

An IOT-Based Environmental Monitoring System Using Adaptive Filtering Techniques

Dr. Pallavi Sharma¹, Rajni Bedi²

¹Assistant Professor, School of Engineering, Design and Automation- E, Department of ECE,
GNA University, Phagwara, Punjab, India

²Assistant Professor, Department of CSE, Lyallpur Khalsa College Technical Campus, Jalandhar, Punjab,
India

E-mail: pallavi.sharma@gnauniversity.edu.in, rajnibedi@lkcengg.edu.in

Abstract: Rapid urban development and growth of urban centres worldwide, combined with changing climate conditions have created an urgent need for improved environmental monitoring systems that are able to provide accurate assessment of environmental conditions in real-time. An Environmental Monitoring System based on the Internet of Things (IoT) is proposed in this paper, and it employs adaptive filtering techniques to increase accuracy of sensor data. The system combines inexpensive environmental sensors measuring air quality, temperature, humidity and light intensity. To reduce system noise and inaccuracies caused by the environment, a combination of adaptive filtering techniques (Least Mean Squares (LMS) and Kalman filters), will be applied to sensor data at the point where measurements are made. The main contribution made by this work is the development of a Hybrid Adaptive Filter (HAF) that automatically switches between the use of LMS and Kalman filtering based on real-time variance of the input signal to achieve improved noise reduction (i.e., mean square error reduction of 58%, signal-to-noise ratio increase of 15 dB) while maintaining negligible additional power consumption. A new Environmental Health Index (EHI) is also proposed to create an actionable score from all the different sensors collecting data on multiple parameters. To demonstrate the success of these techniques, experiments over 14 days were conducted at five geographically dispersed urban sites, and data indicate that both techniques are effective in reducing noise (improved reliability of collected data), and improving both sensor and power efficiency for IoT systems. This work creates a pathway to the development of scalable, thermally and energy-efficient environmental monitoring systems, which may support smart cities, agriculture, and industrial operations.

Keywords: IoT, adaptive filtering, Kalman filter, LMS, HAF, EHI, environmental monitoring, signal processing, smart cities, noise reduction.

1. Introduction

The monitoring of the environment is among the foremost challenges of this century regarding accelerated industrialization and urbanization combined with increasing effects of climate change, requiring robust, scalable, real-time systems for monitoring environmental conditions [1]. IoT can provide a transformative solution with highly dense, low-cost, wireless networked sensors capable of continuously sensing, transmitting and processing the environmental data.

However, although there is considerable promise for the IoT-based sensing of the environment, the raw sensor data from low-cost environmental sensors can often contain corruption due to, for example, noise, temperature drift, electromagnetic interference and aging [2]. Adaptive filtering techniques are one type of signal processing used to separate the desired environmental signal from the noise present in the IoT sensor deployment, providing a principled basis for doing so.

In this paper, we present a complete IoT-based system for monitoring the environment using two traditional forms of adaptive filtering — the Kalman filter and the Least Mean Squares (LMS) filter — along with a new HAF architecture developed for this research. In addition to presenting these findings, we introduce an EHI, which is a composite score based on multiple parameter readings that can then be used as a single actionable number/metric for public or policymaking purposes [3,4]. This system incorporates the use of ESP32 microcontroller-based sensor networks to monitor air quality (CO₂ and PM 2.5), temperature, humidity and light levels, which feed into a cloud-based analytic platform. The key contributions of this research paper are as follows:

- A comprehensive hardware/software architecture for an ESP-32-based multiparameter IoT environmental monitoring system with adaptive filtering at the node level.
- A novel HAF that selects dynamically between Kalman and LMS filtering algorithms based on real-time signal variances, significantly outperforming either Kalman or LMS filtering alone.
- EHI that consolidates the five environmental parameters into a single actionable score with colour-coded alert levels.
- Site stratified and diurnal performance assessment of each filtering method across 5 heterogeneous urban locations over 14 days.
- Quantitative measurement of power consumption and latency of each of the three filtering methods on embedded hardware.

The remaining portions of this research paper are organized as follows: Section two presents a review of related literature, Section three describes the system architecture and methodology, Section four provides experimental results, Section five discusses and analyses results and Section six concludes with a summary of conclusions and recommendations for future research.

2. Literature Review

2.1 IoT-Based Environmental Monitoring Systems

Since at least the beginning of the 2000s, researchers have been actively researching using IoT technology to monitor the environment. The first demonstrations of using wireless sensor networks (WSNs) for measuring things like temperature, humidity, and basic air quality in agriculture and industry went well [5]. Since then, low-power wireless protocols such as Zigbee, LoRaWAN and NB-IoT have come out, allowing WSNs to be used not just in local areas, but also now across an entire continent! Nearly every day that passes, new sensors are being developed that sense CO₂, PM_{2.5}, PM₁₀, many types of VOCs, and ozone (O₃) — providing many insights into air quality in cities across the globe. The development of platforms such as ThingSpeak, AWS IoT, and Microsoft Azure IoT Hub has brought to many developers an opportunity to monitor the environment by providing them with some form of cloud service for data ingestion, cloud analytics, cloud-based visualisation dashboards and authentication [7]. The problems are many however with a wide variety of low-cost sensors being deployed in a wide range of environments causing significant variability in data quality and comparability.

2.2 Noise and Signal Distortion in IoT Sensors

There are several types of distortions caused by inexpensive environmental sensors, which prevent the accurate measurement of data. These include thermal noise, EMI resulting from nearby co-located wireless devices, and sensor drift caused by temperature fluctuations (gradual shifts in calibration due to aging and/or contamination), which all lead to false systematic biases that develop over time (week -month continuous operation) [8,9]. In addition, cross-sensitivity occurs when a sensor detects interfering constituents unrelated to its intended analyte. If appropriate signal processing is not in place, this can cause consistent over- or under-estimation of values associated with contaminants in the environment [10].

2.3 Adaptive Filtering in Sensor Networks

Many researchers have looked into adaptive filtering to enhance the quality of signals in uncertain measurements; non-adaptive filters do not adapt well to the presence of noise [11]. The Kalman Filter, developed by Rudolf E. Kalman in 1960, is the earliest recursive Bayesian estimator that is optimal (minimum mean squared error) under the assumption of Gaussian noise [12]. LMS adaptive filter was introduced by Widrow and Hoff as a simple computationally manageable alternative in 1960 [13]. The normalised LMS (NLMS) algorithm scales the step size based on the input signal power, providing improved convergence stability [14]. The extended Kalman filter (EKF) and unscented Kalman filter (UKF) allow for better tracking of non-linear systems but at a much higher computational cost [15]. Although much of the literature has explored these areas, the authors of this paper did not find anything related to combining

Kalman and LMS adaptive filtering techniques into a single hybrid filter that chooses one or the other based on the current signal/statistics.

2.4 Cloud Integration and Data Management

Elastic storage, distributed analysis, and machine learning capabilities on cloud platforms complement the filtering of data at the edge (Mandy, 2016) [16]. The integration of edge processing with cloud computing (fog computing or edge-cloud collaboration) creates an opportunity for scalable, low-latency and resource-efficient IoT ecosystems (Morris et al., 2016) [17]. There is little in the scientific literature regarding composite environmental indices that are based on multi-parameter-filtered data streams from IoT devices. This work proposes the EHI composite environmental index to address this gap in the literature.

3. Methodology

3.1 System Architecture

The design of the system consists of 4 different hierarchical layers: (1) Sensing Layer; (2) Processing Layer using ESP32 microcontrollers; (3) Communications Layer using Wi-Fi; (4) Cloud Analytics Layer using AWS IoT connected with ThingSpeak integration [7]. By using this layered architecture, we can decouple functionality and upgrade each layer independently. Sensor Nodes include the following: MQ-135 (measuring CO₂/VOC), PMS5003 (measuring PM_{2.5}/PM₁₀), DHT22 (measuring Temperature/Humidity), and BH1750 (measuring Light Intensity).

3.2 Hardware Implementation

The ESP32 microcontroller was chosen due to its dual-core Xtensa LX6 architecture, built-in Wi-Fi and Bluetooth 4.2 support, as well as its wide variety of peripherals [18]. The dual-core design allows for processing of signal data on one core and wireless communication using Wi-Fi on another core; thus reducing the latency associated with performing both tasks on a single core. Each node has a 3.7V 2000mAh lithium polymer battery, and can charge via solar power and maintain accurate RTC (Real Time Clock) timestamps [19].

3.3 Adaptive Filtering Techniques

3.3.1 Kalman Filter

Each environmental parameter was modeled as a first-order linear dynamical system [12] and both the process noise covariance Q and measurement noise covariance R were empirically adjusted. The prediction and update equations are:

$$\begin{aligned}\hat{x}^-(k) &= A \cdot \hat{x}(k-1) \\ P^-(k) &= A \cdot P(k-1) \cdot A^T + Q \\ K(k) &= P^-(k) \cdot H^T \cdot [H \cdot P^-(k) \cdot H^T + R]^{-1} \\ \hat{x}(k) &= \hat{x}^-(k) + K(k) \cdot [z(k) - H \cdot \hat{x}^-(k)] \\ P(k) &= [I - K(k) \cdot H] \cdot P^-(k)\end{aligned}$$

3.3.2 LMS Filter

A 16-tap FIR structure was used to implement the LMS filter [13]. The weight updating equation is:

$$w(k+1) = w(k) + 2\mu \cdot e(k) \cdot x(k)$$

where 0.01 step size μ was obtained using grid search across [0.001, 0.1].

3.3.3 Novel Hybrid Adaptive Filter (HAF)

A major disadvantage of one algorithmic approach will be addressed by HAF. Kalman filtering provides excellent performance for low variance, slowly changing signals, but takes longer to adapt when an above-normal event happens. Whereas, LMS converges quickly, near the end of a transient period but at the expense of a higher amount of steady state error. Real-time variance-guided mode switching provided by the HAF algorithm offers an answer to this issue through the use of variance-based mode switching at each time step K . The HAF algorithm works as follows. For every time step k , the local signal variance $\sigma^2(k)$ is calculated using a sliding window of $W=20$ samples.

$$\sigma^2(k) = \frac{1}{W} \cdot \sum_{i=0}^{W-1} [z(k-i) - \mu(k)]^2$$

where $i = 0, \dots, W-1$ and $\mu(k)$ is the windowed mean. Variance is based on the values of two thresholds (θ_{low} and θ_{high}) used to separate three different operating modes (regimes) from a variance perspective:

1. **Kalman mode** ($\sigma^2(k) < \theta_{\text{low}}$): A quasi-stationary environment. A Kalman filter will be applied as Kalman provides a minimum mean square error (MMSE) estimate, given that the conditions are low-noise and variable slowly over time.
2. **LMS mode** ($\sigma^2(k) > \theta_{\text{high}}$): An abrupt environmental change has occurred (e.g., emissions from vehicles, HVAC startup, etc.). An LMS filter will be used to adapt to the new environment rapidly.
3. **Blended mode** ($\theta_{\text{low}} \leq \sigma^2(k) \leq \theta_{\text{high}}$): A weighted sum of the two estimated system outputs (from the Kalman; from the LMS) will be utilized:

$$y(k) = \alpha(k) \cdot y_{\text{Kalman}}(k) + [1 - \alpha(k)] \cdot y_{\text{LMS}}(k)$$

$$\alpha(k) = \frac{\theta_{\text{high}} - \sigma^2(k)}{\theta_{\text{high}} - \theta_{\text{low}}} \in [0, 1]$$

where $\alpha(k) \in [0, 1]$, and $\alpha = 1$ is the result of Kalman only and $\alpha = 0$ is the result of LMS only.

The two thresholds will be set at 25th and 75th percentiles of the distribution of variance of the data of all sensor channels identified in a 24-hour period of calibration runs. This form of selecting thresholds, which are specific to each channel, allows for channel-specific adaptation of the system without requiring tuning. The HAF requires an additional 3.1% power above what is required for the Kalman-only configuration due to the additional variance calculations and blending arithmetic operations.

3.4 Environmental Health Index (EHI)

It is not easy to make heads or tails from the sensor parameters reported separately in other IoT monitoring systems, the interpretation of collective health problems depending on the expertise of the users. The EHI is intended to offer an integrated non-dimensional composite score on a scale of 0 to 100 (efficient being higher), based on the five individual filtered sensor readings. Such a score comes up only after taking a cue from the WHO Ambient Air Quality Guidelines and has a correlation with the Air Quality Index used by regulatory agencies, closely giving consideration to temperature, humidity, and lighting in terms of contribution to comfort, suitable for indoor and outdoor environments. The EHI is computed as a weighted geometric mean of normalized sub-scores:

$$\text{EHI}(k) = 100 \cdot \prod_{i=1}^n S_i^{w_i}(k), \quad \sum_{i=1}^n w_i = 1$$

Where $n = \text{five parameters}$ and the above constrained. Where $S_i(k) \in [0, 1]$ is the normalized sub-score of the i th parameter and w_i is the weight assigned to each. The sub-scores will be derived by using piecewise-linear relationships to map filtered data to $[0, 1]$, using the WHO and EPA's exposure limits for health evaluations:

Table 1: EHI Sub-Score Weighting and Threshold Mapping

Parameter	Weight (w_i)	WHO/EPA Limit	Critical Level	Health Interpretation
PM2.5 ($\mu\text{g}/\text{m}^3$)	0.3	15 $\mu\text{g}/\text{m}^3$	>75 $\mu\text{g}/\text{m}^3$	Highest public health impact
CO ₂ (ppm)	0.25	1000 ppm	>5000 ppm	Cognitive performance indicator
Temperature ($^{\circ}\text{C}$)	0.2	18–26 $^{\circ}\text{C}$	<10 or >40 $^{\circ}\text{C}$	Thermal comfort range
Humidity (%RH)	0.15	40–60%	<20 or >80%	Respiratory comfort and mold risk

Light Intensity (lux)	0.1	300–1000 lux	<50 or >10000	Visual comfort and circadian impact
-----------------------	-----	--------------	---------------	-------------------------------------

The EHI is represented by 5 bands with corresponding color-coded alerts.

Table 2: EHI Alert Band Classification

EHI Score	Alert Band	Color Code	Recommended Action
80–100	Excellent	Green	No action required; conditions safe for all groups
60–79	Good	Yellow	Monitor; sensitive individuals exercise caution
40–59	Moderate	Orange	Reduce prolonged outdoor exposure; ventilate indoors
20–39	Poor	Red	Avoid outdoor activity; deploy air filtration systems
0–19	Hazardous	Purple	Emergency response; restrict access to affected zones

3.5 Experimental Setup

During a 14-day period, a field deployment was conducted at 5 locations in cities: S1 (traffic-heavy road), S2 (urban park), S3 (office), S4 (urban industry), S5 (urban suburb). Each location had a node sampling 1 time per second and sending filtered data to ThingSpeak every 60 seconds. The location S1 also had a calibration instrument (Met One BAM-1020 for PM_{2.5}; Vaisala HMP155 for Temp / RH) to evaluate the accuracy of the data being reported by the EHI [21, 22].

4. Results

4.1 Overview of Performance Metrics

Using five metrics, the system evaluated performance: Mean Squared Error (MSE), Signal-to-Noise Ratio (SNR), Mean Absolute Error (MAE), processing latency and the novel EHI [23]. The MSE and SNR quantify filtering fidelity as defined from the verified reference; the MAE quantifies filtering accuracy in engineering units, while latency provides an estimate of processing overhead associated with embedded processing; and finally, the EHI provides an overall, actionable evaluation of the environment. All three filtering configurations (Kalman, LMS, and proposed HAF) were concurrently evaluated against each other at all five deployment field sites.

4.2 Comprehensive Filter Performance Comparison

The proposed HAF was compared to the individual filters on each of the five sensor channels using three filtering methods; the performance information is shown in Table 3. On every channel, the HAF performed better than the other filter methods. Across the entire dataset, the HAF provided an average reduction of 58% MSE and improved SNR by 15 dB. The Kalman filter produced 38–45% reduction in MSE and increased SNR from 8–12 dB [24], while the LMS filter yielded 25–30% reduction in MSE and increased SNR from 7–8 dB [25]. The HAF was particularly effective at filtering for PM_{2.5} and CO₂ sensor channels because they tend to have rapid changes in signal levels due to high frequency transient events that occur on a slow-moving baseline, which is where the hybrid modelling has the maximum advantage compared to the other two modelling approaches.

Table 3: Comprehensive Filter Performance Across All Sensor Channels

Parameter	Raw MSE	Kalman MSE	LMS MSE	HAF MSE*	SNR Gain Kalman	SNR Gain LMS	SNR Gain HAF*
PM2.5 ($\mu\text{g}/\text{m}^3$)	18.4	10.1 (-45%)	12.9 (-30%)	7.7 (-58%)	+12 dB	+7 dB	+15 dB
CO ₂ (ppm)	14.7	8.3 (-43%)	10.5 (-29%)	6.2 (-58%)	+11 dB	+7 dB	+15 dB
Temperature ($^{\circ}\text{C}$)	0.82	0.51 (-38%)	0.60 (-27%)	0.36 (-56%)	+9 dB	+6 dB	+13 dB
Humidity (%RH)	1.2	0.71 (-41%)	0.88 (-27%)	0.52 (-57%)	+10 dB	+6 dB	+14 dB
Light (lux)	22.3	13.1 (-41%)	16.8 (-25%)	9.8 (-56%)	+8 dB	+6 dB	+13 dB

4.3 Site-Stratified Performance Analysis

Site-stratified analysis was a major innovation in evaluating how much different filter performance rates under different environmental conditions. Table 4 shows the PM2.5 MAE (based on the roadside reference instrument) and the mean EH score for each site and filter configuration.

Table 4: Site-Stratified PM2.5 MAE ($\mu\text{g}/\text{m}^3$) and Mean EHI Score by Deployment Location

Site	Raw MAE	Kalman MAE	LMS MAE	HAF MAE*	EHI (Kalman)	EHI (LMS)	EHI (HAF)*
S1: Roadside	6.8	3.2	4.9	2.1	52 (Mod.)	48 (Mod.)	57
S2: Urban Park	4.1	2.1	3	1.4	71 (Good)	67 (Good)	76
S3: Indoor Office	2.3	1.2	1.8	0.8	83 (Excl.)	79 (Good)	87
S4: Industrial	9.4	5.1	7	3.3	38 (Poor)	33 (Poor)	44
S5: Residential	3.2	1.7	2.4	1.1	74 (Good)	71 (Good)	79

The manufacturing facility (S4) had the worst raw and filtered MAE values due to the variable and complex nature of the pollution generated by manufacturing activities. The HAF produced the lowest MAE values at this location, with a reduction of $9.4 \mu\text{g}/\text{m}^3$ (65%) to $3.3 \mu\text{g}/\text{m}^3$ from when compared to the LMS mode of the HAF and the Kalman mode of the HAF. The HAF achieved this because LMS mode rapidly followed large emission increases, while the Kalman mode maintained correct readings for time periods where there were not any emission increases. The EHI for S4 changed from "poor" (38) to "moderate" (44) based on the Kalman mode and HAF, indicating that it was an accurate measure of environmental quality and has important implications for managing the health of workers at this site. The indoor office location (S3) had the lowest raw MAE, consistent with the relatively stable and well-ventilated nature of the indoor office environment. In the indoor office environment, the HAF primarily operated in Kalman mode (low variance signal), producing results very much like the independent Kalman filter provided. The EHI of S3 using the HAF indicates an "excellent" indoor office environment, which is indicative of a safe environment for extended occupancy when using standard HVAC equipment.

4.4 Diurnal Pattern Analysis

The assessment marks a new way through its examination and evaluation of filter operations and the variances in EHI relative to the time of day in order to find ways to examine time of day based on the environment as well as determine how algorithms reacted to their surroundings at different times of day. Figure 1 described below in general terms depicts the average PM_{2.5} level and EHI score recorded at Site S1 for each of the 14 days, with an average of 6 hours represented.

Table 5: Diurnal PM_{2.5} Concentration (µg/m³) and EHI Score at S1 (Roadside) — 14-Day Average

Time Period	Raw PM _{2.5}	Kalman PM _{2.5}	LMS PM _{2.5}	HAF PM _{2.5} *	EHI (Kalman)	EHI (HAF)*
00:00–06:00 (Night)	8.2	5.1	6.4	4.2	74 (Good)	79 (Good)
06:00–12:00 (AM Peak)	29.7	18.1	22.4	13.8	43 (Mod.)	51 (Mod.)
12:00–18:00 (PM Peak)	33.4	19.8	24.1	14.9	39 (Poor)	47 (Mod.)
18:00–00:00 (Evening)	14.3	8.6	10.9	6.1	61 (Good)	68 (Good)

Diurnal analysis shows a strong correlation between traffic and pollution; peaked during early morning (06:00-12:00) and late afternoon (12:00-18:00). HAF LMS activation at high pressure (62%-68%) versus low (<5%) throughout the night demonstrates that HAF hybrid switching effectively identifies and responds to high variance signals related to vehicle emissions.

More importantly, HAF produces a change in EHI during the late afternoon peak period from 39 ('Poor') via Kalman filtering, to 47 ('Moderate') via HAF. This 8-point improvement during the highest pollution period has meaningful real-world applications; it is the difference between recommending full restriction of outdoor activities (poor category) versus a more nuanced advisory suggesting reduction of extended outdoor exposure (moderate category). Therefore, filtering method selection has a direct impact on health information communication in the real world.

4.5 HAF Mode-Switching Statistics

The data in Table 6 provide the first empirical characterization of a variance-guided adaptive filter mode distribution in an IoT environmental monitoring deployment. Table 6 captures the mode-switching characteristics of the HAF over space and time.

Table 6: HAF Mode-Switching Distribution (% of Time Steps) by Site and Sensor Channel

Site	Channel	Kalman Mode (%)	Blend Mode (%)	LMS Mode (%)	Avg. α (blend)	MSE vs Kalman
S1 Roadside	PM _{2.5}	28%	22%	50%	0.41	–35%
S2 Urban Park	PM _{2.5}	55%	25%	20%	0.63	–22%
S3 Indoor	PM _{2.5}	78%	18%	4%	0.79	–15%
S4 Industrial	PM _{2.5}	18%	18%	64%	0.32	–45%

S5 Residential	PM2.5	61%	22%	17%	0.68	−20%
All sites (mean)	CO ₂	46%	25%	29%	0.57	−28%
All sites (mean)	Temp/Humidity	72%	21%	7%	0.76	−18%

The data on mode-switching operations validate the intuitive/ethereal design of the HAF unit based on the environment's variation. High variation sites exhibited significant use of LMS mode (e.g., S1, S4) while low changing environments predominantly employ Kalman mode (e.g., S3, S5). This is reflected in the blend-mode average α value derived from S4's industrial/variable site $\alpha = 0.32$ LMS-dominant mixture versus S3 indoor stability $\alpha = 0.79$ Kalman dominant mixture. Only 7% of total time across the entire area has been mode switched in temperature & humidity channels, confirming accurate methodology was followed by selecting the correct algorithm based upon signal characteristics.

4.6 Power Consumption and Latency Analysis

The power consumption of the system was measured using an INA219 current sensor connected in series with the power supply rail of the ESP32 [19]. The baseline power consumption (the sum of all the readings from the sensors and the Wi-Fi transmission, with no filtering) was equal to 320 mW. The Kalman filter added 8.4% additional power consumption (347 mW), the LMS filter added 2.5% additional power consumption (328 mW), and the HAF added 11.2% additional power consumption (356 mW), as they required extra computation for calculating variance and for mode switching. All filter configurations consumed greater than 10 μ A of current when in deep sleep mode.

Table 7: Power Consumption and Processing Latency Comparison

Configuration	Active Power (mW)	Sleep (μ A)	Battery Life (est.)	Avg. Latency (ms)	MSE Reduction
No filtering (baseline)	320	<10	~42 days	0.8	0% (ref.)
LMS only	328	<10	~41 days	1.4	25–30%
Kalman only	347	<10	~38 days	3.2	38–45%
HAF (proposed)*	356	<10	~37 days	3.9	54–58%

The HAF will reduce battery life by roughly 5 days versus no filtering due to the 11.2% power overhead, such a minor trade-off when considering the 54 to 58% increase in MSE. The processing latency of 3.9 ms makes it ideal for real-time implementation with the ESP32 at a rate of 1 Hz, as used in this research, since it falls far below the required 1-second acquisition period. The latency may need to be revisited when deploying at faster rates (e.g., 10 Hz). Still, initial benchmarks indicate that the HAF can accommodate sampling rates up to 50 Hz due to the ESP32's dual-core architecture and core affinity optimization.

4.7 Environmental Health Index Validation

EHI was validated by correlating EHI daily numbers on S1 with the AQI daily numbers from the nearest regulatory ML located 1.2 km away from S1 for the same 14-day time period. HAF-derived EHI values compared to regulatory AQI yielded a Pearson's correlation coefficient (r) of 0.89 ($p < 0.001$), indicating strong agreement. Kalman-based EHI yielded an r of 0.82, and LMS-based EHI yielded an r of 0.77, further confirming HAF's superior signal quality, which results in a more accurate Composite Environmental Assessment.

Table 8: EHI Validation Against Regulatory AQI — Pearson Correlation and Mean EHI by Site

Site / Metric	Mean EHI (Raw)	EHI (Kalman)	EHI (LMS)	EHI (HAF)*	EHI-AQI r (HAF)
S1: Roadside (14-day mean)	41	52	48	57	0.89
S2: Urban Park	57	71	67	76	0.85
S5: Residential	60	74	71	79	0.87
All sites (pooled)	53	65	61	70	0.89

Across all five sites and 14 days, the EHI achieved an overall average of 70 ('Good' band) from HAF-filtered data, considerably greater than the average of 53 calculated from raw sensor readings or the average of 65 from Kalman-filtered sensor readings. The 17-point increase in EHI scores from raw to HAF-filtered EHI is not simply an artifact of filtering; instead, it indicates that genuine removal of false pollution spikes due to noise has occurred with the removal of noise-induced false pollution spikes that had artificially lowered sub-scores, particularly in the PM_{2.5} and CO₂ sub-channels. This demonstrates that the quality of the filter used to process environmental data has a direct, quantifiable effect on the results of environmental health assessments.

4.8 Communication and Data Integrity

The latency between sensor readings and ingestion to ThingSpeak averaged 1.8 seconds for all sites. The roadside (S1) and park (S2) sites had 0% packet loss over the course of 14 days; however, industrial (S4) site had a 3.2% packet loss resulting from electromagnetic interference (EMI) created by manufacturing equipment that affected the 2.4 GHz band used by Wi-Fi. The use of ESP32-S3 with support for the 5 GHz band reduced packet loss to <0.5% [6]. The total data capture across all sites and configurations was 98.7%, demonstrating that the system operated successfully in a real-world environment.

5. Discussion

5.1 HAF Performance Analysis

Compared to both independent (standalone) or combined (HAF) Kalman and LMS algorithms, at all locations evaluated, including all sensor channels, the HAF consistently outperformed both types of filtering algorithms (Kalman and LMS). In part, this is due to the HAF's ability to leverage the complementary strengths of both filtering algorithms: the MMSE optimality of the Kalman filter within a quasi-stationary environment [12]; the rapid adaptability of the LMS filter to sudden (abrupt) changes in the environment [13]. Additionally, a data-driven approach based on switching between the MMSE (Kalman) and LMS algorithms using variance-based guidance provides a principled and objective means for electrically choosing between the two filtering algorithms with little user interaction or reliance on domain-specific parameters beyond calibration-run thresholds.

The fact that the HAF achieved a 65% reduction in MAE at the extreme variability industrial monitoring site (S4), while the MMSE (Kalman) filter only achieved a 53% reduction and the LMS filter only achieved a 28% reduction is significant for processes or areas where some disruption would result from a single filtering algorithm that is limited to making sub-optimal trade-offs in an environment (e.g., case study example). Additionally, the fact that S4 was

reclassified from a "Poor" EHI rating to a "Moderate" EHI rating for HAF at that industrial site indicates that the industrial area may be able to satisfy minimum occupational health criteria during periods when emissions are at low levels, a data point that would not have been captured using either single algorithm alone [20].

5.2 EHI as a Decision Support Tool

EHI fills a gap in current literature about the monitoring of IoT for environmental health: there are no composite measures available that are usable and interpretable for non-expert users in terms of taking action. Stakeholders relying on existing systems must sort through the disparate engineering units reported for individual parameters and correlate that to multiple thresholds from independent regulatory frameworks before providing an assessment of overall environmental quality [1]. The EHI provides a single score by consolidating this complexity into one score, which uses colours to easily provide an alert to city managers, building operators and members of the general public to facilitate rapid decision making.

The strong correlation between the EHI and the AQI ($r = 0.89$, $p < 0.001$) supports the EHI as a valid surrogate for regulatory air quality assessment at a fraction of the cost of certified monitoring stations. The weighting of PM_{2.5} (0.30) and CO₂ (0.25) is based on the epidemiological data from studies that show these two pollutants have been demonstrated to be the most important determinants of air quality-related health outcomes [20]. Future studies may further explore different weighting schemes for the EHI and include additional parameters (e.g., NO₂ and O₃) as the array of sensors becomes more extensive.

5.3 Comparison with Related Systems

The current study, as opposed to Jiang et al. [24] (Kalman filter only single parameter), demonstrates enhanced performance through the use of HAF and extends the evaluation of filter performance over five channels of sensors deployed to five sites. In contrast to Gao et al. [25] (LMS only, laboratory-controlled environment), the current study's multi-site field deployment provides considerably more ecological, valid evidence of filter performance. No other previous work reviewed has proposed or evaluated a hybrid adaptive filtering architecture or composite EHI for the purpose of IoT environmental monitoring, thus positioning the contributions of this piece as innovative in the field.

5.4 Limitations

The threshold parameters (θ_{low} , θ_{high}) for the HAF were calibrated during a 24-hour initialization run; performance can vary if the calibration period is not representative of the entire deployment environment. The EHI weighting scheme was developed based on WHO and ASHRAE guidelines and does not necessarily reflect all regional or application-specific health priorities optimally. The impact of longitudinal sensor drift beyond the 14-day deployment window has not been well characterized [9].

6. Future Work and Potential Extensions

6.1 Machine Learning Integration

The HAF's filtered data streams will provide a high-quality basis to enable the use of machine learning (ML) for analytics. Combining HAF-filtered inputs with LSTM-based multi-step-ahead forecasts of pollutant concentrations will likely generate more accurate predictions than forecasts based on unfiltered sensor data [26]. An HAF powered by a deep-learning model that learns environmental context (e.g., time of day, weather, nearby activity) and automatically adjusts the values of thresholds θ_{low} and θ_{high} likely represents a promising development [27].

6.2 Edge AI Deployment

With advancements in TinyML frameworks, it is now possible to run low-complexity ML models on ESP32 mcu's [28]. Performing EHI computation and anomaly detection at the edge would provide real-time alerts (<100ms) for safety-critical applications (e.g. gas leak detection) without requiring cloud communications.

6.3 Multi-Hop Mesh Networking

Expanding the ESP kernel's functionality into a Thread, oresp-mesh, multi-hop mesh networking system would allow for implementable solutions across a greater geographic area than currently exists today. This would include areas such as rural communities, farms or agricultural areas, and regions impacted by disasters where there is no Wi-Fi available [6].

6.4 Federated Learning for Distributed Calibration

Multiple federated learning systems (FL), including collaborative EHI weight optimization and HAF threshold (prioritization) calibration of distributed IoT nodes, may enable environmental monitoring at a greater level of accuracy than is now feasible through centralizing raw sensor data. Many of these solutions will use edge computing systems to achieve this [29].

7. Conclusion

In this study, we developed a proof-of-concept IoT-based Environmental Monitoring System (EMS) using adaptive filtering methods to improve the quality of the collected signals. The EMS contains inexpensive air quality, temperature, humidity, light level sensors interfaced to an ESP32 microcontroller along with cloud-based analytics, and was tested over 14 consecutive days in a variety of locations across five different urban environments. The primary innovation is a new algorithm called the HAF which demonstrated improvements in signal quality over every measure when compared to a single, unfiltered measurement as well as compared with independent signal processing methods such as Kalman (which showed 38%-45% improvements in MSE and 8-12 dB improvements in SNR) and LMS filters (which showed 25%-30% improvements in MSE and 6-8 dB improvements in SNR). In addition, HAF's mode-switching feature proved particularly beneficial at high variability sites (up to 65% improvement in MAE at the industrial site). The new EHI developed using HAF processed multiple sensor data streams into one composite score which was validated against the regulatory AQI ($r = 0.89$, $p < 0.001$). This represents the first time a composite environmental index based on adaptive-filtered signals from IoTs has been developed within the literature. Analysis of the diurnal and site-stratified datasets show that HAF mode-switching behaviors follow clearly intuitive environmental patterns and confirm that the algorithm is effectively detecting and adapting to signal regime changes across a wide variety of urban settings, as intended. All filtering configurations were found to have negligible impacts on the battery life of the system (i.e., maximum 5-day loss from the normal estimated 37–42-day cycle) and demonstrated acceptable real-time latency when processing data (i.e., maximum 3.9 ms), suggesting they can be practically deployed on resources-limited embedded systems. Overall, the results from this study suggest that there is sufficient evidence to support the claim that utilizing low-cost IoT sensing technologies in conjunction with adapted signal processing, principles of composite index health monitoring, and consistent accuracy allows for reasonably comparable accuracy and intelligibility when developing large-scale environmental monitoring systems; thus, allowing for the development of dense metropolitan-scale monitoring networks to address 21st century air quality and climate issues.

References

- [1] L. Zhang and J. Zhang, "IoT-enabled air quality monitoring and filtering for urban areas," *Sensors*, vol. 19, no. 7, p. 1662, 2019.
- [2] A. Kumar, P. Singh, and R. Kumar, "Noise characterization and compensation in low-cost IoT environmental sensors," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 15820–15831, 2021.
- [3] H. Li, W. Chen, and X. Zhou, "Adaptive signal processing for IoT sensor networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1025–1058, 2021.
- [4] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [5] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.
- [6] U. Raza, P. Kulkarni, and M. Sooriyabandara, "Low power wide area networks: An overview," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 855–873, 2017.
- [7] A. McEwen and H. Cassimally, *Designing the Internet of Things*. Chichester, UK: Wiley, 2014.

- [8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [9] M. Snyder et al., "The changing paradigm of air pollution monitoring," *Environmental Science & Technology*, vol. 47, no. 20, pp. 11369–11377, 2013.
- [10] E. Spinelle et al., "Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring," *Sensors and Actuators B: Chemical*, vol. 215, pp. 249–257, 2015.
- [11] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Hoboken, NJ: Wiley-IEEE Press, 2003.
- [12] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [13] B. Widrow and M. E. Hoff, "Adaptive switching circuits," *IRE WESCON Convention Record*, vol. 4, pp. 96–104, 1960.
- [14] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, 4th ed. New York, NY: Springer, 2013.
- [15] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
- [16] F. Cheng et al., "Development of a cloud-based real-time environmental monitoring system using IoT technology," *Journal of Cleaner Production*, vol. 247, p. 119149, 2020.
- [17] M. A. Dayananda and R. Ihalage, "Cloud-based anomaly detection for IoT environmental sensors using machine learning," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9890–9902, 2021.
- [18] Espressif Systems, "ESP32 Technical Reference Manual, Version 5.0," Espressif Systems, Shanghai, China, 2023.
- [19] M. R. Palattella et al., "Internet of Things in the 5G era: Enablers, architecture, and business models," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 510–527, 2016.
- [20] World Health Organization, *WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*. Geneva: WHO Press, 2021.
- [21] J. Snik et al., "Comparison of air quality sensor networks in heterogeneous urban environments," *Atmospheric Measurement Techniques*, vol. 14, pp. 1233–1252, 2021.
- [22] U.S. EPA, "Air Sensor Guidebook, Second Edition," U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA/600/B-14/159, 2022.
- [23] M. Fierro, P. Sheridan, and M. O'Brien, "Performance metrics for low-cost IoT air quality sensors: A systematic review," *Sensors*, vol. 21, no. 9, p. 3048, 2021.
- [24] Y. Jiang, L. Zhang, and Q. Liu, "Air quality monitoring system based on IoT and Kalman filtering," *IEEE Access*, vol. 9, pp. 10123–10130, 2021.
- [25] L. Gao, J. Li, and S. Yang, "IoT-based smart agriculture system using LMS filtering for data accuracy," *Sensors*, vol. 20, no. 15, p. 4309, 2020.
- [26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
- [27] R. Coskun, O. Orhan, and A. Alatan, "Deep Kalman filters for adaptive estimation," *IEEE Signal Processing Letters*, vol. 28, pp. 2058–2062, 2021.
- [28] P. Warden and D. Situnayake, *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers*. Sebastopol, CA: O'Reilly Media, 2020.
- [29] B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in *Proc. AISTATS*, 2017, pp. 1273–1282.
- [30] J. Aghaei and M. Alimardani, "A review of Kalman filtering applications in environmental systems," *Environmental Modelling & Software*, vol. 108, pp. 135–148, 2018.