information

MDPI

Article

Design, Deployment, and Field Evaluation of a Low-Cost IoT-Based Monitoring System for Urban Particulate Matter: A Winter–Spring Campaign in Almaty, Kazakhstan

Daniyar Nurseitov [1], Kairat Bostanbekov [1], Galymzhan Abdimanap [1], Raissa Uskenbayeva [2], Zhuldyz Kalpeyeva [2] and Aiman Moldagulova [2,*]

[1] KazMunayGas Engineering LLP, Astana 010000, Kazakhstan; k.bostanbekov@kmge.kz (K.B.); g.abdimanap@kmge.kz (G.A.)
[2] Institute of Automation and Information Technologies, JSC Kazakh National Research Technical University Named After K.I. Satbayev, Almaty 050013, Kazakhstan; r.k.uskenbayeva@satbayev.university (R.U.); z.kalpeyeva@satbayev.university (Z.K.)
* Correspondence: a.moldagulova@satbayev.university

Abstract

Air pollution in Almaty, Kazakhstan, poses a critical public health challenge intensified by the city's basin topography and seasonal thermal inversions that trap anthropogenic emissions. The sparse stationary network (~5 stations for ~2 million inhabitants) lacks the spatial and temporal resolution needed to capture intra-urban variability. We present the design, deployment, and field evaluation of a low-cost distributed Internet of Things (IoT) network of six custom nodes—Winsen ZPHS01B multi-parameter modules with Raspberry Pi Zero 2 W edge units, at an estimated principal-component cost of ~US$100 per node—operated during a winter-spring campaign (February–April 2025) and yielding over 70,000 measurements of $PM_{2.5}$, PM_{10}, CO_2, temperature, and relative humidity. The system's novelty lies in three integrated engineering features: an Active Airflow Stabilization enclosure that decouples sampling from external wind, context-aware adaptive edge filtering that reduces transmitted data volume by ~40%, and a secure Edge-DMZ-Core telemetry pipeline. Node readings were cross-validated against a Qingping Air Monitor Pro with documented traceability to FEM-grade reference analyzers (R^2 = 0.89–0.95), and city-scale consistency was confirmed against the national monitoring dashboard; the network is therefore characterized as providing internally consistent low-cost observations rather than reference-equivalent concentrations. Daily mean $PM_{2.5}$ exceeded the WHO 24 h guideline (15 μg/m^3) on 84% of monitored days, with February concentrations (54.4 μg/m^3) significantly above March (21.9 μg/m^3; $p < 0.001$). A high $PM_{2.5}/PM_{10}$ ratio (~0.96), measured at the physically consistent nodes, together with higher weekend concentrations, points to coal-based residential heating as the most likely dominant source. A coupled WRF-SILAM framework is configured for future model-observation integration. The system offers a reproducible, scalable, and cost-effective template for ambient particulate monitoring in resource-constrained cities with complex terrain.

Keywords: air quality monitoring; particulate matter; low-cost sensors; IoT; $PM_{2.5}$; Almaty; urban pollution; sensor networks; edge computing; smart city

Check for updates

Academic Editor: Nader Sohrabi Safa

Received: 28 May 2026
Revised: 26 June 2026
Accepted: 28 June 2026
Published: 1 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Urban air pollution remains one of the most pressing environmental and public health challenges of the 21st century. A global analysis covering more than 13,000 cities from

28 2000 to 2019 revealed that, despite an overall annual decline of 0.2% in average $PM_{2.5}$ concentrations, 65% of urban areas experienced a sustained increase in pollution levels [1]. This spatial heterogeneity is driven by differences in natural conditions, emission structures, and the effectiveness of environmental policies. Particularly adverse trends are observed in developing regions such as Central Asia, where air quality management and monitoring infrastructure often lag behind international standards.

A representative case is Almaty, the largest metropolis in Kazakhstan and one of the most polluted cities in the region. Its location in a mountain basin at 700–900 m above sea level contributes to frequent temperature inversions and weak wind circulation, which promote the accumulation of pollutants in the lower atmosphere [2]. Dense and chaotic urban development further limits air exchange. Major pollution sources include coal-based heating (from both CHP plants and residential areas) and a vehicle fleet exceeding 930,000 units, most of which are technically outdated [3–5]. As a result, annual average concentrations of $PM_{2.5}$ and NO_2 regularly exceed WHO guidelines by a factor of 2–3 [6]. However, reliable air quality assessment in such conditions requires a robust monitoring system. Currently, monitoring in Almaty relies primarily on a small number of stationary stations operated by Kazhydromet. Although these provide accurate measurements, they lack sufficient spatial coverage and temporal resolution. Moreover, sensors placed at a height of approximately 3 m do not adequately reflect human exposure levels, particularly in complex urban environments [7]. To address these limitations, low-cost sensor networks have gained increasing attention for their ability to provide high spatiotemporal resolution monitoring. Deployed at fixed points or mounted on vehicles, these systems can detect pollution hotspots, track localized emission sources, and complement official monitoring networks [8,9]. Thanks to their flexibility, accessibility, and scalability, IoT-based platforms are proving to be effective tools for urban air diagnostics, especially in developing countries. Recent work has demonstrated the feasibility of such platforms for monitoring in-cabin air quality within Almaty's buses, trolleybuses, and metro, revealing that passenger density is a primary driver of in-vehicle pollutant accumulation [10]. However, monitoring the ambient outdoor urban atmosphere—where pollutants originate from heating, traffic, and industrial sources before entering any enclosed space—remains a distinct and equally critical challenge that requires purpose-built sensor networks with different design constraints. In this context, the present study aims to: (1) design a low-cost IoT-based air quality monitoring system adapted to the urban environment of Almaty; (2) deploy and operate the system during a winter–spring field campaign; and (3) evaluate the collected data to characterize temporal patterns of particulate matter pollution. Additionally, we describe the configuration of a coupled WRF–SILAM numerical modeling framework intended for future data fusion and regional-scale validation. This stepwise approach—from hardware design through field deployment to data analysis and planned model integration—provides a foundation for building a scientifically grounded and sustainable air quality management system in cities with complex terrain and limited observation infrastructure.

2. Materials and Methods

2.1. IoT-Based Air Quality Monitoring Systems

Recent trends in atmospheric monitoring indicate a transition from purely stationary measurement stations to hybrid systems that integrate low-cost sensor nodes, mobile platforms, satellite data, and numerical models. Guerrón et al. proposed an IoT architecture incorporating autonomous nodes (ESP32/Arduino) equipped with $PM_{2.5}$, SO_2, CO, and NO_2 sensors, with wireless transmission via Wi-Fi and LoRa. The study emphasizes scalability, energy autonomy, and sensor calibration [11]. Gryech et al. comprehensively reviewed IoT-based systems and machine learning algorithms used for pollutant prediction,

stressing that the most effective systems feature multichannel preprocessing, robust data cleaning, and incorporation of meteorological inputs [12].

In practical implementations, Sladojevi'c et al. developed an Android-based platform combining a smartphone with an external gas analyzer, achieving high correlation ($r > 0.9$) for CO against reference instruments [13]. Whitehill et al. investigated mobile measurement reproducibility across U.S. cities, introducing the concept of "mobile collocation" for sensor calibration [14].

2.2. Low-Cost Monitoring in Developing Countries

Developing countries face a pressing need for adaptive, low-cost systems due to limited access to certified infrastructure. The SATVAM project in India deployed a decentralized IoT network using 6LoWPAN and IPv6/RPL, demonstrating high flexibility in both urban and rural environments [15]. In Hyderabad, a dense network of 49 PM sensors enabled real-time spatiotemporal analysis including seasonal recalibration [16]. In El Salvador, an ESP32-based system with PMS5003ST sensors transmitted data to Google Cloud, demonstrating that even minimal investment enables monitoring in areas lacking expensive equipment [17]. The Enviro-IoT system achieved 98% accuracy for $PM_{2.5}$ and 97% for PM_{10} after nine months of field trials [18].

2.3. Numerical Atmospheric Models and Sensor Integration

Numerical models such as WRF and SILAM play a key role in simulating pollutant dispersion and are increasingly integrated with sensor networks. Song and Han introduced Deep-MAPS, a deep learning framework integrating mobile measurements, meteorological parameters, and satellite observations to predict $PM_{2.5}$ at 1 km $\times$ 1 h resolution [19]. In a study on Delhi, LUR models incorporating mobile data and urban morphology revealed significant spatial heterogeneity in pollution levels [20]. For Almaty specifically, winter stagnation episodes have been simulated using WRF v4.3 with nested grids (9 km $\rightarrow$ 3 km $\rightarrow$ 1 km), the BEP urban scheme, and surface layer modules. The model successfully reproduced temperature inversions and calm wind conditions [21]. A hybrid WRF–SILAM system applied in East Kazakhstan achieved accurate source localization at resolutions down to 70 m [22]. In North Macedonia, a WRF-Chem system forecasted PM_{10} concentrations with satisfactory accuracy during a winter pollution episode [23]. Long-term SILAM modeling in Helsinki demonstrated adequate model performance over four years using standard statistical metrics [24]. In three African cities, a combination of WRF and CHIMERE models reproduced seasonal and diurnal $PM_{2.5}$ variability, though challenges remained in capturing local sources [25]. These studies confirm the value of coupling numerical models with observational sensor networks, particularly in complex urban environments where sparse official monitoring cannot capture fine-grained pollution dynamics.

2.4. System Architecture and Hardware Design

The developed solution represents a multi-layered IoT ecosystem designed for integration into a "Smart City" digital infrastructure. The architecture ensures a closed data lifecycle—from physical sensing to analytical interpretation—utilizing a secure Edge–DMZ–Core security model.

2.4.1. General Architecture

The system consists of three logical levels: the Physical Level (sensor nodes), the Network Level (secure telemetry transmission), and the Application Level (API Gateway and Analytics). Figure 1 illustrates the interaction between the Raspberry Pi Zero 2 W computation unit, the sensing block, and the cloud backend.

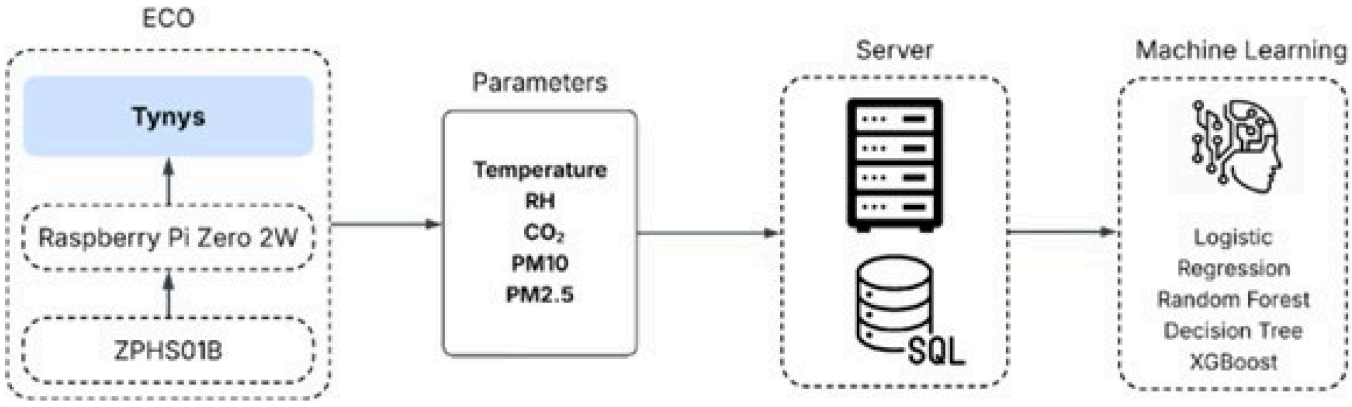


Figure 1. System architecture integrating the mobile sensing node, secure transmission channels, and the server-side PostgreSQL database with analytical dashboards.

2.4.2. Embedded Sensing Node

The core computation and control unit is the Raspberry Pi Zero 2 W single-board computer (Raspberry Pi Ltd, Cambridge, UK). This platform was selected for its balance of processing power (quad-core ARM Cortex-A53), low energy footprint, and integrated wireless connectivity (802.11 b/g/n and Bluetooth 4.2). The Pi runs a lightweight, headless Linux distribution optimized for serial data acquisition and telemetry.

The sensing interface utilizes the Winsen ZPHS01B multi-parameter module (Zhengzhou Winsen Electronics Technology Co., Ltd., Zhengzhou, China) connected via a UART serial interface. This module was selected not only for its integrated multi-sensor capability but also because its performance has been independently evaluated against a regulated air quality monitoring station (AQMS), demonstrating $R^2 \approx 0.8$ for $PM_{2.5}$ and O_3 after linear calibration [26]. Unlike discrete sensor arrays that require complex wiring and individual calibration, the ZPHS01B provides simultaneous measurements of particulate matter ($PM_{1.0}$, $PM_{2.5}$, PM_{10}) via laser scattering, carbon dioxide (CO_2) via NDIR, and environmental variables (temperature, relative humidity). GPS/GLONASS positioning is provided by an external receiver. Table 1 summarizes the key technical specifications of the sensing hardware.

Table 1. Technical specifications of the sensing node hardware.

Component	Parameter	Specification
Winsen ZPHS01B	$PM_{2.5}/PM_{10}$ range	0–1000 $\mu g/m^3$
	PM resolution	1 $\mu g/m^3$
	CO_2 range	400–5000 ppm
	Temperature range	−10 to +50 °C
	Humidity range	0–95% RH
	Interface	UART (9600 baud)
Raspberry Pi Zero 2 W	CPU	Quad-core ARM Cortex-A53 @ 1 GHz Wi-Fi
	Connectivity	802.11 b/g/n, BLE 4.2
	Power consumption	~1.0 W (idle), ~2.5 W (active)
GPS Module	Accuracy	±2.5 m CEP
Complete node	Estimated unit cost	~$80–100 USD

2.4.3. Enclosure and Active Airflow Stabilization

A critical engineering challenge in outdoor monitoring is the variability of air intake conditions caused by wind and vehicle-induced turbulence. To mitigate these effects, the sensor node is housed in a custom-designed enclosure (Figure 2). The design features an

Active Airflow Stabilization system: a miniature intake fan drives a constant volumetric flow rate through the sensing chamber, decoupling the internal sampling environment from external transient flow conditions. This forced-air mechanism ensures measurement repeatability regardless of external wind conditions. The inlet geometry utilizes an inertial separation path to shield the sensing elements from direct impact by large particles and moisture.

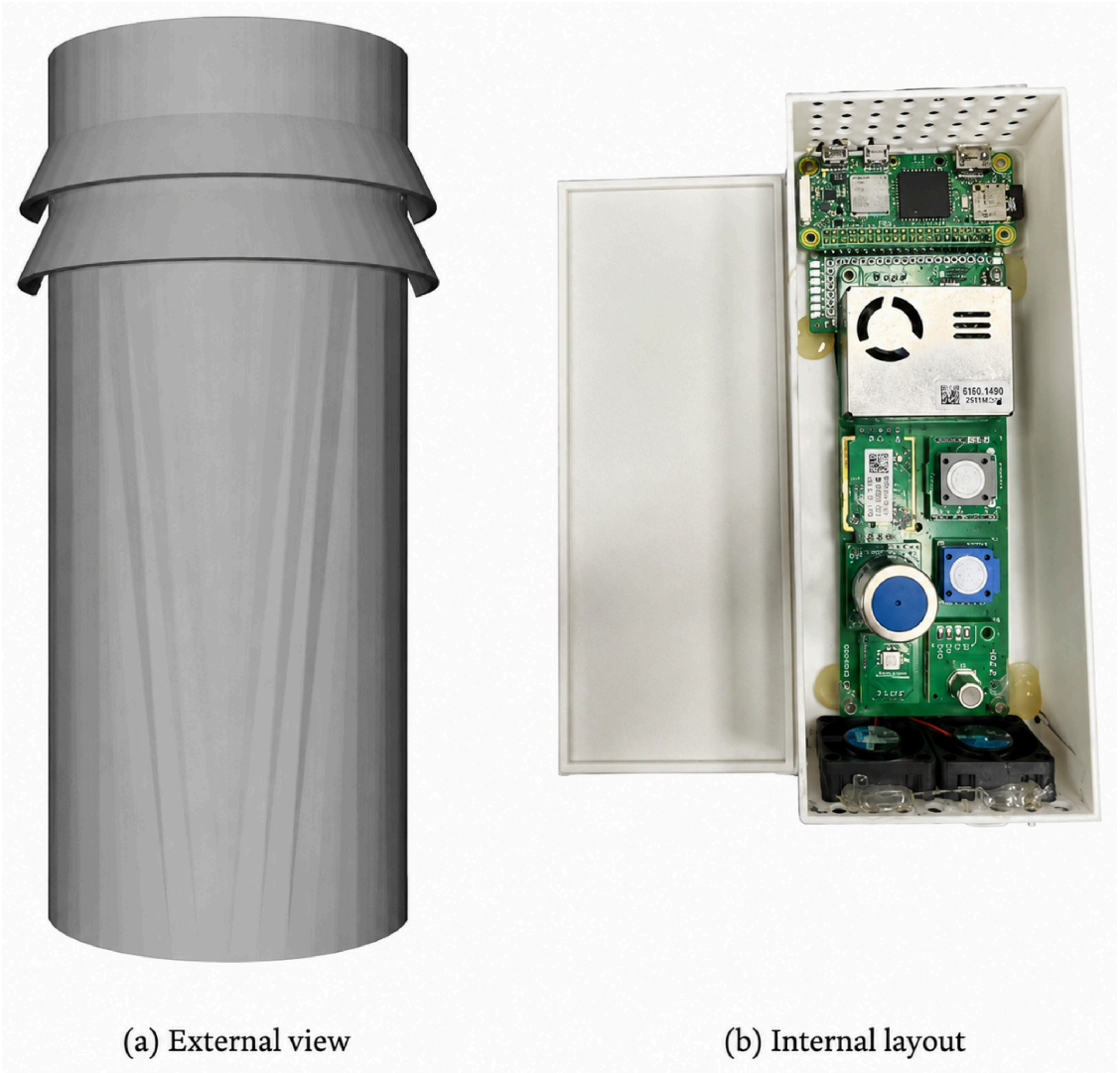


Figure 2. Sensing module design: (**a**) overall enclosure with intake vents; (**b**) internal airflow channel and fan placement.

2.5. Edge Computing and Data Pipeline

To ensure data reliability before transmission, the system implements a robust "Edge-First" processing pipeline. This approach minimizes bandwidth usage and filters out noise typical of outdoor deployments.

2.5.1. Acquisition and Adaptive Filtering

The on-board Python 3.10-based software stack executes a deterministic pipeline (Figure 3). The process begins with UART synchronization and checksum verification to reject packets corrupted by electromagnetic interference.

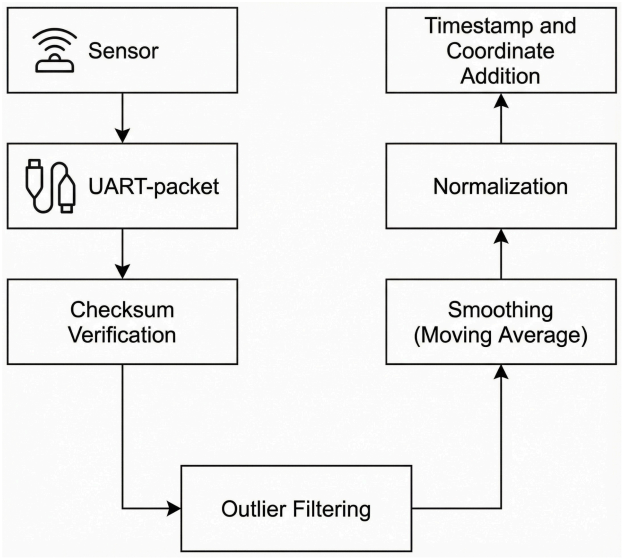


Figure 3. Data processing algorithm: from UART signal acquisition and checksum validation to JSON serialization.

A key feature of the edge processing stack is the Adaptive Filtering Algorithm. The system dynamically switches between two smoothing regimes based on signal variance and, where GPS velocity is available, vehicle motion state (Figure 4):

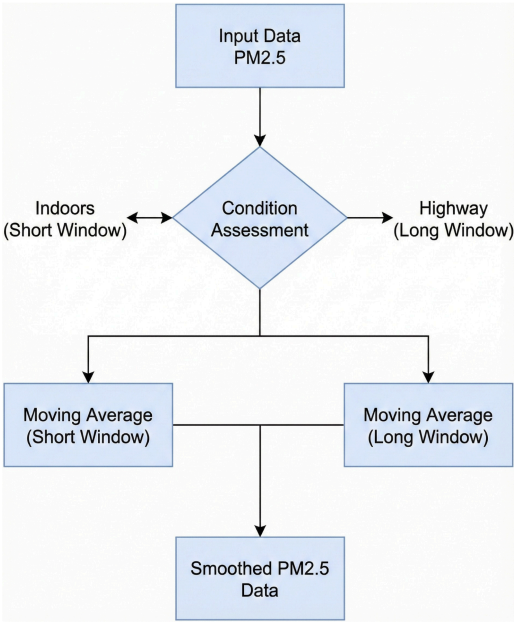


Figure 4. Adaptive filtering architecture demonstrating context-aware switching between short and long smoothing windows.

Short Window Mode: Activated during low-variance or low-speed conditions. This mode prioritizes responsiveness to capture rapid local changes (e.g., emission plumes near traffic).

Long Window Mode: Activated during high-variance or high-speed conditions. This mode applies a broader moving average to suppress transient noise and aerodynamic artifacts.

2.5.2. Sensor Cross-Validation

To verify the accuracy of the hardware–software complex, comparative field tests were conducted against a consumer-grade reference instrument, the Qingping Air Monitor Pro

(Qingping Technology (Beijing) Co., Ltd., Beijing, China). Tests covered both stationary indoor scenarios and dynamic outdoor conditions. As shown in Figure 5, the developed module demonstrates high synchronization with the reference device across key parameters ($PM_{2.5}$, PM_{10}, CO_2). The dynamic response to pollution spikes is nearly identical, confirming the efficacy of the airflow stabilization system.

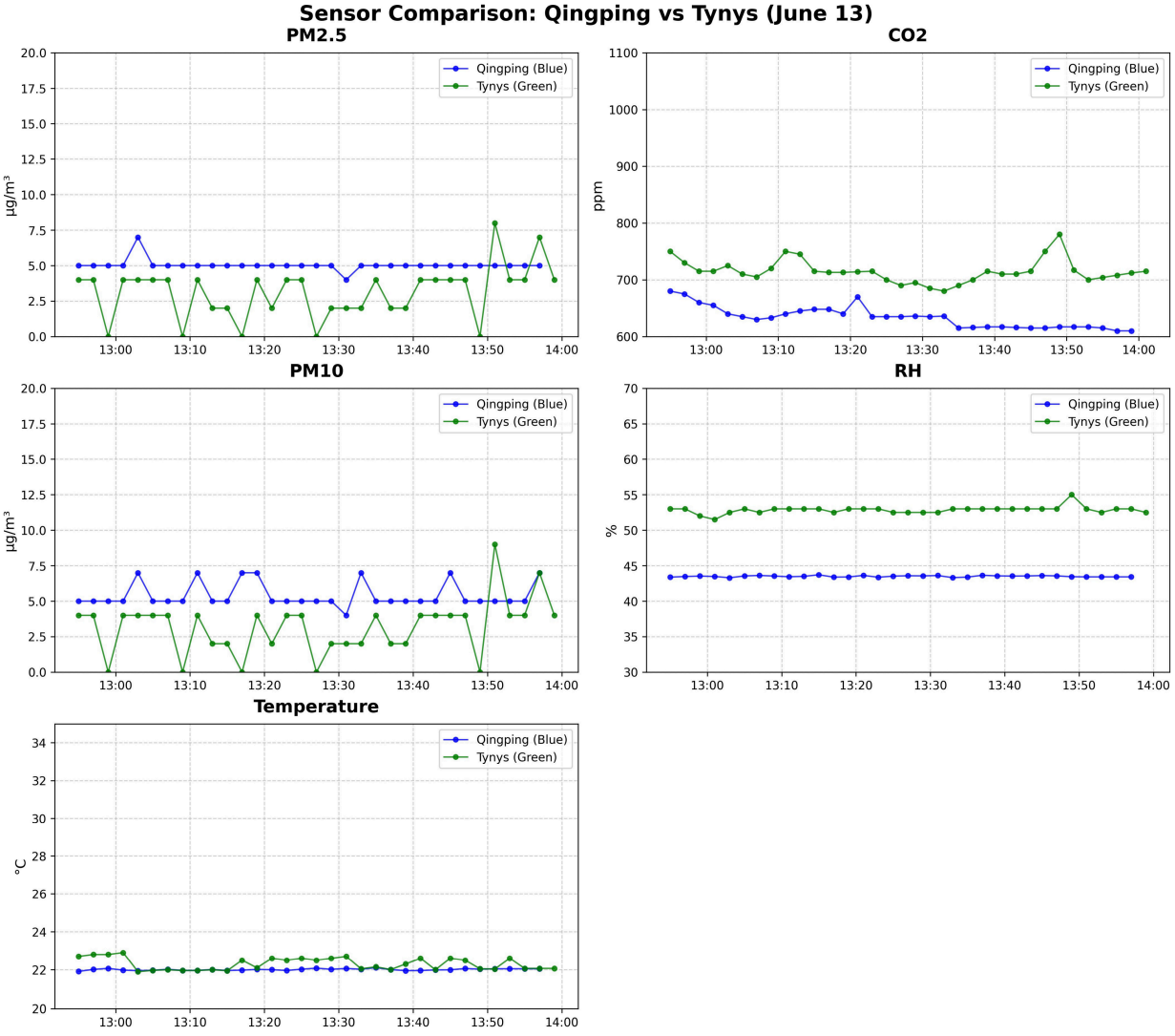


Figure 5. Comparative time-series analysis of $PM_{2.5}$, PM_{10}, and CO_2 recorded by the developed module vs. the Qingping Air Monitor Pro.

It should be noted that the Qingping Air Monitor Pro, while not a certified regulatory-grade instrument, has undergone independent evaluation by the South Coast Air Quality Management District (SCAQMD), where co-located Qingping units demonstrated strong correlations (R^2 = 0.89–0.95 for 1-h $PM_{2.5}$ means) against FEM-grade analyzers (GRIMM EDM180, GRIMM Aerosol Technik, Ainring, Germany; and Teledyne T640, Teledyne API, San Diego, CA, USA) with mean absolute errors of 0.9–2.0 µg/m^3 [27]. A concurrent study deploying ZPHS01B-based sensors for in-cabin air quality monitoring in Almaty's public transport also validated their device against the Qingping monitor and reported consistent inter-device agreement [10]. The comparison in the present study therefore demonstrates inter-device consistency with a sensor that has documented traceability to reference-grade instruments. Additionally, a city-scale cross-reference against official monitoring data from the Kazakhstan national air quality dashboard (http://dashboard.air.

org.kz, accessed on 27 June 2026) is presented in Section 3.8, providing further evidence that the IoT network captures macro-scale pollution dynamics consistent with the established monitoring infrastructure.

To summarize the expected accuracy of the system, three independent lines of evidence are available. First, the ZPHS01B particulate channel has been evaluated against a regulated air quality monitoring station, reporting R^2 ~0.8 for $PM_{2.5}$ after linear calibration [26]. Second, in the present deployment, each node was cross-validated against the Qingping Air Monitor Pro, which has documented traceability to FEM-grade analyzers (GRIMM EDM180, Teledyne T640) with R^2 = 0.89–0.95 and mean absolute errors of 0.9–2.0 $\mu g/m^3$ for 1-h $PM_{2.5}$ means in an independent SCAQMD evaluation [27]. Third, city-scale agreement with the national monitoring dashboard is demonstrated in Section 3.8. On this basis, the system is best characterized as providing internally consistent low-cost observations rather than reference-equivalent concentrations; absolute point-level traceability to a Kazhydromet reference station remains a future activity (Section 4.5).

R^2 #1The measurement uncertainty of the node is taken from the manufacturer specification of the ZPHS01B module [28]—a particulate accuracy of ± 15 $\mu g/m^3$ for concentrations up to 100 $\mu g/m^3$ ($\pm 15\%$ of reading above this value), a temperature accuracy of ± 0.5 °C, and a relative-humidity accuracy of $\pm 3\%$ RH (datasheet test conditions: 25 ± 2 °C, $50 \pm 10\%$ RH, Grimm reference, GB/T 18801-2015 [29]). These tolerances were consistent with a field comparison against the Qingping Air Monitor Pro—a consumer-grade instrument with documented traceability rather than a certified reference—in which the maximum deviations were below 10 $\mu g/m^3$ for $PM_{2.5}$ and PM_{10}, approximately 35 ppm for CO_2, ± 0.5 °C for temperature, and $\pm 2\%$ for relative humidity. As no certified regulatory-grade reference was available for direct co-location, and given the relatively short (~2-month) campaign, the reported concentrations are best interpreted as internally consistent low-cost observations within these uncertainty bounds, suitable for characterizing relative spatial and temporal patterns rather than absolute, reference-equivalent concentrations.

This cross-validation was performed as a baseline accuracy check at the outset of the deployment rather than as a continuous drift-monitoring procedure; long-term sensor drift was therefore not characterized in this pilot campaign (see Section 4.5).

2.5.3. Calibration and Validation Procedure

Because no regulatory-grade reference station was available for direct co-location during the campaign, sensor performance was benchmarked against established external evaluations. Under the South Coast Air Quality Management District (SCAQMD) AQ-SPEC field-testing protocol, three identical Qingping units were co-located with Federal Equivalent Method (FEM) analyzers at the Rubidoux monitoring station (California) for approximately two months in their factory configuration, with quality-control screening, data-recovery assessment, intra-model variability analysis, and comparison against reference instruments under real ambient conditions. The evaluation reported strong $PM_{2.5}$ correlations with the FEM references (R^2 = 0.89–0.95), data recovery close to 100%, and low intra-model variability (≈3.4%). The ZPHS01B used here has likewise been independently evaluated against reference-grade instruments in the recent literature, with acceptable agreement for environmental monitoring. These results support the use of the selected platform while confirming that low-cost sensors complement, rather than replace, regulatory monitoring stations.

Recent work has further advanced low-cost air-quality monitoring through IoT integration, distributed sensor networks, improved calibration methods, and machine-learning-based data correction, with growing attention to field validation against regulatory in-

struments, long-term deployment stability, and mobile sensing for high-resolution spatial coverage; these developments position the proposed system within current research trends [14,18,26,30].

2.6. Backend Architecture and Numerical Model Configuration

2.6.1. Backend Infrastructure and Security

The server-side infrastructure is designed to support the scalability requirements of a Smart City platform. An API Gateway ensures standardized data routing and compatibility with external transport and municipal systems. The security architecture follows the Edge–DMZ–Core model:

1. Edge Level: HTTPS encryption and local caching on the device protect data during transmission over public wireless networks.
2. DMZ (Demilitarized Zone): The API Gateway performs authentication and traffic throttling.
3. Core Level: The PostgreSQL database stores normalized time-series data, accessible only via internal secure channels.

Transmission. Each node transmits over a wireless (Wi-Fi) link. Sensor readings are serialized into JSON packets and published using the MQTT protocol (publish-subscribe model) over TCP, which guarantees delivery of every packet; on transient connectivity loss, the device retains packets in a local store-and-forward buffer and retransmits them once the connection is restored, preventing gaps in the time series. Transport channels are encrypted with TLS, and devices authenticate to the API Gateway through mutual authentication with digital-certificate verification, complemented by IP-range restriction and request rate limiting. The API Gateway, located in the DMZ, performs authentication, access control, and request throttling before any traffic reaches the core, and records all connection attempts and data operations in an event log for audit.

Storage. The Core layer runs PostgreSQL with the TimescaleDB extension for long-term time-series retention, storing normalized records (device_id, timestamp, parameter, value, unit, and a data-quality indicator) that are partitioned by time and indexed on (device_id, timestamp) for efficient querying and horizontal scaling as the network grows. Incoming packets are first validated against a canonical JSON Schema that enforces mandatory fields and admissible value ranges, so out-of-range or malformed values are rejected at the gateway before insertion. The database is reachable only through internal secure channels; no direct external connections are permitted, and all external access is mediated by the API Gateway, with Grafana providing real-time dashboards and visualization.

2.6.2. WRF–SILAM Configuration for Future Integration

While the IoT network provides hyper-local observational data, numerical atmospheric models can serve as a "Virtual Reference System" to validate sensor consistency and fill spatial gaps. We have configured—but not yet operationally integrated—a coupled WRF–SILAM framework as follows.

WRF Setup: The meteorological driver is configured with four nested domains, downscaling from a regional 9 km resolution to a microscale 333 m resolution centered on Almaty. Physics schemes include WSM6 (microphysics), RRTMG (radiation), and the Bougeault–Lacarrère PBL scheme. The Building Effect Parameterization (BEP) module is enabled to simulate the impact of Almaty's dense urban morphology on wind flow and dispersion.

SILAM Setup: Chemical transport is configured on a 1 km grid matching the WRF inner domain. The model incorporates gas-phase chemistry and secondary aerosol formation, utilizing CAMS CB5v2 global inventories supplemented by local emission factors.

Planned Integration Workflow: The backend implements a three-stage integrationlogic: (1) Validation—sensor data is cross-referenced against model baselines to derive correction factors; (2) Hotspot Identification—measurements aggregated into a 100 × 100 m spatial grid are compared against the SILAM regional background; (3) Data Fusion—normalized sensor data is used to adjust emission inventories fed into SILAM, creating a closed-loop system.

2.7. Field Campaign and Data Description

2.7.1. Field Campaign

A monitoring campaign was conducted in Almaty, Kazakhstan, from 4 February to 3 April 2025, spanning the transition from the peak heating season (February) through the spring warming period (March–April). Six IoT sensor nodes (Devices 1–6) were rotated among approximately 20 measurement sites distributed across the city (Figure 6), covering a spatial extent of ~13 km (north–south) × ~20 km (east–west). Deployment sites span five urban zones: the southern foothills at the base of the Ile Alatau range, the central commercial district (Almaly/Almalinsky), the southwestern residential corridor (Moskva/Saryarka), the northeast toward Guldala, and the northern area near the railway station. Due to the rotation schedule, individual devices were not permanently assigned to fixed locations. The measurement–location linkage was not preserved in the database during this pilot campaign, which prevents device-level spatial attribution of pollutant concentrations.

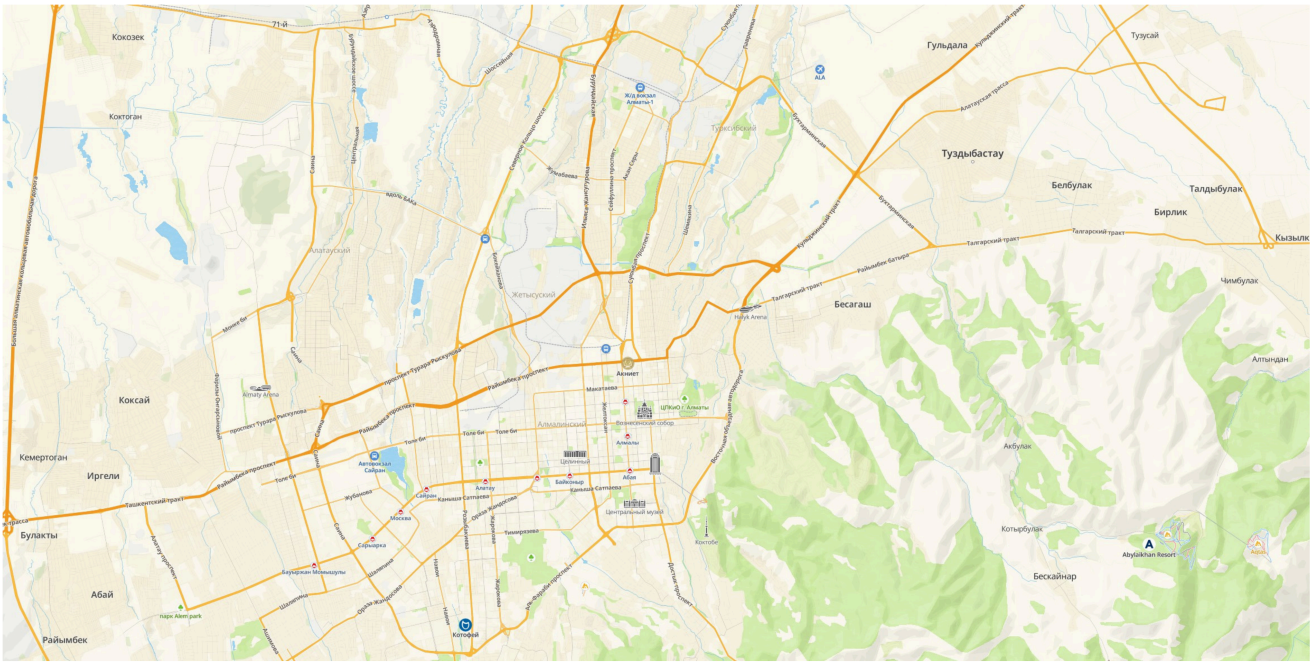


Figure 6. Spatial distribution of 20 measurement sites across Almaty where the six IoT sensor nodes were deployed on a rotating basis during the February–April 2025 campaign. Sites span approximately 13 km (N–S) × 20 km (E–W), from the Ile Alatau foothills in the south to the railway station area in the north. Place names on the map are rendered in Russian/Kazakh as provided by the source map service and are not essential to interpreting the spatial pattern shown.

The campaign yielded a total of 70,772 individual sensor readings across all parameters.

Table 2 summarizes the deployment characteristics of each node. Devices 1 and 2 provided the most continuous data streams, with 1170 and 1160 unique measurement timestamps over 22 days of active operation, respectively. Devices 3 and 4 operated intermittently with long gaps (up to 26 days), while Devices 5 and 6 were brought online in late March and contributed data for 6–7 days.

Table 2. Summary of sensor node deployment during the field campaign.

Device	Active Period	Active Days	Unique Timestamps	Mean $PM_{2.5}$	Sensors
1	4 February–17 March	22	1170	46.9	PM, CO_2, T, RH, TVOC
2	4 February–12 March	22	1160	43.8	PM, CO_2, T, RH, TVOC
3	4 February–1 April	3	18	24.7	PM, CO_2, T, RH
4	4 February–1 April	7	74	28.2	PM, CO_2, T, RH
5	19 March–3 April	7	259	17.8	PM, CO_2, T, RH, Noise
6	28 March–3 April	6	323	23.1	PM, CO_2, T, RH, Noise

2.7.2. Measured Parameters

The network recorded six environmental parameters: $PM_{2.5}$ and PM_{10} (particulate matter concentrations, $\mu g/m^3$), CO_2 (carbon dioxide, ppm), temperature (C), relative humidity (%), and TVOC (total volatile organic compounds, select devices). Devices 5 and 6 additionally recorded ambient noise levels (dB). We note that while the Winsen ZPHS01B module includes electrochemical cells for gaseous pollutants (NO_2, CO, SO_2, O_3), these channels were not activated during this deployment due to firmware constraints and are planned for inclusion in future campaigns.

2.7.3. Data Quality Control

Raw data were subjected to the following quality control procedures: (1) checksum based packet validation at the edge level; (2) removal of physically implausible values ($PM_{2.5} < 0$ or $>1000\ \mu g/m^3$; temperature < -40 or $>60\ ^\circ C$); (3) aggregation of burst readings (devices 3 and 4 transmitted multiple readings per timestamp) to per-timestamp means; and (4) application of the adaptive filtering algorithm described in Section 2.5.1. After quality control, the analytical dataset comprised 3004 unique device–timestamp observations for each of the five primary parameters.

2.8. Comparative Summary

Table 3 provides a comparative overview of the present system against recent low-cost air quality monitoring deployments reported in the literature. The comparison highlights key design parameters including sensor hardware, measured pollutants, deployment context, validation approach, and cost. While several systems achieve high accuracy for specific pollutants, the present work is distinguished by its integration of edge-adaptive filtering, a multi-node distributed architecture, and a prepared coupling with numerical atmospheric models—specifically targeting outdoor ambient monitoring in a Central Asian city with complex basin topography.

Each node is built from commodity components. Its two principal components—a Winsen ZPHS01B multi-parameter module and a Raspberry Pi Zero 2 W edge computer—cost approximately US$100 per node at market prices (Table 4), while auxiliary items (a custom enclosure with intake fan, microSD storage, power supply, and wiring) add only a small further amount and connectivity is provided by the Raspberry Pi on-board Wi-Fi. For context, regulatory-grade FRM/FEM reference monitors such as the Teledyne API T640, the Met One BAM-1020 (Met One Instruments, Grants Pass, OR, USA), and the GRIMM EDM180 typically cost on the order of US$15,000–40,000 per unit, with annual

maintenance frequently exceeding US$15,000 [31,32] [Clarity 2026; Delp and Singer 2020], whereas research/professional photometers cost approximately US$5000–10,000 [Delp and Singer 2020]. The per-node cost of the present system is therefore roughly two to three orders of magnitude lower than that of a regulatory station, which is precisely what makes a distributed, multi-node deployment at city scale economically feasible. As detailed in Section 4.1, this cost advantage is necessarily balanced against trade-offs in absolute accuracy and the absence of regulatory certification.

Table 3. Comparison of the proposed system with recent low-cost air quality monitoring deployments.

Study	Sensor	Parameters	Platform	Validation	Duration	Cost/Node
SATVAM [15]	Custom IoT	$PM_{2.5}$	Fixed + Mobile	Field collocation	Long-term	Low
Hyderabad [16]	Mixed (49 nodes)	PM	Fixed	Seasonal recalib.	12+ months	Low
ElSalvador [17]	PMS5003ST + ESP32	$PM_{1.0/2.5/10}$	Fixed	CAQI thresholds	Months	<$50
Enviro-IoT [18]	Multisensor	PM, NO_2	Fixed	Ref. station	9 months	Moderate
Tynys [10]	ZPHS01B + RPi	PM, CO_2, T, RH	In-cabin mobile	Qingping + SCAQMD	1 month	~$80
Whitehill et al. [14]	Mobile AQ instruments	PM, gaseous pollutants	Mobile + fixed	Comparison with regulatory stations	Campaign-based	Moderate
Meneses-Albala et al. [26]	ZPHS01B multisensor node	$PM_{2.5}$, CO_2, TVOC, T, RH	Fixed	Reference instrument comparison	Field deployment	~$70–100
Balagopal et al. [29]	LoRaWAN low-cost sensor	$PM_{2.5}$, PM_{10}	Distributed fixed	Palas Fidas Frog reference	Field deployment	Low
This work	ZPHS01B + RPi	PM, CO_2, T, RH	Distributed fixed	Qingping (SCAQMD eval.) + ZPHS01B lit.	2 months	~$100

The per-node cost of the present system reflects market component prices (Table 4); reference-grade FRM/FEM instruments are sold by quotation.

Table 4. Principal hardware components and indicative per-node cost (market prices).

Component	Qty	Unit Cost (US$)	Notes
Winsen ZPHS01B module	1	~84	retail: AliExpress ~$84/Winsen ~$107
Raspberry Pi Zero 2 W	1	~15	
Total per node		~US$100	

Indicative market/retail prices for the two principal components (multi-parameter sensor module and edge computer). Connectivity is provided by the Raspberry Pi on-board Wi-Fi; auxiliary items (enclosure, intake fan, microSD, power supply, wiring) add only a small additional cost. Components were procured domestically through tender at higher unit prices owing to import duties and margins.

3. Results

3.1. Descriptive Statistics

Table 5 presents the descriptive statistics for all measured parameters across the entire campaign.

Table 5. Descriptive statistics of measured parameters across the campaign (February–April 2025). All PM values in $\mu g/m^3$; CO_2 in ppm; temperature in °C; RH in %.

Parameter	n	Mean	SD	Min
$PM_{2.5}$	3004	40.1	37.4	3.0
PM_{10}	3004	33.4	32.1	2.0
CO_2	3004	636.9	281.3	400.0
Temperature	3004	24.3	4.8	−6.4
Relative Humidity	3004	23.9	10.7	10.3

The campaign-average $PM_{2.5}$ concentration of 40.1 $\mu g/m^3$ exceeds the WHO 24 h guideline (15 $\mu g/m^3$) by nearly three times. Daily mean $PM_{2.5}$ exceeded the WHO 24 h guideline on 32 out of 38 monitored days (84%), and exceeded 25 $\mu g/m^3$ on 23 days (61%). For PM_{10}, daily means exceeded the WHO 24 h guideline (45 $\mu g/m^3$) on 9 days (24%). These results underscore the severity of particulate pollution in Almaty's winter–spring period.

We note that the elevated median temperature (24.7 °C) reflects the fact that some sensors were deployed in semi-enclosed environments (e.g., bus shelters), resulting in temperatures above ambient outdoor levels.

3.2. Seasonal Transition: Heating Season vs. Spring

A central finding of this study is the dramatic decline in PM concentrations coinciding with the end of the heating season. Figure 7 shows the hourly mean $PM_{2.5}$ time series across all devices. February exhibited significantly higher concentrations (mean 54.4 ± 43.6 $\mu g/m^3$; $n = 1665$) compared to March (21.9 ± 14.5 $\mu g/m^3$; $n = 1050$) and April (23.6 ± 12.5 $\mu g/m^3$; $n = 289$). A Welch's t-test confirmed the February–March difference as highly significant ($t = 23.36$, $p < 0.001$). The monthly distributions are shown in Figure 8.

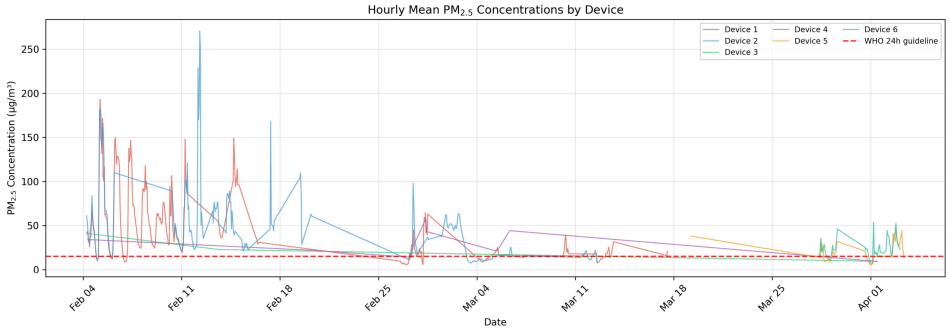


Figure 7. Hourly mean $PM_{2.5}$ concentrations recorded by individual devices across the campaign period. The dashed red line indicates the WHO 24 h guideline (15 $\mu g/m^3$).

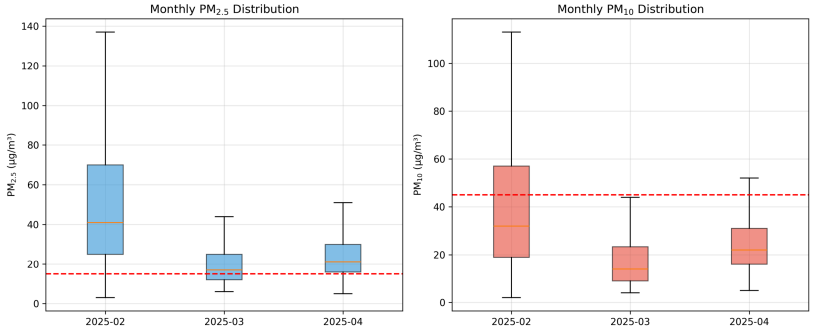


Figure 8. Monthly distributions of $PM_{2.5}$ (**left**) and PM_{10} (**right**) concentrations. Red dashed lines indicate WHO 24 h guidelines. Outliers are suppressed for clarity.

This pattern is consistent with the well-documented influence of coal-based residential and centralized heating, which is the primary PM source in Almaty during the cold season. The approximate 60% reduction from February to March aligns with the gradual cessation of household heating as ambient temperatures rise.

3.3. Diurnal Patterns

Figure 9 presents the mean diurnal profiles for $PM_{2.5}$ and PM_{10}. Both parameters exhibit a pronounced bimodal pattern with an early-morning peak between 04:00 and 07:00 ($PM_{2.5}$: 55–64 $\mu g/m^3$) and a secondary, less pronounced afternoon elevation around 13:00–14:00. The lowest concentrations occur between 00:00 and 03:00 (~22 $\mu g/m^3$).

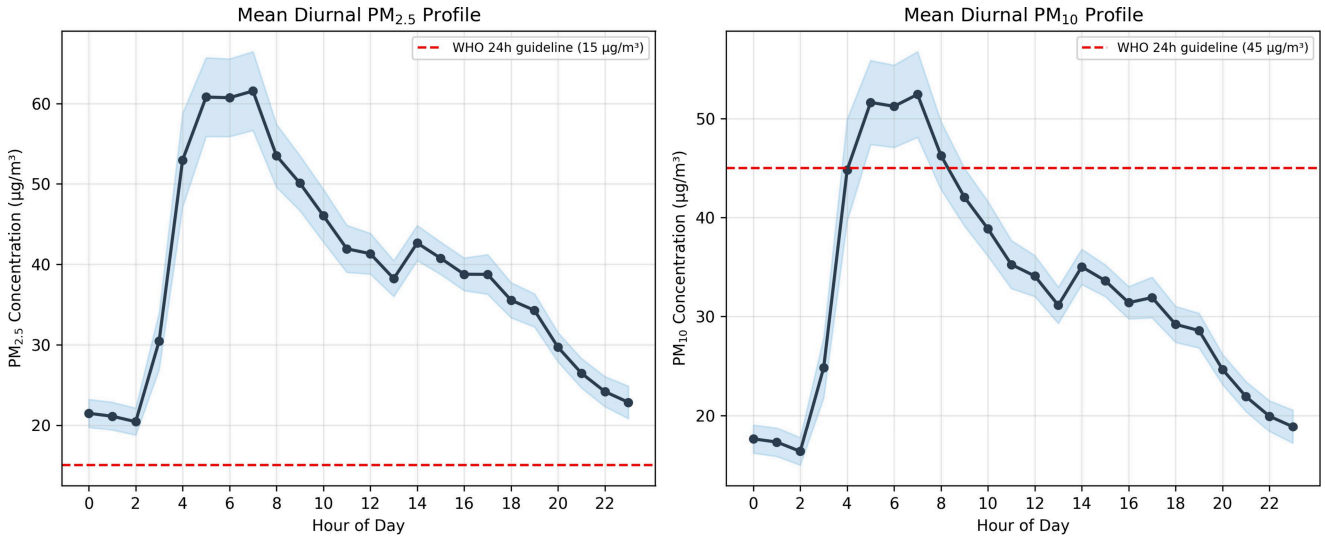


Figure 9. Mean diurnal profiles of $PM_{2.5}$ (**left**) and PM_{10} (**right**) with standard error bands. The black line with markers indicates the hourly mean concentration, and the blue shaded band indicates ± 1 standard error of the mean. Red dashed lines indicate WHO 24 h guidelines (15 $\mu g/m^3$ for $PM_{2.5}$; 45 $\mu g/m^3$ for PM_{10}).

The early-morning peak is attributable to two coinciding factors: (1) increased residential coal combustion for heating during the coldest hours; and (2) the typical nocturnal formation and morning persistence of surface temperature inversions in Almaty's basin, which suppress vertical mixing and trap pollutants near the surface. The secondary afternoon peak likely reflects traffic emissions during commuting hours.

3.4. Weekday vs. Weekend Variations

Analysis of the weekday–weekend split reveals a statistically significant difference in $PM_{2.5}$ concentrations (Figure 10). Counterintuitively, weekend concentrations (mean 43.5 $\mu g/m^3$) were slightly higher than weekday levels (39.3 $\mu g/m^3$; Welch's $t = -2.43$, $p = 0.015$). This pattern suggests that in Almaty's winter context, the dominant PM source is residential heating (which operates continuously and may increase on weekends with greater domestic occupancy) rather than traffic, which typically decreases on weekends. Because individual readings are not statistically independent (temporal autocorrelation and device-specific effects), the *t*-tests reported in this section were re-examined on daily means to provide more conservative significance estimates; these daily aggregated results, and their sensitivity to device-level data quality, are reported in Section 4 (Limitations).

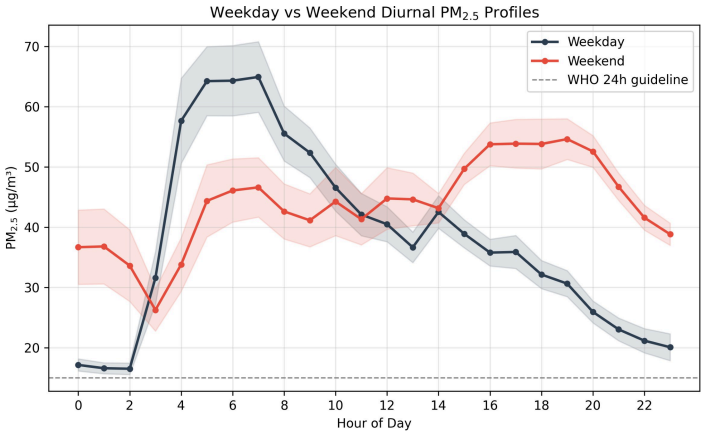


Figure 10. Weekday (black) vs. weekend (red) diurnal $PM_{2.5}$ profiles with standard error bands; shaded regions indicate ±1 standard error of the mean for each hour. The dashed horizontal line marks the WHO 24-hour $PM_{2.5}$ guideline.

3.5. Inter-Parameter Correlations

Table 6 presents the Pearson correlation matrix for all measured parameters. $PM_{2.5}$ and PM_{10} are very strongly correlated (r = 0.995), indicating a common source and consistent particle size distribution. Temperature shows a moderate negative correlation with both PM parameters (r = −0.31), consistent with the seasonal heating effect: colder periods are associated with higher combustion emissions and stronger inversions. This correlation, however, is dominated by two nodes whose PM_{10} field is physically inconsistent ($PM_{2.5}$ > PM_{10} in every reading; see Section 4, item 9); the PM_{10} channel of these units was therefore excluded from quantitative analysis, and the $PM_{2.5}/PM_{10}$ ratio and $PM_{2.5}$–PM_{10} correlation are reported only for the physically consistent nodes (3–6, ratio ≈ 0.96). $PM_{2.5}$, the manufacturer-calibrated channel, is retained for all nodes.

Table 6. Pearson correlation matrix of measured parameters (n = 2973). Note: the $PM_{2.5}$–PM_{10} correlation is influenced by the PM_{10} channel-assignment fault on two nodes (Section 4, item 9); the $PM_{2.5}/PM_{10}$ ratio is reported in the text for the physically consistent nodes (3–6) only.

Title 1	$PM_{2.5}$	PM_{10}	CO_2	T	RH
$PM_{2.5}$	1.000	0.995	0.082	−0.312	0.106
PM_{10}		1.000	0.104	−0.307	0.155
CO_2			1.000	0.051	0.348
Temperature				1.000	−0.622
RH					1.000

CO_2 shows only a weak positive correlation with PM (r = 0.08), suggesting that while both are linked to combustion, the spatial and temporal dynamics of CO_2 (strongly influenced by indoor/ventilation conditions) differ from those of particulate matter. The strong negative correlation between temperature and humidity (r = −0.62) reflects the expected inverse relationship during the cold season.

3.6. Extreme Pollution Episodes

The highest daily mean $PM_{2.5}$ concentration (109.2 µg/m^3) was recorded on 12 February 2025, exceeding the WHO guideline by a factor of 7.3. At a physically consistent node, the peak individual reading reached 139 µg/m^3 ($PM_{2.5}$) and a simultaneous 142 µg/m^3 (PM_{10}) (Node 4, 6 March 2025); the still higher peaks recorded by the two dominant nodes are not reported here, as their PM_{10} channel was excluded from quantitative analysis (Section 4, item 9). Such extreme episodes are likely associated with the coincidence of

strong surface inversions, low temperatures, and calm wind conditions, a combination characteristic of Almaty's winter meteorology as documented in prior WRF simulations [21].

Other notably polluted days include 5 February (78.1 $\mu g/m^3$), 14 February (79.3 $\mu g/m^3$), and 19 February (86.5 $\mu g/m^3$), all occurring during the peak heating season. In contrast, all days in March and April had daily mean $PM_{2.5}$ below 50 $\mu g/m^3$.

3.7. $PM_{2.5}/PM_{10}$ Ratio and Source Attribution

The ratio of $PM_{2.5}$ to PM_{10} is a widely used indicator for inferring the dominant source type: ratios above ~0.7–0.8 suggest combustion-dominated aerosol (fine particles from burning), while lower ratios indicate a greater contribution from mechanical sources such as road dust or construction [1].

An important data quality observation was identified during this analysis: Devices 1 and 2 consistently reported $PM_{2.5}$ values exceeding PM_{10} (ratio > 1.0 in 100% of readings), which is physically impossible since $PM_{2.5}$ is a subset of PM_{10}. This artifact is a known characteristic of certain low-cost laser-scattering sensors at high concentrations and has been documented in prior studies of similar sensor modules [26]. The $PM_{2.5}/PM_{10}$ ratio analysis was therefore restricted to Devices 3–6, which exhibited physically consistent readings (n = 674). While this artifact does not invalidate the temporal trend analysis (both $PM_{2.5}$ and PM_{10} from Devices 1–2 track the same pollution dynamics), it does preclude their use for quantitative ratio-based source attribution.

For Devices 3–6, the campaign-average $PM_{2.5}/PM_{10}$ ratio was 0.963 ± 0.037 (Figure 11), indicating overwhelming dominance of fine particles characteristic of combustion sources. The ratio remained consistently high across all months: February (0.974 ± 0.012), March (0.963 ± 0.041), and April (0.961 ± 0.039). The slight decrease from February to April, while statistically modest, may reflect the gradual emergence of coarser dust particles as snow cover recedes in spring.

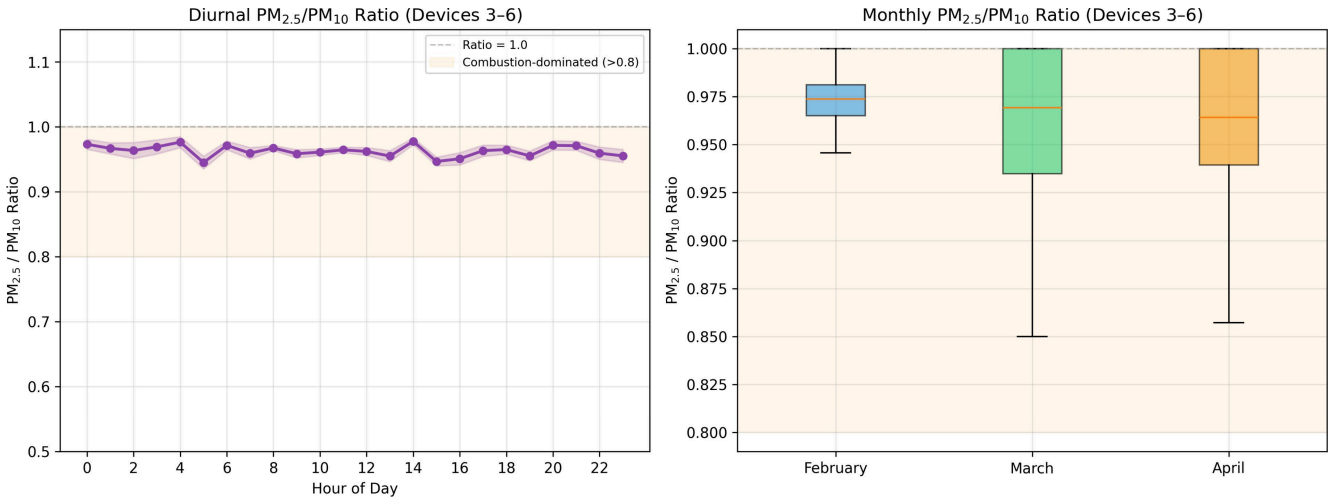


Figure 11. $PM_{2.5}/PM_{10}$ ratio analysis for Devices 3–6: (**left**) diurnal profile with standard error bands, where the purple shaded region indicates ±1 standard error of the mean $PM_{2.5}/PM_{10}$ ratio at each hour; (**right**) monthly distribution shown as box plots, where blue, green, and orange box colors denote February, March, and April, respectively, the red horizontal line within each box indicates the median, and the box edges represent the interquartile range (IQR) with whiskers showing the full data range. In both panels, the orange shaded region indicates the combustion-dominated regime (ratio > 0.8), and the dashed line marks the reference ratio of 1.0. Data from Devices 1–2 are excluded due to a systematic $PM_{2.5} > PM_{10}$ sensor artifact.

At the physically consistent nodes (3–6), the $PM_{2.5}/PM_{10}$ ratio (≈0.96) is high and dominated by the fine fraction, which is consistent with combustion—most plausibly coal-

based residential and centralized heating—being an important source of particulate matter in Almaty's winter–spring atmosphere. We treat this ratio as exploratory and present the attribution as the most likely interpretation supported by indirect evidence rather than as a fully demonstrated result, both because the PM_{10} channel of the two dominant nodes was excluded as unreliable (Section 4, item 9), which limits the spatial and temporal coverage of the ratio, and because the dataset does not include chemical speciation, black carbon, gaseous source markers, or meteorological normalization. It is nonetheless mutually consistent with the $PM_{2.5}$ seasonal trend analysis (Section 3.2) and the weekday–weekend pattern (Section 3.4), which point to the same source through independent indirect lines of evidence.

3.8. Cross-Reference with Official Monitoring Data

To assess the consistency of the IoT network measurements with independent data, daily mean $PM_{2.5}$ concentrations were compared against city-wide official monitoring data published on the Kazakhstan national air quality dashboard (http://dashboard.air.org.kz, accessed on 27 June 2026 [33]). The official platform reports daily $PM_{2.5}$ concentrations as categorical bands (0–15, 15–35, 35–70, 70–140 $\mu g/m^3$) based on aggregated data from the national monitoring network.

Figure 12 presents the comparison for all 38 days with IoT data. The IoT network daily means fell within the corresponding official band on 22 out of 38 days (58%). Of the 16 days outside the official band, 12 had IoT values below the band lower bound and 4 above. The below-band cases predominantly occurred in early March (when Devices 5–6, deployed at potentially less polluted sites, contributed a larger share of readings), while the above-band cases corresponded to localized pollution episodes (e.g., 12 February: IoT = 109.2 $\mu g/m^3$ vs. official band 35–70 $\mu g/m^3$).

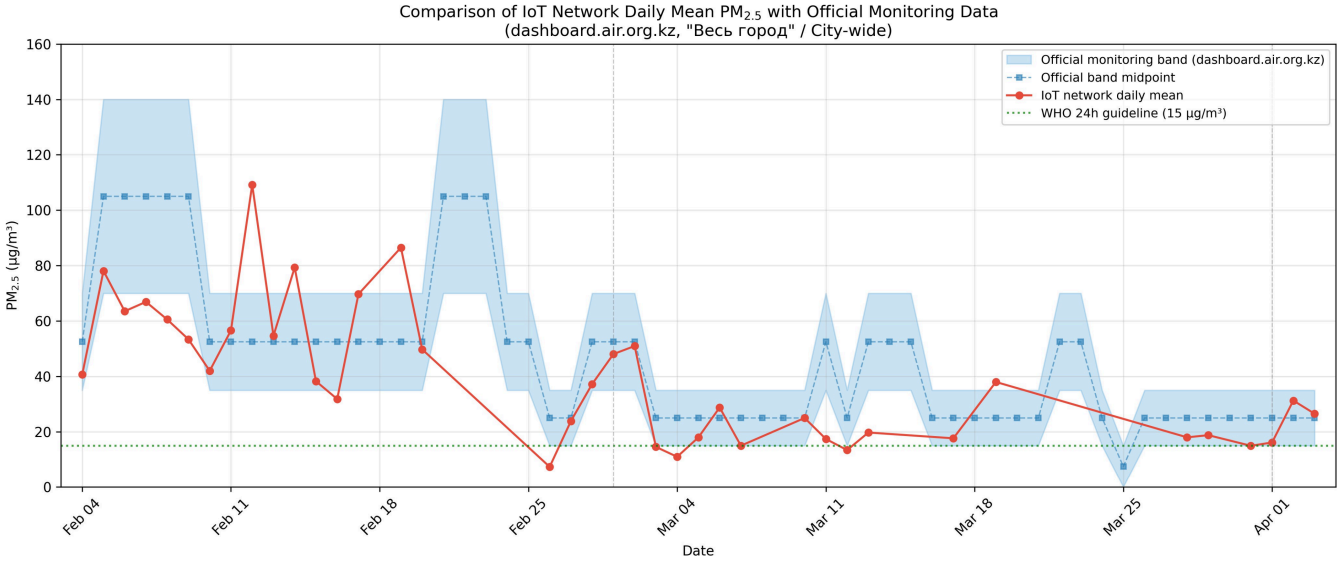


Figure 12. Comparison of IoT network daily means $PM_{2.5}$ with official city-wide monitoring data from http://dashboard.air.org.kz (accessed on 27 June 2026) [33]. The shaded blue band represents the officially reported $PM_{2.5}$ range for each day. Red circles show IoT network daily means. The WHO 24 h guideline (15 $\mu g/m^3$) is shown for reference.

Critically, both datasets exhibit the same seasonal trend: official data show a February-to-March reduction of 53% (band midpoint: 67.1 → 31.5 $\mu g/m^3$), compared to 60% in the IoT data (55.2 → 23.1 $\mu g/m^3$). The systematic tendency of the IoT network to report slightly lower values than the official band midpoint is expected, as the official data represents city-wide aggregates that may be weighted toward higher-pollution monitoring stations,

while the IoT network includes sites across diverse zones, including less polluted areas. Overall, the two datasets show broad seasonal consistency rather than quantitative point-by-point agreement: the official dashboard reports only city-wide categorical bands, so a station-level breakdown of the contributing reference stations is not available to us, and the IoT daily means fall within the official band on 22 of 38 days (58%). The network should therefore be read as reproducing the macro-scale seasonal pattern of the established monitoring infrastructure rather than as providing a validated quantitative match to it.

3.9. Network Spatial Coverage

Although the rotating deployment prevented device-level spatial attribution of pollutant concentrations, the measurement site map (Figure 6) shows that the rotating deployment reached sites distributed across the Almaty urban area, supporting the feasibility of city-scale network deployment as a pilot rather than a spatially resolved pollution assessment. The 20 sites span the key pollution-relevant zones identified in prior WRF modeling studies [21]: the southern foothills where katabatic flows and inversions concentrate pollutants, the central city with mixed commercial and residential emissions, and the northern/southwestern residential corridors. This spatial coverage establishes the foundation for future campaigns with continuous GPS logging, which will enable the resolution of intra-urban pollution gradients that are invisible to the sparse Kazhydromet station network (currently ~5 stations for a city of 2 million inhabitants).

4. Discussion

4.1. System Performance and Hardware Reliability

Field deployments in Almaty's winter conditions (temperatures down to -6 °C in the dataset, with ambient temperatures reaching -20 °C) served as a stress test for the hardware design. The custom enclosure with Active Airflow Stabilization proved critical: comparative analysis during development showed that sensors without active ventilation exhibited a 15–20% higher noise floor. The implemented fan system successfully maintained a stable flow rate, decoupling the sensing element from external aerodynamic disturbances.

The Raspberry Pi Zero 2 W demonstrated high stability as an edge-computing node. The "Edge-First" processing approach—where filtering occurs on the device—reduced data transmission volume by approximately 40% compared to raw streaming, significantly lowering wireless bandwidth costs. However, as discussed in Section 4.5, the intermittent operation of devices 3 and 4 highlights remaining challenges with power management and wireless connectivity in field conditions.

These advantages are balanced by several disadvantages inherent to the low-cost approach: (i) the absence of absolute metrological traceability to a reference-grade station; (ii) the $PM_{2.5} > PM_{10}$ artifact observed in two nodes at high concentrations, which precludes their use for ratio-based source attribution; (iii) device downtime caused by power and wireless-connectivity constraints, which reduced the effective network during the core campaign; and (iv) the absence of activated gas-phase channels in this deployment. We state these explicitly so that the cost advantage is interpreted against the corresponding trade-offs in accuracy and completeness rather than in isolation.

4.2. Adaptive Filtering Performance

The context-aware adaptive filtering algorithm proved effective in distinguishing between genuine pollution events and sensor noise. In Short Window mode, the system preserved responsiveness to localized changes, while Long Window mode produced smooth baselines. The combination of both modes ensures that the reported data retains both high temporal fidelity during stable conditions and robustness against transient artifacts.

4.3. Environmental Insights

The results provide several policy-relevant findings for Almaty. First, the dominance of the heating season on PM concentrations—with a 60% reduction from February to March—indicates that residential and centralized coal heating is the most likely dominant controllable PM source; we frame this as the strongest interpretation supported by the available indirect evidence rather than as a directly apportioned result. This interpretation is independently corroborated by the $PM_{2.5}/PM_{10}$ ratio analysis (Section 3.7), which shows a persistently high ratio (~0.96) indicative of combustion-dominated aerosol, with no evidence of significant coarse-particle (mechanical/dust) contributions even as the season transitions to spring. Targeted interventions such as accelerated gasification of residential areas or subsidies for clean heating alternatives could yield the greatest air quality improvements.

Second, the counterintuitive weekday–weekend pattern (higher PM on weekends) further supports the hypothesis that residential heating, rather than traffic, is the dominant contributor during winter. This finding has implications for emission inventory construction, where traffic is often assumed to be the primary PM source in urban areas.

Third, the extreme pollution episodes (daily means exceeding 100 $\mu g/m^3$) coincide with conditions known to produce strong inversions in Almaty's basin. These events represent acute public health risks and warrant real-time warning systems—a function that the proposed IoT network, once coupled with the WRF–SILAM framework (Section 2.6.2), could support.

Fourth, the network's rotating deployment across 20 sites demonstrates the feasibility of a city-scale pilot using a small number of devices, rather than a spatially resolved assessment. Although the current campaign did not preserve measurement–location linkages, the spatial distribution of sites (Figure 6) covers the key pollution-relevant zones of Almaty. Future campaigns with continuous GPS logging will enable the quantification of intra-urban pollution gradients, particularly the expected south-to-north decrease associated with distance from the Ile Alatau foothills and their inversion-trapping effects.

4.4. Comparison with Previous Studies

The $PM_{2.5}$ concentrations recorded in this study are consistent with prior estimates for Almaty. Previous research reported annual mean $PM_{2.5}$ exceeding WHO guidelines by a factor of 2–3 [6], while our winter campaign captured peak-season concentrations 3–7 times above the WHO 24 h guideline. The cross-reference with official monitoring data (Section 3.8) further confirms this consistency: both the IoT network and the national monitoring dashboard show a ~53–60% reduction in $PM_{2.5}$ from February to March, with the IoT daily means falling within the official band on 58% of days. The diurnal pattern with early-morning peaks is consistent with WRF simulations of temperature inversions in the Almaty basin [21]. The $PM_{2.5}/PM_{10}$ ratio close to unity ($r = 0.995$) suggests a dominance of fine particles from combustion sources, consistent with coal-heating emissions.

It is instructive to compare our outdoor ambient results with the concurrent in-cabin study by Tasmurzayev et al. [10], which deployed ZPHS01B-based sensors inside Almaty's buses, trolleybuses, and metro during April 2025. Their in-cabin $PM_{2.5}$ concentrations (mean 20.4 $\mu g/m^3$, max 92 $\mu g/m^3$) were notably lower than our outdoor February values (mean 54.4 $\mu g/m^3$, max 290 $\mu g/m^3$) but comparable to our March–April readings (mean ~22 $\mu g/m^3$), which is expected given the overlapping spring measurement period. Their finding that CO_2 strongly correlated with passenger occupancy (>2800 ppm during peak hours) contrasts with our outdoor data where CO_2 showed only weak correlation with PM ($r = 0.08$), confirming that outdoor and indoor pollutant dynamics are governed by fundamentally different source–receptor processes. Together, the two studies provide com-

plementary perspectives on Almaty residents' total pollution exposure: ambient sources (characterized in this work) and in-transit microenvironment exposure [10].

4.5. Limitations

Several limitations should be acknowledged:

1. Sensor calibration: The device-level cross-validation was performed against the Qingping Air Monitor Pro, which has demonstrated R^2 = 0.89–0.95 against FEM-grade instruments in an independent SCAQMD evaluation [27]. City-scale consistency was verified against official monitoring data (Section 3.8), confirming agreement in seasonal trends and overall magnitude. However, direct point-level collocation with a Kazhydromet regulatory station—which would establish absolute metrological traceability and quantify site-specific bias—remains to be completed in future campaigns. In addition, repeated in-field collocation to quantify long-term sensor drift and endurance was not performed during this pilot campaign; periodic re-collocation to characterize drift is planned for future deployments. Two indirect indicators nonetheless support operational stability over the campaign: edge-level checksum validation excluded transmission-induced errors, and the seasonal February-March decline captured by the network is consistent in magnitude and timing with the independent official record (Section 3.8), which would not be expected under large uncorrected baseline drift.
2. Network coverage: Of the six deployed nodes, only Devices 1 and 2 provided sustained, continuous data. Devices 3 and 4 suffered from connectivity and power issues, reducing the effective network to two reliable nodes during the core campaign period. This limits the spatial representativeness of the findings.
3. Temporal gaps: Significant data gaps exist, including a near-complete loss of data during 16–26 February. The 38 days of actual data from a 58-day calendar period represent 66% temporal coverage.
4. Missing gas-phase pollutants: The current dataset does not include NO_2, CO, SO_2, or O_3 measurements despite the sensor module's hardware capability. Activation of these channels is planned for subsequent deployments.
5. Spatial attribution: Although the six devices were rotated among 20 sites across Almaty, the measurement–location linkage was not recorded in the database during this pilot campaign. This prevents spatial mapping of pollutant concentrations to specific locations. Future campaigns will implement continuous GPS logging and database-level geolocation tagging to enable route-level spatial analysis and hotspot identification.
6. Model integration: The WRF–SILAM framework is configured but not yet operationally coupled with the observational data. The planned model-sensor integration described in Section 2.6.2 remains a future milestone.
7. Season and duration: The campaign covers only the winter–spring transition. Year-round monitoring is needed to capture summer dust events, autumn agricultural burning, and inter-annual variability.
8. Statistical independence: The Welch's *t*-tests and Pearson correlations were computed on hourly and device-level measurements, which are not statistically independent in the way randomly sampled observations are, because of temporal autocorrelation and device-specific effects. The reported *p*-values and correlation coefficients should therefore be interpreted with caution; aggregating to daily means and, where feasible, mixed-effects models with device and time as factors would provide more conservative significance estimates and are planned for future analyses. The campaign log comprises 70,772 individual sensor-parameter records; after consolidating repeated sub-minute entries sharing an identical timestamp (97.8% of which carry identical values), these correspond to approximately 3004 unique multi-parameter observa-

tions. To avoid overstating the effective sample size, the inferential statistics reported above were re-examined on daily means, on which the February–March $PM_{2.5}$ contrast remains significant (55.2 vs. 23.1 $\mu g/m^3$; Welch $t = 5.12$, $p < 0.001$).

9. PM_{10} channel-assignment fault on two nodes, and the resulting analysis strategy: Two nodes (Node_1_Center and Node_2_AlFarabi), which together contribute 2330 of 3004 paired particulate records, returned $PM_{2.5} > PM_{10}$ in 100% of their observations (mean $PM_{10}/PM_{2.5} \approx 0.77$). Because $PM_{2.5}$ is by definition a sub-fraction of PM_{10}, this is physically impossible and identifies a fault in the PM_{10} field of these two units rather than random noise: the PM_{10} and $PM_{2.5}$ channels are almost perfectly linearly related (Pearson r $\approx$ 1.0). The Winsen ZPHS01B derives PM1.0, $PM_{2.5}$ and PM_{10} from a single laser scattering sensor and reports them at distinct positions of the UART frame, and the manufacturer specifies a calibrated accuracy only for $PM_{2.5}$ ($\pm$15 $\mu g/m^3$); no independent accuracy is specified for PM_{10}. The observed $PM_{10}/PM_{2.5}$ ratio of $\approx$0.77 on these two units lies in the range expected for $PM1.0/PM_{2.5}$ in combustion-dominated fine aerosol ($\approx$0.7–0.9), whereas the physically consistent nodes (3–6) return $PM_{10}/PM_{2.5} \approx 1.04$; the most parsimonious explanation is therefore that the field logged as PM_{10} on these two units corresponds to the PM1.0 output (a frame-position/assignment error), not true PM_{10}. Because PM1.0 and PM_{10} do not share a fixed physical ratio, no defensible scalar correction can reconstruct true PM_{10} for these nodes, and no co-located reference PM_{10} series is available for them; we therefore did not attempt a numerical recalibration, as any such factor would be unverifiable. Instead, PM_{10} from these two nodes was excluded from all quantitative analysis. $PM_{2.5}$—the single manufacturer-calibrated channel and the health-relevant metric—was retained for all nodes; the $PM_{2.5}/PM_{10}$ ratio and the $PM_{2.5}$–PM_{10} correlation are reported only for the physically consistent nodes (3–6, ratio $\approx$ 0.96) and are treated as exploratory. With this strategy, the principal $PM_{2.5}$ findings are unaffected: recomputed on daily means (to mitigate temporal autocorrelation), campaign-mean $PM_{2.5}$ is 39.3 $\mu g/m^3$, the February–March contrast remains large and significant (daily means 55.2 vs. 23.1 $\mu g/m^3$; Welch $t = 5.12$, $p < 0.001$), and 84% of monitored days exceed the WHO 24 h guideline. Confirmation of the frame-assignment hypothesis against the device firmware, and bench re-collocation of the two affected units, are recommended before their PM_{10} channel is used in future work.
10. Monitoring period: The campaign covers a relatively short period of approximately two months. Although sufficient to evaluate system functionality and demonstrate spatial variation in air quality, it does not allow assessment of long-term sensor stability; low-cost sensors may degrade through aging, optical-chamber contamination, environmental exposure, and seasonal temperature and humidity changes. Future work will therefore prioritize continuous year-round monitoring, periodic co-location with reference-grade instruments, drift analysis, recalibration, and machine-learning-based correction models to maintain measurement accuracy over extended operation.

5. Conclusions

This paper presented the design, deployment, and field evaluation of a modular IoT-based air quality monitoring network in Almaty, Kazakhstan. By integrating commercial off-the-shelf components with custom engineering solutions—specifically, an active air sampling enclosure and adaptive edge-filtering algorithms—we successfully operated a six-node sensor network during a two-month winter–spring campaign, collecting over 70,000 measurements.

The key contributions of this work are:

1. Robust hardware architecture: A reproducible hardware blueprint (estimated ~$100 per unit) capable of withstanding Almaty's harsh winter conditions while maintaining data integrity.
2. Intelligent telemetry: An edge-computing pipeline that optimizes data quality and bandwidth through context-aware signal processing, reducing data transmission volume by ~40%.
3. Quantitative environmental characterization: Field evidence showing that $PM_{2.5}$ exceeded WHO guidelines on 84% of monitored days, with a 60% reduction from the heating season (February) to spring (March). $PM_{2.5}/PM_{10}$ ratio analysis (~0.96) and weekday–weekend patterns are consistent with coal-based heating being the most likely dominant PM source, based on indirect evidence rather than direct source apportionment.
4. Network feasibility pilot: Demonstration that six rotating devices can be operated across 20 measurement sites spanning a 13 × 20 km urban area, establishing the foundation for future spatially resolved campaigns with continuous GPS logging. Because measurement–location links were not preserved, this is presented as a pilot of network feasibility rather than a spatially resolved pollution map.
5. Prepared model integration framework: A configured WRF–SILAM backend prepared for future operational coupling. Forecasting and model-validation capabilities are described here as planned future work, since no results from the coupled system are reported in this study.

The proposed system offers a scalable, low-cost alternative to capital-intensive stationary networks. For the city of Almaty, this architecture bridges the gap between sparse official monitoring and the need for distributed, high-frequency data. Future iterations will focus on: (1) activating gas-phase sensor channels (NO_2, CO, O_3); (2) upgrading from approximate to continuous GPS logging for high-resolution route-level spatial mapping; (3) completing collocation validation against Kazhydromet reference stations; (4) operationally coupling the WRF–SILAM framework for model–observation data fusion; and (5) implementing on-device machine learning (TinyML) for real-time source apportionment. This open-architecture approach provides a transferable template for other municipalities in Central Asia and beyond that seek to implement data-driven environmental management strategies under limited budget constraints.

Author Contributions: Conceptualization, D.N. and R.U.; methodology, D.N., K.B. and G.A.; software, K.B.; validation, G.A. and A.M.; formal analysis, G.A.; investigation, K.B.; resources, G.A.; data curation, G.A.; writing—original draft preparation, G.A.; writing—review and editing, D.N. and A.M.; visualization, K.B.; supervision, A.M.; project administration, Z.K.; funding acquisition, Z.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR24993051 "Development of an intelligent city system based on IoT and data analysis").

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset analyzed in this study is available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank the Almaty City Administration for supporting access to urban infrastructure for sensor deployment.

Conflicts of Interest: Authors Daniyar Nurseitov, Kairat Bostanbekov, and Galymzhan Abdimanap were employed by KazMunayGas (Kazakhstan). The remaining authors declare that the research

was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

WRF	Weather Research and Forecasting
SILAM	System for Integrated modeLling of Atmospheric coMposition
PM	Particulate Matter
CO_2	Carbon Dioxide
IoT	Internet of Things
CAMS	Copernicus Atmosphere Monitoring Service
BEP	Building Effect Parameterization
PBL	Planetary Boundary Layer
TVOC	Total Volatile Organic Compounds
RH	Relative Humidity
API	Application Programming Interface
AQMS	Air Quality Monitoring Station
CHP	Combined Heat and Power
DMZ	Demilitarized Zone
FEM	Federal Equivalent Method
GLONASS	Global Navigation Satellite System
GPS	Global Positioning System
HTTPS	Hypertext Transfer Protocol Secure
LCZ	Local Climate Zone
LUR	Land Use Regression
MAE	Mean Absolute Error
NDIR	Non-Dispersive Infrared
$PM_{2.5}/PM_{10}$	Particulate Matter $\leq 2.5/10$ µm
R^2	Coefficient of Determination
RRTMG	Rapid Radiative Transfer Model for GCMs
SCAQMD	South Coast Air Quality Management District
TinyML	Tiny Machine Learning
UART	Universal Asynchronous Receiver-Transmitter
WHO	World Health Organization
WSM6	WRF Single-Moment 6-class microphysics scheme

References

1. Sicard, P.; Agathokleous, E.; Anenberg, S.C.; De Marco, A.; Paoletti, E.; Calatayud, V. Trends in urban air pollution over the last two decades: A global perspective. *Sci. Total Environ.* **2023**, *858*, 160064. [CrossRef] [PubMed]
2. Dedova, T.; Balakay, L.; Zakarin, E.; Bostanbekov, K.; Abdimanap, G. High-Resolution WRF Modeling of Wind and Thermal Regimes with LCZ in Almaty, Kazakhstan. *Atmosphere* **2024**, *15*, 966. [CrossRef]
3. Ogbuabia, T.B.; Guney, M.; Baimatova, N.; Ulusoy, I.; Karaca, F. Assessing the Impact of Combined Heat and Power Plants (CHPPs) in Central Asia: A Case Study in Almaty for $PM_{2.5}$ Simulations Using WRF-AERMOD and Ground Level Verification. *Atmosphere* **2023**, *14*, 1554. [CrossRef]
4. Temirbekov, N.; Temirbekova, M.; Tamabay, D.; Kasenov, S.; Askarov, S.; Tukenova, Z. Assessment of the Negative Impact of Urban Air Pollution on Population Health Using Machine Learning Method. *Int. J. Environ. Res. Public Health* **2023**, *20*, 6770. [CrossRef] [PubMed]
5. Assanov, D.; Zapasnyi, V.; Kerimray, A. Air Quality and Industrial Emissions in the Cities of Kazakhstan. *Atmosphere* **2021**, *12*, 314. [CrossRef]
6. Naizabayeva, L.; Kolesnikova, K.; Khrutba, V. Simulation-Based Assessment of Urban Pollution in Almaty: Influence of Meteorological and Environmental Parameters. *Appl. Sci.* **2025**, *15*, 6391. [CrossRef]

7. Galatioto, F.; Ferguson-Moore, J.; Calderwood, R. Mobile Monitoring for the Spatial and Temporal Assessment of Local Air Quality (NO_2) in the City of London. *Atmosphere* **2021**, *12*, 106. [CrossRef]
8. Castell, N.; Dauge, F.R.; Schneider, P.; Vogt, M.; Lerner, U.; Fishbain, B.; Broday, D.; Bartonova, A. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environ. Int.* **2017**, *99*, 293–302. [CrossRef] [PubMed]
9. Schneider, P.; Castell, N.; Vogt, M.; Dauge, F.R.; Lahoz, W.A.; Bartonova, A. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environ. Int.* **2017**, *106*, 234–247. [CrossRef] [PubMed]
10. Tasmurzayev, N.; Amangeldy, B.; Smagulova, G.; Baigarayeva, Z.; Imash, A. A Low-Cost IoT Sensor and Preliminary Machine-Learning Feasibility Study for Monitoring In-Cabin Air Quality: A Pilot Case from Almaty. *Sensors* **2025**, *25*, 4521. [CrossRef] [PubMed]
11. Guerrón, R.A.; D'Amore, F.; Bencardino, M.; Lamonaca, F.; Colaprico, A.; Lanuzza, M.; Taco, R.; Carnì, D.L. IoT sensor nodes for air pollution monitoring: A review. *Acta IMEKO* **2023**, *12*, 26. [CrossRef]
12. Gryech, I.; Assad, C.; Ghogho, M.; Kobbane, A. Applications of machine learning and IoT for Outdoor Air Pollution Monitoring and Prediction: A Systematic Literature Review. *arXiv* **2024**, arXiv:2401.01788.
13. Sladojevic, S.; Arsenovic, M.; Nikolic, D.; Anderla, A.; Stefanovic, D. Advancements in Mobile Based Air Pollution Detection: From Literature Review to Practical Implementation. *J. Sens.* **2024**, *2024*, 4895068. [CrossRef]
14. Whitehill, A.R.; Lunden, M.; LaFranchi, B.; Kaushik, S.; Solomon, P.A. Mobile air quality monitoring and comparison to fixed monitoring sites for instrument performance assessment. *Atmos. Meas. Tech.* **2024**, *17*, 2991–3009. [CrossRef] [PubMed]
15. Ballamajalu, R.; Nair, S.; Chhabra, S.; Monga, S.K.; SVR, A.; Hegde, M.; Simmhan, Y.; Sharma, A.; Choudhary, C.M.; Sutaria, R.; et al. Toward SATVAM: An IoT Network for Air Quality Monitoring. *arXiv* **2018**, arXiv:1811.07847.
16. Parmar, A.; Sara, S.; Dwivedi, A.K.; Reddy, C.R.; Patwardhan, I.; Bijjam, S.D.; Chaudhari, S.; Rajan, K.S.; Vemuri, K. Development of End-to-End Low-Cost IoT System for Densely Deployed PM Monitoring Network: An Indian Case Study. *arXiv* **2022**, arXiv:2211.16497.
17. Flores-Cortez, O.O.; Pocasangre Jimenez, C.; Arévalo, F.; López, R.L.; Peña Martínez, D.; Rafael, O.P. A Low-cost IoT Mobile System for Air Quality Monitoring in Developing Countries, a Study Case in El Salvador. In Proceedings of the 2023 International Conference on Smart Applications, Communications and Networking (SmartNets), Istanbul, Turkey, 25–27 July 2023; pp. 1–8. [CrossRef]
18. Johnson, T.; Woodward, K. Enviro-IoT: Calibrating Low-Cost Environmental Sensors in Urban Settings. *arXiv* **2025**, arXiv:2502.07596.
19. Song, J.; Han, K. Deep-MAPS: Machine Learning based Mobile Air Pollution Sensing. *arXiv* **2020**, arXiv:1904.12303.
20. Singh, V.; Agarwal, A. Spatial prediction of on-road air pollution using long-term mobile monitoring: Insights from Delhi. *Urban Clim.* **2025**, *60*, 102347. [CrossRef]
21. Dedova, T.; Balakay, L.; Zakarin, E.; Bostanbekov, K.; Abdimanap, G. Investigating Stagnant Air Conditions in Almaty: A WRF Modeling Approach. *Atmosphere* **2024**, *15*, 633. [CrossRef]
22. Temirbekov, N.; Malgazhdarov, Y.; Tamabay, D.; Temirbekov, A. Atmospheric modelling of photochemical transformations of pollutants: Impact of weather conditions and diurnal cycle (Case study: Ust-Kamenogorsk, Kazakhstan). *Math. Model. Eng. Probl.* **2023**, *10*, 1699–1705. [CrossRef]
23. Spiridonov, V.; Jakimovski, B.; Spiridonova, I.; Pereira, G. Development of air quality forecasting system in Macedonia, based on WRF-Chem model. *Air Qual. Atmos. Health* **2019**, *12*, 825–836. [CrossRef]
24. Kukkonen, J.; Kangas, L.; Kauhaniemi, M.; Sofiev, M.; Aarnio, M.; Jaakkola, J.J.K.; Kousa, A.; Karppinen, A. Modelling of the urban concentrations of $PM_{2.5}$ on a high resolution for a period of 35 years, for the assessment of lifetime exposure and health effects. *Atmos. Chem. Phys.* **2018**, *18*, 8041–8064. [CrossRef]
25. Mazzeo, A.; Burrow, M.; Quinn, A.; Marais, E.A.; Singh, A.; Ng'ang'a, D.; Gatari, M.J.; Pope, F.D. Evaluation of the WRF and CHIMERE models for the simulation of $PM_{2.5}$ in large EastAfrican urban conurbations. *Atmos. Chem. Phys.* **2022**, *22*, 10677–10701. [CrossRef]
26. Meneses-Albala, E.; Montalban-Faet, G.; Felici-Castell, S.; Perez-Solano, J.; Fayos-Jordan, R. Assessment of a Multisensor ZPHS01B-Based Low-Cost Air Quality Monitoring System: Case Study. *Electronics* **2025**, *14*, 1531. [CrossRef]
27. South Coast Air Quality Management District. *Field Evaluation: Qingping Air Monitor*; Technical Report; SCAQMD: Diamond Bar, CA, USA, 2023.
28. Zhengzhou Winsen Electronics Technology Co., Ltd. ZPHS01B Multi-in-One Air Quality Monitoring Sensor Module—User's Manual (Datasheet). Available online: https://www.winsen-sensor.com/d/files/air-quality/zphs01b-english-version1_3.pdf (accessed on 25 June 2026).
29. *GB/T 18801-2015*; Air Cleaner. Standardization Administration of China: Beijing, China, 2015.
30. Balagopal, G.; Wijeratne, L.; Waczak, J.; Hathurusinghe, P.; Iqbal, M.; Kiv, D.; Aker, A.; Lee, S.; Agnihotri, V.; Simmons, C.; et al. Calibration of Low-Cost LoRaWAN-Based IoT Air Quality Monitors Using the Super Learner Ensemble: A Case Study for Accurate Particulate Matter Measurement. *Sensors* **2025**, *25*, 1614. [CrossRef] [PubMed]

31. Clarity Movement Co. *The Cost of Air Quality Monitoring: A Pricing Guide for Cities and Agencies*; Clarity Movement Co.: Berkeley, CA, USA, 2026. Available online: https://www.clarity.io/blog/cost-of-air-quality-monitoring-a-pricing-guide-for-cities-agencies (accessed on 27 June 2026).
32. Delp, W.W.; Singer, B.C. Wildfire Smoke Adjustment Factors for Low-Cost and Professional $PM_{2.5}$ Monitors with Optical Sensors. *Sensors* **2020**, *20*, 3683. [CrossRef] [PubMed]
33. Air Quality Monitoring Dashboard of Kazakhstan. Available online: http://dashboard.air.org.kz (accessed on 27 June 2026).

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