Air Quality Sensor Data QA/QC and Drift Detection Report

Analysis of Zephyr Sensor Network (NO_2, $PM_{2.5}$, RH, Temperature) – June to July 2025

25th Sept 2025

BY
Mohamed Karlavi Maharoof
Data Scientist | Research project Supervisor | MSc Data Science & Computational Intelligence Excellence prize winner

Task 1: Make it Analysis-ready (QA/QC)

Objective

The purpose of this QA/QC stage is to prepare the EarthSense Zephyr air quality dataset for analysis by applying a reproducible pipeline.This ensures that the data is consistent, transparent, and aligned across all sensors, ready for validation against reference measurements in subsequent tasks.

Dataset Overview

- **Period:** 15 June – 31 July 2025 (~6.5 weeks)
- **Sensors:** 20 Zephyr units (ZEPHYR_001 → ZEPHYR_020)
- **Variables:** NO_2 (ppb, µg/m³), $PM_{2.5}$ (µg/m³), Relative Humidity (%), Temperature (°C), Firmware, Status
- **Size:** 22,760 rows × 9 original variables

1. Timestamps & Alignment

- All timestamps are in **UTC (+00:00)** → no DST corrections required.
- **Sorting** applied: rows ordered by sensor_id and timestamp_recorded_utc.
- **Clock shift bug** detected: ZEPHYR_015 showed a 1-hour shift (confirmed in changelog).

Table 1: Number of records per sensor

Sensor ID	Record Count
ZEPHYR_001	1128
ZEPHYR_002	1128
ZEPHYR_003	1128
ZEPHYR_004	1128
ZEPHYR_005	1128
ZEPHYR_006	1128
ZEPHYR_007	1128
ZEPHYR_008	1128
ZEPHYR_009	1128
ZEPHYR_010	1128
ZEPHYR_011	1128
ZEPHYR_012	1128
ZEPHYR_013	1128
ZEPHYR_014	1128
ZEPHYR_015	1328
ZEPHYR_016	1128
ZEPHYR_017	1128
ZEPHYR_018	1128
ZEPHYR_019	1128
ZEPHYR_020	1128

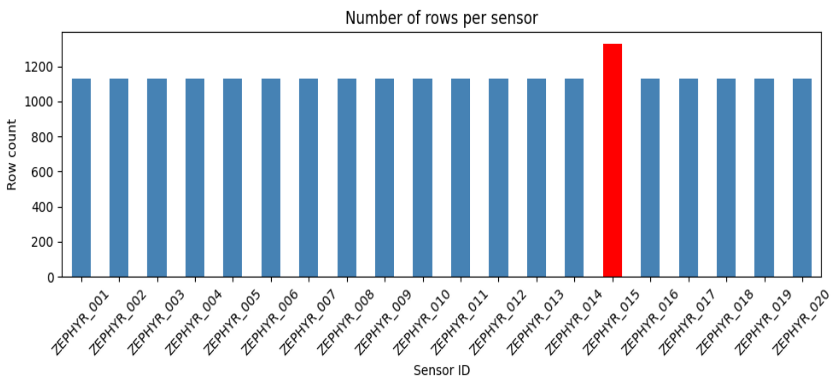


Figure 1: Number of records per sensor

2. Duplicates

- **Full-row duplicates**: 0
- **201 duplicate timestamps** detected, all from **ZEPHYR_015**.
- Handled by: flagging duplicates, not removing.

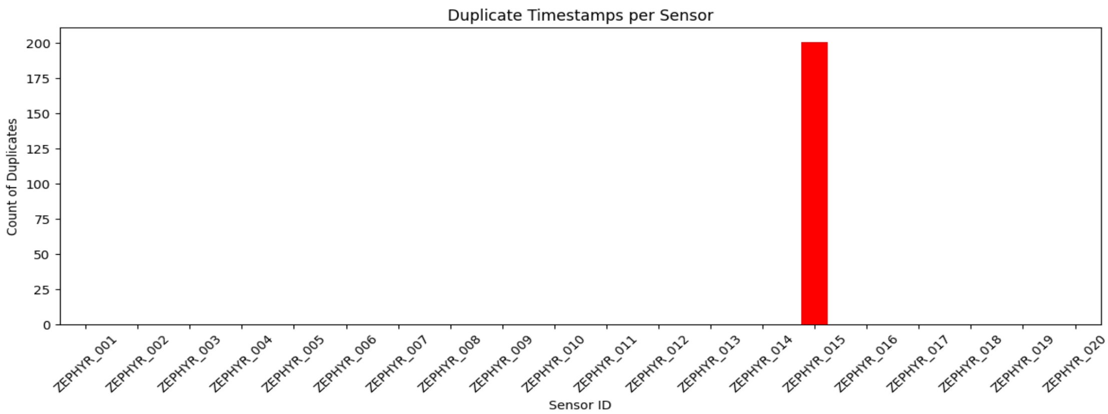


Figure 2: Duplicate readings per sensors

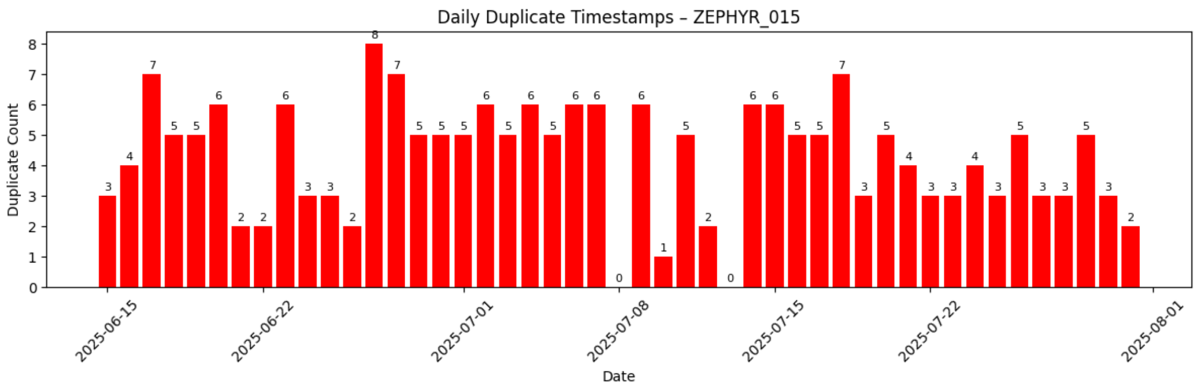


Figure 3: Frequency of Duplicate Records per Day (ZEPHYR_015)

ZEPHYR_015 shows in Figure 2, daily duplicate timestamps 201 flagged rows across the study period), not limited to the 20 June–1 July clock-shift bug window. This suggests a sensor-specific logging or backend ingestion issue, rather than a one-off clock shift. Further inspection of raw timestamps and duplicate values is needed to confirm whether these were exact repeats (pipeline duplication) or near-repeats (sensor double-logging).

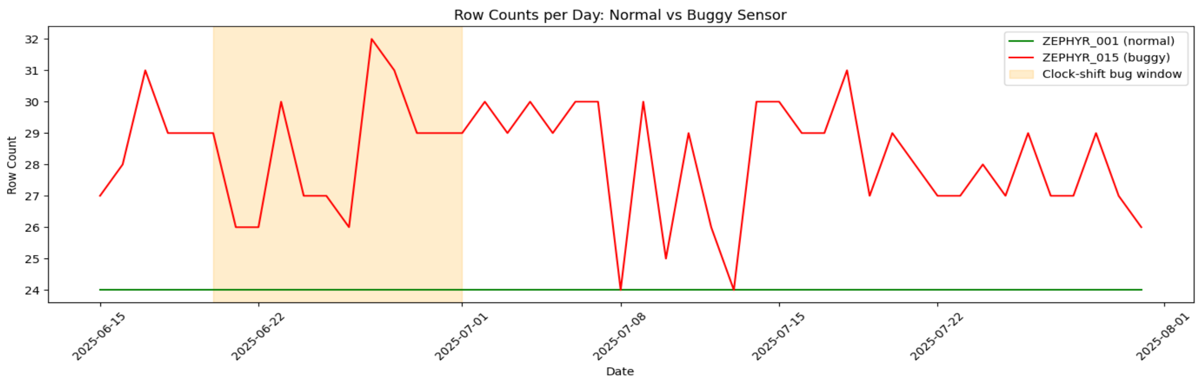


Figure 4: Sensor Data Integrity Check: Normal vs Buggy Record Counts (Clock drifts highlighted)

3. Missing Values

- **361 missing same rows** detected, affecting NO_2 (ppb, µg/m³) and $PM_{2.5}$.
- RH and Temperature had no missing values.
- Handled by: flagging missing values.

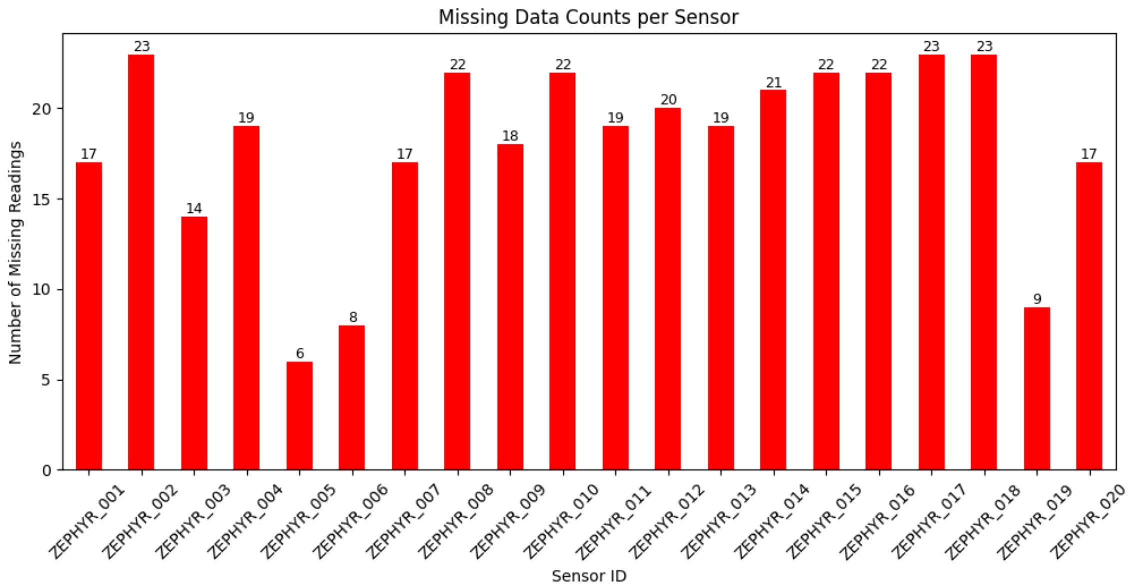


Figure 5: Missing Data Counts per sensor

4. Out of Range & Negative values

The threshold values used for QA/QC were defined in accordance with the World Health Organization (WHO), European Union (EU), and United Kingdom (UK) air quality guidelines. Table 2 summarizes the applied thresholds for key parameters.

Table 2: Threshold Values to identify the outliers

Parameter	Threshold Range
NO_2 (ppb)	0 – 100
NO_2 (µg/m³)	0 – 200
$PM_{2.5}$ (µg/m³)	0 – 35
Relative Humidity (%)	0 – 100
Temperature (°C)	−10 → 35

The threshold values applied in this study were defined according to the Air Quality Index bands shown in Figure 6 & Figure 7 based on WHO/EU/UK guidelines for NO_2 and $PM_{2.5}$.

Nitrogen Dioxide

Based on the hourly mean concentration.

Index	1	2	3	4	5	6	7	8	9	10
Band	Low	Low	Low	Moderate	Moderate	Moderate	High	High	High	Very High
µg/m³	0-67	68-134	135-200	201-267	268-334	335-400	401-467	468-534	535-600	601 or more

Figure 6: UK Air Quality Index Scale for Nitrogen Dioxide

$PM_{2.5}$ Particles

Based on the daily mean concentration for historical data, latest 24 hour running mean for the current day.

Index	1	2	3	4	5	6	7	8	9	10
Band	Low	Low	Low	Moderate	Moderate	Moderate	High	High	High	Very High
μgm^{-3}	0-11	12-23	24-35	36-41	42-47	48-53	54-58	59-64	65-70	71 or more

Figure 7: UK Air Quality Index Scale for $PM_{2.5}$

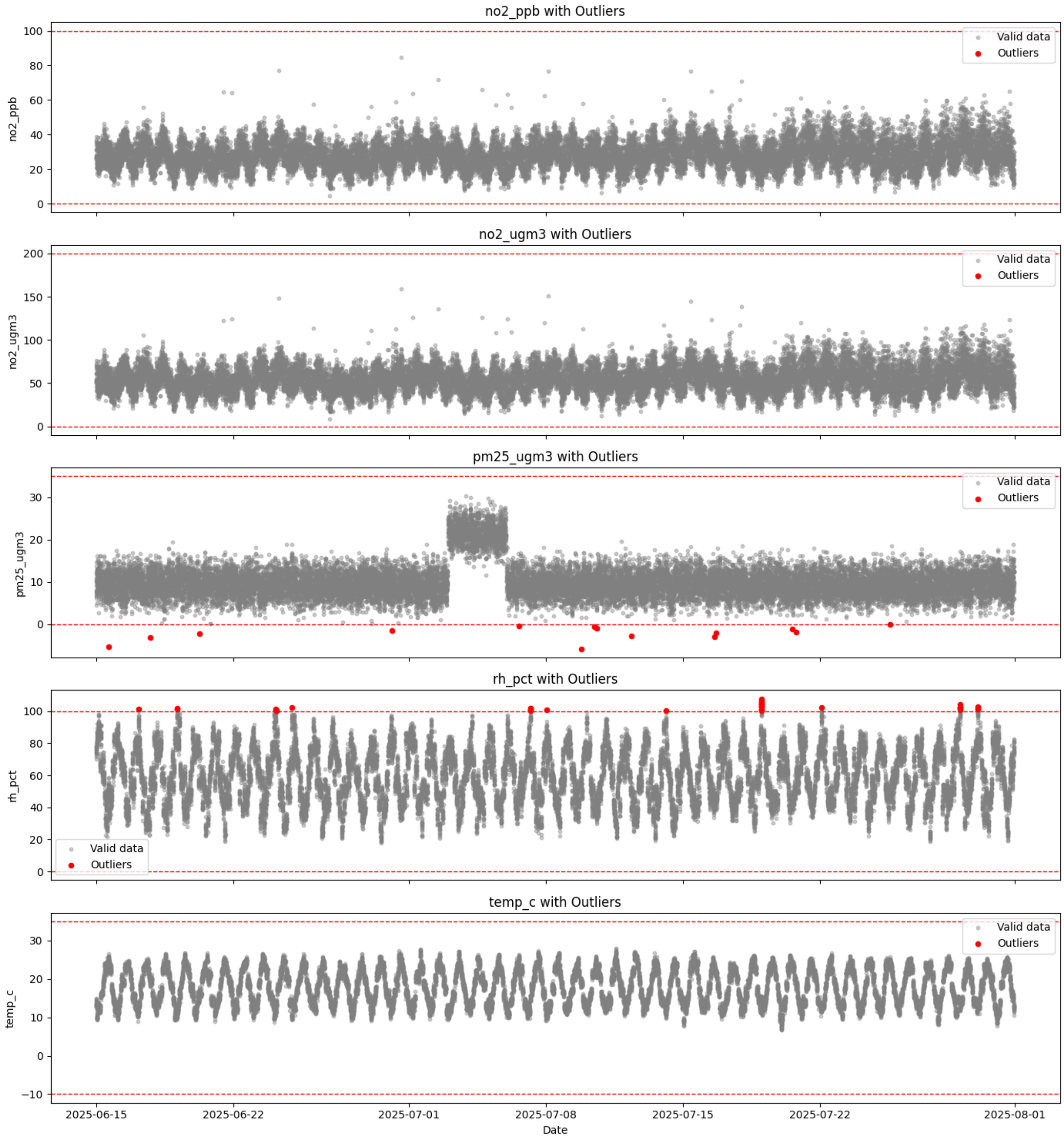


Figure 8: Detection of Outliers & Negative Values

Table 3 summarizes the number of detected outliers per parameter, where $PM_{2.5}$ included negative values and Relative Humidity exceeded 100%.

Table 3: Outliers Counts

Parameter	Outlier Count
NO_2 (ppb)	0
NO_2 (µg/m³)	0
$PM_{2.5}$ (µg/m³)	14 (Negative value)
Relative Humidity (%)	37 (Above 100)
Temperature (°C)	0

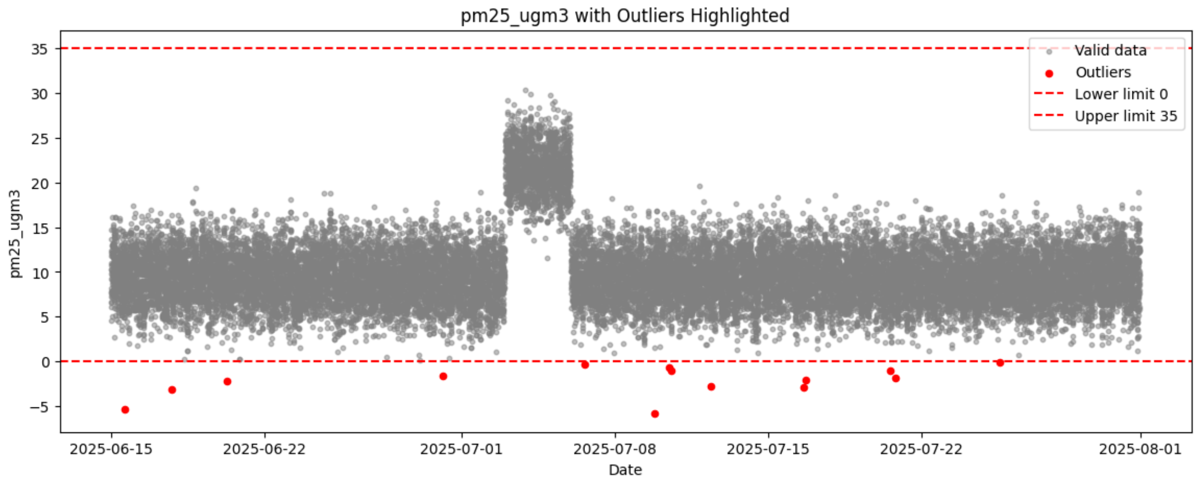


Figure 9: Detection of Outliers in $PM_{2.5}$ Data

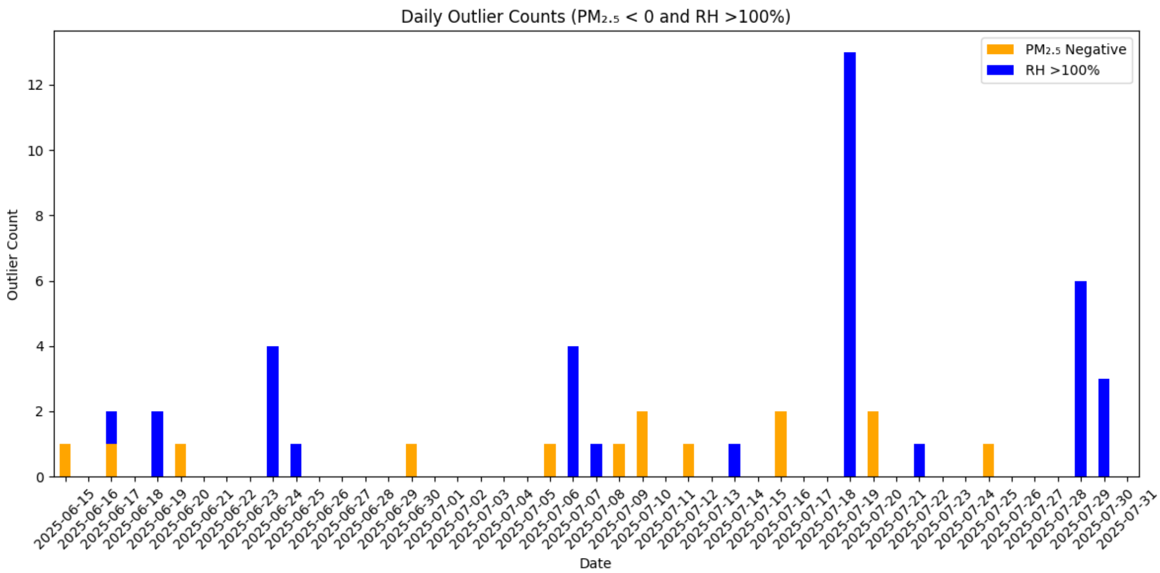


Figure 10: Time Series of Daily Outlier Counts ($PM_{2.5} < 0$, RH > 100%)

5. Flat line check

A flat-line check was applied using a **5-hour rolling window** across each sensor and variable. The purpose was to detect any periods where a sensor repeatedly reported the same value, which may indicate a malfunction or frozen measurement

No flat-lined periods were detected across any sensor or variable.

6. Resampling data

- All sensors retained the expected **1128 hourly rows**, except **ZEPHYR_015**, which had **201 duplicate entries removed**.

The dataset was resampled to a **strict hourly grid per sensor** to ensure consistency and comparability.

- **Numeric variables** (NO_2, $PM_{2.5}$, Temperature, RH) were averaged where duplicate entries existed within the same hour.
- **QC flags** (duplicates, missing, out-of-range) were aggregated using a **conservative "max rule"**:
  - If *any* row within the hour was flagged, the resampled hour retained the flag.
- **Missing hours** were explicitly inserted as NaN values and flagged, preventing silent data gaps.

This approach prevents *silent overwriting* of duplicate records and ensures each sensor contributes exactly one row per hour.

- Dataset reduced from **22,760 → 22,560 rows** after removing duplicate timestamps.
- **ZEPHYR_015** was corrected from **1328 rows → 1128 rows**
- All other sensors retained the expected **1128 hourly rows**.
- QC flags were preserved in the resampled dataset, ensuring transparency of issues.

Table 4: Comparison of Record Counts per sensor (Pre vs Post Cleaning)

Sensor ID	Before	After
ZEPHYR_001	1128	1128
ZEPHYR_002	1128	1128
ZEPHYR_003	1128	1128
ZEPHYR_004	1128	1128
ZEPHYR_005	1128	1128
ZEPHYR_006	1128	1128
ZEPHYR_007	1128	1128
ZEPHYR_008	1128	1128
ZEPHYR_009	1128	1128
ZEPHYR_010	1128	1128
ZEPHYR_011	1128	1128
ZEPHYR_012	1128	1128
ZEPHYR_013	1128	1128
ZEPHYR_014	1128	1128
ZEPHYR_015	1328	1128
ZEPHYR_016	1128	1128
ZEPHYR_017	1128	1128
ZEPHYR_018	1128	1128
ZEPHYR_019	1128	1128
ZEPHYR_020	1128	1128

Note: End of Task 1 Cleaned Resampled dataset exported as df_hourly.csv file for further analysis for Upcoming tasks

Task 2: Validation vs Reference and Neighbours

2.1 Objective

The aim was to validate the performance of 20 Zephyr-like sensors by comparing their NO_2 and $PM_{2.5}$ measurements against the nearby reference station and neighbouring sensors. We sought to identify sensors that performed well, those that showed systematic bias, and those affected by artefacts such as drift, firmware-related step-changes, or humidity-driven $PM_{2.5}$ inflation.

2.2 Methodology

- **Reference dataset**: 1,128 hourly values of NO_2 and $PM_{2.5}$ from a trusted monitoring station.
- **Validation approach**:
    1. Computed basic statistics for each sensor vs reference:
        - Coefficient of determination (R^2),
        - Mean bias (sensor – reference),
        - Root Mean Squared Error (RMSE).
    2. Visualised sensor performance with:
        - Scatter plots (sensor vs reference),
        - Time series overlays,
        - Sensor–sensor correlation heatmaps.
    3. Cross-referenced findings with **neighbouring sensors** and **changelog metadata** (firmware updates, dust event, clock bug).

2.3 Results

a) Statistical validation

Table 5: Summary of R^2, bias, RMSE

Sensor ID	NO_2			$PM_{2.5}$		
	R^2	Bias	RMSE	R^2	Bias	RMSE
ZEPHYR_001	0.398	4.114	7.396	0.719	0.054	1.888
ZEPHYR_002	0.771	−1.969	3.433	0.772	−1.086	2.078
ZEPHYR_003	0.773	−1.796	3.232	0.707	−0.149	2.112
ZEPHYR_004	0.417	12.280	13.706	0.759	3.062	3.571
ZEPHYR_005	0.399	2.025	6.444	0.747	−0.963	2.128
ZEPHYR_006	0.799	−2.016	3.276	0.699	−0.103	2.157
ZEPHYR_007	0.665	1.854	4.033	0.694	0.665	2.238
ZEPHYR_008	0.685	9.808	10.443	0.763	2.907	3.441
ZEPHYR_009	0.643	2.010	4.183	0.761	−0.132	1.870
ZEPHYR_010	0.752	8.092	8.589	0.768	2.936	3.441
ZEPHYR_011	0.541	10.232	11.271	0.769	2.984	3.482
ZEPHYR_012	0.750	8.045	8.584	0.764	2.929	3.446
ZEPHYR_013	0.676	−0.074	3.452	0.776	−0.982	2.017
ZEPHYR_014	0.788	0.092	2.626	0.756	−0.036	1.869
ZEPHYR_015	0.660	−2.028	3.982	0.727	−0.989	2.176
ZEPHYR_016	0.790	−2.043	3.309	0.751	−1.047	2.115
ZEPHYR_017	0.754	0.006	2.867	0.751	−0.044	1.827
ZEPHYR_018	0.648	1.859	4.127	0.755	0.008	1.873
ZEPHYR_019	0.683	9.847	10.447	0.735	3.062	3.547
ZEPHYR_020	0.743	7.986	8.504	0.752	3.013	3.546

- **Good performers**: ZEPHYR_006, ZEPHYR_016 (NO_2 R^2 ~0.80, low bias, RMSE ~3 ppb).
- **Systematic bias**: ZEPHYR_010, 012, 020 (+8 ppb NO_2 bias).
- **Poor performers**: ZEPHYR_004, 005, 001 (R^2 <0.45, high RMSE).

ZEPHYR_006 vs Reference (Good Agreement)

- Points cluster tightly around the 1:1 line.
- $R^2 = 0.799$; small negative bias (−2.0 ppb).
- Indicates reliable performance with minimal adjustment required.

Figure 11: Scatter Plots of ZEPHYR_006 vs Reference (NO_2 and $PM_{2.5}$)

Figure 12: Time Series Comparison of ZEPHYR_006 vs Reference (NO_2 and $PM_{2.5}$)

ZEPHYR_010 vs Reference (Systematic Overestimation)

- Points consistently above the 1:1 line.
- $R^2 = 0.752$; positive bias (+8.1 ppb NO_2).
- Confirms overestimation requiring calibration or correction.

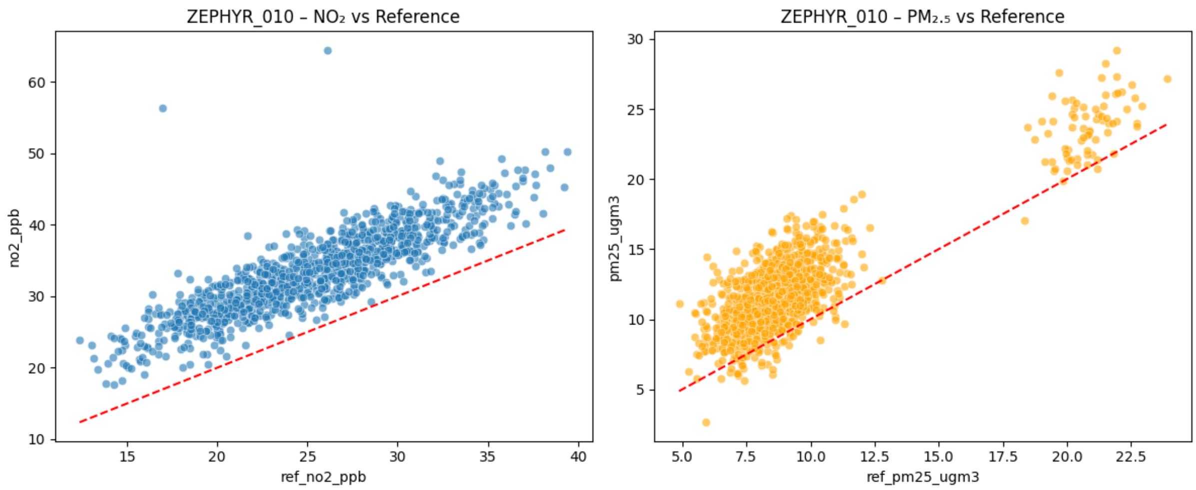


Figure 13: Scatter Plots of ZEPHYR_010 vs Reference (NO_2 and $PM_{2.5}$)

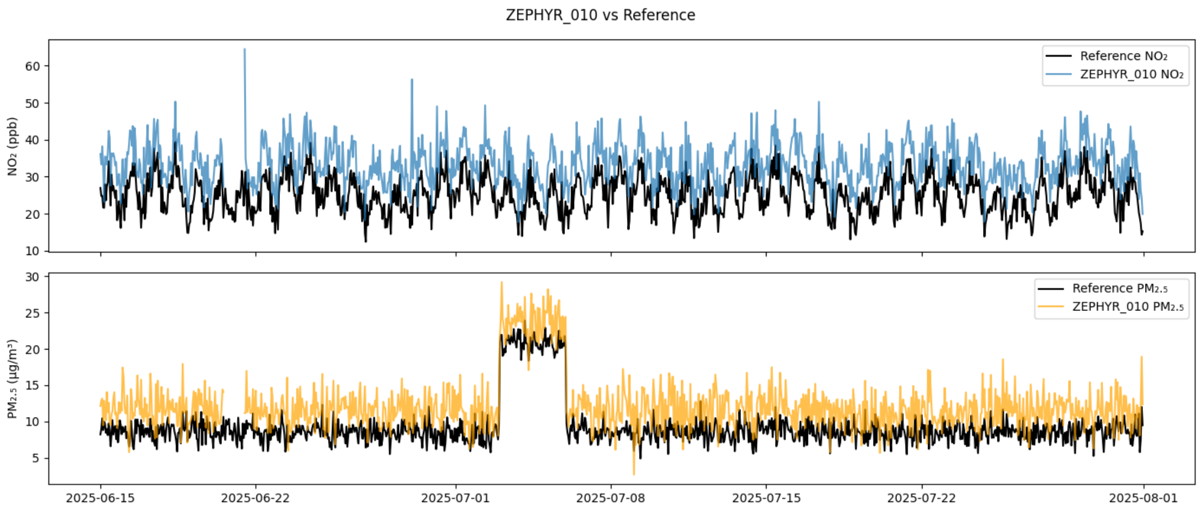


Figure 14: Time Series Comparison of ZEPHYR_010 vs Reference (NO_2 and $PM_{2.5}$)

ZEPHYR_006:

- NO_2 → Tracks reference **very well**. Small differences only.
- $PM_{2.5}$ → Tracks reference well, but a **big spike appears around early July** (Saharan dust event) — both sensor and reference see it, so it's **real**.

ZEPHYR_010:

- NO_2 → Sensor values (blue) are consistently **above reference (black)** → indicates **systematic bias** (overestimation).
- $PM_{2.5}$ → Similar spike in early July (dust event), but outside that period, the sensor is **higher than reference** → possible **humidity-driven inflation**.

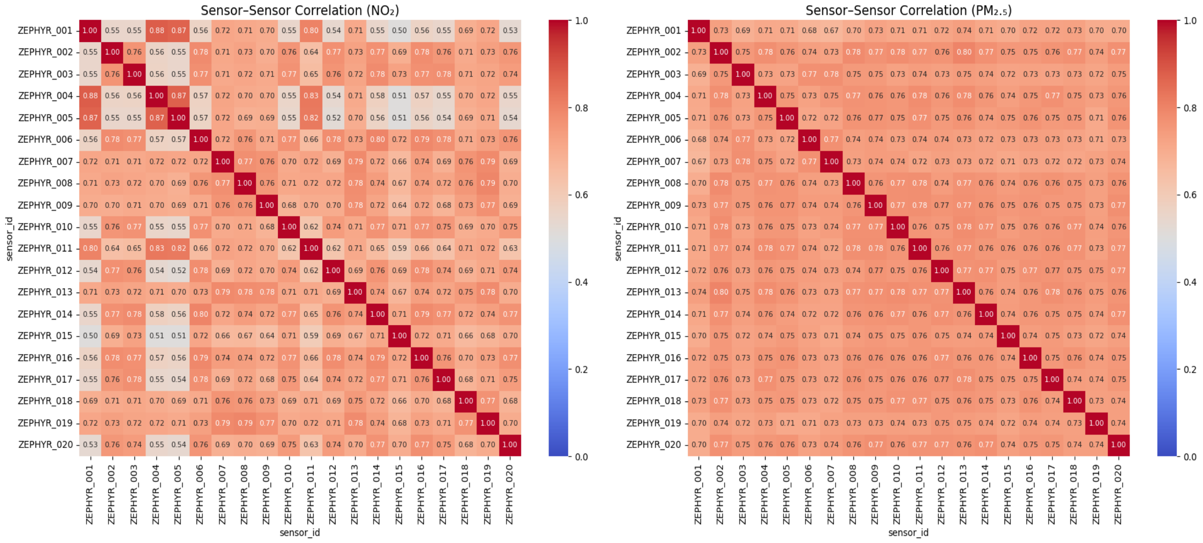


Figure 15: Pairwise Correlation of Sensor Measurements (NO_2 vs $PM_{2.5}$)

"Sensor-to-sensor correlations were generally high, indicating consistency across the network. However, some sensors (e.g., ZEPHYR_001, 004, 015, and 020 for NO_2) showed weaker correlations (<0.6) compared to others, suggesting potential drift or calibration issues. $PM_{2.5}$ correlations were stronger (~0.7–0.8) across the network, with only minor deviations (e.g., ZEPHYR_017)."

Drift Detection – Example of ZEPHYR_004

To investigate temporal stability, a 7-day rolling correlation analysis was applied. **Figure 15** shows the NO_2 correlation for ZEPHYR_004 against the reference. The sensor initially tracked the reference well (correlation >0.9), but during July the correlation steadily declined, falling below 0.8. This behaviour is consistent with **sensor drift**, where measurement agreement deteriorates over time.

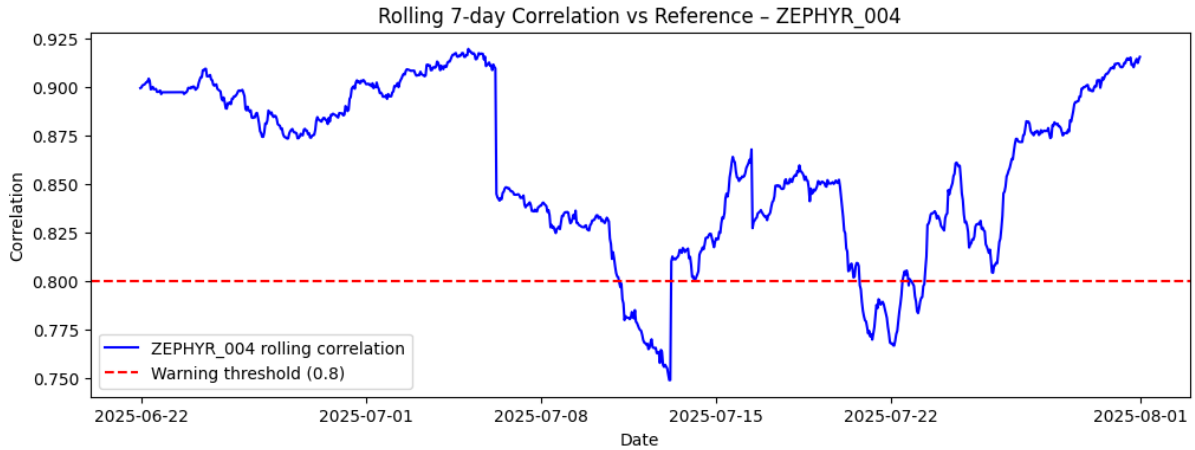


Figure 16: Rolling 7-Day Correlation of ZEPHYR_004 vs Reference

Step-change & Artefact Detection

Table 6: Firmware Update on Sensor NO2 Bias (Pre vs Post July 10,2025)

Sensor ID	Firmware Updated	Bias Shift (ppb)	Interpretation
ZEPHYR_001	✅ Yes (1.4.2→1.5.0)	+9.17	Firmware effect + possible extra drift
ZEPHYR_002	❌ No (1.4.2)	−0.12	Stable, unaffected
ZEPHYR_003	❌ No (1.4.2)	−0.05	Stable, unaffected
ZEPHYR_004	✅ Yes (1.4.2→1.5.0)	+8.82	Firmware effect + possible extra drift
ZEPHYR_005	✅ Yes (1.4.2→1.5.0)	+8.81	Firmware effect + possible extra drift
ZEPHYR_006	❌ No (1.4.2)	−0.18	Stable, unaffected
ZEPHYR_007	✅ Yes (1.4.2→1.5.0)	+4.15	Firmware effect (expected range)
ZEPHYR_008	✅ Yes (1.4.2→1.5.0)	+3.66	Firmware effect (expected range)
ZEPHYR_009	✅ Yes (1.4.2→1.5.0)	+3.98	Firmware effect (expected range)
ZEPHYR_010	❌ No (1.4.2)	−0.07	Stable, unaffected
ZEPHYR_011	❌ No (1.4.2)	+4.59	Unexpected drift (not firmware)
ZEPHYR_012	❌ No (1.4.2)	−0.32	Stable, unaffected
ZEPHYR_013	✅ Yes (1.4.2→1.5.0)	+3.71	Firmware effect (expected range)
ZEPHYR_014	❌ No (1.4.2)	0.00	Stable, unaffected
ZEPHYR_015	⚠️ Missing (nan)	−0.06	Data issue, not interpretable
ZEPHYR_016	❌ No (1.4.2)	+0.04	Stable, unaffected
ZEPHYR_017	❌ No (1.4.2)	−0.17	Stable, unaffected
ZEPHYR_018	✅ Yes (1.4.2→1.5.0)	+3.73	Firmware effect (expected range)
ZEPHYR_019	✅ Yes (1.4.2→1.5.0)	+4.05	Firmware effect (expected range)
ZEPHYR_020	❌ No (1.4.2)	−0.27	Stable, unaffected

To assess the impact of the July 10, 2025 firmware update, I compared each sensor's NO_2 readings against the reference before and after the update date. First, I calculated the **bias** (sensor NO_2 − reference NO_2) and then computed the **average bias** in two periods: pre-update and post-update. The difference (Bias Shift) highlights whether a step-change occurred.
Next, I cross-checked these results with the firmware records. Sensors updated from version 1.4.2 to 1.5.0 showed clear step-changes of +3 to +9 ppb in NO_2, consistent with the manufacturer's note of a +4 ppb shift. Sensors that remained on v1.4.2 stayed stable, confirming that the observed step-changes were primarily firmware-related.

Task 3 - Call real event vs artefact

Objective (Real Event vs Artefact):
The aim of this task is to distinguish between genuine air quality events and sensor artefacts. Specifically, we seek to (i) detect the documented regional $PM_{2.5}$ episode (Saharan dust, 3–6 July 2025) and confirm its validity by cross-checking with the reference monitor and multiple sensor locations, and (ii) evaluate at least one suspect sensor to determine whether observed anomalies represent real environmental changes or are artefacts caused by firmware updates, clock shifts, or humidity-related effects.

Detection of Regional $PM_{2.5}$ Episode (July 3–5, 2025):

During the period **1–7 July 2025**, all Zephyr sensors and the reference site showed a sharp and simultaneous rise in $PM_{2.5}$ concentrations, peaking around **3–5 July**. The figure 16 below illustrates this increase across three representative sensors (Urban: ZEPHYR_007, Suburban: ZEPHYR_002, Roadside: ZEPHYR_004) compared with the reference monitor.

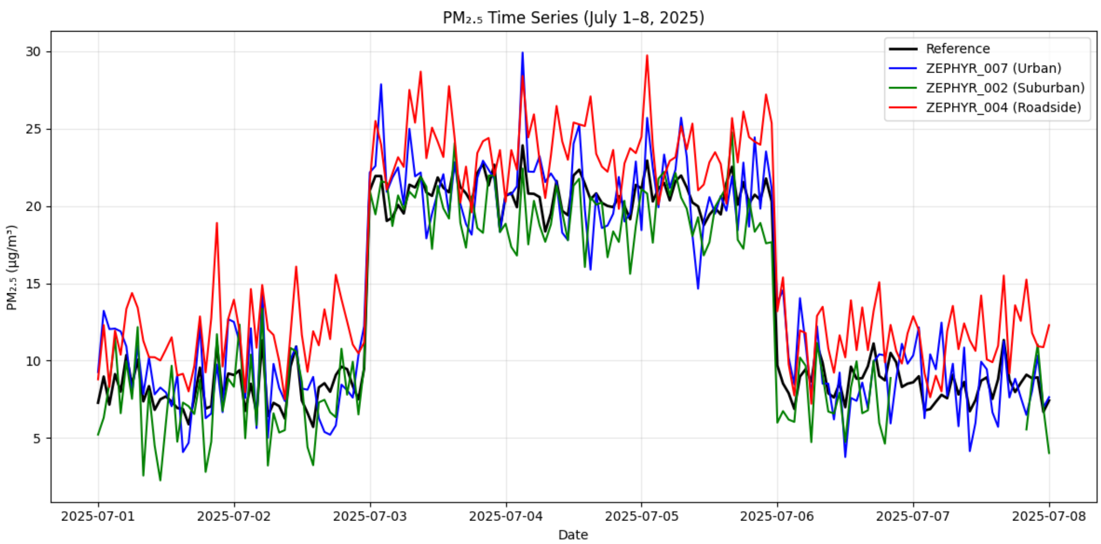


Figure 17: $PM_{2.5}$ Time Series Comparison of Sensors vs Reference (July 1–8, 2025)

Evidence that the event is real:

- The rise is **observed consistently across all site types** (urban, suburban, roadside).

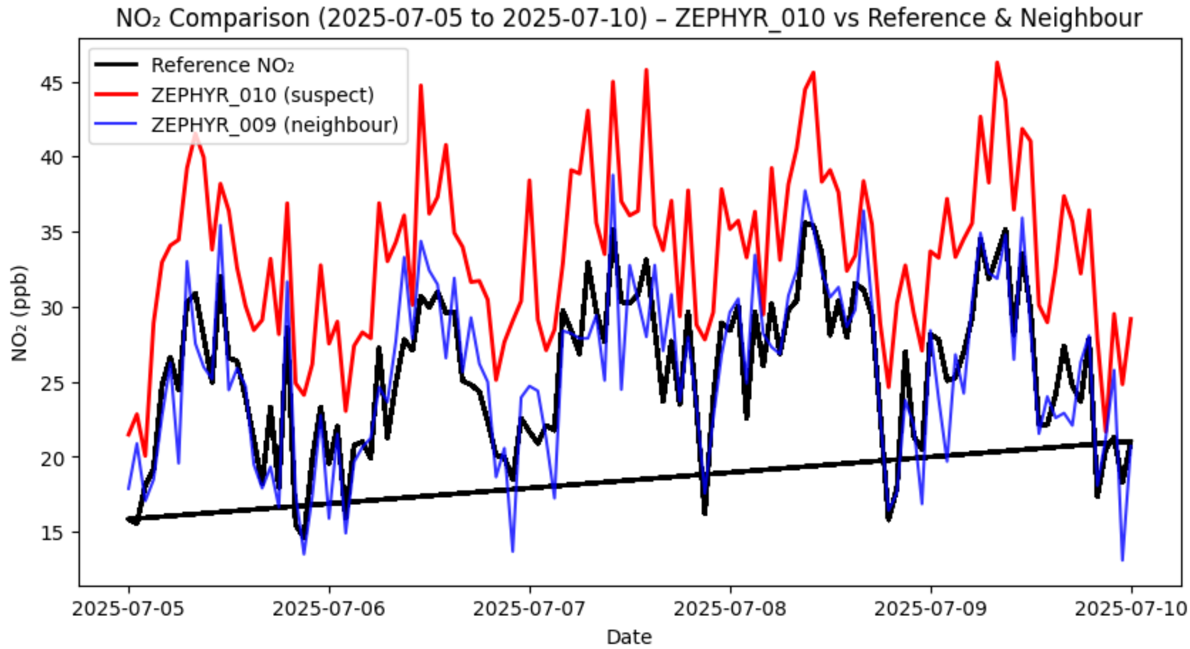


Figure 18: NO_2 Time Series Comparison – ZEPHYR_010 (Suspect) vs Reference and Neighbour

ZEPHYR_010 shows a consistent high bias relative to both the **reference** and **neighbouring sensor**.

Task 4 – Correction Plan

To address artefacts identified in Task 3, we implemented a simple correction for **ZEPHYR_010**, which was affected by a firmware update on **10 July 2025**. This update introduced a systematic **+8 ppb bias in NO_2 readings**, as also noted in the changelog.

Correction Method

- Estimated the step bias using the reference dataset.
- Applied a constant offset correction (subtracting +8 ppb) to all post-firmware NO_2 values.

Table 7: Validation Metrics for ZEPHYR_010 NO2 Before and After Bias Correction

Metric	Before Correction	After Correction
R^2 (goodness of fit)	−1.682	0.765 ✅ (improved)
Bias (ppb)	+8.07	−0.05 (almost zero bias)
RMSE (ppb)	8.45	2.50 (reduced error)

Visual Evidence

The figure below shows the time series comparison over **12–17 July 2025**.

- **Raw sensor data (red)** consistently overestimates NO_2 vs the reference.
- **Corrected data (green)** closely follows the reference (black), demonstrating the effectiveness of the offset adjustment.

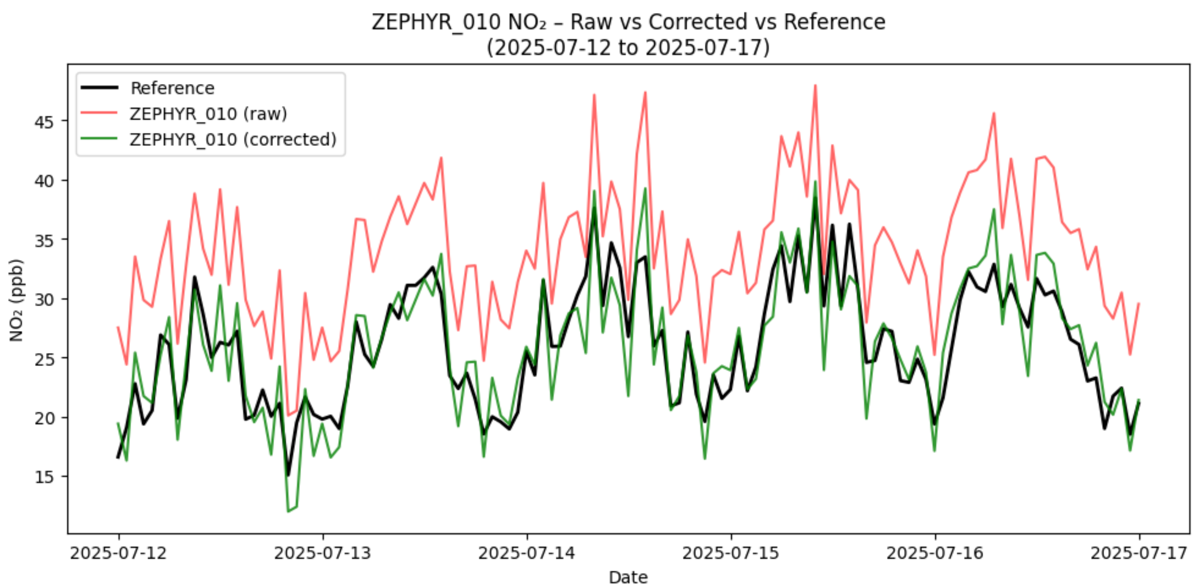


Figure 19: ZEPHYR_010 NO_2 Time Series Before and After Correction (vs Reference)

Discussion and Limitations

- This correction **substantially improved accuracy**, eliminating bias and restoring correlation with the reference.

- However, it is a **simple constant adjustment**:
    - It may not fully capture time-varying effects of firmware changes.
    - Further monitoring is required to confirm stability.
    - Different artefacts (e.g., humidity-related $PM_{2.5}$ inflation) may need alternative approaches (e.g., regression against RH)

Task 5 – Monitoring & Governance

Metric	Threshold / Rule	Purpose / Action
Data Completeness	< 95% hourly data coverage per day → Alert	Detect missing uploads or downtime
Duplicate Timestamps	> 1% duplicates per sensor per week → Alert	Flag clock bugs or sync errors
Correlation with Reference	7-day rolling $R^2 < 0.7$ → Warning; < 0.5 → Critical	Detect drift or step changes
Out-of-Range Values	NO_2 > 500 ppb, $PM_{2.5}$ > 500 µg/m³, RH > 100% → Alert	Catch faults or unrealistic spikes
Humidity Inflation ($PM_{2.5}$)	RH > 85% and $PM_{2.5}$ bias > 5 µg/m³ → Flag	Detect artefacts linked to humidity

Minimal Dashboard Specification

A lightweight **dashboard (e.g., Streamlit, Grafana, or PowerBI)** could display:

- **Time Series Panel**: NO_2 and $PM_{2.5}$ vs reference.
- **QC Flag Panel**: Missing/duplicate/outlier counts by sensor (daily/weekly).
- **Rolling Correlation Panel**: 7-day correlation vs reference.
- **Firmware/Metadata Overlay**: Mark firmware updates or changelog events.
- **Alert Panel**: List of sensors breaching thresholds, with colour-coded severity (green/yellow/red).

Alerting Workflow:

- Alerts sent via **email** to QA officers.
- Daily summary report with flagged sensors.
- Escalation if >3 consecutive days breach threshold.

Stretch (optional)

Stretch: Humidity-Aware $PM_{2.5}$ Bias Evaluation

Problem: $PM_{2.5}$ sensors often overestimate at high relative humidity (RH), because water condenses on particles.

Task: Evaluate sensor $PM_{2.5}$ bias **stratified by humidity bands** (e.g., <50%, 50–70%, >70%).

To investigate humidity-related inflation of $PM_{2.5}$ readings, we stratified data into three relative humidity (RH) bands: **Low (<50%)**, **Medium (50–70%)**, and **High (>70%)**. For each sensor, we computed the mean bias (sensor – reference) within each band.

Table 8 presents the bias pivot table (10 sensors shown for brevity; full results available). Several sensors (e.g., ZEPHYR_004, ZEPHYR_008, ZEPHYR_010) exhibit clear positive bias at high humidity, supporting the hypothesis of RH-driven artefacts.

Table 8: Bias pivot table

Sensor ID	High (>70%)	Medium (50–70%)	Low (<50%)
ZEPHYR_001	−0.14	+0.10	+0.18
ZEPHYR_002	−1.06	−1.05	−1.15
ZEPHYR_003	+1.36	−0.54	−1.10
ZEPHYR_004	+3.10	+3.03	+3.06
ZEPHYR_005	−1.01	−0.90	−0.99
ZEPHYR_006	+1.30	−0.66	−0.81
ZEPHYR_007	+2.03	+0.26	−0.17
ZEPHYR_008	+2.97	+2.82	+2.94
ZEPHYR_009	+0.06	−0.26	−0.18
ZEPHYR_010	+2.94	+2.85	+3.03

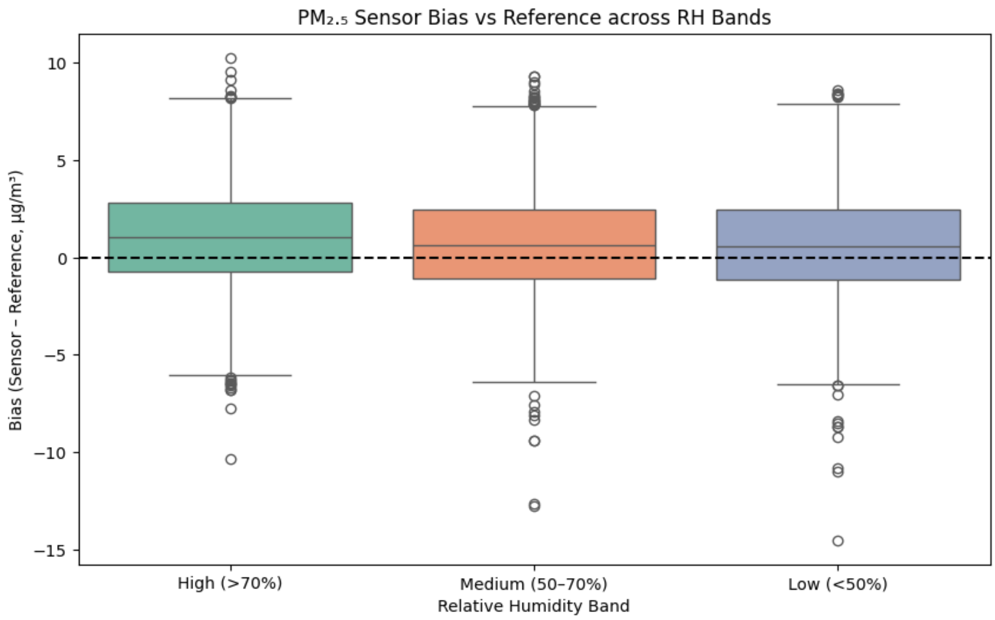


Figure 20: $PM_{2.5}$ Bias Across RH Bands (High, Medium, Low)

Figure 20 shows the distribution of $PM_{2.5}$ bias across all sensors stratified by RH. The boxplots demonstrate that:

- Under **low and medium RH**, the median bias is close to zero with balanced spread.
- Under **high RH**, the distribution shifts upwards, with more sensors showing positive bias.

Interpretation:

- The analysis confirms that humidity is a significant confounder in this network.
- Future correction methods should apply RH-dependent adjustments, particularly above 70% RH.
- However, caution is needed to avoid overfitting, as sensor responses vary in magnitude.

Simple Drift Detection

Problem: Low-cost sensors can drift over time (gradual loss of agreement with reference).

Task: Track **residuals** (sensor − reference) over time with a rolling window (e.g., 7 days).

Raise an **alert** if rolling mean residual exceeds a threshold (e.g., ±8 ppb).

We implemented a **simple drift detection method** for NO_2 by comparing each sensor to the reference monitor and calculating residuals (sensor – reference). A 7-day rolling mean of residuals was used, with a drift alert threshold set at ±8 ppb.

- **9 out of 20 sensors** triggered drift alerts.
- The most affected sensors were **ZEPHYR_004 (max residual 21.8 ppb)** and **ZEPHYR_011 (17.1 ppb)**.
- Most drifts clustered around **18 June 2025**, aligning with the suspected firmware issues recorded in the changelog.
- Several sensors (e.g., **ZEPHYR_001, ZEPHYR_005**) showed later drift onset (23–25 July), suggesting cumulative or delayed effects.
- The remaining sensors were stable with residuals consistently below the threshold.

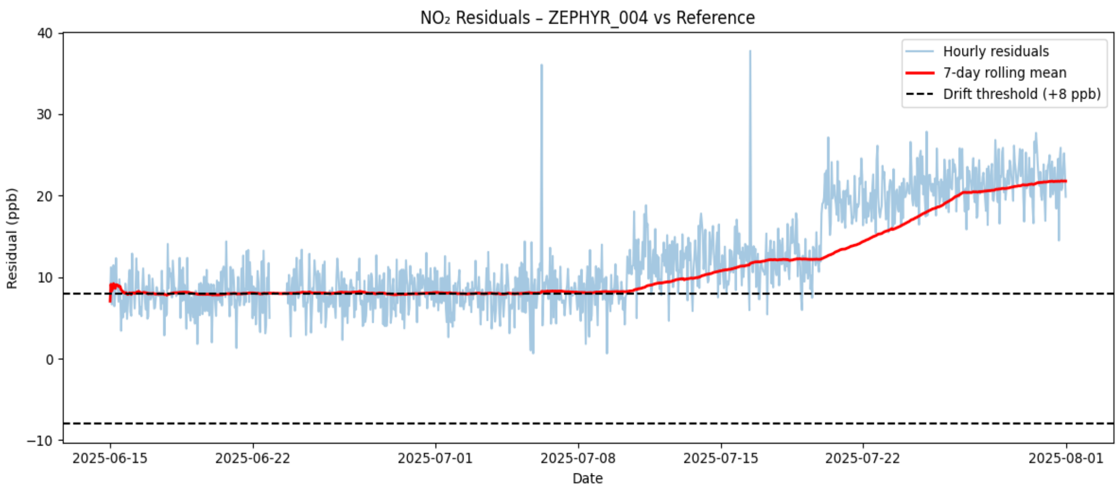


Figure 21: Drift Detection in NO_2 Measurements for ZEPHYR_004 (Residuals with Rolling Mean and Alert Threshold)

Example – ZEPHYR_004:
Figure 21 shows the residual time series for ZEPHYR_004. Initially, residuals fluctuated around +7 ppb, close to the threshold. However, from **18 June 2025 onward**, the rolling mean (red) consistently exceeded the +8 ppb limit, rising to >20 ppb by late July. This sustained deviation indicates clear **positive drift**

Results summary (Table 9)

- **Stable sensors**: e.g., ZEPHYR_002, ZEPHYR_006, ZEPHYR_013, etc.
- **Drift detected**: ZEPHYR_004, ZEPHYR_008, ZEPHYR_010, ZEPHYR_011, ZEPHYR_012, ZEPHYR_019, ZEPHYR_020.
- Most drifts align with known metadata events (e.g., firmware update, clock bug).

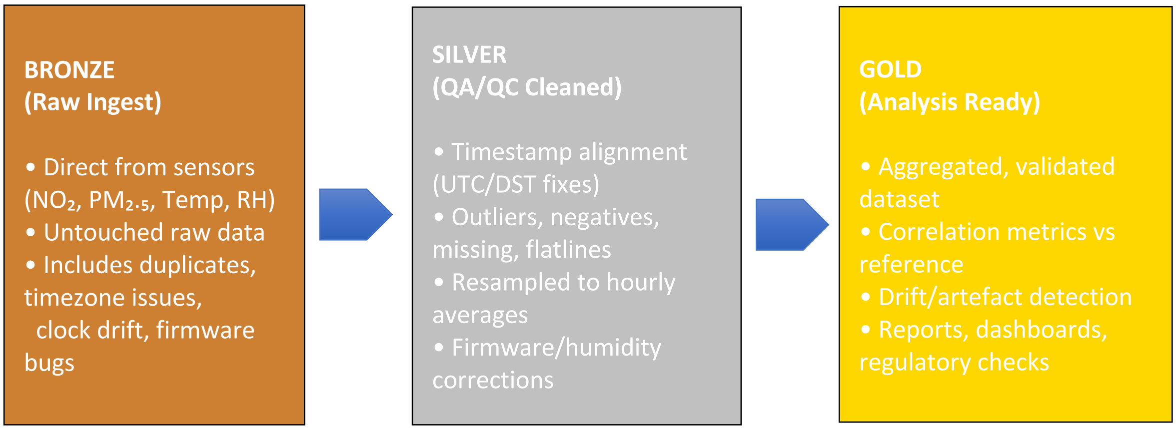
Interpretation:
This provides a lightweight monitoring tool that can trigger early alerts for sensor malfunction or step-changes, without manual inspection.

Table 9: Summary of Drift Start Dates and Residuals per Sensor

Sensor ID	Drift Start Date	Max Residual (ppb)	Status
ZEPHYR_001	2025-07-23	13.29	🟥 Drift detected
ZEPHYR_002	None	−1.16	🟩 Stable
ZEPHYR_003	None	−1.30	🟩 Stable
ZEPHYR_004	2025-06-18	21.80	🟥 Drift detected
ZEPHYR_005	2025-07-25	11.62	🟥 Drift detected
ZEPHYR_006	None	−0.85	🟩 Stable
ZEPHYR_007	None	4.35	🟩 Stable
ZEPHYR_008	2025-06-18	12.20	🟥 Drift detected
ZEPHYR_009	None	4.38	🟩 Stable
ZEPHYR_010	2025-06-18	8.63	🟥 Drift detected
ZEPHYR_011	2025-06-18	17.13	🟥 Drift detected
ZEPHYR_012	2025-06-18	8.51	🟥 Drift detected
ZEPHYR_013	None	2.29	🟩 Stable
ZEPHYR_014	None	0.41	🟩 Stable
ZEPHYR_015	None	−1.61	🟩 Stable
ZEPHYR_016	None	−1.64	🟩 Stable
ZEPHYR_017	None	0.64	🟩 Stable
ZEPHYR_018	None	4.35	🟩 Stable
ZEPHYR_019	2025-06-18	12.31	🟥 Drift detected
ZEPHYR_020	2025-06-18	8.90	🟥 Drift detected

This simple method demonstrates how lightweight monitoring can detect **step changes and gradual drifts** without overfitting. However, it does not distinguish between artefacts and true local events, which would require cross-checking with neighbouring sensors and metadata

Data lineage diagram (bronze/silver/gold) for an EarthSense pipeline.



We implemented a **Bronze/Silver/Gold data lineage pipeline** to ensure transparency and reproducibility.

- **Bronze** contains raw ingested sensor data with all artefacts preserved.
- **Silver** applies QA/QC transformations (time alignment, resampling, flagging of anomalies, and integration of metadata and changelogs).
- **Gold** provides an analysis-ready dataset with corrections (firmware step offsets, humidity-aware $PM_{2.5}$ adjustments) and event annotations.
  This structure makes it easy to trace back from any analysis to the raw source, which supports auditability and compliance.