

An Online Framework for Quality Control, Anomaly Detection and Auto-Calibration of Hourly Air-Quality Data from an IoT Sensor Node

Minas D., Tsiakmakis K., Hatzopoulos, A.T., Papadopoulou M. and Vassios V.

Department of Information and Electronic Engineering
International Hellenic University, Sindos, Thessaloniki, Greece

Abstract— This work presents an end-to-end pipeline designed to improve the quality and trustworthiness of hourly air-quality data collected by a multi-sensor IoT node. The node measures temperature, relative humidity, particulate matter (PM2.5 and PM10), carbon monoxide (CO), carbon dioxide (CO2), nitrogen oxides (NO, NO2) and UV index on a university campus. A Python backend streams the measurements into a database and drives a web dashboard that runs rule-based quality control (QC), multivariate anomaly detection, time-series analysis and short-term forecasting in near real time.

The system operates in two complementary modes. In online mode, the incoming hourly stream is continuously analyzed and visualized with QC flags, anomaly markers, daily statistics, intraday profiles and 24-hour-ahead forecasts. In offline mode, the same tools are applied to a complete stored period, together with a robust, data-driven auto-calibration step that targets spikes, heavy anomalies and small gaps in the sensor signals by replacing only clearly abnormal samples with median-based estimates that are consistent with the typical behavior of the node.

A month-long pilot deployment demonstrates that the proposed approach substantially reduces the impact of rare but severe anomalies on PM2.5 statistics while preserving the underlying diurnal pattern. The overall workflow is lightweight, transparent and well suited both for small-scale deployments that require reliable air-quality information and for educational environments where students can interact with live data, explore anomalies and directly observe their effect on statistics, forecasts and decision support.

Index Terms— Environmental monitoring, IoT sensors, anomaly detection, quality control, auto-calibration, data cleaning.

I. INTRODUCTION

Environmental sensor nodes have become an attractive option for smart cities, campuses and citizen science initiatives, because they provide dense spatial coverage of key air-quality indicators at a fraction of the cost of regulatory stations. However, the resulting data streams are often affected by noise, calibration drift, missing values, communication issues and occasional extreme outliers. If we don't deal with these effects in a careful, structured way, they can give us the wrong picture of how much pollution there really is and make our analytics and forecasting models work much worse than they should. Recent work has highlighted the importance of anomaly detection and repairing of air-quality data, to improve the reliability of monitoring systems and downstream analyses [1]. At the same time, several studies have focused on the design of sensing infrastructures and IoT-based air-quality stations that can be deployed in urban environments and integrated with cloud or fog-computing platforms [2], [3]. Other contributions have analyzed the behavior of air-quality sensors in detail,

discussing their biases, calibration challenges and the need for transparent correction procedures and comparability frameworks [4], [5]. Modern data-analytics techniques from rule-based quality control to machine-learning-based anomaly detection and time-series decomposition offer powerful tools for extracting information from such sensor streams, but they must be combined into clear and transparent workflows that operators and students can understand and reproduce. In this work we focus on a single multi-parameter air-quality node that continuously monitors temperature, humidity, particulate matter, gaseous pollutants and UV index at hourly resolution on the campus of our university. During the month of September, a Python/Flask backend receives the measurements, stores them and feeds an interactive web dashboard that runs QC rules, multivariate anomaly detection, short-term forecasting and an AI-inspired auto-calibration module. The key idea is that the system operates in two complementary modes: online, where it processes and visualizes the incoming stream in real time, and offline, where it re-analyses a full period (e.g. one month) to correct anomalies, compute indicators and evaluate performance. At the end of the month, we use the offline analysis to quantify how much the pipeline reduces the impact of rare but severe anomalies on statistics and forecasts and to prepare summarized indicators and figures that can be used for reporting, teaching and further research. In this work, we use the term auto-calibration to denote an automatic, data-driven correction step that targets clearly abnormal samples in the sensor time series (extreme spikes, heavy anomalies and small gaps). Instead of relying on an external reference instrument, the procedure exploits quality-control rules, machine-learning anomaly flags and typical hourly profiles of the sensor to replace only the problematic values with robust, median based estimates that are consistent with the usual behavior of the node.

II. EXPERIMENTAL SETUP

The environmental node integrates sensors for temperature, relative humidity, particulate matter (PM2.5, PM10), carbon monoxide (CO), carbon dioxide (CO2), nitrogen oxides (NO, NO2) and UV index, which are connected through dedicated electronic signal-conditioning circuits. These include analogue front-end stages for amplification, filtering and level shifting, as well as protection and impedance-matching networks tailored to each sensor type, followed by analogue-to-digital conversion and a DFRobot FireBeetle 2 ESP32-E (dual-core ESP32 with integrated Wi-Fi and Bluetooth) microcontroller unit that handles local preprocessing, timing and packaging of the measurements. The microcontroller aggregates the sensor data every hour, timestamps them in local time and transmits them via a lightweight communication stack to a Python

backend through a simple API for further processing and visualization. The backend ingests the measurements, converts them to a uniform hourly time axis and stores them in a database. Figure 1 illustrates the overall experimental setup and system architecture, including the sensor node, the communication link, the Python/Flask backend and the web dashboard used for online monitoring and offline analysis.

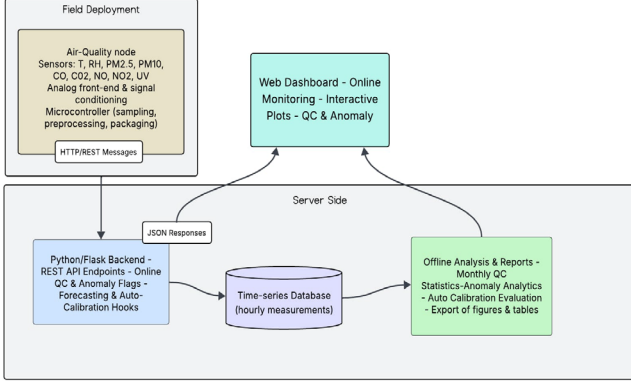


Fig. 1. Experimental setup and system architecture of the air-quality node, communication link, Python/Flask backend and web dashboard used for online monitoring and offline analysis.

For each new hour, the web dashboard updates the main time-series plot, the QC markers, the anomaly scores, the daily summary tables, the intraday profiles and the short-term forecasts. This reproduces the realistic situation where an operator or a group of students monitors a single node in real time and wants to understand what the sensors report and when they misbehave. Over the month of September, we collected 720 hourly records from the node. In the uncalibrated stream the PM2.5 values, for example, range from 6.2 to 203.6 $\mu\text{g}/\text{m}^3$ with a mean of about 21.0 $\mu\text{g}/\text{m}^3$, reflecting both typical background conditions and occasional extreme episodes. After applying the auto-calibration step described below, the corresponding monthly statistics for PM2.5 become much more stable, with values between 6.9 and 28.4 $\mu\text{g}/\text{m}^3$, mean 17.2 $\mu\text{g}/\text{m}^3$ and standard deviation 3.7 $\mu\text{g}/\text{m}^3$.

III. EXPERIMENTAL RESULTS

The analytics pipeline combines rule-based quality control, multivariate anomaly detection, time-series analysis and a robust auto-calibration module, all applied on the live hourly data stream from the node. Fig. 2 shows the main online time-series view and displays the raw measurements together with rule-based QC markers and ML anomaly flags for the selected sensor; this allows the operator to immediately see when values leave their typical range or when the multivariate behavior becomes unusual.

The rule-based QC layer applies simple physical limits, rate-of-change and flatline rules and consistency checks between variables. For example, UV values above 15 or temperatures above 50 $^{\circ}\text{C}$ are flagged as out-of-range, while constant readings over many hours are marked as potential sensor lockups.

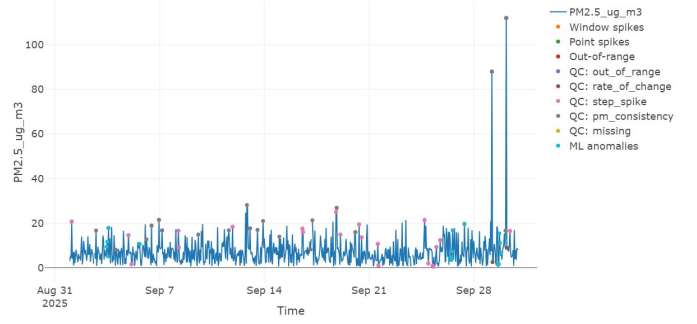


Fig. 2. Online time-series view of hourly PM2.5 over the study month, as seen by the dashboard on the live sensor data stream, with rule-based QC markers and ML anomaly flags overlaid on the raw measurements.

Daily QC statistics are summarized through a heatmap and a simple Quality Index (QI), defined as the mean number of rule violations per sample (Fig. 3). Days with large temperature and UV excursions clearly stand out with higher QC burden, even though most hours remain within normal limits.

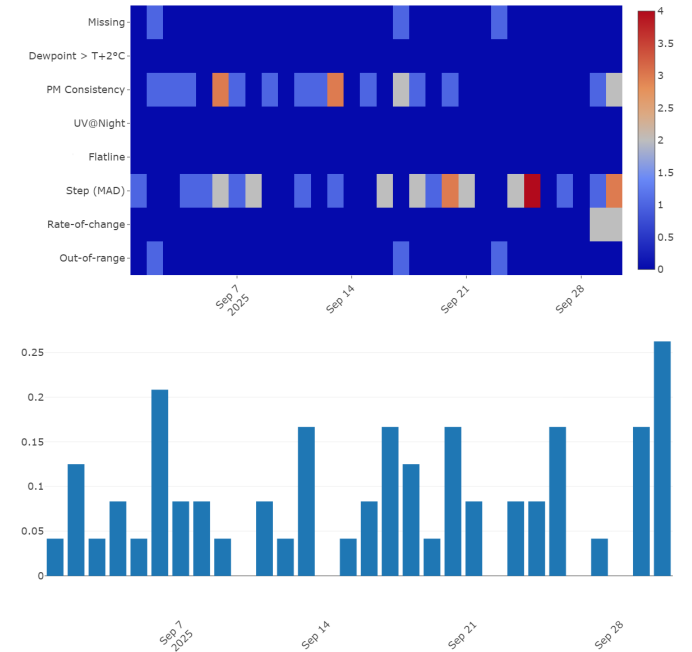


Fig. 3. Daily summary of rule-based QC results: heatmap of rule violations per day and Quality Index (QI) showing days with heavier QC burden due to large excursions in the measurements.

On top of these rules, the system computes multivariate anomaly scores using an Isolation Forest model that takes as input the current values of all sensors, time-of-day features and short rolling-window statistics. The model is trained in an unsupervised way on the same month of data and returns an anomaly score and a binary flag for each hour. This allows the dashboard to highlight not only obvious spikes but also unusual combinations of variables, such as high particulate matter without corresponding changes in gases. The evolution of

anomaly scores and the corresponding daily counts of flagged anomalies for PM2.5 are shown in Fig. 4, where clusters of anomalous behavior appear as clear bursts in the score time series and in the daily anomaly counts.

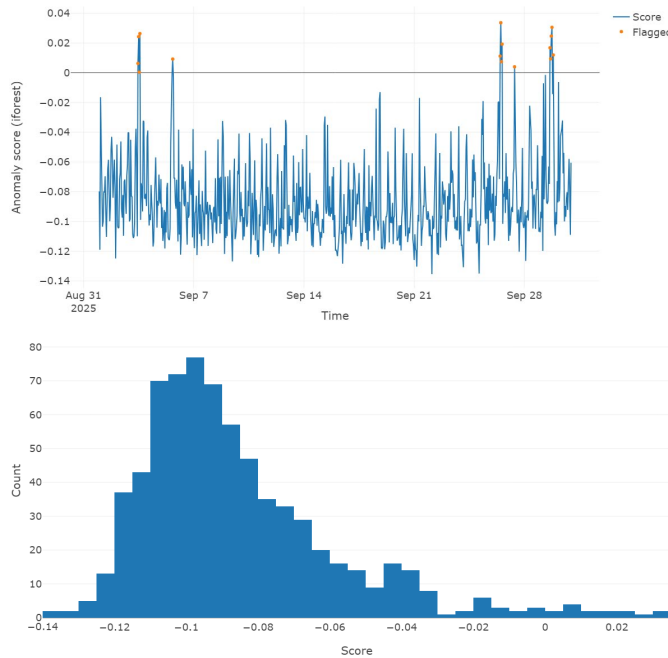


Fig. 4. Multivariate anomaly analytics for PM2.5: evolution of Isolation Forest anomaly scores and daily counts of flagged anomalies that highlight clusters of unusual behavior in the live data stream.

To better understand the relationships between variables, the system also computes a correlation heatmap over the main sensors (temperature, relative humidity, PM2.5, PM10, CO, CO2, NO, NO2 and UV index).

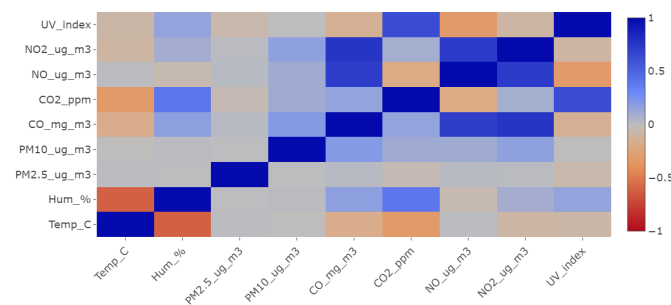


Fig. 5. Correlation heatmap between the main sensors (temperature, relative humidity, PM2.5, PM10, CO, CO2, NO, NO2 and UV index), showing the strength of pairwise linear relationships over the study period.

Figure 5 provides an at-a-glance summary of how strongly the different parameters co-vary over the study period, for example highlighting the strong association between PM2.5 and PM10 or between NO and NO2, as well as weaker or negligible correlations for variables that are driven by different processes.

Such information is useful both for interpretation of events and for guiding the design of multivariate models.

For time-series analysis we apply STL decomposition and PELT-based change-point detection to the hourly PM2.5 series, separating trend, diurnal seasonality and residuals. We also fit a simple ARIMA/SARIMAX model to generate 24-hour forecasts and evaluate them in a rolling back test over the last 48 hours. In the raw stream the forecast errors are dominated by a small number of extreme PM2.5 spikes, which inflate the RMSE and distort the residuals. To address this, we apply a robust auto-calibration step on the same series. The algorithm marks as bad all samples that are out of physical range, flagged by the ML detector, missing or strong outliers relative to a seven-sample rolling median. Clean samples are used to compute a global median and a median per hour-of-day. Each bad sample is then replaced by the median of clean samples at the same hour-of-day, or by the median in a local ± 6 -hour window, with a fallback to the global median, and values are clipped to the physical range of the sensor. Applied to PM2.5, this procedure modifies 113 out of 720 hourly samples. In the raw stream the monthly standard deviation of PM2.5 is $23.6 \mu\text{g}/\text{m}^3$ with 55 hours above $25 \mu\text{g}/\text{m}^3$ and 18 extreme episodes above $100 \mu\text{g}/\text{m}^3$ and after auto-calibration the standard deviation drops to $3.7 \mu\text{g}/\text{m}^3$, the number of hours above $25 \mu\text{g}/\text{m}^3$ falls to 13 and there are no episodes above $100 \mu\text{g}/\text{m}^3$. Visual inspection of the corrected series confirms that the main diurnal pattern and normal day-to-day variability are preserved, while unrealistic spikes and gaps are replaced by values that are consistent with the typical behavior of the node at the same time of day as shown in Figure 6.

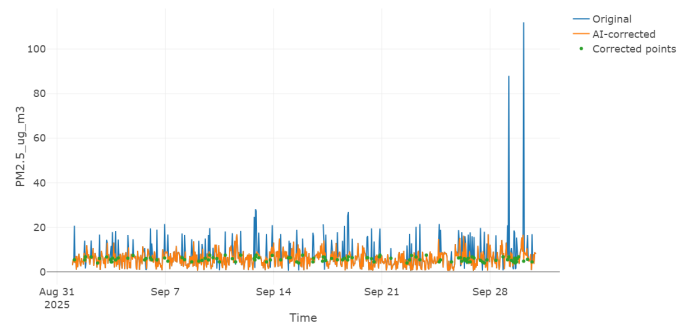


Fig. 6. Auto-calibration of the raw PM2.5 series, showing how robust median based corrections remove unrealistic spikes and gaps while preserving the underlying diurnal pattern and typical variability.

In addition to the monthly statistics, we also evaluated the effect of the auto-calibration step against a reference version of the dataset. For PM2.5, considering the 717 overlapping hourly samples, the RMSE between the raw measurements and the reference is $13.04 \mu\text{g}/\text{m}^3$, while the RMSE after auto-calibration is $12.69 \mu\text{g}/\text{m}^3$. The small change in RMSE reflects the fact that the bulk of the time series remains unaffected, whereas a very small subset of hours is heavily corrected. At the same time, the number of spike hours above the anomaly threshold used by the dashboard decreases from 6 to 0, confirming that the proposed procedure is effective at suppressing isolated extreme events

without distorting the overall diurnal pattern and background variability.

IV. CONCLUSION

This work demonstrates an integrated yet lightweight system for online quality control, anomaly analytics and auto-calibration of hourly air-quality data from a single IoT node. During a month-long deployment the system operates online, receiving and visualizing the live stream of measurements, flagging rule-based and multivariate anomalies and providing short-term forecasts. At the end of the period the same tools are applied offline to a complete dataset, together with a robust auto-calibration step and a quantitative evaluation of how the corrections affect statistics and episode counts. For the PM_{2.5} series the proposed procedure reduces the monthly standard deviation from 23.6 to 3.7 $\mu\text{g}/\text{m}^3$, reduces the number of hours above 25 $\mu\text{g}/\text{m}^3$ from 55 to 13 and removes all extreme episodes above 100 $\mu\text{g}/\text{m}^3$ while preserving the underlying diurnal pattern. These results show that even simple median based correction strategy, combined with QC and multivariate anomaly detection, can substantially improve the trustworthiness of air-quality data without introducing a complex black-box model. The same pipeline can be used both as a practical tool for small deployments that need basic but reliable air-quality information, and as an educational platform where students can experiment with live IoT data, inject anomalies and immediately observe how data quality affects statistics, forecasts and decision making. Future work will explore more advanced forecasting models, uncertainty quantification and the extension of the auto-calibration logic to networks of nodes.

References

- [1] F. Rollo, C. Bachechi and L. Po, "Anomaly Detection and Repairing for Improving Air Quality Monitoring," *Sensors*, vol. 23, no. 2, p. 640, 2023, doi: 10.3390/s23020640.
- [2] L. García, J. Lloret, P. Lorenz and J. Tomás, "Smart Air Quality Monitoring IoT-Based Infrastructure for Smart Cities," *Sensors*, vol. 22, no. 23, p. 9221, 2022.
- [3] A. A. Rahmadani, I. K. Wijaya and N. R. Hidayat, "Real-Time Air Quality Monitoring Through IoT and Web Technologies," *Instruments*, vol. 14, no. 1, p. 2, 2024.
- [4] I. Stergiou, A. Kleanthous and P. Kassomenos, "Bias Correction, Event Detection, and Temporal Pattern Discovery for Air Quality Forecast Ensembles," *Atmosphere*, vol. 16, no. 6, p. 739, 2025.
- [5] S. Shahid, M. A. Raza, M. A. Khan and A. Qamar, "Innovations in Air Quality Monitoring: Sensors, IoT and Emerging Technologies," *Sensors*, vol. 25, no. 7, p. 2070, 2025.