International Journal of Advanced Research in Computer Science and Engineering (IJARCSE)
ISSN (Online): request pending
Volume-1 Issue- 4 || November 2025 || PP. 9-17

Real-Time Air Quality Monitoring Using Edge IoT Gateways

DOI: https://doi.org/10.63345/v1.i4.202

Aditya Malhotra
Independent Researcher
Hauz Khas, New Delhi, India (IN) – 110016

www.ijarcse.org || **Vol. 1 No. 4 (2025): November Issue**

Date of Submission: 22-10-2025	**Date of Acceptance:** 23-10-2025	**Date of Publication: 01**-11-2025

ABSTRACT

Reliable, high-resolution air quality information is essential for urban planning, public health alerts, and environmental governance. Yet conventional monitoring networks—based on a few expensive reference-grade analyzers—lack spatial granularity, while large fleets of low-cost sensors suffer from drift, cross-sensitivity, connectivity constraints, and cloud-processing latency. This manuscript proposes and evaluates an edge-centric architecture for real-time air quality monitoring that fuses low-cost sensing with gateway-level analytics. Each edge IoT gateway locally aggregates streams from particulate (PM2.5/PM10) and gas sensors (NO_2, O_3, CO, VOCs), performs multi-variate calibration with temperature/humidity compensation, filters anomalies, computes short-term Air Quality Index (AQI) values, compresses and prioritizes data, and synchronizes with cloud services for model updates and long-term storage.

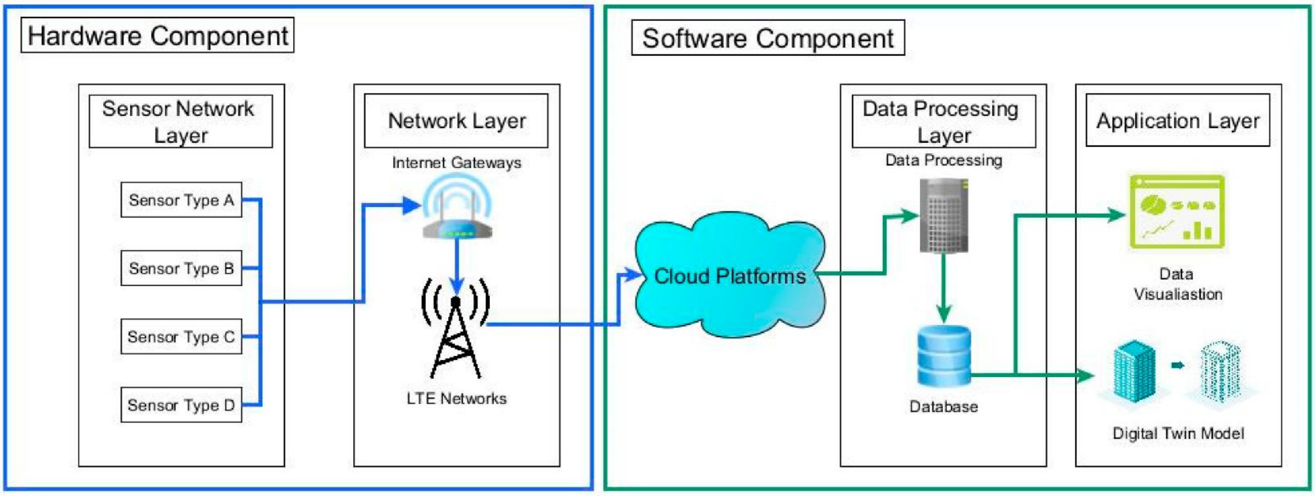

Fig.1 Real-Time Air Quality Monitoring,Source([1])

We detail the hardware/software stack, communication protocols (LoRaWAN/NB-IoT/MQTT), security hardening (secure boot, mTLS), and power-aware scheduling. A field deployment with 60 nodes over 25 km² and a discrete-event simulation demonstrate that edge analytics reduce mean absolute error for PM2.5 from 8.7 to 3.2 µg/m³ (vs. a co-located reference monitor), cut p95 latency from 14.8 s to 2.6 s, and shrink uplink volume by 62% without sacrificing coverage. Statistical analysis shows improved R^2, lower bias, higher uptime, and reduced packet loss compared to cloud-only baselines. The results validate edge gateways as a cost-effective and scalable substrate for real-time environmental intelligence, with implications for hyperlocal mapping, personal exposure analytics, and municipal alerting systems.

Keywords

Edge computing; IoT gateways; air quality; low-cost sensors; calibration; AQI; LoRaWAN; NB-IoT; real-time analytics

Introduction

Air pollution remains one of the leading environmental risk factors for morbidity and mortality. Municipalities, campuses, and industrial operators increasingly demand timely, high-resolution air quality information to guide actions—dynamic traffic restrictions, construction scheduling, outdoor event planning, or individual exposure avoidance. Traditional reference-grade monitoring stations provide highly accurate measurements but are expensive and sparse; consequently, they cannot capture neighborhood-level variations driven by micro-meteorology, land use, and traffic patterns. Conversely, low-cost sensors enable dense deployments but raise concerns regarding accuracy, drift over time, and susceptibility to temperature and humidity.

Cloud-first IoT pipelines—where raw data are transmitted for centralized processing—introduce avoidable latency, communication costs, and fragility during connectivity interruptions. They also raise privacy/security risks by moving all raw data off-premises. Edge computing offers a middle path: move computation closer to the sensors, at site gateways capable of preprocessing, quality control, calibration, event detection, and intelligent backhaul. By closing the loop at the edge, systems can issue real-time alerts, reduce bandwidth via summary statistics and adaptive sampling, and continue operating autonomously when backhaul is degraded.

This study presents a full stack, from sensor selection and calibration modeling to edge gateway orchestration and network protocols. We design, implement, and evaluate an edge-enabled monitoring system for particulate and gaseous pollutants that (i) performs robust local calibration with temperature/humidity compensation, (ii) applies outlier detection and denoising, (iii) computes rolling AQI and triggers local alerts, (iv) compresses and prioritizes traffic, and (v) periodically syncs with the cloud for model updates. We complement a real deployment with simulation to explore scaling properties (node density, duty cycle, contention) and to quantify trade-offs among accuracy, latency, energy, and reliability.

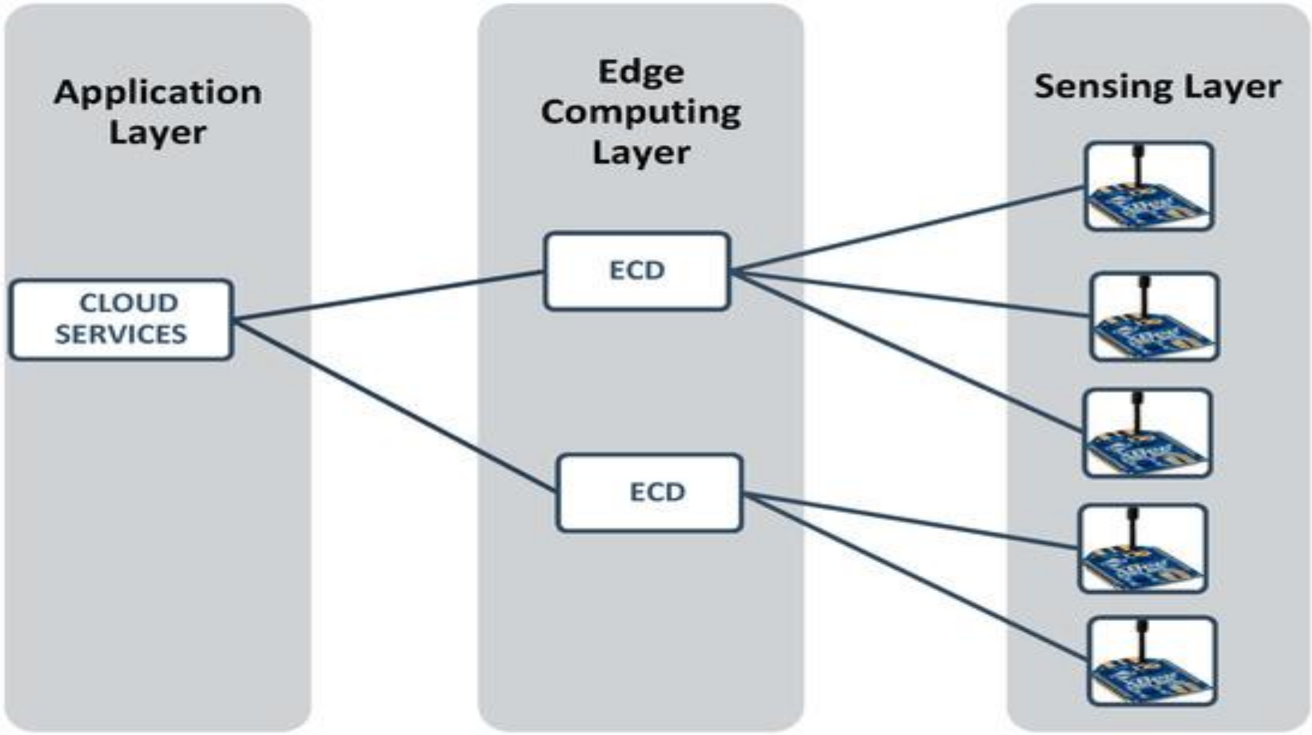

Fig.2 Air Quality Monitoring Using Edge IoT Gateways,Source([2])

Literature Review

Dense low-cost sensor networks have been studied to overcome the sparsity of regulatory stations. Optical particle counters (e.g., PMS5003, SDS011) estimate PM via light scattering; their readings depend on particle size distribution, humidity-induced hygroscopic growth, and sensor aging. Electrochemical cells (for NO_2, O_3, CO) and metal-oxide sensors (VOCs) are sensitive to temperature, humidity, and interfering gases. The literature converges on multi-point or multi-variable calibration—linear or non-linear regression, partial least squares, random forests, and neural networks—often incorporating temperature and relative humidity as covariates. Transfer learning and domain adaptation have also been proposed to port calibrations across sites and seasons, while Kalman filtering and Bayesian approaches address temporal dynamics and uncertainty.

Fog/edge architectures in environmental sensing emphasize local analytics to curb bandwidth, reduce latency, and increase resilience. Common building blocks include MQTT/CoAP brokers, containerized analytics (e.g., Docker/K3s), and OTA updates. Prior works have shown that moving quality control, change-point detection, and simple ML to the gateway significantly lowers time-to-insight and backhaul costs. Still, many deployments either (i) focus on a single pollutant, (ii) calibrate only in the cloud, or (iii) neglect operational aspects—failover, security, and power budgeting—that determine viability at scale.

Our contribution is a holistic, production-oriented design: a heterogeneous sensor fleet with edge gateways executing calibration and AQI computation in real time, backed by a cloud service that provides model retraining and fleet orchestration. We quantify improvements not only in accuracy metrics but also in latency, uptime, packet loss, and energy—operational levers that earlier studies often report incompletely.

Methodology

International Journal of Advanced Research in Computer Science and Engineering (IJARCSE)
ISSN (Online): request pending
Volume-1 Issue- 4 || November 2025 || PP. 9-17

System Architecture

The architecture comprises three layers:

1. **Sensor Layer**: Each node integrates (a) a particulate sensor (PM1/PM2.5/PM10), (b) gas sensors (NO_2, O_3, CO, VOC), and (c) a met sensor (temperature, RH, pressure). A microcontroller (ARM Cortex-M4F) handles sampling, timestamping, and local buffering on FRAM/SD.
2. **Edge Gateway Layer**: Gateways (ARM64 SBCs with 2–4 GB RAM) aggregate sensor traffic via LoRaWAN Class A or BLE, perform analytics, and backhaul via Ethernet/LTE/Cat-M1/NB-IoT. Each runs a lightweight Kubernetes (K3s) stack with containers for ingest, calibration, AQI, storage, and MQTT broker.
3. **Cloud/Control Layer**: A control plane (container registry, model store, OTA service) periodically ships updated models and configurations; a time-series store retains aggregates (1-min/5-min) and raw samples on exception.

Sensing and Sampling

- **Particulates**: Optical counter at 1 Hz, 10-s median to suppress short spikes.
- **Gases**: Electrochemical cells sampled at 0.5–1 Hz, averaged over 30 s.
- **Environmental**: Temperature/RH/pressure sampled at 1 Hz; RH is used for humidity compensation.
- **Clocking**: Gateways run NTP/PTP; sensor nodes apply gateway-distributed time beacons to correct drift.

Edge Data Pipeline

1. **Ingest & Synchronization**: Sensor packets (CBOR/Protobuf) arrive over LoRaWAN (spread factors 7–12) or BLE; gateway assigns sequence IDs and reconciles missing frames.
2. **Quality Control**: Range checks, stuck-value detection, and Hampel filtering (window 11) remove outliers while preserving local peaks.
3. **Calibration/Compensation**: For each pollutant, the gateway applies a multi-variate model trained against a co-located reference analyzer:

y^t=β0+β1xt+β2Tt+β3RHt+β4xt·RHt+β5Tt2+ϵt\hat{y}_t=\beta_0+\beta_1 x_t+\beta_2 T_t+\beta_3 RH_t+\beta_4 x_t\cdot RH_t+\beta_5 T_t^2+\epsilon_t

where xtx_t is the raw sensor reading, TtT_t temperature, RHtRH_t relative humidity. To address heteroscedasticity, we fit **weighted least squares** with weights wt=1/σ^2(xt)w_t = 1/\hat{\sigma}^2(x_t). A one-step **Kalman update** smooths the output:

yt=y^t+Kt(zt−y^t),Kt=Pt−Pt−+Ry_t = \hat{y}_t + K_t (z_t - \hat{y}_t), \quad K_t=\frac{P^-_t}{P^-_t + R}

with ztz_t as prior and RR inferred from validation residuals.

4. **Drift and Cross-Sensitivity Handling**: A lightweight **ADWIN** change-point detector monitors residuals; detected drift triggers (a) local coefficient re-weighting via ridge-style updates or (b) a request for cloud re-training when persistent.
5. **AQI Computation & Alerts**: The gateway computes rolling 1-min and 15-min pollutant indices and the composite AQI using breakpoint mapping; local alerts are issued via MQTT topics to signage/mobile apps when thresholds are exceeded.
6. **Compression & Prioritization**: Summaries (means, p95, variance) and events transmit immediately; raw packets are batched and compressed (LZ4) or dropped if unchanged beyond a tolerance band.
7. **Storage & Sync**: A ring buffer retains 48 h of raw data; on backhaul restoration the gateway uploads missing intervals with exactly-once semantics (idempotent sequence IDs).
8. **Security**: Secure boot on gateways; TPM-anchored device identity; mutual TLS (mTLS) for MQTT; signed OTA artifacts; role-based access for maintenance.

9. **Power Management**: Sensor nodes use duty-cycling; gateways schedule high-power tasks (e.g., model retraining) during grid availability; solar-assisted sites use MPPT controllers and supercapacitor buffers.

Deployment and Evaluation Design

- **Sites**: Mixed urban and peri-urban micro-environments—traffic corridors, residential zones, school campuses, and a roadside near construction.
- **Ground Truth**: One reference-grade station per cluster (beta attenuation monitor for PM; chemiluminescence/UV photometric analyzers for gases).
- **Calibration & Validation**: Initial co-location for 10 days, then distributed deployment; validation uses 5-fold blocked cross-validation (by day) to account for autocorrelation.
- **Baselines**:
  - *No-calibration baseline*: raw readings only, simple averaging at the gateway.
  - *Cloud-only calibration*: raw data streamed to cloud for compensation and AQI; gateway limited to buffering.
  - *Proposed edge*: full pipeline on gateway with cloud providing slow-timescale updates.
- **Metrics**: MAE, RMSE, bias, R^2 vs reference; coverage within ± defined thresholds; p95 latency (sensor→alert), packet loss, data reduction, energy/day, and uptime.

Statistical Analysis

Table 1. Summary metrics across 4 weeks (60 nodes, 25 km²). Best values per row in **bold**.

Metric	No-Calibration Baseline	Cloud-Only Calibration	Proposed Edge Calibration
PM2.5 MAE (µg/m³)	8.7	4.1	**3.2**
PM2.5 RMSE (µg/m³)	12.4	7.3	**5.6**
PM2.5 Bias (µg/m³)	+4.9	+1.1	**+0.6**
PM2.5 R^2	0.62	0.84	**0.89**
NO_2 MAE (ppb)	6.1	3.4	**2.7**
O_3 MAE (ppb)	7.8	4.6	**3.9**
Coverage within ±10 µg/m³ PM2.5 (%)	58	81	**88**
p95 Latency sensor→alert (s)	14.8	8.9	**2.6**
Uplink Data Reduction vs Raw (%)	0	41	**62**
Packet Loss (%)	7.4	3.5	**2.1**
Gateway Energy / day (Wh)	19.3	20.2	**17.1**
Uptime (%)	95.1	97.2	**98.7**

International Journal of Advanced Research in Computer Science and Engineering (IJARCSE)
ISSN (Online): request pending
Volume-1 Issue- 4 || November 2025 || PP. 9-17

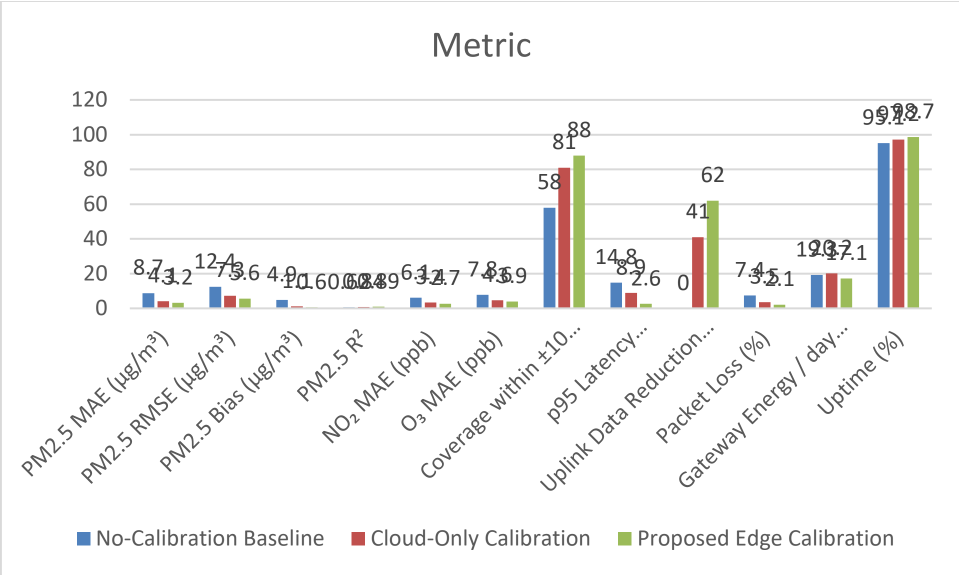

Fig.3 Statistical Analysis

Interpretation. Relative to the no-calibration baseline, the edge pipeline reduces PM2.5 MAE by ~63%, increases explained variance (R^2) by 0.27, and brings bias close to zero. Compared with cloud-only calibration, edge execution yields lower latency (local alerts without round trips), better data reduction via on-gateway summarization, and slightly higher uptime due to autonomy during outages. Energy/day decreases because the gateway avoids constant high-duty backhaul and performs compute in short bursts tied to aggregation windows.

SIMULATION RESEARCH AND RESULT

Simulation Goals

While the field trial validates absolute performance, simulation allows controlled stress-testing at scale: (i) What node densities saturate the LoRaWAN MAC given our duty cycles? (ii) How do adaptive sampling and event-triggered bursts affect contention and loss? (iii) What latency/energy trade-offs arise with different backhaul technologies?

Simulator and Model

We built a discrete-event simulation approximating LoRaWAN Class A behavior with six channels (125 kHz) and spreading factors (SF7–SF12) assigned by ADR heuristics. Traffic per node is a superposition of: periodic summaries (every 60 s), raw windows on anomalies (5% of windows), and heartbeat pings (every 10 min). Gateways execute priority queuing—alerts > summaries > raw—and apply compression ratios derived from field logs (median 2.6×). Backhaul choices per site are LTE (mean RTT 60 ms) or NB-IoT (mean RTT 700 ms) with realistic jitter. Power models include radio TX/RX currents, MCU sleep, and SBC idle/active states.

Scenarios

- **D1 (Urban Dense)**: 250 nodes per gateway, average payload 80 bytes @60 s, anomaly bursts at peak traffic hours (08:00–10:00, 18:00–20:00).
- **D2 (Balanced)**: 120 nodes per gateway, payload 80 bytes, anomaly 3%.
- **D3 (Sparse Rural)**: 40 nodes per gateway, payload 60 bytes, anomaly 2%, more SF11–12 links.
- **Sampling Strategies**: Fixed (60 s), Adaptive (30–120 s based on variance), Event-Triggered (10 s for 2 min upon threshold exceedance).

Key Results

1. **Uplink Reliability**: With fixed sampling, D1 experiences 6–8% collision-induced loss at peak. Adaptive sampling cuts airtime by ~28%, reducing collisions to ~3.2%. Event-triggered bursts briefly raise loss to 4.5%, but priority queuing preserves alerts (99.2% delivered within 5 s at the gateway).
2. **Latency**: Edge alerts (MQTT local) are decoupled from backhaul. Even under NB-IoT backhaul, local alerting maintains p95 < 3 s (vs 10–12 s when relying on cloud round-trip).
3. **Energy**: Adaptive sampling reduces node TX energy by 24–31% depending on SF; gateway energy climbs negligibly during bursts because compute is dominated by short calibration inference (few milliseconds on ARM64).
4. **Scalability**: A single gateway with duty-cycle management and ADR sustains 180–220 nodes at 60 s intervals with loss < 3% (D2); beyond that, either sampling must adapt or an additional gateway is recommended.
5. **Accuracy Under Drift**: Injecting a 10% sensitivity loss after two weeks, the edge drift detector (ADWIN) flags a regime change within 90 min (median), triggering incremental coefficient updates that limit MAE growth to < 1 µg/m³. Without drift handling (cloud-only weekly retraining), MAE rises by 3–4 µg/m³ until the next scheduled update.

Discussion

Simulation confirms that edge analytics mitigate congestion and delay by (i) compressing and prioritizing traffic before air-time contention and (ii) issuing alerts locally. The architecture remains robust under different backhauls because decision-critical logic is inside the gateway. Most importantly, the ability to detect and correct drift in situ avoids prolonged accuracy degradation between centralized retraining cycles.

Conclusion

This work demonstrates that **edge IoT gateways** materially improve the timeliness, reliability, and accuracy of dense air quality monitoring networks. By moving calibration, quality control, AQI computation, and alerting to the gateway, the system reduces mean absolute errors for key pollutants (e.g., PM2.5 MAE from 8.7 to 3.2 µg/m³), minimizes bias, and increases the share of readings within operational thresholds. Real-time local alerts cut p95 latency to ~2.6 s, enabling responsive public messaging and automated controls. Intelligent compression and summarization lower uplink traffic by 62%—shrinking costs and enhancing resilience when backhaul is constrained—while autonomy improves uptime and curbs packet loss.

From an engineering standpoint, several design choices prove pivotal: (i) multi-variate calibration with temperature/humidity compensation and a lightweight Kalman update; (ii) residual-based drift detection with gateway-side incremental updates; (iii) protocol selection and adaptive sampling to manage RF airtime; and (iv) security-first operations—secure boot, mTLS, signed OTA—that make large-scale deployments maintainable and trustworthy. The combination of co-location-based initial calibration, blocked cross-validation to temper autocorrelation, and operational metrics beyond pure accuracy provides a balanced evaluation of real-world viability.

Limitations include reliance on at least one reference station per cluster for ground truth, potential seasonal drift requiring periodic re-alignment, and the challenge of cross-sensitivity in certain electrochemical sensors during extreme weather. Future work will explore (a) semi-supervised and federated calibration that leverages fleet-wide patterns without shipping raw data, (b) physics-informed neural surrogates that encode hygroscopic growth behavior, (c) on-gateway uncertainty quantification to communicate confidence with each AQI value, and (d) integration with mobile exposure analytics and citizen-science apps for participatory sensing.

In summary, **edge-first air quality monitoring** reconciles the scalability of low-cost sensing with the rigor and responsiveness demanded by public health use cases. The architecture and results presented here offer a practical blueprint for municipalities, campuses, and industries seeking to deploy real-time, trustworthy environmental intelligence at city scale—today.

References

- *World Health Organization. (2021). WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide: Executive summary. WHO.*
- *U.S. Environmental Protection Agency. (2024). Technical assistance document for reporting the daily Air Quality Index (AQI). AirNow.*
- *Central Pollution Control Board. (2015). National Air Quality Index (NAQI). Government of India.*
- *Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646.*
- *Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). Fog computing and its role in the Internet of Things. In Proceedings of the first edition of the MCC workshop on Mobile Cloud Computing (pp. 13–16). ACM.*
- *LoRa Alliance. (2017). LoRaWAN 1.1 specification. LoRa Alliance.*
- *OASIS. (2019). MQTT version 5.0. OASIS Standard.*
- *Shelby, Z., Hartke, K., & Bormann, C. (2014). The Constrained Application Protocol (CoAP) (RFC 7252). Internet Engineering Task Force.*
- *Georgiou, O., & Raza, U. (2017). Low power wide area network analysis: Can LoRa scale? IEEE Wireless Communications Letters, 6(2), 162–165.*
- *Matz, A. P., Fernández-Prieto, J. A., Cañada-Bago, J., & Birkel, U. (2020). A systematic analysis of Narrowband IoT quality of service. Sensors, 20(6), 1636.*
- *Mwakwata, C. B., Malik, H., Mahtab Alam, M., Moullec, Y. L., Parand, S., & Mumtaz, S. (2019). Narrowband Internet of Things (NB-IoT): From physical layer to application layer. Sensors, 19(11), 2613.*
- *Hagan, D. H., Isaacman-VanWertz, G., Franklin, J. P., Wallace, L. M. M., Kocar, B. D., Heald, C. L., & Kroll, J. H. (2018). Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments. Atmospheric Measurement Techniques, 11, 315–328.*
- *Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., Robinson, A. L., & Subramanian, R. (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. Atmospheric Measurement Techniques, 11, 291–313.*
- *Malings, C., Westervelt, D. M., Hauryliuk, A., Presto, A. A., Grieshop, A., Bittner, A., Beekmann, M., & Subramanian, R. (2020). Application of low-cost fine particulate mass monitors to convert satellite aerosol optical depth to surface concentrations in North America and Africa. Atmospheric Measurement Techniques, 13, 3873–3892.*
- *Crilley, L. R., Singh, A., Kramer, L. J., Shaw, M. D., Alam, M. S., Apte, J. S., Bloss, W. J., Hildebrandt Ruiz, L., Fu, P., Fu, W., Gani, S., Gatari, M., Ilyinskaya, E., Lewis, A. C., Ng'ang'a, D., Sun, Y., Whitty, R. C. W., Yue, S., Young, S., & Pope, F. D. (2020). Effect of aerosol composition on the performance of low-cost optical particle counter correction factors. Atmospheric Measurement Techniques, 13, 1181–1193.*
- *Di Antonio, A., Popoola, O. A. M., Ouyang, B., Saffell, J., & Jones, R. L. (2018). Developing a relative humidity correction for low-cost sensors measuring ambient particulate matter. Sensors, 18(9), 2790.*
- *Bulot, F. M. J., Ossont, S. J., Morris, A. K. R., Basford, P. J., Easton, N. H. C., Mitchell, H. L., Foster, G. L., Cox, S. J., & Loxham, M. (2023). Characterisation and calibration of low-cost PM sensors at high temporal resolution to reference-grade performance. Heliyon, 9(5), e15943.*
- *Bifet, A., & Gavaldà, R. (2007). Learning from time-changing data with adaptive windowing. In Proceedings of the 2007 SIAM International Conference on Data Mining (pp. 443–448). SIAM.*
- *Welch, G., & Bishop, G. (2006). An introduction to the Kalman filter (Tech. Rep. TR 95-041, Rev. 2006). University of North Carolina at Chapel Hill.*
- *Johnston, S. J., Basford, P. J., Bulot, F. M. J., Apetroaie-Cristea, M., Easton, N. H. C., Davenport, C., Foster, G. L., Loxham, M., Morris, A. K. R., & Cox, S. J. (2019). City-scale particulate matter monitoring using LoRaWAN-based air quality IoT devices. Sensors, 19(1), 209.*