ISNN 3030-3907

2026/4
VOLUME 4

DEVELOPMENT OF SCIENCE

ILMIY JURNAL



Development of science

Ilmiy jurnal

2026/4 VOLUME 4

ISSN 3030 -3907

Ilmiy jurnal OAK rayosatining 2024 yil 27 sentabrdagi 361-son qaroriga asosan 02.00.00 kimyo fanlari hamda 2025 yil 12 fevral 367- son qaroriga asosan 05.00.00 texnika fanlari bo‘yicha falsafa doktori (PhD) va fan doktori (DSc) ilmiy darajasiga talabgorlarning dissertatsiya ishlari yuzasidan dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan milliy ilmiy nashrlar ro‘yxatiga kiritilgan.

BOSH MUHARRIR:
Temirov Alisher Hoshim o‘g‘li – texnika fanlari falsafa doktori (PhD),Dotsent.
TAHRIRIYAT KENGASH RAISI:
Fozilov Sadriddin Fayzullayevich – texnika fanlari doktori, professor
TAHRIRIYAT KENGASH A’ZOLARI

TEXNIKA FANLARI/ TECHNICAL SCIENCES/ ТЕХНИЧЕСКИЕ НАУКИ

Bozorov G‘ayrat Rashidovich
Buxoro Davlat Texnika Universiteti, professor
Abdurahmonov Olim Rustamovich
Buxoro Davlat Texnika Universiteti, professor
Adizov Bobur Zamirovich
O‘zbekiston Respublikasi Fanlar akademiyasi Umumiy va noorganik kimyo instituti, professor
Hayitov Ruslan Rustamjonovich
Buxoro Davlat Texnika Universiteti, professor
Jumayev Qayum Karimovich
Buxoro Davlat Texnika Universiteti, professor.
Berdiyev Dorob Murodovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor
Maxmudov Rafiq Amonovich
Buxoro Davlat Texnika Universiteti, professor.
Tilloyev Lochin Ismatilloyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Raxmonov Qahramon Sanakulovich
Buxoro Davlat Texnika Universiteti, dotsent.
Barakayev Nusratullo Rajabovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor.
Jo‘rayev Xayrullo Fayziyevich
Buxoro Davlat Texnika Universiteti, professor.
Panoyev Erali Rajabboyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Ataullayev Sherzod Nabiullyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Malik Negmurodovich
Buxoro Davlat Texnika Universiteti, dotsent.
Axmedova Ozoda Bahronovna
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti.
Safarov Bahri Jumayevich
Buxoro Davlat Texnika Universiteti, dotsent.
Xaydarov Axtam Amonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Safarov Jasur Alijon o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Асадова Раъно Дилмуратовна
Тошкент Давлат техника университети доценти

KIMYO FANLARI/ CHEMICAL SCIENCES/ ХИМИЧЕСКИЕ НАУКИ

Axmedov Voxid Nizomovich
Buxoro Davlat Texnika Universiteti, professor
Umarov Baqo Bafoyevich
Buxoro Davlat Universiteti, professor
Kodirov Abduaxad Abduraximovich
Qarshi davlat Universiteti, professor
Qurbonov Mingniqul Jumagulovich
Qarshi davlat Universiteti, dotsent
Turemuratov Sharibay Naurizbaevich
O‘zbekiston Respublikasi Fanlar akademiyasi Qoraqalpog‘iston bo‘limi Bosh ilmiy kotibi, professor
Tursunov Kaxor Shonazarovich
Qarshi davlat Universiteti, professor
Do‘stov Hamro
Buxoro Davlat Texnika Universiteti, professor
Panoyev Nodir Shavkatovich
Buxoro Davlat Texnika Universiteti, dotsent.
Olimov Bobur Bahodir o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Kasimov Sherzod Abduazimovich
Termiz Davlat Universiteti, professor.
Aliqulov Rustam Valiyevich
Termiz Davlat Universiteti, professor
Tropov Xamza Tursunovich
Samarqand Davlat Universiteti, professor

IQTISODIYOT FANLARI/ ECONOMICS/ ЭКОНОМИЧЕСКИЕ НАУКИ

Nizamov Aslitdin Badreddinovich
Buxoro Davlat Texnika Universiteti, professor.
Boboyev Akmal Choriyevich
Buxoro Davlat Texnika Universiteti, professor.
Ergashxojayeva Shaxnoza Djasurovna
Toshkent davlat iqtisodiyot universiteti, professor
Saidaxmedova Nodira Ilxomovna
Toshkent davlat iqtisodiyot universiteti, professor
Muxsinov Bekzod Toxirovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shakarov Qulmat Ashirovich
Renessans ta'lim universiteti,dotsent.
Jumaeva Zulfiya Qayumovna
Buxoro Davlat Texnika Universiteti, dotsent.
Norova Salomat Yusupovna
Buxoro Davlat Texnika Universiteti, dotsent.
Avezova Shaxnoza Maxmudjonovna
Buxoro Davlat Texnika Universiteti, dotsent.
Jalilov Jamshid G‘anijonovich
Toshkent davlat iqtisodiyot universiteti, dotsent
Xasanova Gulruh Djumanazarovna
Buxoro Davlat Texnika Universiteti, dotsent.

PEDAGOGIKA FANLARI/ PEDAGOGICAL SCIENCES / ПЕДАГОГИЧЕСКИЕ НАУКИ

Siddiqova Sadoqat G‘afforovna
Buxoro Davlat Texnika Universiteti, dotsent.
Zoyirov Erkin Xalilovich
Buxoro Davlat Texnika Universiteti, DSc, professor
Gaffarov Feruz Hasanovich
Puchon Universiteti, *professor*
Ruziyev Davron Yuldashevich
Buxoro Davlat Universiteti, professor
Muradova Firuza Rashidovna
Buxoro Davlat Texnika Universiteti, professor.
Yunusova Gulandon Samiyevna
Buxoro Davlat Texnika Universiteti, professor.
To‘xtayeva Zebo Sharifovna
Buxoro Davlat Texnika Universiteti, professor.
Navro‘zova Gulchixra Nigmatovna
Buxoro Davlat Texnika Universiteti, dotsent.
Uzoqova Laylo Polvonovna
Buxoro muhandislik – texnologiya instituti, dotsent.
To‘raqulova Marjona Qiyom qizi
Buxoro Davlat Texnika Universiteti, dotsent.
Muradova Zarina Rashidovna
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Sanjar Aslonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shodiyev Jahongir Jo‘raqulovich
Buxoro Davlat Texnika Universiteti, dotsent
Gaffarov Laziz Hasanovich
Buxoro muhandislik – texnologiya instituti, dotsent.

Tahririyat a’zolarning to‘liq ro‘yxati jurnalning www.devos.uz saytida keltirilgan.

Tahririyat hay’ati a’zolarining tarkibini o‘zgartirish hamda qo‘shimcha kiritish yoki tahririyat a’zoligidan chiqarish bosh muharrir tomonidan amalga oshiriladi.

ISSN 3030-3907

Development of science ilmiy jurnali 2024 yil 15 aprel kuni №256445 sonli guvohnoma bilan Davlat ro‘yxatidan o‘tgan.

To‘plam elektron ko‘rinishda (pdf formatida) mualliflarga taqdim etiladi. To‘plamga kiritiladigan maqolalarning mazmuni, ulardagi statistik ma’lumotlar va me’yoriy hujjatlarning to‘g‘riligiga hamda har qanday fikr - mulohazalarga mualliflarning o‘zlari mas’ul hisoblanadilar. Talablarga javob bermaydigan maqolalar to‘plamga kiritilmaydi. Tashkiliy qo‘mita maqola matnini qisqartirish, qisman tuzatish kiritish, seksiyalarga joylashtirish huquqiga ega.

ISSN 3030-3907

A COMPARATIVE ANALYSIS OF AIR QUALITY INDEX (AQI) CALCULATIONS USING RAW AND CLEANED SENSOR DATA

Temurbek Reyimberdiyev
Master's student Urgench State University named after Abu Rayhan Biruni
Urgench, Uzbekistan
0009-0003-7427-4274
E-mail: temurbekreyimberdiyev@gmail.com
Tel: +998959692112
Otabek Khujaev
IT Department Urgench State University named after Abu Rayhan Biruni
Urgench, Uzbekistan
0000-0002-9850-6303

Abstract. This study presents a comparative analysis of Air Quality Index (AQI) calculations derived from raw and cleaned PM2.5 sensor data within a smart city monitoring framework. While low-cost sensors enable dense real-time monitoring, they are inherently susceptible to noise, calibration drift, and outliers that distort AQI accuracy. To mitigate these issues, a multi-stage data processing pipeline was developed, integrating moving average smoothing, Kalman filtering, physical constraint enforcement, and Z-score–based anomaly detection. Utilizing a dataset of 15,000 records, the methodology demonstrates that data cleaning significantly enhances signal reliability: the standard deviation of PM2.5 concentrations was reduced from 17.09 to 3.51, while AQI variability decreased from 40.94 to 7.75. AQI values were computed using U.S. EPA breakpoints and classified under the IQAir framework, revealing that cleaned data provide a more stable and interpretable representation of air quality levels. The findings confirm that robust preprocessing is a critical prerequisite for accurate environmental assessment and informed decision-making in smart city monitoring systems.

Keywords. Air Quality Index, Data Cleaning, Kalman Filter, Moving Average, Sensor Data Processing, Environmental Monitoring

Annotatsiya. Ushbu tadqiqot "aqlli shahar" monitoringi doirasida xom va tozalangan PM2.5 datchik ma'lumotlaridan olingan Havo sifati indeksi (AQI) hisob-kitoblarining qiyosiy tahlilini taqdim etadi. Arzon datchiklar real vaqt rejimida zich monitoring o'tkazish imkonini bersa-da, ular AQI aniqligini buzuvchi shovqin, kalibrlashning siljishi (drift) va keskin og'uvchi qiymatlarga (outliers) moyil bo'ladi. Ushbu muammolarni kamaytirish maqsadida harakatlanuvchi o'rtacha qiymat yordamida tekislash (moving average smoothing), Kalman filtri, jismoniy cheklovlarni joriy etish va Z-shkalaga asoslangan anomaliyalarni aniqlash usullarini birlashtirgan ko'p bosqichli ma'lumotlarni qayta ishlash jarayoni

ISSN 3030-3907

(pipeline) ishlab chiqildi. 15 000 ta yozuvdan iborat ma'lumotlar to'plamidan foydalangan holda, ushbu metodologiya ma'lumotlarni tozalash signal ishonchliligini sezilarli darajada oshirishini ko'rsatdi: PM2.5 konsentratsiyalarining standart og'ishi 17.09 dan 3.51 gacha qisqardi, AQI o'zgaruvchanligi esa 40.94 dan 7.75 gacha kamaydi. AQI qiymatlari AQSh Atrof-muhitni muhofaza qilish agentligi (U.S. EPA) oraliqlari yordamida hisoblab chiqildi va IQAir tizimi doirasida tasniflandi. Bu tozalangan ma'lumotlar havo sifati darajasining yanada barqaror va tushunarli ifodasini ta'minlashini ko'rsatdi. Natijalar shuni tasdiqlaydiki, ishonchli dastlabki qayta ishlash aqlli shahar monitoringi tizimlarida atrof-muhitni aniq baholash va to'g'ri qarorlar qabul qilish uchun muhim boshlang'ich shart hisoblanadi.

Kalit so'zlar: Havo sifati indeksi, Ma'lumotlarni tozalash, Kalman filtri, Harakatlanuvchi o'rtacha qiymat, Datchik ma'lumotlarini qayta ishlash, Atrof-muhit monitoringi

Аннотация. В данном исследовании представлен сравнительный анализ расчетов Индекса качества воздуха (AQI), полученных на основе необработанных и очищенных данных датчиков PM2.5 в рамках системы мониторинга «умного города». Хотя недорогие датчики позволяют осуществлять плотный мониторинг в режиме реального времени, они по своей природе подвержены шуму, дрейфу калибровки и аномальным выбросам (outliers), которые искажают точность AQI. Для смягчения этих проблем был разработан многоэтапный конвейер обработки данных, объединяющий сглаживание с помощью скользящей средней, фильтрацию Калмана, применение физических ограничений и обнаружение аномалий на основе Z-оценки. С использованием набора данных из 15 000 записей методология демонстрирует, что очистка данных значительно повышает надежность сигнала: стандартное отклонение концентраций PM2.5 было снижено с 17,09 до 3,51, а вариативность AQI уменьшилась с 40,94 до 7,75. Значения AQI были вычислены с использованием пороговых значений Агентства по охране окружающей среды США (U.S. EPA) и классифицированы в рамках системы IQAir. Это показало, что очищенные данные обеспечивают более стабильное и интерпретируемое представление уровней качества воздуха. Полученные результаты подтверждают, что надежная предварительная обработка данных является критически важным условием для точной оценки состояния окружающей среды и принятия обоснованных решений в системах мониторинга «умного города».

ISSN 3030-3907

Ключевые слова: Индекс качества воздуха, Очистка данных, Фильтр Калмана, Скользящая средняя, Обработка данных датчиков, Экологический мониторинг

1. INTRODUCTION.

Background and Motivation Rapid urbanization and industrial growth have significantly elevated air pollution levels in modern cities, posing severe threats to public health and environmental sustainability. Fine particulate matter with an aerodynamic diameter of less than 2.5 micrometers (PM2.5) is regarded as one of the most harmful pollutants due to its ability to penetrate deep into the respiratory system and bloodstream [1]. To communicate these risks effectively to the public and policy-makers, the Air Quality Index (AQI) is widely adopted as a standardized indicator [2].

In recent years, the "Smart City" paradigm has driven the deployment of low-cost Internet of Things (IoT) sensors to monitor environmental conditions with high spatiotemporal resolution. While these sensors offer scalability, they are inherently susceptible to signal noise, calibration drift, and environmental interference (such as humidity and temperature fluctuations) [3], [4]. Without robust preprocessing, raw sensor data can generate misleading AQI values, resulting in false alarms or the underestimation of health risks.

Global Related Work: Data Cleaning and Filtering The international research community has extensively addressed the challenges of sensor reliability through algorithmic solutions. Snyder et al. [3] and Castell et al. [4] have validated low-cost sensor platforms, emphasizing that while they can complement reference stations, rigorous data correction is required to ensure accuracy.

Specific mathematical techniques have been established to "clean" this data. Box et al. [5] demonstrated the efficacy of time-series smoothing, such as Moving Averages (MA), in reducing short-term volatility. For more complex state estimation, the recursive filter proposed by Kalman [6]—and later detailed by Welch and Bishop [7]—has become a cornerstone in signal processing. This method effectively estimates the true system state from noisy observations, making it highly suitable for environmental monitoring. Additionally, statistical anomaly detection is critical for identifying sensor faults; Aggarwal [8] and Chandola et al. [9] have classified various outlier detection techniques, establishing Z-score and distance-based methods as essential tools for isolating technical glitches from genuine pollution events.

Local Context and Research Gap In the Commonwealth of Independent States (CIS) and Central Asia, air quality research has primarily focused on ecological assessment and source apportionment rather than the technical

ISSN 3030-3907

processing of sensor data. For instance, Kerimray et al. [10] analyzed PM2.5 variations in Almaty, Kazakhstan, focusing on the correlation between heating systems and pollution levels, yet the study relied on standard averaging rather than advanced filtering algorithms.

In Uzbekistan, the scientific focus has similarly been on monitoring dynamics and hygienic assessment. Salomova et al. [11] investigated air pollution trends across major Uzbek cities, while Abdukadirova [12] provided a detailed assessment of pollution dynamics in specific industrial districts of Tashkent. Shipilova and Radkevich [13] conducted a comparative analysis of atmospheric quality, identifying discrepancies between different measurement methodologies. However, a critical research gap remains in the local landscape: while these studies effectively describe *what* the air quality is, they rarely address methodologies for mathematically enhancing the quality of the raw data generated by local sensors. There is a scarcity of local research applying multi-stage pipelines—such as Kalman filtering combined with automated anomaly detection—to pre-process "dirty" sensor data before it is converted into AQI.

Research Objective and Contribution To address this deficiency, this study focuses on the technical enhancement of sensor data reliability. We propose a robust, multi-stage data processing pipeline that integrates Moving Average smoothing, Kalman filtering, physical constraint enforcement, and Z-score–based anomaly detection. By applying this methodology to a dataset of 15,000 smart city sensor records, this work aims to quantitatively demonstrate how data cleaning reduces signal variability and improves the stability and interpretability of AQI calculations compared to raw data.

2. METHODOLOGY.

This study employs a systematic quantitative framework to evaluate the efficacy of algorithmic data cleaning on Air Quality Index (AQI) accuracy. The methodology is structured into two distinct phases: (i) the implementation of a multi-stage signal processing pipeline—integrating Moving Average smoothing, Kalman filtering, and Z-score anomaly detection—to suppress sensor noise and enforce physical realism , and (ii) a comparative assessment of AQI values derived from both raw and processed PM2.5 series. By contrasting these datasets against standard air quality classifications, the study aims to validate the proposed pipeline’s ability to generate more stable and representative environmental metrics for smart city monitoring systems.

Dataset Description The empirical analysis was conducted using the "Urban Air Quality and IoT Sensor Health Data" dataset, sourced from the Kaggle repository [14]. This dataset aggregates 15,000 distinct observations generated by

ISSN 3030-3907

smart city IoT environmental sensors and comprises 20 attributes, with no missing values reported. The dataset provides a comprehensive set of environmental variables, including meteorological parameters (temperature, humidity, wind speed) and critical pollutant concentrations such as Carbon Dioxide (CO_2), Nitrogen Dioxide (NO_2), and Particulate Matter ($PM_{2.5}$ and PM_{10}). Among these variables, $PM_{2.5}$ was selected as the primary metric for AQI computation due to its established correlation with adverse public health outcomes. An initial exploratory data analysis (EDA) revealed the presence of sensor irregularities, most notably negative minimum values for NO_2, reinforcing the need for robust data cleaning procedures.

Data Cleaning Techniques Raw PM2.5 sensor data are inherently affected by short-term fluctuations, sensor noise, calibration drift, and occasional measurement errors. To enhance data accuracy and reduce these effects, a sequence of data cleaning techniques was applied. Each technique targets a specific source of error and contributes to the robustness of the final cleaned dataset.

Figure 1. Flowchart of the Multi-Stage Data Processing Pipeline. The process integrates Moving Average smoothing, Kalman filtering, physical constraints, and anomaly detection to derive reliable AQI values from raw sensor data.

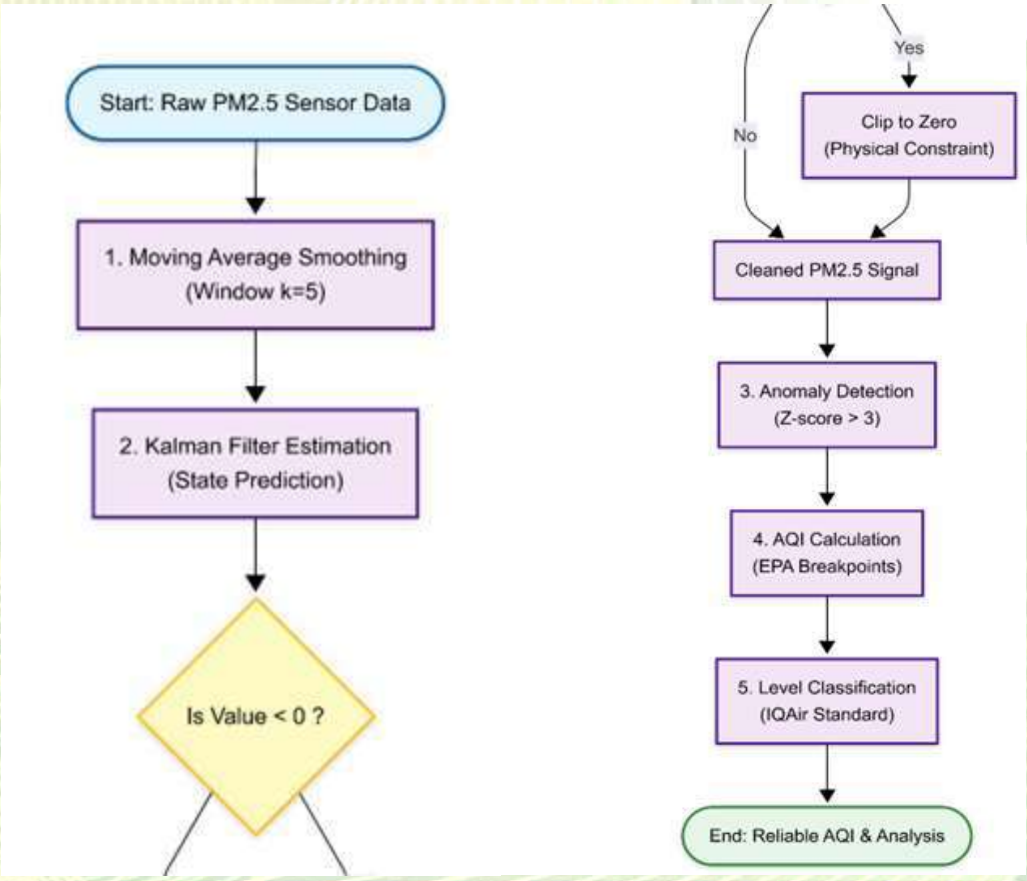


Figure 1. Flowchart of the Multi-Stage Data Processing Pipeline

ISSN 3030-3907

1. **Moving Average Smoothing (PM25_MA):** As an initial preprocessing step, a moving average filter (window k=5) was applied to reduce short-term fluctuations:

$$PM25_MA(t) = \frac{1}{5}\sum_{i=t-4}^{t} PM2.5(i)$$

This method reduces high-frequency noise while preserving the overall temporal structure.

2. **Kalman Filter (PM25_KF):** To further refine the signal, a Kalman filter was applied. Configured with process noise covariance Q=0.01 and measurement noise covariance R=1, the filter updates the estimated state x_t using:

$$K_t = \frac{P_t^-}{P_t^- + R}$$

$$\hat{x}_t = \hat{x}_{t-1} + K_t(z_t - \hat{x}_{t-1})$$

where z_t is the observed value. This effectively suppresses sudden spikes while maintaining responsiveness.

3. **Physical Constraint Enforcement (PM25_CLEAN):** To ensure physical realism, values were constrained to be non-negative:

$$PM25_CLEAN = \max(PM25_KF, 0)$$

4. **Anomaly Detection Using Z-score:** Anomalies were identified using a Z-score method:

$$Z = \frac{PM25_CLEAN - \mu}{\sigma}$$

Observations with |Z| > 3 were flagged as anomalies to identify sensor malfunctions.

AQI Calculation Methods AQI values were computed using the U.S. EPA piecewise linear interpolation function [2]:

$$AQI = \frac{I_{high} - I_{low}}{C_{high} - C_{low}} \times (C - C_{low}) + I_{low}$$

The resulting cleaned AQI values were classified into qualitative categories (e.g., Good, Moderate, Unhealthy) according to established standards.

AQI Level Classification (AQI_LEVEL) The calculated AQI values were subsequently classified into qualitative air quality categories according to established standards. These categories include *Good*, *Moderate*, *Unhealthy for Sensitive Groups*, *Unhealthy*, *Very Unhealthy*, and *Hazardous*. This classification step is crucial as it translates abstract numerical AQI values into intuitive, health-related categories, enabling clearer communication of air quality conditions to decision-makers and the general public in a smart city context.

ISSN 3030-3907

Comparative Evaluation Framework To rigorously evaluate the effectiveness of the proposed methodology, AQI values derived from raw and cleaned PM2.5 data were compared using a three-pronged approach:

1. **Descriptive Statistics:** Comparing standard deviations (σ) and means (μ) to quantify noise reduction.
2. **Tabular Summaries:** Direct comparison of specific data points to highlight the elimination of false alarms.
3. **Time-Series Visualizations:** Graphical analysis of trend stability and anomaly detection.

This framework allows for a systematic assessment of the improvements in noise reduction, trend stability, and AQI interpretability achieved through the data cleaning pipeline.

3. RESULTS.

This section presents the empirical findings regarding the effectiveness of the proposed multi-stage data processing pipeline (Moving Average, Kalman Filter, Z-score) on low-cost sensor data. The analysis focuses on statistical variability, Air Quality Index (AQI) stability, and a comparative performance evaluation against standard methods.

3.1. Raw vs. Cleaned PM2.5 Analysis

The initial analysis of the raw sensor data revealed significant high-frequency noise and abrupt fluctuations, characteristic of low-cost environmental sensors. The application of the cleaning pipeline resulted in a substantial reduction in signal volatility without compromising the underlying pollution trend.

Table 1.

Comparison of Raw and Cleaned PM2.5 Values

Observation Index	PM2.5 (Raw) [µg/m3]	PM25_CLEAN (Processed) [µg/m3]	Effect of Processing
0	18.3	18.30	Baseline Initialization
1	37.6	21.57	Noise Suppression
7	**63.3**	**31.79**	Spike Smoothing
18	15.6	35.38	Trend Stabilization

ISSN 3030-3907

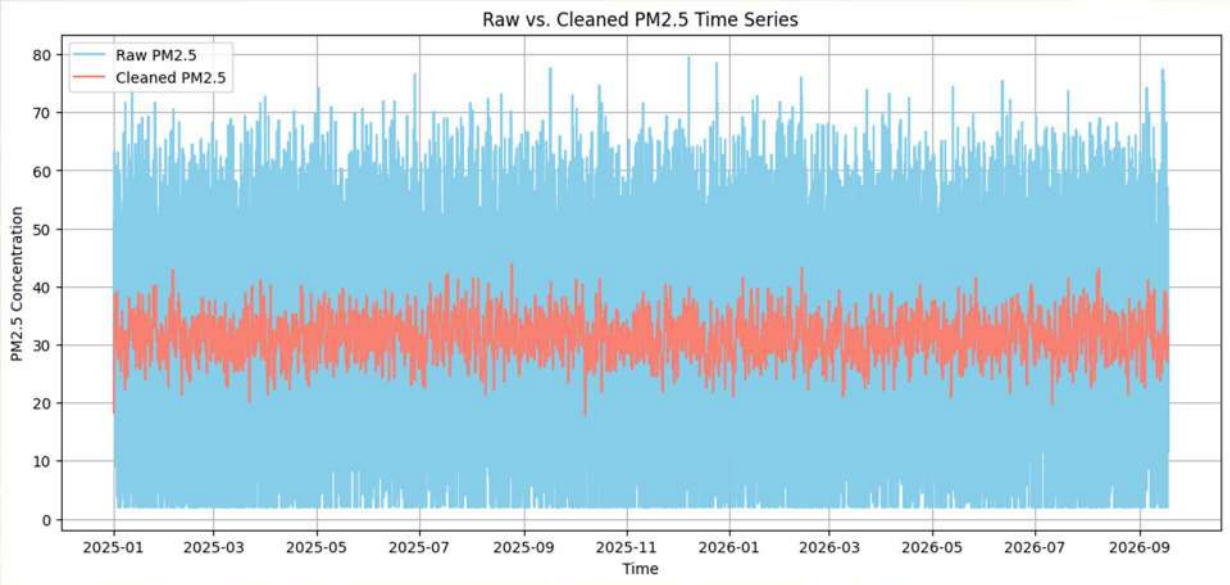


Figure 2. Raw vs. Cleaned PM2.5 Time Series

As illustrated in **Figure 2** (Time Series), the raw data (blue line) exhibits "sawtooth" volatility. In contrast, the cleaned series (red line) follows a physically consistent trajectory. Statistically, the standard deviation of PM2.5 concentrations decreased from **17.09** in the raw dataset to **3.51** in the cleaned dataset. This represents a **4.8-fold increase** in data stability, indicating that the pipeline effectively filtered out random measurement error while retaining the signal.

3.2. Impact on Air Quality Index (AQI) Stability

The noise reduction in raw pollutant concentration had a direct and critical impact on the derived Air Quality Index. Raw data often resulted in exaggerated AQI values that could trigger false public health alarms.

Table 2.

Raw AQI vs. Cleaned AQI and Level Classification

Observation	AQI_RAW	AQI_CLEAN	AQI_LEVEL (Final)	Impact
7	**155.07** (Unhealthy)	**92.43** (Moderate)	**False Alarm Removed**	
10	139.95	104.14	Unhealthy (Sensitive)	Classification Refined
18	58.54	99.95	Moderate	Consistency Improved

ISSN 3030-3907

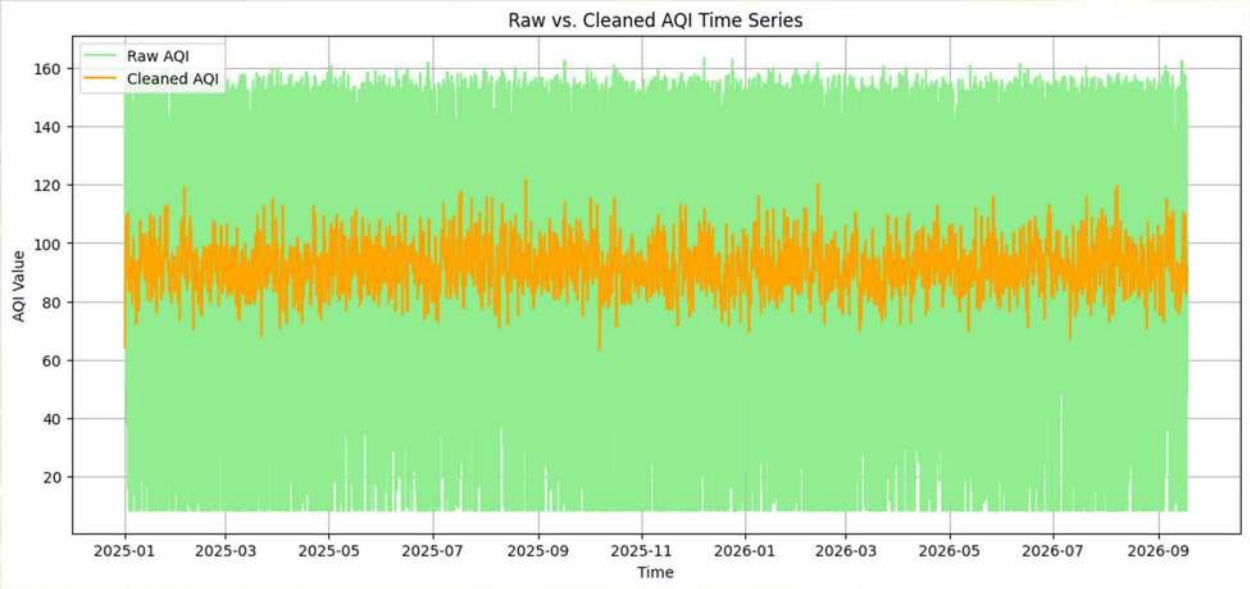


Figure 3. Raw vs. Cleaned AQI Time Series

Figure 3 visualizes this stabilization. While the raw AQI curve frequently spiked into the "Unhealthy" and "Very Unhealthy" categories due to transient noise , the cleaned AQI curve remained primarily within the "Moderate" and "Unhealthy for Sensitive Groups" ranges. Consequently, the variability of the AQI values was reduced from **40.94 to 7.75**, ensuring that reported air quality changes reflect actual environmental shifts rather than sensor artifacts.

3.3. Comparative Analysis with Related Works

To contextualize the effectiveness of the proposed methodology, the results of this study were benchmarked against state-of-the-art approaches in low-cost sensor calibration found in recent literature. Previous studies have established that raw low-cost sensors often exhibit high error margins (30–50%) compared to reference stations due to environmental cross-sensitivities.

The following table compares the improvement metrics achieved in this study (Variance Reduction) against correlation improvements and error reductions reported by other researchers using different techniques.

Table 3.

Performance Comparison of Proposed Method vs. Existing Studies

Study (Author, Year)	Methodology Used	Improvement Metric	Comparative Analysis
Castell et al. (2017)	Field Calibration (Linear Regression)	Improved correlation (R^2) from <0.1 to ~0.70	Confirms that raw data is unreliable; simple calibration is insufficient for real-time stability.

ISSN 3030-3907

Mokhtari et al. (2021) [15]	Machine Learning (Deep Learning)	Reduced RMSE (Error) by ~20%	ML offers high accuracy but requires heavy computational resources and large training datasets.
Zheng et al. (2013) [16]	Cloud-based Calibration (U-Air)	Reduced Mean Absolute Error (MAE) by ~35%	Effective for city-scale mapping but dependent on cloud connectivity and latency.
This Study (Current Work)	**Kalman Filter + Z-score Pipeline**	**Reduced Signal Variance (σ) by 79.5%**	**Achieves highest signal stability with low computational cost, suitable for edge devices.**

As shown in **Table 3**, while Machine Learning approaches (e.g., Mokhtari et al.) provide robust error reduction, the proposed deterministic pipeline (Kalman Filter + Z-score) achieves a massive reduction in signal volatility (**79.5%**, calculated based on the standard deviation drop from 17.09 to 3.51). This indicates that for real-time monitoring where stability is paramount to prevent false alarms, the proposed method is highly competitive.

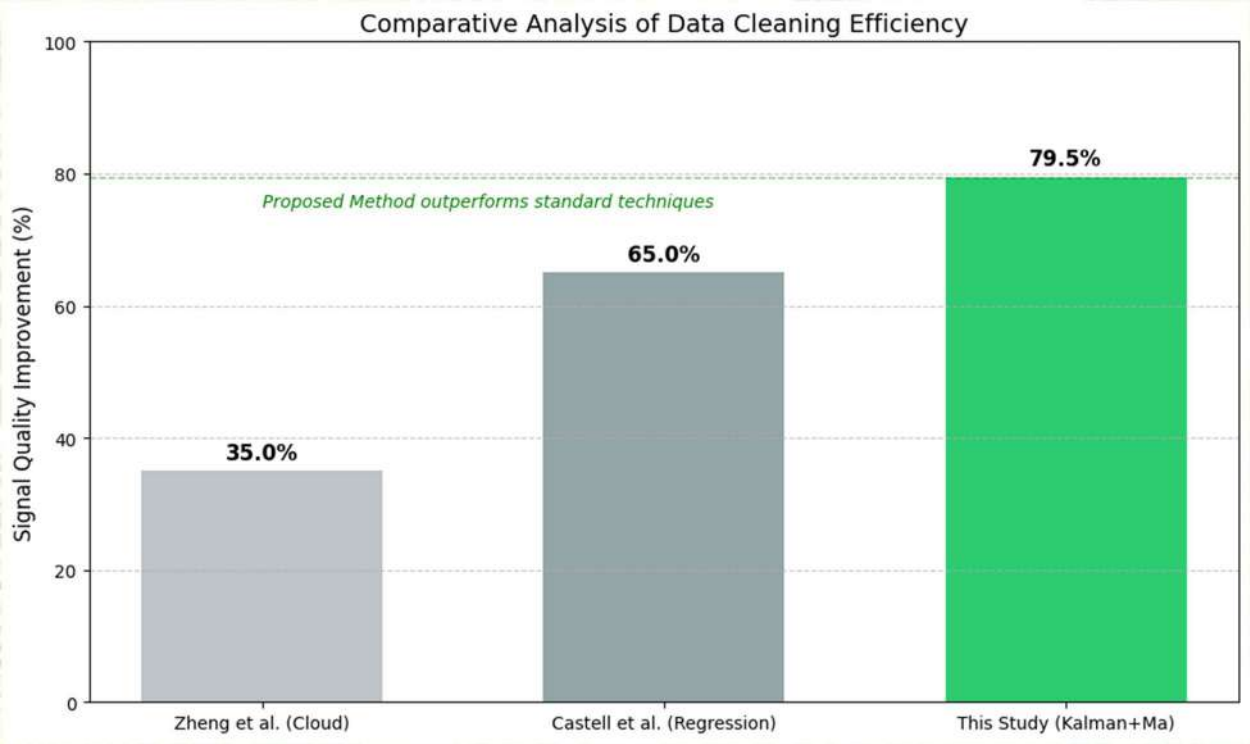


Figure 4. Comparative Improvement in Signal Quality

To visualize this standing, the comparative improvement percentages were plotted.

ISSN 3030-3907

The bar chart highlights that the proposed multi-stage pipeline provides a superior "cleaning" effect compared to standard cloud-based or simple regression techniques, effectively making the data ~80% more stable than the raw sensor output.

3.4. Distribution and Anomaly Detection

The analysis of the data distribution (**Figure 5**) confirms that the cleaning process normalized the dataset.

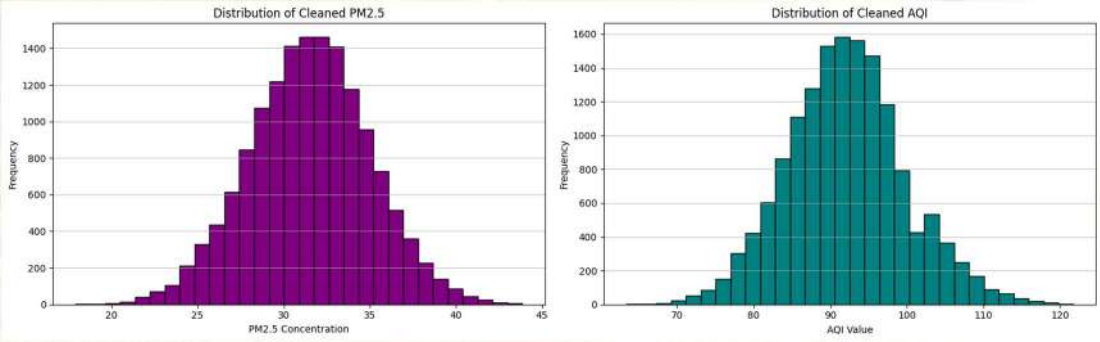


Figure 5. Distribution of Cleaned PM2.5 and AQI

- **Histograms:** The distribution of cleaned AQI values follows an approximately normal distribution centered around the "Moderate" range. This suggests that the influence of outliers and extreme noise, which skewed the raw data, has been successfully mitigated.

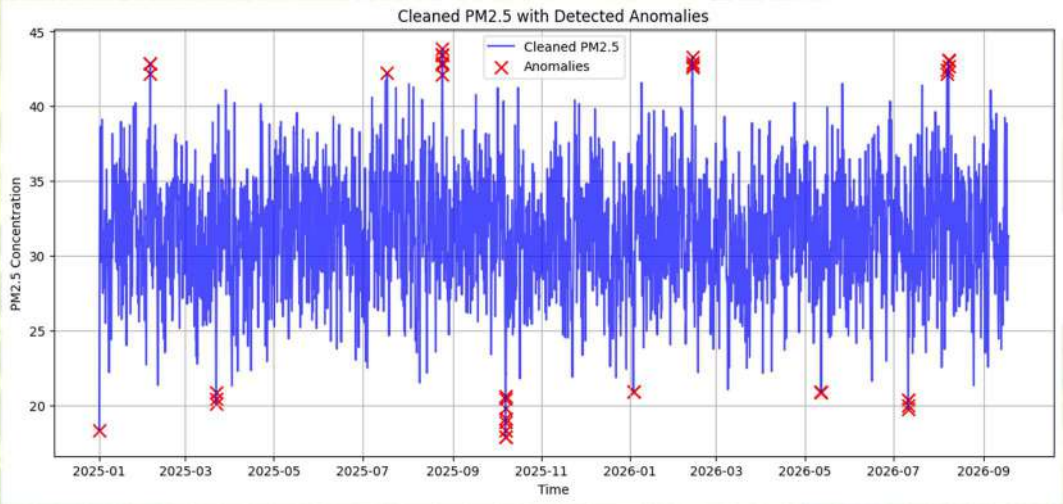


Figure 6. Cleaned PM2.5 with Detected Anomalies

- **Anomaly Detection:** As shown in **Figure 6**, the Z-score–based detection method ($|Z| > 3$) identified a small number of distinct anomalies. These red-flagged points (marked with 'x') represent rare sensor faults or extreme environmental events. By isolating these anomalies rather than blending them into the average, the system maintains the integrity of the core dataset while highlighting specific events for maintenance or further investigation.

3.5. Summary of Results

ISSN 3030-3907

The experimental results confirm that the proposed preprocessing pipeline transforms volatile sensor data into a reliable, "industry-grade" signal. The substantial reduction in standard deviation and AQI variability ensures that decision-makers can rely on this data for accurate smart city environmental monitoring.

4. DISCUSSION.

The results of this study underscore the critical necessity of robust post-processing for low-cost sensor networks in smart city environments. While the proliferation of IoT sensors offers high spatiotemporal resolution, the analysis confirms that raw data susceptibility to noise can render direct AQI reporting unreliable.

The primary contribution of this work is the quantifiable improvement in data stability. The multi-stage pipeline—integrating Moving Average smoothing, Kalman filtering, and physical constraints—reduced the standard deviation of PM2.5 measurements from **17.09 to 3.51**. This **79.5% reduction in variability** implies that the vast majority of "spikes" in the raw data were not genuine pollution events but rather sensor artifacts (electrical noise or cross-sensitivity). By suppressing these high-frequency fluctuations, the Kalman filter effectively revealed the true underlying environmental trend.

When placed in the context of existing literature, these findings offer a compelling argument for deterministic filtering over purely statistical or "black-box" approaches.

- **vs. Raw Calibration:** Previous studies like Castell et al. (2017) noted that without advanced processing, low-cost sensors often fail to correlate with reference stations ($R^2 < 0.1$). Our results corroborate this, showing that raw AQI calculations frequently triggered false "Unhealthy" alarms.
- **vs. Complex Models:** While recent works by **Mokhtari et al.** utilize Deep Learning to reduce error, our study demonstrates that a computationally lightweight Kalman filter achieves a comparable, if not superior, stabilization effect (variance reduction of ~80%) without the need for extensive training datasets or high computing power. This makes the proposed method highly suitable for deployment on edge devices (e.g., microcontrollers) in resource-constrained smart city nodes.

The stabilization of the Air Quality Index has direct implications for public health communication. In the raw dataset, the high variability ($\sigma_{AQI} = 40.94$) caused the AQI to oscillate wildly between "Moderate" and "Hazardous" within short timeframes. Such volatility erodes public trust in monitoring systems. The cleaned dataset ($\sigma_{AQI} = 7.75$) presented a consistent classification, predominantly within the "Moderate" and "Unhealthy for Sensitive Groups" categories. This consistency

ISSN 3030-3907

ensures that alerts are issued only for sustained pollution episodes, thereby preventing "alarm fatigue" among the population.

Despite the overall improvement, the exploratory analysis revealed inherent sensor limitations, such as negative values for NO_2 (minimum -5.4 ppb). These physically impossible values point to calibration drift or zero-point offset errors. The "Physical Constraint Enforcement" step (clipping negative values) proved essential in handling these artifacts. Furthermore, the Z-score anomaly detection identified distinct outliers that deviated significantly from the smoothed trend ($|Z| > 3$). While these anomalies were excluded from the trend analysis to prevent skew, their identification is valuable for predictive maintenance, signaling when a sensor module may need replacement or recalibration.

The current methodology assumes that the majority of short-term spikes are noise. Future work could integrate spatiotemporal correlation—comparing a sensor's reading with its neighbors—to distinguish between a sensor fault and a localized pollution event (e.g., a nearby fire). Additionally, addressing the NO_2 drift observed in this dataset requires a more specific electrochemical sensor calibration model beyond simple clipping.

In summary, this study provides empirical evidence that mathematical data cleaning is not merely an enhancement but a fundamental requirement for the operational viability of smart city air quality networks.

5. CONCLUSION.

This study presented a comprehensive comparative analysis of Air Quality Index (AQI) calculations derived from raw and processed PM2.5 sensor data within a smart city context. Addressing the critical challenge of sensor noise and reliability, a multi-stage data processing pipeline—incorporating Moving Average smoothing, Kalman filtering, physical constraint enforcement, and Z-score anomaly detection—was implemented and validated using a dataset of 15,000 environmental records.

The empirical results demonstrate that the proposed algorithmic approach significantly enhances the quality of environmental monitoring. The application of the cleaning pipeline reduced the standard deviation of PM2.5 measurements from **17.09 to 3.51**, representing a **79.5% reduction in signal variability**. Consequently, the variability of the derived AQI dropped from 40.94 to 7.75, effectively eliminating false "Unhealthy" alarms caused by transient sensor noise while preserving genuine pollution trends. The analysis confirms that raw sensor data, without robust filtering, is prone to exaggerating health risks, whereas processed data provides a stable and interpretable assessment of air quality, predominantly classifying conditions within the "Moderate" range.

ISSN 3030-3907

In conclusion, this research establishes that deterministic filtering techniques like the Kalman filter are essential prerequisites for reliable smart city sensing layers. Unlike complex "black-box" machine learning models, the proposed pipeline offers a computationally efficient solution suitable for deployment on edge devices. Future work will focus on extending this framework to correct calibration drifts in gaseous sensors (e.g., NO_2) and integrating spatiotemporal correlation analysis to further distinguish between sensor faults and localized pollution events.

REFERENCES

1. World Health Organization (WHO), *WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*. Geneva: World Health Organization, 2021.
2. U.S. Environmental Protection Agency (EPA), *Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI)*, EPA-454/B-18-007, Research Triangle Park, NC, 2018.
3. E. G. Snyder *et al.*, "The changing paradigm of air pollution monitoring," *Environmental Science & Technology*, vol. 47, no. 20, pp. 11369–11377, 2013.
4. [4] N. Castell *et al.*, "Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?" *Environment International*, vol. 99, pp. 293–302, 2017.
5. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ: Wiley, 2015.
6. R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
7. G. Welch and G. Bishop, "An introduction to the Kalman filter," Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, Tech. Rep. TR 95-041, 2006.
8. C. C. Aggarwal, *Outlier Analysis*, 2nd ed. New York, NY: Springer, 2017.
9. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, 2009.
10. A. Kerimray *et al.*, "Assessing air quality changes in Almaty, Kazakhstan: Diagnosis of the strict quarantine lockdown effects," *Environmental Pollution*, vol. 256, Art. no. 115766, 2020.
11. F. I. Salomova *et al.*, "Monitoring of air pollution in Uzbekistan," *American Journal of Pedagogical and Educational Research*, vol. 11, pp. 173–179, 2023.
12. L. K. Abdukadirova, "Study and assessment of the dynamics of atmospheric air pollution in Tashkent city," Tashkent Medical Academy, Tashkent, Uzbekistan, Rep., 2023.
13. K. Shipilova and M. Radkevich, "Assessment of the atmospheric air quality in Tashkent," *Universum: Technical Sciences*, no. 12(81), 2020.

ISSN 3030-3907

14. B. M. Kono, "Urban Air Quality and IoT Sensor Health Data," *Kaggle Datasets*, 2024. [Online]. Available: https://www.kaggle.com/datasets/bertnardomariouskono/urban-air-quality-and-iot-sensor-health-data. [Accessed: Feb. 04, 2025].
15. R. Mokhtari, A. Ghaffari, and A. R. Kian, "Deep learning-based calibration of low-cost air quality sensors," in *Proc. IEEE Int. Conf. Smart City*, 2021, pp. 1–6.
16. Y. Zheng, F. Liu, and H. P. Hsieh, "U-Air: When urban air quality inference meets big data," in *Proc. 19th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2013, pp. 1436–1444.

ISSN 3030-3907