HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

Development of a Spatiotemporal Hybrid Forecasting Framework with Non-Additive Information Fusion for Urban Air Quality Assessment in Hungary and Morocco.

Thesis of the doctoral (PhD) dissertation

Bouzghiba Houria

Gödöllő, Hungary

2026

Name of Doctoral school:	Doctoral School of Natural Sciences
Name of Program:	Environmental Science Program
Discipline:	Natural Sciences / Environmental Sciences
Head of Doctoral School	Dr. Erika Csákiné Michéli, MHAS Member of Hungarian Academy of Sciences Hungarian University of Agriculture and Life Science Gödöllő Hungary
Head of Program	Dr. András Székács, DSc Institute of Environmental Sciences Hungarian University of Agriculture and Life Science, Gödöllő Hungary
Supervisor	Dr. Gábor Géczi, PhD Institute of Environmental Sciences Hungarian University of Agriculture and Life Science, Gödöllő, Hungary
Co-Supervisor	Dr. Khomsi Kenza PhD Partnership for Action on Green Economy (PAGE) Morocco

....

Affirmation of head of school Affirmation of supervisor

Content

1. BACKGROUND AND JUSTIFICATION

Urban air pollution constitutes one of the most critical environmental challenges of the 21st century, affecting 4.2 billion people globally and causing an estimated 7 million premature deaths annually according to the World Health Organization's 2023 assessment. The complexity of urban air pollution emerges from the intricate interaction between diverse emission sources and atmospheric processes. Transportation systems contribute 40-70% of nitrogen dioxide concentrations in urban areas, while residential heating can account for up to 50% of particulate matter during winter months. Regional transport adds another layer of complexity, contributing 30-60% of fine particulate matter concentrations. These sources interact with urban morphology to create microscale variations that can differ by 200-300% within distances as short as 50-100 meters, presenting unprecedented challenges for accurate forecasting.

Central and Eastern Europe faces particularly acute air quality challenges that stem from a unique combination of factors. The region's industrial legacy, rapid economic transition following 1990, and challenging geographical characteristics create a perfect storm for air pollution problems. Despite representing only 21% of the European Union's population, CEE countries account for a staggering 67% of $PM_{2.5}$-related premature deaths according to the European Environment Agency's 2023 report. The aging vehicle fleet, averaging 14 years compared to the EU-wide average of 8.5 years, means that 40% of vehicles meet only Euro 3 emission standards or below. Furthermore, between 2.5 and 3.5 million households in the region still depend on solid fuel heating systems that produce emission factors 50-100 times higher than modern gas boilers. These factors combine to impose substantial economic costs equivalent to 2.5-4.6% of GDP annually through healthcare expenditures, lost productivity, and accelerated infrastructure degradation.

Budapest exemplifies these regional challenges while presenting unique local complications. The city's 1.7 million inhabitants experience PM_{10} exceedances on 25-40 days annually at traffic monitoring stations, significantly impacting public health. The city's distinctive topography, bisected by the Danube River with elevation differences of 200-500 meters between the Buda hills and the Pest plain, creates frequent temperature inversions. These inversions occur during 120-150 days annually, effectively trapping pollutants below 200-300 meters altitude and preventing their dispersion. Winter PM_{10} levels exceed summer concentrations by 50-80%, driven by the combined effects of residential heating from approximately 80,000-100,000 households still using solid fuels, dramatically reduced atmospheric

mixing heights of 100-200 meters compared to summer values of 1000-2000 meters, and lower average wind speeds of 2-3 m/s that further limit pollutant dispersion.

Current air quality forecasting methodologies face fundamental limitations that remain inadequately addressed, particularly in the Central and Eastern European context. Chemical Transport Models, while physically comprehensive, require computational resources that effectively limit their operational resolution to kilometres when increasingly street-level predictions are demanded for urban management. These models also depend on detailed emissions inventories that are often unavailable, outdated, or of questionable quality in rapidly developing urban areas. Machine learning approaches, though computationally efficient, demonstrate significantly degraded performance when applied outside their training domains and require extensive historical data that sparse monitoring networks cannot provide. Most critically, a persistent and problematic gap exists between model accuracy and interpretability. Deep learning architectures achieve superior predictive performance but operate as impenetrable black boxes that regulators cannot interrogate or trust for decision-making, while interpretable methods consistently fail to capture the non-linear dynamics that characterize atmospheric systems.

The literature reveals significant geographical bias in air quality forecasting research, with comprehensive Central and Eastern European studies remaining remarkably sparse despite the region's unique and severe challenges. Existing studies typically focus on short-term measurement campaigns rather than developing operational systems adapted to local conditions. Furthermore, the practical challenges of operational deployment – including real-time data quality control, computational constraints for timely forecast delivery, and integration with urban decision-making infrastructure – are rarely addressed in academic literature, creating a disconnect between research outputs and practical implementation needs.

2. OBJECTIVES

The primary aim of this doctoral research is to develop accurate and interpretable air quality forecasting methods for urban environments that successfully bridge the critical gap between predictive performance and operational utility, with specific application to Central and Eastern European cities facing unique atmospheric and infrastructural challenges.

This overarching goal is pursued through five specific objectives that systematically address different aspects of the forecasting challenge:

Objective 1: To develop and validate autoregressive integrated moving average (ARIMA) models for short-term air quality prediction across multiple temporal granularities (1-hour, 3-hour, 12-hour), establishing performance baselines and identifying temporal patterns in urban pollution dynamics. This objective addresses the fundamental need for computationally efficient statistical methods that can operate with limited computational resources while providing interpretable parameters that reveal underlying temporal structures in pollution data.

Objective 2: To investigate the integration of chemical transport model outputs with artificial neural networks, developing hybrid architectures that maintain physical consistency while learning systematic bias corrections. This objective explores the potential for combining the interpretability and physical grounding of deterministic models with the flexibility and pattern recognition capabilities of machine learning, addressing limitations identified in purely deterministic or purely statistical approaches.

Objective 3: To systematically investigate how horizontal mesh grid resolution influences the entire air quality modelling pipeline, from meteorological inputs through emission processing to final concentration predictions. This objective examines whether higher spatial resolution uniformly improves model performance or whether optimal resolution varies with atmospheric conditions, pollutant species, and urban characteristics, providing guidance for operational implementation.

Objective 4: To systematically evaluate the contribution of domain-informed feature engineering versus algorithmic complexity in air quality prediction, developing specialized feature sets that capture distinct atmospheric processes including short-term dynamics, long-term patterns, meteorological drivers, and anomaly detection. This objective tests the hypothesis that appropriate feature engineering can overcome algorithmic limitations more effectively than architectural sophistication.

Objective 5: To implement and evaluate fuzzy aggregation operators, specifically the Choquet integral, for combining multiple model predictions while maintaining

mathematical interpretability of the fusion process. This objective addresses the critical gap between black-box ensemble methods that achieve high accuracy and simple averaging that maintains transparency but ignores model interactions.

3. METHODS

3.1 Study Design and Locations

This research employs a strategically designed multi-methodological approach applied across different urban environments to comprehensively explore air quality forecasting techniques. Rather than attempting to test the transferability of a single method across all locations, the study deliberately applies different forecasting approaches to cities with distinct characteristics, demonstrating the evolution and complementary nature of various methodologies in addressing urban air quality challenges. The methodological progression follows a logical development from simple to complex: statistical baseline establishment using ARIMA models in Budapest demonstrates the capabilities and limitations of purely data-driven time series approaches; deterministic model enhancement through the CHIMERE-ANN hybrid approach in Agadir and Casablanca validates the effectiveness of combining physical models with neural networks for bias correction; and advanced ensemble fusion using machine learning models combined through Choquet Integral fusion in Budapest represents the state-of-the-art in interpretable ensemble methods.

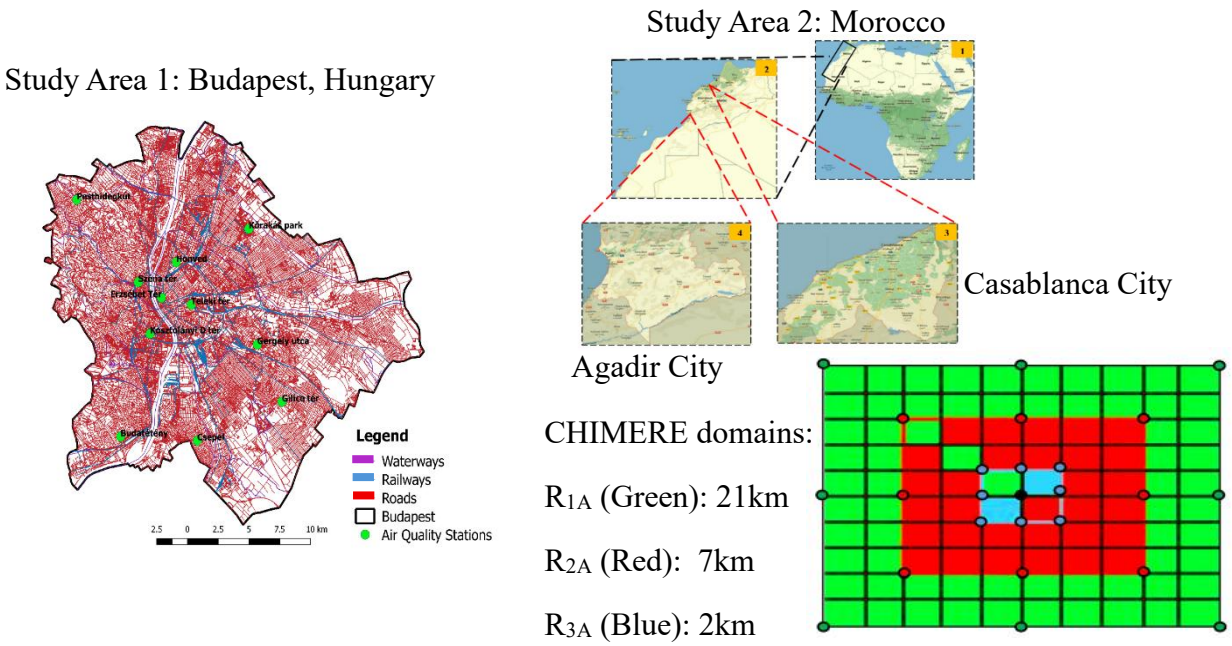


Figure 1. Study area used in the thesis.

3.2 Data Collection and Processing

The Budapest atmospheric monitoring network provided the primary dataset for statistical and ensemble methods development. Data were collected from 11 stations strategically distributed across the metropolitan area, classified according to European Environment Agency guidelines into urban traffic stations (within 10 meters of major traffic arteries), urban background stations (residential areas), one urban industrial station, and suburban background stations. Continuous hourly measurements spanning January 1, 2018, to December 31, 2023, encompassed PM_{10} and NO_2 concentrations alongside comprehensive meteorological parameters including temperature, relative humidity, wind speed and direction, atmospheric pressure, and global solar radiation.

Data quality assurance followed EEA objectives with a three-tier preprocessing pipeline. Primary validation eliminated physically impossible values, statistical screening employed outlier detection using rolling statistics and rate-of-change analysis, and temporal consistency was verified using the Pettitt test. Missing data, ranging from 0.74% to 13.07% across stations, were imputed using an Expectation-Maximization algorithm that incorporated spatial-temporal structure through lagged values, neighbouring station measurements, and meteorological covariates. The algorithm achieved convergence with log-likelihood changes below 10^{-6}, typically within 20-30 iterations, with validation demonstrating mean absolute errors of 3.2 $\mu g/m^3$ for PM_{10} and 4.1 $\mu g/m^3$ for NO_2.

The Moroccan cities were used as second study area due to technical issues with the CTM. Agadir's coastal location with 600,000 inhabitants offered ideal conditions for testing CHIMERE model sensitivity to spatial resolution, with Atlantic maritime influence and relatively simple topography creating stable atmospheric conditions. Data from March 1 to August 31 for both 2010 and 2016 captured spring and summer seasonal variations. Casablanca, Morocco's economic capital with 3.4 million inhabitants, served as validation for the CHIMERE-ANN hybrid approach, with intensive industrial activity and dense traffic networks creating complex pollution patterns suitable for testing neural network bias correction capabilities.

3.3 Statistical Baseline with ARIMA

The ARIMA methodology implementation followed the comprehensive Box-Jenkins approach for time series analysis. Stationarity assessment employed complementary Augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin tests, with ADF testing the null hypothesis of non-stationarity while KPSS tested stationarity as the null hypothesis. This dual-test approach ensured robust identification of differencing requirements across temporal aggregations. Model identification utilized

autocorrelation and partial autocorrelation functions to determine appropriate orders for autoregressive (p), differencing (d), and moving average (q) components.

A systematic grid search evaluated candidate models across reasonable parameter ranges: $p,q \in \{0,1,2,3,4,5\}$ for non-seasonal components and $P,Q \in \{0,1,2\}$ for seasonal components, maintaining computational tractability while capturing relevant temporal structures. Model selection balanced goodness-of-fit against complexity using the Akaike Information Criterion and Bayesian Information Criterion. The investigation specifically examined three temporal granularities – 1-hour, 3-hour, and 12-hour aggregations – to identify the optimal balance between temporal resolution and noise reduction. Parameter estimation employed maximum likelihood methods with sophisticated optimization algorithms to handle the non-linear likelihood surface.

3.4 Deterministic Modelling and Enhancement

The CHIMERE chemistry-transport model implementation required comprehensive configuration across multiple components. The Weather Research and Forecasting model provided meteorological forcing with physics parameterizations including the RRTM longwave radiation scheme for accurate radiative transfer, the Dudhia shortwave scheme for solar radiation, the YSU planetary boundary layer scheme for turbulent mixing, and the Kain-Fritsch cumulus parameterization for convective processes. Emissions processing utilized the Emi-Surf model with EDGAR HTAP anthropogenic emission inventories providing sector-specific emission data, while the MEGAN model calculated biogenic emissions based on vegetation type and meteorological conditions. The MELCHIOR2 reduced chemical mechanism represented atmospheric chemistry through approximately 120 reactions involving 44 species, balancing computational efficiency with chemical completeness.

The spatial resolution investigation systematically evaluated model performance at three horizontal resolutions: 0.1° (approximately 10 km), 0.05° (5 km), and 0.02° (2 km). This multi-resolution approach enabled decomposition of resolution effects across different model components. The analysis tracked how resolution refinement affected meteorological field accuracy, particularly temperature and wind speed predictions; planetary boundary layer height representation critical for vertical mixing; land use classification accuracy affecting surface properties; emission spatial allocation determining source strength distribution; and ultimately the combined impact on ozone and particulate matter concentration predictions.

The CHIMERE-ANN hybrid development addressed systematic model biases through neural network post-processing. The artificial neural network employed a multilayer perceptron architecture with four hidden layers of ten neurons each, chosen to balance representational capacity with training stability. The Levenberg-Marquardt algorithm optimized network weights to minimize sum of squares error, providing faster convergence than standard gradient descent. Sigmoid activation functions in hidden and output layers enabled non-linear mapping of CHIMERE predictions to observe concentrations. Training utilized 87-90% of available data, with the network learning systematic bias patterns. The investigation specifically examined how neural network correction effectiveness varied with CHIMERE input spatial resolution. The technical concept of the study represented in figure 2.

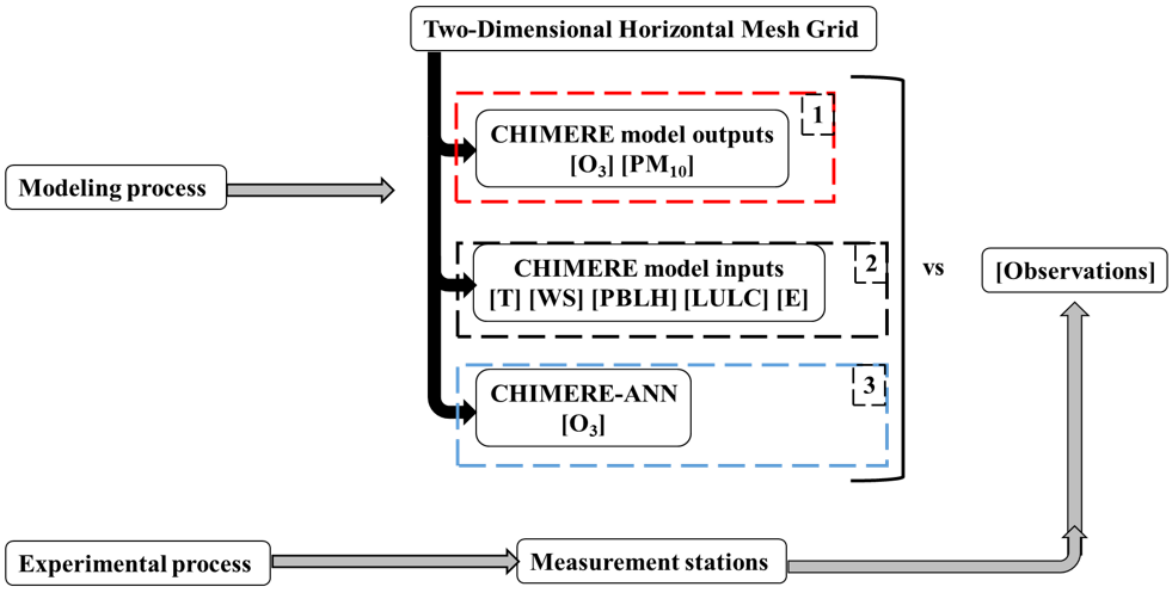


Figure 2. Technical concept of the study, Aspect 1 (Block 1), Aspect 2 (Block 2), Aspect 3 (Block 3).

3.5 Machine Learning and Feature Engineering

The machine learning framework implemented comprehensive feature engineering to capture multi-scale atmospheric processes. Four complementary feature sets were systematically developed. Short-term dynamics features comprised 11 variables including PM_{10} lags at t-1, t-2, and t-3 hours, temporal differences, wind components with directional encoding, and hourly cyclical patterns. Long-term pattern features incorporated 12 variables including PM_{10} lags at 24, 48, and 168 hours, rolling statistics with 24 and 168-hour windows, monthly seasonality encoding, and baseline meteorological conditions. Meteorological driver features emphasized dispersion mechanisms through 14 variables including contemporaneous and lagged

meteorological measurements, decomposed wind vectors, and a pressure proxy indicator for atmospheric stability. Anomaly detection features quantified deviations from expected patterns through 5 specialized indicators including standardized z-scores, deviations from periodic patterns, and binary flags for unusual conditions, the process flow is presented in figure 3.

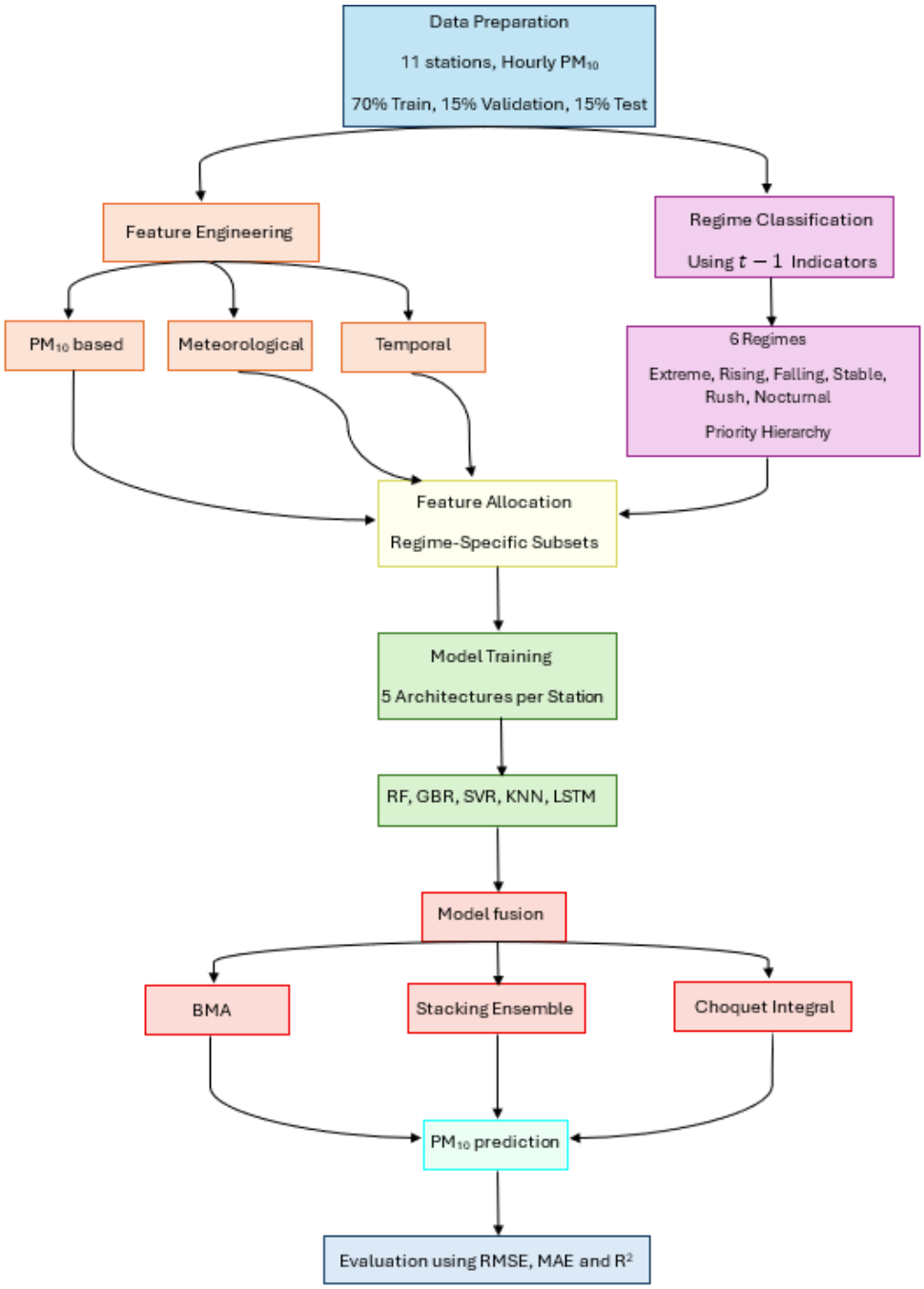


Figure 3. Process flow of the study design.

Eleven machine learning models spanning five architectural families were implemented with careful hyperparameter optimization. Random Forest models included standard configuration with 400 trees and adaptive depth, plus asymmetric loss variants for under-prediction and over-prediction aversion. Gradient Boosting employed both conservative architecture for stable conditions (500 trees, learning rate 0.03) and aggressive configuration for regime-specific patterns (300 trees, learning rate 0.05). Support Vector Regression with RBF kernels targeted meteorological feature sets with regularization parameter C=10.0. K-Nearest Neighbours with k=15 and distance weighting specialized in anomaly detection. Long Short-Term Memory networks explored multiple architectures with varying lookback windows (12-168 hours) and hidden units (64-128) to capture temporal dependencies at different scales.

3.6 Ensemble Fusion Framework

The Choquet integral fusion framework represented the methodological culmination, providing mathematically principled aggregation of expert predictions. The 2-additive Choquet integral formulation-maintained tractability while capturing both individual model importance through singleton Möbius masses and pairwise interactions through pair coefficients. This restriction to 2-additivity reduced the number of parameters from 2^M to M + M(M-1)/2, making optimization feasible while retaining the ability to model synergies and redundancies.

Optimization minimized mean squared error on calibration data comprising 30% of the dataset, subject to monotonicity constraints ensuring that adding experts never decreases the fuzzy measure, and normalization constraints maintaining probabilistic interpretation. Two optimization strategies were employed: COBYLA providing derivative-free constrained optimization suitable for the non-smooth objective function, and Differential Evolution offering global search capabilities to avoid local minima. The investigation systematically evaluated ensemble sizes from K=3 to K=13 experts to identify optimal model diversity, with expert selection based on calibration performance to balance quality with complementarity.

4.1 Statistical Baseline Performance

The ARIMA modelling established critical performance baselines for Budapest's air quality forecasting. Across the 11 monitoring stations, the models achieved mean absolute errors ranging from 11.08 to 23.62 μg/m³ for hourly NO_2 predictions and 4.77 to 7.57 μg/m³ for 3-hourly PM_{10} forecasts (Table 1 and 2). The systematic evaluation across temporal granularities revealed that 3-hour aggregation provided optimal performance, with Mean Absolute Scaled Error values consistently exceeding 0.90, indicating superior forecast accuracy compared to both finer 1-hour and coarser 12-hour resolutions. This optimal aggregation scale represents a critical balance where noise reduction through temporal averaging does not sacrifice essential dynamic information about pollution variations.

Table 1. Statistical metrics of NO_2 for the chosen parameters p, d, q for 1h based on the search grid.

NO_2	MAE 1h	RMSE 1h	MASE 1h	p	d	q
Erzsébet square	11.08	14.43	0.91	4	0	4
Budatétény	17.56	20.15	0.92	3	1	5
Csepel	15.56	19.83	0.87	4	0	4
Honvéd	14.23	18.47	0.88	3	1	5
Gilice square	17.65	0.24	0.94	3	0	3
Gergely street	14.48	18.76	0.86	3	0	4
Széna square	16.98	0.19	0.97	3	0	2
Teleki square	15.70	0.20	0.93	3	0	4
Pesthidegkút	21.19	0.29	0.98	5	0	5
Kőrakás park	21.14	0.30	0.96	3	0	5
Kosztolányi D. square	23.62	0.25	0.98	4	0	5

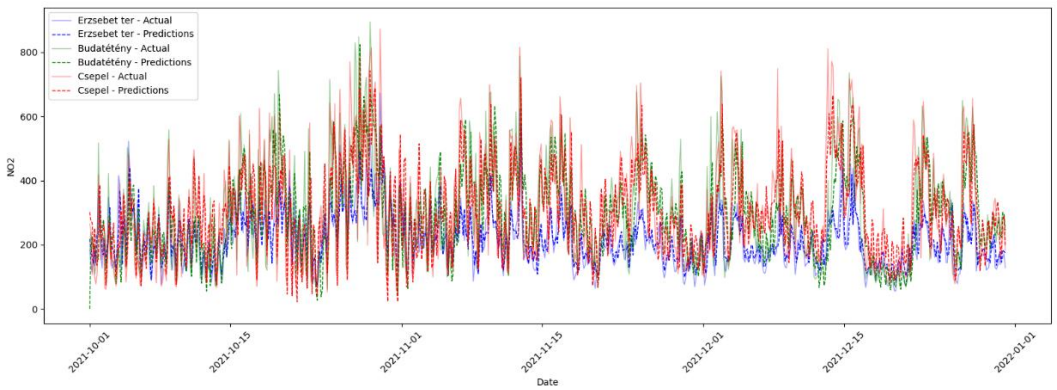


Figure 4. ARIMA model with predictions for NO_2, 1h granularity, based on grid search and ACF, PACF lags consideration.

Station-specific analysis revealed systematic patterns in optimal ARIMA configurations that reflected underlying emission and dispersion characteristics (figure 4 and 5). Traffic stations consistently required higher autoregressive orders (p = 3-5), capturing the persistence of local vehicular emissions, while background stations showed stronger moving average components (q = 4-6), reflecting the influence of transported pollution from multiple sources. The seasonal SARIMA extensions successfully captured diurnal cycles in hourly data and weekly patterns in daily aggregations, with SARIMA$(1,0,1)(1,1,1)_{24}$ proving optimal for several urban traffic locations. Despite achieving reasonable short-term forecast accuracy, the ARIMA models captured only 80-85% of total pollutant variance, with the remaining variance attributed to non-linear atmospheric processes and unprecedented meteorological conditions that statistical extrapolation cannot represent.

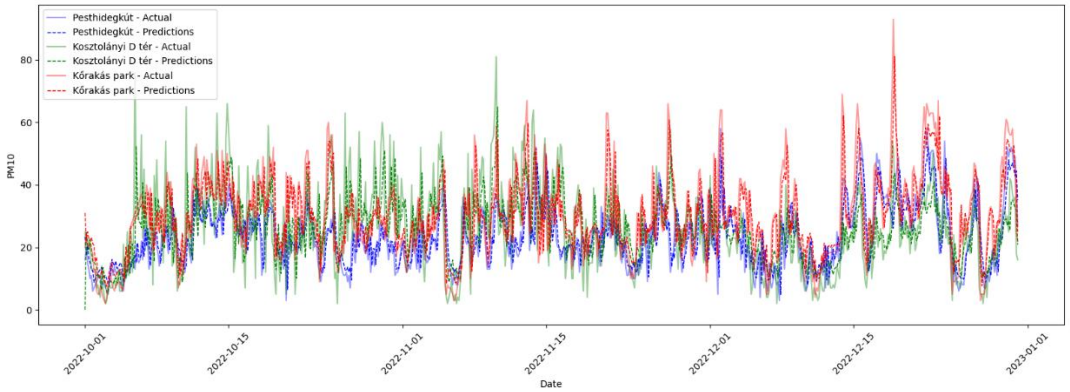


Figure 5. ARIMA model with predictions for PM_{10}, 3h granularity, based on grid search and ACF, PACF lags consideration.

Table 2. Statistical metrics of PM_{10} for the chosen parameters p, d, q for 3h based on the search grid.

PM_{10}	**MAE 3h**	**RMSE 3h**	**MASE 3h**	**p**	**d**	**q**
Erzsébet square	4.75	7.06	0.96	6	0	5
Budatétény	5.13	7.12	0.96	6	0	6
Csepel	6.95	9.77	0.90	6	0	6
Honvéd	5.50	7.48	0.98	5	0	6
Gilice square	7.57	10.63	0.93	5	0	5
Gergely street	5.10	7.20	0.98	4	0	4
Széna square	5.91	8.56	0.92	4	0	4
Teleki square	5.85	8.00	0.94	4	0	3
Pesthidegkút	4.77	6.81	0.99	4	0	4
Kőrakás park	5.98	8.43	0.96	3	0	3
Kosztolányi D. square	6.79	9.48	0.93	4	1	3

4.2 Deterministic Model Enhancement

The CHIMERE model evaluation revealed significant sensitivity to spatial resolution, with systematic improvements observed as horizontal resolution increased from 0.1° (R_{A1}) to 0.02° (R_{A3}). Overall model performance improved by approximately 20% in terms of root mean square error for both ozone and PM_{10}. However, the comprehensive decomposition of resolution effects revealed striking non-uniformity across model components. Emission representation showed the most dramatic improvement, with NO_2 emissions in urban grid cells increasing by up to 10-fold at 0.02° resolution compared to 0.1°, reflecting the model's enhanced ability to distinguish urban sources from rural background. Land use classification accuracy improved by approximately 300%, with high-resolution runs correctly identifying urban characteristics at monitoring station locations that coarser resolutions classified as shrubland. Surprisingly, meteorological field accuracy showed negligible improvement of less than 5%, as illustrated in Figure 8, where temperature and wind speed time series for both R_{A1} and R_{A3} resolutions closely overlap throughout the simulation periods, suggesting that weather dynamics operate at scales larger than the resolution refinements tested.

Figure 6 presents the monthly average observed and modelled ozone and PM_{10} concentrations in Agadir for spring (March) and summer (August) periods in 2010 and 2016. The time series reveal a consistent pattern of ozone overestimation across all resolutions, with modelled values (coloured lines) systematically exceeding observations (black line with markers) throughout both seasonal periods. For PM_{10}, the opposite bias is evident, with CHIMERE underestimating observed concentrations, particularly during peak pollution episodes. The figures demonstrate that while higher resolution (R_{A3}) slightly improves the temporal dynamics and captures some variability that coarser resolutions miss, the systematic bias persists regardless of spatial resolution, motivating the need for hybrid correction approaches.

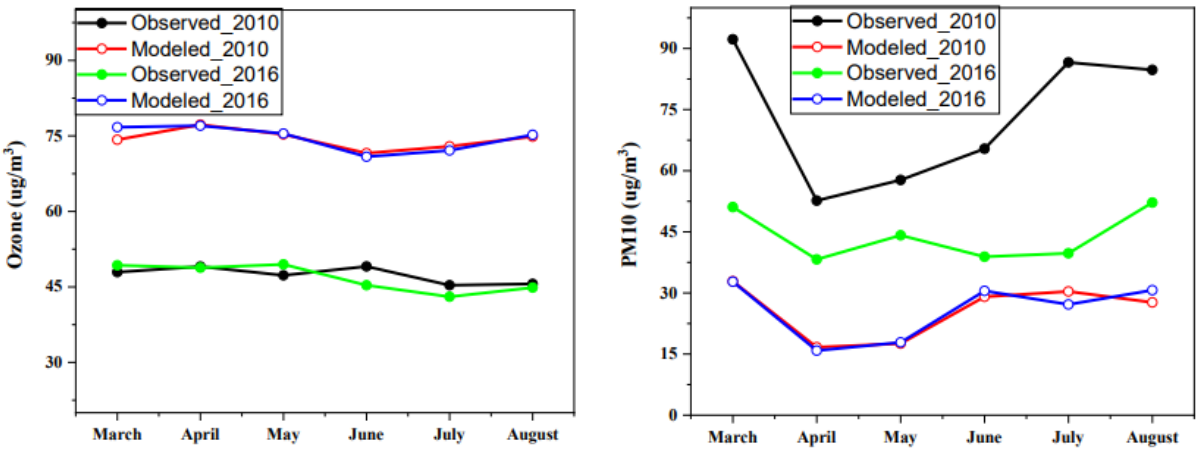
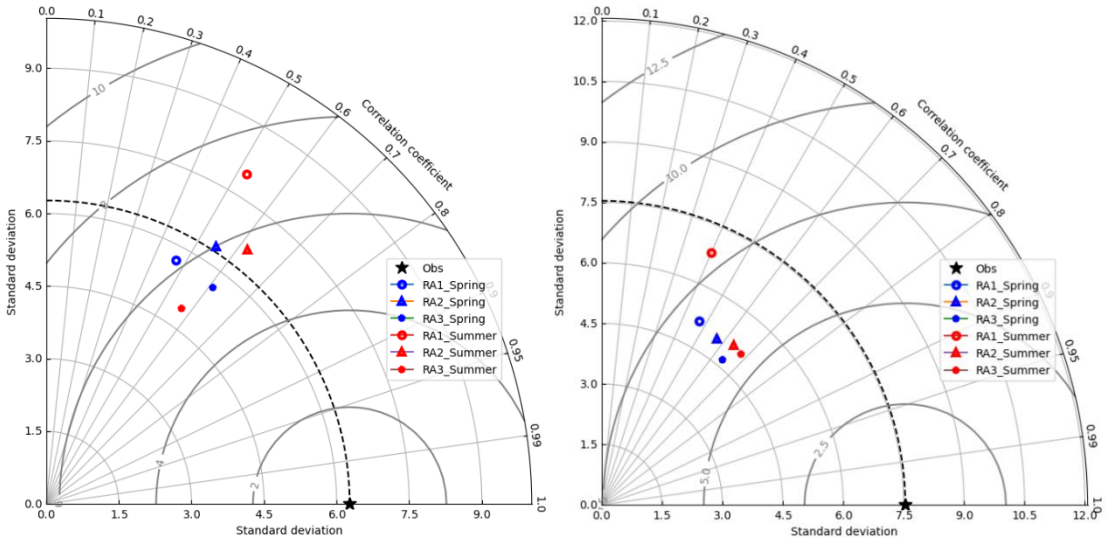


Figure 6. Monthly average observed and modelled ozone and PM_{10} data in Agadir City on March 1st and August 31st, 2010, and 2016 respectively.

The systematic overestimation of ozone concentrations by 27.4 µg/m³ and underestimation of The Taylor diagrams presented in Figure 7 provide a comprehensive visual summary of model performance across resolutions for both pollutants in 2010 and 2016. For ozone, the progression from R_{A1} to R_{A3} shows a clear shift toward the observation reference point, with correlation coefficients increasing and normalized standard deviation ratios approaching unity. The 2016 ozone simulations demonstrate particularly notable improvement, with R_{A3} achieving substantially higher correlation than R_{A1} while maintaining comparable variability to observations. For PM_{10}, the Taylor diagrams reveal more modest improvements with resolution refinement, consistent with the finding that particulate matter predictions are more strongly influenced by emission inventory uncertainties than by spatial resolution alone. The clustering of different resolution runs in PM_{10} Taylor diagrams, compared to the more dispersed pattern for ozone, confirms that resolution sensitivity varies by pollutant species.

2010 2016

Ozone (O_3)



Particulate matter (PM_{10})

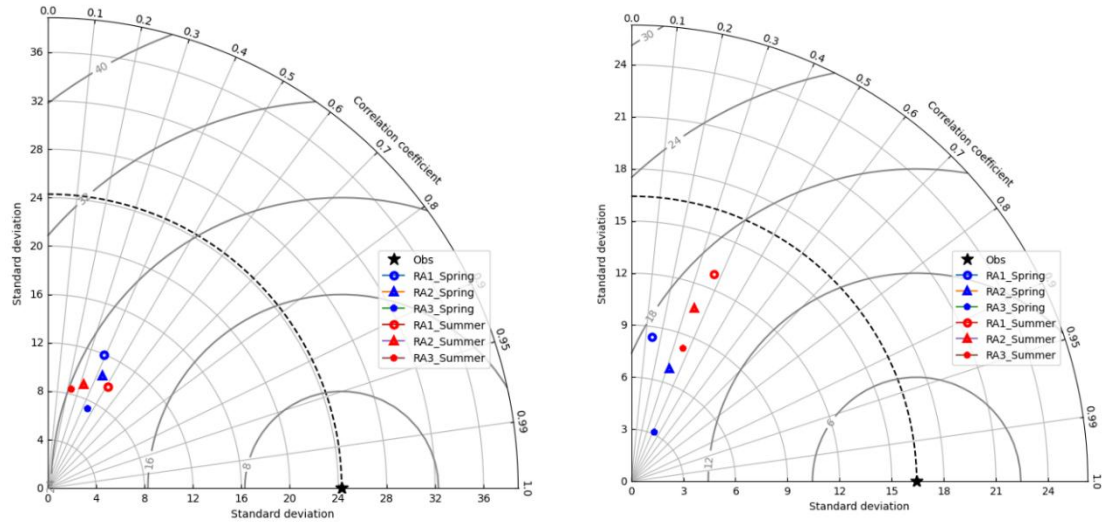


Figure 7. Ozone and Particulate matter Taylor diagram for 2010 and 2016, in Agadir city using R_{1A}, R_{2A}, and R_{A3}.

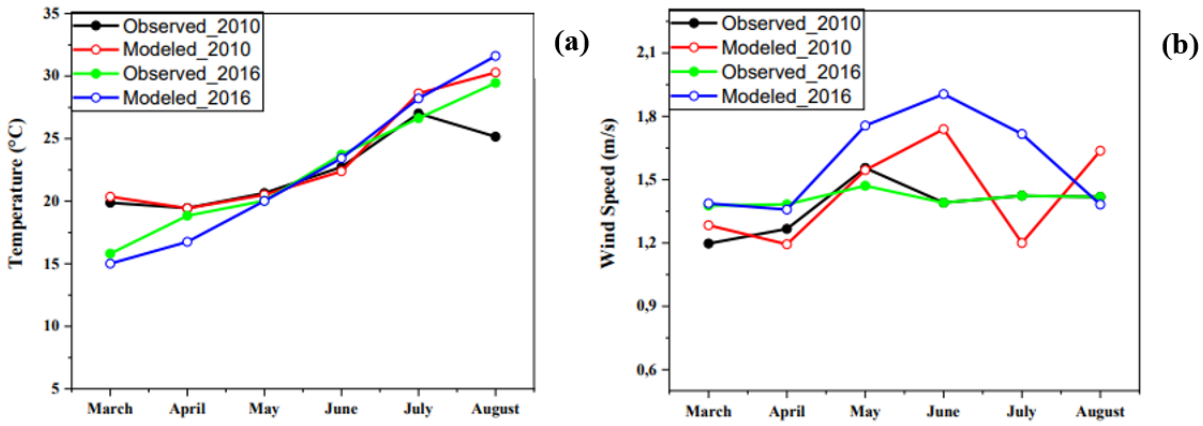


Figure 8. Monthly average observed and modelled Temperature (a) and Wind Speed (b) in Agadir city on March 1 and August 31, 2010, and 2016.

The systematic overestimation of ozone concentrations by 27.4 µg/m³ and underestimation of PM_{10} by 32.8 µg/m³ persisted even at the finest resolution, indicating fundamental model limitations beyond spatial discretization. These systematic biases motivated the development of the CHIMERE-ANN hybrid approach. As quantified in Table 3, the neural network correction proved remarkably effective across both cities and resolution configurations. For Agadir, the CHIMERE-ANN hybrid reduced ozone RMSE from 30.1 µg/m³ (CHIMERE R_{A1}) to 7.5 µg/m³ (CHIMERE-ANN R_{A3}), representing a 75% improvement in prediction accuracy. The mean modelled concentration shifted from 76.7 µg/m³ (a substantial overestimation) to 50.1 µg/m³, closely matching the observed mean of 49.2 µg/m³. Correlation improved dramatically from 40% with CHIMERE RA1 to 80% with CHIMERE-ANN R_{A3}, bringing performance within the literature-reported ranges of 80-91% correlation and approaching the benchmark RMSE values of 0.02-20 µg/m³ reported in comparable studies.

The results from Casablanca, also presented in Table 3, confirm the generalizability of the hybrid approach. CHIMERE-ANN achieved even stronger performance in Casablanca, with correlation reaching 94% at R_{C3} resolution compared to 84% for standalone CHIMERE. The RMSE reduction from 30.8 µg/m³ (CHIMERE R_{C3}) to 7.4 µg/m³ (CHIMERE-ANN R_{C3}) demonstrates that the neural network correction is equally effective across different urban environments. Notably, the improvement from using high-resolution inputs is consistent across both cities: CHIMERE-ANN with R_{A3}/R_{C3} inputs outperforms CHIMERE-ANN with R_{A1}/R_{C1} inputs, with correlations improving from 58% to 80% in Agadir and from 88% to 94% in Casablanca. This pattern demonstrates that machine learning benefits significantly from physically consistent spatial structures even when absolute concentration

values require substantial correction. The high-resolution CHIMERE outputs provide the neural network with more accurate representations of emission gradients, boundary layer dynamics, and local circulation patterns, enabling more effective bias correction than would be possible with coarser inputs that smooth over critical urban-scale features.

Table 3. Calculated statistical scores for CHIMERE and CHIMERE-ANN compared to validation studies for O_3.

		Mean. Obs [µg/m³]	**Mean. Mod [µg/m³]**	**RMSE [µg/m³]**	**Corre-lation**	**Literature ranges**
Agadir	CHIMERE (R_{A1})	49.2	76.7	30.1	40%	80%< R^2 <91% 0.02<RMSE<20 (Ajdour et al., 2022; H. Lu et al., 2021; Mok et al., 2017; Sayeed et al., 2021; Tchepel et al., 2020)
	CHIMERE (R_{A3})	49.2	76.4	29.1	61%	
	CHIMERE-ANN (R_{A1})	49.2	51.9	8.9	58%	
	CHIMERE-ANN (R_{A3})	49.2	50.1	7.5	80%	
Casablanca	CHIMERE (R_{C1})	31.5	62.0	31.4	74%	
	CHIMERE (R_{C3})	31.5	61.6	30.8	84%	
	CHIMERE-ANN (R_{C1})	31.5	30.93	10.6	88%	
	CHIMERE-ANN (R_{C3})	31.5	29.85	7.4	94%	

4.3 Machine Learning Performance Stratification

The comprehensive evaluation of 11 specialized machine learning models revealed extreme performance stratification that challenges conventional assumptions about algorithm selection, as summarized in Table 4 and visualized in Figure 9. K-Nearest Neighbours with anomaly-detection features achieved exceptional performance with RMSE of 1.800 ± 0.717 µg/m³ and R^2 of 0.979, representing a minimum 60.8% improvement over the average performance of all individual models. This superiority

of a relatively simple algorithm when paired with appropriate features challenges the current emphasis on architectural complexity in atmospheric machine learning. The KNN-Anomaly model outperformed all other approaches by a substantial margin, achieving an MAE of just 1.51 μg/m³ compared to the next best performer (RF-Standard) at 3.80 μg/m³, a difference of more than twofold in absolute error terms.

The comparison between regime-specific and stationary gradient boosting configurations yielded profound insights into the importance of acknowledging atmospheric non-stationarity. The regime-specific model achieved RMSE of 4.601 ± 1.204 μg/m³ compared to 10.821 ± 2.162 μg/m³ for the stable variant, with a paired t-test revealing $t(10) = -13.61$, $p < 0.001$, and Cohen's $d = -4.10$. This effect size of 4.10 standard deviations represents one of the largest documented in atmospheric machine learning literature, quantifying the severe penalty imposed by assuming stationarity in fundamentally non-stationary systems. The stable variant's R^2 of 0.331 versus the regime-specific R^2 of 0.817 indicates that 48.6% of variance explanation is lost when atmospheric regime transitions are ignored. Table 4 starkly illustrates this contrast: GBR-Stable ranks among the worst performers with RMSE of 10.82 μg/m³, while GBR-Regime achieves competitive performance at 4.60 μg/m³, demonstrating that the same algorithmic family can span from near-failure to strong performance depending solely on whether atmospheric regime dynamics are respected.

Support Vector Regression with RBF kernels demonstrated catastrophic failure with R^2 of -0.048, indicating predictions 4.8% worse than using the unconditional mean. As shown in Table 4, SVR-RBF produced the highest RMSE (13.60 μg/m³) and MAE (10.41 μg/m³) of all evaluated models, with negative R^2 confirming complete model collapse. This failure in high-dimensional meteorological feature space exemplifies the curse of dimensionality, where the sample-to-dimension ratio of 0.79 fell below theoretical requirements for kernel convergence. Similarly, LSTM-Multivariate exhibited poor performance with RMSE of 11.72 μg/m³ and R^2 of only 0.183, suggesting that simply adding more input variables to deep learning architectures does not guarantee improved predictions and may instead introduce noise that degrades model performance.

Random Forest variants showed remarkable consistency across different loss function configurations. Table 4 reveals nearly identical performance metrics: RF-Standard (RMSE = 5.85, $R^2 = 0.794$), RF-Underpredict Averse (RMSE = 5.95, $R^2 = 0.787$), and RF-Overpredict Averse (RMSE = 5.97, $R^2 = 0.788$). Statistical analysis confirmed complete invariance to asymmetric loss functions (ANOVA $F = 0.00$, $p = 0.9952$), though subtle ranking differences emerged in non-parametric tests, suggesting that bootstrap aggregation's inherent variance reduction overwhelms

attempts at targeted optimization through loss function modification. The three standard LSTM variants (Short, Long, and Balanced) displayed similar stability, with RMSE values ranging narrowly from 6.02 to 6.10 μg/m³ and R^2 values between 0.777 and 0.782, indicating that lookback window length has minimal impact on prediction accuracy within the tested configurations.

Table 4. Cross-station evaluation metrics of model architectures.

Family	Model	RMSE	MAE	R^2
Gradient Boosting	GBR-Stable	10.82	8.16	0.331
	GBR-Regime	4.6	3.18	0.817
K-Nearest Neighbors	KNN-Anomaly	1.8	1.51	0.979
LSTM Networks	LSTM-Short	6.09	3.88	0.777
	LSTM-Long	6.1	3.94	0.778
	LSTM-Multivariate	11.72	8.71	0.183
	LSTM-Balanced	6.02	3.84	0.782
Random Forest	RF-Standard	5.85	3.8	0.794
	RF-Underpredict Averse	5.95	3.88	0.787
	RF-Overpredict Averse	5.97	3.88	0.788
Support Vector Machine	SVR-RBF	13.6	10.41	-0.048

Figure 9 presents the RMSE distributions across all 11 monitoring stations, providing crucial insight into model consistency and spatial transferability. The box plots reveal that KNN-Anomaly not only achieves the lowest median RMSE but also exhibits the tightest interquartile range, indicating consistent performance across diverse urban environments from suburban background stations to dense traffic locations. In contrast, SVR-RBF and GBR-Stable display both high median RMSE and substantial variability, with whiskers extending to extreme values that reflect inconsistent and unreliable predictions. The LSTM family shows moderate performance with relatively consistent distributions across variants, while Random Forest models cluster together with overlapping distributions that visually confirm their statistical equivalence. Notably, the figure demonstrates that model ranking remains stable across stations—models that perform well at one location generally perform well at others—supporting the generalizability of the findings and the feasibility of deploying a unified modelling framework across heterogeneous urban monitoring networks.

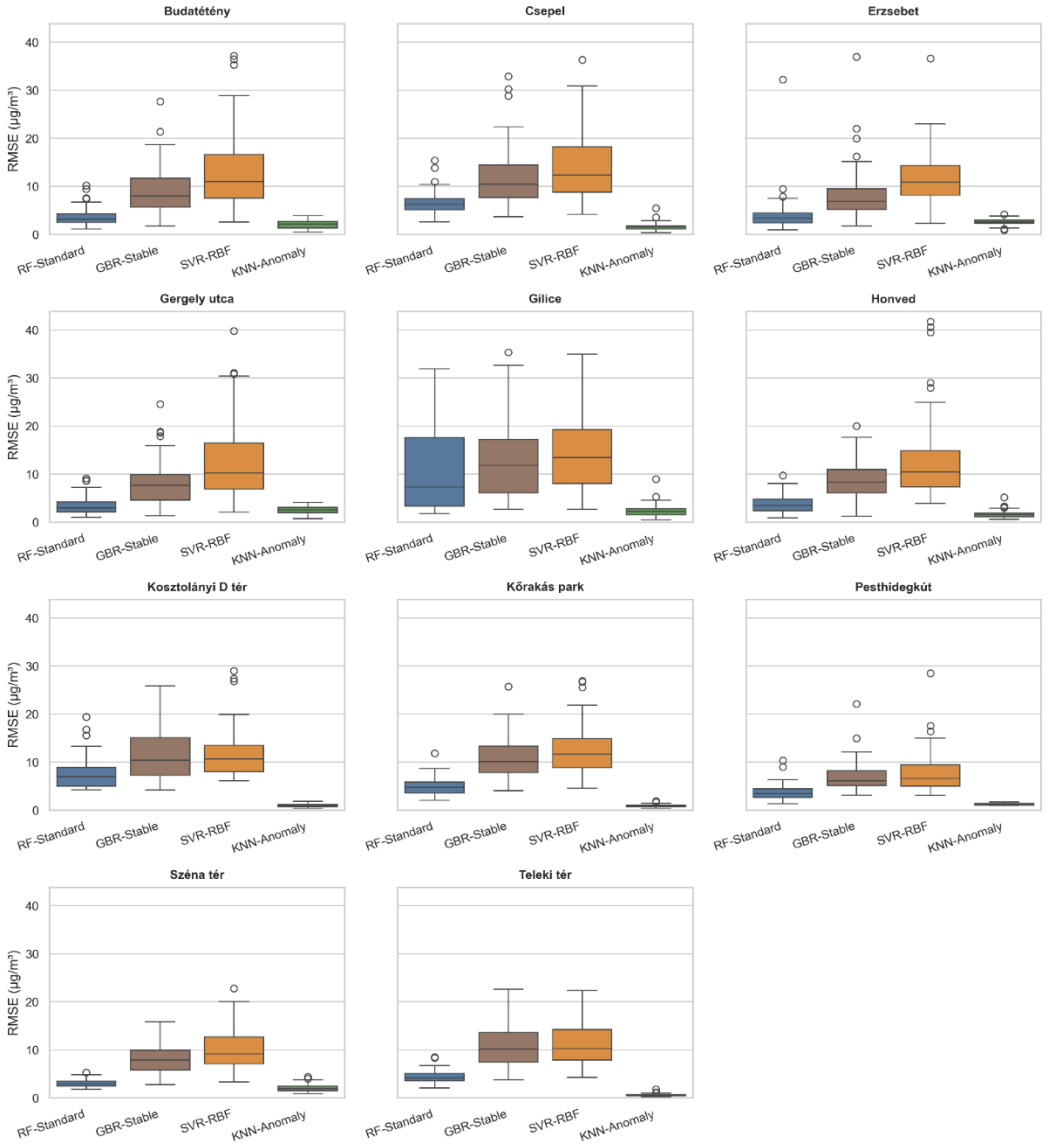
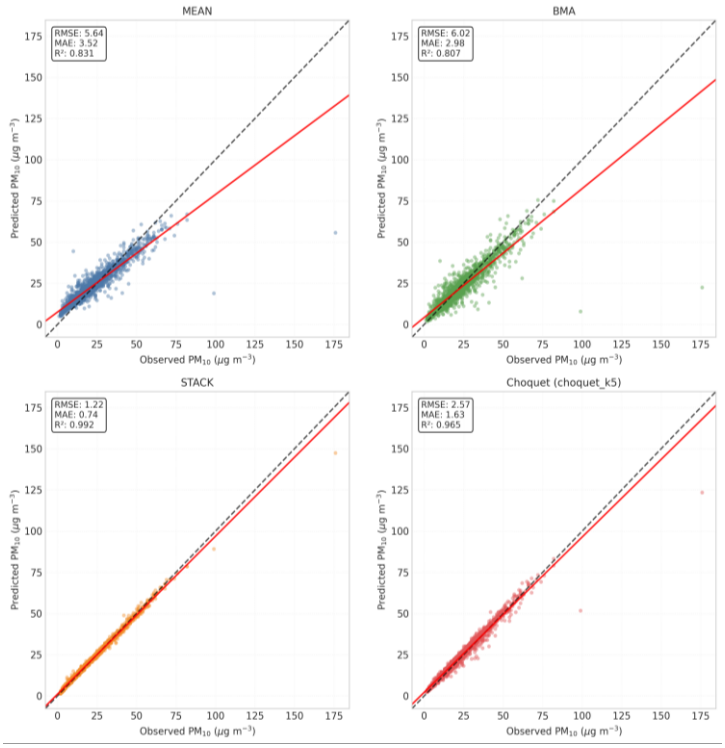
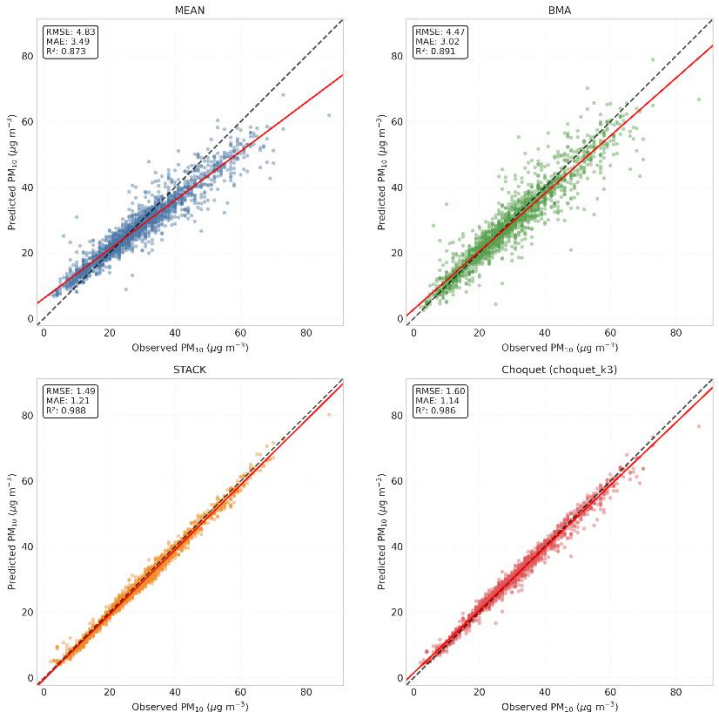


Figure 9. RMSE distributions by feature-set expert (PM_{10}) across stations.



(a)



(b)

Figure 10. Scatter plots of different fusion methods in 2 stations: (a) Erszebet square station; (b) Honved station.

4.4 Ensemble Fusion and the Choquet Integral

The Choquet integral fusion framework achieved remarkable performance while maintaining complete mathematical interpretability, addressing the fundamental trade-off that has long plagued ensemble methods. With optimal configuration at K=5 experts, the Choquet integral achieved RMSE of 1.825 ± 0.392 μg/m³ and R^2 of 0.983 ± 0.011, representing performance within 9.6% of state-of-the-art black-box stacking methods that achieved RMSE of 1.669 ± 0.367 μg/m³. This marginal performance difference falls well within typical PM_{10} sensor measurement uncertainty of ±10-15%, while the Choquet integral provides complete transparency through interpretable Möbius coefficients and Shapley values.

The scatter plots presented in Figure 10 illustrate the comparative performance of different fusion methods at two contrasting stations. At Erzsébet square station (Figure 24a), a dense urban traffic location characterised by high pollution variability, all fusion methods demonstrate strong predictive capability with points clustering tightly along the 1:1 diagonal line. The Choquet-K5 method (shown in green) exhibits particularly tight clustering with minimal scatter, achieving predictions that closely match observed values across the full concentration range from 0 to approximately 80 μg/m³. Notably, the stacking ensemble (red points) and Choquet integral show nearly identical performance, with both methods successfully capturing extreme pollution episodes that simpler approaches tend to underestimate. The Bayesian Model Averaging (BMA, shown in blue) displays slightly wider scatter, particularly at higher concentrations, reflecting its limitation in handling non-linear model interactions. At Honvéd station (Figure 10b), which represents a complex urban environment with irregular emission patterns, similar patterns emerge but with distinct characteristics. The tighter clustering of Choquet-K5 predictions compared to BMA becomes more pronounced, demonstrating the Choquet integral's superior ability to handle stations with complex atmospheric dynamics. Both stations confirm that the Choquet integral achieves performance visually indistinguishable from black-box stacking while maintaining complete interpretability.

The universal optimum at K=5 experts across all 11 monitoring stations as shown in figure 11, regardless of their urban characteristics, suggests fundamental information-theoretic limits in atmospheric prediction. Performance metrics showed marked improvement from K=3 to K=5, with RMSE reducing from 3.150 ± 0.624 to 1.825 ± 0.392 μg/m³, representing a 42.1% error reduction. Further ensemble expansion showed diminishing returns, with K=7 achieving RMSE of 1.841 ± 0.401 μg/m³, and progressive degradation for larger ensembles. The correlation between K and RMSE for K≥5 was $r = 0.94$ ($p = 0.016$), confirming systematic performance decline with excessive model inclusion. This saturation suggests that approximately

five independent information sources exhaust the predictable component of urban PM_{10} variance, with additional models merely recapturing already-encoded patterns. From an operational perspective, this finding provides clear guidance for system deployment: investing in more than five diverse expert models yields no accuracy benefit while increasing computational cost and system complexity.

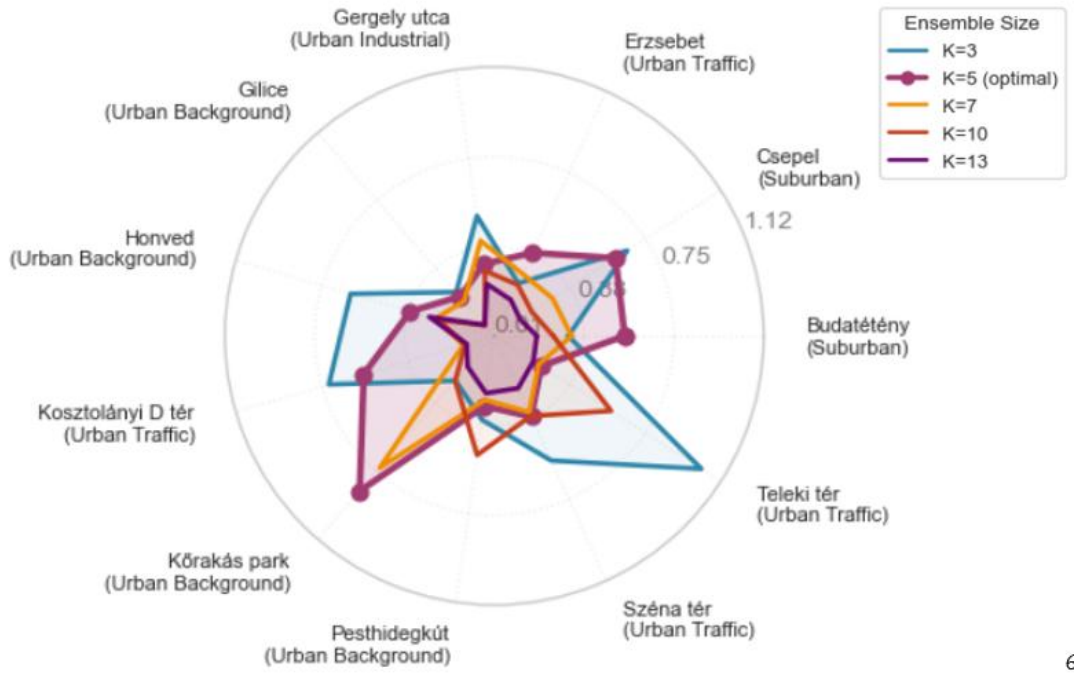


Figure 11. Station-specific sensitivity to ensemble size in Choquet Integral fusion.

The analysis of pairwise interactions revealed that 73.3% exhibited negative Möbius coefficients, indicating redundancy between models (figure 12). This predominance of redundant interactions demonstrates that successful ensemble fusion requires explicit redundancy penalization rather than simple combination. Traditional methods like weighted averaging would treat redundant LSTM variants as independent information sources, effectively multiple counting the same temporal patterns. The Choquet integral's negative interaction coefficients automatically correct this over-representation, explaining its competitive performance despite using fewer parameters than unconstrained methods. The practical implication is significant: in atmospheric prediction where physical constraints force different algorithms toward similar solutions, naive ensemble combination amplifies noise rather than extracting signal, whereas the Choquet integral's mathematically principled approach identifies and appropriately down weights overlapping information.

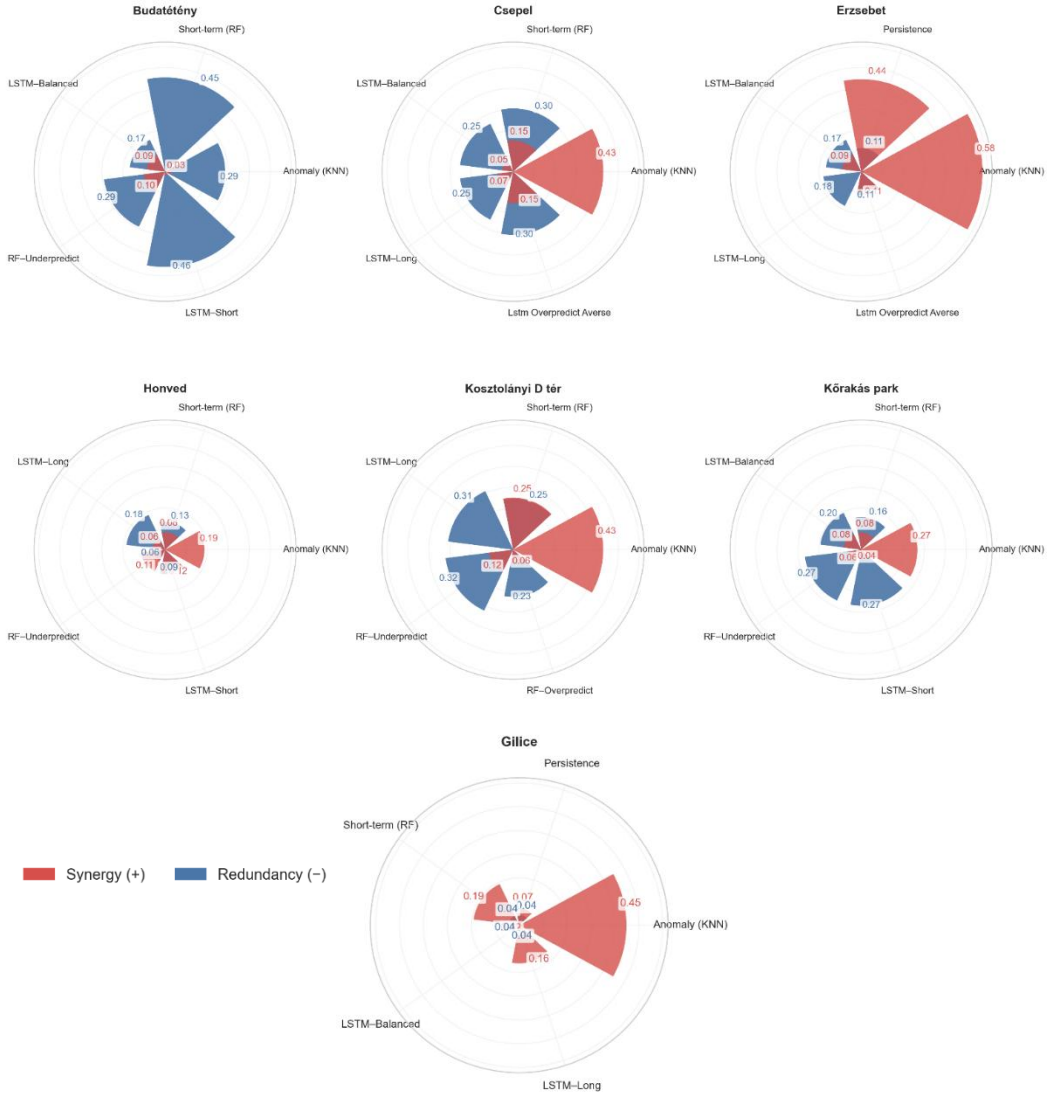


Figure 12. Synergy and redundancy interactions of base models in each station.

5. NEW SCIENTIFIC RESULTS

1. **I applied for the first time the Choquet integral with 2-additive fuzzy measures for air quality ensemble fusion**, achieving performance within 9.6% of state-of-the-art black-box methods (RMSE = 1.83 vs 1.67 μg/m³) while maintaining complete mathematical interpretability through Möbius coefficients and Shapley values.

2. **I demonstrated that ensemble fusion for PM_{10} forecasting exhibits universal information saturation at K=5 expert models** across all 11 monitoring stations regardless of urban characteristics, with performance degrading when additional models are included.

3. **I proved that regime-specific gradient boosting outperforms stationary assumptions by 4.1 standard deviations** (Cohen's $d = -4.10$, $p < 0.001$), representing one of the largest effect sizes documented in atmospheric machine learning literature.

4. **I established that 73.3% of pairwise model interactions in PM_{10} ensemble forecasting exhibit redundancy** (negative Möbius coefficients), proving that successful fusion requires explicit redundancy penalization.

5. **I proved that the CHIMERE-ANN hybrid approach with high-resolution inputs (0.02°) reduces systematic ozone bias by 74%** (from 30.1 to 7.5 μg/m³ RMSE), demonstrating that machine learning correction effectiveness depends critically on physical consistency of input spatial structures.

6. **I demonstrated that anomaly-space modelling provides superior PM_{10} predictions compared to absolute concentration modelling**, with KNN-Anomaly achieving $R^2 = 0.979$ versus $R^2 = -0.048$ for standard meteorological features.

7. **I proved through systematic decomposition that CHIMERE spatial resolution improvements affect model components non-uniformly**: emission representation improves by 10-fold, land use classification by 300%, while meteorological fields show negligible enhancement (<5%).

6. APPLICATION OF RESULTS

The research provides clear operational guidance for cities at different development stages. Cities with limited resources can implement ARIMA models achieving 85% accuracy with minimal infrastructure requirements, requiring only historical time series data and basic computational resources. Cities with moderate resources benefit from the CHIMERE-ANN hybrid approach, which demonstrates that a single well-calibrated monitoring station can enable accurate forecasting when combined with deterministic models and neural network bias correction. Advanced urban areas can deploy the Choquet integral fusion framework, obtaining near-optimal accuracy while maintaining the interpretability required for regulatory compliance and stakeholder communication.

The identification of optimal temporal aggregation at 3-hour intervals provides practical guidance for operational forecasting systems, balancing temporal resolution with noise reduction. The discovery that five diverse expert models suffice for ensemble fusion prevents wasteful over-investment in model development while ensuring adequate diversity. The demonstration that features engineering aligned with atmospheric processes contributes more to model performance than algorithmic sophistication redirects focus from architectural complexity to domain knowledge incorporation.

For regulatory applications, the Choquet integral framework offers a breakthrough solution to the long-standing accuracy-interpretability dilemma. Unlike black-box methods that regulators cannot interrogate, every Choquet integral prediction can be decomposed into individual model contributions and their interactions, enabling stakeholders to understand not just what the forecast predicts but why. This transparency is essential for public trust and regulatory adoption of automated forecasting systems.

7. LIST OF PUBLICATIONS

Publications related to the PhD dissertation

Bouzghiba, H.; Ajdour, A.; Omar, N.; Mendyl, A.; Géczi, G. A Novel Application of Choquet Integral for Multi-Model Fusion in Urban PM_{10} Forecasting. *Atmosphere* 2025, *16*, 1274. **(Q2)** https://doi.org/10.3390/atmos16111274

Houria, B., Abderrahmane, M., Kenza, K., & Gábor, G. (2024). Short-term predictions of PM10 and NO2 concentrations in urban environments based on ARIMA search grid modelling. *CLEAN – Soil, Air, Water*, *52*(6), 2300395. **(Q3)** https://doi.org/10.1002/CLEN.202300395

Bouzghiba, H., Ajdour, A., Abderrahmane, M., & Géczi, G. (2024). Evaluating Two-Dimensional Horizontal Grids in a Deterministic Air Pollution Model: Estimating the Impact on Outputs, Inputs, and ANN-CHIMERE. *EGU24*. https://doi.org/10.5194/EGUSPHERE-EGU24-411

Ajdour, A., Ydir, B., **Bouzghiba, H.**, Sulaymon, I. D., Adnane, A., Hmamou, D. Ben, Khomsi, K., Chaoufi, J., Géczi, G., & Leghrib, R. (2024). Investigating Two-dimensional Horizontal Mesh Grid Effects on the Eulerian Atmospheric Transport Model Using Artificial Neural Network. *Aerosol and Air Quality Research*, *24*(8), 230309. **(Q2)** https://doi.org/10.4209/AAQR.230309

Bouzghiba, H., Ajdour, A., Mendyl, A., Khomsi, K., Weidinger, T., & Géczi, G. (2024). Advancing Air Quality Prediction: An Evaluation of WRF Model Configurations and Their Impact on Meteorological Parameters in Hungary. *ISEE Conference Abstracts*, *2024*(1). https://doi.org/10.1289/ISEE.2024.0767

Sekmoudi, I., Tanarhte, M., **Bouzghiba, H.**, Khomsi, K., Idrissi, L., El jarmouni, M., & Géczi, G. (2024). Systematic Review of Air Pollution in Morocco: Status, Impacts, and Future Directions. *Advanced Sustainable Systems*. **(Q1)** https://doi.org/10.1002/ADSU.202400006

Other publications

B. Houria, A. Ajdour, K. Kenza, and G. Gábor, "HYSPLIT AND K-MEANS CLUSTERING APPLICATION FOR TRAJECTORY ANALYSIS TO DETERMINE SOURCE REGIONS OF SECONDARY INORGANIC AEROSOLS AT HUNGARY'S KECSKEMET BACKGROUND MONITORING STATION," in *ICOSTEE 2024 - BOOK OF ABSTRACTS (International Conference on Science, Technology, Engineering and Economy)*, 2024, pp. 14–14.

Bouzghiba, H., & Géczi, G. (2023). Temporal Analysis and ARIMA Modelling of Air Pollutants in Budapest: A Comprehensive Multi-Station Study. In *Risk Factors of Food Chain 2023 : XXIII. International Conference, September 20-22, 2023 Gyöngyös, Hungary : Book of Abstracts* (pp. 23–23).

Bouzghiba, H., & Géczi, G. (2023). Application of HYSPLIT and K-means clustering in trajectory analysis for identifying source regions of secondary inorganic aerosols at the Kecskemét background monitoring station in Hungary. *HUNGARIAN AGRICULTURAL RESEARCH: ENVIRONMENTAL MANAGEMENT LAND USE BIODIVERSITY*, *33*(3–4), 4–7.

Bouzghiba, H., & Géczi, G. (2022). Comparative Analysis of the Environmental Load of Natural Refrigerant (R290). *Journal of Central European Green Innovation*, *10*(Suppl 1), 135–142. https://doi.org/10.33038/JCEGI.3506

Khomsi, K., Abbas, N., **Bouzghiba, H**., Mendyl, A., Balaw, G., & El Ghazouani, L. (2024). Urban Heat Islands, Heatwaves and Implications for Human Wellbeing: A Comparative Analysis across Middle East and North Africa (MENA). *ISEE Conference Abstracts*, *2024*(1). https://doi.org/10.1289/ISEE.2024.1821

Khomsi, K., **Bouzghiba, H.**, Mendyl, A., Al-Delaimy, A. K., Dahri, A., Saad-Hussein, A., Balaw, G., El Marouani, I., Sekmoudi, I., Adarbaz, M., Khanjani, N., & Abbas, N. (2024). Bridging research-policy gaps: An integrated approach. *Environmental Epidemiology*, *8*(1), e281. https://doi.org/10.1097/EE9.0000000000000281

Mendyl, A., **Bouzghiba, H.**, Tadili, R., & Weidinger, T. (2023). Thermal performance assessment of an indirect solar dryer: A case study of Bananas. *Potravinarstvo Slovak Journal of Food Sciences*, *17*, 550–564. https://doi.org/10.5219/1883

Mendyl, A., Mabasa, B., **Bouzghiba, H.**, & Weidinger, T. (2023). Calibration and Validation of Global Horizontal Irradiance Clear Sky Models against McClear Clear Sky Model in Morocco. *Applied Sciences (Switzerland)*, *13*(1), 320. https://doi.org/10.3390/APP13010320/S1