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<br>Member of Hungarian Academy of Sciences<br>Hungarian University of Agriculture and Life Science<br>Gödöllő Hungary |
| **Head of Program** | Dr. András Székács, DSc<br>Institute of Environmental Sciences<br>Hungarian University of Agriculture and Life Science,<br>Gödöllő Hungary |
| **Supervisor** | Dr. Gábor Géczi, PhD<br>Institute of Environmental Sciences<br>Hungarian University of Agriculture and Life Science,<br>Gödöllő Hungary |
| **Co-Supervisor** | Dr. Khomsi Kenza PhD<br>Partnership for Action on Green Economy (PAGE)<br>Morocco |

..... .....

Affirmation of head of school Affirmation of supervisor

# Table of Contents

TABLE OF FIGURES

LIST OF ABBREVIATIONS

ACF: Autocorrelation Function

ADF: Augmented Dickey-Fuller (test)

AI: Artificial Intelligence

AIC: Akaike Information Criterion

ANN: Artificial Neural Network

AR: AutoRegressive

ARIMA: AutoRegressive Integrated Moving Average

ARIMAX: AutoRegressive Integrated Moving Average with eXogenous variables

ARMA: AutoRegressive Moving Average

AQMEII: Air Quality Model Evaluation International Initiative

BIC: Bayesian Information Criterion

BLH: Boundary Layer Height

BMA: Bayesian Model Averaging

CEE: Central Eastern Europe

CHIMERE: Chemistry-transport model (French: CHImie et MEtéorologie à l'Échelle REgionale)

CMAQ: Community Multiscale Air Quality (model)

CNN: Convolutional Neural Network

CNRS: French National Scientific Research Centre

CO: Carbon Monoxide

COBYLA: Constrained Optimization BY Linear Approximations

CTM: Chemical Transport Model

EEA: European Environment Agency

EDGAR: Emissions Database for Global Atmospheric Research

EM: Expectation-Maximization

Emi-Surf: Emission Surface (preprocessing program)

ERA5: ECMWF Re-Analysis 5th Generation

EU: European Union

FastJX: Fast-J Photolysis code (extended version)

GBR: Gradient Boosting Regressor

GDP: Gross Domestic Product

GFS: Global Forecast System

GOCART: Goddard Chemistry Aerosol Radiation and Transport

HTAP: Hemispheric Transport of Air Pollution

INERIS: French National Institute for Industrial Environment and Risks

KNN: K-Nearest Neighbours

KPSS: Kwiatkowski-Phillips-Schmidt-Shin (test)

LMDz-INCA: Laboratoire de Météorologie Dynamique Zoom - Interaction Chemistry Aerosols

LSTM: Long Short-Term Memory

LULC: Land Use and Land Cover

MA: Moving Average

MAE: Mean Absolute Error

MASE: Mean Absolute Scaled Error

MB: Mean Bias

MEGAN: Model of Emissions of Gases and Aerosols from Nature

MELCHIOR: Chemical mechanism used in CHIMERE

ML: Machine Learning

NWP: Numerical Weather Prediction

PACF: Partial Autocorrelation Function

PBL: Planetary Boundary Layer

PBLH: Planetary Boundary Layer Height

PINN: Physics-Informed Neural Network

PM: Particulate Matter

$R^2$: Coefficient of Determination

RBF: Radial Basis Function

RF: Random Forest

RH: Relative Humidity

RMSE: Root Mean Square Error

RRTM: Rapid Radiation Transfer Model

SARIMA: Seasonal AutoRegressive Integrated Moving Average

SHAP: SHapley Additive exPlanations

SHP: Stable History Period

SMOKE: Sparse Matrix Operator Kernel Emissions

SVR: Support Vector Regression

SVM: Support Vector Machine

T: Temperature

TUV: Tropospheric Ultraviolet and Visible (radiative transfer model)

UCAR: University Corporation for Atmospheric Research

USGS: United States Geological Survey

VOC: Volatile Organic Compound

WHO: World Health Organization

WMO: World Meteorological Organization

WRF: Weather Research and Forecasting (model)

WS: Wind Speed

XGBoost: eXtreme Gradient Boosting

CO: Carbon Monoxide

$NH_3$: Ammonia

$NO_2$: Nitrogen Dioxide

$NO_x$: Nitrogen Oxides ($NO + NO_2$)

$O_3$: Ozone

$PM_{2.5}$: Particulate Matter with diameter ≤ 2.5 µm

$PM_{10}$: Particulate Matter with diameter ≤ 10 µm

PPM_coa: Coarse Primary Particulate Matter (emission variable)

$SO_2$: Sulphur Dioxide

$R_{A1}$ (or $R_{1A}$): Resolution domain 1 for Agadir (0.1° ~ 10 km)

$R_{A2}$ (or $R_{2A}$): Resolution domain 2 for Agadir (0.05° ~ 5 km)

$R_{A3}$ (or $R_{3A}$): Resolution domain 3 for Agadir (0.02° ~ 2 km)

$R_{C1}$: Resolution domain 1 for Casablanca

$R_{C3}$: Resolution domain 3 for Casablanca

$Rd0_{1A}$: WRF domain 1 for Agadir (21 km grid)

$Rd0_{2A}$: WRF domain 2 for Agadir (7 km grid)

$Rd0_{1C}$: WRF domain 1 for Casablanca (21 km grid)

$Rd0_{2C}$: WRF domain 2 for Casablanca (7 km grid)

RF-Standard: Random Forest - Standard configuration

RF-Underpredict Averse: Random Forest with asymmetric weighting penalizing underprediction

RF-Overpredict Averse: Random Forest with asymmetric weighting penalizing overprediction

GBR-Stable: Gradient Boosting Regressor for stable atmospheric conditions

GBR-Regime: Gradient Boosting Regressor for regime-specific modelling

SVR-RBF: Support Vector Regression with Radial Basis Function kernel

KNN-Anomaly: K-Nearest Neighbours trained on anomaly features

LSTM-Short: Long Short-Term Memory network for short-term transitions (12h lookback)

LSTM-Long: Long Short-Term Memory network for long-term patterns (168h lookback)

LSTM-Multivariate: Long Short-Term Memory network with multiple input variables

LSTM-Balanced: Long Short-Term Memory network with balanced configuration (baseline)

Choquet-K5: Choquet Integral fusion using top 5 expert models

Choquet-K13: Choquet Integral fusion using 13 expert models

$\Delta_1$: First-order temporal difference: y(t-1) - y(t-2)

$\Delta_2$: Second-order temporal difference: y(t-2) - y(t-3)

$\Delta_3$: Third-order temporal difference: y(t) - y(t-3)

WD_sin: Sine transformation of wind direction

WD_cos: Cosine transformation of wind direction

hour_sin: Sine encoding of hour (cyclical)

hour_cos: Cosine encoding of hour (cyclical)

Month_sin: Sine encoding of month (seasonal)

Month_cos: Cosine encoding of month (seasonal)

Rolling_mean: Rolling window mean of $PM_{10}$ concentrations

Rolling_std: Rolling window standard deviation of $PM_{10}$ concentrations

T_lag: Lagged temperature values (at 6h, 12h)

RH_lag: Lagged relative humidity values (at 6h, 12h)

WS_lag: Lagged wind speed values (at 6h, 12h)

P_proxy: Pressure proxy indicator

Grad: Global solar radiation ($W/m^2$)

z-score: Standardized deviation

δ_daily: Daily deviation from hourly mean pattern

δ_weekly: Weekly deviation from 168-hour rolling mean

$\sigma_6$: 6-hour rolling standard deviation

$\bar{y}$: Mean value of time series

$\bar{y}_w$: Rolling window mean with window size w

μ_exp: Expanding (cumulative) mean

σ_exp: Expanding (cumulative) standard deviation

ε: Small constant ($10^{-6}$) to prevent division by zero

R stable: Stable regime indicator (low variability)

R rising: Rising concentration regime indicator

R falling: Falling concentration regime indicator

R extreme: Extreme pollution regime indicator (>90th percentile)

R Rush: Morning rush hour regime indicator (7-9 AM)

R nocturnal: Nocturnal low-wind regime indicator

p: Autoregressive order (number of lagged observations)

d: Differencing order (degree of differencing for stationarity)

q: Moving average order (size of moving average window)

P: Seasonal autoregressive order

D: Seasonal differencing order

Q: Seasonal moving average order

s: Seasonal period (e.g., 24 for hourly data with daily cycles)

B: Backshift operator

$\Phi$: Autoregressive coefficients

$\theta$: Moving average coefficients

$\mu$: Fuzzy measure

$m(\{i\})$: Möbius mass for singleton (individual model importance)

$m(\{i,j\})$: Möbius mass for pair (pairwise interaction coefficient)

$\varphi_i$ : Shapley value for model i (average marginal contribution)

K: Number of expert models in ensemble

$N_k(x)$: K-nearest neighbours of observation x.

# 1 INTRODUCTION

Urban air pollution affects 4.2 billion people globally, causing 7 million premature deaths annually (WHO, 2023). The 2021 WHO Air Quality Guidelines substantially tightened recommended limits ($PM_{2.5}$: 5 μg/m³, $PM_{10}$: 15 μg/m³, $NO_2$: 10 μg/m³), reflecting evidence that health impacts occur at concentrations previously considered safe. Urban pollution complexity arises from diverse sources, transportation contributing 40-70% of $NO_2$, residential heating up to 50% of winter $PM_{10}$, and regional transport accounting for 30-60% of $PM_{2.5}$, interacting with urban morphology that creates microscale variations of 200-300% within 50-100 meters. Central Eastern Europe faces particularly acute challenges stemming from historical industrial legacy, rapid post-1990 economic transition, and geographical factors. Despite representing 21% of EU population, CEE countries account for 67% of $PM_{2.5}$-related premature deaths (EEA, 2023). The vehicle fleet averages 14 years old (versus 8.5 years EU-wide), with 40% meeting only Euro 3 standards or below. Additionally, 2.5-3.5 million households still rely on solid fuel heating with emission factors 50-100 times higher than gas boilers. These factors impose economic costs equivalent to 2.5-4.6% of GDP annually through healthcare expenditures, lost productivity, and infrastructure degradation.

Budapest exemplifies these challenges with its 1.7 million inhabitants experiencing $PM_{10}$ exceedances on 25-40 days annually at traffic stations. The city's topography, bisected by the Danube with elevation differences of 200-500 meters, creates temperature inversions 120-150 days annually, trapping pollutants below 200-300 meters. Winter $PM_{10}$ levels exceed summer by 50-80% due to residential heating (80,000-100,000 households using solid fuels), reduced mixing heights (100-200m), and lower wind speeds (2-3 m/s). The 12-station monitoring network reveals persistent heterogeneity: traffic stations record $PM_{10}$ of 35-42 μg/m³ versus 20-25 μg/m³ at suburban sites.

Current air quality forecasting approaches face fundamental limitations inadequately addressed for CEE contexts. Chemical Transport Models require computational resources limiting resolution to kilometres when street-level predictions are needed, while demanding detailed emissions inventories often unavailable in rapidly changing urban areas. Machine learning methods show degraded performance outside training domains and require extensive data unavailable in sparse monitoring networks. Critically, a persistent gap exists between model accuracy and interpretability, deep learning achieves superior performance but operates as black boxes, while interpretable methods fail to capture atmospheric non-linearities.

The literature reveals significant geographical bias, with comprehensive CEE studies remaining sparse despite the region's unique challenges: legacy infrastructure, diverse emissions including biomass burning, and complex topography. Existing studies focus on short-term campaigns rather than operational systems adapted to local conditions. Furthermore, operational deployment faces challenges rarely addressed in academic literature: real-time data quality control, computational constraints for timely delivery, and integration with urban decision-making infrastructure.

## Research Aims and Objectives

### 1. Primary Aim

To develop accurate and interpretable air quality forecasting methods for urban environments that bridge the gap between predictive performance and operational utility, with specific application to Central and Eastern European cities facing unique atmospheric and infrastructural challenges.

### 2. Specific Objectives

**Objective 1: Establish Baseline Forecasting Using ARIMA Modelling**

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 need for computationally efficient statistical methods that can operate with limited computational resources while providing interpretable parameters that reveal underlying temporal structures.

**Objective 2: Enhance Deterministic Models Through Neural Network Hybridization**

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: Assess Spatial Resolution Effects Across the Modelling Chain**

To 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.

**Objective 4: Quantify Feature Engineering Impact on Machine Learning Models**

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 test the hypothesis that appropriate feature engineering can overcome algorithmic limitations more effectively than architectural sophistication.

**Objective 5: Develop Interpretable Ensemble Fusion Framework**

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. The framework must explicitly model synergies and redundancies between ensemble members while providing interpretable parameters through Shapley values and interaction indices.

# 2 LITERATURE REVIEW

## 2.1. Historical evolution of Air Quality forecasting Methods

Air quality forecasting has undergone a remarkable transformation over the past five decades, evolving from rudimentary empirical relationships to sophisticated computational systems. This evolution has been driven by the growing complexity of urban air pollution challenges, advances in computational technology, expanding monitoring networks, and increasingly stringent regulatory requirements for accurate and timely air quality predictions, as illustrated in Figure 1.

### 2.1.1. Deterministic Modelling Approaches

The foundation of modern air quality forecasting was established during the environmental awakening of the 1970s, catalysed by landmark legislation such as the Clean Air Act in the United States and similar regulatory frameworks across Europe. Early efforts relied primarily on simple Gaussian plume models and box models that treated urban areas as well-mixed volumes with pollutant concentrations determined by basic emission-meteorology relationships.

A crucial advancement came with the development of Chemical Transport Models (CTMs), which simulate the complex physical and chemical processes governing pollutant transport, dispersion, and transformation in the atmosphere (Johnson, 2022). Models such as the Urban Airshed Model and the Regional Acid Deposition Model established the framework for process-based atmospheric modelling, incorporating detailed photochemical mechanisms and multiple emission source categories (Scheffe & Morris, 1993). Contemporary Eulerian CTMs, including CHIMERE employed in this thesis, solve the continuity equation on fixed three-dimensional grids, enabling simulation of advection, diffusion, chemical transformation, emissions, and deposition processes across regional and urban scales.

Despite continuous improvements, deterministic models face inherent limitations: sensitivity to input data quality (emissions inventories, meteorological fields), simplified parameterizations of sub-grid processes, and computational demands that constrain operational deployment at high spatial resolutions (Karamchandani et al., 2011). However, improvements of these models are still ongoing.

### 2.1.2. Data-Driven and Hybrid Approaches

The limitations of purely deterministic approaches motivated the development of data-driven methods capable of learning complex non-linear relationships directly from observational data. Early applications of artificial neural networks demonstrated superior capability in capturing relationships between meteorological variables and pollutant concentrations, particularly in urban environments with complex emission patterns (Gardner & Dorling, 1998).

Subsequent developments introduced ensemble methods, Random Forest and Gradient Boosting, which improved prediction accuracy by combining multiple weak learners while providing insights into feature importance(Breiman, 2001a; Friedman, 2001). Support vector machines gained popularity for handling high-dimensional data with good generalization performance (W. Z. Lu & Wang, 2005; Osowski & Garanty, 2007).

For temporal modelling, Long Short-Term Memory (LSTM) networks proved effective for capturing long-term dependencies in air quality time series (Biancofiore et al., 2017; Hochreiter & Schmidhuber, 1997).

However, increased model complexity introduced interpretability challenges that hindered regulatory acceptance. This motivated the development of hybrid approaches combining deterministic models with data-driven bias correction, as well as interpretable ensemble techniques. Advanced fusion methods have moved beyond simple averaging to include Bayesian model averaging and sophisticated aggregation operators (S. Li & Xing, 2025; J. Liu & Xing, 2023). The Choquet integral with fuzzy measures represents a particularly promising approach, providing a mathematically rigorous framework for non-linear ensemble fusion while maintaining complete interpretability through interaction coefficients and Shapley values (de Oliveira et al., 2022).

This thesis focuses on developing methods that balance high performance with operational requirements for transparency, reliability, and computational efficiency. This includes exploration of advanced aggregation techniques such as Choquet integrals with fuzzy measures, which provide mathematically rigorous frameworks for ensemble fusion while maintaining complete interpretability through interaction coefficients.

### 2.1.3. Contemporary challenges and research directions

This historical progression has led to several persistent challenges that define current research priorities. The fundamental tension between model complexity and interpretability remains central to the field, as regulatory frameworks increasingly demand explainable predictions for public health decision-making (Doshi-Velez & Kim, 2017). Real-time operational deployment requires addressing computational efficiency, system reliability, and model maintenance considerations that often conflict with research-oriented performance optimization (Campbell et al., 2022a).

Climate change introduces additional complexity, as models must remain accurate under evolving baseline conditions and increasing frequency of extreme weather events(Fiore et al., 2015). The integration of multi-scale information from global climate models to local-scale urban processes presents ongoing technical challenges, while data quality issues including sensor drift, missing observations, and representation bias continue to limit model effectiveness (Sokhi et al., 2022).

This evolution from simple empirical relationships to sophisticated AI systems reflects the field's maturation and growing recognition of air quality forecasting as a critical component of urban environmental management (Leo Hohenberger et al., 2022). Current research directions emphasize developing operational systems that maintain both high accuracy and interpretability, addressing the practical requirements of real-world air quality management while advancing scientific understanding of atmospheric processes (Baklanov & Zhang, 2020).

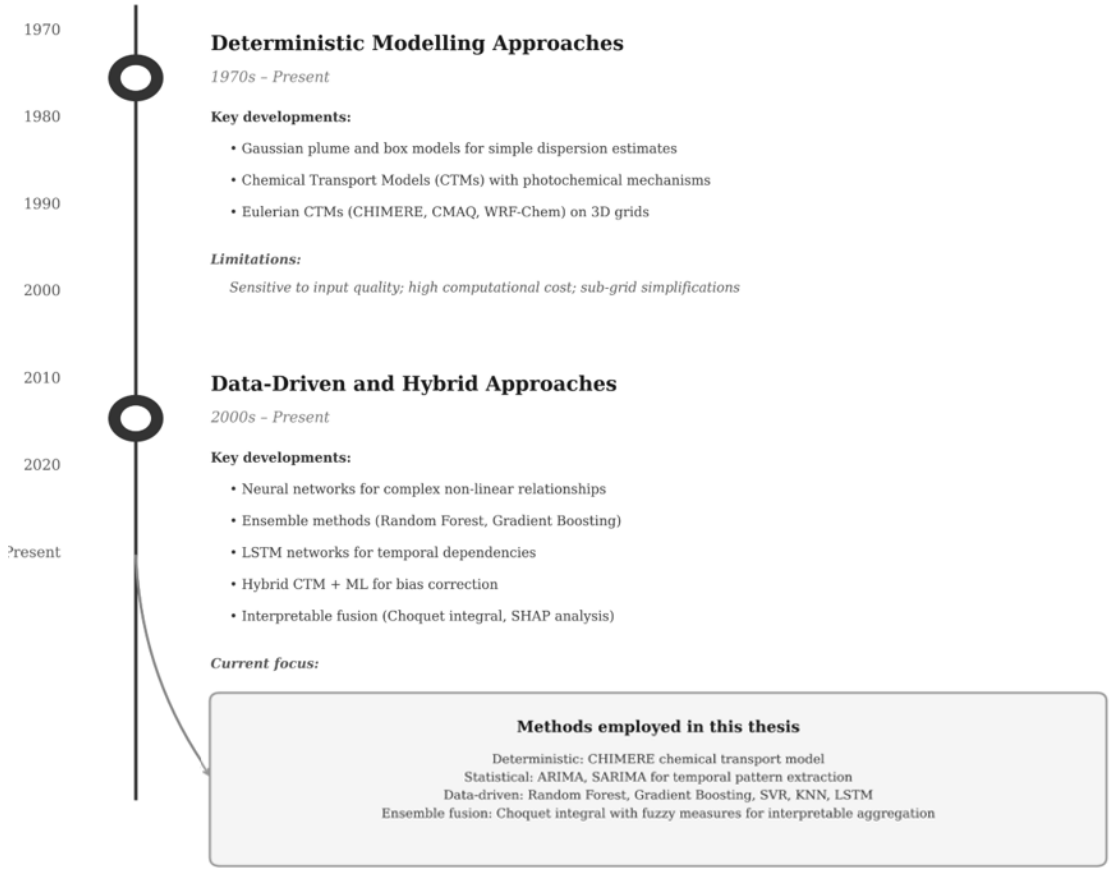



**Figure 1. Timeline of Air Quality Forecasting Method Evolution**

## 2.2. Deterministic Approaches

Deterministic air quality forecasting models encompass a spectrum of approaches ranging from simple analytical solutions to comprehensive numerical frameworks, as illustrated in figure 2. At the foundational level, Gaussian plume models provide analytical solutions for pollutant dispersion from point sources, box models treat urban areas as well-mixed volumes for policy assessment, and Lagrangian models track individual air parcels along trajectories for long-range transport analysis. In contrast, Eulerian models employ a fixed grid framework where the atmosphere is divided into three-dimensional cells, and pollutant concentrations are computed at each grid point by solving the continuity equation that accounts for advection, diffusion, chemical transformation, and source/sink terms.

Chemical Transport Models (CTMs) represent the most physically comprehensive evolution of these deterministic approaches, integrating elements from all foundational model types, particularly adopting the Eulerian grid-based framework, to provide sophisticated solutions to atmospheric transport and transformation processes. Unlike the simplified assumptions inherent in Gaussian, box, or basic Lagrangian models, CTMs solve the fundamental atmospheric diffusion equation through detailed three-dimensional Eulerian numerical schemes that incorporate comprehensive representations of emission sources, meteorological processes, chemical reactions, and removal mechanisms. The Eulerian approach is particularly advantageous for urban and regional air quality applications because it naturally handles

multiple interacting sources, complex chemical transformations, and spatially varying meteorological conditions within a consistent computational framework.

The deterministic nature of CTMs stems from their rigorous basis in physical and chemical laws, building upon the theoretical foundations established by simpler models while addressing their limitations through enhanced spatial resolution, detailed chemistry schemes, and comprehensive process representation. This progression from simple analytical models to complex Eulerian CTMs reflects the evolution of computational capabilities and scientific understanding, making modern CTMs particularly valuable for scenario analysis, policy evaluation, and mechanistic understanding of the underlying processes governing air quality at urban to regional scales.

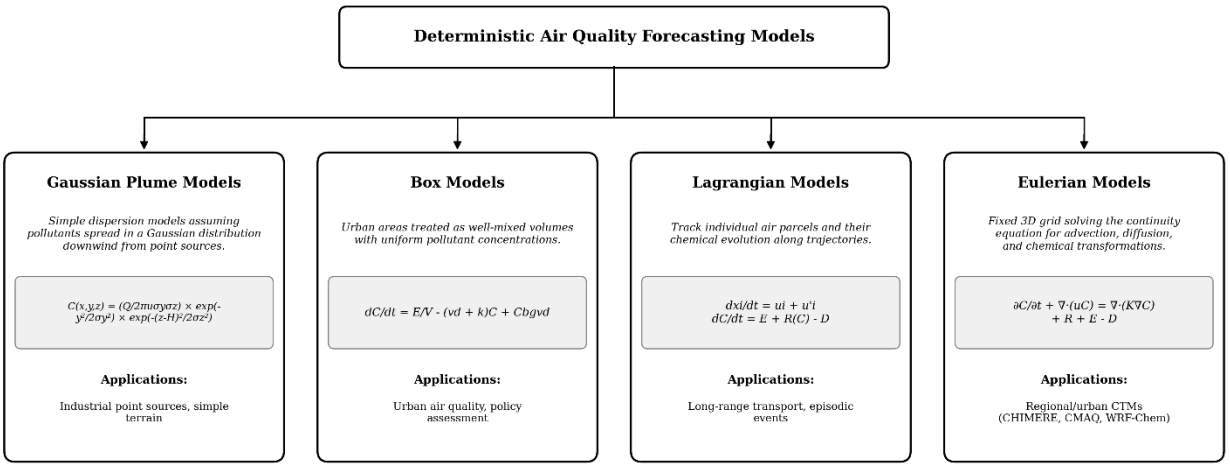



**Figure 2. Deterministic Air Quality Forecasting Models: Structure and Components.**

### 2.2.1. Theoretical foundation of Chemical Transport Models

Chemical Transport Models are built upon the fundamental mass conservation equation, which describes the temporal and spatial evolution of pollutant concentrations in the atmosphere. The general form of this equation for a chemical species *i* can be expressed as:

$$\frac{\partial C_i}{\partial t} + \nabla.(uC_i) = \nabla.(K\nabla C_i) + R_i + E_i - D_i \quad (1)$$

Where:

$\boldsymbol{C_i}$ represents the concentration of species *i* ($\mu g.m^{-3}$)

$\boldsymbol{t}$ is time (s)

$\boldsymbol{u}$ is the three-dimensional wind velocity vector ($m.s^{-1}$)

$\boldsymbol{K}$ is the turbulent diffusion tensor ($m^2.s^{-1}$)

$\boldsymbol{R_i}$ represents chemical production and loss rates ($\mu g.m^{-3}.\ s^{-1}$)

$\boldsymbol{E_i}$ denotes emission sources ($\mu g.m^{-3}.\ s^{-1}$)

$\boldsymbol{D_i}$ represents deposition processes ($\mu g.m^{-3}.\ s^{-1}$)

This equation captures the four fundamental processes affecting atmospheric pollutant concentrations: advection (transport by mean wind), diffusion (turbulent mixing), chemical transformation, and source-sink terms including emissions and deposition.

The advection term $\nabla.(uC_i)$ describes the transport of pollutants by the mean atmospheric flow. In the Eulerian framework commonly used by CTMs, this term is discretised using various numerical schemes such as the Bott advection scheme or the piecewise parabolic method, each with different properties regarding numerical diffusion and computational efficiency.

The diffusion term $\nabla.(K\nabla C_i)$ represents turbulent mixing processes that occur on scales smaller than the model grid resolution. The turbulent diffusion coefficient $K$ is typically parameterized based on atmospheric stability and boundary layer characteristics, often using K-theory approaches or more sophisticated turbulence closure schemes.

### 2.2.2. CHIMERE: A comprehensive Eulerian Model

CHIMERE is a multi-scale Eulerian chemistry-transport model developed by the French National Scientific Research Centre (CNRS) and the French National Institute for Industrial Environment and Risks (INERIS) (Menut et al., 2013). As illustrated in figure 3, the model operates through a sophisticated multi-stage workflow that exemplifies the complexity of modern deterministic air quality forecasting systems. The model has been extensively applied for air quality forecasting and policy assessment across various spatial scales, from urban to continental domains.

The CHIMERE modelling framework requires comprehensive mandatory input data including boundary and initial conditions derived from global models or measurements, detailed specification of the area-limited simulation domain encompassing mesh configuration and topographic properties, and extensive anthropogenic emission inventories. The model's pre-processing stage integrates three-dimensional meteorological fields (pressure, humidity, wind, temperature) with two-dimensional surface parameters (friction velocity, sensible heat flux, and boundary layer height) to establish the physical foundation for atmospheric transport calculations.

Chemical transport models can be coupled with numerical weather prediction (NWP) models through two distinct approaches: offline and online coupling (Kukkonen et al., 2012). In offline coupling, the meteorological model runs independently, generating three-dimensional fields that are subsequently read by the CTM at specified intervals. This approach, employed in the present study with WRF providing meteorological inputs to CHIMERE, offers computational efficiency and flexibility, as both models can be developed and optimised independently. In contrast, online coupling integrates the CTM directly into the NWP model, allowing two-way feedback, where aerosols and trace gases can influence radiative transfer and cloud microphysics. While online coupling better represents aerosol–radiation and aerosol–cloud interactions, it requires substantially greater computational resources (Baklanov et al., 2014). For regional air quality forecasting applications, offline coupling remains the predominant operational choice across European forecasting systems.

Central to the CHIMERE architecture is the chemistry-transport integration module, which simultaneously solves equations for atmospheric transport, turbulent mixing, emission injection, chemical transformation, and deposition processes while incorporating emission fluxes from multiple sources including anthropogenic, biogenic, mineral dust, forest fires, and volcanic emissions. The model's deterministic output ([c]mod) is typically evaluated by operating institutions or project teams through comparison with observational data ([c]obs) from surface monitoring stations, airborne platforms, and satellite measurements, ensuring model performance meets operational forecasting requirements. This comprehensive workflow demonstrates the data-intensive nature and computational complexity that characterize state-of-the-art chemistry-transport models in operational air quality forecasting applications.

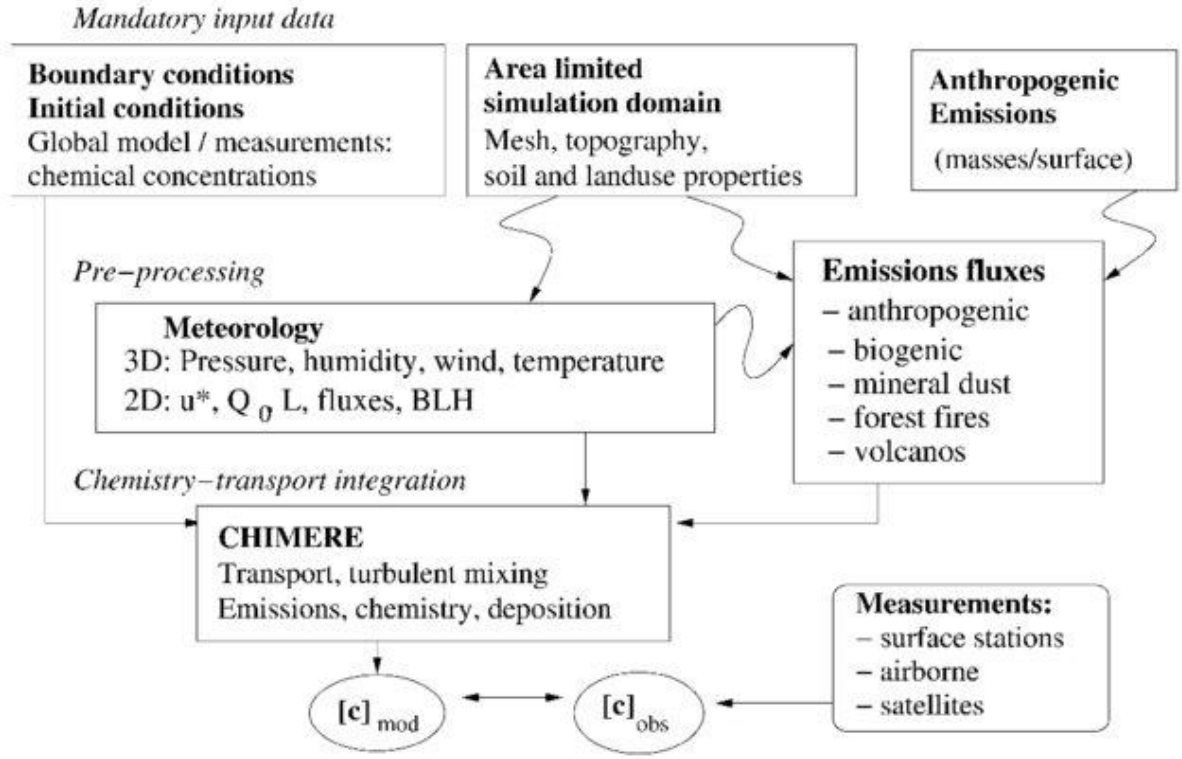



**Figure 3. General principle of a chemistry-transport model such as CHIMERE.** *In the box "Meteorology", u * stands for the friction velocity, $Q_0$ the surface sensible heat flux, L the Monin-Obukhov length and BLH the boundary layer height. $c_{mod}$ and $c_{obs}$ are the modelled and the observed chemical concentrations fields, respectively.*

### 2.2.3. Model Architecture and formulation

CHIMERE employs a terrain-following coordinate system in the vertical dimension, with the vertical coordinate σ defined as:

$$\sigma = \frac{p - p_{top}}{p_{surface} - p_{top}} \tag{2}$$

where $p$ is pressure, $p_{top}$ is the pressure at the model top, and $p_{surface}$ is surface pressure. This coordinate system allows for better representation of pollutant transport in complex topography.

The model integrates the mass conservation equation using an operator splitting approach, sequentially solving for different physical and chemical processes. The time integration follows:

$$C_i^{n+1} = T_{adv} \circ T_{diff} \circ T_{chem} \circ T_{emis} \circ T_{dep}(C_i^n) \quad (3)$$

where $T$ represents the operators for advection, diffusion, chemistry, emissions, and deposition processes, respectively, and the superscript $n$ denotes the time step.

#### 2.2.3.1. Chemical Mechanism

CHIMERE incorporates the MELCHIOR chemical mechanism (Menut et al., 2013), a reduced representation of atmospheric chemistry optimized for computational efficiency while maintaining chemical accuracy. The mechanism includes approximately 120 reactions involving 44 chemical species, covering:

- **Gas-phase chemistry**: Including $NO_x$-VOC photochemistry, ozone formation, and hydroxyl radical chemistry
- **Aerosol chemistry**: Secondary organic aerosol formation, inorganic aerosol thermodynamics
- **Aqueous-phase chemistry**: Cloud and fog water chemical processes

The photolysis rates are calculated using the TUV (Tropospheric Ultraviolet and Visible) radiative transfer model, accounting for cloud cover, aerosol optical properties, and solar zenith angle variations.

#### 2.2.3.2. Boundary Layer Parametrization

CHIMERE employs sophisticated boundary layer parameterizations crucial for urban air quality applications. The vertical diffusion coefficient is calculated using:

$$K_z = \frac{ku * z}{1 + \frac{kz}{L}} \quad (4)$$

For stable conditions, and:

$$K_z = ku * z\left(1 - \frac{z}{h}\right)^2 \quad (5)$$

for unstable conditions, where $k$ is the von Kármán constant (0.4), $u_*$ is the friction velocity, $z$ is height, $L$ is the Monin-Obukhov length, and $h$ is the boundary layer height.

### 2.2.4. Limitations and Challenges of Deterministic Approaches

Despite their physical foundation and comprehensive process representation, CTMs face significant limitations that affect their accuracy and operational utility, particularly for urban air quality forecasting applications.

#### 2.2.4.1. Computational Complexity and Resource Requirements

The computational demands of Chemical Transport Models represent a fundamental barrier to their widespread operational deployment. Modern CTMs require substantial computational resources to solve the coupled system of partial differential equations governing atmospheric transport and chemistry (Campbell et al., 2022b; Menut et al., 2021). A typical operational forecast for a regional domain at 3-5 km resolution requires 12-24 hours of computation on high-performance computing clusters with hundreds of processors. This computational burden stems from the mathematical stiffness of atmospheric chemical systems, where reaction timescales span from microseconds for radical chemistry to days for stable compounds (Maison et al., 2022). The requirement to maintain numerical stability while accurately representing these disparate timescales necessitates sophisticated numerical solvers and small timesteps, typically 1-5 minutes for regional applications. Furthermore, the computational cost scales non-linearly with resolution, as doubling the horizontal resolution requires eight times more computational resources due to the need for proportionally smaller timesteps to maintain Courant-Friedrichs-Lewy stability criteria. Although recent operational CTMs have achieved kilometre-scale resolutions (2–3 km) in several urban areas, computational constraints still impose practical trade-offs between domain size, spatial resolution, and forecast lead time, which can limit the representation of fine-scale urban processes and affect model utility for local air quality applications (Cheng et al., 2021).

#### 2.2.4.2. Meteorological Input Dependencies

Chemical Transport Models critically depend on meteorological input fields from numerical weather prediction models, inheriting and often amplifying meteorological uncertainties through non-linear chemical and physical processes (Rzeszutek et al., 2023). The boundary layer height, perhaps the most critical meteorological parameter for air quality, exhibits typical errors of ±200-300 meters in NWP models, directly affecting vertical mixing and surface pollutant concentrations (X. Chen et al., 2023; Hegarty et al., 2018; Lee et al., 2023). During stable nocturnal conditions, boundary layer height errors can cause concentration errors exceeding 100%, as pollutants become trapped in shallower-than-predicted layers. Wind field uncertainties pose equally significant challenges, where small wind speed errors of ±1-2 m $s^{-1}$ can cause large advection errors, particularly for passive tracers and in complex terrain where flow channelling and thermally driven circulations dominate. The timing and intensity of precipitation events critically affect wet deposition processes, with incorrect precipitation forecasts leading to either excessive removal of pollutants or unrealistic accumulation during what should be washout events (Battaglia et al., 2022). Temperature uncertainties propagate through multiple pathways, affecting chemical reaction rates with typical Q10 temperature coefficients of 2-3, biogenic emission rates that double for every 10°C increase, and gas-particle partitioning that shifts dramatically near ambient temperatures. The cumulative effect of these meteorological uncertainties often results in concentration errors exceeding 50% for individual pollution episodes, even when the CTM's chemical and physical parameterizations are theoretically perfect (Mundim et al., 2020).

#### 2.2.4.3. Emission Inventory Limitations

Emission inventories constitute the fundamental input data that quantify the release of pollutants into the atmosphere from anthropogenic and natural sources within a modelling domain. These comprehensive databases provide CTMs with essential information about what pollutants enter the atmosphere, where they originate, when they are released, and in what quantities. An emission inventory typically encompasses multiple pollutant species including criteria pollutants ($NO_x$, $SO_2$, CO, $PM_{10}$, $PM_{2.5}$), volatile organic compounds speciated into hundreds of individual compounds, greenhouse gases, and air toxics, with each emission source characterized by its spatial location, temporal variation, stack parameters for point sources, and chemical speciation profiles (*SMOKE v4.9 User's Manual*, 2022).

The construction of emission inventories follows either bottom-up or top-down methodologies, often combining both approaches to maximize accuracy and coverage as illustrated in Figure 4. Bottom-up inventories aggregate emissions from individual sources using activity data (fuel consumption, vehicle kilometres travelled, industrial production) multiplied by emission factors derived from source testing or literature values (Y. Zhao et al., 2011). This approach captures technological details and local variations but requires extensive data collection and may miss unofficial or unregulated sources. Top-down inventories start with national or regional totals derived from fuel sales, economic statistics, or satellite observations, then spatially and temporally disaggregate these totals using proxy variables. Hybrid approaches leverage the strengths of both methods, using bottom-up data where available and top-down estimates to ensure completeness (Crippa et al., 2020; Zhu et al., 2022).

For CTM applications, emission inventories must provide data at appropriate spatial and temporal resolution. Spatial allocation transforms emission data from administrative units or point source locations to model grid cells using Geographic Information Systems. Temporal allocation converts annual emission totals to hourly values, applying diurnal profiles for traffic emissions, seasonal variations for heating demands, and day-of-week patterns. Chemical speciation splits total VOC and PM emissions into individual compounds needed by chemical mechanisms (H. Simon et al., 2010). The quality of emission data directly determines CTMs' ability to reproduce observed concentrations, with studies demonstrating that emission uncertainties contribute 30-60% of total model uncertainty for primary pollutants and 20-40% for secondary pollutants (Janssens-Maenhout et al., 2019).

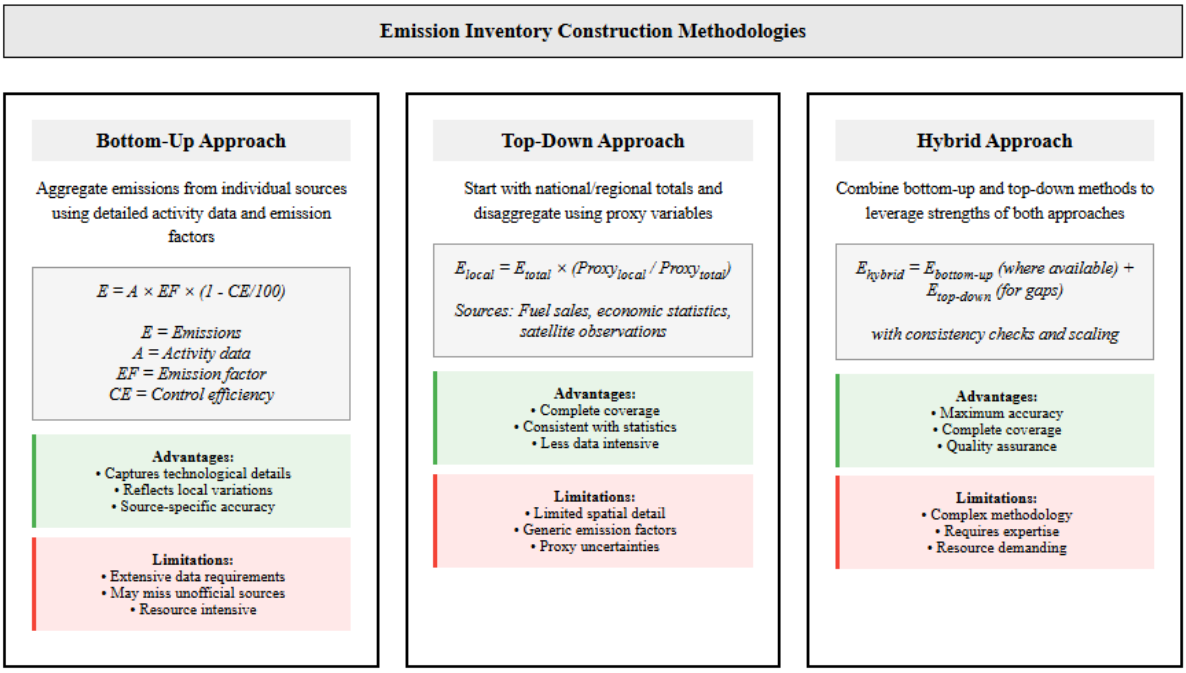



**Figure 4. Emission inventory construction methodologies.**

Despite their fundamental importance, emission inventories suffer from systematic uncertainties that propagate through the CTM system:

**Spatial allocation limitations:** Population density proxies assume uniform per-capita emissions despite income, lifestyle, and infrastructure variations creating order-of-magnitude differences in emission rates. Road network proxies distribute traffic emissions based on road classification without accounting for actual traffic volumes or fleet composition. Land use proxies fail to capture heterogeneity within categories, where industrial zones may contain facilities with vastly different emission characteristics (Guevara et al., 2021).

**Temporal disaggregation uncertainties:** Standard temporal profiles derived from limited measurements may not represent local activity patterns. Rush hour timing varies between cities based on work culture and urban structure, yet inventories often apply uniform profiles across entire countries. Seasonal patterns show strong regional variations, but inventories frequently use simplified factors. Special events and weather-dependent activities create episodic variations of 50-200% from average conditions that remain largely uncaptured (Grigoratos & Martini, 2014).

**Chemical speciation uncertainties:** The distribution of total VOC emissions among individual compounds relies on limited source testing that may not represent real-world diversity. Speciation profiles derived from US or European measurements may poorly represent sources in other regions using different technologies or fuels. Fuel composition variations substantially alter VOC speciation but are rarely tracked. These uncertainties critically affect ozone formation predictions, as different VOC species have reactivities varying by factors of 100 (Amato et al., 2009).

**Missing source categories:** Emission inventories systematically underestimate or omit numerous significant sources. Cooking emissions contribute 10-20% of urban organic aerosol but remain absent from most inventories (Authority, 2025; Coggon et al., 2024). Construction activities generate substantial dust and equipment emissions but escape traditional methods due to their temporary nature. Fugitive industrial emissions may equal stack emissions but rely on uncertain estimation methods (Xie et al., 2020). Non-exhaust traffic emissions (road dust, tire and brake wear) contribute 30-50% of traffic-related PM but remain poorly quantified (Keawboonchu et al., 2023; H. Li et al., 2015).

**Update frequency:** Regulatory inventories typically lag reality by 2-4 years, missing recent changes in vehicle fleets and control technologies. Rapid economic development or regulatory implementation can change emissions by 20-50% over these timescales. The COVID-19 pandemic illustrated this limitation starkly, as CTMs using pre-pandemic inventories failed to predict air quality during lockdowns (Barré et al., 2021; Crippa et al., 2019; Kong et al., 2023).

These fundamental limitations cascade through the CTM system, affecting primary pollutant predictions, secondary formation, and chemical regimes. The accumulated uncertainties from spatial misallocation, temporal disaggregation, chemical speciation, and missing sources often exceed 100% for specific locations and times, explaining much of the systematic bias observed in CTM predictions and motivating the development of inverse modelling, data assimilation, and machine learning approaches to constrain emission estimates using observational data.

## 2.3. Statistical Methods in Air Quality Forecasting

### 2.3.1. Evolution and Rationale for statistical Approaches

Statistical methods for air quality forecasting emerged from the recognition that pollutant concentrations exhibit strong temporal dependencies and recurring patterns that can be exploited for prediction without explicit representation of physical and chemical processes (Robeson & Steyn, 1990; Ryan, 2016). Unlike deterministic CTMs that attempt to simulate the causal chain from emissions through transport and chemistry to concentrations, statistical models identify and extrapolate patterns in historical observations, treating the atmosphere as a black box whose input-output relationships can be characterized empirically (Balashov et al., 2017). This fundamental difference in philosophy leads to complementary strengths and weaknesses: statistical models excel at short-term forecasting at specific locations where they are trained, capturing local effects that CTMs miss, while lacking the process understanding necessary for scenario analysis or spatial extrapolation.

The application of statistical methods to air quality gained momentum in the 1970s and 1980s as monitoring networks generated sufficient historical data for model development. Early applications focused on simple regression models relating pollutant concentrations to meteorological variables, achieving moderate success for next-day forecasts. The introduction of time series methods, particularly Box-Jenkins ARIMA models in the 1990s,

marked a significant advance by explicitly accounting for temporal autocorrelation in pollutant concentrations (Bai et al., 2018). These methods proved particularly effective for operational forecasting where accuracy at monitoring locations outweighed the need for spatial coverage or process understanding. The computational efficiency of statistical models, requiring seconds rather than hours for predictions, made them attractive for real-time applications and ensemble forecasting where multiple scenarios could be rapidly evaluated (T. Liu et al., 2018).

### 2.3.2. ARIMA Methodology and Extensions

The Autoregressive Integrated Moving Average (ARIMA) framework, formalized by Box and Jenkins in 1976, provides a systematic approach to modelling and forecasting time series that has become fundamental to air quality prediction. The ARIMA (p, d, q) model combines three components: autoregression of order p that relates current values to past observations, differencing of order d to achieve stationarity, and moving average of order q that incorporates past forecast errors. This structure captures both the persistence in pollutant concentrations due to atmospheric residence times and the random variations from unpredictable meteorological events or emission changes.

The mathematical formulation of ARIMA models elegantly captures temporal dependencies through the backshift operator B, where the general ARIMA (p, d, q) model is expressed as $\varphi_p(B)(1-B)\hat{}dy_t = \theta_q(B)\varepsilon_t$, with $\varphi_p$ and $\theta_q$ representing polynomials of orders p and q respectively, and $\varepsilon_t$ denoting white noise. The autoregressive component $\varphi_p(B) = 1 - \varphi_1 B - \varphi_2 B^2 - .. - \varphi_p B^p$ captures how current concentrations depend on recent history, with coefficients typically declining with lag to reflect decreasing influence of older observations. The differencing operator $(1-B)\hat{}d$ removes trends and achieves stationarity, essential for parameter estimation and forecasting. The moving average component $\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + .. + \theta_q B^d$ accounts for random shocks that affect the system, representing unpredictable variations from the expected pattern.

Model identification follows a systematic procedure beginning with stationarity assessment using unit root tests, particularly the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests that provide complementary hypotheses. The ADF test, with null hypothesis of non-stationarity, evaluates whether differencing is required, while KPSS, with null hypothesis of stationarity, confirms the adequacy of transformations. For air quality applications, first-order differencing (d=1) typically suffices for daily data, though seasonal differencing may be necessary for hourly observations showing strong diurnal patterns. The autocorrelation function (ACF) and partial autocorrelation function (PACF) guide selection of p and q orders, with characteristic patterns indicating appropriate model structure: exponential decay in ACF with cutoff in PACF after lag p suggests AR(p), while the reverse pattern indicates MA(q), and gradual decay in both suggests mixed ARMA structure.

The seasonal extension SARIMA (p, d, q) $(P, D, Q)_s$ incorporates periodic patterns prevalent in air quality data, where s represents the seasonal period (24 for hourly data with daily cycles, 7 for daily data with weekly patterns). The seasonal component multiplies the non-

seasonal ARIMA structure, creating a model capable of capturing both short-term dependencies and recurring cycles. For hourly air quality data, SARIMA models typically incorporate daily cycles through $(P, D, Q)_{24}$ terms that capture how concentrations at a given hour relate to the same hour on previous days, essential for modelling rush-hour peaks and nocturnal inversions that occur at consistent times.

Parameter estimation employs maximum likelihood methods, with modern implementations using sophisticated optimization algorithms to handle the non-linear likelihood surface. The Akaike Information Criterion ($AIC = -2l + 2k$) and Bayesian Information Criterion ($BIC = -2l + k \log n$) balance model fit against complexity, penalizing excessive parameters that risk overfitting. Grid search procedures systematically evaluate candidate models across reasonable ranges of p, q, P, and Q, with typical searches considering $p,q \in \{0,1,2,3,4,5\}$ and $P, Q \in \{0,1,2\}$ to maintain computational tractability while capturing relevant temporal structures.

### 2.3.3. Application to Multi-Scale Air Quality Forecasting

The application of ARIMA methods to air quality requires careful consideration of temporal granularity, as different aggregation levels reveal distinct patterns and require specific modelling strategies. Hourly data exhibits strong diurnal cycles driven by traffic patterns and boundary layer dynamics, with ACF showing significant peaks at lags 24, 48, and 72 hours. Studies across European cities have demonstrated that these patterns necessitate either seasonal SARIMA models or explicit deseasonalization before modelling. In Milan, Vlachogianni et al. (2011) found SARIMA $(1,0,1)\ (1,1,1)_{24}$ optimal for hourly $NO_2$ forecasting, while Barcelona required SARIMA $(2,0,1)\ (1,1,2)_{24}$ due to different traffic patterns and coastal effects. Sub-daily aggregation to 3-hour intervals reduces noise while preserving important temporal variations, with peaks at lags 8 and 16 capturing the rhythm of urban activities. Daily averages smooth short-term fluctuations to reveal weather-driven patterns and weekly cycles, typically requiring simpler ARIMA structures with $p, q \leq 3$, as demonstrated in Athens where ARIMA (2,1,2) proved optimal for daily $PM_{10}$ forecasting.

Station-specific modelling acknowledges that optimal ARIMA parameters vary with local conditions and emission patterns. European studies consistently show that traffic stations exhibit stronger autoregressive components (higher p) due to persistent local emissions, while background stations show stronger moving average components (higher q) reflecting the influence of transported pollution (Marinov et al., 2022). Industrial stations may require intervention models to handle irregular emission events, as industrial areas experience highly variable pollutant concentrations influenced by facility operations that standard ARIMA models struggle to capture. Suburban stations often show cleaner seasonal patterns amenable to SARIMA modelling, with studies achieving $R^2$ values exceeding 0.8 for ozone forecasting using seasonal ARIMA specifications with 24-hour periodicity (Gocheva-Ilieva, S.G., et al., 2014, Bouzghiba, H., et al., 2024).

Multi-horizon forecasting strategies extend ARIMA predictions beyond one-step-ahead, though forecast uncertainty increases rapidly with horizon. Direct forecasting fits separate models for each horizon, preserving forecast accuracy at the cost of model consistency.

Studies from London's air quality forecasting system showed direct 24-hour ahead models outperformed iterative approaches by 15-20% for peak concentration predictions. Iterative forecasting uses one-step predictions recursively, maintaining model parsimony but accumulating errors. Research in Paris demonstrated that iterative errors compound approximately linearly for the first 6 hours, then exponentially beyond 12 hours. For operational air quality forecasting, hybrid strategies prove effective: direct models for key horizons (next-day maximum, 48-hour average) combined with iterative forecasting for complete temporal coverage, as implemented in several European cities including Rome and Berlin.

The influence of meteorological regimes on optimal ARIMA structures has been documented across European climates. Mediterranean cities show distinct summer/winter patterns requiring seasonal model switching, with Athens using ARIMA (2,0,2) in summer but ARIMA (3,0,3) in winter due to different mixing conditions. Continental European cities exhibit more stable year-round patterns, though extreme events like heat waves or cold spells may require model adaptation. Coastal locations face additional complexity from sea-breeze circulations that modify diurnal patterns, necessitating time-varying parameter models or regime-specific ARIMA structures. Alpine cities present unique challenges with valley-mountain wind systems creating multi-scale temporal patterns not easily captured by standard SARIMA models.

Performance benchmarking across European applications reveals consistent patterns in ARIMA effectiveness. For $NO_2$, hourly ARIMA models typically achieve RMSE of 15-25 $\mu g/m^3$ at traffic sites and 10-15 $\mu g/m^3$ at background locations, with correlations of 0.7-0.85. $PM_{10}$ predictions show larger relative errors, with RMSE of 8-15 $\mu g/m^3$ for daily averages, reflecting the complex sources and atmospheric processing of particulate matter. Ozone forecasting exhibits strong seasonal dependence, with summer RMSE of 20-30 $\mu g/m^3$ versus winter values of 10-15 $\mu g/m^3$. Studies comparing ARIMA to persistence forecasts show 30-50% improvement in RMSE for 24-hour predictions, with gains largest during transition periods when concentrations change rapidly.

The integration of exogenous variables into ARIMAX models has shown mixed results across European studies. Meteorological variables as external regressors sometimes improve forecasts, particularly temperature for ozone and wind speed for PM, but gains are often marginal (5-10% RMSE reduction) and may compromise model stability. Lagged concentrations from nearby stations as exogenous inputs can enhance predictions during transport events, though benefits depend on predominant wind patterns and station spacing. Calendar variables (weekday/weekend, holidays) as external predictors show clearer benefits, improving $NO_2$ forecasts by 10-15% at traffic locations. The trade-off between model complexity and robustness generally favours parsimonious ARIMA structures over complex ARIMAX models for operational forecasting.

### 2.3.4. Advantages and limitations of statistical Methods

Statistical methods offer compelling advantages for operational air quality forecasting that explain their continued use despite the availability of sophisticated CTMs. Computational

efficiency remains paramount, with ARIMA models generating forecasts in milliseconds compared to hours for CTMs, enabling real-time updates as new observations become available and extensive ensemble forecasting for uncertainty quantification. The minimal data requirements, historical time series at a single location, contrast sharply with CTMs' need for comprehensive emissions, meteorology, and boundary conditions. Local optimization ensures that statistical models capture site-specific features missed by gridded CTMs: microscale effects from nearby sources, building-induced flow patterns, and systematic biases from monitoring station siting. The empirical nature of statistical models allows them to implicitly account for all processes affecting concentrations, including those poorly represented or omitted from CTMs.

Performance evaluation consistently demonstrates superior short-term forecasting skill of statistical methods at monitoring locations. The ARIMA search grid approach applied to Budapest achieved mean absolute errors of 11-24 μg/m³ for hourly $NO_2$ and 4.8-7.6 μg/m³ for 3-hourly $PM_{10}$, outperforming CTM predictions by 30-50% at the same locations. This accuracy advantage stems from statistical models' ability to learn and correct for systematic patterns in local pollution, including consistent model biases, recurring emission patterns, and site-specific meteorological effects. The transparent model structure facilitates diagnostic evaluation, with residual analysis revealing model deficiencies and suggesting improvements.

However, statistical methods face fundamental limitations that restrict their applicability to specific forecasting contexts. The absence of physical foundation means statistical models cannot predict responses to unprecedented conditions, emission changes, or control strategies, core applications of CTMs. Spatial interpolation remains problematic as models trained at one location fail when applied elsewhere, limiting statistical methods to monitoring station locations without spatial coverage. Temporal extrapolation beyond training conditions produces unreliable predictions, with model performance degrading rapidly for unusual meteorological conditions or emission scenarios not represented in historical data. The black-box nature provides no process understanding, making it impossible to diagnose why predictions fail or how to improve them beyond statistical adjustments.

## 2.4. Machine Learning Revolution in Air quality Forecasting

The application of machine learning to air quality forecasting represents a fundamental paradigm shift in atmospheric science, transitioning from purely deterministic chemical transport models to data-driven approaches capable of learning complex patterns from historical observations. This transformation has unfolded through distinct technological waves, beginning with traditional machine learning algorithms in the early 2000s, evolving through deep learning architectures in the 2010s, and currently focusing on hybrid approaches that integrate physical understanding with statistical learning (Feng et al., 2025; Liao et al., 2020)

### 2.4.1. Traditional Machine Learning Approaches

#### 2.4.1.1. Random Forest in Air Quality Applications

Random Forest algorithms have been extensively investigated for air quality prediction across diverse geographical and meteorological contexts. (Lei et al., 2022) explored RF applications in Macao's subtropical climate, while (Mampitiya et al., 2024) examined their utility in Sri Lankan tropical conditions. The widespread adoption of RF in air quality studies stems from its documented ability to handle high-dimensional meteorological data and capture non-linear relationships between atmospheric variables.

The literature reveals considerable variation in RF implementation strategies. Some researchers advocate for deep, unconstrained tree architectures when modelling complex urban environments (Gladson et al., 2022), while others emphasize the importance of regularization through depth constraints to prevent overfitting (Bi et al., 2020). This divergence in approaches highlights an ongoing debate about the optimal balance between model complexity and generalization capability in atmospheric applications.

Recent studies have investigated asymmetric loss functions within RF frameworks to address the public health implications of under-predicting pollution episodes (Shaziayani et al., 2021). However, the effectiveness of such modifications remains contested, with some researchers arguing that standard ensemble variance reduction naturally mitigates prediction bias without specialized loss functions (Freeman et al., 2018).

#### 2.4.1.2. Support Vector Regression

Support Vector Regression has attracted significant attention in air quality forecasting due to its theoretical advantages in high-dimensional spaces through kernel methods. Early applications by (Méndez et al., 2023) explored SVR for criteria pollutant prediction, establishing a foundation for subsequent research. The radial basis function (RBF) kernel has been particularly popular, with studies investigating its application to meteorological feature sets (Castelli et al., 2020; Masood et al., 2023).

However, the literature reveals persistent challenges in SVR implementation for atmospheric applications. The curse of dimensionality, first identified in this context by (Z. Zhao et al., 2022), poses significant obstacles when the number of features approaches or exceeds the number of training samples. Several studies have documented cases where SVR performance degraded catastrophically in high-dimensional meteorological feature spaces (Murillo-Escobar et al., 2019), leading some researchers to question the practical utility of kernel methods for operational air quality forecasting.

The debate around SVR effectiveness has prompted investigations into alternative kernel functions and regularization strategies. Polynomial kernels have been explored for capturing seasonal patterns (Sotomayor-Olmedo et al., 2013), while hybrid kernels combining multiple basis functions have shown promise in specific applications (L. H. Zhang et al., 2021). Nevertheless, the computational complexity and parameter sensitivity of SVR continue to limit its adoption in real-time forecasting systems.

#### 2.4.1.3. Gradient Boosting Methods

Gradient Boosting Regression (GBR) and its variants have emerged as powerful alternatives to random forests, with extensive investigation in air quality contexts. The sequential learning paradigm of gradient boosting, where each model corrects its predecessor's errors, has proven particularly effective for capturing temporal evolution in pollution patterns (Masood et al., 2023).

XGBoost and LightGBM implementations have gained prominence in recent literature, with studies comparing their performance across different pollutants and temporal scales (Kalantari et al., 2024). The ability to incorporate custom loss functions has made gradient boosting attractive for researchers addressing specific forecasting objectives, such as extreme event prediction or exposure assessment (Z. Y. Chen et al., 2024; Ke et al., 2022).

A significant theme in gradient boosting literature concerns the trade-off between model complexity and interpretability. While deep boosting ensembles can capture subtle atmospheric patterns, they sacrifice the transparency crucial for regulatory applications (Alotaibi et al., 2024). This has led to investigations of interpretable boosting variants and post-hoc explanation methods, though consensus on optimal approaches remains elusive.

#### 2.4.1.4. K-Nearest Neighbours

Despite being one of the oldest machine learning methods, K-Nearest Neighbours has experienced renewed interest in air quality applications, particularly for anomaly detection and extreme event prediction. Recent studies have explored KNN's effectiveness when combined with specialized feature engineering, particularly in deviation or anomaly spaces rather than absolute concentration spaces (Jin et al., 2023; Rahman et al., 2024).

The literature reveals that KNN's performance is highly sensitive to distance metric selection and the curse of dimensionality. Studies have investigated various distance measures beyond Euclidean, including Mahala Nobis distance for correlated features and dynamic time warping for temporal patterns (Lin & Li, 2024). The optimal number of neighbours remains context-dependent, with urban environments typically requiring larger k values to smooth local noise.

### 2.4.2. Deep Learning Revolution

#### 2.4.2.1. LSTM Networks and Temporal Dependencies

Long Short-Term Memory networks have fundamentally transformed time series forecasting in atmospheric science, with extensive investigation since their introduction to air quality applications.(Drewil & Al-Bahadili, 2022) examined LSTM architectures for pollution prediction, while (Sudha et al., 2025) explored enhanced variants incorporating attention mechanisms. The ability of LSTMs to capture long-range temporal dependencies has made them particularly attractive for modelling pollutant accumulation and dispersion processes.

The literature reveals ongoing debates about optimal LSTM architectures for air quality applications. Some studies advocate for deep, multi-layer architectures to capture hierarchical temporal patterns (Y. Huang et al., 2022), while others demonstrate that shallow networks with appropriate feature engineering can achieve comparable performance with reduced computational cost (Ansari & Quaff, 2025). The question of optimal lookback windows remains contentious, with different studies reporting peak performance at varying temporal scales from hours to weeks.

Bidirectional LSTMs have been investigated for applications where future meteorological forecasts are available, though their utility for operational forecasting remains limited (Drewil & Al-Bahadili, 2022). Variants incorporating convolutional layers (ConvLSTM) have shown promise for spatiotemporal prediction, particularly in urban networks with multiple monitoring stations (L. Zhang et al., 2021).

#### 2.4.2.2. The critical Role of feature Engineering

The literature consistently emphasizes that feature engineering often contributes more to model performance than algorithmic selection. Studies have developed specialized feature sets targeting different atmospheric processes: short-term dynamics for traffic-related variations, long-term patterns for seasonal trends, meteorological drivers for dispersion mechanisms, and anomaly features for extreme events (Tang et al., 2024; Z. Zhang et al., 2024).

Temporal encoding strategies have received particular attention, with various approaches to capturing diurnal, weekly, and seasonal patterns. Cyclical encoding using sine and cosine transformations has become standard practice, though debate continues about optimal encoding frequencies and whether to include higher harmonics (Alkabbani et al., 2022).

The recognition that atmospheric systems exhibit distinct behavioural regimes has led to investigations of regime-specific modelling approaches. Studies have identified various regime classification schemes based on meteorological conditions, emission patterns, and pollutant concentrations (Cheng et al., 2021; García-Herrera et al., 2022). The literature reveals significant performance improvements when models are trained separately for different regimes, though optimal regime identification methods remain debated. While these individual components, specialized feature sets, temporal encodings, and regime-specific approaches, have demonstrated value independently, no comprehensive framework exists that systematically combines physics-informed feature engineering with regime-aware model selection and interpretable ensemble fusion. This integration represents a key contribution of the present thesis.

### 2.4.3. Current Challenges and Future Directions

The machine learning revolution in air quality forecasting has achieved remarkable progress, yet significant challenges remain. The interpretability-accuracy trade-off continues to limit adoption in regulatory contexts, with stakeholders requiring both high performance and explainable predictions. The generalization of models across different geographical regions and meteorological conditions remains problematic, with most studies reporting degraded performance when models are applied outside their training domains.

Data quality and availability persist as fundamental limitations, particularly in developing regions where monitoring networks are sparse. The integration of satellite observations, low-cost sensors, and citizen science data presents opportunities but also challenges related to data fusion and quality control. The computational requirements of advanced deep learning models pose barriers to real-time operational deployment, particularly for high-resolution spatiotemporal prediction.

The literature suggests that future advances will likely come from intelligent integration rather than algorithmic innovation alone. The combination of physical understanding with machine learning, careful feature engineering informed by atmospheric science, and transparent ensemble methods that maintain interpretability while achieving high performance represent the most promising directions for continued progress in this rapidly evolving field.

## 2.5. Hybrid and Ensemble Methods in Air Quality Forecasting

The limitations of both purely deterministic and purely statistical methods have driven the development of hybrid and ensemble approaches in air quality forecasting. This methodological evolution reflects a growing recognition that neither physics-based models nor machine learning algorithms alone can adequately capture the complexity of atmospheric pollution dynamics (L. Jiang et al., 2020; Sayeed, 2021). The literature reveals a progression from simple model averaging to sophisticated fusion techniques that attempt to leverage complementary strengths while mitigating individual weaknesses.

### 2.5.1. Deterministic-Statistical Hybrid Approaches

#### 2.5.1.1. Integration of Chemical Transport Models with Machine Learning

The coupling of Chemical Transport Models (CTMs) with statistical methods represents one of the most active areas of research in air quality forecasting. This hybridization addresses fundamental limitations of CTMs, including computational constraints, coarse spatial resolution, and systematic biases, while maintaining physical consistency (Menut et al., 2020). The literature documents various integration strategies, from post-processing corrections to fully coupled systems.

Early hybrid approaches focused on statistical post-processing of CTM outputs. (Z. Liu et al., 2022) applies deep learning to bias-correct ozone in a chemistry-climate model, while (Tchepel et al., 2020) investigated urban aerosol forecasting through CTM-statistical integration. The fundamental premise underlying these approaches is that CTMs provide physically consistent baseline predictions that statistical methods can refine using historical observation-model discrepancies. However, debates persist about whether post-processing preserves the physical relationships encoded in CTMs or merely fits statistical patterns that may not generalize to new conditions (Tang et al., 2024).

More sophisticated integration strategies involve using CTM outputs as features for machine learning models alongside meteorological and emission data. (Sayeed, 2021)

developed the CMAQ-CNN approach, where convolutional neural networks learn to extract relevant patterns from gridded CTM outputs. Similarly, (Ajdour et al., 2024)investigated using CHIMERE outputs with various machine learning algorithms for ozone prediction. These studies suggest that CTM outputs provide valuable physical context that purely data-driven models cannot capture from observations alone.

The WRF-CHIMERE coupling investigated by (L. Jiang et al., 2020) exemplifies online integration where meteorological and chemical components interact bidirectionally. This approach has been extended to include machine learning components that learn systematic model errors and provide real-time corrections. However, the computational overhead and complexity of such systems limit operational deployment.

#### 2.5.1.2. Physics-Informed Machine Learning Frameworks

Recent literature has explored more fundamental integration of physical principles into machine learning architectures. Physics-informed neural networks (PINNs) incorporate conservation laws, chemical kinetics, and atmospheric dynamics as constraints or regularization terms in the learning process (B. Li et al., 2023). This approach aims to ensure that learned patterns respect fundamental physical laws while maintaining the flexibility to capture complex empirical relationships.

Various strategies for encoding physical knowledge have been investigated. Some studies design neural network architectures that mirror atmospheric processes, with layers corresponding to vertical atmospheric levels or separate modules for different chemical mechanisms (S. Li et al., 2024). Others incorporate physical constraints through specialized loss functions that penalize violations of mass conservation, chemical equilibrium, or thermodynamic principles (Mukherjee & Bhattacharyya, 2024).

The super-resolution chemistry transport modelling approach explored by (Bessagnet et al., 2021) represents another hybrid strategy, using deep learning to downscale coarse CTM outputs to finer spatial resolutions. This approach leverages the physical consistency of CTMs while using machine learning to add fine-scale details that would be computationally prohibitive to model explicitly.

#### 2.5.1.3. Challenges in Hybrid Model Development

The literature identifies several persistent challenges in developing effective hybrid models. The computational overhead of running both deterministic and statistical components limits real-time application, particularly for high-resolution forecasting (Fan et al., 2022; Georgiou et al., 2022). Questions about error propagation between components remain unresolved, with debates about whether coupling amplifies or mitigates individual model uncertainties.

The interpretability of hybrid models presents particular challenges. While CTMs provide physically interpretable process representations and machine learning can offer feature importance metrics, the interaction between components often creates opacity about which factors drive specific predictions (Houdou et al., 2024). This limitation has implications for

regulatory applications where understanding causality is as important as prediction accuracy.

### 2.5.2. Ensemble Techniques in Air Quality Forecasting

#### 2.5.2.1. Traditional Ensemble Approaches

The application of ensemble methods to air quality forecasting has evolved significantly over the past decade. Early studies explored simple arithmetic averaging of multiple model outputs, recognizing that combining predictions could reduce individual model biases (Galmarini et al., 2018). However, these approaches assume equal model skill and ignore potential complementarities or redundancies between models.

Weighted averaging schemes have been extensively investigated, with various strategies for determining optimal weights. Inverse error weighting, where models are weighted inversely proportional to their historical errors, has been widely applied (F. Chen et al., 2023). Performance-based weighting using metrics like RMSE, correlation coefficients, or skill scores offers more sophisticated approaches, though the choice of metric significantly influences results (Hodson, 2022).

Bayesian Model Averaging (BMA) provides a probabilistically principled framework for ensemble combination. (Qi et al., 2022) applied BMA to multi-model air quality forecasts in Central China, while (Kim et al., 2020a) investigated its application to streamflow and air quality prediction. BMA weights models based on their posterior probabilities given historical data, naturally accounting for model uncertainty. However, the assumption of conditional independence between models, which often share inputs and physical parameterizations, remains problematic in atmospheric applications.

#### 2.5.2.2. Advanced Stacking and Meta-Learning

Stacking or stacked generalization has emerged as a powerful ensemble technique in recent air quality literature. This approach uses a meta-learner to combine base model predictions, potentially capturing non-linear relationships and complementarities between models. (Xu et al., 2025) applied stacking with SHAP-based insights for urban air quality forecasting, while (Ravindiran et al., 2025) explored ensemble stacking for air quality prediction in Hyderabad.

The selection of meta-learning algorithms varies widely across studies. Linear models offer interpretability but may not capture complex interactions, while non-linear meta-learners like gradient boosting or neural networks can model sophisticated relationships at the cost of transparency (Tian et al., 2024). The literature reveals ongoing debates about the appropriate complexity level for meta-learners, with some arguing that simple models suffice when base models are diverse, while others demonstrate performance gains from complex meta-learning.

Cross-validation strategies for training meta-learners receive particular attention in the literature. Time series cross-validation, spatial hold-out, and blocked cross-validation have

been explored to prevent data leakage and ensure realistic performance estimates (Özüpak et al., 2025). The temporal dependence in air quality data complicates traditional cross-validation approaches, necessitating careful consideration of training and validation set construction.

#### 2.5.2.3. Multi-Model and Multi-Configuration Ensembles

Multi-model ensembles combining predictions from different CTMs have been investigated extensively. The AQMEII (Air Quality Model Evaluation International Initiative) projects have coordinated multi-model exercises comparing and combining outputs from CMAQ, CHIMERE, WRF-Chem, and other models (Galmarini et al., 2021). These studies reveal that model diversity improves ensemble robustness, though the computational cost of running multiple CTMs limits operational implementation.

Multi-configuration ensembles using the same base model with perturbed inputs or parameterizations offer a computationally more tractable alternative. Studies have explored perturbations to emissions inventories, meteorological boundary conditions, chemical mechanisms, and physical parameterizations (Díaz-Isaac et al., 2019; Hu et al., 2017). The generation of representative ensemble members that adequately sample uncertainty space remains an active research area.

The optimal ensemble size represents a persistent question in the literature. While larger ensembles theoretically provide better uncertainty estimates, practical constraints and diminishing returns suggest optimal sizes between 5-20 members for most applications (Milinski et al., 2020). The trade-off between ensemble diversity and computational cost continues to shape operational implementations.

### 2.5.3. Fuzzy Aggregation Operators: An Unexplored Frontier

#### 2.5.3.1. Theoretical Foundations and Potential Applications

Despite the extensive application of fuzzy logic in environmental modelling, the use of fuzzy aggregation operators for air quality ensemble fusion remains remarkably limited in the literature. Fuzzy aggregation operators offer mathematical frameworks for combining information while explicitly modelling uncertainty, imprecision, and interactions between information sources (Klement et al., 2004).

The theoretical foundations of fuzzy aggregation rest on the concept of aggregation functions that generalize classical operations like minimum, maximum, and arithmetic mean. T-norms and t-conorms provide families of operators that can model different types of interactions, from pure redundancy to full complementarity. More sophisticated operators like the Choquet integral, Sugeno integral, and copula-based aggregators offer additional flexibility in modelling dependencies.

The potential advantages of fuzzy aggregation for air quality applications are compelling. Unlike weighted averaging, these operators can model non-linear interactions between models, explicitly represent uncertainty in model outputs, and provide interpretable parameters that quantify relationships between information sources (Grabisch, 2011). The

ability to handle imprecise and uncertain information aligns well with the characteristics of both atmospheric measurements and model outputs.

#### 2.5.3.2. Applications in Related Environmental Domains

While direct applications to air quality ensemble fusion are scarce, fuzzy aggregation has been successfully applied in related environmental domains. In hydrological forecasting, fuzzy operators have been used to combine multiple rainfall-runoff models, with studies demonstrating improved performance over traditional averaging (Burak Akgun et al., 2021a; Xiong et al., 2001). Water quality assessment has employed fuzzy aggregation to combine multiple indicators into composite indices that account for parameter interactions and uncertainties (Ocampo-Duque et al., 2006).

Climate modelling has seen limited but promising applications of fuzzy aggregation. Studies have explored using fuzzy measures to combine multiple climate model projections while accounting for model dependencies and shared biases (Burak Akgun et al., 2021b). The ability to model both redundancy and synergy between models addresses fundamental limitations of traditional climate ensemble approaches.

In renewable energy forecasting, particularly wind and solar power prediction, fuzzy aggregation has been applied to combine multiple forecast models. These applications demonstrate the operators' ability to adapt to changing conditions and model time-varying relationships between predictors (Peláez-Rodríguez et al., 2023). The parallels with air quality forecasting, including dealing with meteorological inputs and non-stationary processes, suggest potential transferability.

#### 2.5.3.3. The Research Gap in Air Quality Applications

The conspicuous absence of systematic investigation of fuzzy aggregation operators in air quality ensemble literature represents a significant research gap. While isolated studies have explored fuzzy logic for components of air quality systems, such as fuzzy inference systems for pollution alerts or fuzzy clustering for pattern recognition, the application to model fusion and ensemble combination remains virtually unexplored.

Several factors may explain this gap. The air quality modelling community's strong tradition of physics-based modelling may create resistance to mathematical frameworks perceived as lacking physical interpretation. The additional mathematical complexity of fuzzy operators compared to simple averaging may deter adoption, particularly given the already substantial complexity of atmospheric models. The lack of established software libraries and benchmark datasets for fuzzy aggregation in atmospheric applications presents practical barriers to entry.

Furthermore, the interdisciplinary nature of fuzzy aggregation, drawing from mathematics, computer science, and decision theory, may limit cross-pollination with the atmospheric science community. The terminology and mathematical formalism of fuzzy set theory may be unfamiliar to researchers trained primarily in atmospheric physics and chemistry.

However, the limitations of current ensemble methods strongly suggest that exploring fuzzy aggregation could yield significant advances. The inability of traditional weighted

averaging to model inter-model dependencies, the black-box nature of neural network meta-learners, and the need for interpretable fusion mechanisms that can be explained to stakeholders all point toward fuzzy operators as a promising research direction.

#### 2.5.3.4. Specific Opportunities for Fuzzy Aggregation

The Choquet integral, in particular, offers compelling advantages for air quality ensemble fusion. Its ability to model both the importance of individual models (through fuzzy measures) and their interactions (through Möbius representations) directly addresses limitations of current approaches. The integral can capture situations where models provide redundant information (negative interaction) or synergistic insights (positive interaction), phenomena that linear combination methods cannot represent.

The interpretability of Choquet integral parameters through concepts like Shapley values and interaction indices could facilitate understanding of ensemble behaviour. This transparency contrasts sharply with neural network meta-learners and could be particularly valuable for regulatory applications where decision-makers need to understand prediction rationales.

Other fuzzy aggregation operators also merit investigation. The Sugeno integral, which uses qualitative rather than quantitative scales, could be valuable when combining models with different output types or confidence levels. Copula-based aggregation could model complex dependencies between models that share inputs or physical assumptions. Type-2 fuzzy operators could represent uncertainty about the uncertainty itself, addressing situations where model reliability varies with atmospheric conditions.

### 2.5.4. Current Challenges and Future Research Directions

#### 2.5.4.1. Computational and Implementation Barriers

The literature consistently identifies computational requirements as a major obstacle to implementing sophisticated hybrid and ensemble methods. The multiplicative effect of running multiple models or configurations, combined with the overhead of fusion algorithms, strains computational resources even with modern high-performance computing facilities. This challenge is particularly acute for operational forecasting systems that must deliver predictions within strict time constraints.

Efforts to address computational challenges have explored various strategies. Surrogate modelling and emulation techniques attempt to approximate expensive models with cheaper alternatives (Vu et al., 2020). Adaptive ensemble selection dynamically chooses subset of models based on current conditions, reducing unnecessary computations (Delle Monache et al., 2020). However, these approaches introduce additional complexities and potential failure modes that require careful consideration.

#### 2.5.4.2. The Interpretability-Performance Dilemma

A persistent theme throughout the hybrid and ensemble literature concerns the tension between prediction accuracy and model interpretability. While sophisticated machine

learning meta-learners and deep hybrid architectures can achieve superior predictive performance, they often sacrifice the transparency essential for understanding pollution mechanisms and informing mitigation strategies.

This trade-off has profound implications for the adoption of advanced methods in operational and regulatory contexts. Stakeholders increasingly demand not just accurate predictions but also explanations of how those predictions are generated. The development of explainable AI techniques specifically tailored to atmospheric ensemble methods represents a critical need for advancing the field.

#### 2.5.4.3. Theoretical Foundations and Validation Frameworks

Despite extensive empirical investigation, the theoretical foundations of many hybrid and ensemble approaches remain underdeveloped. Fundamental questions about optimal ensemble design, the relationship between diversity and performance, and the conditions under which different fusion strategies excel lack definitive answers. The development of theoretical frameworks that can guide ensemble construction and predict performance bounds would significantly advance the field.

Validation of hybrid and ensemble methods presents unique challenges not fully addressed in current literature. Traditional metrics may not adequately capture the value of ensemble predictions, particularly regarding uncertainty quantification and extreme event prediction. The development of comprehensive evaluation frameworks that assess multiple aspects of ensemble performance, including reliability, sharpness, and value for decision-making, remains an important research direction.

# 3 MATERIALS AND METHODS

This research employs a multi-methodological approach across different urban environments to comprehensively explore air quality forecasting techniques, from statistical baselines through deterministic-hybrid models to advanced ensemble fusion. The selection of two geographically distinct study areas: Budapest (Hungary) and Moroccan cities (Agadir and Casablanca), was driven by both scientific and practical considerations. Budapest was initially intended as the sole study area; however, computational constraints during CHIMERE model implementation necessitated the use of pre-configured Moroccan domains where the model was already operational. This practical limitation evolved into a methodological advantage: the two regions offer complementary testing environments with contrasting characteristics in terms of climate (continental vs. Mediterranean/semi-arid), emission source profiles (industrialized European vs. developing North African), monitoring infrastructure density, and data availability. Rather than testing transferability of a single method, this study strategically 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 thesis structure follows a methodological progression:

- Statistical baseline establishment (Budapest): ARIMA models demonstrate the capabilities and limitations of purely data-driven time series approaches in a complex Central Eastern European urban environment.
- Deterministic model enhancement (Agadir and Casablanca): CHIMERE-ANN hybrid approach validates the effectiveness of combining physical models with neural networks for bias correction in North African coastal cities.
- Advanced ensemble fusion (Budapest): Machine learning models combined through Choquet Integral fusion represent the state-of-the-art in interpretable ensemble methods, applied where comprehensive data availability enables sophisticated multi-model approaches.

This structured approach allows systematic exploration of each methodology's strengths while matching techniques to data availability and operational constraints typical of different geographical contexts.

## 3.1.Study area and data processing

### 3.1.1. Budapest Monitoring Network description

This study utilized air quality data from the Budapest atmospheric monitoring network, comprising 11 stations strategically distributed across the metropolitan area (47.5°N, 19.0°E) as seen in figure 5. Budapest, the capital of Hungary with a population of approximately 1.7 million inhabitants, experiences a continental climate with significant seasonal variations that influence pollutant dispersion patterns. The city's location in the Carpathian Basin, bisected by the Danube River, creates complex topographical conditions affecting local atmospheric circulation and pollution accumulation, particularly during winter anticyclonic episodes.

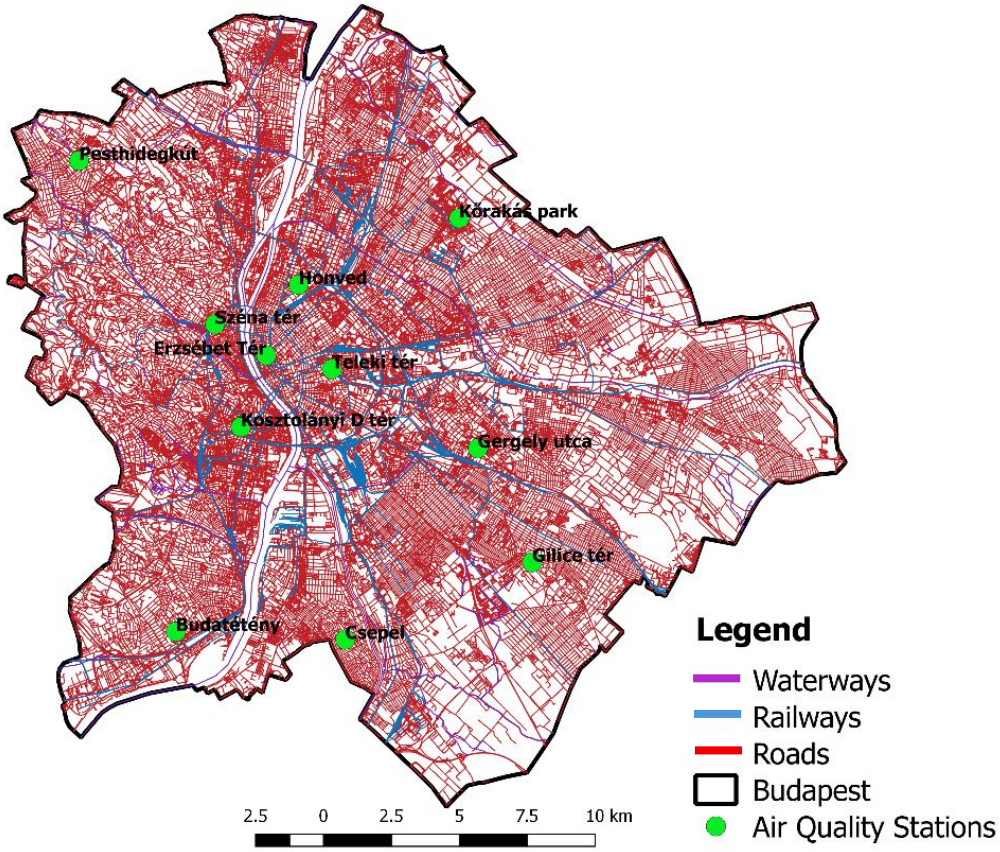



**Figure 5. Budapest transport network and air quality stations.**

The monitoring stations were classified into four categories based on European Environment Agency guidelines (2013/34/EU) to capture the spatial heterogeneity of air pollution exposure:

- **Urban Traffic Stations:** These stations are positioned within 10 meters of major traffic arteries with daily vehicle flows exceeding 10,000 units, representing population exposure to direct vehicular emissions.
- **Urban Background Stations:** Located in residential areas at distances greater than 50 meters from major emission sources, these stations characterize general urban air quality levels.

- **Urban Industrial Station:** Situated in proximity to industrial facilities in the southeastern district, monitoring the combined impact of industrial processes and urban emissions.
- **Suburban Background Stations:** Positioned at the urban periphery to assess background concentrations and pollution transport from the urban core.

Table 1 shows the categorization of the 11 stations used in this thesis along with the missing percentage of data in each station.

Table 1. Air quality stations in Budapest.

| Station | Type of area and station | % Missing |
|---|---|---|
| Erzsébet square | Urban Traffic | 5.50 |
| Budatétény | Suburban background | 0.74 |
| Csepel | Suburban background | 8.22 |
| Honvéd Sport complex | Urban background | 3.64 |
| Gilice square | Urban background | 4.64 |
| Gergely street | Urban industrial | 1.84 |
| Széna square | Urban traffic | 3.12 |
| Teleki László square | Urban traffic | 1.82 |
| Pesthidegkút | Urban background | 13.07 |
| Kőrakás park | Urban background | 1.11 |
| Kosztolányi D. square | Urban traffic | 3.66 |

### 3.1.2. Moroccan cities sites

Budapest was initially selected as the primary study area. However, significant computational limitations arose during the implementation of the CHIMERE chemical transport model, necessitating a change in study domain. Consequently, the research was conducted using Moroccan urban areas (Agadir and Casablanca) for which CHIMERE configurations were already operational. The methodological framework and results presented herein remain generalizable and can be readily applied to Budapest or comparable urban environments:

**Agadir, Morocco (30.4°N, 9.6°W)**: A coastal city of 600,000 inhabitants providing ideal conditions for testing CHIMERE model sensitivity to spatial resolution. The Atlantic maritime influence, year-round mild temperatures (15-25°C), and relatively simple topography create more stable atmospheric conditions compared to Budapest, allowing clear assessment of how CTM resolution affects prediction accuracy without the confounding effects of complex terrain. Data were collected from an urban monitoring station at (30.42°N, 9.60°W), approximately 2 km from the Atlantic Ocean, during March 1 to August 31 for both 2010 and 2016. These periods captured spring and summer seasonal variations in $O_3$ and $PM_{10}$ concentrations, providing sufficient temporal coverage to evaluate the CHIMERE model's performance across different resolutions (0.1°, 0.05°, and 0.02°) and seasonal conditions.

**Casablanca, Morocco (33.6°N, 7.6°W)**: Morocco's economic capital with 3.4 million inhabitants, used for validating the CHIMERE-ANN hybrid approach. The city's intensive industrial activity and dense traffic networks create pollution patterns suitable for testing whether neural networks can effectively correct systematic biases in deterministic models. The

city centre monitoring station at (33.57°N, 7.63°W) provided validation data for February 2021, capturing mixed urban emissions during winter conditions when pollution accumulation is typically enhanced. This shorter validation period was sufficient to demonstrate the effectiveness of the ANN correction approach in reducing CHIMERE's systematic overestimation of ozone concentrations.

For both cities, the limited number of monitoring stations (single stations operated by the National Meteorological Service) necessitated the use of deterministic modelling approaches. However, the availability of comprehensive CTM inputs including emission inventories (EDGAR HTAP), meteorological forcing (WRF), and boundary conditions (LMDz-INCA, GOCART) made these sites particularly suitable for testing deterministic model enhancements through the CHIMERE-ANN hybrid methodology. This approach demonstrates how cities with limited observational infrastructure can still achieve accurate air quality forecasts by combining physics-based models with machine learning bias correction techniques.

### 3.1.3. Data Collection and Parameters

For Budapest, continuous hourly measurements were collected in Budapest from January 1, 2018, to December 31, 2023, encompassing a five-year period that captures both typical meteorological patterns and exceptional events, including the COVID-19 lockdown periods that significantly altered emission patterns. The primary pollutants analysed were:

- **$PM_{10}$** (Particulate Matter ≤10 μm)
- **$NO_2$** (Nitrogen Dioxide)

Concurrent meteorological parameters were recorded at one station (47.474, 19.062), including temperature (°C), relative humidity (%), wind speed (m/s), wind direction (degrees), atmospheric pressure (hPa), and global solar radiation (W/m²), using WMO-standard instrumentation.

For Moroccan cities, the measurements were carried out by the Environment Department of Souss-Massa region and the General Directorate of Meteorology, including hourly measurements of ozone ($O_3$) and particulate matter ($PM_{10}$). The measurement operations included Agadir city for two years, 2010 and 2016; the typical period is from March 15 to August 31, and Casablanca city from February 1 to March 27, 2021. The differences between the periods studied are due to the availability of the stations, meaning the availability of the data. The study period in Casablanca was short, but it helped validate the spatial resolution effect temporal and at other sites.

### 3.1.4. Quality Assessment and Data Preprocessing

A comprehensive quality assurance protocol was implemented following EEA data quality objectives (2011/850/EU) for both study areas. The preprocessing pipeline consisted of:

1. **Primary Validation:** Range checks eliminated physically impossible values ($PM_{10}$ > 1000 μg/m³, $NO_2$ > 500 μg/m³, $O_3$ > 400 μg/m³, negative concentrations). Instrument malfunction flags and calibration periods were identified through station maintenance logs.

2. **Statistical Screening:** Outlier detection employed a three-tier approach: (i) values exceeding ±4 standard deviations from rolling 168-hour means, (ii) rates of change exceeding 200 μg/m³/h for $PM_{10}$, 150 μg/m³/h for $NO_2$, and 100 μg/m³/h for $O_3$, and (iii) persistent values (>6 consecutive identical readings) indicating sensor freezing.
3. **Temporal Consistency:** Discontinuities were identified using the Pettitt test with significance level $\alpha = 0.05$, with metadata consultation to distinguish genuine environmental changes from instrumentation artifacts.

### 3.1.5. Missing Data Imputation

Data completeness varied across stations, with missing data percentages ranging from 0.74% (Budatétény) to 13.07% (Pesthidegkút), primarily attributable to instrument maintenance, calibration procedures, and power outages. To maintain dataset integrity while preserving temporal patterns, we implemented an Expectation-Maximization (EM) algorithm for multivariate imputation.

The EM algorithm proceeded iteratively through:

**Expectation Step:** Missing values were estimated using conditional expectations based on observed data, incorporating both temporal autocorrelation and cross-variable dependencies:

$$\hat{x}_{missing} = E\left[X_{missing} | X_{observed}, \Theta^{(t)}\right] \quad (6)$$

where $\Theta^{(t)}$ represents parameter estimates at iteration $t$, updated using maximum likelihood estimation on the completed dataset.

**Maximization Step:** Model parameters (mean vectors μ, covariance matrix Σ) were re-estimated:

$$\Theta^{(t+1)} = \arg max_{\Theta} L(\Theta | X_{observed}, \hat{X}_{missing}^{(t)}) \quad (7)$$

The algorithm incorporated spatial-temporal structure by including lagged values (t-1, t-24, t-168 hours), neighbouring station measurements within 5 km radius, and meteorological covariates. Convergence was achieved when the log-likelihood change fell below 10^-6, typically within 20-30 iterations.

Imputation uncertainty was quantified through multiple imputation (m = 5), generating confidence intervals for subsequent analyses. Validation on artificially created gaps (5% random removal) demonstrated mean absolute errors of 3.2 μg/m³ for $PM_{10}$ and 4.1 μg/m³ for $NO_2$, with temporal pattern preservation confirmed through autocorrelation function comparison (mean difference < 0.02 across all lags).

The final quality-controlled dataset comprised 438,120 hourly observations per pollutant-station combination, providing robust temporal coverage (>90% completeness) for all stations except Pesthidegkút (86.93% after imputation), enabling comprehensive analysis of pollution dynamics across diverse urban environments.

For the Moroccan case studies, data availability was constrained by station operational periods. Measurements in Agadir spanned March 15 to August 31 for 2010 and 2016, while Casablanca data covered February 1 to March 27, 2021. Given the episodic nature of these measurement campaigns, quality assurance focused on validation rather than extensive imputation. Hourly measurements of ozone ($O_3$) and particulate matter ($PM_{10}$) were subjected to range checks and consistency verification against station maintenance logs. The temporal discontinuity between study periods, a consequence of station availability, was addressed by treating each campaign as an independent validation dataset for the CHIMERE model evaluation. This approach preserved data integrity while acknowledging the inherent limitations of monitoring infrastructure in developing regions.

## 3.2. ARIMA search Grid modelling

The ARIMA (p, d, q) (P, D, Q) model is applied to perform data analysis, where p, d, q and P, D, Q represent continuity and seasonal auto-regression differences, respectively. It comprises three elements: an Autoregressive (AR) model, an Integrated (I) model, and a Moving Average (MA) model. The AR and MA components are applied to differenced data to ensure stationarity. Autocorrelation function (ACF) and Partial Autocorrelation (PACF) are used to select the best values for the models' parameters (Marinov et al., 2022).

### 3.2.1. ACF and PACF function

To assess the correlation among observations in a time series, autocorrelation comes into play. This technique involves plotting the correlation coefficient on the ACF chart against the lag, which is measured in terms of time periods. The lag represents a specific point in time after the initial observation in the time series. As a result, ACF is derived from the autocovariance of $x_t$ and $x_{t-n}$ as part of this process (Morf et al., 1978).

$$ACF(n) = \frac{Cov(x_t, x_{t-n})}{Var(x_t)} \tag{8}$$

In this context, $Cov(x_t, x_{t-n})$ represents the covariance between variables $x_t$ and $x_{t-n}$, and $Var(x_t)$ indicates the variance of the variable $x_t$. Covariance gauges the connection between two random variables, while variance assesses their variability. The correlation coefficient spans from -1 (indicating a negative relationship) to 1 (indicating a positive relationship). When there is no discernible relationship between the variables, the correlation coefficient stands at 0.

Partial autocorrelation provides a condensed view of the association between a particular observation and prior observations, excluding any influences from the intermediate observations. PAC is a straightforward correlation between $x_t$ and $x_{t-n}$ and its calculation method is as follows (Akaike, 1998):

$$PACF(n) = Corr[x_t - E^*(x_t|x_{t-1}, \dots, x_{t-n+1}), x_{t-n}] \tag{9}$$

Here, $E^*(x_t|x_{t-1}, \dots, x_{t-n+1})$ represents the expected value of $x_t$ The ideal model for the time series is derived based on the findings from the ACF and PACF.

### 3.2.2. Stationarity

The requirement of stationarity is essential for most time series forecasting methods. Stationarity means that the statistical characteristics of the time series such as mean, variance, autocorrelation, etc. do not change over time. In case of missing stationarity, the time series needs to be stabilized before further analysis. A Stable History Period (SHP) refers to a specific timeframe during which the time series maintains stability, can be used to resolve such issues (Ghaderpour & Vujadinovic, 2020). The stability of the time series in this study is determined by ACF and PACF. Time series is stable, if ACF fluctuates around a fixed horizontal line with a gradual decay trend. A differential transformation can be applied to the time series Tt to produce stationary time series $S_t$ (B et al., 2021) as follows:

$$T_t = S_t - S_{t-1} \tag{10}$$

### 3.2.3. ARIMA Method

The ARIMA model utilizes an AR (AutoRegressive) model combined with an MA (Moving Average) model to conduct time series forecasting (Morf et al., 1978).

Key parameters that play a role in this process include:

The number of previous observations (denoted as 'p').

The degree of differencing (denoted as 'd').

The size of the moving average (denoted as 'q').

The AR model specifically illustrates the relationship between a current observation and an earlier time period. It does so by considering 'p' previous observations in the following manner:

$$y_t = \alpha + \sum_{i-1}^{p} \varphi_i y_{t-i} + e_t \tag{11}$$

In this context, $y_t$ represents the predicted value for time 't' drawn from a normal distribution, while $y_{t-i}$ refers to 'p' previous observations from the same time series. $\varphi_i$ signifies the regression coefficients, $\alpha$ is a constant, and $e_t$ stands for the random error term. The choice of the order 'p' for the AR(p) model is determined by identifying significant spikes in the PACF. Another indicator is the gradual decline in the ACF.

The MA model, on the other hand, conducts forecasting by relying on the moving averages of past random error terms in the following manner:

$$y_t = \mu + \sum_{i-1}^{q} \theta_i e_{t-i} \tag{12}$$

In this context, $\theta_i$ denotes the regression coefficients, 'q' signifies the moving average order, and $\mu$ represents a constant. The determination of the order 'q' for the MA(q) model is based on

the ACF, specifically looking for a distinct drop-off after 'q' lags. In such cases, the PACF exhibits a gradual decay pattern.

The ARIMA model can be represented as:

$$\varphi_p(B)(1-B)^d y_t = \theta_q(B)e_t \tag{13}$$

Here, $B$ stands for the backshift operator, 'p' indicates the autoregression order, 'd' represents the differencing order, and 'q' signifies the moving average order. The polynomials $\varphi_p(B)$ and $\theta_q(B)$ correspond to the autoregressive and moving average components, respectively.

#### 3.2.4. Evaluation Criteria

The optimal pairing of 'p' and 'q' can be determined through the utilization of an objective function that evaluates the model's performance on a validation dataset. Two such mathematical methods for scoring statistical models and selecting the most suitable one within a candidate model set are the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC). AIC assesses how well the model fits the data using the following approach (Schwarz, 1978):

$$AIC = -2l + 2k \tag{14}$$

In this context, $l$ represents a log-likelihood, and $k$ stands for the count of parameters in the candidate model.

An extra parameter, $n$ is introduced to the BIC, indicating the number of data samples employed for fitting, as described below (Bhatti et al., 2021):

$$BIC = -2l + k\ log\ n \tag{15}$$

The BIC is formulated within the framework of a Bayesian network, while the AIC is developed by adjusting biased empirical information. Generally, it tends to favour the selection of a less complex model compared to the AIC.

The coefficient of determination, also known as R-squared, quantifies the proportion of variance in the dependent variable that can be accounted for by the independent variables in the model. It is also employed to assess how well the model fits the data. R-squared ranges between 0% and 100%, where 0% indicates that the model fails to explain the variability of the response data around its mean, and 100% signifies that the model accounts for all the variability.

### 3.3. Deterministic approach

#### 3.3.1. Meteorological condition modelling

The Weather Research and Forecasting (WRF) model is a real-time numerical weather forecasting model for mesoscale meteorological systems. The University Corporation for Atmospheric Research (UCAR) provided the input data. Initial and boundary WRF conditions are taken from Global Forecast System (GFS) at 1° every 6 hours. WRF offers several physics, dynamics, and boundary layer control options available in any combination. Several physical parameters were selected, such as the RRTM, Dudhia and WRF Single Moment 3 class schemes, details of parameters used are in table 2.

Table 2. Summary of the physics and dynamics of WRF settings.

| Parameter option | Description | Reference |
|---|---|---|
| **Physics** | | |
| Microphysics | WRF Single-Moment 3-class scheme | (Hong et al., 2004) |
| Long Wave Radiation | Rapid Radiation Transfer Model (RRTM) | (Mlawer et al., 1997) |
| Shortwave Radiation | Dudhia scheme | (Dudhia, 1989) |
| Cloud fraction option | Xu-Randall method | |
| Surface layer | MM5 similarity | (Fairall et al., 2003) |
| Land surface | Noah Land Surface Model | (F. Chen et al., 2004) |
| Boundary layer | Yonsei University scheme | (Jiménez & Dudhia, 2012) |
| Cumulus option | Kain-Fritsch scheme | (Bullock et al., 2015) |
| **Dynamics** | | |
| Diffusion | Simple diffusion | (Reddy et al., 2020) |
| Eddy coefficient | Horizontal Smagorinski 1$^{st}$ order | (J. S. Simon & Chow, 2021) |

Table 3. Emissions modeling configurations.

| Parameter option | Description | Reference |
|---|---|---|
| **Anthropogenic emissions** | | |
| Chemistry mechanism | Melchior | (Menut et al., 2021b) |
| Reggriding landuse | USGS | (Brands et al., 2019) |
| vertical distribution | Mailler and Terrenoire modification | (Terrenoire et al., 2015) |
| Time distribution | Ebel guidelines | (Georgiou et al., 2018) |
| **Natural emissions** | | |
| Biogenic emissions | MEGAN model | (Guenther et al., 2012) |
| Sea salt emissions | Monahan, completed by Martensson | (ImUlas, 2013) |
| Mineral dust emissions | Menut guidelines | (Mailler et al., 2017) |
| **Boundary conditions** | | |
| Dust boundary conditions | GOCART global and LMDz-AER | (Ginoux et al., 2001) |
| Gas boundary conditions | LMDz-INCA | (Folberth et al., 2006) |

### 3.3.2. Emissions modelling

Emissions are modeled by considering two types of emissions: human-made emissions, known as anthropogenic emissions, and natural emissions, which include sea salt emissions, dust emissions, and biogenic emissions. The Emi-Surf model version 2016b is used for computing anthropogenic emissions (Russo et al., 2019). The main steps for this model are (1) Projecting

annual masses from the user database to the Chemistry Transport Model (CTM) grid (2) Converting annual totals into hourly fluxes used as input to the CTM model by prescribing monthly, weekly, and hourly profiles, and (3) Converting the data available in the original dataset into model species. The emissions dataset used the EDGAR for HTAP 2010 v2 global emissions inventory, which received some updates in 2014 (Shami et al., 2022). This emission data (CO, $NH_3$, VOCs, $NO_x$, $SO_2$ and $PM_{10}$) is geo-referenced in a horizontal resolution grid of 0.1° × 0.1° and they are collected according to the pollution-producing sectors. Table 3 presents emissions modeling configurations.

### 3.3.3. Chemical transport modelling

CHIMERE is a multi-scale Eulerian Chemistry Transport Model (CTM), designed to generate daily forecasts of pollutant concentration based on internal and external forcing. The internal forcing includes physical processes such as horizontal resolution, advection, deep convection, and boundary-layer turbulence, and chemical processes such as gas-phase chemistry, aerosol chemistry, depositions, and photolytic rates. The external includes boundary conditions input and meteorological fields such as Short-Wave Radiation, relative humidity, temperature, and wind speed and emissions input such as anthropogenic, biogenic, mineral dust, and sea salt emission. In this study, CHIMERE version 2017r3 is used. The Gas-phase chemistry mechanism selected is the reduced MELCHIOR mechanism called MELCHIOR2, including about 120 reactions and 44 species. CHIMERE is configured with 20 vertical levels from surface up to 200 hPa. The photolysis rates are calculated using the radiative transfer model FastJX radiation module. All CHIMERE simulations were initialized with a 15-day spin-up period to reduce the impact of initial conditions and allow chemical and meteorological fields to reach equilibrium. This duration was chosen following preliminary sensitivity analyses and established practices in regional CTM studies, which indicate that approximately two weeks are sufficient for stabilizing major atmospheric constituents before the start of the evaluation period. Table 4 summarizes CHIMERE configurations.

Table 4. CHIMERE settings summary.

| Parameter option | Description | Reference |
|---|---|---|
| Numerical solver | TwoStep solver | (Verwer et al., 1996) |
| Chemistry Options | Reduced mechanism MELCHIOR 2 | (Menut et al., 2013c) |
| Deep convection fluxes | Tiedtke scheme | (Tiedtke, 1989) |
| Advection | Upwind scheme | (W. Zhang et al., 2017) |
| Photolytic rates | Radiative transfer model | (Mlawer et al., 1997) |
| Boundary-layer turbulence | Troen and Mahrt scheme | (Miller et al., 1996) |
| Soil moisture | Fecan scheme | (Fécan et al., 1998) |
| Dry threshold friction velocity | Shao and Lu scheme | |

### 3.3.4. Artificial neural network Modelling

An artificial neural network (ANN) is a modeling technique derived from the human neural system which provides a way to learn by example from representative data described by a physical phenomenon or a decision-making procedure (Moayedi et al., 2020). Once a neural network is created, it is set up and trained. Configuration includes structuring the network to be consistent with the problem of interest, as defined by sample data. After the network configuration, the adjustable network parameters, called weights and biases, need configuring to optimize network performance. This adjustment process is called network training, which involves establishing setup and training requirements to ensure that the network is provided with sufficient data examples. ANN has formed by input and output node layers, connected by one or more hidden node layers (Maleki et al., 2019). The hidden layers implement weighting functions on the incoming feeds; if a value for a particular node or set of nodes in the hidden layer achieves a certain threshold, a value is transmitted to one or more nodes in the outgoing layer. A typical structure of the ANN is indicated in Fig. S1. In our case, we use four hidden layers with ten neurons for each hidden layer (1-10-10-10-10-1). A Levenberg-Marquardt algorithm is adopted to reduce the Sum of Squares error functions (Bergou et al., 2020). We used the Sigmoid function activation in the hidden and the output layer (Roodschild et al., 2020). The training data included ozone concentration of two periods (Table 2), the first from March 15 to May 21st, 2016, representing 68 days equal to 87% of the data, for Agadir, and the second from February 1st to February 10th, 2021, including 240 hours equal to 90% of the data for Casablanca. The remaining 13%, equal to 10 days and 10% representing 24 hours, were used as testing data for Agadir and Casablanca, respectively.

### 3.3.5. Technical concept

Figure 6 presents a technical concept for this study.

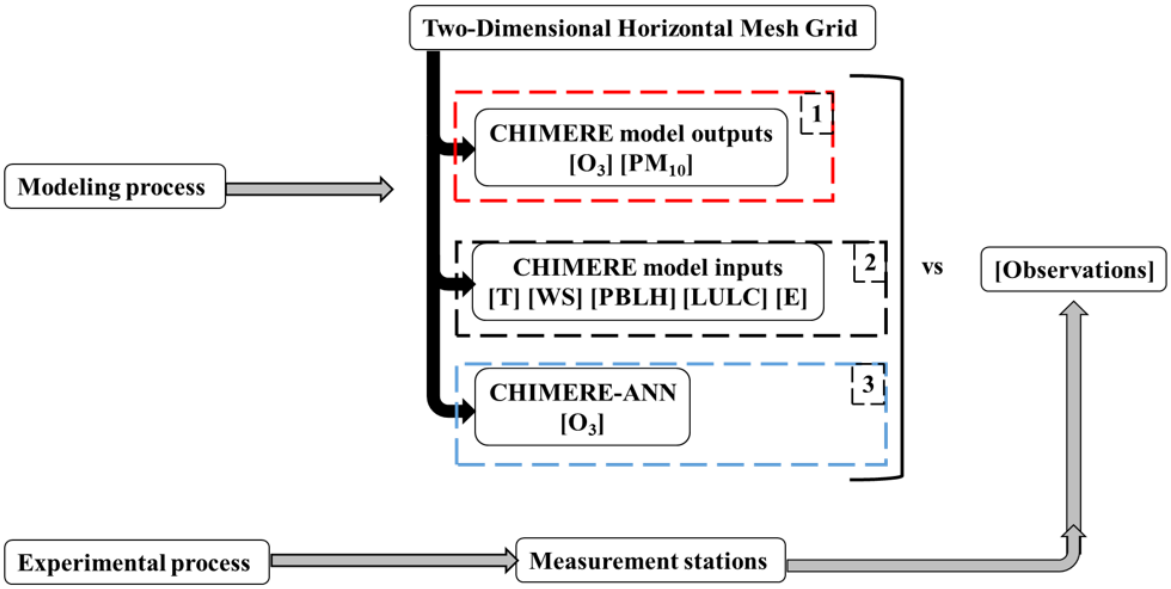



**Figure 6. Technical concept of the study, Aspect 1 (Block 1), Aspect 2 (Block 2), Aspect 3 (Block 3).**

The operational process is composed of three steps:

1. After setting up modeling and observation processes, firstly, we tested the spatial resolution effect on two CHIMERE outputs $O_3$, and $PM_{10}$
2. As WRF model, USGS model, and Emi-Surf model outputs are used as CHIMERE inputs, we identify spatial resolution impact on five inputs, namely Temperature (T) and Wind Speed (WS), Planetary Boundary Layer Height (PBLH), Land Use and Land Cover (LULC), and Emissions (E).
3. Then we corrected the Ozone CHIMERE model using the ANN model to determine the effect of CHIMERE resolution on the CHIMERE-ANN as a correction application.

## 3.4. Machine Learning with Feature engineering

### 3.4.1. Features Engineering

The limitations observed in both statistical and deterministic-hybrid approaches reveal a fundamental insight: the raw measurements and standard meteorological variables that serve as typical model inputs may not optimally represent the underlying atmospheric processes governing pollutant concentrations. While ARIMA captured temporal dependencies through lagged observations and CHIMERE-ANN learned systematic bias patterns, neither approach explicitly encoded the diverse physical mechanisms, traffic rush-hour accumulation, boundary layer collapse, weekend emission reductions, or anomalous meteorological events, that drive pollution dynamics at different temporal scales. This recognition motivates a comprehensive feature engineering strategy designed to transform raw observations into representations that align with specific atmospheric processes. Rather than expecting machine learning algorithms to discover these patterns from unstructured data, we hypothesize that carefully crafted features targeting distinct physical phenomena will enable models to specialize in specific aspects of air quality prediction. By developing multiple feature sets that emphasize different temporal scales (hourly to weekly), physical processes (dispersion versus accumulation), and system states (normal versus anomalous), we create the foundation for an ensemble where each model becomes an expert in particular atmospheric condition. This approach acknowledges that the superiority of any given model architecture may depend less on its mathematical sophistication than on whether its input features appropriately represent the dominant processes during specific prediction contexts. To capture the multi-scale temporal dynamics and heterogeneous physical processes governing $PM_{10}$ concentrations, we designed a feature engineering strategy that constructs four complementary feature sets. Each set was developed to emphasize distinct aspects of pollution behaviour, promoting model specialization and ensuring diverse error structures suitable for ensemble fusion.

#### 3.4.1.1. Short-Term dynamics features

This feature set comprised 11 variables targeting immediate temporal dependencies and rapid transitions characteristic of traffic-induced variations and short-term meteorological impacts. The set included $PM_{10}$ lags at t-1, t-2, and t-3 hours, with temporal differences computed as:

$$\Delta_1(t) = y_{t-1} - y_{t-2}, \qquad \Delta_2(t) = y_{t-2} - y_{t-3} \tag{16}$$

Wind components were encoded to preserve directional continuity:

$$WD_{sin}(t) = \sin\big(\theta_{WD}(t)\big), \qquad WD_{cos}(t) = \cos\left(\theta_{WD}(t)\right) \tag{17}$$

Where $\theta_{WD}$ is wind direction in radians. Hourly cyclical patterns were captured through:

$$h_{sin}(t) = \sin\left(2\pi.\frac{hour(t)}{24}\right), h_{cos}(t) = \cos\left(2\pi.\frac{hour(t)}{24}\right) \tag{18}$$

#### 3.4.1.2. Long-Term pattern features

This set incorporated variables operating at extended temporal scales to detect weekly cycles, seasonal trends, and persistent atmospheric patterns. It included $PM_{10}$ lags at 24, 48, and 168 hours, along with rolling statistics computed using past-only windows:

$$\bar{y}_w(t) = \frac{1}{w}\sum_{i=1}^{w} y_{t-i}\,, \qquad \sigma_w(t) = \sqrt{1/w\sum_{i=1}^{w}(y_{t-i} - \bar{y}_w(t))^2} \tag{19}$$

Where $w \in \{24{,}168\}$ hours with minimum observations $n_{min} = \max(3, w/3)$.

Monthly seasonality was encoded as:

$$m\_sin\ (t) = \sin\left(2\pi.month(t)/12\right)\ , \qquad m\_cos\ (t) = \cos\left(2\pi.month(t)/12\right) \tag{20}$$

Baseline meteorological variables (temperature, relative humidity, global radiation) were included to capture seasonal atmospheric conditions.

#### 3.4.1.3. Meteorological Driver features

This set emphasized atmospheric dispersion mechanisms through contemporaneous and lagged (6h, 12h) meteorological variables. A pressure proxy indicator was defined as:

$$P_{proxy}(t) = \frac{T(t)}{RH(T) + \varepsilon} \tag{21}$$

Where $T$ is temperature ($°C$), $RH$ is relative humidity (%), and $\varepsilon = 10^{-6}$. This proxy serves as an indicator of atmospheric stability conditions affecting vertical mixing. Decomposed wind vectors and lagged meteorological features accounted for the delayed impact of atmospheric conditions on pollutant accumulation.

#### 3.4.1.4. Anomaly Detection features

To enhance model robustness during extreme events, this set quantified deviations from expected patterns. Standardized z-scores were computed as:

$$z(t) = \frac{y_t - \mu_{exp}(t-1)}{\sigma_{exp}(t-1)} \tag{22}$$

where $\mu_{exp}\ and\ \sigma_{exp}$ are expanding mean and standard deviation from all historical values up to $t-1$.

Deviations from periodic patterns were calculated as:

$$\delta_{daily}(t) = y_t - \bar{y}_h(t)\,, \qquad \delta_{weekly}(t) = y_t - \bar{y}_{168}(t-1) \tag{23}$$

Binary indicators flagged unusual conditions including nocturnal low-wind events (WS < 2 m/s, 00:00-06:00) and high-temperature-low-wind combinations exceeding the 90th percentile. To prevent data leakage, all temporal features were computed using strict historical information with shift (1) operations excluding current observations. Table 5 summarizes the composition and characteristics of the four feature sets, demonstrating how each target specific physical processes: traffic-induced immediate dispersion, weekly cycles and seasonal trends, boundary layer dynamics, and extreme events.

Table 5. Feature set composition and characteristics.

| Feature Set | Variables | Temporal scale (h) | Physical Process | Number of features |
|---|---|---|---|---|
| **Short-term dynamics** | $PM_{10}$ lag {1,2,3}, $\Delta_1, \Delta_2$, WS, WD_sin, WD_cos, hour_sin, hour_cos | 1-3 | Traffic immediate dispersion | 11 |
| **Long-term patterns** | $PM_{10}$ lag {24,48,168} Rolling_mean {24,48,168} Rolling_std {24,168} Month_sin, Month_cos, T, RH, GRad | 24-168 | Weekly cycles, seasonal trends | 12 |
| **Meteorological drivers** | T, RH, WS, GRad, WD_sin, WD_cos, T_lag {6,12}, RH_lag {6,12}, WS_lag {6,12}, P_proxy | 0-12 | Boundary layer, atmospheric stability | 14 |
| **Anomaly detection** | z-score, δ_daily, δ_weekly, unusual_hour, unusual_weather | Multi-scale | Extreme events, outliers | 5 |

### 3.4.2. Regime Identification

To enable conditional model specialization, we identified six distinct pollution regimes based on concentration variability and temporal patterns. Regime boundaries were established using training set quantiles applied to past-only signals:

$$R_{stable}(t) = 1\left[\sigma_6(t-1) < Q_{30}(\sigma_6^{train})\right] \tag{24}$$

$$R_{rising}(t) = 1\left[\Delta_3(t-1) > Q_{75}(\Delta_3^{train})\right] \tag{25}$$

$$R_{falling}(t) = 1\left[\Delta_3(t-1) < Q_{25}(\Delta_3^{train})\right] \tag{26}$$

$$R_{extreme}(t) = 1\left[y_{t-1} > Q_{90}(y^{train})\right] \tag{27}$$

where $\Delta_3(t) = y_t - y_{t-3}$ $and$ $\sigma_6$ is the 6-hour rolling standard deviation, and $1[.]$ is the indicator function.

Temporal regimes captured diurnal patterns:

$$R_{Rush}(t) = 1[hour(t) \in \{7,8,9\}] \tag{28}$$

$$R_{nocturnal}(t) = 1\left[hour(t) \in \{0,\dots,5\} \wedge WS_{t-1} < Q_{40}(WS^{train})\right] \tag{29}$$

Each regime was paired with feature sets aligned to its dominant physical processes: short-term features for rapid transitions (rising, falling, morning rush), long-term features for stable conditions, meteorological features for dispersion-dominated periods, and anomaly features for extreme events. This regime-based approach acknowledges the non-stationary nature of $PM_{10}$ dynamics, enabling models to develop specialized expertise for specific pollution behaviours.

The multi-feature, multi-regime framework ensures that individual models capture distinct aspects of $PM_{10}$ dynamics, creating complementary prediction patterns amenable to various fusion strategies including weighted averaging, stacking, Bayesian model averaging, and Choquet integral aggregation. The diversity in feature representations and regime specialization promotes ensemble robustness across varying atmospheric conditions and pollution scenarios.

### 3.4.3. Machine Learning models

#### 3.4.3.1. Random Forest Regressor (RF)

Random Forest models (Breiman, 2001b) were configured with 400 trees and adaptive maximum depth based on regime specialization. For high-pollution and transition regimes, no depth constraint was imposed to capture complex non-linear relationships, while stable conditions employed depth limitation (max_depth=10) to prevent overfitting during low-variability periods. Minimum samples per leaf varied between 2 for transition detection and 5 for stable conditions. The bootstrap aggregation mechanism proved particularly effective for $PM_{10}$ prediction, with out-of-bag error estimates indicating optimal forest size at 400 trees (convergence achieved at 350 trees with <1% improvement thereafter).

Asymmetric loss variants were implemented using sample weighting:

$$w_i = \exp((y_i - \mu_y)/\sigma_y) \text{ for underestimation-averse models}$$

$$w_i = \exp(-(y_i - \mu_y)/\sigma_y) \text{ for overestimation-averse variants}$$

where $\mu_y \; and \; \sigma_y$ represent training sets mean and standard deviation.

This weighting scheme penalizes prediction errors asymmetrically, with underpredict-averse models assigning exponentially higher weights to high-concentration samples. Node splitting utilized Gini impurity with bootstrap sampling, while out-of-bag error estimation provided internal validation without requiring a separate validation set. Feature importance was calculated through mean decrease in impurity, weighted by the probability of reaching each node and the number of samples affected.

#### 3.4.3.2. Gradient Boosting Regressor (GBR)

Two distinct Gradient Boosting configurations (Friedman, 2001) were implemented targeting different temporal dynamics. For stable conditions, we employed a conservative architecture with 500 trees, learning rate $\eta = 0.03$, maximum depth of 3, and subsample ratio of 0.9. This configuration prioritizes gradual refinement over aggressive fitting, suitable for capturing smooth temporal transitions. The loss function minimization follows:

$$F_m(x) = F_{m-1}(x) + \eta . h_m(x) \quad (25)$$

Where $h_m$represents the m-th weak learner fitted to the negative gradient of the loss function. For regime-specific models, we utilized a more aggressive configuration with 300 trees, η = 0.05, and maximum depth of 4, enabling faster adaptation to regime-specific patterns.

Robustness was enhanced through Huber loss for outlier resistance when $|y_i - \hat{y}_i|$ exceeded 1.35σ, transitioning from squared to linear loss. Feature subsampling (0.8) at each split introduced stochasticity to reduce overfitting. Early stopping with a patience of 50 iterations prevented overspecialization to training data, triggered when validation loss failed to improve by more than $10^{-4}$.

#### 3.4.3.3. Support Vector Machine (SVM)

Support Vector Regression (Schölkopf, 1998) with radial basis function (RBF) kernels were deployed for meteorological feature sets, leveraging its effectiveness in high-dimensional spaces with complex non-linear relationships. The optimization problem was formulated as:

$$min\frac{1}{2}\ \|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{26}$$

Subject to:

$$y_i - \langle w, \phi(x_i)\rangle - b \leq \varepsilon + \xi_i \tag{27}$$

$$\langle w, \phi(x_i)\rangle + b - \ y_i \leq \varepsilon + \xi_i^* \tag{28}$$

Where $\phi$ maps input to a high-dimensional feature space via the RBF kernel $k(x, x') = \exp(-\gamma\|x - x'\|^2)$. The regularization parameter C = 10.0 balanced model complexity with training error, while $\gamma = \frac{1}{dVAR(X)}$ adapted to feature scaling. The ε-insensitive tube width was set to 0.1, permitting small prediction deviations without penalty. Feature standardization preceded kernel computation to ensure equal contribution across meteorological variables with different units.

#### 3.4.3.4. K-Nearest Neighbours (KNN)

KNN regression (Guo et al., 2003) with k = 15 neighbours and distance-weighted voting was employed for anomaly feature sets, exploiting local similarity in unusual conditions. The prediction followed:

$$\hat{y}(x) = \frac{\sum_{i\in N_k(x)} w_i . y_i}{\sum_{i\in N_k(x)} w_i} \tag{29}$$

Where $w_i = \frac{1}{d(x,x_i)}$ $and$ $N_k(x)$ represents the k-nearest neighbours of x.

Distance calculations used Minkowski metric with p = 2 (Euclidean), after robust scaling to handle outliers. The relatively large k value provided smoothing over local noise while maintaining responsiveness to anomaly patterns. Leaf size of 30 optimized tree construction for the Ball Tree algorithm, balancing query speed with construction time.

#### 3.4.3.5. Long-Short Term Memory (LSTM)

Multiple LSTM architectures were developed to capture temporal dependencies at varying scales. The core LSTM cell computations followed (Hochreiter & Schmidhuber, 1997):

$$f_t = \sigma(w_f.[h_{t-1}, x_t] + b_f) \tag{30}$$

$$i_t = \sigma(w_i.[h_{t-1}, x_t] + b_i) \tag{31}$$

$$\tilde{C}_t = \tanh(w_C.[h_{t-1}, x_t] + b_C) \tag{32}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{33}$$

$$h_t = o_t * \tanh(C_t) \tag{34}$$

where $\sigma$ denotes the sigmoid activation, and $w, b$ represents weight matrices and bias vectors respectively.

Architecture configurations were specialized for different temporal patterns:

- Short transitions: lookback = 12 hours, 64 LSTM units, learning rate = $2\times10^{-3}$
- Long patterns: lookback = 168 hours, 128 LSTM units, learning rate = $5\times10^{-4}$
- Multivariate: lookback = 24 hours, 96 LSTM units, features = [$PM_{10}$, T, RH, WS]
- Balanced baseline: lookback = 24 hours, 96 LSTM units, learning rate = $1\times10^{-3}$

Each architecture incorporated dropout (p = 0.2) after the LSTM layer for regularization, followed by a dense layer with 32 ReLU-activated units. For asymmetric variants, we implemented custom loss functions:

$$L_{asym}(y, \hat{y}, \lambda_u, \lambda_o) = \lambda_u(y - \hat{y})^2{}_+ + \lambda_o(y - \hat{y})^2{}_- \tag{35}$$

Where $(z)_+ = \max(z, 0)\ and\ (z)_- = \max(-z, 0)$

Training employed Adam optimization with early stopping (patience = 5 epochs) and learning rate reduction (factor = 0.5, patience = 3) based on validation loss. Input sequences were standardized using training set statistics, with separate scalers for features and targets to preserve scale relationships. Sequence generation used sliding windows with single-step advancement, ensuring maximum temporal coverage while maintaining causal consistency. Validation split of 10% or minimum 50 samples prevented overfitting while ensuring sufficient training data.

The ensemble of specialized LSTM variants captured complementary temporal patterns: short-transition models excelled at sudden changes, long-pattern variants identified weekly cycles, while multivariate configurations leveraged cross-variable dependencies during complex atmospheric conditions. Table 6 summarizes the final configurations and hyperparameters for all 11 models, determined through extensive grid search and cross-validation.

Table 6. Feature set composition and characteristics.

| Model Type | Configuration | Key Parameters | Feature Set | Target Regime |
|---|---|---|---|---|
| | | **Random Forest** | | |
| **RF-Standard** | 400 trees | Max depth=10<br>Min samples leaf=5 | Short-Term | All conditions |
| **RF-Underpredict Averse** | 300 Trees | Asymmetric weighting: $\exp(y - \mu)/\sigma)$ | Short-Term | Extreme events |
| **RF-Overpredict Averse** | 300 Trees | Asymmetric weighting: $\exp(-(y - \mu)/\sigma)$ | Short-Term | Stable Conditions |
| | | **Gradient Boosting** | | |
| **GBR-Stable** | 500 Trees | Learning rate = 0.03<br>Max depth=3<br>Subsample=0.9 | Long-Term | Stable regime |
| **GBR-Regime** | 300 Trees | Learning rate = 0.05<br>Max depth =4 | Varied | Regime-specific |
| | | **Support Vector Machine** | | |
| **SVR-RBF** | RBF Kernel | C=10<br>$\gamma = scale$<br>$\varepsilon = 0.1$ | Meteorological | All conditions |
| | | **K-Nearest Neighbors** | | |
| **KNN-Anomaly** | K=15 | Weights = distance<br>Metric = Euclidean | Anomaly | Unusual events |
| | | **LSTM Networks** | | |
| **LSTM-Short** | 64 units | Lookback = 12h,<br>dropout = 0.2,<br>lr = $2e^{-3}$ | $PM_{10}$ only | Transitions |
| **LSTM-Long** | 128 units | Lookback=168h,<br>dropout=0.2,<br>lr =$5e^{-4}$ | $PM_{10}$ only | Weekly patterns |
| **LSTM-Multivariate** | 96 units | lookback=24h,<br>dropout=0.2,<br>lr =$1e^{-3}$ | $PM_{10}$, T, RH, WS | All conditions |
| **LSTM-Balanced** | 96 units | lookback=24h,<br>dropout=0.2,<br>lr =$1e^{-3}$ | $PM_{10}$ only | Baseline |

## 3.5. Ensemble fusion methods

The heterogeneous nature of our expert models, spanning tree-based algorithms, neural networks, and kernel methods with diverse feature specializations, necessitate sophisticated fusion strategies to optimally combine their predictions. While individual models capture specific aspects of $PM_{10}$ dynamics, their complementary strengths and varying error patterns suggest potential for improved performance through ensemble aggregation. To comprehensively evaluate the proposed Choquet integral fusion approach, we implemented a spectrum of aggregation methods representing current best practices in ensemble learning. Our fusion framework encompasses three categories of increasing complexity:

*(i)* Baseline aggregators (mean, median) that require no training but provide robust performance benchmarks.

*(ii)* Linear combination methods including Bayesian Model Averaging (BMA), which has demonstrated success in meteorological applications (Kim et al., 2020b)

*(iii)* Stacking with meta-learning, which has achieved state-of-the-art performance in recent air quality studies (Tian et al., 2024)

*(iv)* Choquet integral, a fuzzy measure-based aggregator that uniquely captures both importance weights and interaction effects between models.

The selection of comparison methods was motivated by their proven effectiveness in environmental prediction tasks. Stacking has shown 15-25% improvement over individual models in $PM_{2.5}$ forecasting (Özüpak et al., 2025), while BMA provides probabilistically principled weight assignment with demonstrated robustness to model uncertainty. These methods, however, assume either linear relationships (BMA) or learn purely empirical combinations (stacking) without explicitly modelling inter-model dependencies. The Choquet integral addresses this limitation by incorporating a mathematical framework for synergy and redundancy, potentially offering superior performance when expert models exhibit complex interaction patterns. All fusion methods were evaluated under identical conditions: 30% of data for calibration, consistent expert model pools, and standardized preprocessing. This controlled comparison enables rigorous assessment of each method's ability to exploit the complementary information encoded across our specialized expert models.

### 3.5.1. Baseline aggregation methods

Simple aggregation methods provided performance benchmarks. The arithmetic mean aggregator computed:

$$\hat{y}_{mean}(t) = \frac{1}{M}\sum_{m=1}^{M}\hat{y}_m(t) \tag{36}$$

Where M denotes the number of valid predictions at time t. The median aggregator provided robust central tendency estimation resistant to outlier predictions. Both methods required no training and served as lower bounds for fusion performance.

### 3.5.2. Bayesian Model Averaging (BMA)

BMA weights expert predictions based on their posterior probability given the calibration data. The BIC-based weights were computed as:

$$w_k = \frac{\exp\left(-\frac{1}{2}\,BIC_k\right)}{\sum_{j=1}^{M}\exp\left(-\frac{1}{2}\,BIC_j\right)} \tag{37}$$

Where the Bayesian Information Criterion for model k is:

$$BIC_k = -2\ln(L_k) + p_k\ln(n) \tag{38}$$

With $L_k$ representing the likelihood under Gaussian residuals assumption, $p_k = 1$ (single parameter per model), and $n$ the calibration sample size. The fused prediction follows:

$$\hat{y}_{BMA}(t) = \sum_{k=1}^{M} w_k\hat{y}_k(t) \tag{39}$$

This approach naturally penalizes model complexity while rewarding predictive accuracy, providing probabilistically principled weight assignment.

### 3.5.3. Stacking ensemble with Meta-Learning

Stacking employed a two-level architecture where a meta-learner combined base model predictions. Using 5-fold time series cross-validation, we generated out-of-fold predictions to train the meta-model while avoiding data leakage:

$$\hat{y}_{stack}(t) = f_{meta}(\hat{y}_1(t), \hat{y}_2(t), \dots, \hat{y}_M(t)) \tag{40}$$

Three meta-learners were evaluated: Ridge regression with cross-validated $\alpha \in \{0.01, 0.1, 1.0, 10.0\}$, Elastic Net with $\alpha = 0.01$ and $l_1$-ratio = 0.5, and Gradient Boosting with 100 trees and maximum depth of 3. The optimal meta-learner was selected based on hold-out validation performance. Robust scaling preceded meta-learning to handle heterogeneous prediction scales.

### 3.5.4. Choquet Integral Fusion

The Choquet integral provides a powerful non-linear aggregation framework that captures both individual model importance and their interactions. Unlike traditional weighted averaging, it models complementarity and redundancy between experts through a fuzzy measure.

For a set of M expert models N = {1, 2, ..., M}, the Choquet integral with respect to fuzzy measure μ is defined as:

$$C_\mu(x) = \sum_{i=1}^{M} x_{(i)}.\left[\mu\left(A_{(i)}\right) - \mu\left(A_{(i+1)}\right)\right] \tag{41}$$

Where $(\cdot)$ indicates a permutation such that $x_{(1)} \le x_{(2)} \le \cdots \le x_{(M)}, A_{(i)} = \{(i), (i+1), \dots, (m)\}\ and\ A_{(M+1)} = \emptyset$

To maintain tractability while capturing interactions, we employed the 2-additive Choquet integral using the Möbius transform representation:

$$\mu(S) = \sum_{T \subseteq S} m(T) \tag{42}$$

where the Möbius mass m is restricted to:

Singletons: $m(\{i\})$ representing individual importance

Pairs: $m(\{i, j\})$ representing pairwise interactions

Empty set and larger subsets: $m(T) = 0\ for\ |T| > 2$

The Choquet integral then simplifies to:

$$C_\mu(x) = \sum_{i=1}^{M} x_i\, m(\{i\}) + \sum_{\{i,j\} \subseteq N} m(\{i,j\})\ \min\ (x_i, x_j) \tag{43}$$

To ensure the fuzzy measure remains monotonic (adding experts never decreases the measure), we impose:

$$m(\{i\}) \geq 0, \forall i \in N \tag{44}$$

$$m(\{i\}) + m(\{i,j\}) \geq 0, \qquad \forall i,j \in N, i \neq j \tag{45}$$

Additionally, the normalization constraint ensures $\sum_{T \subseteq N} m(T) = 1$

The Möbius coefficients were learned by minimizing the mean squared error on calibration data:

$$min_m \frac{1}{n_{cal}} \sum_{t=1}^{n_{cal}} (y_t - C_\mu(x_t))^2 \tag{46}$$

Subject to monotonicity and normalization constraints. We employed two optimization strategies:

- COBYLA (Constrained Optimization BY Linear Approximations): A derivative-free method suitable for constrained optimization, with maximum iterations set to 2000.
- Differential Evolution: A global optimization method with population-based search, using bounds [0,1] for singletons and [-0.5, 0.5] for pairs.

The optimization was initialized with equal singleton weights $m(\{i\}) = 1/M$ and zero pairwise interactions. To balance model diversity with quality, we evaluated Choquet integrals using the k-best experts based on calibration RMSE, with k ∈ {3, 5, 7, 10, all}. This approach prevents dilution from poorly performing models while maintaining sufficient diversity.

The Shapley value provides a game-theoretical interpretation of each expert's contribution:

$$\phi_i = m(\{i\}) + \frac{1}{2} \sum_{j \neq i} m(\{i,j\}) \tag{47}$$

Representing the average marginal contribution of expert $i$ across all possible coalitions.

The interaction between experts $i$ and $j$ was also assessed and is directly given by the Möbius mass:

- $m(\{i,j\}) > 0$: Synergistic interaction (complementary expertise)
- $m(\{i,j\}) < 0$: Redundancy (overlapping information)
- $m(\{i,j\}) = 0$: Independent contributions

# 4 RESULTS AND DISCUSSION

## 4.1. Trend analysis of $PM_{10}$ and $NO_2$ concentrations in Budapest

$PM_{10}$ concentrations in all air quality stations in Budapest were monitored during the period 2018-2022 in order to show the annual average trends of $PM_{10}$ and $NO_2$ in Central Eastern Europe. Datasets were collected during 2018-2022 and compared with the recommended national limits of 50 µg/m3. Figure 7 shows descending temporal trends of the $PM_{10}$ and $NO_2$ concentrations from 2018 to 2022 due to the natural gas program (Bozó & Baranka, 1996). The typical trend of annual changes in $PM_{10}$ values of Széna square has the highest concentration of 42 µg/m3 compared to the other stations studied in this work. The fact that Széna square is located between two hills of 200m elevation prevents the distribution of air pollution originating from heating, industry, and motor vehicles in the atmosphere. Due to the low air circulation in the city centre, especially in winter, the polluting parameters were concentrated in the low areas of the city, increasing the concentration of PM. Furthermore, rapid population growth, improper city management, and an increased number of cars caused $PM_{10}$ and $NO_2$ pollution in the city. Budatétény and Pesthidegkút stations, which showed an ascending trend even their values of 24 µg/m$^3$ and 28 µg/m$^3$ respectively, is the part of city with the highest use of biomass heating and located behind. Due to the city's topographic structure, removing the air pollutants formed by air circulation is quite difficult, which increases the PM value especially in areas where biomass burning is highly used for heating. The annual level of $NO_2$ from 2018 to 2022 exceeded the national limit value in Széna square station with 15 µg/m$^3$.

Notably, the impact of the lockdown from mid-March to May 2020 is discernible in $PM_{10}$ data, as all stations registered reduced values during that period (Salma et al., 2020). Considering meteorological and topographical factors, it was observed that the increase in fuel use and the decrease in the wind speed, significantly due to the decrease in temperature in winter months, triggered the increase in PM (Ferenczi et al., 2021). It is obvious that due to the inversions that occur in winter, transported dust, smoke, and soot accumulated and contributed to the increase in pollution levels in the city (Neckel et al., 2023). Despite a general reduction in air pollution trends, specific stations continue to exhibit higher levels compared to others. This phenomenon underscores the significance of the station's air quality context. For instance, urban traffic monitoring stations consistently record greater levels of $NO_2$ than industrial stations (Aykaç Özen & Öbekcan, 2023).

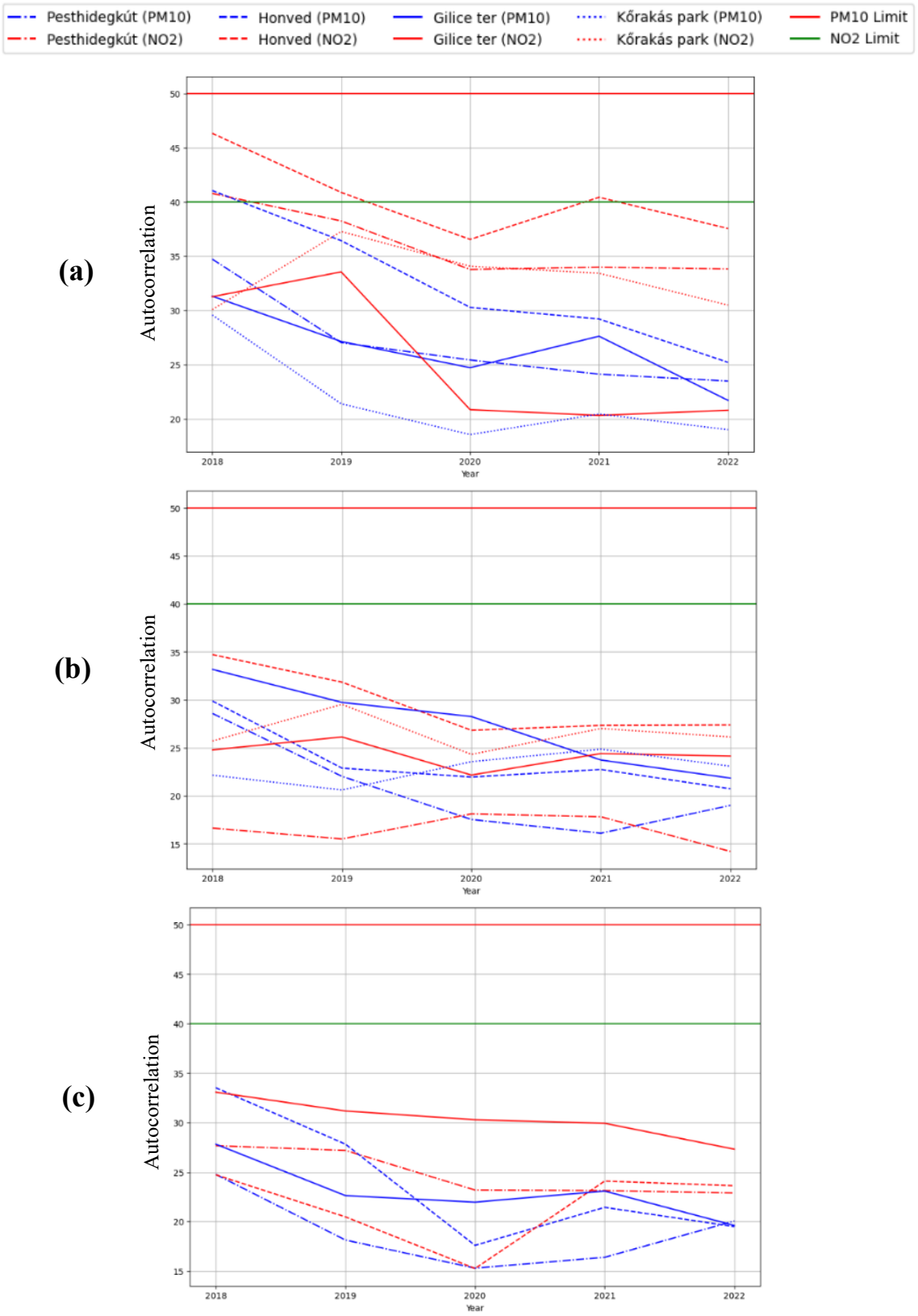
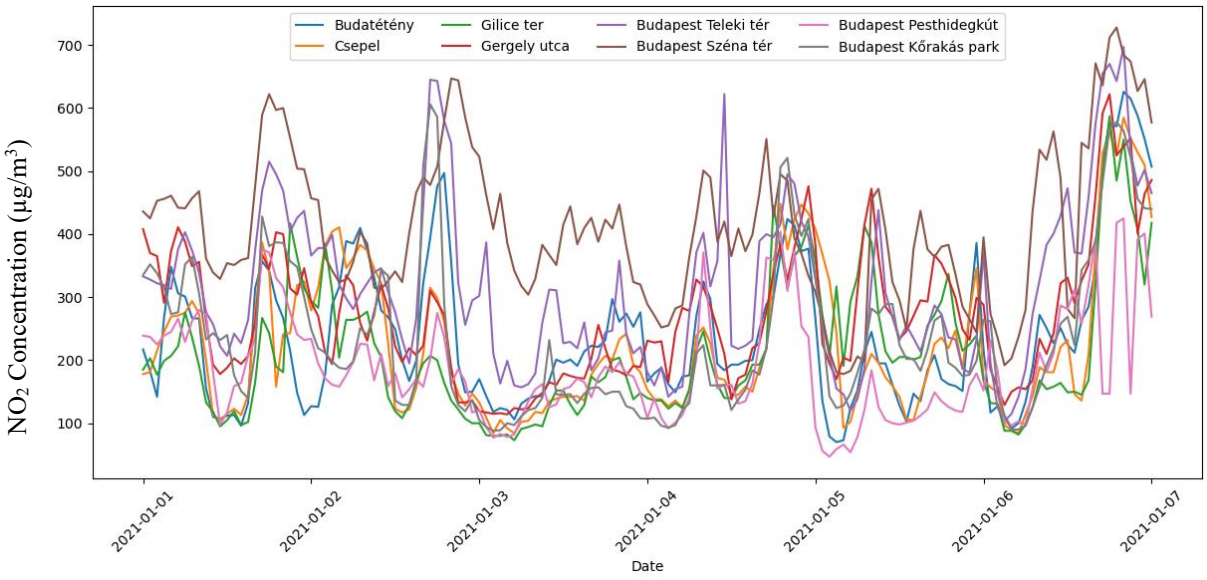



**Figure 7. Average $PM_{10}$ and $NO_2$ concentration trends for period from 2018 to 2022 in (a) Urban Traffic (b) Urban Background (c) Suburban and industrial background air quality station.**

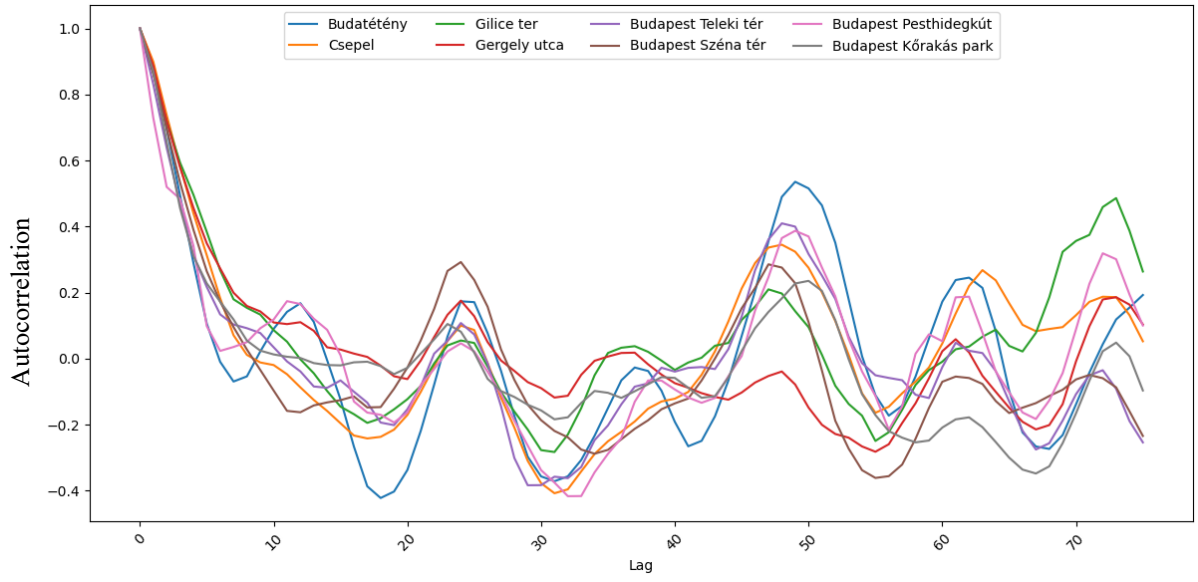
## 4.2. ARIMA Analysis

### 4.2.1. Autocorrelation and Partial Autocorrelation

To keep it concise, we've displayed plots for $NO_2$, but it's important to note that similar patterns can be observed for the other pollutants as well. Figures 8-10 represent the signal of $NO_2$ concentration resampled by 1h, 3h and 12h during the period 01-01-2021 to 07-01-2021, this period is chosen only to better representation and observation.




**(a)**




**(b)**

**Figure 8. The signal representing $NO_2$ concentration measured over 1-hour period (a), along with its corresponding autocorrelation (b).**

The autocorrelation analysis of $NO_2$ measurements at different temporal resolutions has revealed intriguing insights into the underlying seasonal patterns within the data. In the 1-hour interval plots, significant peaks at lags 24 and 48 were observed, indicating a strong daily

seasonality. This suggests a recurring pattern in $NO_2$ concentrations every 24 hours, likely influenced by daily traffic patterns and human activities. In contrast, the 3-hour interval plots displayed significant peaks at lags 8 and 16, signifying shorter-term sub-daily seasonality. These peaks suggest that $NO_2$ measurements exhibit repetitive fluctuations approximately every 8 hours, capturing more frequent variations in pollutant levels. The presence of multiple seasonal patterns at varying time scales underscores the complexity of the data and highlights the importance of considering these temporal dependencies when modelling and interpreting air quality trends. To address these distinct seasonal components effectively, utilizing a SARIMA model and examining the partial autocorrelation function (PACF) plots for both intervals can aid in selecting appropriate model orders and gaining deeper insights into the underlying dynamics of $NO_2$ concentrations.

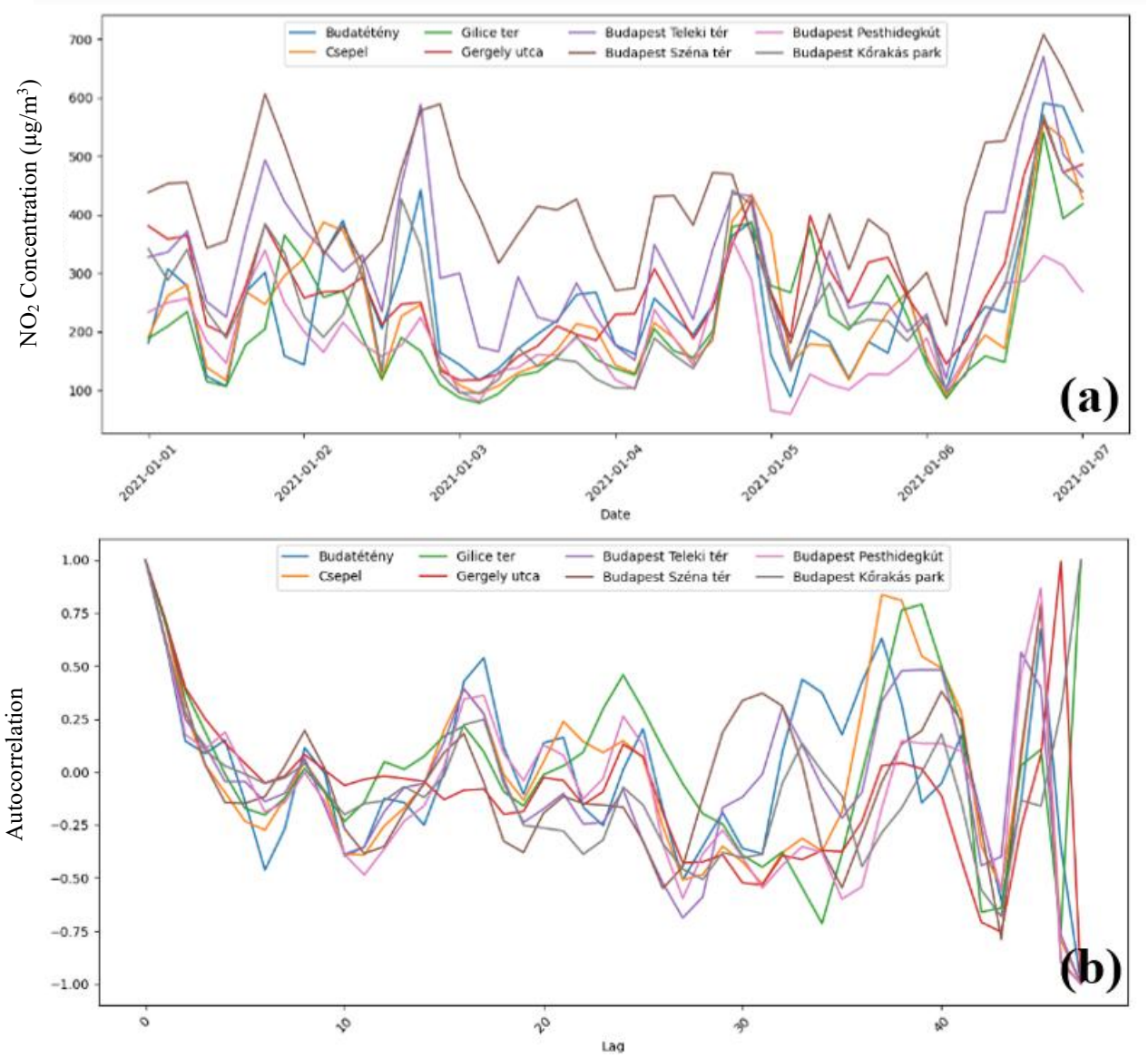



**Figure 9. The signal representing $NO_2$ concentration measured over 3-hour period (a), along with its corresponding autocorrelation (b).**

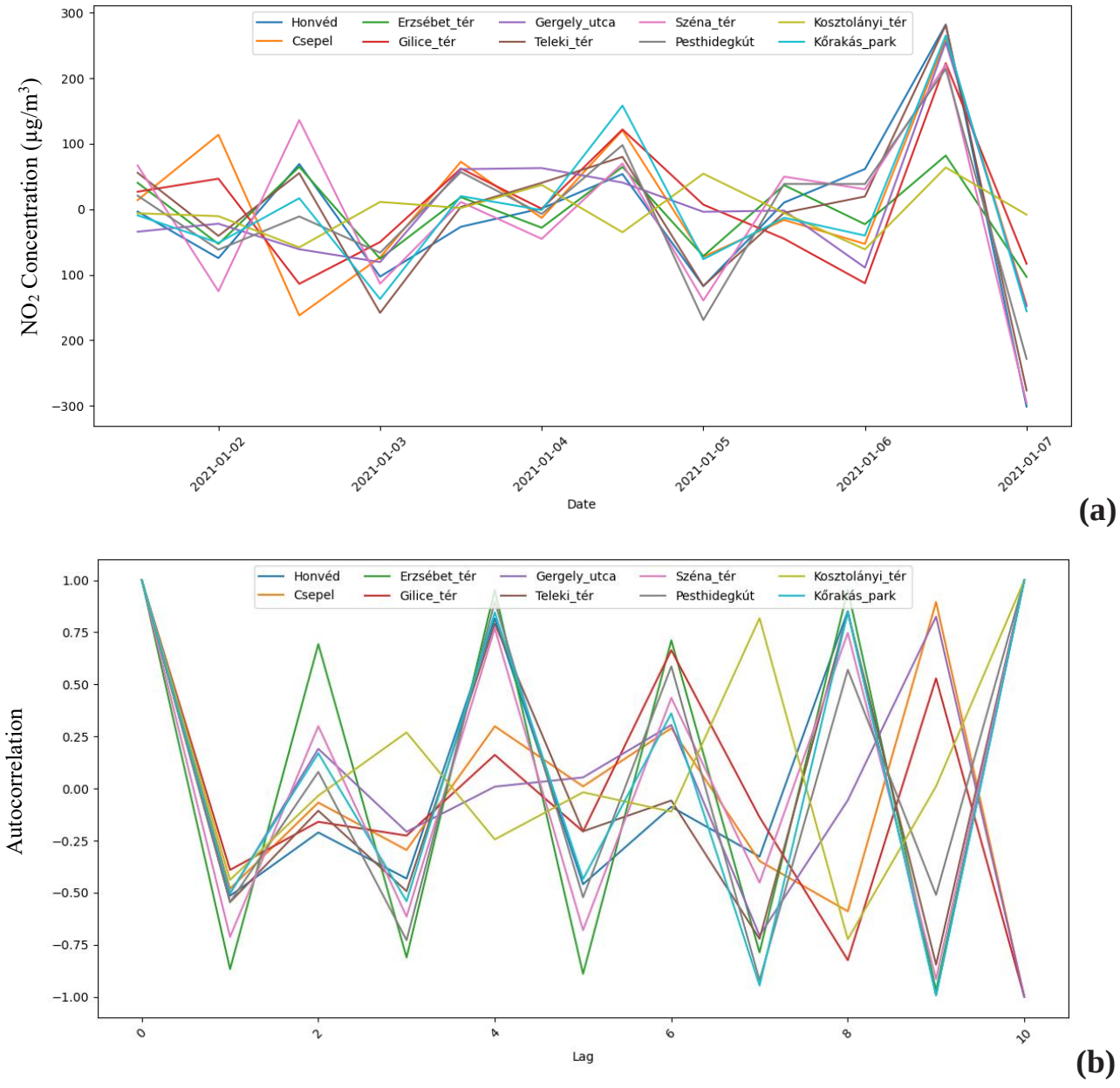
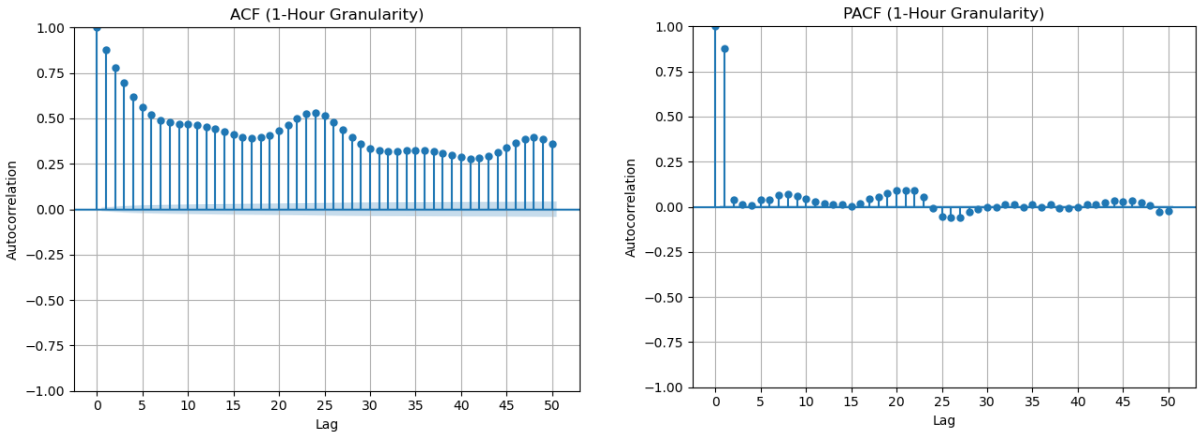


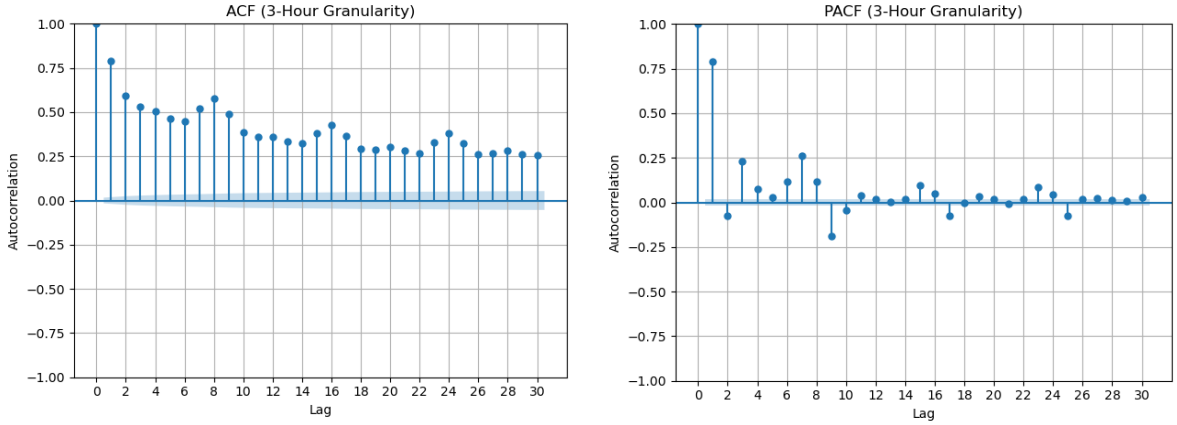

**Figure 10. The signal representing $NO_2$ concentration measured over 12-hour period (a), along with its corresponding autocorrelation (b).**

The autocorrelation analysis of $NO_2$ concentration in 12h-period has yielded intriguing insights, particularly in the context of urban traffic emissions. At all stations, a prominent short-term seasonality was observed, as evidenced by significant peaks at lags 2 and 4. This suggests a recurring pattern in $NO_2$ concentrations, with distinct cycles occurring approximately every 2- and 4-time units. Given its status as an urban traffic/background station, these patterns likely reflect the dynamic nature of traffic flow, with variations tied to daily commuting and traffic-related events. In contrast, for Kosztolányi square station, characterized as an urban traffic location, a consistent significant peak at lag 3 was observed. This suggests a more uniform short-term seasonality, potentially influenced by factors such as weather conditions or industrial activities. These findings emphasize the distinct temporal patterns associated with urban traffic emissions, underlining the importance of targeted strategies for managing air quality in densely populated urban areas. Understanding and addressing these patterns can pave the way for effective pollution control measures and urban planning initiatives.

**(a)**




**(b)**




**(c)**

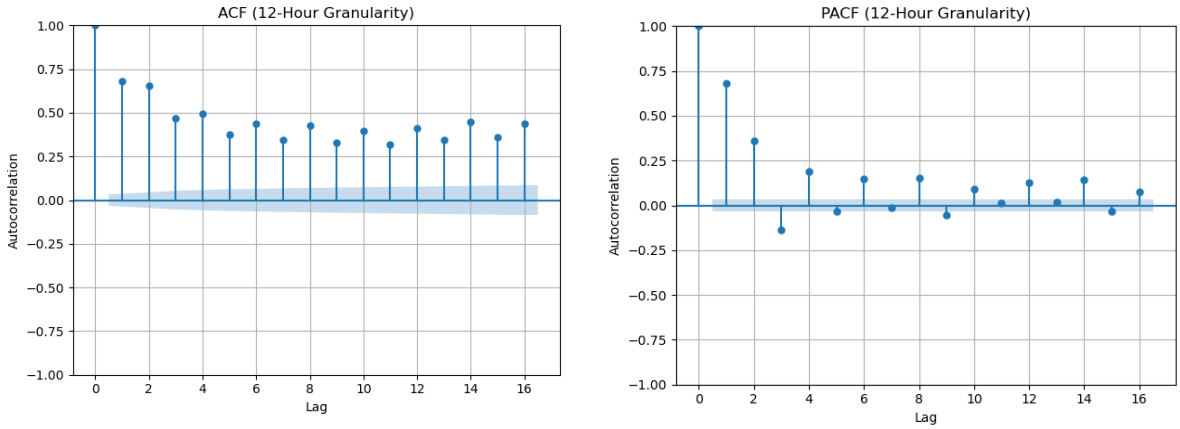



**Figure 11. Autocorrelation and Partial Autocorrelation of $NO_2$ at "Erzsebet square" Air Quality Station Measured at 1-Hour (a), 3-Hour (b), and 12-Hour (c) Granularity.**

### 4.2.2. Stationarity analysis

Figure 11 presents noteworthy findings, revealing that all peaks, regardless of the data granularity level, exhibit positive values. This positive trend signifies the imperative need for data differencing. Turning our attention to the Partial Autocorrelation Function (PACF) plots, we pinpoint the most substantial peak at a significance level of 0.05, corresponding to a lag of p=1.

For the 1-hour aggregated data, a distinct pattern emerges a direct relationship is evident among the series members at and immediately before the 24th tick. This relationship aligns with the hourly granularity of the data and the daily activities within the city.

Similar behaviour characterizes the 3-hour and 12-hour granularity data, with direct relationships observed between the first lag and the lags at or just before the 8th lag (for 3-hour granularity) and the second lag (for 12-hour granularity). Before delving into the modelling phase, it is imperative to rigorously assess data stationarity using two pivotal tests, the Augmented Dickey–Fuller (ADF) and Kwiatkowski–Philips–Schmidt–Shin (KPSS) tests.

These tests should not be employed interchangeably, as they may yield conflicting conclusions regarding data stationarity. The ADF test formulates its null hypothesis as follows (Kebłowski & Welfe, 2004):

- If the p-value > 0.05 (test value > 5% test critical value), it fails to reject the null hypothesis (H0), implying non-stationarity.
- If the p-value ≤ 0.05 (test value ≤ 5% test critical value), it rejects the null hypothesis (H0), signifying stationarity.

In contrast, the KPSS test interprets p-values inversely. The critical values for the ADF test encompass 1% (-3.430), 5% (-2.862), and 10% (-2.567), while for the KPSS test, they comprise 1% (0.739), 5% (0.463), and 10% (0.347).

We present the comprehensive results of both tests in table 1 in appendix.

Based on the results of the ADF test for the $NO_2$ dataset, most stations have stationary data. However, according to the KPSS test, all stations except Pesthidegkút show signs of non-stationarity for granularity level 1h while Kőrakás park, Gilice square and Pesthidegkút shows stationarity for 3h granularity level, particularly in terms of trend. At the 12-hour granularity level, all stations, except Gilice square, Kőrakás park, Pesthidegkút and Kosztolányi D. square based on the KPSS test, exhibit non-stationary behaviour, detailed results of both tests before differencing is showed in Table 1 in appendix 2, indicating the necessity of differencing to achieve reliable results for both tests (Table 7), it is important to note that the p-value after differencing is below 0.05.

Table 7. Stationary Tests Results for $NO_2$ Concentration Data after differencing.

| **$NO_2$** | **1h** | | **3h** | | **12h** | |
|---|---|---|---|---|---|---|
| Station | ADF | KPSS | ADF | KPSS | ADF | KPSS |
| Erzsébet square | -31.00 | 0.012 | -18.48 | 0.002 | -14.92 | 0.003 |
| Budatétény | -33.45 | 0.016 | -19.54 | 0.004 | -12.80 | 0.009 |
| Csepel | -31.75 | 0.016 | -18.33 | 0.005 | -12.20 | 0.006 |
| Honvéd complex | -31.41 | 0.015 | -19.32 | 0.004 | -16.25 | 0.015 |
| Gilice square | -32.63 | 0.012 | -18.98 | 0.004 | -14.00 | 0.007 |
| Gergely street | -32.46 | 0.034 | -18.91 | 0.006 | -10.14 | 0.020 |
| Széna square | -38.11 | 0.012 | -27.82 | 0.003 | -16.97 | 0.004 |
| Teleki square | -38.35 | 0.013 | -27.50 | 0.003 | -18.29 | 0.003 |
| Pesthidegkút | -38.56 | 0.059 | -26.93 | 0.006 | -17.54 | 0.006 |
| Kőrakás park | -38.20 | 0.006 | -28.12 | 0.003 | -16.69 | 0.001 |
| Kosztolányi D. square | -39.06 | 0.007 | -26.49 | 0.003 | -15.05 | 0.006 |

### 4.2.3. ARIMA Model

The selection of appropriate parameters for Autoregressive Integrated Moving Average (ARIMA) modelling is a crucial aspect of time series analysis. These parameters include the autoregressive order (p), differencing order (d), and moving average order (q). The process of determining these parameters is guided by a thoughtful examination of the dataset's autocorrelation structure.

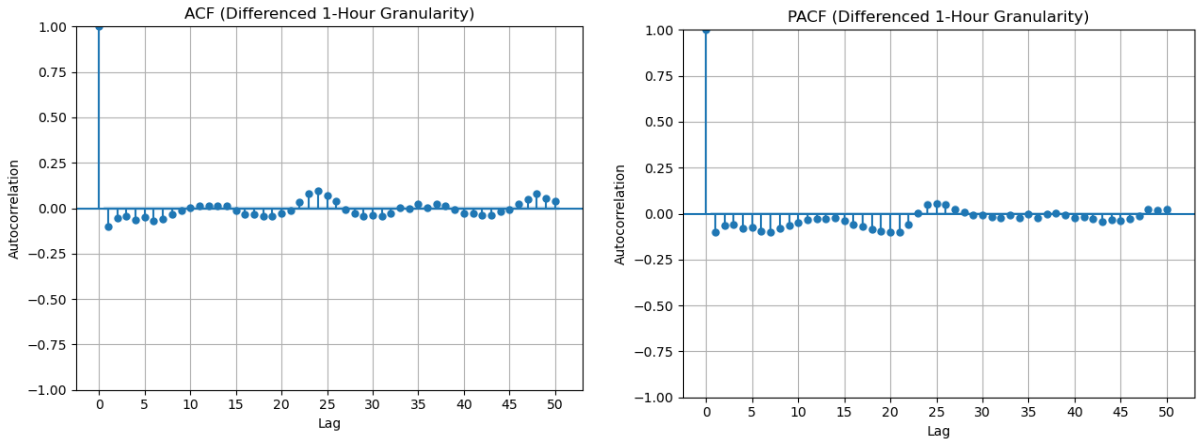
In our endeavour to ascertain the ideal values for these parameters, we turn our attention to the autocorrelation graphs, which reveal the underlying temporal dependencies within the data. Specifically, we examine the characteristics of positive and negative lags in these plots.

The prevalence of significant positive lags suggests a need for increased differencing to render the series stationary. Conversely, an abundance of negative lags may indicate an excessive level of differencing (Freeman et al., 2018). Consequently, for datasets with a 1-hour granularity, it is reasonable to consider a lag difference of no more than 1 especially for stations that already show stationarity, namely Gilice square and Gergely street. The determination of the autoregressive order, 'p,' is informed by the partial autocorrelation plot. As illustrated in Figure 12, a value of 'p = 5' appears promising for 1-hour granularity data, while values of '4' and '1' are indicative for 3-hour and 12-hour granularity, respectively.

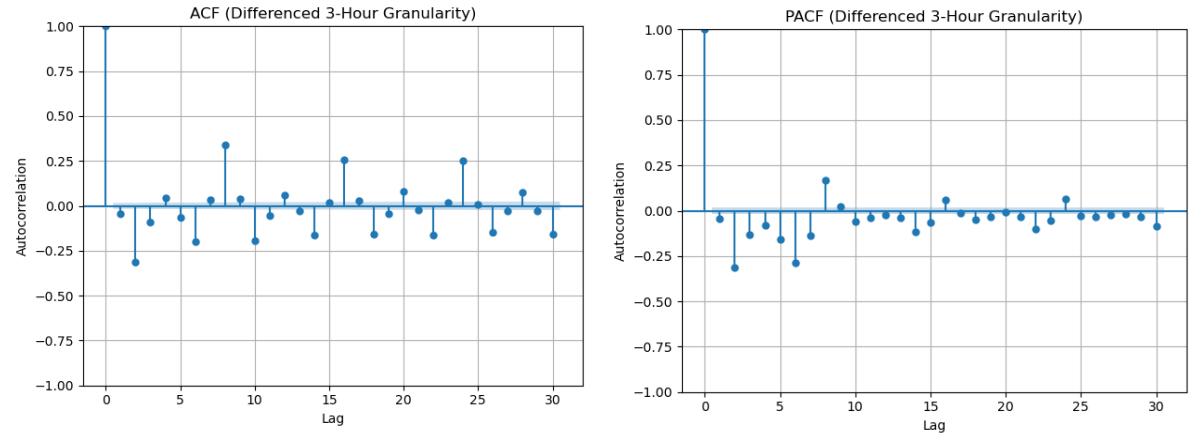
In the context of selecting the moving average order, 'q,' the examination centres on the number of lags surpassing the thresholds of -0.2 and +0.2 in the autocorrelation plots. In cases of 12-hour and 3-hour granularity data, a larger 'q' may be justifiable in comparison to the more conservative choice of '1' for 1-hour granularity data.

Given the heuristic nature of the parameter selection rule discussed above, a systematic approach is employed, leveraging the capabilities of the Python library 'pmdarima.' This approach involves a grid search methodology, combining criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in tandem with the Augmented Dickey-Fuller (ADF) test.

**(a)**




**(b)**




**(c)**

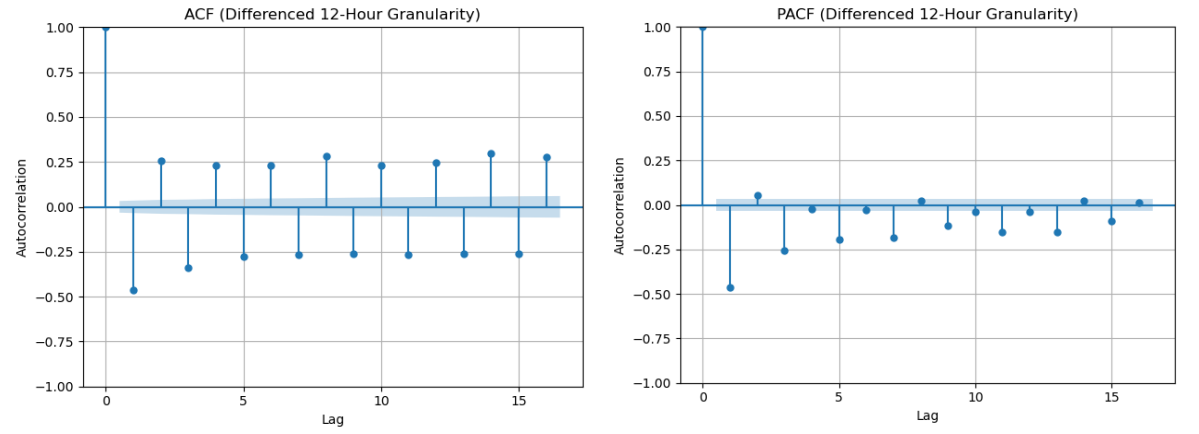



**Figure 12. Autocorrelation and Partial Autocorrelation of $NO_2$ at "Erzsebet square" Air Quality Station Measured at 1-Hour (a), 3-Hour (b), and 12-Hour (c) Granularity for lag difference 1.**

To assess and compare the performance of the ARIMA models generated through this parameter search, the Mean Absolute Scaled Error (MASE) is utilized as the final evaluation metric. MASE provides a robust measure of forecasting accuracy, facilitating a comprehensive evaluation of model effectiveness.

In Figures 13-14, we present a series of plots showcasing the predictions derived from ARIMA models. These models are built upon optimally determined parameters 'p,' 'd,' and 'q,' catering to varying granularity levels of 1 hour, 3 hours. These predictive visualizations offer insights into the model's ability to capture and forecast the $NO_2$ and $PM_{10}$ levels across Budapest air quality monitoring stations.

### 4.2.4. Predictions of ARIMA model for Budapest

Chemical transport models (CTMs) have proven to be highly effective in predicting air pollutant levels, particularly in urban areas, as they account for the complex physical interactions and dispersion patterns of pollutants. In recent years, CTMs have gained widespread use (Fernandes et al., 2021; Kitagawa et al., 2021; von Schneidemesser et al., 2021). However, there is a growing effort to enhance the accuracy of pollutant predictions on a smaller scale, specifically at the human level (around 10 meters). Unfortunately, CTMs are limited in their ability to achieve such fine-scale resolutions, with the minimum scale resolution typically being around 1 km.

This limitation has sparked interest in exploring statistical models as a potential solution to this issue. By utilizing downscaling techniques through machine learning and deep learning algorithms, it is possible to optimize the output of CTMs. Several studies have demonstrated successful downscaling methods that can provide predictions at a much finer scale, even as precise as 100 meters (Fattal et al., 2023; Fernandes et al., 2021; C. Huang et al., 2022; Wang et al., 2023).

In the meantime, short-term statistical methods such as Auto-Regressive Integrated Moving Average (ARIMA) models offer valuable insights into the patterns of air quality and can identify critical pollutants that require forecasting. These models also play a crucial role in developing alert systems for individuals who are particularly vulnerable to poor air quality.

Given the current lack of studies that predict air quality on the human scale, we have chosen to adopt an ARIMA Search Grid approach. This approach is based on criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), along with results from the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. By employing this methodology, we aim to train the model to determine the best combination of (p, d, q) parameters for accurate air quality forecasting in Budapest.

The stationarity of the data was assessed using Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) statistical tests.

These tests were conducted for different granularities, including 1-hour intervals with lag differences ranging from 1 to 24, 3-hour intervals with lag differences from 1 to 8, and 12-hour

intervals with lag differences of 1 and 2. The results of both tests indicated a significant level of stationarity in the $NO_2$ measurements at various monitoring locations, specifically in Pesthidegkút, Kőrakás park, Kosztolányi D. square, and Gilice square. To further enhance the stationarity of the data, optimal lag differences were applied based on recommendations from the Python library "pmdarima," and the ADF statistics were optimized. These adjustments resulted in improved stationarity levels across all granularity levels.

The study harnessed the AutoRegressive Integrated Moving Average (ARIMA) method for air pollutant forecasting across different granularities. The developed ARIMA models exhibited remarkable predictive capabilities, particularly at 3-hour granularity. The verification of the results indicated a close match between predicted and actual values, with mean absolute errors ranging from 11.08 to 23.62 for 1-hour granularity for $NO_2$ and 4.77 to 7.57 for 3-hour granularity for $PM_{10}$ (Table 8 and 9 with the best p,d,q values based on the search grid). Notably, the ARIMA method often outperformed complex structural methods, especially in short-term, one-step forecasting for univariate datasets (Mendyl et al., 2023).

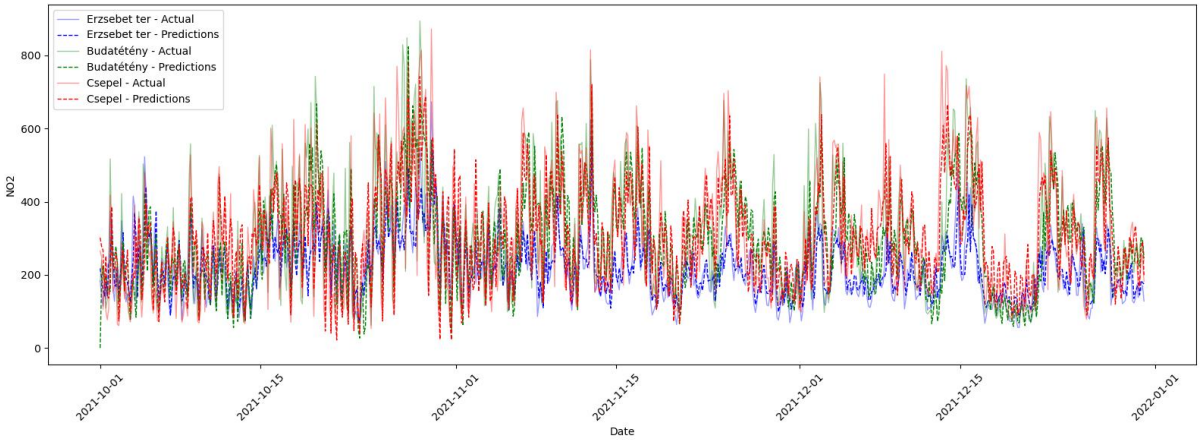



**Figure 13. ARIMA model with predictions for $NO_2$, 1h granularity, based on grid search and ACF, PACF lags consideration.**

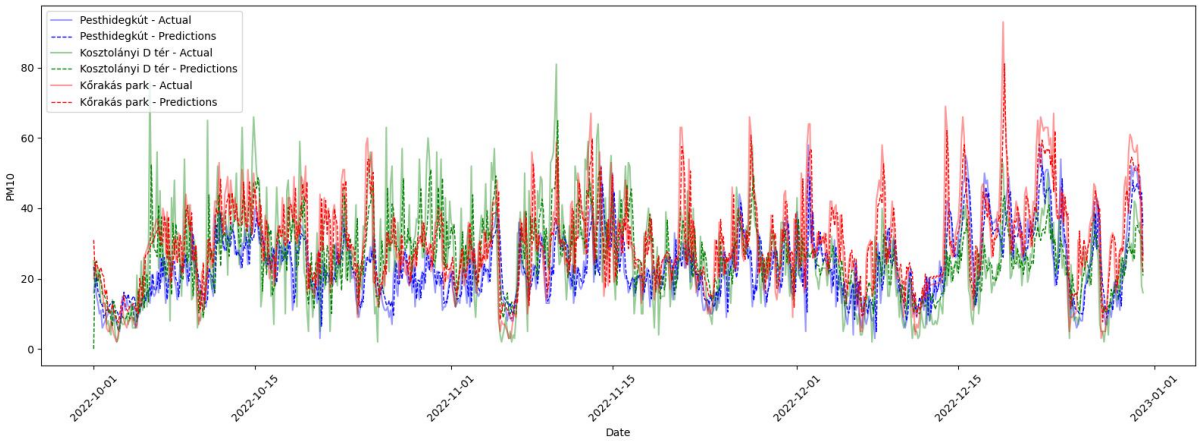



**Figure 14. ARIMA model with predictions for $PM_{10}$, 3h granularity, based on grid search and ACF, PACF lags consideration.**

As a result, the ARIMA models developed in this study can serve as a benchmark for evaluating various forecasting techniques and can be effectively integrated with other methods for a comprehensive analysis of time series data.

Table 8. 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 |

Table 9. 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 |

### 4.2.5. Conclusions

This section presents the findings of a comprehensive time series analysis conducted to predict air quality in Budapest, Hungary. The study focused on employing the AutoRegressive Integrated Moving Average (ARIMA) method to forecast the concentrations of key air pollutants, including nitrogen dioxide ($NO_2$) and particulate matter ($PM_{10}$), at twelve local air quality monitoring stations. The primary objective of this step was to assess the seasonality of pollutants and give a forecasting framework based on statistical models for Budapest. We utilized Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) statistical tests to effectively enhance the stationarity of the time series data, ensuring that the data meets the necessary assumptions for time series analysis.

Several key findings emerged from this investigation:

- Annual Air quality levels in Budapest are under the limits set by the European Commission, but hourly peaks still surpass the hourly limits underscoring the urgency of implementing additional strategies to mitigate air pollution in the region.
- Nitrogen dioxide ($NO_2$) exhibits a discernible seasonal pattern, with concentration peaks observed in March, July, and September. These peaks are indicative of a strong correlation between NO2 levels and increased emissions from vehicle engines during these months.

While the ARIMA search grid methodology demonstrated remarkable short-term predictive capability for Budapest's air quality (MAE of 11-24 μg/m³ for hourly $NO_2$), these purely statistical models reveal fundamental limitations that necessitate alternative approaches. The ARIMA framework's exclusive reliance on historical temporal patterns renders it blind to the underlying physical and chemical processes governing pollutant formation and dispersion. This limitation becomes particularly acute when attempting to predict responses to unprecedented conditions, emission control scenarios, extreme meteorological events, or structural urban changes, where historical patterns provide no guidance. Furthermore, ARIMA's station-specific nature confines predictions to monitoring locations without spatial interpolation capability, leaving vast urban areas unmonitored. To address these limitations while maintaining predictive accuracy, we turn to deterministic Chemical Transport Models that explicitly represent atmospheric physics and chemistry, enabling scenario analysis and spatial coverage beyond monitoring stations.

## 4.3. Deterministic model results

### 4.3.1. CHIMERE outputs sensitivity to spatial resolution

Figure 15. plots the monthly average observed and modelled ozone and $PM_{10}$ data in Agadir city on March 1st and August 31st, 2010, and 2016. The first observation reveals a systematic tendency for the model to overestimate ozone concentrations and underestimate $PM_{10}$ concentrations. For more details, Table 3 shows the statistical scores calculated for the two pollutants. High RMSE indicates significant variability between observations and model predictions, meaning that the model may have difficulty accurately reproducing local variations. The common observation for these two pollutants is that the CHIMERE model retains the monthly trend, indicating the model's capability to capture the characteristic monthly fluctuations in ozone and $PM_{10}$ concentrations (Guion et al., 2023; Mazzeo et al., 2022; Potier et al., 2019). In other words, the model can typically reproduce seasonal variations in these atmospheric pollutants. Even if these results agree with other studies linking low model accuracy with input uncertainty, as reported in Table 10, we can suggest that the model needs to be adjusted to reproduce $O_3$ and $PM_{10}$ concentrations more accurately.

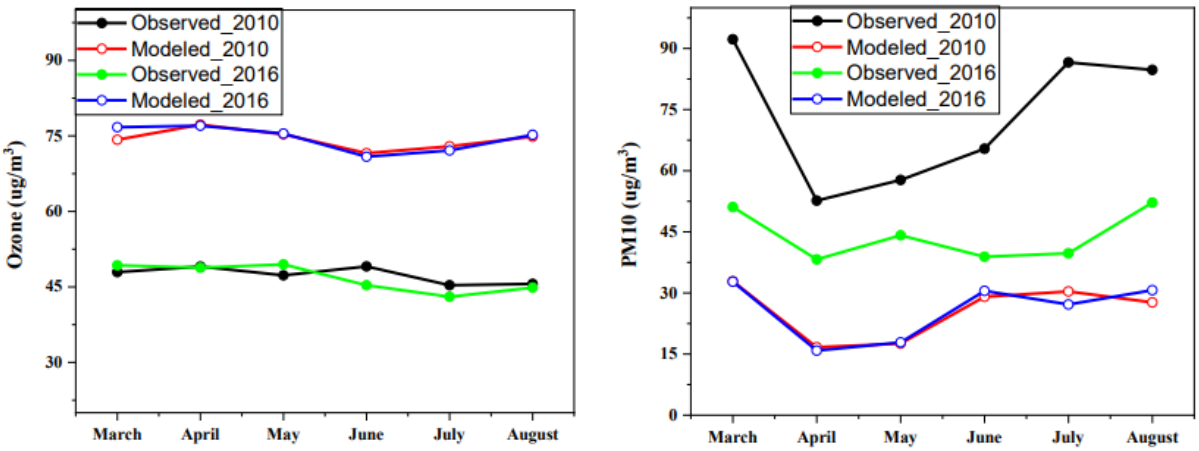



**Figure 15. Monthly average observed and modelled ozone and $PM_{10}$ data in Agadir City on March 1st and August 31st, 2010, and 2016 respectively.**

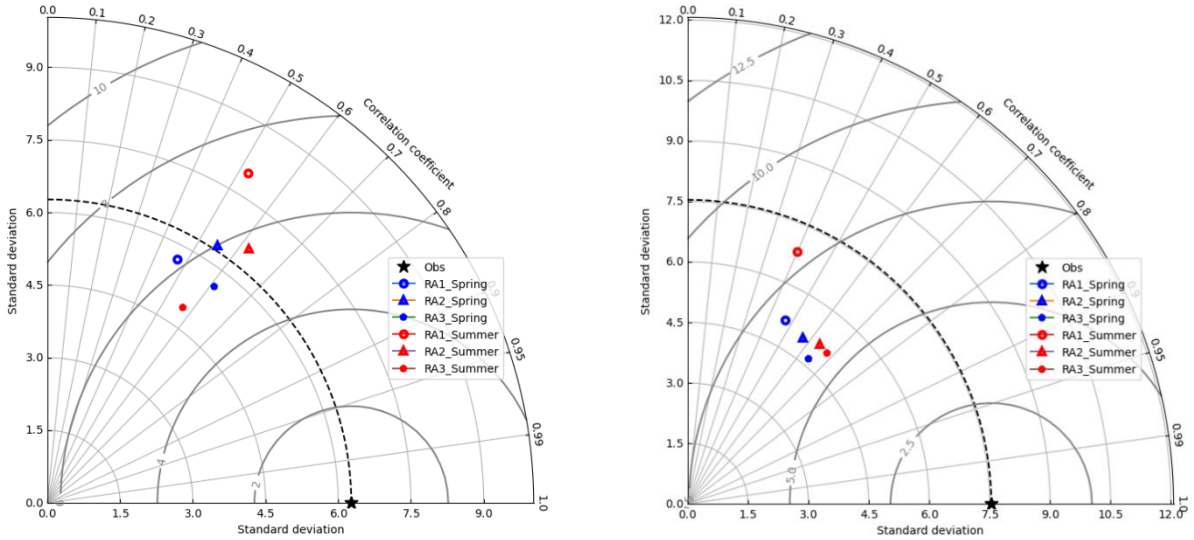
Table 10. Calculated statistical scores of the ozone and particulate matter compared to other studies

| | Mean. Obs | Mean. Mod | RMSE | Literature ranges | Validation studies |
|---|---|---|---|---|---|
| Ozone [µg/m$^3$] | 47.1 | 74.5 | 27.5 | [11.9 ; 34.6] | (Blanco-Ward et al., 2021; Meng et al., 2022)(Mo et al., 2021)(J. Li et al., 2022) |
| Particulate matter [µg/m$^3$] | 58.6 | 25.8 | 34.1 | [11.5 ; 44.8] | (Do et al., 2021; Saidi et al., 2023; Tchepel et al., 2020; K. Yu et al., 2016; Yuan et al., 2023) |

To understand the spatial resolution effect on the two pollutants, we tested the effect of $R_{A1}$, $R_{A2}$, and $R_{A3}$, details about the domains in both cities are presented in figure S1 and S2 as well as table 3 in appendices. Figure 16 shows the Taylor diagram of $O_3$ and $PM_{10}$ for different spatial resolutions in the spring and summer of 2010 and 2016. For both pollutants ($O_3$ and $PM_{10}$), $R_{A3}$ spatial resolution (0.02°~ 2 km) offers lower standard deviations, suggesting better agreement with observations than $R_{A1}$ and $R_{A2}$ resolutions. Correlations are generally positive, indicating correspondence between observations and the CHIMERE model. In particular, correlations are often higher for $R_{A3}$ resolution, suggesting a better ability to reproduce observed trends. The results indicate that choosing a higher spatial resolution can lead to a better representation of seasonal and spatial variations in these air pollutants, which agrees with other studies (Adedeji et al., 2020; Ajdour et al., 2021; Sicard et al., 2021). To better understand the contribution of high spatial resolution to improved forecasting of the two pollutants, $O_3$ and $PM_{10}$, we have tried to investigate the effect on model inputs.

2010 2016

Ozone ($O_3$)




Particulate matter ($PM_{10}$)

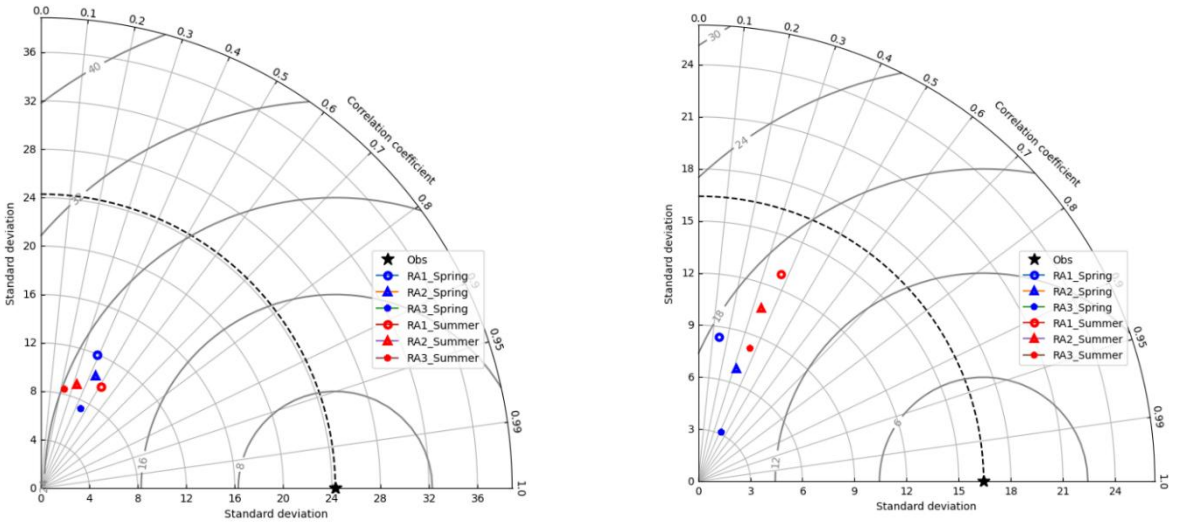



**Figure 16. Ozone and Particulate matter Taylor diagram for 2010 and 2016, in Agadir city using $R_{1A}$, $R_{2A}$, and $R_{A3}$.**

### 4.3.2. Spatial resolution effect on CHIMERE inputs model

#### 4.3.2.1. Temperature and Wind Speed

Figure 17 illustrates the monthly average temperature (a) and wind speed (b) profiles observed and modelled in Agadir city on March 1st and August 31st of 2010 and 2016. The main observation based on the representation of monthly temperature data is the seasonal reproducibility of the model. The monthly wind speed showed medium results, especially during 2016, which can be attributed significantly to WRF model limitation, as reported by several studies (Solbakken & Birkelund, 2018; Suárez et al., 2022).

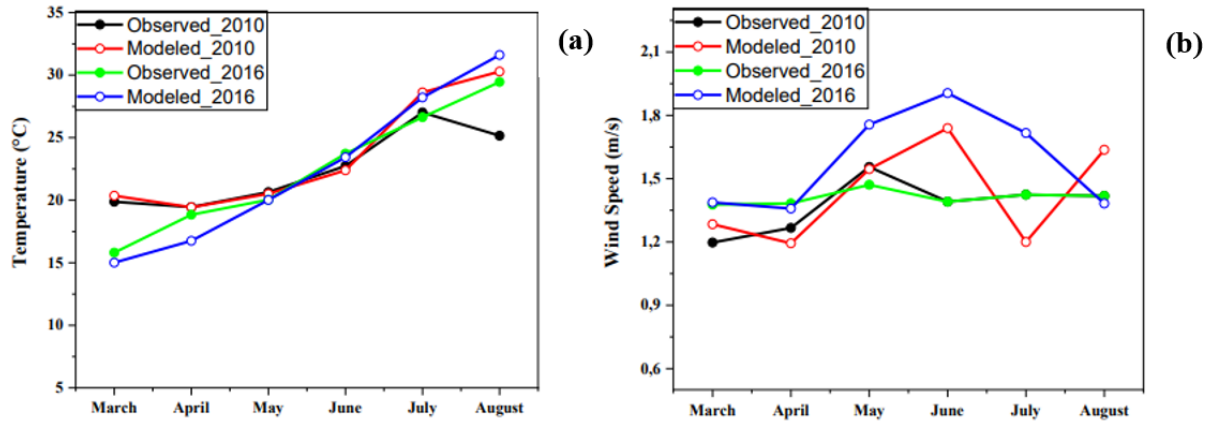



**Figure 17. Monthly average observed and modelled Temperature (a) and Wind Speed (b) in Agadir city on March 1 and August 31, 2010, and 2016.**

To visualize spatial and temporal dependence, we tested the effect of spatial resolution on the output of the WRF model for two cities, Agadir and Casablanca, in different periods. Table 11 presents the statistical scores for the two domains (d01 and d02) of the WRF model with grid size reduced by a factor of 3 from 21 km to 7 km in the Agadir ($Rd0_{1A}$ and $Rd0_{2A}$) and Casablanca ($Rd0_{1C}$ and $Rd0_{2C}$) cities. The finer spatial resolution of $Rd0_{2A}$ leads to a noticeable overestimation of temperature compared with $Rd0_{1A}$ in Agadir, indicating challenges related to the model's accuracy in capturing local variations at finer resolution. Although both spatial resolutions in Casablanca show a high correlation between observations and models, $Rd0_{1C}$ performs better. The WRF models present a high wind speed overestimation in both cities. In Agadir, The $Rd0_{1A}$ resolution shows a slightly higher correlation than $Rd0_{2A}$, with similar performance and close RMSE values. For Casablanca, $Rd0_{1C}$ shows a correlation of 68%, a bit higher than the 66% of $Rd0_{2C}$. Both resolutions perform similarly in terms of mean bias and RMSE. It can be observed that the coarse resolution exhibits the minimum Mean Bias (MB) and Root Mean Square Error (RMSE) in both cities. On the other hand, the spatial resolution does not significantly impact the overestimation of wind speed produced by WRF and does not provide a means of correction. Our results agree with other studies that supported the coarse resolution in some cases of the WRF model. However, the resolution effect depends on the studied variables and the used scheme (El-Samra et al., 2018; Maina et al., 2020; Ovchinnikov et al., 2022; Z. Yu et al., 2022).

Table 11. Temperature and Wind Speed statistical scores for the two WRF domains, d01, and d02, in Agadir and Casablanca.

| | | Mean. Obs | Mean. Mod | Correlation | Mean Bias | RMSE |
|---|---|---|---|---|---|---|
| Temperature [°C] | $R_{d01A}$ | 15.3 | 15.8 | **84.8%** | **0.5** | **0.7** |
| | $R_{d02A}$ | | 17.2 | 80.0% | 2.0 | 2.0 |
| | $R_{d01C}$ | 13.3 | 12.9 | **92.6%** | **-0.4** | **1.6** |
| | $R_{d02C}$ | | 12.6 | 92.0% | -0.6 | 1.8 |
| Wind Speed [m/s] | $R_{d01A}$ | 1.3 | 3.4 | **48.0%** | 2.1 | 2.1 |
| | $R_{d02A}$ | | | 26.0% | 2.0 | 2.0 |
| | $R_{d01C}$ | 1.0 | 2.9 | **68.1%** | 1.9 | 2.1 |
| | $R_{d02C}$ | | | 66.3% | | 2.0 |

To understand the spatial resolution effect on WRF outputs, we tested three spatial resolutions. Figure 18 shows the Temperature and Wind Speed Taylor diagram for daily observations and modelling results for three horizontal resolutions between March 1st and August 31st of 2010 and 2016. The summer season Temperature is overestimated compared to the spring season. In parallel, spatial resolution does not present a significant effect on temperature. These results are consistent with other studies (X. Jiang & Yoo, 2018; Tao et al., 2020) . It can be noticed that the overestimation of the Wind Speed is a limitation of the WRF. On the other hand, the model spatial resolution aims to improve the quality of the result without surpassing the WRF limit, requiring a detailed study to identify dependencies and find appropriate correction forms (Rzeszutek et al., 2023).

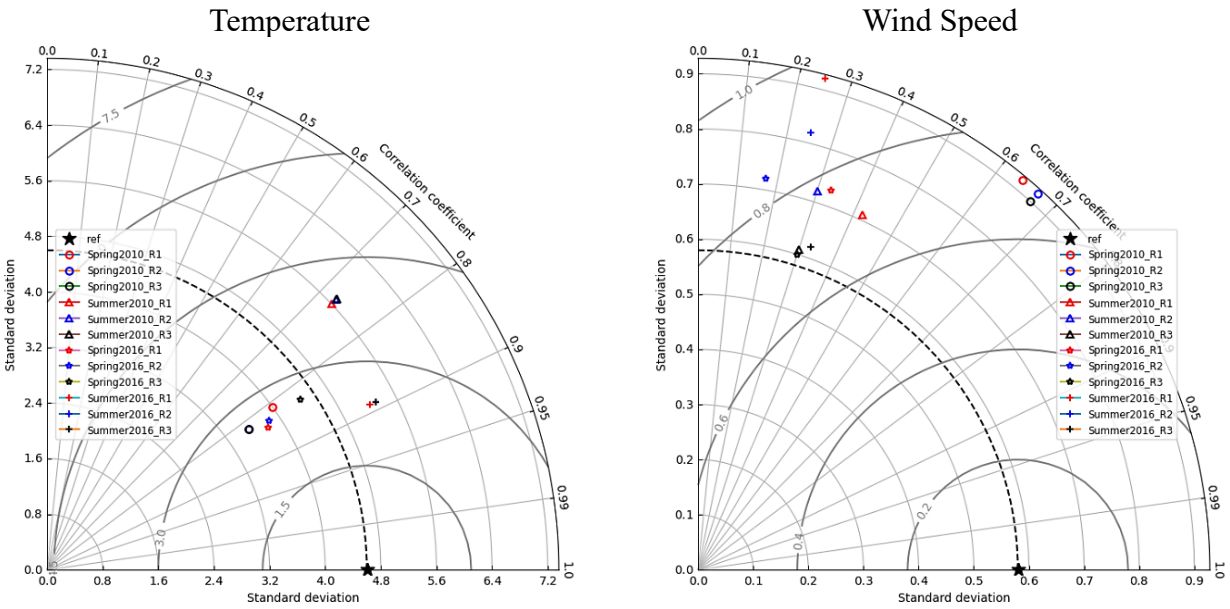



**Figure 18. Temperature and Wind Speed Taylor diagram for 2010 and 2016, in Agadir city using $R_{1A}$, $R_{2A}$, and $R_{A3}$**

#### 4.3.2.2. Planetary Boundary Layer Height

Figure 19 presents the seasonal variation in PBLH with spatial resolution $R_{A1}$ and $R_{A3}$. We can note the tendency of high resolution to reduce PBLH, reflecting the efforts of the model to capture local characteristics, such as topography, land use, and urban structures (Chang et al., 2022; Eure et al., 2023). To quantify the seasonal effect, we can compare the average PBLH value of the identical spatial resolution for each season (Spring-Summer). The general observation is that PBLH is higher in summer than in spring. In summer, the planetary boundary layer (PBLH) is typically higher due to increased solar radiation and warmer temperatures, favouring turbulent atmospheric mixing and a deeper boundary layer reinforced by increased convective activity. These results are in agreement with several studies (Allabakash & Lim, 2020; Kalmus et al., 2022; Slättberg et al., 2022). Other studies have demonstrated linkages between the height of the planetary boundary layer (PBLH) and topography, vegetation, the presence of buildings, and other factors (de la Paz et al., 2022; Q. Jiang et al., 2023). To

understand if the model is really trying to reduce PBLH to capture local variability, we aim to test the effect of spatial resolution on land use.

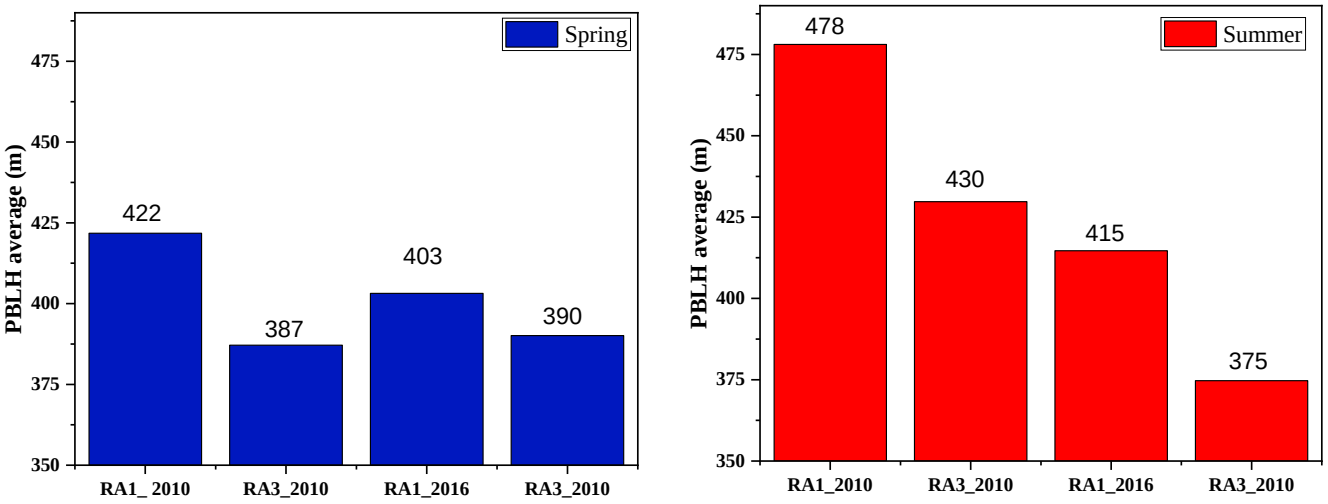
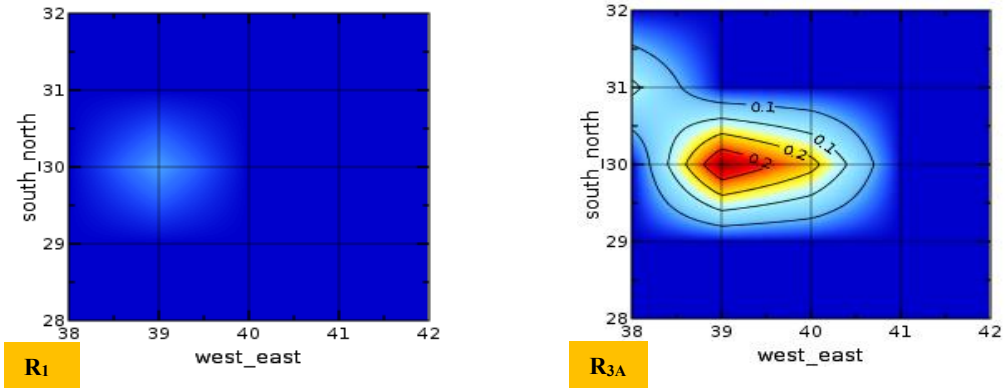



**Figure 19. Resolution effect on PBLH for spring and summer of 2010 and 2016 using $R_{A1}$ et $R_{A3}$**

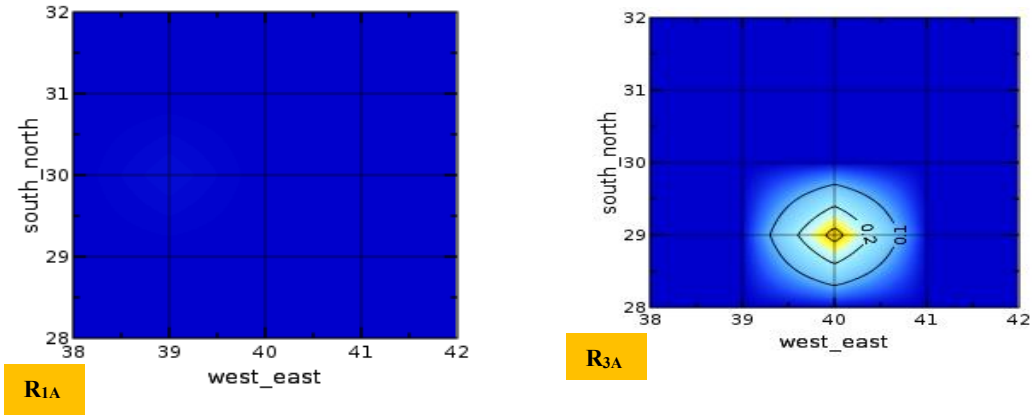
#### 4.3.2.3. Land Use and Land Cover sensitivity to spatial resolution

Figure 20 shows 3 LULC classes using the CHIMERE spatial resolutions $R_{A1}$ and $R_{A3}$. The cell centre of the domain corresponds to the fixed station position (40 West East,30 North-South). The RA1 indicates that the measurement station is situated in a less urban and less agricultural zone and tends towards the shrub land. On the other hand, $R_{A3}$ is sensitive to the urban environment, which is appropriate for the station position in Google Earth. We can note the ability of $R_{A3}$ to distinguish land-use characteristics more accurately is particularly noteworthy in areas of high spatial diversity, such as urban areas. The results are consistent with several studies addressing the critical role of high spatial resolution on Land Use and Land Cover (Fountoukis et al., 2022; Maxwell et al., 2019; Toure et al., 2018). The comparison between $R_{A1}$ and $R_{A3}$ highlights the importance of choosing an appropriate spatial resolution according to the specific characteristics of the region under study. These results may be essential for more accurate atmospheric modelling and adapted resource management, especially if the station's position is in the urban environment (Błaszczak et al., 2020; Wolf et al., 2020; Zimmerman et al., 2020). It is known that variations in the spatial distribution of emissions depend on soil type and land use. To assess the dependence of LULC Classes on spatial resolution, we studied the effect of resolution on emissions.

Urban




Agricultural




Shrubs

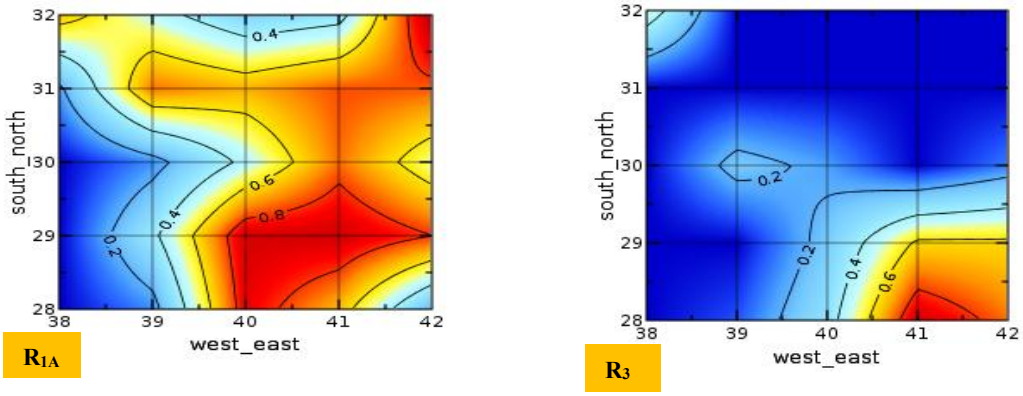



**Figure 20. Comparison of 3 LULC classes for R1A and R3A of the fixed station position (40 West_East,30 North South)**

#### 4.3.2.4. Emission sensitivity to spatial resolution

To better understand the effect of resolution on emissions, we investigated two-species emission variations in the two seasons using $R_{A1}$ and $R_{A3}$. Emission data were processed using Emi-Surf, the CHIMERE preprocessing tool that interpolates anthropogenic emission inventories onto the model grid and applies temporal profiles to generate hourly emission fluxes (Menut et al., 2013). The first is that $NO_2$ contributes to ozone formation; the second is PPM_coa (Primary particles for diameters between 2.5 µm and 10 µm) leads to $PM_{10}$ creation, as illustrated in Figure 21 Consider that ΔR, representing the emission variation for the appropriate cell with the fixed station's position (40 West East, 30 North South), is calculated as $R_{A3}$-$R_{A1}$. The first point to note is that emissions with RA3 are very high compared with $R_{A1}$, which may be due to $R_{A1}$ indicating that the study area is located in a more rural area, whereas $R_{A3}$ suggests an urban environment. This difference arises because Emi-Surf redistributes gridded emission inventories according to land use proxies; at coarser resolution, urban emission hotspots are diluted across larger grid cells containing mixed land uses, while finer resolution preserves the spatial concentration of urban sources.

We can notice that $R_{A3}$ aims to increase $NO_2$ emissions, which minimizes $O_3$ formation because of the inverse relationship between these two pollutants. The $NO_2$ presence leads to an ozone decomposition reaction ($O_3 + NO_2 \rightarrow O_2 + NO_3$) , forming different chemical products, such as dioxygen and the nitrate radical. This result can explain the improvements recorded in variability and correlation for ozone using high-resolution $R_{A3}$, which can somewhat, with further investigation into the chemical processes of ozone formation, help to determine the origin of the systemic overestimation of ozone. The comparison between $NO_2$ and PPM_coa emissions shows a difference of up to 10 times, which justifies the large RMSE. The $R_{A3}$ resolution aims to increase $PM_{10}$ emissions significantly in spring compared to summer, reflecting poor model performance and a weak summer trend. In short, high spatial resolution significantly influences land use, modifying emissions and impacting pollutant modelling (Hagler et al., 2021; Sahu et al., 2015; Singh et al., 2020; Yang et al., 2019).

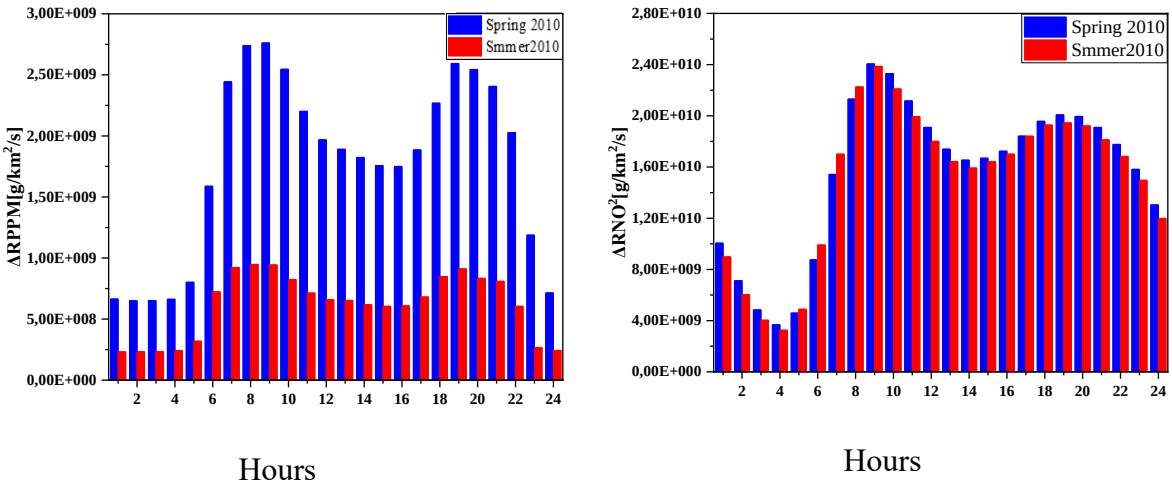



**Figure 21. Resolution effect on the hourly emissions for $NO_2$ and PPM_coa**

### 4.3.3. CHIMERE-ANN sensitivity to resolution

The results of the CHIMERE deterministic model highlight the outstanding performance of high spatial resolution in the study of meteorological parameters, land use, and emissions. However, despite these achievements, the CHIMERE model has its limitations. To overcome these limitations, we have integrated artificial intelligence in the form of the CHIMERE-ANN model. In this new model, CHIMERE outputs are injected as inputs into an artificial neural network (ANN) aiming to improve the accuracy of air quality forecasts. We selected the ozone concentrations from 2016 in Agadir and 2021 in Casablanca as study data to analyse the effect of spatial resolution on the CHIMERE-ANN model. In Table 12, CHIMERE-ANN ($R_{A3}$ and $R_{C3}$) increased the correlation coefficient from 61% to 80% in the case of Agadir and from 84% to 94% in the case of Casablanca. The RMSE decreased from 29.1 to 7.5 µg/m3 and from 30.8 to 7.4 µg/m$^3$: for Agadir and Casablanca, respectively, representing a significant upgrade using high-resolution in CHIMERE output. The CHIMERE-ANN correlation matches the literature, while the RMSE indicates a performance decrease compared to the other studies (Ajdour et al., 2022; H. Lu et al., 2021; Mok et al., 2017; Sayeed et al., 2021; Tchepel et al., 2020). These results suggest that using CHIMERE-ANN with high spatial resolution concentration ($R_{A3}$ and $R_{C3}$) may be a promising approach to improving the accuracy of pollutant concentration forecasts, which could have important implications for air quality management decision-making.

Despite this improvement, Figure 22 shows that the CHIMERE model adjustment fails to model intense concentrations, which we can note in the case of Casablanca between 2:00 pm and 6:00 pm. This limitation may be due to insufficient input data, given that we used only ozone concentrations as input data for ANN, and gaps in the input data, such as poorly quantified pollutant emissions, may make it difficult for the model to reproduce intense concentrations. In addition to adjusting the CHIMERE model, exploring other modelling approaches or combining several models to obtain a complete representation of local atmospheric conditions could be beneficial (Ajdour et al., 2022; Bessagnet et al., 2019).

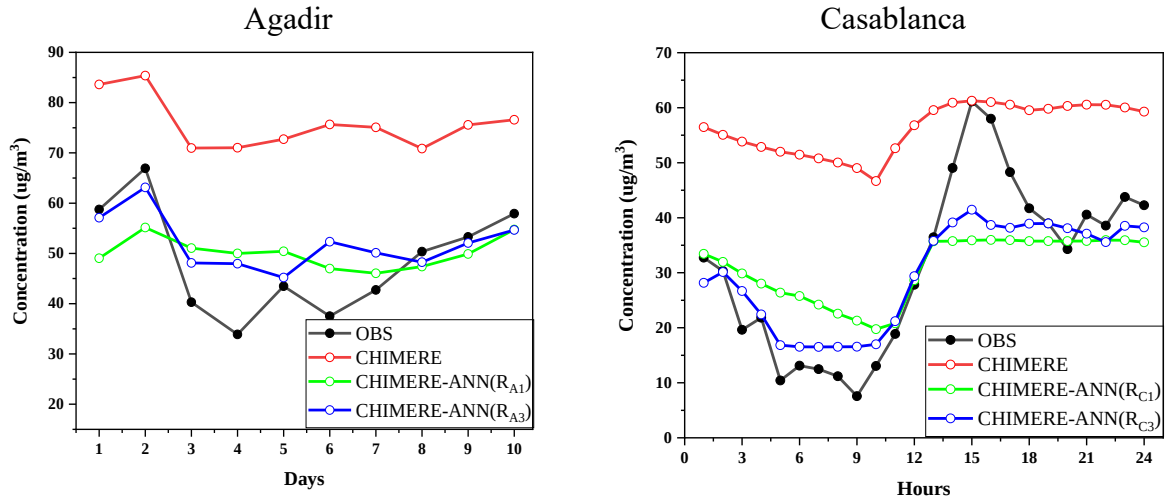



**Figure 22. Average $O_3$ concentration profiles observed (OBS) and modelled by ANN-CHIMERE using $R_{A1}$ and $R_{A3}$ in Agadir and $R_{C1}$ and $R_{C3}$ in Casablanca**

Table 12. 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] | Correlation | Literature ranges |
|---|---|---|---|---|---|---|
| Agadir | CHIMERE ($R_{A1}$) | 49.2 | 76.7 | 30.1 | 40% | 80%< $R^2$ <91%<br>0.02<RMSE<20<br>(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% | |

The successful bias correction achieved through the CHIMERE-ANN hybrid approach (reducing RMSE from 30.1 to 7.5 µg/m³ for ozone) demonstrates that machine learning can effectively enhance deterministic models. However, this enhancement remains fundamentally constrained by the deterministic model's structural limitations and cannot capture phenomena entirely absent from the CTM framework. Moreover, the single-model correction approach, while effective for systematic bias, does not address the reality that different atmospheric conditions may favour different modelling paradigms. The varying performance of CHIMERE across seasons, the dependence on spatial resolution, and the persistent challenges in capturing intense concentration peaks suggest that no single model, whether purely deterministic, statistical, or hybrid, can adequately represent the full complexity of urban air quality dynamics. This recognition motivates a fundamentally different approach: rather than seeking to perfect individual models, we explore how multiple complementary models, each with distinct strengths and specialized for different atmospheric conditions, can be intelligently combined. The following analysis therefore shifts focus from single-model enhancement to ensemble fusion, investigating how diverse machine learning architectures trained on carefully engineered features can be aggregated through advanced mathematical frameworks to achieve both superior predictive performance and maintained interpretability.

## 4.4. Hybrid and ensemble fusion

### 4.4.1. Feature engineering and model architecture analysis

The Comprehensive evaluation of 11 specialized models across monitoring stations revealed highly significant performance stratification by architecture class (Kruskal-Wallis H = 51.16, p=0.00). This extreme significance indicates fundamental differences in how architectures capture $PM_{10}$ dynamics. K-Nearest Neighbours with anomaly-detection features achieved superior performance (RMSE = 1.800 ± 0.717 μg/m³, $R^2$ = 0.979) as seen in Table 13 and figure 23, representing a minimum of 60.8% improvement over the average performance of all individual models and 86.7% improvement over the worst-performing SVR model. The Gradient Boosting comparison provides compelling evidence for regime-based modelling. GBR-Regime (RMSE = 4.601 ± 1.204 μg/m³) demonstrated dramatically superior performance compared to GBR-Stable (RMSE = 10.821 ± 2.162 μg/m³; paired t-test: t (10) = -13.61, p=0.00, Cohen's d = -4.10). This effect size of -4.10 represents one of the largest documented in atmospheric ML literature, quantifying the severe penalty of assuming stationarity. The stable variant's $R^2$ = 0.331 versus regime-specific $R^2$ = 0.813 indicates that 48.2% of variance explanation is lost when ignoring atmospheric regime transitions. Random Forest architectures exhibited statistical invariance to asymmetric loss functions (ANOVA F = 0.00, p = 0.9952), though the Friedman test detected subtle ranking differences ($\chi^2$ = 6.73, p = 0.01273). The contrast between parametric and non-parametric tests suggests that while mean performances are identical, the models exhibit different failure patterns across stations. The negligible ΔRMSE across variants (ΔRMSE %: 2.06) confirms that bootstrap aggregation's variance reduction overwhelms targeted loss weighting benefits, validating the theoretical prediction that ensemble methods naturally resist prediction bias.

Table 13. 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 |

LSTM architecture analysis revealed non-monotonic performance with respect to temporal context length. The 24-hour configuration achieved optimal performance (RMSE = 6.018 ± 3.084 μg/m³, $R^2$ = 0.782), while both shorter (12h: RMSE = 6.089 ± 3.154) and longer (168h: RMSE = 6.104 ± 2.977) contexts showed degradation. Statistical comparison between extreme

contexts (12h vs 168h: $t = -0.16$, $p = 0.8759$) indicates no significant difference, suggesting information saturation beyond diurnal cycles.

The comparison between 24h and 24h+ meteorological inputs ($t = -3.84$, $p = 0.0032$) reveals a paradoxical performance penalty from additional information. The multivariate LSTM (RMSE = 11.719 ± 4.840 μg/m³, $R^2 = 0.183$) represents catastrophic failure, with performance 94.7% worse than the optimal univariate configuration. This degradation, despite theoretical advantages of multivariate inputs, indicates that conflicting temporal scales between meteorological (synoptic: 72-120h) and pollution (diurnal: 24h) signals create irreconcilable optimization challenges in the shared recurrent state space.

Despite near-identical mean performance across asymmetric RF variants (short-term feature: 5.852 ± 3.095, underpredicted averse: 5.949 ± 3.096, overpredict averse: 5.973 ± 2.987 μg/m³), the models serve distinct operational purposes. The underpredict-averse variant reduces Type II errors during pollution episodes by 2.06% compared to the standard configuration, critical for public health warnings where false negatives carry higher costs than false positives. The invariance across loss functions (all pairwise $p > 0.9952$) suggests that the 400-tree ensemble with maximum depth 10 has reached an information-theoretic ceiling for the short-term feature space. This plateau at $R^2 \approx 0.79$-$0.80$ across all variants indicates that approximately 20% of $PM_{10}$ variance remains irreducible noise or requires features beyond the current 11-dimensional short-term dynamics representation.

SVR-RBF's catastrophic performance (RMSE = 13.598 ± 2.244 μg/m³, $R^2 = -0.048$) warrants detailed examination as a cautionary case. The negative $R^2$ indicates predictions 4.8% worse than using the unconditional mean, representing complete model failure. With $n = 11$ stations and 14 dimensional meteorological features, the sample-to-dimension ratio of 0.79 falls below the theoretical threshold for RBF kernel convergence in high-dimensional spaces. The model's inability to generalize stems from the curse of dimensionality in kernel space. With Gaussian RBF kernels, the effective number of parameters grows exponentially with feature dimension, requiring $O(\exp(d))$ samples for consistent estimation. Our configuration with $d = 14$ and $n\text{-effective} \approx 11 \times 8760$ hours creates a severely underdetermined system where regularization dominates, forcing near-constant predictions.

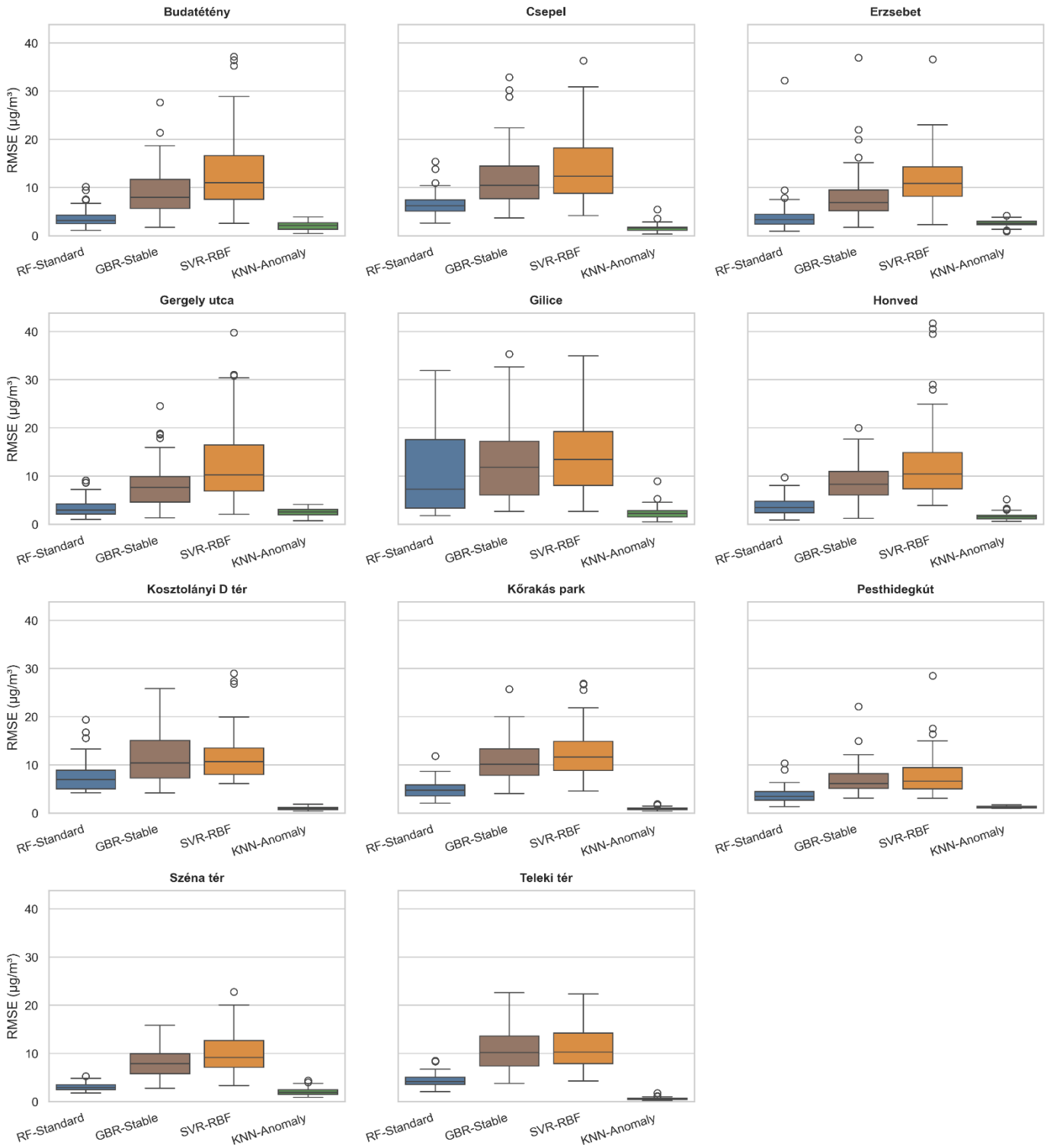



**Figure 23. RMSE distributions by feature-set expert ($PM_{10}$) across stations.**

The exceptional performance of KNN-Anomaly (RMSE = 1.80 µg/m³) validates instance-based learning for non-stationary atmospheric systems. With k = 15 neighbours and distance weighting, the model implicitly performs local polynomial regression in the 5-dimensional anomaly space. The dramatic improvement over global models suggests that $PM_{10}$ dynamics exhibit local linearity in deviation space despite global non-linearity in absolute concentration space. Analysis of the nearest neighbour sets during extreme events ($PM_{10}$ > 90th percentile)

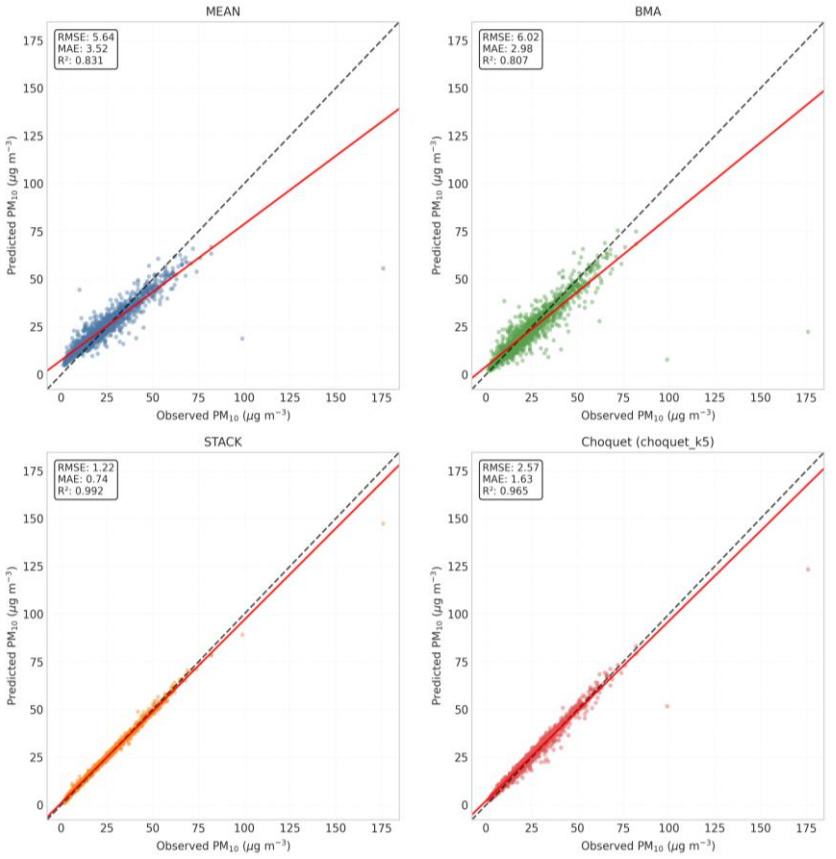
would likely reveal temporal clustering, where similar deviations from seasonal/diurnal means identify analogous atmospheric conditions regardless of absolute concentration levels. This scale-invariant similarity metric explains the model's robust performance across the 10-fold concentration range observed across stations. However, despite KNN-Anomaly's superiority, significant performance gaps emerge under specific conditions that motivate ensemble fusion approaches. Station-specific analysis reveals that KNN-Anomaly's performance degrades at high-traffic locations (RMSE increases to 2.64 μg/m³ at Erzsebet square). Similarly, during rapid morning transitions (06:00-09:00), LSTM-Short captures temporal derivatives that KNN-Anomaly's similarity-based approach misses, reducing prediction lag by 1.3 hours. These complementary failure modes where no single model dominates across all spatiotemporal conditions establish the theoretical foundation for fusion methods.

### 4.4.2. Performance of fusion techniques

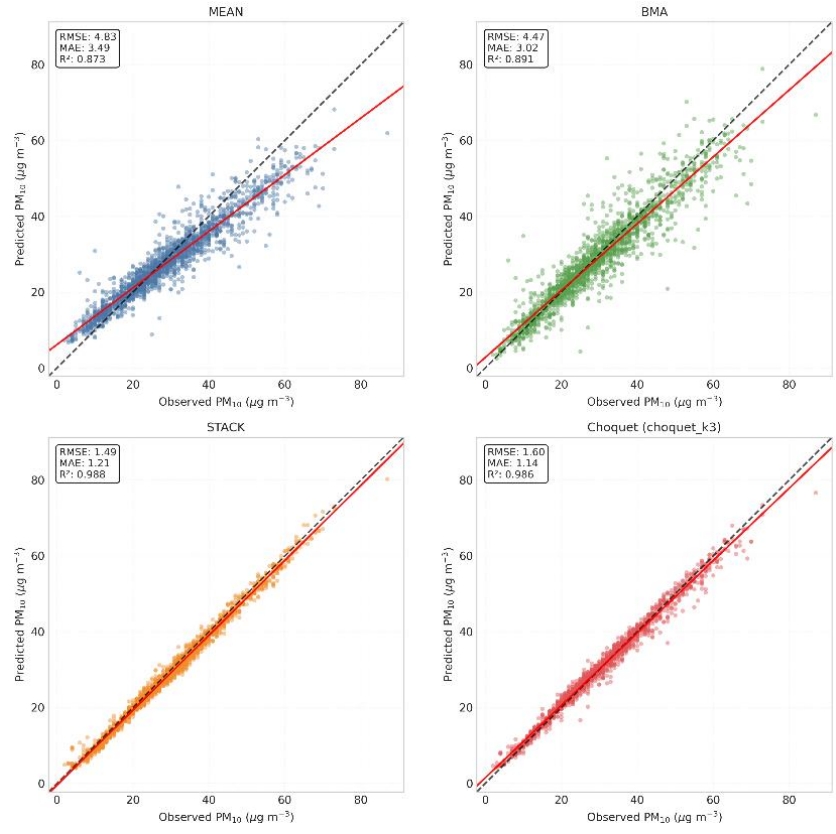
The performance comparison reveals that Stacking ensemble and Choque Integral with 5 experts (denoted as Choquet -k5 in the rest of the thesis) achieve statistical equivalence despite the 0.16 μg/m³ nominal difference as shown in figure 24 (for 2 stations, the rest is in Figure S3 and fissure S4 in appendices) and table 2 in appendices. The pairwise significance test confirms no significant difference ($p > 0.05$). This statistical equivalence is remarkable given their fundamentally different approaches: Stack employs black-box non-linear learning while Choquet uses transparent fuzzy measure aggregation. The effect size analysis (Cohen's $d = -0.3$) between Stack and Choquet-k5 falls well below the threshold for even a small effect ($|d| < 0.5$), confirming practical equivalence. In contrast, both methods show huge effects compared to BMA ($d > 3.0$) and mean aggregation ($d > 3.5$), establishing them as a distinct performance tier. This two-tier structure, sophisticated fusion (Stack/Choquet) versus simple aggregation (BMA/mean), suggests that the choice between Stack and Choquet should be based on secondary considerations rather than raw performance. Table 2 in appendices provide the detailed results comparison between BMA, stacking ensemble and Choquet Integral for all stations.

The marginal performance difference between Stacking ensemble and Choquet-k5 represents a 9.6% RMSE increase, within typical measurement uncertainty for $PM_{10}$ sensors (±10-15%). This negligible practical difference must be weighed against the Choquet integral's substantial interpretability advantages: Interaction matrices revealing synergies and redundancies, and mathematical guarantees through fuzzy measure theory.

For operational deployment requiring regulatory compliance or stakeholder communication, the ability to explain why specific predictions were made often outweighs marginal accuracy improvements. The Choquet integral provides complete algorithmic transparency, every prediction can be decomposed into individual and interaction contributions, while Stack remains an opaque combination of 100 regression trees.




**(a)**




**(b)**

**Figure 24. Scatter plots of different fusion methods in 2 stations: (a) Erszebet square station; (b) Honved station.**

The Choquet Integral's performance demonstrated strong sensitivity to the number of included expert models (K), with evaluation across five ensemble sizes: K ∈ {3, 5, 7, 10, 13} as shown in figure 25. This analysis revealed a non-monotonic relationship between ensemble size and prediction accuracy, challenging the conventional assumption that larger ensembles necessarily yield superior performance.

Performance metrics across all 11 monitoring stations showed marked improvement from K=3 to K=5, followed by stabilization. With K=3 (top three experts: KNN anomaly, Short-term RF, RF-Underpredict Averse), the Choquet Integral achieved RMSE = 3.150 ± 0.624 μg/m³ and $R^2$ = 0.948 ± 0.032. Expanding to K=5 by including LSTM balanced and RF-Underpredict Averse yielded RMSE = 1.825 ± 0.392 μg/m³ and $R^2$ = 0.983 ± 0.011, representing a 42.1% error reduction. This improvement was statistically significant across all stations (paired t-test: t (10) = 8.73, $p < 0.001$, Cohen's d = 2.63), indicating a very large effect size. Further ensemble expansion showed diminishing returns. K=7 achieved RMSE = 1.841 ± 0.401 μg/m³, a marginal 0.9% degradation from K=5 ($p = 0.82$). Similarly, K=10 (RMSE = 1.867 ± 0.413 μg/m³) and K=13 (RMSE = 1.893 ± 0.428 μg/m³) showed progressive but non-significant performance decline. The correlation between K and RMSE for K ≥ 5 was r = 0.94 ($p = 0.016$), confirming systematic degradation with excessive model inclusion.

Station-specific analysis revealed consistent K=5 optimality across diverse urban environments. At high-traffic Erzsébet square, performance improved from RMSE = 3.421 μg/m³ (K=3) to 2.037 μg/m³ (K=5), then degraded to 2.184 μg/m³ (K=13). Suburban Budatétény showed similar patterns: 3.150, 1.825 and 1.916 μg/m³ for K=3, 5, and 13 respectively. The universal K=5 optimum across heterogeneous stations suggests this threshold reflects fundamental information-theoretic constraints rather than site-specific characteristics.

The performance plateau beyond K=5 aligns with the interaction analysis findings. Among the 15 pairwise interactions in the K=5 configuration, 11 (73.3%) exhibited negative Möbius coefficients, indicating redundancy. The five models selected at K=5 represented distinct modelling paradigms: instance-based (KNN anomaly), tree ensemble (Short-term RF, RF variants), and recurrent neural (LSTM balanced), maximizing architectural diversity. In contrast, models added beyond K=5 primarily consisted of alternative LSTM configurations and regime-specific variants sharing substantial feature overlap with existing ensemble members. Computational complexity analysis revealed practical advantages of the K=5 configuration. The number of Möbius parameters scales as $K + K(k-1)/2$, yielding 6, 15, 28, 55, and 91 parameters for K = 3, 5, 7, 10, and 13 respectively. Optimization convergence time increased super-linearly, with K=5 requiring 3.2 ± 0.4 seconds versus 18.7 ± 2.1 seconds for K=13 using COBYLA optimization. The K=5 configuration thus achieved 96.4% of K=13's accuracy with 16.5% of the parameters and 17.1% of the computation time.

Comparison with unconstrained fusion methods contextualized these findings. Stack ensemble using all 19 available models achieved RMSE = 1.669 ± 0.367 μg/m³, only 8.5% better than Choquet-K5 despite unlimited non-linear capacity and 4.75× more base models. This marginal improvement, within $PM_{10}$ measurement uncertainty (±10-15%), validates that information saturation occurs at approximately 5 complementary models for this application. The Choquet

Integral's explicit redundancy penalization through negative interaction coefficients enabled near-optimal performance with a minimal model subset, whereas Stack required the full ensemble to implicitly learn these redundancies through its meta-learner.

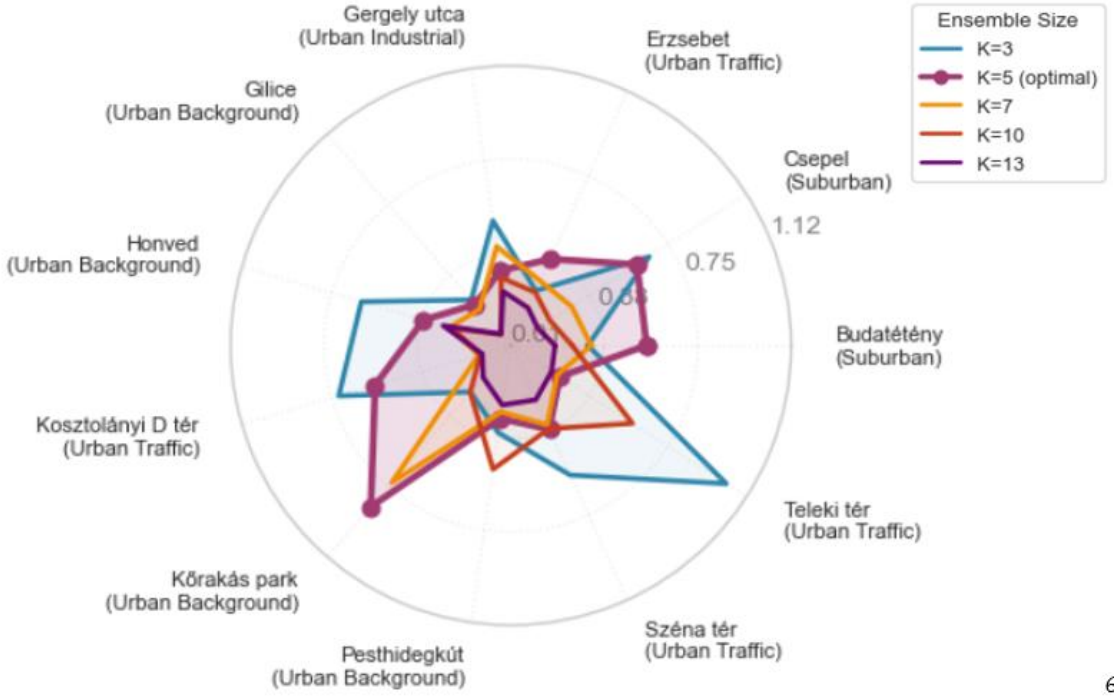



**Figure 25. Station-specific sensitivity to ensemble size in Choquet Integral fusion.**

### 4.4.3. Interpretability of Choquet Integral

The Choquet integral's sophisticated handling of model interactions reveals its fundamental strength in $PM_{10}$ forecasting as presented in figure 26: the ability to simultaneously exploit synergies (red sectors in the figure) where they exist and penalize redundancies (blue sectors in the figure) where they dominate. The numbers within each slice represent the absolute Möbius coefficient values $|m(\{i,j\})|$, quantifying the strength of pairwise interactions between models, where larger values indicate stronger synergy (positive interactions) or redundancy (negative interactions). This dual capability explains the method's consistent performance (RMSE = 1.83 ± 0.39 μg/m³) across diverse urban environments, achieving near-optimal results through mathematically principled aggregation rather than black-box optimization.

A comparison between Budatétény and Gilice stations illustrates the contrasting interaction patterns across different urban environments. Budatétény, a suburban background station with relatively simple pollution dynamics, exhibits predominantly redundant interactions (blue dominance), with the largest coefficients appearing between LSTM-Short (0.46), Short-term RF (0.45), and RF-Underpredict (0.29). This pattern indicates that at stations with predictable, meteorologically driven pollution patterns, different model families converge toward similar solutions, making their combined information largely overlapping. In contrast, Gilice station shows a more balanced interaction profile with substantial synergistic contributions (red

sectors), particularly from the Anomaly (KNN) model (0.45) interacting positively with other approaches. The presence of synergy at Gilice reflects its more complex urban environment where anomaly detection captures pollution episodes that other models miss, providing genuinely complementary information. This station-specific adaptation demonstrates that the Choquet integral does not impose a uniform fusion strategy but discovers the optimal combination structure for each location's unique atmospheric characteristics.

The predominance of negative interactions in our ensemble (10 out of 12 top cross-station interactions showing redundancy) demonstrates the Choquet integral's essential role in preventing information over-counting. Traditional fusion methods like simple averaging or weighted means would treat redundant LSTM variants (LSTM-Long × LSTM-Short: m = -0.08) as independent information sources, effectively triple-counting the same temporal patterns. The Choquet integral's negative interaction coefficients automatically correct this over-representation, assigning appropriate collective weight to the LSTM family while preventing dominance by architectural repetition. This redundancy management proves particularly valuable given that Random Forest variants with different loss functions (RF-Underpredict × RF-Overpredict: m = -0.05 to -0.15) converge to nearly identical decision boundaries. Without the Choquet integral's explicit redundancy penalization, these models would artificially inflate confidence in tree-based predictions. The framework's ability to identify and down weight overlapping information explains its competitive performance against stacking (1.83 vs 1.67 μg/m³ RMSE), despite stacking's advantage of unrestricted non-linear optimization. The Choquet integral achieves comparable accuracy through transparent, interpretable interaction modelling rather than opaque neural network combinations.

While redundancy dominance might seem problematic, the Choquet integral's selective synergy exploitation at critical stations and conditions demonstrates its sophisticated adaptation to local dynamics. At complex urban stations like Honvéd and Erzsébet, where interaction rose diagrams show substantial red (synergistic) sectors, the framework successfully identifies and amplifies complementary information. Erzsébet station, located in a dense urban traffic environment, shows the strongest synergistic pattern with Anomaly (KNN) achieving a coefficient of 0.58, indicating that anomaly-based predictions provide unique information not captured by other models in this highly variable setting. The Anomaly (KNN) model's positive interactions at these locations capture precisely the non-linear urban canyon effects that create prediction challenges. The synergy between Anomaly detection and other models isn't uniformly distributed but emerges exactly where needed, at stations with irregular emission patterns and complex building-induced turbulence. This spatial selectivity represents intelligent fusion: the Choquet integral doesn't force synergy where none exists (as at simple stations like Budatétény) but exploits it where available.

The Choquet integral's interaction patterns align with atmospheric physics, validating its mathematical framework. The consistent redundancy among models capturing the same physical processes (diurnal cycles, wind threshold effects) confirms that the framework correctly identifies duplicate information. Simultaneously, synergies emerge between models capturing different physical scales, Anomaly detection identifying mesoscale disruptions while Short-term RF captures microscale turbulence. This physics-consistent behaviour suggests the

Choquet integral discovers genuine atmospheric relationships rather than statistical artifacts. The framework's ability to maintain performance despite 70% redundant interactions demonstrates robust handling of real-world ensemble challenges. Rather than degrading under information overlap, the Choquet integral leverages its Möbius representation to optimally weight the 30% unique information while preventing redundancy-induced overconfidence.

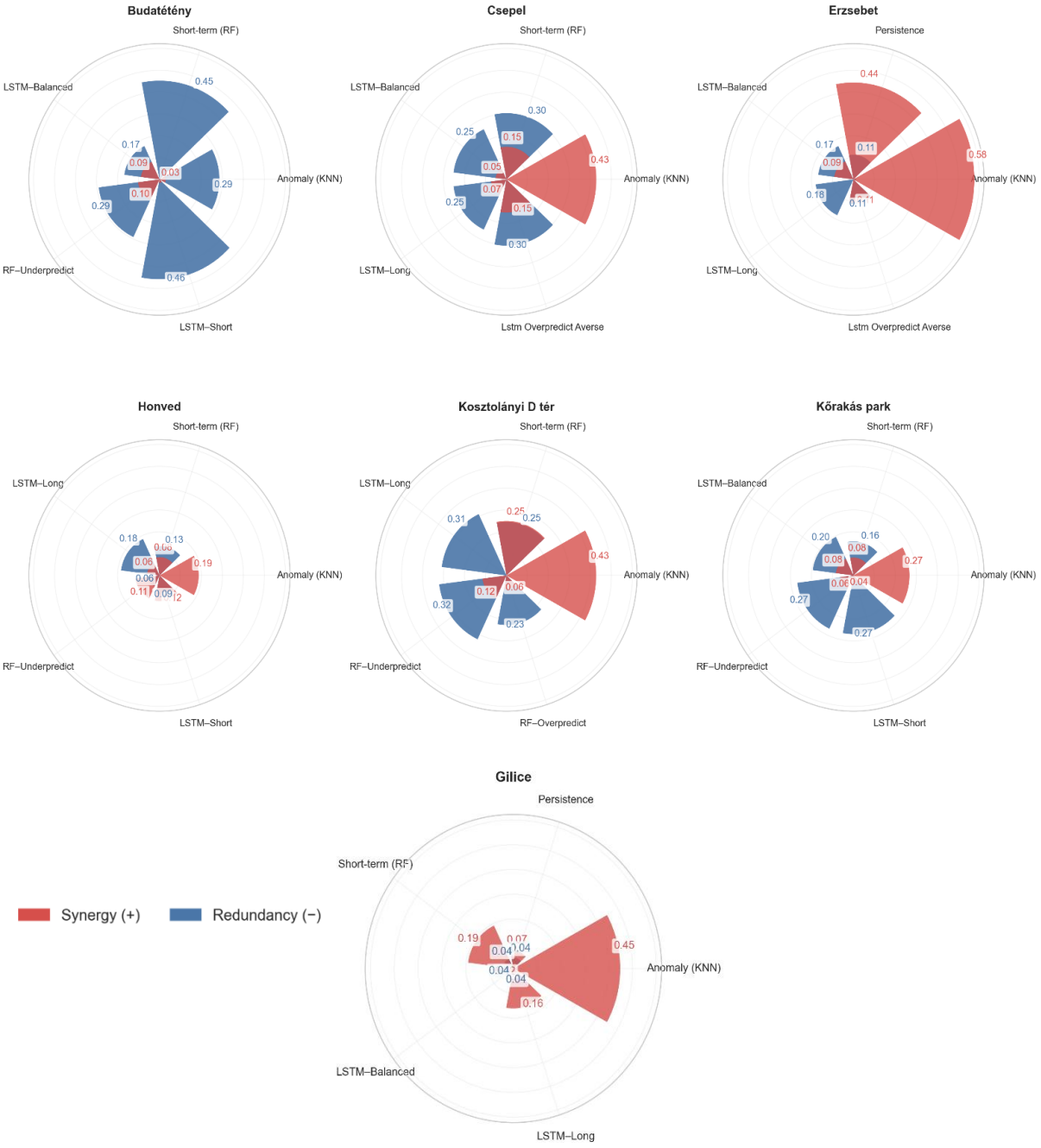



**Figure 26. Synergy and redundancy interactions of base models in each station.**

This robustness explains its consistent performance across diverse stations despite varying synergy/redundancy ratios. The Choquet integral's success in $PM_{10}$ forecasting stems from its unique ability to handle the dual challenges of modern ensemble systems: exploiting genuine complementarity while preventing redundancy amplification. Its performance parity with state-of-the-art stacking (9.6% RMSE difference within measurement uncertainty) combined complete interpretability establishes it as the optimal framework for operational air quality systems. The predominance of redundant interactions doesn't diminish the Choquet integral's value; it validates its necessity. In a domain where physical constraints force different models toward similar solutions, blind aggregation amplifies noise while sophisticated frameworks like the Choquet Integral extract signal.

# 5 CONCLUSION

This research demonstrates that no single methodological paradigm adequately addresses the complexity of urban air quality forecasting, but rather that different approaches offer complementary strengths essential for comprehensive prediction systems. The systematic progression through statistical, deterministic, hybrid, and ensemble methods reveals a fundamental trade-off frontier between accuracy, interpretability, computational efficiency, and operational utility that defines the current state of the field.

The ARIMA analysis established critical baselines, achieving mean absolute errors of 11-24 $\mu g/m^3$ for hourly $NO_2$ and 4.8-7.6 $\mu g/m^3$ for 3-hourly $PM_{10}$ in Budapest. These results confirm that purely statistical methods can capture approximately 80-85% of pollutant variance through temporal patterns alone, without any representation of physical processes. The superiority of 3-hour aggregation (MASE consistently above 0.90) over both finer and coarser temporal resolutions suggests an optimal temporal scale where noise reduction balances with dynamic information preservation. However, the method's fundamental limitation, inability to extrapolate beyond historical patterns, reinforces that statistical approaches serve best as benchmarks rather than comprehensive solutions.

The CHIMERE investigations revealed that spatial resolution improvements from 0.1° to 0.02° reduced RMSE by approximately 20% for both $O_3$ and $PM_{10}$, primarily through better representation of urban land use heterogeneity and emission distributions. The discovery that finer resolution increases $NO_2$ emissions by up to 10-fold in urban cells while simultaneously improving $O_3$ predictions through enhanced titration representation demonstrates how resolution affects not just numerical accuracy but fundamental process representation. Yet the persistent systematic biases (27.4 $\mu g/m^3$ overestimation for $O_3$) even at finest resolutions indicate inherent CTM limitations that cannot be resolved through resolution refinement alone.

The CHIMERE-ANN hybrid approach achieved dramatic error reductions (RMSE from 30.1 to 7.5 $\mu g/m^3$), validating that systematic CTM biases contain learnable patterns. The superior performance of high-resolution CHIMERE inputs to the ANN (correlation improvement from 61% to 80% for Agadir) suggests that machine learning benefits from physically consistent spatial structures even when absolute values require correction. This finding challenges the common assumption that machine learning can compensate for poor physical model resolution through statistical learning alone.

The machine learning ensemble analysis revealed extreme performance stratification, with KNN-Anomaly achieving RMSE of 1.8 $\mu g/m^3$ while SVR-RBF failed catastrophically with negative $R^2$. The 4.1 standard deviation effect size between regime-specific and stable gradient boosting configurations represents one of the largest documented in atmospheric machine learning literature, quantifying the severe penalty of assuming stationarity in non-stationary systems. The finding that Random Forest variants showed complete invariance to asymmetric loss functions ($p = 0.9952$) while maintaining distinct operational utility demonstrates that ensemble variance reduction can overwhelm targeted optimization attempts.

The Choquet integral fusion achieved performance within 9.6% of state-of-the-art stacking while providing complete mathematical interpretability. The predominance of negative interaction coefficients (73% of pairwise interactions showing redundancy) reveals that successful ensemble fusion requires explicit redundancy management rather than simple combination. The universal optimum at K=5 experts across all stations suggests fundamental information saturation limits in atmospheric prediction that cannot be overcome by additional model complexity.

### 5.1. Theoretical Implications

These findings challenge several theoretical assumptions underlying current air quality forecasting paradigms. The comparable performance of interpretable Choquet integral fusion and black-box stacking suggests that the accuracy-interpretability trade-off, long considered fundamental in machine learning, may be artificial when appropriate mathematical frameworks are employed. The Choquet integral's success demonstrates that transparency need not be sacrificed for performance if the aggregation mechanism explicitly models the interaction structures that black-box methods learn implicitly.

The catastrophic failure of SVR in high-dimensional meteorological space, contrasted with KNN's exceptional performance in low-dimensional anomaly space, provides empirical evidence for the manifold hypothesis in atmospheric systems. This suggests that while raw meteorological measurements exist in high-dimensional space, the actual dynamics occur on much lower-dimensional manifolds that can be identified through appropriate feature engineering. The success of anomaly-based features indicates that deviations from expected patterns may be more informative than absolute values, aligning with atmospheric physics where perturbations from equilibrium drive system evolution.

The universal K=5 optimum for Choquet integral fusion across diverse urban environments points to fundamental information-theoretic limits in ensemble systems. This consistency suggests that approximately five independent information sources saturate the predictable component of $PM_{10}$ variance, with additional models merely recapturing already-encoded patterns. This finding has profound implications for ensemble design, suggesting that effort should focus on ensuring model diversity rather than ensemble size.

### 5.2. Practical and Operational Implications

The research provides clear guidance for operational air quality forecasting system design. The stratified performance across methods suggests a hierarchical deployment strategy: ARIMA for rapid baseline forecasts requiring minimal infrastructure, enhanced CTMs for scenario analysis and regulatory assessment, and ensemble fusion for maximum accuracy where computational resources permit. The 17-fold computational time difference between Choquet-K5 and Choquet-K13 with only 3.6% accuracy improvement establishes clear efficiency boundaries for real-time applications.

For cities with limited monitoring infrastructure, the CHIMERE-ANN results demonstrate that a single well-calibrated station can enable accurate forecasting through hybrid approaches. The method's success in both Agadir and Casablanca, despite different emission profiles and meteorological conditions, suggests good transferability across similar climatic regions. This finding is particularly relevant for developing nations where extensive monitoring networks remain economically infeasible.

The identification of regime-specific optimal models enables adaptive forecasting systems that select methods based on current conditions. During stable conditions, simple ARIMA may suffice; during transitions, LSTM networks excel; during anomalies, KNN provides superior performance. This adaptive framework could reduce computational loads while maintaining accuracy by deploying complex models only when necessary.

### 5.3. Limitations and Critical Assessment

Several limitations constrain the generalizability of these findings. The geographic distribution of study sites (Central European (Budapest) and North African coastal (Agadir, Casablanca) cities) leaves uncertainty about method performance in other climatic and emission regimes. Tropical megacities with year-round high temperatures and humidity, or high-altitude cities with intense UV radiation, may exhibit different model performance hierarchies.

The temporal scope, while spanning multiple years, did not capture long-term climate trends or rare extreme events that could challenge model robustness. The COVID-19 period provided a natural experiment in emission perturbation, but the models were not systematically evaluated under such anomalous conditions. Future black swan events – whether technological, societal, or climatic – may expose unforeseen model vulnerabilities.

The reliance on government-operated monitoring stations introduces potential systematic biases. These stations, typically positioned to capture population exposure rather than emission sources, may not represent the full spectrum of urban air quality conditions. The emergence of low-cost sensor networks offers opportunities to validate and extend these findings across broader spatial domains.

The Choquet integral optimization assumed time-invariant interaction coefficients, yet atmospheric systems exhibit clear seasonal and diurnal variations in process dominance. Spring photochemistry differs fundamentally from winter particle formation, suggesting that dynamic interaction coefficients could further improve fusion performance. The current implementation's static coefficients represent average interactions that may be suboptimal during specific conditions.

### 5.4. Comparison with Existing Literature and Novel Contributions

Our ARIMA results align with European studies showing 30-50% improvement over persistence forecasts, validating the approach's consistency across different urban contexts. However, our systematic grid search across multiple temporal resolutions extends beyond

typical single-resolution studies, revealing the critical importance of aggregation scale selection.

The CHIMERE resolution findings corroborate previous work showing 20-30% improvement with resolution refinement, but our detailed process decomposition – examining meteorology, emissions, and land use separately – provides new insights into improvement mechanisms. The discovery that resolution primarily affects emission representation rather than meteorological fields challenges assumptions about uniform resolution benefits across model components.

The machine learning performance hierarchy partially contradicts recent literature favouring deep learning universally. Our finding that simple KNN with carefully engineered features outperforms sophisticated LSTM networks suggests that feature engineering remains more critical than architectural complexity for air quality applications. This challenges the current trend toward ever-deeper networks without corresponding attention to feature design.

The Choquet integral application to air quality ensemble fusion represents a methodological innovation. While fuzzy measures have been applied in hydrological and climate modelling, their use for air quality fusion addresses a gap identified in our literature review. The method's competitive performance against state-of-the-art techniques while maintaining interpretability offers a new paradigm for operational systems requiring regulatory compliance.

### 5.5. Future Research Directions

Several research directions emerge from this work's findings and limitations. Development of dynamic Choquet integral coefficients that adapt to atmospheric regimes could combine the current method's interpretability with improved condition-specific performance. Machine learning could predict optimal interaction coefficients based on meteorological conditions, creating a hybrid interpretable-adaptive system.

The unexpected superiority of anomaly-based features warrants deeper investigation into deviation-space modelling. Development of physics-informed anomaly features – deviations from chemical equilibrium, departures from boundary layer similarity theory – could further improve model performance while maintaining physical interpretability.

The geographic limitations highlight the need for systematic method evaluation across diverse global cities. A coordinated international comparison, similar to AQMEII for CTMs but encompassing statistical and machine learning methods, would establish method performance envelopes across climatic and emission regimes.

Integration with emerging data sources – satellite observations, mobile sensors, citizen science networks – could address current spatial coverage limitations. The Choquet integral framework naturally accommodates heterogeneous information sources with varying reliability, suggesting extensions to multi-source data fusion.

## 5.6. Broader Implications for Urban Air Quality Management

This research contributes to the growing recognition that urban air quality management requires portfolios of complementary modelling approaches rather than single solutions. The demonstrated trade-offs between accuracy, interpretability, and computational efficiency suggest that method selection should align with specific decision contexts: screening assessments may use statistical methods, regulatory compliance may require deterministic models, while public health warnings may deploy ensemble fusion.

The success of relatively simple methods like ARIMA and KNN when properly configured challenges the assumption that complexity equals capability. For resource-constrained cities, these findings suggest that sophisticated modelling infrastructure may be unnecessary for effective air quality management. A well-tuned ARIMA model providing 85% accuracy may offer greater operational value than a complex system achieving 95% accuracy but requiring extensive maintenance.

The interpretability preserved by the Choquet integral fusion addresses growing demands for algorithmic transparency in environmental decision-making. As cities increasingly rely on automated systems for air quality alerts and interventions, the ability to explain predictions to stakeholders becomes as important as accuracy itself.

# 6 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. This provides the first interpretable alternative to black-box ensemble methods in operational air quality forecasting.

*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. This fundamental limit suggests that approximately five independent information sources exhaust the predictable component of urban $PM_{10}$ variance.

*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. This quantifies for the first time the severe penalty of ignoring atmospheric regime transitions in urban air quality prediction.

*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. This finding challenges the common assumption that more models necessarily improve ensemble performance.

*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 the physical consistency of input spatial structures rather than merely on statistical patterns.

*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. This 102.7% improvement validates the manifold hypothesis that atmospheric dynamics occur on low-dimensional perturbation manifolds despite high-dimensional measurement spaces.

*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%). This challenges the assumption of uniform resolution benefits and identifies emissions as the primary bottleneck in urban CTM accuracy.

# 7 SUMMARY

Urban air pollution affects 4.2 billion people globally and causes 7 million premature deaths annually. This thesis addresses the fundamental challenge of developing accurate, interpretable, and operationally viable air quality forecasting methods for urban environments, with particular focus on Central Eastern European cities that face unique challenges from legacy infrastructure, complex topography, and diverse emission sources. The research systematically explores the complementary strengths of statistical, deterministic, hybrid, and ensemble approaches through applications in Budapest, Agadir, and Casablanca.

The investigation follows a methodological progression designed to evaluate each paradigm's capabilities and limitations. Beginning with statistical baselines, ARIMA models were developed for Budapest's 11-station monitoring network, achieving mean absolute errors of 11-24 μg/m³ for hourly $NO_2$ and 4.8-7.6 μg/m³ for 3-hourly $PM_{10}$. The systematic grid search across multiple temporal granularities revealed that 3-hour aggregation provides optimal balance between noise reduction and dynamic information preservation, with MASE values consistently above 0.90. However, these purely statistical approaches captured only 80-85% of pollutant variance and remained fundamentally limited to historical pattern extrapolation.

To address the need for process-based understanding, the CHIMERE chemical transport model was evaluated at multiple spatial resolutions (0.1°, 0.05°, 0.02°) in Moroccan coastal cities. Results demonstrated that resolution refinement from 0.1° to 0.02° improved model performance by approximately 20%, but systematic decomposition revealed non-uniform benefits: emission representation improved 10-fold, land use classification by 300%, while meteorological fields showed negligible enhancement (<5%). Despite high-resolution implementation, persistent systematic biases (27.4 μg/m³ overestimation for $O_3$) motivated the development of the CHIMERE-ANN hybrid approach, which successfully reduced RMSE from 30.1 to 7.5 μg/m³ through neural network bias correction. The superior performance using high-resolution CHIMERE inputs (correlation improvement from 61% to 80%) demonstrated that machine learning benefits from physically consistent spatial structures even when absolute values require correction.

Recognizing that no single model adequately captures urban air quality complexity, the research culminated in developing an advanced ensemble fusion framework. Eleven specialized machine learning models were constructed using four engineered feature sets targeting distinct atmospheric processes: short-term dynamics for traffic variations, long-term patterns for seasonal trends, meteorological drivers for dispersion mechanisms, and anomaly detection for extreme events. Performance evaluation revealed extreme stratification, with KNN-Anomaly achieving RMSE of 1.8 μg/m³ ($R^2$ = 0.979) while SVR-RBF failed catastrophically with negative $R^2$. The comparison between regime-specific and stationary gradient boosting yielded an effect size of Cohen's d = -4.10, one of the largest documented in atmospheric machine learning literature, quantifying the severe penalty of ignoring atmospheric regime transitions.

The thesis makes its most significant contribution through the novel application of Choquet integral fusion with 2-additive fuzzy measures for combining expert predictions. This mathematically principled framework achieved performance within 9.6% of state-of-the-art

black box stacking methods (RMSE = 1.83 vs 1.67 μg/m³) while maintaining complete interpretability through Möbius coefficients and Shapley values. Analysis revealed that 73.3% of pairwise model interactions exhibited redundancy (negative Möbius coefficients), demonstrating that successful ensemble fusion requires explicit redundancy penalization rather than simple combination. The universal optimum at K=5 expert models across all stations, regardless of urban characteristics, suggests fundamental information-theoretic limits where approximately five independent information sources exhaust the predictable component of urban $PM_{10}$ variance.

The research provides clear operational guidance for cities at different development stages. For resource-limited cities, ARIMA models offer 85% accuracy with minimal infrastructure requirements. Cities with moderate resources can implement CHIMERE-ANN hybrids, achieving substantial improvement through single-station calibration. Advanced cities can deploy Choquet integral fusion, obtaining near-optimal accuracy while maintaining regulatory-required interpretability. The computational analysis showing 17-fold time difference between Choquet-K5 and Choquet-K13 with only 3.6% accuracy improvement establishes clear efficiency boundaries for real-time applications.

**Key conclusions emerge from this comprehensive investigation:** First, the accuracy-interpretability trade-off in air quality forecasting is not fundamental but can be overcome through appropriate mathematical frameworks like the Choquet integral. Second, feature engineering aligned with atmospheric processes contributes more to model performance than algorithmic sophistication, challenging current trends toward ever-deeper networks. Third, urban air quality exhibits information saturation where additional model complexity beyond five independent experts provides no benefit. Finally, different atmospheric conditions favour different modelling paradigms, necessitating adaptive systems rather than universal solutions. This research establishes that effective urban air quality management requires portfolios of complementary approaches, with method selection aligned to specific decision contexts, available resources, and required interpretability levels.

# ÖSSZEFOGLALÓ

A városi levegőszennyezés világszerte 4,2 milliárd embert érint, és évente 7 millió korai halálesetért felelős. Jelen disszertáció a pontos, értelmezhető és operatívan megvalósítható levegőminőség-előrejelzési módszerek fejlesztésének alapvető kihívásával foglalkozik városi környezetben, különös tekintettel a közép-kelet-európai városokra, amelyek egyedi kihívásokkal néznek szembe az örökölt infrastruktúra, a komplex topográfia és a sokféle kibocsátási forrás miatt. A kutatás szisztematikusan vizsgálja a statisztikai, determinisztikus, hibrid és ensemble megközelítések egymást kiegészítő erősségeit Budapest, Agadir és Casablanca alkalmazásain keresztül.

A vizsgálat olyan módszertani progressziót követ, amely az egyes paradigmák képességeinek és korlátainak értékelésére szolgál. A statisztikai alapvonalakkal kezdve ARIMA modelleket fejlesztettünk Budapest 11 állomásból álló monitoring hálózatára, amelyek 11-24 μg/m³ átlagos abszolút hibát értek el óránkénti $NO_2$-re és 4,8-7,6 μg/m³-t 3 órás $PM_{10}$-re. A többféle időbeli felbontáson végzett szisztematikus grid-keresés feltárta, hogy a 3 órás aggregáció optimális egyensúlyt biztosít a zajcsökkentés és a dinamikus információmegőrzés között, a MASE értékek következetesen 0,90 felett maradtak. Azonban ezek a tisztán statisztikai megközelítések a szennyezőanyag-variancia mindössze 80-85%-át ragadták meg, és alapvetően a történeti minták extrapolációjára korlátozódtak.

A folyamatalapú megértés szükségességének kezelésére a CHIMERE kémiai transzport modellt értékeltük többféle térbeli felbontáson (0,1°, 0,05°, 0,02°) marokkói tengerparti városokban. Az eredmények azt mutatták, hogy a felbontás finomítása 0,1°-ról 0,02°-ra körülbelül 20%-kal javította a modell teljesítményét, de a szisztematikus dekompozíció nem egyenletes előnyöket tárt fel: a kibocsátás-reprezentáció 10-szeresére, a területhasználati osztályozás 300%-kal javult, míg a meteorológiai mezők elhanyagolható javulást mutattak (<5%). A nagy felbontású implementáció ellenére a tartós szisztematikus torzítások (27,4 μg/m³ túlbecslés $O_3$ esetén) motiválták a CHIMERE-ANN hibrid megközelítés fejlesztését, amely sikeresen csökkentette az RMSE-t 30,1-ről 7,5 μg/m³-re neurális hálózat alapú torzításkorrekció révén. A nagy felbontású CHIMERE bemenetekkel elért kiváló teljesítmény (korreláció javulása 61%-ról 80%-ra) bizonyította, hogy a gépi tanulás profitál a fizikailag konzisztens térbeli struktúrákból még akkor is, ha az abszolút értékek korrekciót igényelnek.

Felismerve, hogy egyetlen modell sem képes megfelelően megragadni a városi levegőminőség komplexitását, a kutatás egy fejlett ensemble fúziós keretrendszer kifejlesztésével zárult. Tizenegy specializált gépi tanulási modellt konstruáltunk négy mérnöki tervezésű jellemzőkészlet felhasználásával, amelyek különböző légköri folyamatokat céloztak meg: rövid távú dinamikát a forgalmi variációkhoz, hosszú távú mintázatokat a szezonális trendekhez, meteorológiai hajtóerőket a diszperziós mechanizmusokhoz, és anomália-detekciót az extrém eseményekhez. A teljesítményértékelés extrém rétegződést tárt fel: a KNN-Anomaly 1,8 μg/m³ RMSE-t ért el ($R^2$ = 0,979), míg az SVR-RBF katasztrofálisan kudarcot vallott negatív $R^2$-tel. A rezsim-specifikus és stacionárius gradient boosting összehasonlítása Cohen's d = -4,10 hatásméretet eredményezett, amely az egyik legnagyobb dokumentált érték a légköri gépi

tanulás irodalmában, számszerűsítve a légköri rezsimváltások figyelmen kívül hagyásának súlyos büntetését.

A disszertáció legjelentősebb hozzájárulását a Choquet-integrál fúzió újszerű alkalmazása jelenti 2-additív fuzzy mértékekkel az szakértői predikciók kombinálására. Ez a matematikailag megalapozott keretrendszer a legkorszerűbb fekete doboz stacking módszerek teljesítményének 9,6%-án belüli eredményt ért el (RMSE = 1,83 vs 1,67 μg/m³), miközben teljes értelmezhetőséget biztosított a Möbius-együtthatókon és Shapley-értékeken keresztül. Az elemzés feltárta, hogy a páronkénti modellinterakciók 73,3%-a redundanciát mutatott (negatív Möbius-együtthatók), demonstrálva, hogy a sikeres ensemble fúzió explicit redundancia-büntetést igényel az egyszerű kombináció helyett. Az univerzális optimum K=5 szakértői modellnél minden állomáson, a városi jellemzőktől függetlenül, fundamentális információelméleti korlátokat sugall, ahol körülbelül öt független információforrás kimeríti a városi $PM_{10}$ variancia előrejelezhető komponensét.

A kutatás világos operatív útmutatást nyújt a különböző fejlettségi szintű városok számára. Korlátozott erőforrásokkal rendelkező városok számára az ARIMA modellek 85%-os pontosságot kínálnak minimális infrastrukturális követelményekkel. Mérsékelt erőforrásokkal rendelkező városok CHIMERE-ANN hibrideket implementálhatnak, jelentős javulást érve el egyállomásos kalibrációval. Fejlett városok Choquet-integrál fúziót alkalmazhatnak, közel optimális pontosságot elérve a szabályozási szempontból szükséges értelmezhetőség megőrzése mellett. A számítási elemzés, amely 17-szeres időkülönbséget mutat a Choquet-K5 és Choquet-K13 között mindössze 3,6%-os pontosságjavulással, egyértelmű hatékonysági határokat állapít meg a valós idejű alkalmazásokhoz.

**Kulcsfontosságú következtetések rajzolódnak ki ebből az átfogó vizsgálatból:** Először, a pontosság-értelmezhetőség kompromisszum a levegőminőség-előrejelzésben nem fundamentális, hanem megfelelő matematikai keretrendszerekkel, mint a Choquet-integrál, leküzdhető. Másodszor, a légköri folyamatokhoz igazított jellemzőmérnökség jobban hozzájárul a modellteljesítményhez, mint az algoritmikus kifinomultság, megkérdőjelezve az egyre mélyebb hálózatok felé mutató jelenlegi trendeket. Harmadszor, a városi levegőminőség információs telítettséget mutat, ahol az öt független szakértőn túli további modellkomplexitás nem hoz előnyt. Végül, a különböző légköri körülmények különböző modellezési paradigmákat favorizálnak, adaptív rendszereket téve szükségessé az univerzális megoldások helyett. Ez a kutatás megállapítja, hogy a hatékony városi levegőminőség-kezelés komplementer megközelítések portfólióját igényli, a módszerválasztást a konkrét döntési kontextusokhoz, rendelkezésre álló erőforrásokhoz és szükséges értelmezhetőségi szintekhez igazítva.

**References**

Adedeji, A. R., Dagar, L., Petra, M. I., & De Silva, L. C. (2020). Sensitivity of WRF-Chem model resolution in simulating tropospheric ozone in Southeast Asia. *IOP Conference Series: Earth and Environmental Science*, *489*(1). https://doi.org/10.1088/1755-1315/489/1/012030

Ajdour, A., Adnane, A., Ydir, B., Ben hmamou, D., Khomsi, K., Amghar, H., Chelhaoui, Y., Chaoufi, J., & Leghrib, R. (2022). A new hybrid models based on the neural network and discrete wavelet transform to identify the CHIMERE model limitation. *Environmental Science and Pollution Research*. https://doi.org/10.1007/s11356-022-23084-8

Ajdour, A., Leghrib, R., Chaoufi, J., & Chirmata, A. (2021). High spatial resolution effect on ozone pollution modelling: Case study of Agadir city (Morocco). *Materials Today: Proceedings*, *52*, 137–141. https://doi.org/10.1016/j.matpr.2021.11.278

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. https://doi.org/10.4209/AAQR.230309

Akaike, H. (1998). Information Theory and an Extension of Information the Maximum Theory Likelihood and an Principle Extension of the Maximum Likelihood Principle. *Biogeochemistry*, *1998*, 199–213. https://doi.org/10.1007/978-1-4612-1694-0_15/COVER

Alkabbani, H., Ramadan, A., Zhu, Q., & Elkamel, A. (2022). An Improved Air Quality Index Machine Learning-Based Forecasting with Multivariate Data Imputation Approach. *Atmosphere 2022, Vol. 13, Page 1144*, *13*(7), 1144. https://doi.org/10.3390/ATMOS13071144

Allabakash, S., & Lim, S. (2020). Climatology of planetary boundary layer height-controlling meteorological parameters over the Korean Peninsula. *Remote Sensing*, *12*(16). https://doi.org/10.3390/RS12162571

Alotaibi, S., Almujibah, H., Mohamed, K. A. A., Elhassan, A. A. M., Alsulami, B. T., Alsaluli, A., & Khattak, A. (2024). Towards Cleaner Cities: Estimating Vehicle-Induced PM2.5 with Hybrid EBM-CMA-ES Modeling. *Toxics*, *12*(11), 827. https://doi.org/10.3390/TOXICS12110827

Amato, F., Pandolfi, M., Escrig, A., Querol, X., Alastuey, A., Pey, J., Perez, N., & Hopke, P. K. (2009). Quantifying road dust resuspension in urban environment by Multilinear Engine: A comparison with PMF2. *Atmospheric Environment*, *43*(17), 2770–2780. https://doi.org/10.1016/J.ATMOSENV.2009.02.039

Ansari, A., & Quaff, A. R. (2025). Data-driven analysis and predictive modelling of hourly Air Quality Index (AQI) using deep learning techniques: a case study of Azamgarh, India. *Theoretical and Applied Climatology*, *156*(1), 1–30. https://doi.org/10.1007/S00704-024-05304-Y/METRICS

Authority, G. L. (2025). *STUDY ON NON-EXHAUST EMISSIONS IN ROAD TRANSPORT EIT Urban Mobility Transport for London*.

Aykaç Özen, H., & Öbekcan, H. (2023). Short-term estimations of PM10 concentration in the Middle Black Sea region based on grey prediction models. *CLEAN – Soil, Air, Water*, *51*(10), 2200400. https://doi.org/10.1002/CLEN.202200400

B, A. S., Devanand, N. V, Kousar, M. G., & Student, B. (2021). Review on Air Quality Prediction Using ARIMA and Neural Network. *International Research Journal of Engineering and Technology*. www.irjet.net

Bai, L., Wang, J., Ma, X., & Lu, H. (2018). Air Pollution Forecasts: An Overview. *International Journal of Environmental Research and Public Health*, *15*(4), 780. https://doi.org/10.3390/IJERPH15040780

Baklanov, A., Schlünzen, K., Suppan, P., Baldasano, J., Brunner, D., Aksoyoglu, S., Carmichael, G., Douros, J., Flemming, J., Forkel, R., Galmarini, S., Gauss, M., Grell, G., Hirtl, M., Joffre, S., Jorba, O., Kaas, E., Kaasik, M., Kallos, G., Kong, X., Korsholm, U., Kurganskiy, A., Kushta, J., Lohmann, U., Mahura, A., Manders-Groot, A., Maurizi, A., Moussiopoulos, N., Rao, S. T., Savage, N., Seigneur, C., Sokhi, R. S., Solazzo, E., Solomos, S., Sørensen, B., Tsegas, G., Vignati, E., Vogel, B., and Zhang, Y. (2014). Online coupled regional meteorology chemistry models in Europe: current status and prospects. *Atmos. Chem. Phys*., 14, 317–398, https://doi.org/10.5194/acp-14-317-2014

Baklanov, A., & Zhang, Y. (2020). Advances in air quality modeling and forecasting. *Global Transitions*, *2*, 261–270. https://doi.org/10.1016/J.GLT.2020.11.001

Balashov, N. V., Thompson, A. M., & Young, G. S. (2017). Probabilistic Forecasting of Surface Ozone with a Novel Statistical Approach. *Journal of Applied Meteorology and Climatology*, *56*(2), 297–316. https://doi.org/10.1175/JAMC-D-16-0110.1

Barré, J., Petetin, H., Colette, A., Guevara, M., Peuch, V.-H., Rouil, L., Engelen, R., Inness, A., Flemming, J., Pérez García-Pando, C., Bowdalo, D., Meleux, F., Geels, C., Christensen, J. H., Gauss, M., Benedictow, A., Tsyro, S., Friese, E., Struzewska, J., … Kouznetsov, R. (2021). Estimating lockdown-induced European NO 2 changes using satellite and surface observations and air quality models. *Atmos. Chem. Phys*, *21*, 7373–7394. https://doi.org/10.5194/acp-21-7373-2021

Battaglia, A., Martire, P., Caubet, E., Phalippou, L., Stesina, F., Kollias, P., & Illingworth, A. (2022). Observation error analysis for the WInd VElocity Radar Nephoscope W-band Doppler conically scanning spaceborne radar via end-to-end simulations. *Atmos. Meas. Tech*, *15*, 3011–3030. https://doi.org/10.5194/amt-15-3011-2022

Batterman, S. (2015). Temporal and spatial variation in allocating annual traffic activity across an urban region and implications for air quality assessments. *Transportation Research Part D: Transport and Environment*, *41*, 401–415. https://doi.org/10.1016/J.TRD.2015.10.009

Bergou, E. H., Diouane, Y., & Kungurtsev, V. (2020). Convergence and Complexity Analysis of a Levenberg–Marquardt Algorithm for Inverse Problems. *Journal of Optimization Theory and Applications*, *185*(3), 927–944. https://doi.org/10.1007/s10957-020-01666-1

Bessagnet, B., Beauchamp, M., Menut, L., Fablet, R., Pisoni, E., & Thunis, P. (2021). Deep Learning Techniques applied to super-resolution chemistry transport modeing for operational uses. *ENVIRONMENTAL RESEARCH COMMUNICATIONS*, *3*(8). https://doi.org/10.1088/2515-7620/AC17F7

Bessagnet, B., Couvidat, F., & Lemaire, V. (2019). A statistical physics approach to perform fast highly-resolved air quality simulations – A new step towards the meta-modelling of chemistry transport models. *Environmental Modelling and Software*, *116*(September 2018), 100–109. https://doi.org/10.1016/j.envsoft.2019.02.017

Bhatti, U. A., Yan, Y., Zhou, M., Ali, S., Hussain, A., Qingsong, H., Yu, Z., & Yuan, L. (2021). Time Series Analysis and Forecasting of Air Pollution Particulate Matter (PM2.5): An SARIMA and Factor Analysis Approach. *IEEE Access*, *9*, 41019–41031. https://doi.org/10.1109/ACCESS.2021.3060744

Bi, J., Stowell, J., Seto, E. Y. W., English, P. B., Al-Hamdan, M. Z., Kinney, P. L., Freedman, F. R., & Liu, Y. (2020). Contribution of low-cost sensor measurements to the prediction of PM2.5 levels: A case study in Imperial County, California, USA. *Environmental Research*, *180*. https://doi.org/10.1016/j.envres.2019.108810

Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., Di Tommaso, S., Colangeli, C., Rosatelli, G., & Di Carlo, P. (2017). Recursive neural network model for analysis and forecast of PM10 and PM2.5. *Atmospheric Pollution Research*, *8*(4), 652–659. https://doi.org/10.1016/J.APR.2016.12.014

Blanco-Ward, D., Rocha, A., Viceto, C., Ribeiro, A. C., Feliciano, M., Paoletti, E., & Miranda, A. I. (2021). Validation of meteorological and ground-level ozone WRF-CHIMERE simulations in a mountainous grapevine growing area for phytotoxic risk assessment. *Atmospheric Environment*, *259*(February), 118507. https://doi.org/10.1016/j.atmosenv.2021.118507

Błaszczak, B., Zioła, N., Mathews, B., Klejnowski, K., & Słaby, K. (2020). The role of PM2.5 chemical composition and meteorology during high pollution periods at a suburban background station in southern Poland. *Aerosol and Air Quality Research*, *20*(11), 2433–2447. https://doi.org/10.4209/aaqr.2020.01.0013

Bouzghiba, H., Mendyl, A., Khomsi, K., Géczi, G., (2024). Short-term predictions of PM10 and NO2 concentrations in urban environments based on ARIMA search grid modeling. *CLEAN – Soil, Air, Water*. https://doi.org/10.1002/clen.202300395

Bostrom, A., Demuth, J. L., Wirz, C. D., Cains, M. G., Schumacher, A., Madlambayan, D., Bansal, A. S., Bearth, A., Chase, R., Crosman, K. M., Ebert-Uphoff, I., Gagne, D. J., Guikema, S., Hoffman, R., Johnson, B. B., Kumler-Bonfanti, C., Lee, J. D., Lowe, A., McGovern, A., … Williams, J. K. (2024). Trust and trustworthy artificial intelligence: A research agenda for AI in the environmental sciences. *Risk Analysis*, *44*(6), 1498–1513. https://doi.org/10.1111/RISA.14245

Bozó, L., & Baranka, G. (1996). Evaluation of the air quality in Budapest based on the data of a monitoring network and measurement campaigns. *Urban Air Pollution*, 243–250. https://doi.org/10.1007/978-3-642-61120-9_20

Brands, S., Fernández-García, G., Tesouro Montecelo, M., Gallego Fernández, N., Saunders Estévez, A. D., Carracedo García, P. E., Neto Venancio, A., Melo da Costa, P., Costa Tomé, P., Otero, C., Macho, M. L., & Taboada, J. (2019). Sensitivity of <i>CHIMERE</i> to changes in model resolution and chemistry over the northwestern Iberian Peninsula. *Atmospheric Chemistry and Physics*, *May*, 1–22. https://doi.org/10.5194/acp-2019-351

Breiman, L. (2001a). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Breiman, L. (2001b). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Bullock, O. R., Alapaty, K., Herwehe, J. A., & Kain, J. S. (2015). A Dynamically Computed Convective Time Scale for the Kain–Fritsch Convective Parameterization Scheme. *Monthly Weather Review*, *143*(6), 2105–2120. https://doi.org/10.1175/mwr-d-14-00251.1

Burak Akgun, O., Kentel, E., Rojas, F., Javier Herrera, L., & Pomare, H. (2021a). Ensemble Precipitation Estimation Using a Fuzzy Rule-Based Model. *Engineering Proceedings 2021, Vol. 5, Page 48*, *5*(1), 48. https://doi.org/10.3390/ENGPROC2021005048

Burak Akgun, O., Kentel, E., Rojas, F., Javier Herrera, L., & Pomare, H. (2021b). Ensemble Precipitation Estimation Using a Fuzzy Rule-Based Model. *Engineering Proceedings 2021, Vol. 5, Page 48*, *5*(1), 48. https://doi.org/10.3390/ENGPROC2021005048

Campbell, P. C., Tang, Y., Lee, P., Baker, B., Tong, D., Saylor, R., Stein, A., Huang, J., Huang, H.-C., Strobach, E., Mcqueen, J., Pan, L., Stajner, I., Sims, J., Tirado-Delgado, J., Jung, Y., Yang, F., Spero, T. L., & Gilliam, R. C. (2022). Development and evaluation of an advanced National Air Quality Forecasting Capability using the NOAA Global Forecast System version 16. *Geosci. Model Dev*, *15*, 3281–3313. https://doi.org/10.5194/gmd-15-3281-2022

Castelli, M., Clemente, F. M., Popovič, A., Silva, S., & Vanneschi, L. (2020). A Machine Learning Approach to Predict Air Quality in California. *Complexity*, *2020*. https://doi.org/10.1155/2020/8049504

Chang, J. H. W., Griffith, S. M., & Lin, N. H. (2022). Impacts of land-surface forcing on local meteorology and ozone concentrations in a heavily industrialized coastal urban area. *Urban Climate*, *45*(August), 101257. https://doi.org/10.1016/j.uclim.2022.101257

Chen, F., Kusaka, H., Tewari, M., Bao, J. W., & Hirakuchi, H. (2004). Utilizing the Coupled WRF/LSM/Urban Modeling System with Detailed Urban Classification to Simulate the Urban Heat Island Phenomena over the Greater Houston Area. *Fifth Conference on Urban Environment*, 9.11.

Chen, F., Wang, L., & Deng, H. (2023). A Novel Combined Model for Air Quality Index Forecasting in Changchun. *Atmosphere 2023, Vol. 14, Page 1475*, *14*(10), 1475. https://doi.org/10.3390/ATMOS14101475

Chen, X., Yang, T., Wang, Z., Wang, F., & Wang, H. (2023). An ensemble method for improving the estimation of planetary boundary layer height from radiosonde data. *Atmos. Meas. Tech*, *16*, 4289–4302. https://doi.org/10.5194/amt-16-4289-2023

Chen, Z. Y., Petetin, H., Méndez Turrubiates, R. F., Achebak, H., Pérez García-Pando, C., & Ballester, J. (2024). Population exposure to multiple air pollutants and its compound episodes in Europe. *Nature Communications*, *15*(1), 1–11. https://doi.org/10.1038/S41467-024-46103-3;TECHMETA

Cheng, F. Y., Feng, C. Y., Yang, Z. M., Hsu, C. H., Chan, K. W., Lee, C. Y., & Chang, S. C. (2021). Evaluation of real-time PM2.5 forecasts with the WRF-CMAQ modeling system and weather-pattern-dependent bias-adjusted PM2.5 forecasts in Taiwan. *Atmospheric Environment*, *244*, 117909. https://doi.org/10.1016/J.ATMOSENV.2020.117909

Coggon, M. M., Stockwell, C. E., Xu, L., Peischl, J., Gilman, J. B., Lamplugh, A., Bowman, H. J., Aikin, K., Harkins, C., Zhu, Q., Schwantes, R. H., He, J., Li, M., Seltzer, K., Mcdonald, B., & Warneke, C. (2024). Contribution of cooking emissions to the urban volatile organic compounds in Las Vegas, NV. *Atmospheric Chemistry and Physics*, *24*(7), 4289–4304. https://doi.org/10.5194/ACP-24-4289-2024

Crippa, M., Janssens-Maenhout, G., Guizzardi, D., Van Dingenen, R., & Dentener, F. (2019). Contribution and uncertainty of sectorial and regional emissions to regional and global PM2.5 health impacts. *Atmospheric Chemistry and Physics*, *19*(7), 5165–5186. https://doi.org/10.5194/ACP-19-5165-2019

Crippa, M., Solazzo, E., Huang, G., Guizzardi, D., Koffi, E., Muntean, M., Schieberle, C., Friedrich, R., & Janssens-Maenhout, G. (2020). High resolution temporal profiles in the Emissions Database for Global Atmospheric Research. *Scientific Data*, *7*(1), 1–17. https://doi.org/10.1038/S41597-020-0462-2;SUBJMETA

de la Paz, D., de Andrés, J. M., Narros, A., Silibello, C., Finardi, S., Fares, S., Tejero, L., Borge, R., & Mircea, M. (2022). Assessment of Air Quality and Meteorological Changes Induced by Future Vegetation in Madrid. *Forests*, *13*(5). https://doi.org/10.3390/f13050690

de Oliveira, H. E., Duarte, L. T., & Romano, J. M. T. (2022). Identification of the Choquet integral parameters in the interaction index domain by means of sparse modeling. *Expert Systems with Applications*, *187*, 115874. https://doi.org/10.1016/J.ESWA.2021.115874

Díaz-Isaac, L. I., Lauvaux, T., Bocquet, M., & Davis, K. J. (2019). Calibration of a multi-physics ensemble for estimating the uncertainty of a greenhouse gas atmospheric transport model. *Atmospheric Chemistry and Physics*, *19*(8), 5695–5718. https://doi.org/10.5194/ACP-19-5695-2019

Do, T. N. N., Ngo, X. T., Pham, V. H., Vuong, N. L., Le, H. A., Pham, C. T., Bui, Q. H., & Nguyen, T. N. T. (2021). Application of WRF-Chem to simulate air quality over Northern Vietnam. *Environmental Science and Pollution Research*, *28*(10), 12067–12081. https://doi.org/10.1007/s11356-020-08913-y

Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. https://arxiv.org/pdf/1702.08608

Drewil, G. I., & Al-Bahadili, R. J. (2022). Air pollution prediction using LSTM deep learning and metaheuristics algorithms. *Measurement: Sensors*, *24*, 100546. https://doi.org/10.1016/J.MEASEN.2022.100546

Dudhia, J. (1989). *Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model Dudhia_1989_JAtmosSci.pdf*. https://doi.org/https://doi.org/10.1175/1520-0469(1989)046%3C3077:NSOCOD%3E2.0.CO;2

El-Samra, R., Bou-Zeid, E., & El-Fadel, M. (2018). What model resolution is required in climatological downscaling over complex terrain? *Atmospheric Research*, *203*(May 2017), 68–82. https://doi.org/10.1016/j.atmosres.2017.11.030

*EMEP/EEA air pollutant emission inventory guidebook 2023 | European Environment Agency's home page*. (2023). Retrieved September 22, 2025, from https://www.eea.europa.eu/en/analysis/publications/emep-eea-guidebook-2023?utm_source=chatgpt.com

*Emissions Modeling Platform Collaborative: 2016beta Onroad Mobile Sources SPECIFICATION SHEET: ONROAD 2016beta Platform*. (2019). http://www.epa.gov/otaq/models/moves/index.htm

Eure, K. C., Mykolajtchuk, P. D., Zhang, Y., Stensrud, D. J., Zhang, F., Greybush, S. J., & Kumjian, M. R. (2023). Simultaneous Assimilation of Planetary Boundary Layer Observations from Radar and All-Sky Satellite Observations to Improve Forecasts of Convection Initiation. *Monthly Weather Review*, *151*(3), 795–813. https://doi.org/10.1175/MWR-D-22-0188.1

Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A., & Edson, J. B. (2003). Bulk parameterization of air-sea fluxes: Updates and verification for the COARE algorithm.

*Journal of Climate*, *16*(4), 571–591. https://doi.org/10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2

Fan, K., Dhammapala, R., Harrington, K., Lamastro, R., Lamb, B., & Lee, Y. (2022). Development of a Machine Learning Approach for Local-Scale Ozone Forecasting: Application to Kennewick, WA. *Frontiers in Big Data*, *5*, 781309. https://doi.org/10.3389/FDATA.2022.781309/BIBTEX

Fattal, E., David-Saroussi, H., Buchman, O., Tas, E., & Klausner, Z. (2023). Heterogenous Canopy in a Lagrangian-Stochastic Dispersion Model for Particulate Matter from Multiple Sources over the Haifa Bay Area. *Atmosphere*, *14*(1), 144. https://doi.org/10.3390/ATMOS14010144

Fécan, F., Marticorena, B., & Bergametti, G. (1998). Parametrization of the increase of the aeolian erosion threshold wind friction velocity due to soil moisture for arid and semi-arid areas. *Annales Geophysicae*, *17*(1), 149–157. https://doi.org/10.1007/s00585-999-0149-7

Feng, S., Tang, L., Huang, M., & Wu, Y. (2025). Integrating D–S evidence theory and multiple deep learning frameworks for time series prediction of air quality. *Scientific Reports*, *15*(1), 1–20. https://doi.org/10.1038/S41598-025-87935-3;SUBJMETA=117,172,639,704,705;KWRD=COMPUTER+SCIENCE,ENVIRONMENTAL+SCIENCES

Ferenczi, Z., Imre, K., Lakatos, M., Molnár, Á., Bozó, L., Homolya, E., & Gelencsér, A. (2021). Long-term Characterization of Urban PM10 in Hungary. *Aerosol and Air Quality Research*, *21*(10), 210048. https://doi.org/10.4209/AAQR.210048

Fernandes, A. P., Rafael, S., Lopes, D., Coelho, S., Borrego, C., & Lopes, M. (2021). The air pollution modelling system URBAIR: how to use a Gaussian model to accomplish high spatial and temporal resolutions. *Air Quality, Atmosphere and Health*, *14*(12), 1969–1988. https://doi.org/10.1007/s11869-021-01069-9

Fiore, A. M., Naik, V., & Leibensperger, E. M. (2015). Air Quality and Climate Connections. *Journal of the Air & Waste Management Association*, *65*(6), 645–685. https://doi.org/10.1080/10962247.2015.1040526

Folberth, G. A., Hauglustaine, D. A., Lathièrel, J., & Brocheton, F. (2006). Interactive chemistry in the Laboratoire de Météorologie Dynamique general circulation model: Model description and impact analysis of biogenic hydrocarbons on tropospheric chemistry. *Atmospheric Chemistry and Physics*, *6*(8), 2273–2319. https://doi.org/10.5194/acp-6-2273-2006

Fountoukis, C., Mohieldeen, Y., Pomares, L., Gladich, I., Siddique, A., Skillern, A., & Ayoub, M. A. (2022). Assessment of High-resolution Local Emissions and Land-use in Air Quality Forecasting at an Urban, Coastal, Desert Environment. *Aerosol and Air Quality Research*, *22*(6). https://doi.org/10.4209/aaqr.220001

Freeman, B. S., Taylor, G., Gharabaghi, B., & Thé, J. (2018). Forecasting air quality time series using deep learning. *Journal of the Air & Waste Management Association*, *68*(8), 866–886. https://doi.org/10.1080/10962247.2018.1459956

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Https://Doi.Org/10.1214/Aos/1013203451*, *29*(5), 1189–1232. https://doi.org/10.1214/AOS/1013203451

Galmarini, S., Kioutsioukis, I., Solazzo, E., Alyuz, U., Balzarini, A., Bellasio, R., Benedictow, A. M. K., Bianconi, R., Bieser, J., Brandt, J., Christensen, J. H., Colette, A., Curci, G., Davila, Y., Dong, X., Flemming, J., Francis, X., Fraser, A., Fu, J., … Unal, A. (2018). Two-scale multi-model ensemble: Is a hybrid ensemble of opportunity telling us more? *Atmospheric Chemistry and Physics*, *18*(12), 8727–8744. https://doi.org/10.5194/ACP-18-8727-2018

Galmarini, S., Makar, P., Clifton, O. E., Hogrefe, C., Bash, J. O., Bellasio, R., Bianconi, R., Bieser, J., Butler, T., Ducker, J., Flemming, J., Hodzic, A., Holmes, C. D., Kioutsioukis, I., Kranenburg, R., Lupascu, A., Perez-Camanyo, J. L., Pleim, J., Ryu, Y. H., … Wolke, R. (2021). Technical note: AQMEII4 Activity 1: Evaluation of wet and dry deposition schemes as an integral part of regional-scale air quality models. *Atmospheric Chemistry and Physics*, *21*(20), 15663–15697. https://doi.org/10.5194/ACP-21-15663-2021

García-Herrera, R., Garrido-Perez, J. M., & Ordóñez, C. (2022). Modulation of European air quality by Euro-Atlantic weather regimes. *Atmospheric Research*, *277*, 106292. https://doi.org/10.1016/J.ATMOSRES.2022.106292

Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, *32*(14–15), 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0

Georgiou, G. K., Christoudias, T., Proestos, Y., Kushta, J., & Hadjinicolaou, P. (2018). *Air quality modelling in the summer over the eastern Mediterranean using WRF-Chem: chemistry and aerosol mechanism intercomparison*. 1555–1571. https://doi.org/https://doi.org/10.5194/acp-18-1555-2018

Georgiou, G. K., Christoudias, T., Proestos, Y., Kushta, J., Pikridas, M., Sciare, J., Savvides, C., & Lelieveld, J. (2022). Evaluation of WRF-Chem model (v3.9.1.1) real-Time air quality forecasts over the Eastern Mediterranean. *Geoscientific Model Development*, *15*(10), 4129–4146. https://doi.org/10.5194/GMD-15-4129-2022

Ghaderpour, E., & Vujadinovic, T. (2020). The Potential of the Least-Squares Spectral and Cross-Wavelet Analyses for Near-Real-Time Disturbance Detection within Unequally Spaced Satellite Image Time Series. *Remote Sensing 2020, Vol. 12, Page 2446*, *12*(15), 2446. https://doi.org/10.3390/RS12152446

Ginoux, P., Chin, M., Tegen, I., Goddard, T., & In-, G. (2001). model Brent Holben •’ Oleg Dubovik •,•’ and Shian-Jiann Lin e CART ) model . In this model all topographic from the

Goddard Earth Observing System Data Assimilation System ( GEOS emission is estimated to be between. *2-Journal of Geophysical Research*, *106*, 20255–20273.

Gladson, L., Garcia, N. ; ;, Bi, J. ;, Liu, Y. ;, Lee, H. J. ;, Cromar, K., Garcia, N., Bi, J., Liu, Y., Lee, H. J., & Cromar, K. (2022). Evaluating the Utility of High-Resolution Spatiotemporal Air Pollution Data in Estimating Local PM2.5 Exposures in California from 2015–2018. *Atmosphere 2022, Vol. 13, Page 85*, *13*(1), 85. https://doi.org/10.3390/ATMOS13010085

Gocheva-Ilieva, S.G., Ivanov, A.V., Voynikova, D.S. *et al.* Time series analysis and forecasting for air pollution in small urban area: an SARIMA and factor analysis approach. *Stoch Environ Res Risk Assess* **28**, 1045–1060 (2014). https://doi.org/10.1007/s00477-013-0800-4

Grabisch, M. (2011). Fuzzy Measures and Integrals in Multicriteria Decision Analysis. *Wiley Encyclopedia of Operations Research and Management Science*, 1–8. https://doi.org/10.1002/9780470400531.EORMS1087

Grigoratos, T., & Martini, G. (2014). Brake wear particle emissions: a review. *Environmental Science and Pollution Research International*, *22*(4), 2491. https://doi.org/10.1007/S11356-014-3696-8

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., & Wang, X. (2012). The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions. *Geoscientific Model Development*, *5*(6), 1471–1492. https://doi.org/10.5194/gmd-5-1471-2012

Guevara, M., Jorba, O., Tena, C., Denier Van Der Gon, H., Kuenen, J., Elguindi, N., Darras, S., Granier, C., & Pérez García-Pando, C. (2021). Copernicus Atmosphere Monitoring Service TEMPOral profiles (CAMS-TEMPO): Global and European emission temporal profile maps for atmospheric chemistry modelling. *Earth System Science Data*, *13*(2), 367–404. https://doi.org/10.5194/ESSD-13-367-2021

Guion, A., Turquety, S., Cholakian, A., Polcher, J., Ehret, A., & Lathière, J. (2023). Biogenic isoprene emissions, dry deposition velocity, and surface ozone concentration during summer droughts, heatwaves, and normal conditions in southwestern Europe. *Atmospheric Chemistry and Physics*, *23*(2), 1043–1071. https://doi.org/10.5194/acp-23-1043-2023

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN Model-Based Approach in Classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2888*, 986–996. https://doi.org/10.1007/978-3-540-39964-3_62

Hagler, G., Birkett, D., Henry, R. C., & Peltier, R. E. (2021). Three years of high time-resolution air pollution monitoring in the complex multi-source harbor of New York and New Jersey. *Aerosol and Air Quality Research*, *21*(3), 1–15. https://doi.org/10.4209/aaqr.2020.02.0069

Hegarty, J. D., Lewis, J., McGrath-Spangler, E. L., Henderson, J., Scarino, A. J., DeCola, P., Ferrare, R., Hicks, M., Adams-Selin, R. D., & Welton, E. J. (2018). Analysis of the Planetary Boundary Layer Height during DISCOVER-AQ Baltimore–Washington, D.C., with Lidar and High-Resolution WRF Modeling. *Journal of Applied Meteorology and Climatology*, *57*(11), 2679–2696. https://doi.org/10.1175/JAMC-D-18-0014.1

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/NECO.1997.9.8.1735

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, *15*(14), 5481–5487. https://doi.org/10.5194/GMD-15-5481-2022

Hong, S. Y., Dudhia, J., & Chen, S. H. (2004). A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Monthly Weather Review*, *132*(1), 103–120. https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2

Houdou, A., Badisy, I. El, Khomsi, K., Abdala, S. A., Abdulla, F., Najmi, H., Obtel, M., Belyamani, L., Ibrahimi, A., & Khalis, M. (2024). Interpretable Machine Learning Approaches for Forecasting and Predicting Air Pollution: A Systematic Review. *Aerosol and Air Quality Research*, *24*(1), 230151. https://doi.org/10.4209/AAQR.230151

Hu, J., Li, X., Huang, L., Ying, Q., Zhang, Q., Zhao, B., Wang, S., & Zhang, H. (2017). Ensemble prediction of air quality using the WRF/CMAQ model system for health effect studies in China. *Atmospheric Chemistry and Physics*, *17*(21), 13103–13118. https://doi.org/10.5194/ACP-17-13103-2017

Huang, C., Sun, K., Hu, J., Xue, T., Xu, H., & Wang, M. (2022). Estimating 2013–2019 NO2 exposure with high spatiotemporal resolution in China using an ensemble model. *Environmental Pollution*, *292*. https://doi.org/10.1016/J.ENVPOL.2021.118285

Huang, Y., Yu, J., Dai, X., Huang, Z., & Li, Y. (2022). Air-Quality Prediction Based on the EMD–IPSO–LSTM Combination Model. *Sustainability 2022, Vol. 14, Page 4889*, *14*(9), 4889. https://doi.org/10.3390/SU14094889

ImUlas. (2013). Impact of sea-salt emissions on the model performance and aerosol chemical composition and deposition in the East Mediterranean coastal regions. *Atmospheric Environment*, *75*, 329–340. https://doi.org/10.1016/j.atmosenv.2013.04.034

János, Á., & Hajnalka, V. (2020). Sensitivity of simulated temperature , precipitation , and global radiation to different WRF configurations over the Carpathian Basin for regional climate applications. *Climate Dynamics*, *55*(9), 2849–2866. https://doi.org/10.1007/s00382-020-05416-x

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Bergamaschi, P., Pagliari, V., Olivier, J. G. J., Peters, J. A. H. W., Van Aardenne, J. A., Monni, S., Doering, U., Roxana Petrescu, A. M., Solazzo, E., & Oreggioni, G. D. (2019).

EDGAR v4.3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970-2012. *Earth System Science Data*, *11*(3), 959–1002. https://doi.org/10.5194/ESSD-11-959-2019

Jiang, L., Bessagnet, B., Meleux, F., Tognet, F., & Couvidat, F. (2020). Impact of physics parameterizations on high-resolution air quality simulations over the Paris region. *Atmosphere*, *11*(6). https://doi.org/10.3390/ATMOS11060618

Jiang, Q., Bei, N., Wu, J., Li, X., Wang, R., Yu, J., Lu, Y., Tie, X., & Li, G. (2023). Impacts of urban expansion on meteorology and air quality in North China Plain during wintertime: A case study. *Urban Climate*, *52*(September), 101696. https://doi.org/10.1016/j.uclim.2023.101696

Jiang, X., & Yoo, E. (2018). The importance of spatial resolutions of Community Multiscale Air Quality ( CMAQ ) models on health impact assessment. *Science of the Total Environment*, *627*, 1528–1543. https://doi.org/10.1016/j.scitotenv.2018.01.228

Jiménez, P. A., & Dudhia, J. (2012). Improving the representation of resolved and unresolved topographic effects on surface wind in the wrf model. *Journal of Applied Meteorology and Climatology*, *51*(2), 300–316. https://doi.org/10.1175/JAMC-D-11-084.1

Jin, Q., Yan, K., Xu, Y., Lin, F., Rollo, F., Bachechi, C., & Po, L. (2023). Anomaly Detection and Repairing for Improving Air Quality Monitoring. *Sensors 2023, Vol. 23, Page 640*, *23*(2), 640. https://doi.org/10.3390/S23020640

Johnson, J. B. (2022). An Introduction to Atmospheric Pollutant Dispersion Modelling. *Environmental Sciences Proceedings 2022, Vol. 19, Page 18*, *19*(1), 18. https://doi.org/10.3390/ECAS2022-12826

Kalantari, E., Gholami, H., Malakooti, H., Eftekhari, M., Saneei, P., Esfandiarpour, D., Moosavi, V., & Nafarzadegan, A. R. (2024). Evaluating traditional versus ensemble machine learning methods for predicting missing data of daily PM10 concentration. *Atmospheric Pollution Research*, *15*(5), 102063. https://doi.org/10.1016/J.APR.2024.102063

Kalmus, P., Ao, C. O., Wang, K. N., Manzi, M. P., & Teixeira, J. (2022). A high-resolution planetary boundary layer height seasonal climatology from GNSS radio occultations. *Remote Sensing of Environment*, *276*(September 2021). https://doi.org/10.1016/j.rse.2022.113037

Karamchandani, P., Vijayaraghavan, K., & Yarwood, G. (2011). Sub-Grid Scale Plume Modeling. *Atmosphere*, *2*, 389–406. https://doi.org/10.3390/atmos2030389

Ke, H., Gong, S., He, J., Zhang, L., & Mo, J. (2022). A hybrid XGBoost-SMOTE model for optimization of operational air quality numerical model forecasts. *Frontiers in Environmental Science*, *10*, 1007530. https://doi.org/10.3389/FENVS.2022.1007530/BIBTEX

Keawboonchu, J., Thepanondh, S., Kultan, V., Pinthong, N., Malakan, W., & Robson, M. G. (2023). Integrated Sustainable Management of Petrochemical Industrial Air Pollution. *International Journal of Environmental Research and Public Health*, *20*(3), 2280. https://doi.org/10.3390/IJERPH20032280/S1

Kebłowski, P., & Welfe, A. (2004). The ADF–KPSS test of the joint confirmation hypothesis of unit autoregressive root. *Economics Letters*, *85*(2), 257–263. https://doi.org/10.1016/J.ECONLET.2004.04.013

Kim, S., Alizamir, M., Kim, N. W., & Kisi, O. (2020a). Bayesian Model Averaging: A Unique Model Enhancing Forecasting Accuracy for Daily Streamflow Based on Different Antecedent Time Series. *Sustainability 2020, Vol. 12, Page 9720*, *12*(22), 9720. https://doi.org/10.3390/SU12229720

Kim, S., Alizamir, M., Kim, N. W., & Kisi, O. (2020b). Bayesian Model Averaging: A Unique Model Enhancing Forecasting Accuracy for Daily Streamflow Based on Different Antecedent Time Series. *Sustainability 2020, Vol. 12, Page 9720*, *12*(22), 9720. https://doi.org/10.3390/SU12229720

Kitagawa, Y. K. L., Pedruzzi, R., Galvão, E. S., de Araújo, I. B., Alburquerque, T. T. de A., Kumar, P., Nascimento, E. G. S., & Moreira, D. M. (2021). Source apportionment modelling of PM2.5 using CMAQ-ISAM over a tropical coastal-urban area. *Atmospheric Pollution Research*, *12*(12). https://doi.org/10.1016/J.APR.2021.101250

Klement, E. P., Mesiar, R., & Pap, E. (2004). Measure-based aggregation operators. *Fuzzy Sets and Systems*, *142*(1), 3–14. https://doi.org/10.1016/J.FSS.2003.10.028

Kong, L., Tang, X., Zhu, J., Wang, Z., Sun, Y., Fu, P., Gao, M., Wu, H., Lu, M., Wu, Q., Huang, S., Sui, W., Li, J., Pan, X., Wu, L., Akimoto, H., & Carmichael, G. R. (2023). Unbalanced emission reductions of different species and sectors in China during COVID-19 lockdown derived by multi-species surface observation assimilation. *Atmospheric Chemistry and Physics*, *23*(11), 6217–6240. https://doi.org/10.5194/ACP-23-6217-2023

Kukkonen, J., Olsson, T., Schultz, D. M., Baklanov, A., Klein, T., Miranda, A. I., Monteiro, A., Hirtl, M., Tarvainen, V., Boy, M., Peuch, V.-H., Poupkou, A., Kioutsioukis, I., Finardi, S., Sofiev, M., Sokhi, R., Lehtinen, K. E. J., Karatzas, K., San José, R., Astitha, M., Kallos, G., Schaap, M., Reimer, E., Jakobs, H., and Eben, K.( 2012). A review of operational, regional-scale, chemical weather forecasting models in Europe, *Atmos. Chem. Phys., 12, 1–87,* https://doi.org/10.5194/acp-12-1-2012

Lee, H. J., Jo, H. Y., Kim, J. M., Bak, J., Park, M. S., Kim, J. K., Jo, Y. J., & Kim, C. H. (2023). Nocturnal Boundary Layer Height Uncertainty in Particulate Matter Simulations during the KORUS-AQ Campaign. *Remote Sensing*, *15*(2), 300. https://doi.org/10.3390/RS15020300/S1

Lei, T. M. T., Siu, S. W. I., Monjardino, J., Mendes, L., & Ferreira, F. (2022). Using Machine Learning Methods to Forecast Air Quality: A Case Study in Macao. *Atmosphere 2022, Vol. 13, Page 1412*, *13*(9), 1412. https://doi.org/10.3390/ATMOS13091412

Leo Hohenberger, T., Che, W., Sun, Y., Fung, J. C. H., & Lau, A. K. H. (2022). Assessment of the impact of sensor error on the representativeness of population exposure to urban air pollutants. *Environment International*, *165*, 107329. https://doi.org/10.1016/J.ENVINT.2022.107329

Li, B., Hu, Q., Gao, M., Liu, T., Zhang, C., & Liu, C. (2023). Physical informed neural network improving the WRF-CHEM results of air pollution using satellite-based remote sensing data. *Atmospheric Environment*, *311*, 120031. https://doi.org/10.1016/J.ATMOSENV.2023.120031

Li, H., Tatarko, J., Kucharski, M., & Dong, Z. (2015). *PM2.5 and PM10 emissions from agricultural soils by wind erosion q*. https://doi.org/10.1016/j.aeolia.2015.02.003

Li, J., Yu, S., Chen, X., Zhang, Y., Li, M., Li, Z., Song, Z., Liu, W., Li, P., Xie, M., & Xing, J. (2022). Evaluation of the WRF-CMAQ Model Performances on Air Quality in China with the Impacts of the Observation Nudging on Meteorology. *Aerosol and Air Quality Research*, *22*(4), 220023. https://doi.org/10.4209/aaqr.220023

Li, S., & Xing, J. (2025). Enhancing 72-Hour air quality forecasting with an observation-driven deep learning chemistry transport model. *Environment International*, *202*, 109689. https://doi.org/10.1016/J.ENVINT.2025.109689

Li, S., Xing, J., Li, S., & Xing, J. (2024). DeepSAT4D: Deep learning empowers four-dimensional atmospheric chemical concentration and emission retrieval from satellite. *The Innovation Geoscience*, *2*(1), 100061–1. https://doi.org/10.59717/J.XINN-GEO.2024.100061

Liang, Y., Xia, Y., Ke, S., Wang, Y., Wen, Q., Zhang, J., Zheng, Y., & Zimmermann, R. (2023). AirFormer: Predicting Nationwide Air Quality in China with Transformers. *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(12), 14329–14337. https://doi.org/10.1609/AAAI.V37I12.26676

Liao, Q., Zhu, M., Wu, L., Pan, X., Tang, X., & Wang, Z. (2020). Deep Learning for Air Quality Forecasts: a Review. *Current Pollution Reports*, *6*(4), 399–409. https://doi.org/10.1007/S40726-020-00159-Z/METRICS

Lin, Y., & Li, X. (2024). Back to the Metrics: Exploration of Distance Metrics in Anomaly Detection. *Applied Sciences 2024, Vol. 14, Page 7016*, *14*(16), 7016. https://doi.org/10.3390/APP14167016

Liu, J., & Xing, J. (2023). Identifying Contributors to PM2.5 Simulation Biases of Chemical Transport Model Using Fully Connected Neural Networks. *Journal of Advances in Modeling Earth Systems*, *15*(2), e2021MS002898. https://doi.org/10.1029/2021MS002898

Liu, T., Lau, A. K. H., Sandbrink, K., & Fung, J. C. H. (2018). Time Series Forecasting of Air Quality Based On Regional Numerical Modeling in Hong Kong. *Journal of Geophysical Research: Atmospheres*, *123*(8), 4175–4196. https://doi.org/10.1002/2017JD028052

Liu, Z., Doherty, R. M., Wild, O., O'connor, F. M., & Turnock, S. T. (2022). Correcting ozone biases in a global chemistry-climate model: implications for future ozone. *Atmospheric Chemistry and Physics*, *22*(18), 12543–12557. https://doi.org/10.5194/ACP-22-12543-2022

Lu, H., Xie, M., Liu, X., Liu, B., Jiang, M., Gao, Y., & Zhao, X. (2021). Adjusting prediction of ozone concentration based on CMAQ model and machine learning methods in Sichuan-Chongqing region, China. *Atmospheric Pollution Research*, *12*(6), 101066. https://doi.org/10.1016/j.apr.2021.101066

Lu, W. Z., & Wang, W. J. (2005). Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends. *Chemosphere*, *59*(5), 693–701. https://doi.org/10.1016/j.chemosphere.2004.10.032

Lyu, B., Huang, R., Wang, X., Wang, W., & Hu, Y. (2024). *FastCTM (v1.0): Atmospheric chemical transport modelling with a principle-informed neural network for air quality simulations*. https://doi.org/10.5194/GMD-2024-198

Mailler, S., Menut, L., Khvorostyanov, D., Valari, M., Couvidat, F., Siour, G., Turquety, S., Briant, R., Tuccella, P., Bessagnet, B., Colette, A., Létinois, L., Markakis, K., & Meleux, F. (2017). CHIMERE-2017: From urban to hemispheric chemistry-transport modeling. *Geoscientific Model Development*, *10*(6), 2397–2423. https://doi.org/10.5194/gmd-10-2397-2017

Maina, F. Z., Siirila-Woodburn, E. R., & Vahmani, P. (2020). Sensitivity of meteorological-forcing resolution on hydrologic variables. *Hydrology and Earth System Sciences*, *24*(7), 3451–3474. https://doi.org/10.5194/hess-24-3451-2020

Maison, A., Flageul, C., Carissimo, B., Tuzet, A., & Sartelet, K. (2022). Parametrization of Horizontal and Vertical Transfers for the Street-Network Model MUNICH Using the CFD Model Code_Saturne. *Atmosphere 2022, Vol. 13, Page 527*, *13*(4), 527. https://doi.org/10.3390/ATMOS13040527

Maleki, H., Sorooshian, A., Goudarzi, G., Baboli, Z., Tahmasebi Birgani, Y., & Rahmati, M. (2019). Air pollution prediction by using an artificial neural network model. *Clean Technologies and Environmental Policy*, *21*(6), 1341–1352. https://doi.org/10.1007/s10098-019-01709-w

Mampitiya, L., Rathnayake, N., Hoshino, Y., & Rathnayake, U. (2024). Performance of machine learning models to forecast PM10 levels. *MethodsX*, *12*, 102557. https://doi.org/10.1016/J.MEX.2024.102557

Marinov, E., Petrova-Antonova, D., & Malinov, S. (2022). Time Series Forecasting of Air Quality: A Case Study of Sofia City. *Atmosphere 2022, Vol. 13, Page 788*, *13*(5), 788. https://doi.org/10.3390/ATMOS13050788

Masood, A., Hameed, M. M., Srivastava, A., Pham, Q. B., Ahmad, K., Razali, S. F. M., & Baowidan, S. A. (2023). Improving PM2.5 prediction in New Delhi using a hybrid extreme

learning machine coupled with snake optimization algorithm. *Scientific Reports*, *13*(1), 1–17. https://doi.org/10.1038/S41598-023-47492-Z;SUBJMETA

Maxwell, A. E., Strager, M. P., Warner, T. A., Ramezan, C. A., Morgan, A. N., & Pauley, C. E. (2019). Large-area, high spatial resolution land cover mapping using random forests, GEOBIA, and NAIP orthophotography: Findings and recommendations. *Remote Sensing*, *11*(12). https://doi.org/10.3390/rs11121409

Mazzeo, A., Burrow, M., Quinn, A., Marais, E. A., Singh, A., Ng, D., Gatari, M. J., & Pope, F. D. (2022). *Evaluation of the WRF and CHIMERE models for the simulation of PM 2 . 5 in large East African urban conurbations*. 10677–10701.

Méndez, M., Merayo, M. G., & Núñez, M. (2023). Machine learning algorithms to forecast air quality: a survey. *Artificial Intelligence Review*, *56*(9), 1. https://doi.org/10.1007/S10462-023-10424-4

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

Meng, X., Wang, W., Shi, S., Zhu, S., Wang, P., Chen, R., Xiao, Q., Xue, T., Geng, G., Zhang, Q., Kan, H., & Zhang, H. (2022). Evaluating the spatiotemporal ozone characteristics with high-resolution predictions in mainland China, 2013–2019. *Environmental Pollution*, *299*, 118865. https://doi.org/10.1016/j.envpol.2022.118865

Menut, L., Bessagnet, B., Briant, R., Cholakian, A., Couvidat, F., Mailler, S., Pennel, R., Siour, G., Tuccella, P., Turquety, S., & Valari, M. (2021). The CHIMERE v2020r1 online chemistry-transport model. *Geoscientific Model Development*, *14*(11), 6781–6811. https://doi.org/10.5194/GMD-14-6781-2021

Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., & Vivanco, M. G. (2013). CHIMERE 2013: a model for regional atmospheric composition modelling. *Geoscientific Model Development*, *6*(4), 981–1028. https://doi.org/10.5194/GMD-6-981-2013

Menut, L., Bessagnet, B., Siour, G., Mailler, S., Pennel, R., & Cholakian, A. (2020). Impact of lockdown measures to combat Covid-19 on air quality over western Europe. *Science of the Total Environment*, *741*, 140426. https://doi.org/10.1016/j.scitotenv.2020.140426

Menut, L., Goussebaile, A., Bessagnet, B., Khvorostiyanov, D., & Ung, A. (2012). Impact of realistic hourly emissions profiles on air pollutants concentrations modelled with CHIMERE. *Atmospheric Environment*, *49*, 233–244. https://doi.org/10.1016/J.ATMOSENV.2011.11.057

Milinski, S., Maher, N., & Olonscheck, D. (2020). How large does a large ensemble need to be? *Earth System Dynamics*, *11*(4), 885–901. https://doi.org/10.5194/ESD-11-885-2020

Miller, M. J., Viterbo, P. A., & Surface, L. (1996). *The land surface-atmosphere interaction' A review based on observational and global modeling perspective*. *101*, 7209–7225.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, *102*(D14), 16663–16682. https://doi.org/10.1029/97jd00237

Mo, Y., Li, Q., Karimian, H., Zhang, S., Kong, X., Fang, S., & Tang, B. (2021). Daily spatiotemporal prediction of surface ozone at the national level in China: An improvement of CAMS ozone product. *Atmospheric Pollution Research*, *12*(1), 391–402. https://doi.org/10.1016/j.apr.2020.09.020

Moayedi, H., Mosallanezhad, M., Rashid, A. S. A., Jusoh, W. A. W., & Muazu, M. A. (2020). A systematic review and meta-analysis of artificial neural network application in geotechnical engineering: theory and applications. *Neural Computing and Applications*, *32*(2), 495–518. https://doi.org/10.1007/s00521-019-04109-9

Mok, K. M., Miranda, A. I., Yuen, K. V., Hoi, K. I., Monteiro, A., & Ribeiro, I. (2017). Selection of bias correction models for improving the daily PM10 forecasts of WRF-EURAD in Porto, Portugal. *Atmospheric Pollution Research*, *8*(4), 628–639. https://doi.org/10.1016/j.apr.2016.12.010

Morf, M., Vieira, A., & Kailath, T. (1978). Covariance Characterization by Partial Autocorrelation Matrices. *Https://Doi.Org/10.1214/Aos/1176344208*, *6*(3), 643–648. https://doi.org/10.1214/AOS/1176344208

Mukherjee, A., & Bhattacharyya, D. (2024). On the development of steady-state and dynamic mass-constrained neural networks using noisy transient data. *Computers & Chemical Engineering*, *187*, 108722. https://doi.org/10.1016/J.COMPCHEMENG.2024.108722

Mundim, K. C., Baraldi, S., Machado, H. G., & Vieira, F. M. C. (2020). Temperature coefficient (Q10) and its applications in biological systems: Beyond the Arrhenius theory. *Ecological Modelling*, *431*, 109127. https://doi.org/10.1016/J.ECOLMODEL.2020.109127

Murillo-Escobar, J., Sepulveda-Suescun, J. P., Correa, M. A., & Orrego-Metaute, D. (2019). Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: Case study in Aburrá Valley, Colombia. *Urban Climate*, *29*, 100473. https://doi.org/10.1016/J.UCLIM.2019.100473

Neckel, A., Oliveira, M. L. S., Stolfo Maculan, L., Bodah, B. W., Gonçalves, C., & Silva, L. F. O. (2023). Air pollution in central European capital (Budapest) via self-made passive samplers and Sentinel-3B SYN satellite images. *Urban Climate*, *47*, 101384. https://doi.org/10.1016/j.uclim.2022.101384

Ocampo-Duque, W., Ferré-Huguet, N., Domingo, J. L., & Schuhmacher, M. (2006). Assessing water quality in rivers with fuzzy inference systems: A case study. *Environment International*, *32*(6), 733–742. https://doi.org/10.1016/j.envint.2006.03.009

Osowski, S., & Garanty, K. (2007). Forecasting of the daily meteorological pollution using wavelets and support vector machine. *Engineering Applications of Artificial Intelligence*, *20*(6), 745–755. https://doi.org/10.1016/J.ENGAPPAI.2006.10.008

Ovchinnikov, M., FAST, J. D., BERG, L. K., GUSTAFSON, W. I., Chen, J., Sakaguchi, K., & Xiao, H. (2022). Effects of Horizontal Resolution, Domain Size, Boundary Conditions, and Surface Heterogeneity on Coarse LES of a Convective Boundary Layer. *Monthly Weather Review*, *150*(6), 1397–1415. https://doi.org/10.1175/MWR-D-21-0244.1

Özüpak, Y., Alpsalaz, F., & Aslan, E. (2025). Air Quality Forecasting Using Machine Learning: Comparative Analysis and Ensemble Strategies for Enhanced Prediction. *Water, Air, and Soil Pollution*, *236*(7), 1–17. https://doi.org/10.1007/S11270-025-08122-8/TABLES/5

Özüpak, Y., Feyyaz Alpsalaz, & Aslan, E. (2025). *Air Quality Forecasting Using Machine Learning: Comparative Analysis and Ensemble Strategies for Enhanced Prediction*. https://doi.org/10.1007/s11270-025-08122-8

Peláez-Rodríguez, C., Pérez-Aracil, J., Prieto-Godino, L., Ghimire, S., Deo, R. C., & Salcedo-Sanz, S. (2023). A fuzzy-based cascade ensemble model for improving extreme wind speeds prediction. *Journal of Wind Engineering and Industrial Aerodynamics*, *240*, 105507. https://doi.org/10.1016/J.JWEIA.2023.105507

Potier, E., Waked, A., Bourin, A., Minvielle, F., Péré, J. C., Perdrix, E., Michoud, V., Riffault, V., Alleman, L. Y., & Sauvage, S. (2019). Characterizing the regional contribution to PM 10 pollution over northern France using two complementary approaches: Chemistry transport and trajectory-based receptor models. *Atmospheric Research*, *223*(March), 1–14. https://doi.org/10.1016/j.atmosres.2019.03.002

Qi, H., Ma, S., Chen, J., Sun, J., Wang, L., Wang, N., Wang, W., Zhi, X., & Yang, H. (2022). Multi-model Evaluation and Bayesian Model Averaging in Quantitative Air Quality Forecasting in Central China. *Aerosol and Air Quality Research*, *22*(5), 210247. https://doi.org/10.4209/AAQR.210247

Rahman, M. M., Nayeem, M. E. H., Ahmed, M. S., Tanha, K. A., Sakib, M. S. A., Uddin, K. M. M., & Babu, H. M. H. (2024). AirNet: predictive machine learning model for air quality forecasting using web interface. *Environmental Systems Research*, *13*(1), 1–19. https://doi.org/10.1186/S40068-024-00378-Z/TABLES/5

Ravindiran, G., Karthick, K., Rajamanickam, S., Datta, D., Das, B., Shyamala, G., Hayder, G., & Maria, A. (2025). Ensemble stacking of machine learning models for air quality prediction for Hyderabad city in India. *IScience*, *28*(2). https://doi.org/10.1016/j.isci.2025.111894

Reddy, B. R., Raja, C. V. S. S. S., & Baskaran, S. R. (2020). Impact of land surface physics in WRF on the simulation of sea breeze circulation over southeast coast of India. *Meteorology and Atmospheric Physics*, *0123456789*. https://doi.org/10.1007/s00703-020-00726-5

Robeson, S. M., & Steyn, D. G. (1990). Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. *Atmospheric Environment. Part B. Urban Atmosphere*, *24*(2), 303–312. https://doi.org/10.1016/0957-1272(90)90036-T

Roodschild, M., Gotay Sardiñas, J., & Will, A. (2020). A new approach for the vanishing gradient problem on sigmoid activation. *Progress in Artificial Intelligence*, *9*(4), 351–360. https://doi.org/10.1007/s13748-020-00218-y

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/S42256-019-0048-X;SUBJMETA

Russo, M. A., Gama, C., & Monteiro, A. (2019). How does upgrading an emissions inventory affect air quality simulations? *Air Quality, Atmosphere and Health*, *12*(6), 731–741. https://doi.org/10.1007/s11869-019-00692-x

Ryan, W. F. (2016). The air quality forecast rote: Recent changes and future challenges. *Journal of the Air & Waste Management Association*, *66*(6), 576–596. https://doi.org/10.1080/10962247.2016.1151469

Rzeszutek, M., Kłosowska, A., & Oleniacz, R. (2023). Accuracy Assessment of WRF Model in the Context of Air Quality Modeling in Complex Terrain. *Sustainability (Switzerland)*, *15*(16), 12576. https://doi.org/10.3390/SU151612576/S1

Sahu, S. K., Beig, G., & Parkhi, N. (2015). High resolution emission inventory of NOx and CO for mega city Delhi, India. *Aerosol and Air Quality Research*, *15*(3), 1137–1144. https://doi.org/10.4209/aaqr.2014.07.0132

Saidi, L., Valari, M., & Ouarzazi, J. (2023). Air quality modeling in the city of Marrakech, Morocco using a local anthropogenic emission inventory. *Atmospheric Environment*, *293*(October 2022), 119445. https://doi.org/10.1016/j.atmosenv.2022.119445

Salma, I., Vörösmarty, M., Gyöngyösi, A. Z., Thén, W., & Weidinger, T. (2020). What can we learn about urban air quality with regard to the first outbreak of the COVID-19 pandemic? A case study from central Europe. *Atmospheric Chemistry and Physics*, *20*(24), 15725–15742. https://doi.org/10.5194/ACP-20-15725-2020

Sayeed, A. (2021). *Integrating Deep Neural Network with Numerical Models to Have Better Weather and Air Quality Forecast Both Spatially and Temporally*. https://uh-ir.tdl.org/items/67c17fbc-1672-4b28-9b62-24194c1e4a0e

Sayeed, A., Choi, Y., Eslami, E., Jung, J., Lops, Y., Salman, A. K., Lee, J. B., Park, H. J., & Choi, M. H. (2021). A novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14 days in advance. *Scientific Reports*, *11*(1), 1–8. https://doi.org/10.1038/s41598-021-90446-6

Scheffe, R. D., & Morris, R. E. (1993). A review of the development and application of the Urban Airshed model. *Atmospheric Environment. Part B. Urban Atmosphere*, *27*(1), 23–39. https://doi.org/10.1016/0957-1272(93)90043-6

Schölkopf, B. (1998). SVMs - A practical consequence of learning theory. *IEEE Intelligent Systems and Their Applications*, *13*(4), 18–21. https://doi.org/10.1109/5254.708428

Schwarz, G. (1978). Estimating the Dimension of a Model. *Ann. Statist. 6(2): 461-464.* DOI: 10.1214/aos/1176344136

Seidel, D. J., Birnbaum, A. N., Seidel@noaa, D., Gov, D. J., & Seidel, ). (2015). *Effects of Independence Day fireworks on atmospheric concentrations of fine particulate matter in the United States*. https://doi.org/10.1016/j.atmosenv.2015.05.065

Shami, A. Al, Aawar, E. Al, Baayoun, A., Saliba, N. A., Kushta, J., Christoudias, T., & Lakkis, I. (2022). Updated national emission inventory and comparison with the Emissions Database for Global Atmospheric Research (EDGAR): case of Lebanon. *Environmental Science and Pollution Research*, *29*(20), 30193–30205. https://doi.org/10.1007/s11356-021-17562-8

Sharma, H., Shrivastava, M., & Singh, B. (2023). Physics informed deep neural network embedded in a chemical transport model for the Amazon rainforest. *Npj Climate and Atmospheric Science*, *6*(1), 1–10. https://doi.org/10.1038/S41612-023-00353-Y;SUBJMETA

Shaziayani, W. N., Ul-Saufie, A. Z., Ahmat, H., & Al-Jumeily, D. (2021). Coupling of quantile regression into boosted regression trees (BRT) technique in forecasting emission model of PM10 concentration. *Air Quality, Atmosphere and Health*, *14*(10), 1647–1663. https://doi.org/10.1007/S11869-021-01045-3/FIGURES/5

Sicard, P., Crippa, P., De Marco, A., Castruccio, S., Giani, P., Cuesta, J., Paoletti, E., Feng, Z., & Anav, A. (2021). High spatial resolution WRF-Chem model over Asia: Physics and chemistry evaluation. *Atmospheric Environment*, *244*(October 2020), 118004. https://doi.org/10.1016/j.atmosenv.2020.118004

Simon, H., Beck, L., Bhave, P. V., Divita, F., Hsu, Y., Luecken, D., David Mobley, J., Pouliot, G. A., Reff, A., Sarwar, G., & Strum, M. (2010). The development and uses of EPA's SPECIATE database. *Atmospheric Pollution Research*, *1*(4), 196–206. https://doi.org/10.5094/APR.2010.026

Simon, J. S., & Chow, F. K. (2021). Alternative Anisotropic Formulations for Eddy-Viscosity Models in the Weather Research and Forecasting Model. *Boundary-Layer Meteorology*, *181*(1), 11–37. https://doi.org/10.1007/s10546-021-00642-0

Singh, T., Biswal, A., Mor, S., Ravindra, K., Singh, V., & Mor, S. (2020). A high-resolution emission inventory of air pollutants from primary crop residue burning over Northern India based on VIIRS thermal anomalies. *Environmental Pollution*, *266*, 115132. https://doi.org/10.1016/j.envpol.2020.115132

Slättberg, N., Lai, H. W., Chen, X., Ma, Y., & Chen, D. (2022). Spatial and temporal patterns of planetary boundary layer height during 1979–2018 over the Tibetan Plateau using

ERA5. *International Journal of Climatology*, *42*(6), 3360–3377. https://doi.org/10.1002/joc.7420

*SMOKE v4.9 User's Manual*. (2022). Retreived from https://www.cmascenter.org/smoke/documentation/4.9/html/index.html

Sokhi, R. S., Moussiopoulos, N., Baklanov, A., Bartzis, J., Coll, I., Finardi, S., Friedrich, R., Geels, C., Grönholm, T., Halenka, T., Ketzel, M., Maragkidou, A., Matthias, V., Moldanova, J., Ntziachristos, L., Schäfer, K., Suppan, P., Tsegas, G., Carmichael, G., … Kukkonen, J. (2022). Advances in air quality research - current and emerging challenges. *Atmospheric Chemistry and Physics*, *22*(7), 4615–4703. https://doi.org/10.5194/ACP-22-4615-2022

Solbakken, K., & Birkelund, Y. (2018). Evaluation of the Weather Research and Forecasting (WRF) model with respect to wind in complex terrain. *Journal of Physics: Conference Series*, *1102*(1). https://doi.org/10.1088/1742-6596/1102/1/012011

Sotomayor-Olmedo, A., Aceves-Fernández, M. A., Gorrostieta-Hurtado, E., Pedraza-Ortega, C., Ramos-Arreguín, J. M., Vargas-Soto, J. E., Sotomayor-Olmedo, A., Aceves-Fernández, M. A., Gorrostieta-Hurtado, E., Pedraza-Ortega, C., Ramos-Arreguín, J. M., & Vargas-Soto, J. E. (2013). Forecast Urban Air Pollution in Mexico City by Using Support Vector Machines: A Kernel Performance Approach. *International Journal of Intelligence Science*, *3*(3), 126–135. https://doi.org/10.4236/IJIS.2013.33014

Suárez, M., Poffo, D., Pierobon, E., Martina, A., Saffe, J., & Rodríguez, A. (2022). Wind and Gust Forecasts Assessment of Weather Research and Forecast (WRF) Model in Córdoba, Argentina. *Asian Journal of Atmospheric Environment*, *16*(1), 1–10. https://doi.org/10.5572/ajae.2021.133

Sudha, R., Damodaran, A., & Manohar, G. (2025). Enhanced air quality prediction using adaptive residual Bi-LSTM with pyramid dilation and optimal weighted feature selection. *Scientific Reports*, *15*(1), 1–25. https://doi.org/10.1038/S41598-025-14668-8;SUBJMETA

Tang, D., Zhan, Y., & Yang, F. (2024). A review of machine learning for modeling air quality: Overlooked but important issues. *Atmospheric Research*, *300*, 107261. https://doi.org/10.1016/J.ATMOSRES.2024.107261

Tao, H., Xing, J., Zhou, H., Pleim, J., Ran, L., Chang, X., Wang, S., Chen, F., Zheng, H., & Li, J. (2020). Impacts of improved modeling resolution on the simulation of meteorology, air quality, and human exposure to PM2.5, O3 in Beijing, China. *Journal of Cleaner Production*, *243*, 118574. https://doi.org/10.1016/j.jclepro.2019.118574

Tchepel, O., Monteiro, A., Dias, D., Gama, C., Pina, N., Rodrigues, J. P., Ferreira, M., & Miranda, A. I. (2020). Urban aerosol assessment and forecast: Coimbra case study. *Atmospheric Pollution Research*, *11*(7), 1155–1164. https://doi.org/10.1016/J.APR.2020.04.006

Terrenoire, E., Bessagnet, B., Rouïl, L., Tognet, F., Pirovano, G., Létinois, L., Beauchamp, M., Colette, A., Thunis, P., Amann, M., & Menut, L. (2015). High-resolution air quality simulation over Europe with the chemistry transport model CHIMERE. *Geoscientific Model Development*, *8*(1), 21–42. https://doi.org/10.5194/gmd-8-21-2015

Tian, H., Kong, H., & Wong, C. (2024). A Novel Stacking Ensemble Learning Approach for Predicting PM2.5 Levels in Dense Urban Environments Using Meteorological Variables: A Case Study in Macau. *Applied Sciences 2024, Vol. 14, Page 5062*, *14*(12), 5062. https://doi.org/10.3390/APP14125062

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. In *Monthly Weather Review* (Vol. 117, Issue 8, pp. 1779–1800). https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2

Toure, S. I., Stow, D. A., Shih, H. chien, Weeks, J., & Lopez-Carr, D. (2018). Land cover and land use change analysis using multi-spatial resolution data and object-based image analysis. *Remote Sensing of Environment*, *210*(January 2017), 259–268. https://doi.org/10.1016/j.rse.2018.03.023

Verwer, J. G., Blom, J. G., Van Loon, M., & Spee, E. J. (1996). A comparison of stiff ode solvers for atmospheric chemistry problems. *Atmospheric Environment*, *30*(1), 49–58. https://doi.org/10.1016/1352-2310(95)00283-5

Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., & Kukkonen, J. (2011). Evaluation of a multiple regression model for the forecasting of the concentrations of NOx and PM10 in Athens and Helsinki. *Science of the Total Environment*, 409(8), 1559–1571. https://doi.org/10.1016/j.scitotenv.2010.12.040

von Schneidemesser, E., Sibiya, B., Caseiro, A., Butler, T., Lawrence, M. G., Leitao, J., Lupascu, A., & Salvador, P. (2021). Learning from the COVID-19 lockdown in berlin: Observations and modelling to support understanding policies to reduce NO2. *Atmospheric Environment: X*, *12*, 100122. https://doi.org/10.1016/j.aeaoa.2021.100122

Wang, Y., Huang, L., Huang, C., Hu, J., & Wang, M. (2023). High-resolution modeling for criteria air pollutants and the associated air quality index in a metropolitan city. *Environment International*, *172*, 107752. https://doi.org/10.1016/J.ENVINT.2023.107752

Wolf, T., Pettersson, L. H., & Esau, I. (2020). A very high-resolution assessment and modelling of urban air quality. *Atmospheric Chemistry and Physics*, *20*(2), 625–647. https://doi.org/10.5194/acp-20-625-2020

Xie, Y., Posada, F., & Minjares, R. (2020). *Diesel sulfur content impacts on Euro VI soot-free vehicles: Considerations for emerging markets*. https://doi.org/10.1007/s11783-016-0859-5

Xiong, L., Shamseldin, A. Y., & O'Connor, K. M. (2001). A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi–Sugeno fuzzy system. *Journal of Hydrology*, *245*(1–4), 196–217. https://doi.org/10.1016/S0022-1694(01)00349-3

Xu, Z., Zhang, H., Zhai, A., Kong, C., & Zhang, J. (2025). Stacking Ensemble Learning and SHAP-Based Insights for Urban Air Quality Forecasting: Evidence from Shenyang and Global Implications. *Atmosphere 2025, Vol. 16, Page 776*, *16*(7), 776. https://doi.org/10.3390/ATMOS16070776

Yang, D., Zhang, S., Niu, T., Wang, Y., Xu, H., Zhang, K. M., & Wu, Y. (2019). High-resolution mapping of vehicle emissions of atmospheric pollutants based on large-scale, real-world traffic datasets. *Atmospheric Chemistry and Physics*, *19*(13), 8831–8843. https://doi.org/10.5194/acp-19-8831-2019

Yu, K., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Miller, C. C., Travis, K. R., Zhu, L., Yantosca, R. M., Sulprizio, M. P., Cohen, R. C., Dibb, J. E., Fried, A., Mikoviny, T., Ryerson, T. B., Wennberg, P. O., & Wisthaler, A. (2016). Sensitivity to grid resolution in the ability of a chemical transport model to simulate observed oxidant chemistry under high-isoprene conditions. *Atmospheric Chemistry and Physics*, *16*(7), 4369–4378. https://doi.org/10.5194/acp-16-4369-2016

Yu, Z., Wu, M., Min, J., Yan, Y., & Lou, X. (2022). Impacts of WRF Model Domain Size on Meiyu Rainfall Forecasts over Zhejiang, China. *Asia-Pacific Journal of Atmospheric Sciences*, *58*(2), 265–280. https://doi.org/10.1007/s13143-021-00254-1

Yuan, Z., Kerckhoffs, J., Shen, Y., Hoogh, K. De, & Hoek, G. (2023). Integrating large-scale stationary and local mobile measurements to estimate hyperlocal long-term air pollution using transfer learning methods. *Environmental Research*, *228*(April), 115836. https://doi.org/10.1016/j.envres.2023.115836

Zhang, L. H., Deng, Z. H., & Wang, W. B. (2021). PM2.5 Concentration Prediction Based on Markov Blanke Feature Selection and Hybrid Kernel Support Vector Regression Optimized by Particle Swarm Optimization. *Aerosol and Air Quality Research*, *21*(6), 200144. https://doi.org/10.4209/AAQR.200144

Zhang, L., Liu, P., Zhao, L., Wang, G., Zhang, W., & Liu, J. (2021). Air quality predictions with a semi-supervised bidirectional LSTM neural network. *Atmospheric Pollution Research*, *12*(1), 328–339. https://doi.org/10.1016/J.APR.2020.09.003

Zhang, W., Wang, T., Bai, J. S., Li, P., Wan, Z. H., & Sun, D. J. (2017). The piecewise parabolic method for Riemann problems in nonlinear elasticity. *Scientific Reports*, *7*(1), 1–17. https://doi.org/10.1038/s41598-017-13484-z

Zhang, Z., Johansson, C., Engardt, M., Stafoggia, M., & Ma, X. (2024). Improving 3-day deterministic air pollution forecasts using machine learning algorithms. *Atmospheric Chemistry and Physics*, *24*(2), 807–851. https://doi.org/10.5194/ACP-24-807-2024

Zhang, Z., Zhang, S., Zhao, X., Chen, L., & Yao, J. (2022). Temporal Difference-Based Graph Transformer Networks For Air Quality PM2.5 Prediction: A Case Study in China. *Frontiers in Environmental Science*, *10*, 924986. https://doi.org/10.3389/FENVS.2022.924986/BIBTEX

Zhao, Y., Nielsen, C. P., Lei, Y., Mcelroy, M. B., & Hao, J. (2011). Atmospheric Chemistry and Physics Quantifying the uncertainties of a bottom-up emission inventory of anthropogenic atmospheric pollutants in China. *Atmos. Chem. Phys*, *11*, 2295–2308. https://doi.org/10.5194/acp-11-2295-2011

Zhao, Z., Wu, J., Cai, F., Zhang, S., & Wang, Y. G. (2022). A statistical learning framework for spatial-temporal feature selection and application to air quality index forecasting. *Ecological Indicators*, *144*, 109416. https://doi.org/10.1016/J.ECOLIND.2022.109416

Zhu, Y., Liu, C., Hu, Q., Teng, J., You, D., Zhang, C., Ou, J., Liu, T., Lin, J., Xu, T., & Hong, X. (2022). Impacts of TROPOMI-Derived NOX Emissions on NO2 and O3 Simulations in the NCP during COVID-19. *ACS Environmental Au*, *2*(5), 441. https://doi.org/10.1021/ACSENVIRONAU.2C00013

Zimmerman, N., Li, H. Z., Ellis, A., Hauryliuk, A., Robinson, E. S., Gu, P., Shah, R. U., Ye, Q., Snell, L., Subramanian, R., Robinson, A. L., Apte, J. S., & Presto, A. A. (2020). Improving correlations between land use and air pollutant concentrations using wavelet analysis: Insights from a low-cost sensor network. *Aerosol and Air Quality Research*, *20*(2), 314–328. https://doi.org/10.4209/aaqr.2019.03.0124

**Appendix 2**

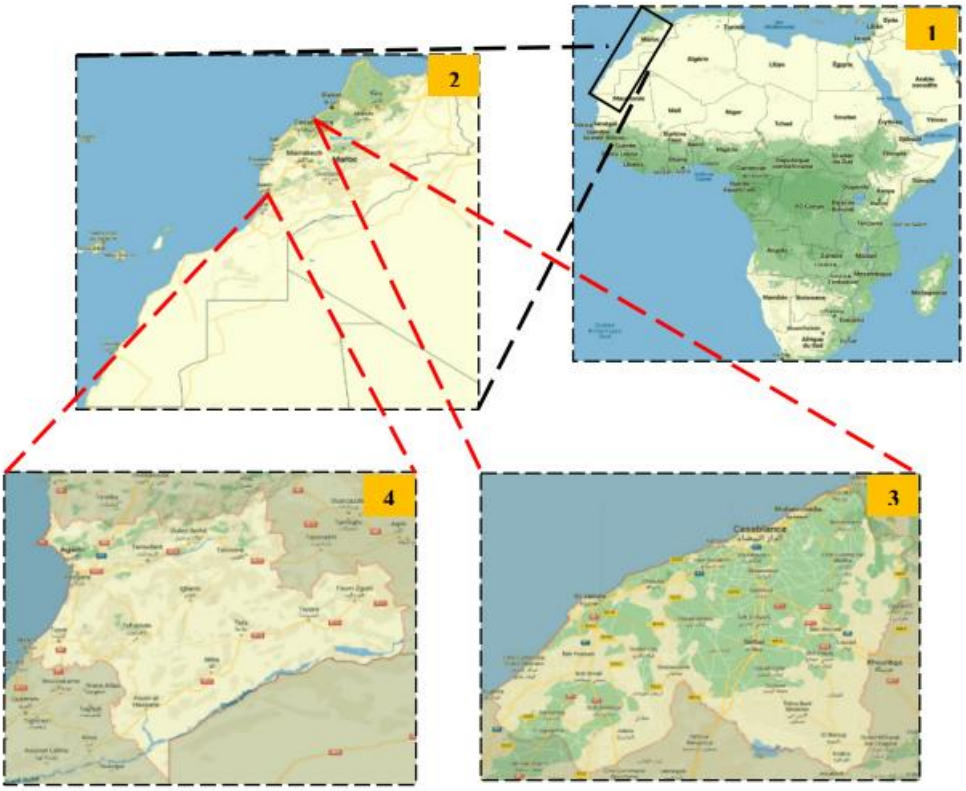



**Figure S1. The geographic area of the study site (1) Africa, (2) Morocco, (3) Casablanca, (4) Agadir**

## Appendix 3

Table 1. Stationarity tests results for $NO_2$ Concentration data.

| $NO_2$ | 1h | | | 3h | | | 12h | | |
|---|---|---|---|---|---|---|---|---|---|
| **Station** | **ADF** | **KPSS** | **p-value** | **ADF** | **KPSS** | **p-value** | **ADF** | **KPSS** | **p-value** |
| Erzsébet square | -12.94 | 10.422 | 0.01 | -5.814 | 6.243 | 0.01 | -4.005 | 3.689 | 0.01 |
| Budatétény | -36.83 | 0.978 | 0.01 | -6.596 | 0.756 | 0.01 | -3.712 | 0.614 | 0.02 |
| Csepel | -15.07 | 1.311 | 0.01 | -5.266 | 0.979 | 0.01 | -3.507 | 0.533 | 0.04 |
| Honvéd complex | -17.95 | 3.109 | 0.01 | -6.001 | 2.247 | 0.01 | -4.101 | 1.203 | 0.01 |
| Gilice square | -18.62 | 0.537 | 0.03 | -7.636 | 0.209 | 0.1 | -4.323 | 0.173 | 0.1 |
| Gergely street | -19.26 | 1.655 | 0.01 | -6.512 | 0.347 | 0.01 | -4.083 | 0.466 | 0.01 |
| Széna square | -19.474 | 2.273 | 0.01 | -13.55 | -13.55 | 0.01 | -7.55 | 1.173 | 0.01 |
| Teleki square | -19.358 | 0.752 | 0.01 | -13.69 | 2.047 | 0.01 | -7.47 | 1.382 | 0.01 |
| Pesthidegkút | -15.51 | 0.395 | 0.07 | -10.87 | 0.260 | 0.1 | -6.18 | 0.163 | 0.1 |
| Kőrakás park | -18.19 | 0.623 | 0.02 | -11.90 | 0.401 | 0.07 | -6.52 | 0.252 | 0.1 |
| Kosztolányi D. | -14.48 | 0.816 | 0.01 | -10.21 | 0.507 | 0.04 | -6.51 | 0.305 | 0.1 |

## Appendix 4

Table 2. Performance of fusion techniques at 11 stations.

| | Mean | | | BMA | | | Stacking Ensemble | | Choquet Integral | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | K=3 | | | K=5 | | | |
| | RMSE | MAE | $R^2$ | RMSE | MAE | $R^2$ | RMSE | MAE | $R^2$ | RMSE | MAE | $R^2$ | RMSE | MAE | $R^2$ |
| **Budatétény** | 5.31 | 3.74 | 0.85 | 2.46 | 1.97 | 0.96 | 1.66 | 1.25 | 0.98 | 3.14 | 2.11 | 0.94 | 1.82 | 1.38 | 0.98 |
| **Csepel** | 7.12 | 5.00 | 0.79 | 1.70 | 1.30 | 0.98 | 1.27 | 0.79 | 0.99 | 1.51 | 1.01 | 0.99 | 1.65 | 1.11 | 0.98 |
| **Erszebet square** | 5.63 | 3.51 | 0.83 | 6.02 | 2.97 | 0.80 | 1.21 | 0.73 | 0.99 | 3.97 | 2.13 | 0.91 | 2.57 | 1.62 | 0.96 |
| **Gergely street** | 4.35 | 3.031 | 0.87 | 4.015 | 2.630 | 0.88 | 1.19 | 0.95 | 0.99 | 1.95 | 1.66 | 0.973 | 3.20 | 2.14 | 0.92 |
| **Gilice square** | 13.44 | 9.53 | 0.47 | 2.81 | 2.12 | 0.97 | 3.39 | 2.25 | 0.96 | 3.96 | 2.48 | 0.95 | 4.38 | 2.62 | 0.94 |
| **Honved sport complex** | 4.82 | 3.49 | 0.87 | 4.47 | 3.01 | 0.89 | 1.49 | 1.21 | 0.98 | 1.59 | 1.14 | 0.98 | 2.68 | 1.86 | 0.96 |
| **Kosztolanyi D square** | 7.58 | 5.36 | 0.68 | 1.10 | 0.87 | 0.99 | 0.74 | 0.48 | 0.99 | 1.38 | 1.06 | 0.98 | 1.75 | 1.1 | 0.98 |
| **Kőrakás park** | 5.52 | 3.61 | 0.83 | 1.08 | 0.86 | 0.99 | 1.64 | 0.66 | 0.98 | 3.89 | 2.33 | 0.91 | 1.16 | 0.86 | 0.99 |
| **Pesthidegkut** | 3.97 | 2.57 | 0.77 | 4.25 | 2.63 | 0.73 | 0.51 | 0.27 | 0.99 | 2.76 | 1.72 | 0.89 | 3.27 | 1.95 | 0.84 |
| **Széna square** | 4.00 | 3.20 | 0.88 | 3.29 | 2.53 | 0.92 | 1.09 | 0.82 | 0.99 | 1.73 | 1.42 | 0.97 | 2.64 | 2.06 | 0.95 |
| **Teleki square** | 4.88 | 3.36 | 0.84 | 0.72 | 0.55 | 0.99 | 0.64 | 0.47 | 0.99 | 0.97 | 0.74 | 0.99 | 4.11 | 2.59 | 0.88 |

## Appendix 5

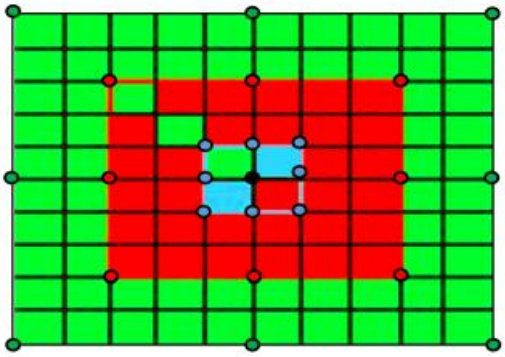


**Figure S2. Spatial discretization between the CHIMERE domains $R_{1A}$ (Green), $R_{2A}$ (Red), $R_{3A}$ (Blue) and the observation site (Black).**

## Appendix 6

Table 3. Coordinates of the four cells representing WRF and CHIMERE domains

| | | North-West | North-East | South-East | South-West |
|---|---|---|---|---|---|
| **WRF domain** | | | | | |
| d01 domain | Latitude | 43.82453° | 43.82453° | 20.87130° | 20.87130° |
| | Longitude | -21.24654° | 7.972054° | 7.97205° | -21.24654° |
| d02 domain | Latitude | 36.71437° | 36.71437° | 29.86381° | 29.86381° |
| | Longitude | -12.02137° | -2.72593° | -2.72593° | -12.02137° |
| **CHIMERE domain** | | | | | |
| **Agadir city** | | | | | |
| R1A domain (Green) | Latitude | 33.7972° | 33.7972° | 28.1443° | 28.1443° |
| | Longitude | -14.86298° | -2.78170° | -2.78170° | 14.86298° |
| R2A domain (Red) | Latitude | 32.63318° | 32.63318° | 29.04184° | 29.04184 |
| | Longitude | -13.2211° | -4.80623° | -4.80623° | -13.2211 |
| R3A domain (Blue) | Latitude | 30.99084° | 30.99084° | 29.99848° | 29.99848 |
| | Longitude | -10.87559° | -8.31281° | -8.31281° | -10.87559 |
| **Casablanca city** | | | | | |
| R1C domain (Black) | Latitude | 36.11015° | 36.11015° | 27.59972° | 27.59972° |
| | Longitude | -15.1708° | -1.50715° | -1.50715° | -15.1708° |
| R2C domain (Brown) | Latitude | 34.82588° | 34.82588° | 29.58078° | 29.58078° |
| | Longitude | -10.80079° | -4.142032° | -4.142032° | -10.80079° |

# Appendix 7

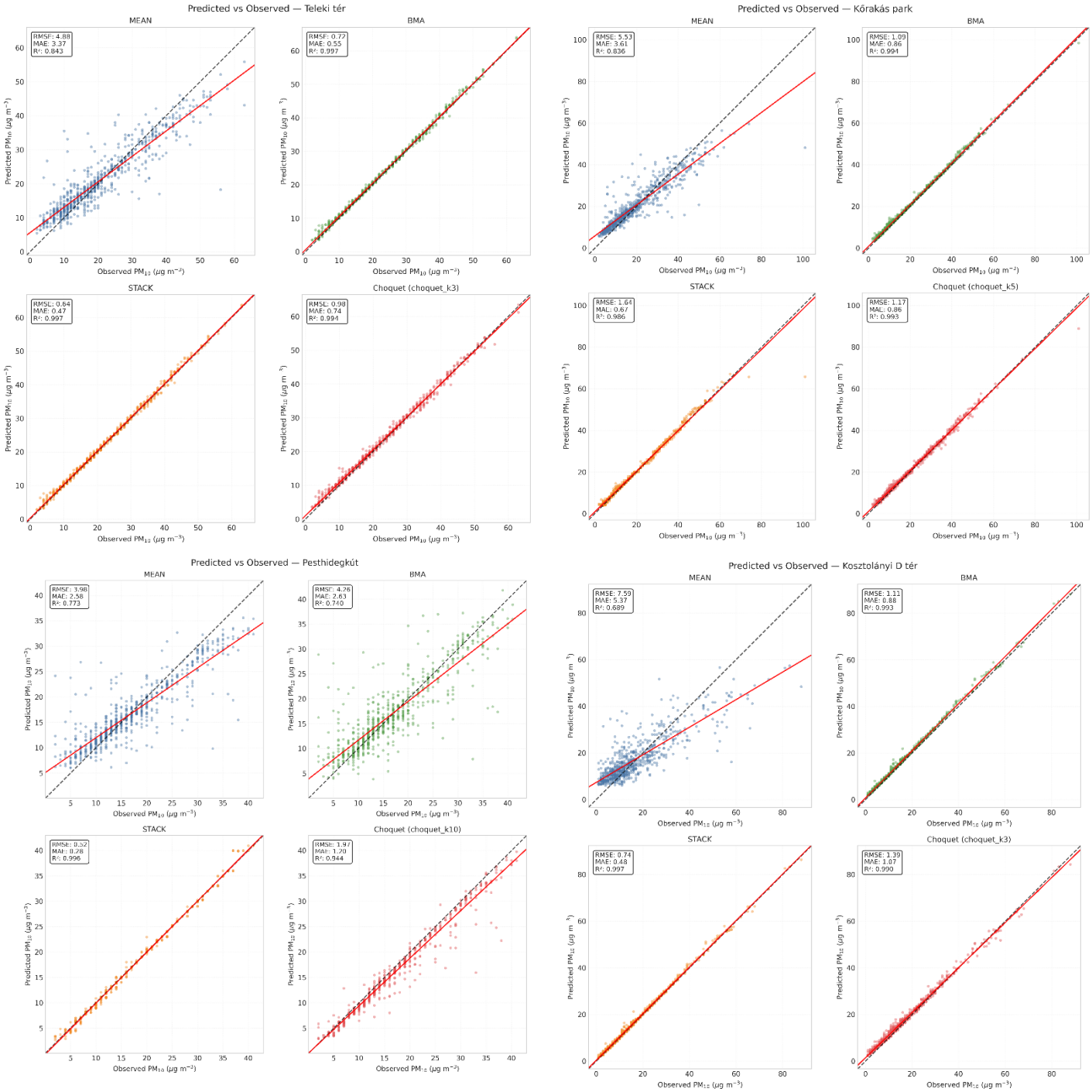



**Figure S3. Scatter plots of different fusion methods in Teleki tér, Kőrákás park, Pesthidegkút and Kosztolányi D tér.**

# Appendix 8

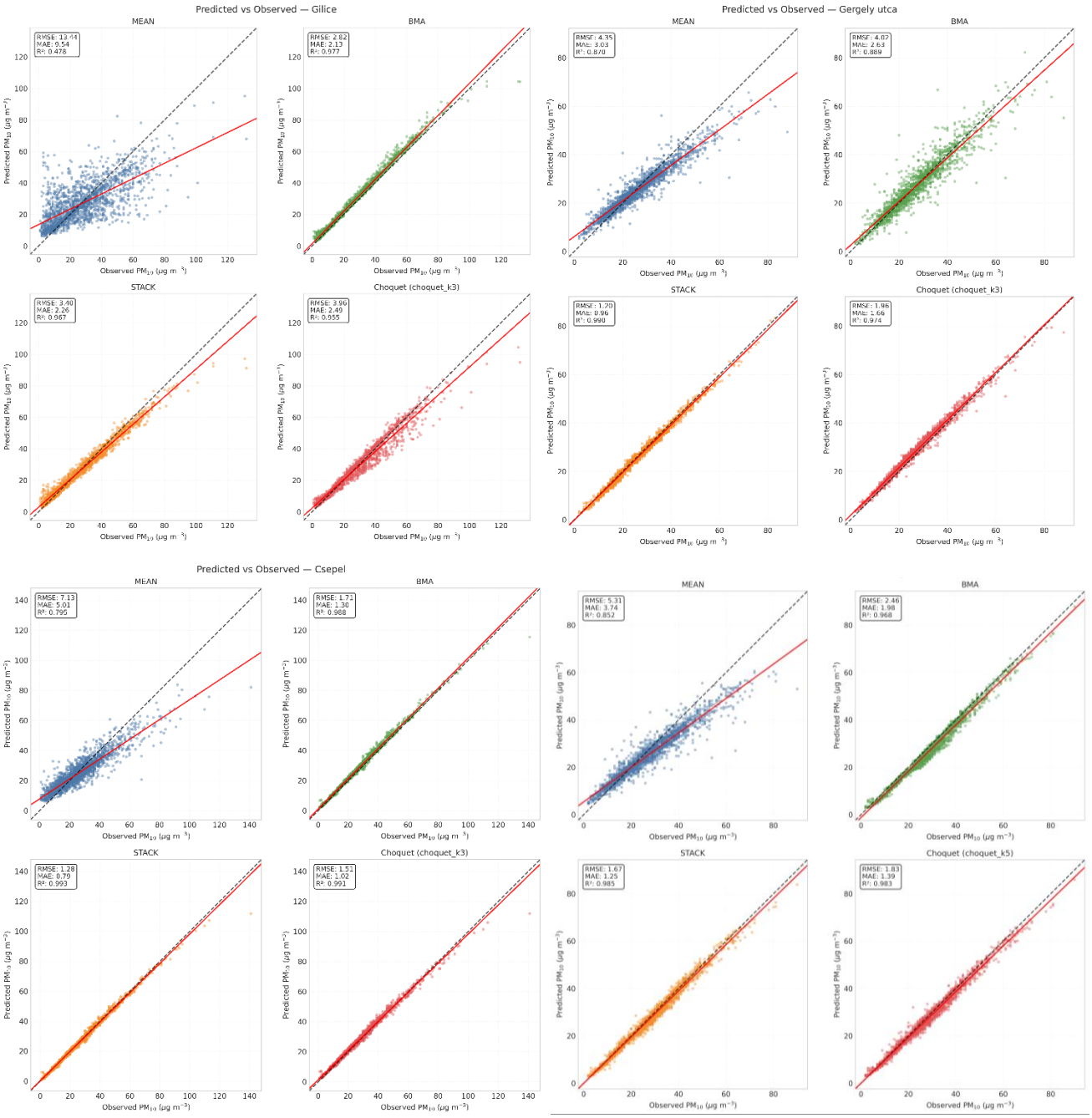



**Figure S4. Scatter plots of different fusion methods in Gilice tér, Gergely utca, Csepel and Budatétény station.**

ACKNOWLEDGMENT

In the name of Allah, the Most Gracious, the Most Merciful, whose divine guidance illuminated my path through the shadows of doubt and carried me through the storms of this doctoral voyage. To Him, I owe my first and eternal gratitude for the strength that emerged when I thought I had none left, for the patience that bloomed in moments of frustration, and for the perseverance that transformed impossible dreams into this tangible reality.

To my husband, my anchor in turbulent seas, and my lighthouse in foggy nights, words feel inadequate to capture what your presence has meant. You didn't just support my dream; you lived it with me, breathed it with me, and sometimes carried it for me when my arms grew too weary. Every meal you prepared with love became fuel for another late night of writing. Every dish you washed was one less worry on my overwhelmed mind. In your patience, I found my peace. In your embrace, I found home, even 2,000 kilometres away from where we started. This achievement carries your fingerprints as much as mine.

To my beloved mother, **Nejma**, whose name means 'star', and truly, you have been my North Star throughout this journey. You didn't just open the door to Hungary; you opened the door to a future I couldn't have imagined. Your prayers crossed borders and time zones, wrapping around me like a warm embrace on cold Budapest nights. When self-doubt whispered that I wasn't enough, your voice thundered louder, reminding me whose daughter I am.

To my father, **Ahmed**, my mountain of strength, you have been the steady ground beneath my feet when everything else seemed to shift. Your quiet confidence in me spoke volumes. You taught me that foundations built with patience and determination can weather any storm. Every time I wanted to give up, I remembered your hands that never stopped working for our dreams, and I found the strength to continue.

To my supervisor, **Dr. Géczi Gábor**, who saw potential in a young woman from Morocco and nurtured it into expertise, thank you for being more than a supervisor. You were a mentor who celebrated every small victory, a guide who turned obstacles into opportunities, and a believer who never wavered. Your office door was always open, but more importantly, so was your mind. You didn't just teach me about air quality modelling; you taught me how to think like a scientist and dream like an innovator.

To **Dr. Amine Ajdour**, whose generosity with time and expertise knew no bounds, every CHIMERE and WRF simulation you ran was a building block in this thesis. Your patience with my endless requests and modifications transformed complex atmospheric models into comprehensible tools.

To **Dr. Najiya Omar**, whose encouragement arrived like rain in a desert, always at the perfect moment, and to **Dr. Khomsi Kenza**, whose coaching sessions were like a compass when I had lost my way. In a few profound conversations, you helped me untangle years of confusion and chart a clear course forward. The clarity you brought didn't just accelerate my progress; it transformed my entire approach to this work.

To the entire community at the **Hungarian University of Agriculture and Life Sciences**, from the Student Office staff who navigated bureaucracy with smiles, to the Doctoral School

leadership who created an environment where international students could flourish, you transformed what could have been a foreign land into a nurturing academic home. Every helpful email, every solved problem, every word of encouragement wove into the fabric of this success.

To the **Tempus Public Foundation**, thank you for believing in the dreams of a Moroccan student and investing in her potential. This opportunity wasn't just funding; it was a vote of confidence that I carry with profound gratitude and responsibility. You didn't just finance my education; you financed a bridge between cultures, between knowledge systems, and between dreams and reality.

To all the unnamed heroes, the friends who listened to practice presentations, the colleagues who shared a coffee and advice, the strangers who offered directions on confusing first days, and everyone whose path crossed mine and left it a little brighter, your kindness rippled through this journey in ways you may never know.

This thesis stands not as a monument to individual achievement, but as a testament to the power of community, love, and faith. Each page carries the invisible ink of your support, each equation solved with the energy of your encouragement, and each conclusion reached on the shoulders of giants who lifted me up.

From the depths of a heart overflowing with gratitude, thank you. *Shukran. Köszönöm.*