteorological modulation. The SHAP analysis (Figures 5 and S6) identified the one-hour autoregressive term of the target pollutant as the dominant predictor in every model, followed by a seasonal day-of-year indicator and the meteorological covariates. Comparable LightGBM-based SHAP frameworks for PM_{2.5} forecasting in megacities have similarly identified meteorological covariates as dominant predictors [31]. Precipitation-delay features (1, 2, 3, and 6-h delays) were added to the model to introduce the delayed wet deposition effect, which manifests as a decrease in pollutant concentrations in the hours following precipitation events. As expected, the strongest contribution to predictive power was concentrated in short precipitation delays, while long-delay terms provided only a marginal improvement; this trend is physically consistent with the fact that the atmospheric persistence of the wet-deposition signal diminishes as the precipitation event moves further into the past. Similar SHAP-based interpretable ML frameworks have been demonstrated for transferable black carbon concentration estimation across European urban and traffic sites, including the Helsinki Mäkeläkatu street canyon [55]. Meteorological integration, rather than model architecture, is the main factor driving model skill; this result quantitatively explains the failure of the Helsinki framework in the Bosphorus and is consistent with similar ML-based studies [31,48,56].

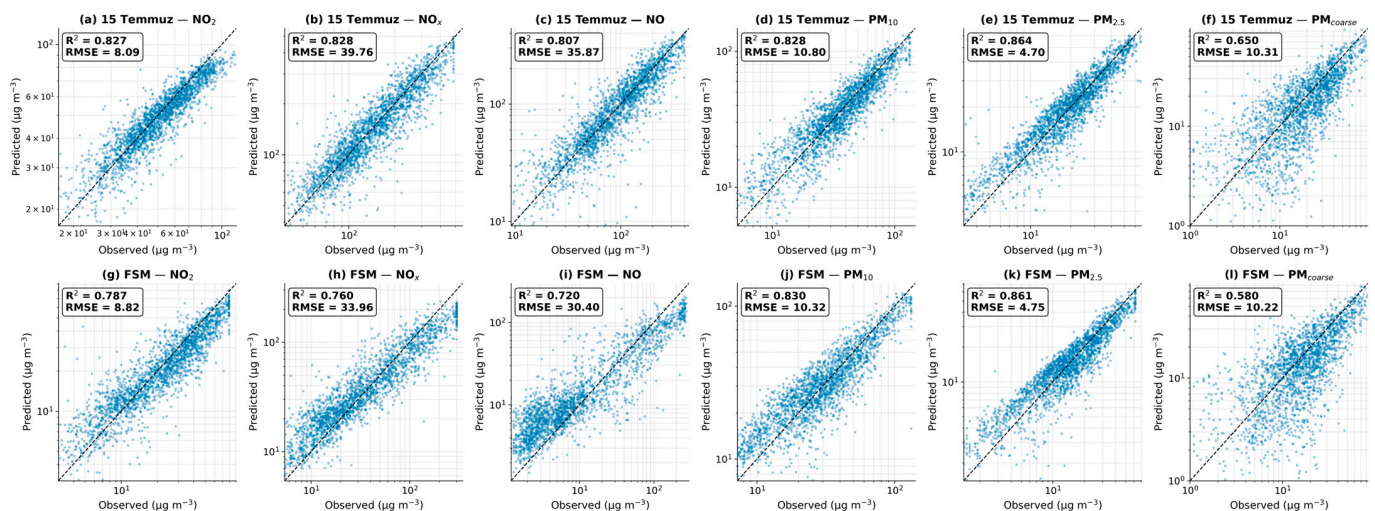


Figure 4. Observed versus predicted scatter plots from the LightGBM model on the hold-out test set (October–December 2023) for the twelve bridge $\times$ pollutant combinations (six pollutants $\times$ two bridges).

Table 2. Hold-out (October–December 2023) performance of LightGBM and XGBoost for the twelve bridge $\times$ pollutant combinations.

Bridge	Pollutant	R ² _{Baseline}	R ² _{LightGBM}	RMSE _{LightGBM}	MAE _{LightGBM}	R ² _{XGBoost}	RMSE _{XGBoost}	MAE _{XGBoost}	NRMSE (%)	rMAE (%)
15 Temmuz	NO ₂	−0.13	0.82	8.09	5.94	0.8	8.6	5.98	15.2	11.1
	NO _x	0	0.82	39.76	27.3	0.82	40.18	27.67	24.5	16.8
	NO	−0.03	0.8	35.87	23.67	0.8	35.81	23.82	32.4	21.3
	PM ₁₀	−0.02	0.82	10.8	7.84	0.82	10.78	7.78	25.5	18.5
	PM _{2.5}	−0.01	0.86	4.7	3.49	0.86	4.74	3.53	22	16.4
	PM _{coarse}	−0.01	0.65	10.31	7.52	0.63	10.48	7.48	49.4	35.7
FSM	NO ₂	−0.16	0.78	8.82	6.12	0.75	9.55	6.51	33.2	22.9
	NO _x	0.01	0.76	33.96	18.83	0.72	36.13	19.48	57.1	31.8
	NO	0.03	0.72	30.4	14.41	0.69	31.94	15.18	86.8	43.1
	PM ₁₀	0.01	0.83	10.32	7.51	0.81	10.68	7.53	30.8	21.7
	PM _{2.5}	0	0.86	4.75	3.49	0.86	4.69	3.47	27.7	20.3
	PM _{coarse}	0.05	0.58	10.22	7.23	0.56	10.39	7.3	54.3	38.4

Table 3. Feature-group ablation of the LightGBM model under 8-fold quarterly block cross-validation (2022–2023). Values are mean hold-out R² across folds. $\Delta R^2 = R^2$ (full) $- R^2$ (without traffic).

Bridge	Pollutant	With Autoregressive Lags			Without Autoregressive Lags		
		R ² Full	R ² No-Traffic	ΔR^2 Traffic	R ² Full	R ² No-Traffic	ΔR^2 Traffic
15 Temmuz	NO ₂	0.870	0.869	0.001	−0.556	−0.617	0.061
	NO _x	0.851	0.838	0.013	−1.063	−1.096	0.033
	NO	0.815	0.793	0.022	−1.310	−1.432	0.122
	PM ₁₀	0.769	0.765	0.004	−0.010	−0.036	0.026
	PM _{2.5}	0.824	0.820	0.004	−0.009	−0.032	0.023
	PM _{coarse}	0.565	0.564	0.001	−0.206	−0.215	0.009
FSM	NO ₂	0.808	0.804	0.004	−0.194	−0.361	0.167
	NO _x	0.770	0.758	0.012	−0.198	−0.315	0.117
	NO	0.732	0.710	0.022	−0.231	−0.475	0.244
	PM ₁₀	0.781	0.777	0.004	−0.268	−0.313	0.045
	PM _{2.5}	0.846	0.848	−0.002	0.024	−0.007	0.031
	PM _{coarse}	0.558	0.548	0.010	−0.566	−0.678	0.112

3.4. SHAP Feature Importance

Shapley values obtained with TreeExplainer reveal a consistent feature importance ranking among the twelve models per learner (twenty-four in total across LightGBM and XGBoost). In all models, the one-hour prior value (lag-1) of the target pollutant had the highest absolute Shapley value, followed by the day-of-year seasonal proxy and meteorological features (wind direction, temperature, relative humidity, wind speed) (Figures 5 and S6). This ranking indicates that seasonal contrast (winter inversions, heating season) and active atmospheric mixing are among the model's main learning signals. Traffic features fall in the middle of the ranking; notably, the number of long (heavy) vehicle entries provides a higher Shapley contribution than total traffic. This result is consistent with heavy-vehicle mobility being a prominent predictor of NO/NO_x channels and supports focusing traffic-related exposure-reduction policies on heavy-vehicle time windows [4]. The added precipitation-delay features (rain_mm_lag1/2/3/6) appear among the top 10 key features in PM channels, indicating that the model reflects delayed wet-removal processes after precipitation. The similar performance of tree-based gradient boosting (LightGBM, XGBoost) and attention-augmented sequence (LSTM-Attention) models on the same dataset within the $|\Delta R^2| \leq 0.07$ limit suggests that, for the architectures tested here, performance is limited more by data and feature representation than by model architecture. This finding is consistent with performance being governed primarily by data quality [7,8].

As a practical recommendation, LightGBM can be preferred for hourly air quality panels with a table structure, such as the Bosphorus. Training cost is about one-tenth that of LSTM, and prediction performance is practically equivalent.

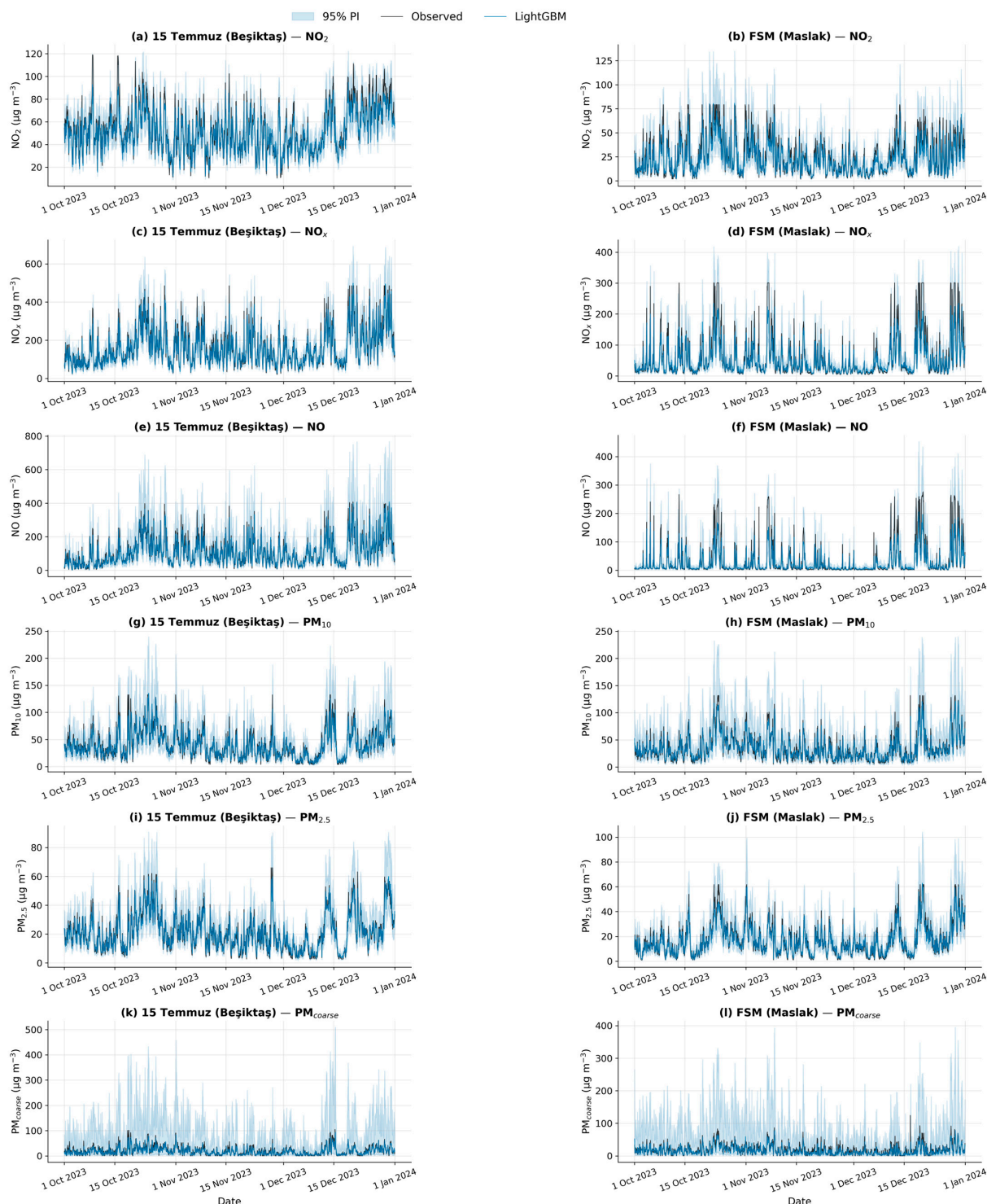


Figure 5. Test-period (Q4 2023) observed (black) and LightGBM-predicted (blue) hourly time series for six pollutants at both bridges.

3.5. Conformal Prediction Intervals

The empirical coverage ratios calculated under the 95% nominal coverage target using split conformal prediction [44,45] ranged from approximately 0.83 to 0.97 for twelve bridge $\times$ pollutant combinations (Table S6); calibration close to the nominal value is

exhibited in PM channels. The overlap of the 95% PI bands with the observations for all combinations is presented in twelve panels in Figure 6 (LightGBM), Figure S4 (XGBoost), and Figure S5 (LSTM-Attention). For NO_2 , NO_x , and NO , empirical coverage remained below the target value; the main reason for this is that the hold-out test set was fixed to the October–December period and the calibration set represents relatively warmer months. As a result, a distribution shift occurs between the calibration and test sets, and the standard split conformal structure cannot fully reflect this shift [46,57]. This situation is an honest methodological limitation.

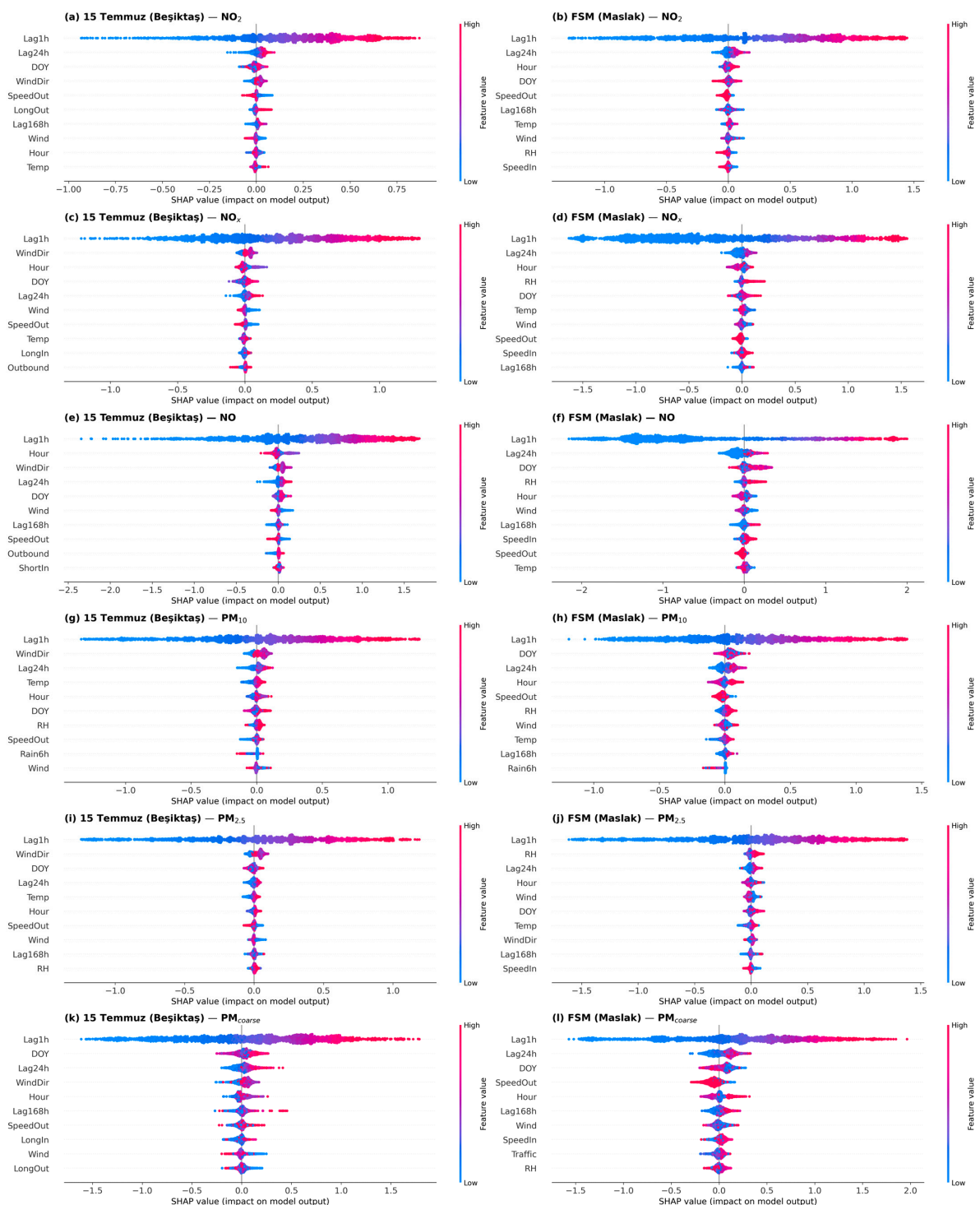


Figure 6. Global SHAP (TreeExplainer) beeswarm plots from the LightGBM models for six pollutants at both bridges.

3.6. LSTM-Attention vs. LightGBM Comparison

The performance difference between the LSTM-Attention sequencing model and the LightGBM gradient boosting model on the same hold-out was negligible in practice. Mean hold-out R^2 across the twelve bridge $\times$ pollutant combinations was 0.78 for LightGBM, 0.76 for XGBoost, and 0.77 for LSTM-Attention; individual differences remained below the $|\Delta R^2| \leq 0.08$ limit (Figure 7). Also, the complete per-combination R^2 and ΔR^2 values for all three models are listed in Table S7. LSTM showed the greatest superiority for FSM NO_x (+0.06), followed by FSM NO (+0.05) and 15 Temmuz NO_2 (+0.02); in PM channels, the two models performed nearly identically. This pattern is consistent with LSTM's ability to better capture the clock-scale dynamics of short-lived gas-phase pollutants thanks to its sequence structure [33,40]. To assess spatial transferability, a LightGBM model was trained on data from one bridge and evaluated on held-out test data from the other, using a harmonized feature set (Table 4). Because particle concentrations at both monitoring sites are driven by regional and meteorological conditions, the particle channels proved almost fully transferable in both directions (PM_{10} $R^2 = 0.81$, $\text{PM}_{2.5}$ 0.85–0.86, $\text{PM}_{\text{coarse}}$ 0.54–0.58, retaining 87–99% of the in-bridge performance). In contrast, nitrogen oxide channels exhibited asymmetric transferability: while the model trained on Beşiktaş data fully maintained its performance in Maslak (NO_2 0.79; NO_x 0.83; NO 0.84), the model trained on Maslak data lost a significant portion of its performance in Beşiktaş (0.49, 0.22, and 0.27, respectively). Extrapolation fails only in this direction because the model based on the background station is never exposed to the high NO_x levels found at the roadside station. Therefore, if surrogate models for primary traffic pollutants are to be transferred within the same airshed, they should be trained using data from the monitoring site with the higher pollution load.

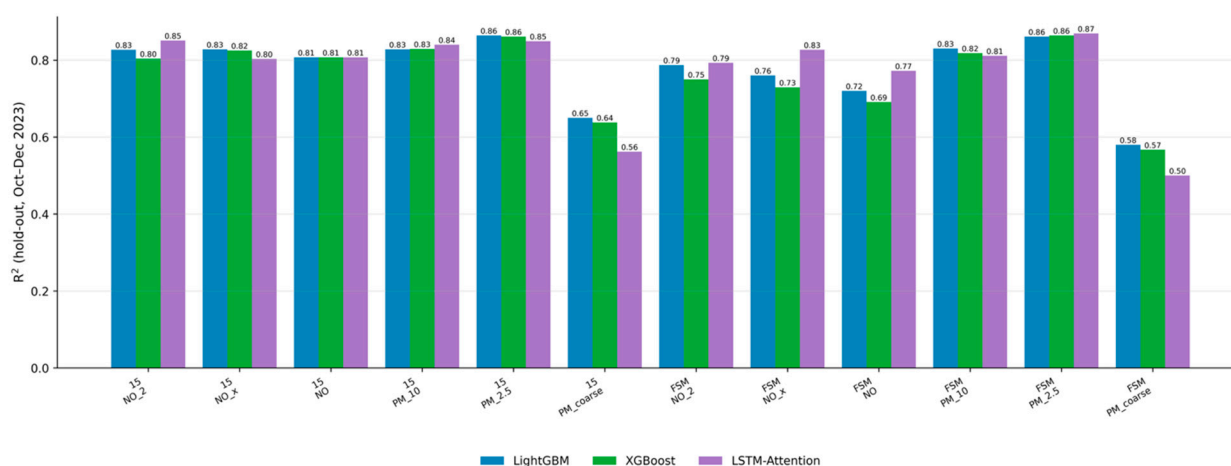


Figure 7. Hold-out R^2 comparison of three model families—LightGBM (blue), XGBoost (green), and LSTM-Attention (purple)—across the twelve bridge $\times$ pollutant combinations for the Oct–Dec 2023 test period.

3.7. Limitations

First, using a two-year dataset from 2022 to 2023 as a single full cold-season (October–December 2023) hold-out results in conformal calibration in gas-phase channels (NO_2 , NO_x , NO) falling below the target value; longer data series are expected to resolve this limitation. Furthermore, the NO and $\text{PM}_{\text{coarse}}$ variables were not directly measured but derived as $\text{NO} = \text{NO}_x - \text{NO}_2$ and $\text{PM}_{\text{coarse}} = \text{PM}_{10} - \text{PM}_{2.5}$, respectively; this process relatively increases measurement noise and explains the modeling difficulty of the NO channel ($R^2 = 0.77$ –0.81) and the $\text{PM}_{\text{coarse}}$ ($R^2 = 0.58$ –0.67). Second, the study was conducted on only two AQMSs; cross-bridge transfer (Table 4) shows that particle models

transfer fully, whereas nitrogen oxide models transfer only from the higher-load to the lower-load site, so generalisation to other Istanbul air basins remains to be tested. Beşiktaş and Maslak AQMS are located approximately 2.1 km and 3.2 km from the mid-span of the respective bridges, and Maslak represents urban-background rather than roadside conditions. Consequently, the models quantify the statistical association between bridge traffic and receptor-site concentrations rather than directly attributing concentrations to bridge emissions. The dominance of the lag-1 autoregressive term in the SHAP rankings (Figures 5 and S6) further confirms that atmospheric persistence, rather than instantaneous traffic, carries the largest share of the predictive signal; traffic features should therefore be read as modulating rather than determining factors. Winsorization at the 99.5th percentile limits the models' ability to reproduce extreme pollution episodes; peak-focused applications would require either an uncapped target or a dedicated extreme-value model. The current model is purely data-driven and has not been hybridized with chemistry-transport outputs (CMAQ, WRF-Chem); this hybrid approach could provide additional gains in medium-to-long-term forecasting horizons. Finally, due to official confidentiality restrictions on RTMS sensor locations, only approximate coordinates are shared in this article; an official data sharing protocol with the General Directorate of Highways is required for sensor location-specific analyses.

Table 4. Spatial transferability of the LightGBM model between the two bridges. In-bridge values refer to models trained and tested at the same bridge, while cross-bridge values refer to models trained at one bridge and tested at the other.

Test Bridge	Pollutant	R ² in-Bridge	R ² Cross-Bridge	Retained Skill (%)
15 Temmuz	NO ₂	0.826	0.492	59.6
	NO _x	0.827	0.219	26.5
	NO	0.809	0.272	33.6
	PM ₁₀	0.835	0.811	97.1
	PM _{2.5}	0.863	0.853	98.8
	PM _{coarse}	0.668	0.584	87.4
FSM	NO ₂	0.798	0.793	99.4
	NO _x	0.775	0.829	107
	NO	0.747	0.841	112.6
	PM ₁₀	0.816	0.812	99.5
	PM _{2.5}	0.863	0.857	99.3
	PM _{coarse}	0.578	0.538	93.1

4. Conclusions

In this study, a meteorology-informed machine learning proxy framework was developed for six target pollutants (NO₂, NO_x, NO, PM₁₀, PM_{2.5}, and the derived coarse particulate channel (PM₁₀ – PM_{2.5})) by integrating traffic count data, air quality measurements, and meteorological observations from a 17,520-h panel of two intercontinental bridges over the Bosphorus Strait during the 2022–2023 period. The study's main findings can be summarized in four points.

First, the simple traffic proxy model produced a negative R² (between –0.53 and –2.71) for all ten bridge × pollutant combinations; this result revealed that the traffic proxy approach, valid at the compact city scale, cannot be directly transferred to megacity basins. Secondly, the gradient boosting models (LightGBM and XGBoost), enriched with meteorological features and historical observational values, produced positive coefficients of determination in the range of R² = 0.58 to 0.86 on the same hold-out, resulting in a 47–63% reduction in RMSE relative to the baseline and a positive skill score in all twelve combinations. Thirdly, the attention-augmented sequence learning model (LSTM-Attention)

and the two tree-based gradient boosting models (LightGBM, XGBoost) yielded practically identical performance (mean hold-out $R^2 = 0.77, 0.78$, and 0.76 , respectively); the greatest sequence-model superiority was observed in the FSM NO_x (+0.06) and FSM NO (+0.05) channels. For the three model families examined here, this result suggests that performance is constrained more by the information available in the data and features than by the choice of architecture. Whether this also holds for other architectures, such as transformer-based or graph neural models, remains an open question. Three approaches are proposed for future studies: (i) expanding to medium-term forecasting horizons by establishing a hybrid structure with chemical-transport model outputs; (ii) fusion of satellite-derived NO_2 column data with ground measurements; and (iii) testing the spatial generalization capacity by extending the framework to other monitoring stations in different air basins of Istanbul. The main contribution of this study is to quantitatively demonstrate that a simple air quality proxy framework valid in compact cities cannot be directly transferred to a megacity scale, and that meteorological integration substantially improves skill.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/atmos17101024/s1>. Table S1: Hyperparameter values used for the LightGBM and XGBoost regression models across all twelve bridge $\times$ pollutant combinations; Table S2: Imputation validation results from the artificial-masking experiment; Table S3: Architecture and training configuration of the LSTM-Attention sequence model; Table S4: Three-stage descriptive statistical comparison of hourly air quality and meteorological variables (raw, after quality control, after imputation); Table S5: Performance of the Prophet + 9th-degree polynomial traffic proxy for ten bridge $\times$ pollutant combinations; Table S6: Split conformal prediction coverage analysis for twelve combinations; Table S7: Performance comparison of LightGBM, XGBoost, and LSTM-Attention across the twelve bridge $\times$ pollutant combinations; Figure S1: Diurnal profiles split by weekday and weekend for six pollutants at the two bridges; Figure S2: Monthly mean concentrations of six pollutants at the two bridges; Figure S3: Observed versus predicted scatter plots from the XGBoost model on the hold-out test set; Figure S4: Test-period observed and XGBoost-predicted hourly time series; Figure S5: Test-period observed and LSTM-Attention-predicted hourly time series; Figure S6: Global SHAP beeswarm plots from the XGBoost models; Equations S1–S4: Performance metrics (R^2 , RMSE, MAE, MAPE).

Author Contributions: Conceptualization: E.B. and T.H.; methodology: E.B. and T.H.; software: E.B.; validation: E.B. and O.S.; formal analysis: E.B. and Q.A.D.; investigation: E.B. and Q.A.D.; data curation: E.B. and H.O.; writing—original draft preparation: E.B.; writing—review and editing: A.D., T.H., Q.A.D., O.S. and H.O.; visualization: E.B.; supervision: A.D. and T.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scientific Research Projects Department of Istanbul Technical University (ITU), grant number 47312.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Air pollution data are publicly available online at <https://havakalitesi.ibb.gov.tr/>, accessed on 5 September 2026.

Acknowledgments: This article is based on the ongoing Ph.D. dissertation of the first author (E.B.) at Istanbul Technical University. This research was also part of a close collaboration between the Environmental and Atmospheric Research Laboratory (EARL, University of Jordan) and the Institute for Atmospheric and Earth System Research (INAR/Physics, University of Helsinki). Generative AI tools (Claude Opus 5.5, Anthropic) assisted with code development for data processing and model implementation. All scientific content, analyses, interpretations, and conclusions were developed, verified, and approved by the authors, who take full responsibility for the content of the manuscript.

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

References

1. Beelen, R.; Raaschou-Nielsen, O.; Stafoggia, M.; Andersen, Z.J.; Weinmayr, G.; Hoffmann, B.; Wolf, K.; Samoli, E.; Fischer, P.; Nieuwenhuijsen, M.; et al. Effects of long-term exposure to air pollution on natural-cause mortality: An analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* **2014**, *383*, 785–795. [CrossRef] [PubMed]
2. Luo, Z.; Wang, Y.; Lv, Z.; He, T.; Zhao, J.; Wang, Y.; Gao, F.; Zhang, Z.; Liu, H. Impacts of vehicle emission on air quality and human health in China. *Sci. Total Environ.* **2022**, *813*, 152655. [CrossRef] [PubMed]
3. WHO. *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. Available online: <https://apps.who.int/iris/handle/10665/345329> (accessed on 10 October 2026).
4. Boogaard, H.; Patton, A.; Atkinson, R.; Brook, J.; Chang, H.; Crouse, D.; Fussell, J.; Hoek, G.; Hoffmann, B.; Kappeler, R.; et al. Long-term exposure to traffic-related air pollution and selected health outcomes: A systematic review and meta-analysis. *Environ. Int.* **2022**, *164*, 107262. [CrossRef] [PubMed]
5. Hoek, G.; Beelen, R.; De Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42*, 7561–7578. [CrossRef]
6. Im, U.; Christensen, J.H.; Geels, C.; Hansen, K.M.; Brandt, J.; Solazzo, E.; Alyuz, U.; Balzarini, A.; Baro, R.; Bellasio, R.; et al. Influence of anthropogenic emissions and boundary conditions on multi-model simulations of major air pollutants over Europe and North America in the framework of AQMEII3. *Atmos. Chem. Phys.* **2018**, *18*, 8929–8952. [CrossRef] [PubMed]
7. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef] [PubMed]
8. Méndez, M.; Merayo, M.G.; Núñez, M. Machine learning algorithms to forecast air quality: A survey. *Artif. Intell. Rev.* **2023**, *56*, 10031–10066. [CrossRef] [PubMed]
9. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
10. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
11. Hu, X.; Belle, J.H.; Meng, X.; Wildani, A.; Waller, L.A.; Strickland, M.J.; Liu, Y. Estimating PM_{2.5} concentrations in the conterminous United States using the random forest approach. *Environ. Sci. Technol.* **2017**, *51*, 6936–6944. [CrossRef] [PubMed]
12. Liao, Q.; Zhu, M.; Wu, L.; Pan, X.; Tang, X.; Wang, Z. Deep Learning for Air Quality Forecasts: A Review. *Curr. Pollut. Rep.* **2020**, *6*, 399–409. [CrossRef]
13. Pak, U.; Ma, J.; Ryu, U.; Ryom, K.; Juhyok, U.; Pak, K.; Pak, C. Deep learning-based PM_{2.5} prediction considering the spatiotemporal correlations: A case study of Beijing, China. *Sci. Total Environ.* **2020**, *699*, 133561. [CrossRef] [PubMed]
14. Račić, N.; Petrić, V.; Mureddu, F.; Portin, H.; Niemi, J.V.; Hussein, T.; Lovrić, M. A Proxy Model for Traffic Related Air Pollution Indicators Based on Traffic Count. *Atmosphere* **2025**, *16*, 538. [CrossRef]
15. Taylor, S.J.; Letham, B. Forecasting at scale. *Am. Stat.* **2018**, *72*, 37–45. [CrossRef]
16. Sun, M.; Rao, C.; Hu, Z. Air quality prediction using a novel three-stage model based on time series decomposition. *Environ. Dev. Sustain* **2026**, *28*, 307–332. [CrossRef]
17. Panaite, F.A.; Rus, C.; Leba, M.; Ionica, A.C.; Windisch, M. Enhancing air-quality predictions on university campuses: A machine-learning approach to PM_{2.5} forecasting at the University of Petroșani. *Sustainability* **2024**, *16*, 7854. [CrossRef]
18. Singh, S.; Suthar, G. Machine learning and deep learning approaches for PM_{2.5} prediction: A study on urban air quality in Jaipur, India. *Earth Sci. Inf.* **2025**, *18*, 97. [CrossRef]
19. Chauhan, B.V.S.; Berg, M.J.; Smallbone, K.L.; Rautela, I.; Ballal, S.; Wyche, K.P. Machine Learning Driven Prediction and Analysis of NO₂ and its Catalyst Based Reduction in Urban Environments. *Top. Catal.* **2025**, *68*, 2089–2108. [CrossRef]
20. Priyan, S.; Guo, Y.; McNabola, A.; Broderick, B.; Caulfield, B.; O'Mahony, M.; Gallagher, J. Detecting and quantifying PM_{2.5} and NO₂ contributions from train and road traffic in the vicinity of a major railway terminal in Dublin, Ireland. *Environ. Pollut.* **2024**, *361*, 124903. [CrossRef] [PubMed]
21. Eren, B.; Aksangür, İ.; Erden, C. Predicting next hour fine particulate matter (PM_{2.5}) in the Istanbul Metropolitan City using deep learning algorithms with time windowing strategy. *Urban Clim.* **2023**, *48*, 101418. [CrossRef]
22. Yagmur Aydin, N. Machine Learning-Based Ground-Level NO₂ Estimation in Istanbul: A Comparative Analysis of Sentinel-5P and GEOS-CF. *Appl. Sci.* **2025**, *15*, 10997. [CrossRef]
23. Birinci, E.; Ekmekcioğlu, Ö.; Ozdemir, H.; Deniz, A. Interpretable machine learning framework for air quality prediction in Istanbul using Shapley additive explanations (SHAP). *Stoch. Environ. Res. Risk Assess.* **2026**, *40*, 37. [CrossRef]
24. Zeydan, Ö.; Tariq, S. Satellite-based mapping of PM_{2.5} pollution in Türkiye: Overcoming gaps in ground-level measurements. *J. Atmos. Chem.* **2026**, *83*, 4. [CrossRef]

25. Yavuz, V.; Özen, C.; Çapraz, Ö.; Özdemir, E.T.; Deniz, A.; Akbayır, İ.; Temur, H. Analysing of atmospheric conditions and their effects on air quality in Istanbul using SODAR and CEILOMETER. *Environ. Sci. Pollut. Res.* **2022**, *29*, 16213–16232. [[CrossRef](#)] [[PubMed](#)]
26. Özdemir, E.T.; Birinci, E.; Deniz, A. Multi-source observations on the effect of atmospheric blocking on air quality in İstanbul: A study case. *Environ. Monit. Assess.* **2024**, *196*, 698. [[CrossRef](#)] [[PubMed](#)]
27. Shikhovtsev, M.Y.; Shikhovtsev, A.Y.; Kovadlo, P.G.; Obolkin, V.A.; Molozhnikova, Y.V. Influence of Air Movement Structure on Microphysical Properties of the Atmosphere over Listvyanka. *Atmos. Ocean Opt.* **2025**, *38*, 180–187. [[CrossRef](#)]
28. Shikhovtsev, M.; Aslamov, I.; Shikhovtsev, A.; Molozhnikova, Y. Spatial distribution and factors determining the level of particulate matter over Lake Baikal. *J. Gt. Lakes Res.* **2026**, 102937. [[CrossRef](#)]
29. Makhdoomi, A.; Sarkhosh, M.; Ziaei, S. PM2.5 concentration prediction using machine learning algorithms: An approach to virtual monitoring stations. *Sci. Rep.* **2025**, *15*, 8076. [[CrossRef](#)] [[PubMed](#)]
30. Lee, J.; Barquilla, C.A.M.; Park, K.; Hong, A. Urban form and seasonal PM2.5 dynamics: Enhancing air quality prediction using interpretable machine learning and IoT sensor data. *Sustain. Cities Soc.* **2024**, *117*, 105976. [[CrossRef](#)]
31. Zhang, Y.; Sun, Q.; Liu, J.; Petrosian, O. Long-term forecasting of air pollution particulate matter (PM2.5) and analysis of influencing factors. *Sustainability* **2023**, *16*, 19. [[CrossRef](#)]
32. Ma, X.; Chen, T.; Ge, R.; Xv, F.; Cui, C.; Li, J. Prediction of PM2.5 concentration using spatiotemporal data with machine learning models. *Atmosphere* **2023**, *14*, 1517. [[CrossRef](#)]
33. Yin, C.; Li, W.; Li, T.; Yuan, S.; Tang, D.; Dong, X. Air quality prediction model based on deep learning hybrid framework. *Sci. Rep.* **2026**, *16*, 7084. [[CrossRef](#)] [[PubMed](#)]
34. Junger, W.L.; De Leon, A.P. Imputation of missing data in time series for air pollutants. *Atmos. Environ.* **2015**, *102*, 96–104. [[CrossRef](#)]
35. Hadeed, S.J.; O’rourke, M.K.; Burgess, J.L.; Harris, R.B.; Canales, R.A. Imputation methods for addressing missing data in short-term monitoring of air pollutants. *Sci. Total Environ.* **2020**, *730*, 139140. [[CrossRef](#)] [[PubMed](#)]
36. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
37. 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. [[CrossRef](#)]
38. Bergmeir, C.; Benítez, J.M. On the use of cross-validation for time series predictor evaluation. *Inf. Sci.* **2012**, *191*, 192–213. [[CrossRef](#)]
39. WMO. *Guide to Climatological Practices* (WMO-No. 100), 3rd ed.; World Meteorological Organization: Geneva, Switzerland, 2011. Available online: <https://library.wmo.int/idurl/4/60113> (accessed on 10 October 2026).
40. Hochreiter, S. *Long Short-Term Memory*; Neural Computation MIT-Press: Cambridge, MA, USA, 1997.
41. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6000.
42. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
43. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [[CrossRef](#)] [[PubMed](#)]
44. Vovk, V.; Gammerman, A.; Shafer, G. *Algorithmic Learning in a Random World*; Springer: Boston, MA, USA, 2005. [[CrossRef](#)]
45. Lei, J.; G’Sell, M.; Rinaldo, A.; Tibshirani, R.J.; Wasserman, L. Distribution-free predictive inference for regression. *J. Am. Stat. Assoc.* **2018**, *113*, 1094–1111. [[CrossRef](#)]
46. Angelopoulos, A.N.; Bates, S. Conformal prediction: A gentle introduction. *Found. Trends Mach. Learn.* **2023**, *16*, 494–591. [[CrossRef](#)]
47. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [[CrossRef](#)]
48. Stirnberg, R.; Cermak, J.; Kotthaus, S.; Haeffelin, M.; Andersen, H.; Fuchs, J.; Kim, M.; Petit, J.-E.; Favez, O. Meteorology-driven variability of air pollution (PM 1) revealed with explainable machine learning. *Atmos. Chem. Phys.* **2021**, *21*, 3919–3948. [[CrossRef](#)]
49. Stull, R.B. *An Introduction to Boundary Layer Meteorology*; Springer Science & Business Media: Dordrecht, The Netherlands, 2012.
50. Wang, Y.; Cai, L.; Shi, G.; Liu, C.; Yao, L.; Li, H.; Du, L.; Ni, X.; Hu, Y.; Pang, D. From traffic flow to ozone surges: Exploring threshold effects and interactive drivers of urban ozone pollution in arid regions. *Urban Clim.* **2026**, *65*, 102801. [[CrossRef](#)]
51. Olufemi, A.P.; Fawole, O.G.; Owoade, O.K.; Popoola, O.A.M.; Jones, R.L.; Toyeye, A.B.; Abiodun, P.O.; Omokungbe, O.R.; Abiye, O.E.; Sunmonu, L.A. Spatiotemporal distribution of pollutants and impact of local meteorology on source influence on pollutants’ level in a traffic air-shed in Lagos megacity, Nigeria. *Environ. Monit. Assess.* **2023**, *195*, 1126. [[CrossRef](#)] [[PubMed](#)]
52. Grigoratos, T.; Martini, G. Brake wear particle emissions: A review. *Environ. Sci. Pollut. Res.* **2015**, *22*, 2491–2504. [[CrossRef](#)] [[PubMed](#)]

53. Zhu, W.; Du, N.; Wang, L.; Zuo, Y.; Deng, X.; Geng, Y.; Qian, L. Joint Estimation and Driving Mechanism Analysis of NO₂ and PM_{2.5} in Northwest China Based on a Hybrid Modeling Approach. *Atmos. Pollut. Res.* **2025**, *16*, 102787. [[CrossRef](#)]
54. Singh, V.; Singh, S.; Sharma, N.; Singh, A.; Srivastava, A.; Srivastava, A.K.; Bisht, D.S.; Patel, K.; Singh, N.; Almazroui, M.; et al. Estimation of surface PM_{2.5} over the Indo-Gangetic Basin using MERRA-2 reanalysis and machine learning. *Sci. Rep.* **2026**, *16*, 13755. [[CrossRef](#)] [[PubMed](#)]
55. Fung, P.L.; Savadkoobi, M.; Zaidan, M.A.; Niemi, J.V.; Timonen, H.; Pandolfi, M.; Alastuey, A.; Querol, X.; Hussein, T.; Petäjä, T. Constructing transferable and interpretable machine learning models for black carbon concentrations. *Environ. Int.* **2024**, *184*, 108449. [[CrossRef](#)] [[PubMed](#)]
56. Alrashidi, H.; Sibai, F.N.; Abonamah, A.; Alrashidi, M.; Alsaber, A. PM_{2.5}: Air Quality Index Prediction Using Machine Learning: Evidence from Kuwait's Air Quality Monitoring Stations. *Sustainability* **2025**, *17*, 9136. [[CrossRef](#)]
57. Romano, Y.; Patterson, E.; Candès, E. Conformalized quantile regression. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 3538–3548.

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.