

IoT-Based Fault-Aware Tropical Microclimate Data Evaluation: Single-Day Preliminary Study

Carmen Claudia

Pradita University

carmen.claudia@student.pradita.ac.id

Submitted : Jul 28, 2026 | Accepted : Sep 7, 2026 | Published : Oct 1, 2026

Abstract: The deployment of affordable Personal Weather Stations (PWS) may expose a set of Citizen IoT Vulnerabilities, including hardware siting bias, sensor faults, and temporal-resolution mismatches, that remain under-documented. This preliminary single-day, Single-Site Case Study explores the cataloguing of those gaps in a tropical Microclimate setting. This paper evaluates a custom ESP32-based IoT station deployed for a single day (4 May 2026) in Curug Sangereng, Tangerang, Indonesia, sampling at 30-second intervals. Using a Fault-Aware Quality Control protocol and an autocorrelation-adjusted effective sample size (n_{eff}) to correct for Temporal Pseudo-Replication, the IoT record was benchmarked against Open-Meteo's 15-minute forecast, BMKG (3-hourly forecast), and Meteostat (hourly archive) datasets. Variance decomposition showed that sub-15-minute fluctuations account for up to 53.9% of total-day wind-speed variance (and up to 67.1% for sub-3-hour fluctuations), compared to under 1.5% and 38% for temperature, pressure, and relative humidity, demonstrating that high-frequency sampling value is variable-dependent rather than universally necessary. Cross-dataset comparisons revealed severe hardware siting biases: Open-Meteo ran 3.74°C cooler than the IoT station, while BMKG ran warmer than Meteostat ($+2.48^{\circ}\text{C}$) but cooler than IoT (-1.39°C). This sign reversal reflects unshielded rooftop heat bias at the IoT site. Poor wind-speed agreement across all pairs further indicated local anemometer obstruction. These single-day, single-site results are offered as a catalogue of PWS setup pitfalls and a worked diagnostic protocol. They motivate four falsifiable hypotheses for a multi-day confirmatory study rather than establishing general conclusions.

Keywords: Citizen IoT Vulnerabilities, Fault-Aware Quality Control, Microclimate, Single-Site Case Study, Temporal Pseudo-Replication

INTRODUCTION

The proliferation of open-access weather APIs and low-cost IoT hardware has democratized hyper-local microclimate monitoring. Decentralized Personal Weather Station (PWS) networks increasingly supplement official meteorological networks (Brousse et al., 2024). Three methodological gaps undermine existing citizen-IoT evaluation studies and motivate this pilot. First, unshielded, uncalibrated sensors co-located with heat-generating microcontrollers introduce thermal contamination. This contamination is often misattributed to forecast-model failure rather than the hardware itself (DeFord et al., 2025). Second, comparing high-frequency IoT readings against step-constant forecasts induces temporal pseudo-replication. Uncorrected autocorrelation understates standard errors and yields spuriously narrow confidence intervals. Correcting for it requires an effective sample size (n_{eff}) or common-grid resampling (Afyouni et al., 2019; Chen et al., 2021). Third, studies routinely conflate predictive forecasts, such as BMKG's 3-hourly sub-district predictions, with observational baselines (Garcia-Marti et al., 2026; Herlambang, 2023). They also often evaluate against only one reference instrument. This creates a blind spot in which an uncorrected sensor-siting error can invert the apparent sign of a forecast product's bias. Underneath all three sits a largely unasked question. Does a PWS's very high sampling rate actually strengthen anything analytically, or is it discarded once the record is resampled to a coarser, WMO-consistent reference?

Answering that question matters because PWS adoption is outpacing the methodological tools available to validate it. Without a fault-aware protocol, hardware and statistical pitfalls risk being reported as forecast-model

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

error, misleading downstream users. This study presents an exploratory, single-day pilot deployment of a custom ESP32-based PWS node in Curug Sangereng, Tangerang, Banten. It produces local IoT observations rather than a calibrated ground reference. The study delivers three contributions distinguishing it from prior single-pair, single-reference PWS evaluations. First, a fault-aware analytical framework integrates physical fault-window isolation, autocorrelation-adjusted n_{eff} , common-grid resampling, and OLS error decomposition to decouple hardware artifacts from forecast-model discrepancy. Second, a granularity-value test applies variance decomposition to check whether sub-15-minute and sub-3-hour sampling resolves meaningful signal or only redundant noise relative to WMO-consistent cadences. Third, a multi-source design evaluates BMKG's forecast against three independent sources, the IoT node, Open-Meteo, and Meteostat, rather than one, so that single-instrument bias misattribution can be exposed rather than assumed away.

To operationalize this protocol for a future multi-day, multi-site confirmatory study, four candidate hypotheses are formulated from this pilot's results. H1 (granularity) posits that high-frequency PWS sampling resolves a non-trivial, variable-dependent share of sub-hourly variability invisible to coarser products. This holds only where sampling that finely is worth its cost. H2 (hardware/siting bias) posits that correcting fault windows and pseudo-replication isolates a systematic additive bias exceeding 1°C for temperature. Pressure is least sensitive to micro-siting. H3 (single-reference) posits that BMKG's forecast-bias direction stays consistent against two independent sources. It should not be an artifact of whichever single instrument a one-station pilot uses. H4 (spatial representativeness) posits that coarse-cadence products under-represent short-lived convective extrema that a high-frequency PWS can resolve. This can hold even when the PWS's own location fails to corroborate the event. The pilot data below show only that each pitfall is measurable, not confirmed.

LITERATURE REVIEW

Temporal Resolution: Undersampling versus Standards-Compliant Adequacy

Tropical convection and boundary-layer processes vary on sub-3-hour timescales, so coarse sampling risks missing peak intensities and diurnal-cycle structure. Large-eddy simulations show that only hourly-or-finer sampling reveals the multi-peak diurnal structure that coarser records collapse into one broad maximum (Wang et al., 2024). This is consistent with earlier Jakarta radiosonde work showing that 6-hourly sampling undersamples diurnal evolution relative to 3-hourly (Katsumata et al., 2018). Maritime Continent evidence also shows that high-resolution satellite precipitation products struggle with peak timing rather than totals (Ramadhan et al., 2024), and reanalysis products such as ERA5 similarly lag the observed diurnal peak (Mustafa et al., 2024). Conversely, WMO guidance sanctions computing daily climate normals from eight 3-hourly observations (World Meteorological Organization, 2017). So 3-hourly sampling is neither uniformly adequate nor inadequate; it depends on the application's own timescale, operationalized directly by the Granularity-Value Method below.

Low-Cost IoT Sensors and Statistical Agreement Methodology

Decentralized networks of low-cost, ESP32-class IoT sensors and citizen-operated PWS are increasingly deployed to fill gaps left by official meteorological infrastructure (Mazzoglio et al., 2026), including agriculture- and API-oriented systems for Indonesian contexts (Idris et al., 2025). Crowdsourced PWS networks have also been used to forecast urban temperatures (Romero Rodríguez et al., 2024), broadly the use this paper puts its own IoT record to. Measurement uncertainty and calibration remain their most persistent limitations (Al Kalaany et al., 2025). Dedicated QC and bias correction, including machine-learning post-processing, are essential before raw low-cost data are deployment-ready (Beele et al., 2022; Nan et al., 2025), though under favorable siting they can achieve accuracies within a few tenths of a degree for temperature (Chen et al., 2021). Strong linear relationships can coexist with large constant offsets, so environmental sensor validation relies on the Bland-Altman limits-of-agreement method (Bland & Altman, 1986); recent guidance notes that differences should be plotted against a designated reference standard rather than pairwise means when one source serves as baseline (Cesana & Antonelli, 2024).

Open-access APIs and data-aggregation platforms such as Open-Meteo and Meteostat have become standard sources across engineering, urban-climate, and machine-learning research. Smith et al. (2025) used Meteostat observations to evaluate gridded datasets, among other applications. Open-Meteo alone has supplied inputs for HVAC simulation (Dadras Javan et al., 2025) and deep-learning weather-forecasting benchmarks (Pavithran et al., 2025). This reliance is not risk-free. Internal model ensembles within one provider can show non-uniform accuracy (Jankauskas et al., 2024). This raises a question: is the margin between global, model-derived weather products and micro-regional conditions better characterized at the forecast's own cadence, or at the finer cadence a PWS can sample. This paper reports a single-day, exploratory case study that deploys one affordable IoT station to measure, not assume, the value of its higher sampling rate.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

METHOD

Study Design and Data Sources

Four sources are compared for Curug Sangereng, Tangerang, Banten. The first is a custom IoT station positioned 6 m above ground, sampled every 30 seconds for a single day, 4 May 2026. The second is BMKG's public forecast product, retrieved at its native 3-hourly cadence for the same date. The third is Meteostat, an hourly archive of recorded observations. The fourth is Open-Meteo, sampled at its own native 15-minute forecast cadence rather than the hourly cadence used in an earlier revision (Zippenfenig, 2023). The BMKG file is a forecast issued in advance of its valid time, not a recorded measurement. This differs from Meteostat's or the IoT station's direct observations. IoT and Meteostat therefore serve as observational references. BMKG's forecast and Open-Meteo are both treated as predictive products under evaluation.

Of six possible source pairs, three are reported here. BMKG–Meteostat ($n = 8$, both full days) is unchanged from the prior revision. IoT–Open-Meteo is re-run on a fault-aware common grid built from Open-Meteo's update timestamps (median spacing ≈ 15 minutes), giving $n = 61$ for temperature, pressure, and humidity and $n = 116$ for wind speed. IoT–BMKG is new: IoT readings are averaged within ± 90 -minute windows centered on each of BMKG's eight 3-hourly valid times, giving $n = 8$ for wind speed and rainfall and $n = 5$ for temperature and humidity once fault-affected windows are excluded.

Hardware, Acquisition, and Siting Limitations

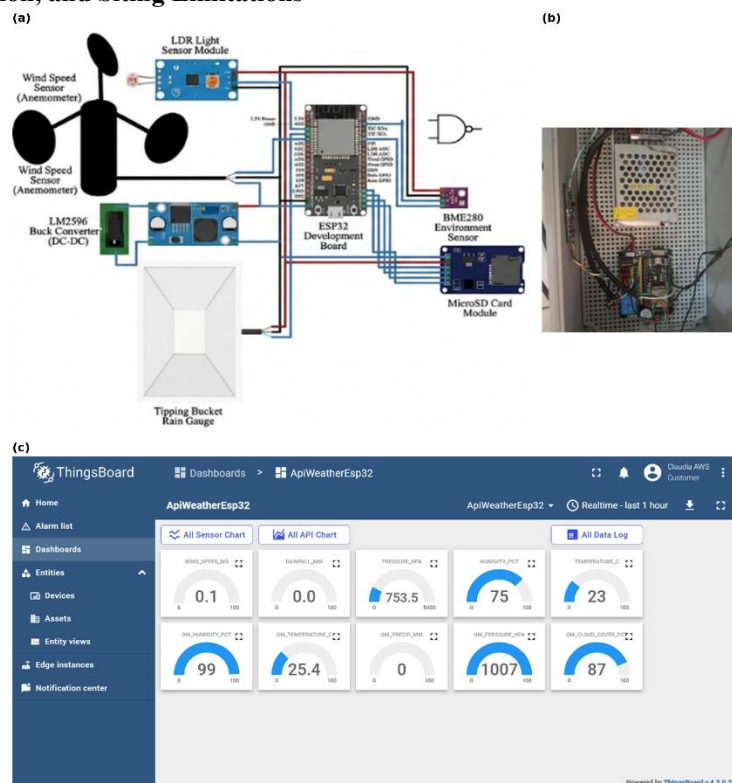


Fig. 1 (a) IoT hardware schematic with sensor and module identification; (b) internal component configuration and wiring layout of the deployment enclosure; (c) ThingsBoard dashboard for the data collection and integration platform

The IoT station used an ESP32-S3 microcontroller in an IP67 enclosure with a BME280 (temperature, pressure, humidity), a cup anemometer, an LDR light sensor, MicroSD backup logging, and NTP-synchronized 30-second telemetry to a ThingsBoard dashboard (Fig. 1). It was mounted roughly 6–7 m above ground on a residential rooftop for a single 24-hour window and was not independently calibrated. It lacked a dedicated radiation shield (DeFord et al., 2025), sat below WMO reference height, shared an enclosure with heat-generating electronics, and had no documented anemometer clearance. These are all recognized bias sources in low-cost deployments (Kim et al., 2022). These limitations mean this deployment functions as a preliminary proof-of-concept, not a calibrated or generalizable instrument.

Temporal Alignment and Fault-Aware Quality Control

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Two comparison grids are used throughout: a native-cadence grid (finest resolution common to a pair) to probe sub-3-hour fidelity, and a common 3-hourly grid (simple-mean resampling to BMKG's eight-observation schedule, per World Meteorological Organization, 2017) to probe daily-aggregate adequacy.

Predictive weather data were retrieved from the Open-Meteo API endpoint, <https://api.open-meteo.com/v1/forecast>, for the study location in Curug Sangereng, Tangerang, Banten, Indonesia ($\sim 6.2542^{\circ}$ S, 106.6218° E) for the evaluation period of 4 May 2026. Open-Meteo was evaluated as a predictive model product, utilizing its operational "Seamless / Best-Match" model selection mode that dynamically combines global numerical weather prediction systems, primarily ECMWF IFS, DWD ICON-Global, and NOAA GFS, operating at native spatial resolutions of 9 km to 13 km. Terrain downscaling was automatically applied by Open-Meteo using the 90-meter Copernicus Digital Elevation Model (GLO-90) to adjust model fields to the site's ground surface elevation of approximately 29 m above sea level, corresponding to the physical deployment site where the local IoT station was mounted on a residential house's roof 6 to 7 m above ground level. Global model runs initialize every 6 to 12 hours (00:00, 06:00, 12:00, 18:00 UTC) across a standard 7-day forecast horizon. Sub-hourly forecast data were queried at a native 15-minute temporal update cadence, which applies linear and physical temporal interpolation across the underlying 1-hour and 3-hour raw model time steps. The retrieved API parameters, temperature_2m, relative_humidity_2m, surface_pressure, wind_speed_10m, precipitation, and cloud_cover, were localized to Western Indonesia Time (WIB, UTC+7) and mapped directly to corresponding ground-station metrics. These API streams were parsed in real time alongside local ESP32 telemetry for side-by-side visualization and fault-aware comparative evaluation within the open-source ThingsBoard IoT dashboard.

ThingsBoard was selected as the open-source IoT device management and visualization platform. Telemetry collected by the ESP32 telemetry was transmitted to ThingsBoard via HTTP GET requests, where local sensor data and Open-Meteo API variables were rendered side-by-side on graphical widgets. A sensor-communications fault beginning at 12:52:27 WIB affected temperature, pressure, and humidity (detailed under Temporal Alignment and Fault-Aware Quality Control); wind speed, on an independent circuit, remained plausible throughout. Paired records for each source pair and variable are built only from timestamps valid in both members after fault-aware screening; n is reported for every statistic since it varies by variable, pair, and grid.

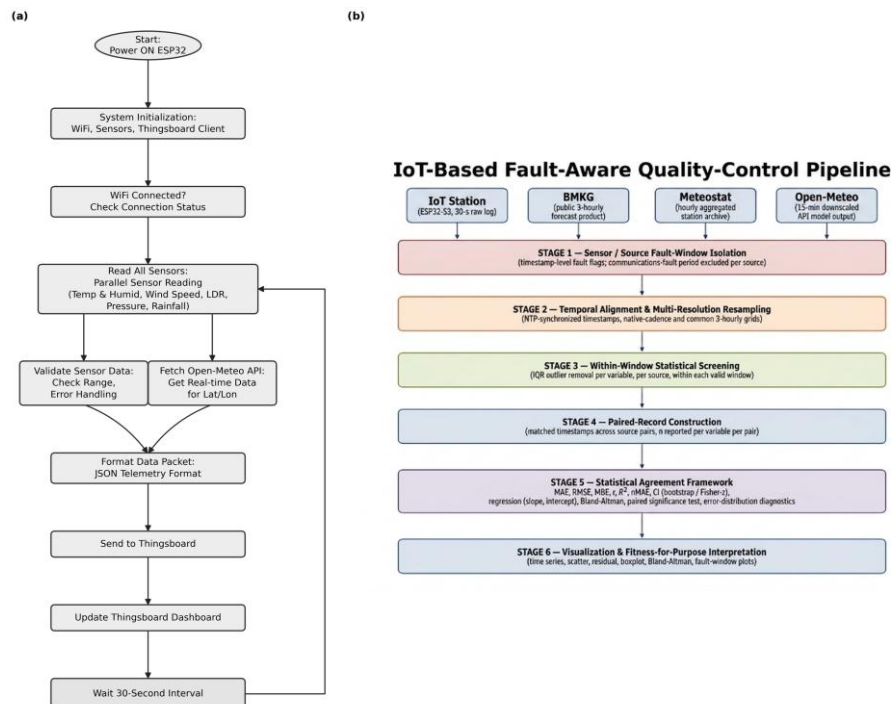


Fig. 2 (a) firmware execution flowchart (single-station, single-day preliminary deployment); (b) fault-aware, multi-source quality-control pipeline

Raw 30-second IoT logs showed a sensor-communications fault beginning at 12:52:27 WIB. Temperature readings collapsed to two fixed, alternating values (22.52°C and 182.23°C), pressure readings jumped to physically impossible values (including negative hPa), and humidity flattened to a near-constant value with intermittent excursions to 100%. Because a stuck reading can fall within a physiologically plausible range, as 22.52°C is an ordinary tropical night-time temperature. IQR filtering alone cannot isolate it. A two-stage pipeline

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

is used for every source: (1) fault-window isolation by inspection, excluding flagged timestamp ranges per variable per source, then (2) within-window IQR screening, consistent with established QC practice for crowdsourced, low-cost sensor data (Mazzoglio et al., 2026). Temperature, pressure, and humidity are restricted to 00:00–12:52 WIB ($n = 1,540$ of 2,873 raw records). Wind speed, on an independent circuit, showed no corresponding discontinuity, with the second stage removing 112 records ($n = 2,761$ retained).

For the IoT–Open-Meteo pair specifically, the native-cadence grid is retained only to demonstrate the pseudo-replication problem directly, as shown below under IoT versus Open-Meteo. The common grid is used for every inferential statistic reported afterward, since it avoids treating up to thirty repeated copies of the same 15-minute forecast value as independent observations.

Sensor Fault Detection and Data Quality Control

Fig. 3 visualizes this fault episode, plotting the verified valid window against the flagged fault period for each channel. It makes the two-stage protocol's necessity visible rather than asserted: a purely statistical outlier test applied to the whole record would not isolate this window, since the stuck values it flags fall inside the marked period rather than outside the overall distribution.

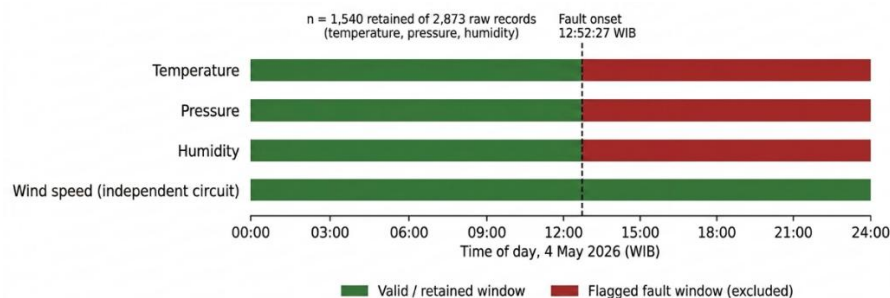


Fig. 3 Sensor-fault window across monitored variables, 4 May 2026. The shaded band marks the post-12:52:27 WIB fault period excluded for temperature, pressure, and humidity; wind speed remained on an independent, unaffected circuit throughout.

Beyond the fault episode, each channel's overall record was still screened with the Interquartile Range (IQR) method for residual statistical outliers, consistent with established quality-control practice for crowdsourced, low-cost sensor data (Mazzoglio et al., 2026). Model performance was then assessed using correlation (r), Mean Absolute Error (MAE), Mean Bias Error (MBE), and the coefficient of determination (R^2) for temperature, humidity, wind speed, and atmospheric pressure.

Granularity-Value Method (H1)

To test whether a PWS's higher sampling rate is itself informative, the 30-second IoT record for each variable is grouped into two sets of non-overlapping windows. The Open-Meteo update grid described above (median width ≈ 15 minutes) and fixed 3-hour bins aligned to local WIB, matching BMKG's issuance schedule. Within each window, the sample variance of the 30-second readings is computed and pooled across windows, weighted by reading count, to obtain a within-window variance, expressed as a percentage of the variable's unconditional full-day variance. A high percentage means a forecast product sampled only at that cadence would structurally miss most of the variability a 30-second PWS can see. A low percentage means the coarser cadence already captures nearly all of it. This procedure is purely descriptive. It does not require a forecast value to compare against, only the IoT record itself.

IoT–BMKG Matching Method

For the new IoT–BMKG comparison, IoT readings are averaged within a ± 90 -minute window centered on each of BMKG's eight 3-hourly valid times, the same simple-mean resampling used for BMKG–Meteostat (World Meteorological Organization, 2017). Because the sensor fault removes valid temperature, pressure, and humidity data after 12:52 WIB, only windows wholly or partly before the fault contribute to those two variables: four windows (01:00, 04:00, 07:00, 10:00 WIB) are complete, the 13:00 WIB window is partial (82 of 180 minutes), and the 16:00, 19:00, and 22:00 WIB windows have no valid data, giving $n = 5$ rather than $n = 8$. Wind speed and rainfall, on the fault-unaffected circuit, retain all $n = 8$. Given $n = 5 - 8$, the confidence intervals and significance tests reported for this pair use the nominal n rather than an autocorrelation-adjusted n_{eff} . Lag-1 autocorrelation cannot be reliably estimated from so few points. Results should be read with the same caution as BMKG–Meteostat.

Statistical Framework

Root Mean Square Error (RMSE) is chosen to measure the typical magnitude of disagreement between a reference series (IoT or Meteostat) and a compared series (Open-Meteo or BMKG's forecast), in the variable's own unit ($^{\circ}\text{C}$, hPa , $\%$, m/s). It is computed as the square root of the mean of the squared paired differences. Because a sensor fault or a missed forecast event can produce large outliers, RMSE will always be $\geq$ Mean Absolute Error (MAE), and the *gap* between the two is itself informative. A wide RMSE-MAE gap signals that a handful of large errors, not consistent small ones, are driving the discrepancy. For paired records $i = 1, \dots, n$ between a reference series O_i (the observational baseline: IoT or Meteostat) and a compared series P_i (the forecast or model product: Open-Meteo or BMKG), with $d_i = P_i - O_i$. Equations (1) to (5) are as follows:

$$O = (1/n)\Sigma O_i, P = (1/n)\Sigma P_i \quad (1)$$

$$MAE = (1/n)\Sigma |d_i|, RMSE = \sqrt{[(1/n)\Sigma d_i^2]}, MBE = (1/n)\Sigma d_i = \bar{d} \quad (2)$$

$$r = \Sigma(O_i - O)(P_i - P) / \sqrt{[\Sigma(O_i - O)^2 \cdot \Sigma(P_i - P)^2]}, R^2 = 1 - \Sigma d_i^2 / \Sigma(O_i - O)^2 \quad (3)$$

$$nMAE\% = 100 \cdot MAE/O, CV(RMSE)\% = 100 \cdot RMSE/O \quad (4)$$

$$P_i = a + b \cdot O_i + \varepsilon_i, b = \Sigma(O_i - O)(P_i - P) / \Sigma(O_i - O)^2, a = P - b \cdot O \quad (5)$$

$b \neq 1$ indicates proportional (scale) bias; $a \neq 0$ indicates additive bias. Because high-frequency d_i series are typically autocorrelated, an effective sample size $n_{\text{eff}} \leq n$ is used for intervals and tests:

$$n_{\text{eff}} = n / [1 + 2\Sigma(1 - k/n)\rho_k], \text{ capped at } n \quad (6)$$

$$CI(MBE) = \bar{d} \pm t_h(n_{\text{eff}} - 1) \cdot (sd/\sqrt{n_{\text{eff}}}) \quad (7)$$

where sd is the SD of d_i . MAE/RMSE confidence intervals use a percentile bootstrap ($B = 5,000$ resamples); r uses the Fisher z -transform, $z = 0.5 \cdot \ln[(1+r)/(1-r)]$, $SEz = 1/\sqrt{(n_{\text{eff}} - 3)}$. Bland-Altman bias and 95% limits of agreement (LoA) are plotted against the reference value, per Cesana and Antonelli (2024):

$$bbias = \bar{d}, LoA = \bar{d} \pm 1.96 \cdot s_d, LoA \text{ 95\% CI} = \pm 1.96\sqrt{(3sd^2/n_{\text{eff}})} \quad (8)$$

Proportional bias is checked from the OLS slope b . Paired significance uses $t = \bar{d}/(sd/\sqrt{n_{\text{eff}}})$, $df = n_{\text{eff}} - 1$, with normality checked by Shapiro-Wilk (substituting Wilcoxon signed-rank where rejected) and effect size by Cohen's $d_z = \bar{d}/sd$. Error shape is summarized by sample skewness and excess kurtosis of d_i . Two caveats apply: at $n \leq 8$, Shapiro-Wilk, the bootstrap, and skewness/kurtosis lose most of their power and are reported for transparency rather than as trustworthy evidence, and roughly fifteen paired tests are run with no multiple-comparisons correction, so p -values should be read descriptively.

One further refinement is noted for a larger confirmatory study. The percentile bootstrap above resamples paired differences d_i as if exchangeable, the same independence assumption the n_{eff} correction elsewhere in this framework already relaxes. Since d_i is a sequential weather series with non-negligible lag-1 autocorrelation, a block bootstrap, resampling contiguous blocks of d_i rather than single points, would better preserve that dependence and avoid understating the true sampling variability of MAE and RMSE (Künsch, 1989). At $n = 5-8$ for the BMKG-involving pairs, a defensible block length would leave too few independent blocks to resample, so it is not applied here; it is recommended for the multi-day confirmatory study this pilot motivates.

RESULTS

Granularity Value: Does a Higher Sampling Rate Capture Real Information? (H1)

Per the Granularity-Value Method, for temperature, pressure, and humidity, the share of total-day variance occurring within a single Open-Meteo update window (≈ 15 minutes) is small ($0.7 - 1.4\%$) meaning Open-Meteo's own cadence already resolves nearly all the variability a 30-second PWS can see for these variables. At BMKG's coarser 3-hourly cadence, $17.4 - 37.9\%$ of variance in the same three variables occurs within a single window and is invisible to any product reporting only at that cadence. Wind speed is the clear outlier at both grids: 53.9% of its variance occurs within a single 15-minute window, and 67.1% within a single 3-hour window. Even Open-Meteo's finer cadence discards over half of the true wind-speed signal here, and a 3-hourly product discards just over two-thirds.

Table 1. Share of total-day variance in the IoT record occurring within a single Open-Meteo (OM, 15-min) or BMKG (3-hour) update window, the granularity-value test for H1

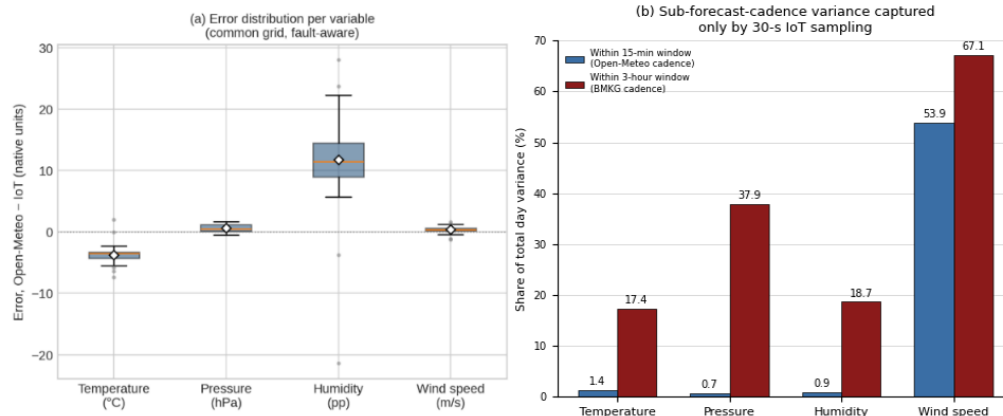
Variable	Total-day SD	Within 15-min SD (share)	Within 3-hour SD (share)	Interpretation
Temperature ($^{\circ}\text{C}$)	2.91	0.34 (1.4%)	1.21 (17.4%)	OM cadence sufficient; BMKG cadence loses info
Pressure (hPa)	0.68	0.06 (0.7%)	0.42 (37.9%)	OM cadence sufficient; BMKG cadence loses info
Humidity (%)	14.01	1.34 (0.9%)	6.05 (18.7%)	OM cadence sufficient; BMKG cadence loses info
Wind Speed (m/s)	0.34	0.25 (53.9%)	0.28 (67.1%)	Neither cadence sufficient

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Fig. 4 (a) Per-variable error distribution, IoT vs. Open-Meteo, common grid. (b) Share of total-day variance occurring within a single Open-Meteo (15-min) or BMKG (3-hour) update window, H1's granularity-value test.



This pattern is interpreted for H1 in the Discussion below. One caveat applies here: the elevated within-window share for pressure at the 3-hour grid (37.9%, versus 0.7% at 15 minutes) is at least partly attributable to un-averaged instrument noise in the uncalibrated BME280 rather than true atmospheric fluctuation.

IoT versus Open-Meteo

Of 2,873 raw IoT rows, temperature, pressure, and humidity are restricted to the pre-fault window (00:00–12:52 WIB, $n = 1,540$), and wind speed retains the full day after IQR screening ($n = 2,761$). Pairing every 30-second IoT record directly against Open-Meteo's repeated 15-minute value (the native-cadence, pseudo-replicated grid) reproduces the inflation problem the framework is designed to catch: nominal n of 1,540–2,761 collapses to an autocorrelation-adjusted n_{eff} of only 5.3–23.6 across the four variables. Reassuringly, resampling to the common grid built from Open-Meteo's own update timestamps ($n = 61$ for temperature/pressure/humidity, $n = 116$ for wind speed) yields an almost identical n_{eff} (5.3–23.4). The two independent corrections agree on how much real information the record contains, which cross-validates the n_{eff} adjustment itself. All inferential statistics below use the common grid.

Table 2 and Table 3 report the full diagnostic battery. Open-Meteo reads 3.74°C cooler than the IoT station ($MBE = -3.74^\circ\text{C}$, 95% CI $[-4.28, -3.20]$, $t(22.4) = -14.25$, $p < .001$, $dz = -2.94$), tied to the sensor's unshielded, enclosure-adjacent placement. Humidity mirrors this: 11.75 pp wetter ($MBE = +11.75$ pp, 95% CI $[8.87, 14.64]$, $p < .001$). Wind speed shows a smaller bias ($MBE = +0.40$ m/s) but the weakest correlation ($r = 0.555$, 95% CI $[0.188, 0.786]$) and largest relative error ($nMAE = 113.6\%$, $CV(RMSE) = 141.4\%$). Pressure showed the closest agreement ($nMAE = 0.07\%$) once the fault was excluded, but n_{eff} collapses to 5.3, so the paired t -test is not significant ($p = .086$) though Wilcoxon is ($p < .001$).

Table 2. IoT vs. Open-Meteo, fault-aware common grid, basic metrics, rounded to 2 decimal places.

* $nMAE$ unstable near a zero reference mean, read alongside absolute MAE

Variable	n	Mean IoT ($\bar{O}$)	Mean OM ($\bar{P}$)	r [95% CI]	MAE [95% CI]	nMAE%	MBE [95% CI]
Temperature (°C)	61	29.35	25.62	0.888 [0.753, 0.951]	3.80 [3.54, 4.07]	12.96	-3.74 [-4.28, -3.20]
Pressure (hPa)	61	1006.50	1007.14	0.778 [-0.244, 0.981]	0.74 [0.60, 0.88]	0.07	+0.64 [-0.14, 1.41]
Humidity (%)	61	79.90	91.66	0.879 [0.730, 0.949]	12.58 [11.42, 13.78]	15.74	+11.75 [8.87, 14.64]
Wind Speed (m/s)	116	0.41	0.81	0.555 [0.188, 0.786]	0.46 [0.40, 0.53]	113.6*	+0.40 [0.22, 0.58]

Table 3. IoT vs. Open-Meteo, fault-aware common grid, regression, Bland–Altman, significance, and distribution diagnostics, rounded to 2 decimal places. ** $p < .01$.

Variable	OLS (b, a) [95% CI]	BA bias [95% LoA] ($\pm$ LoA 95% CI)	Paired test	Cohen's dz	Skew/Kurt
Temperature (°C)	0.74 [0.64, 0.84], 3.90 [0.85, 6.95]	-3.74 [-6.23, -1.25] (± 0.89)	$t = -14.25$, $df = 22.4$, $p < .001$ **	-2.94	1.21 / 7.04
Pressure (hPa)	1.19 [0.81, 1.56], -186.26 [-565.43, 192.92]	+0.64 [-0.66, 1.93] (± 0.97)	$t = 2.22$, $df = 4.3$, $p = .086$	0.96	0.08 / -1.07
Humidity (%)	0.67 [0.57, 0.76], 38.47 [30.56, 46.39]	+11.75 [-1.15, 24.65] (± 4.71)	$t = 8.46$, $df = 21.5$, $p < .001$ **	1.79	-1.82 / 9.67
Wind Speed (m/s)	0.60 [0.43, 0.78], 0.56 [0.46, 0.66]	+0.40 [-0.42, 1.22] (± 0.29)	$t = 4.62$, $df = 22.2$, $p < .001$ **	0.96	-0.42 / 2.82

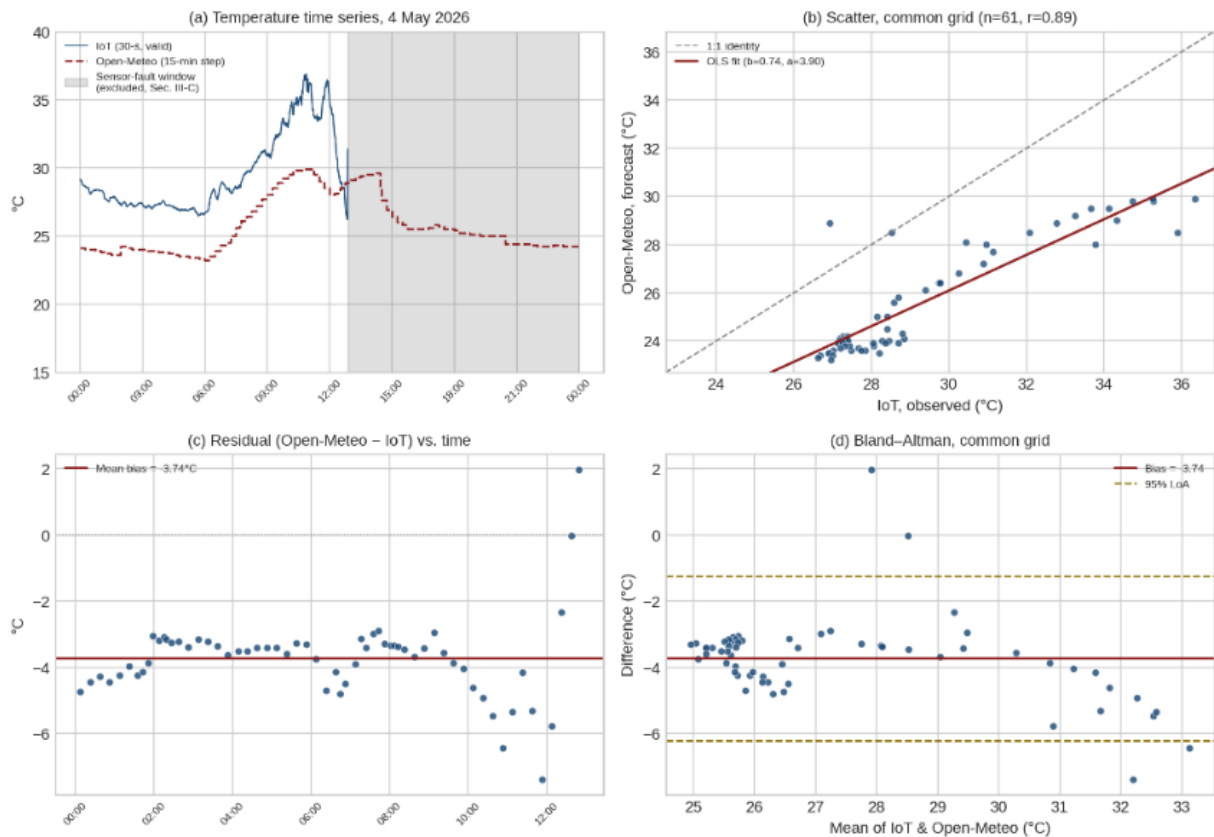
*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Fig. 5 IoT vs. Open-Meteo temperature diagnostics, 4 May 2026: (a) time series, fault window shaded; (b) scatter with 1:1 identity and OLS fit; (c) residual vs. time; (d) Bland–Altman plot.

IoT vs. Open-Meteo Temperature Diagnostics, 4 May 2026 (fault-aware, common grid)



IoT versus BMKG (H3)

The diagnostics below use $n = 5-8$; read tests descriptively, per the Statistical Framework. Table 4 reports the new IoT–BMKG comparison, including the corresponding regression, Bland–Altman, and distribution diagnostics. BMKG’s forecast runs 1.39°C cooler than the IoT station ($MBE = -1.39^{\circ}\text{C}$, $95\% \text{ CI } [-3.01, 0.23]$, $n = 5$, $t(4) = -2.38$, $p = .076$) and 2.76 pp wetter ($95\% \text{ CI } [-3.21, 8.73]$, $n = 5$, $p = .269$). Both are opposite in sign to the BMKG versus Meteostat results below ($+2.48^{\circ}\text{C}$ warmer, -8.25 pp drier). This reversal likely reflects the IoT sensor’s own heat bias (see H3 in Discussion) rather than a genuine change in BMKG’s forecast error. Wind speed is the one variable where BMKG’s bias is large and consistent with both references: BMKG forecasts 2.45 m/s faster than the IoT station ($95\% \text{ CI } [1.85, 3.04]$, $n = 8$, $t(7) = 9.76$, $p < .001$, $dz = 3.45$) and 0.65 m/s faster than Meteostat. No precipitation statistics are computed for this pair since the IoT gauge recorded 0.0 mm all day.

Table 4. IoT vs. BMKG, regression, Bland–Altman, significance, and distribution diagnostics.

Temperature/humidity $n = 5^a$; wind speed $n = 8$. $**p < .01$.

Variable	OLS (b, a) [95% CI]	BA bias [95% LoA] ($\pm$ LoA 95% CI)	Paired test	Cohen's dz	Skew/Kurt
Temperature ($^{\circ}\text{C}$)	1.03 [0.17, 1.90], -2.31 [-28.19, 23.57]	-1.39 [-3.95, 1.17] (± 1.98)	$t = -2.38$, $df = 4$, $p = .076$	-1.06	0.71 / -0.47
Humidity (%)	0.77 [0.27, 1.27], 20.76 [-18.32, 59.85]	+2.76 [-6.67, 12.19] (± 7.30)	$t = 1.28$, $df = 4$, $p = .269$	0.57	1.44 / 0.16
Wind speed (m/s)	3.70 [2.33, 5.07], 1.49 [0.92, 2.06]	+2.45 [1.06, 3.84] (± 0.85)	$t = 9.76$, $df = 7$, $p < .001^{**}$	3.45	0.19 / -1.53

BMKG versus Meteostat (H3)

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Meteostat provided a full 24-hour hourly record. BMKG provided eight 3-hourly valid times, yielding $n = 8$ matched instants. With n_{eff} collapsing to the nominal $n = 8$ since lag-1 autocorrelation cannot be reliably estimated from eight points. Temperature and humidity biases were large and significant, and they survived non-parametric checks where Shapiro–Wilk rejected normality. Wind-speed and precipitation biases were not distinguishable from zero at $n = 8$. Precipitation showed the widest relative disagreement ($nMAE = 77.1\%$), driven by a single missed event: Meteostat recorded 7.3 mm at 13:00 WIB against a same-instant BMKG forecast of 0.2 mm and a daily total of 3.0 mm against Meteostat's 22.5 mm, coincident with BMKG's largest temperature overshoot (32.8°C forecast vs. 27.5°C observed). The IoT station recorded 0.0 mm all day, illustrating how spatially localized tropical convective rainfall can be across the few kilometers separating the three sources (H4).

BMKG Forecast vs. Meteostat and IoT Observational References, 4 May 2026

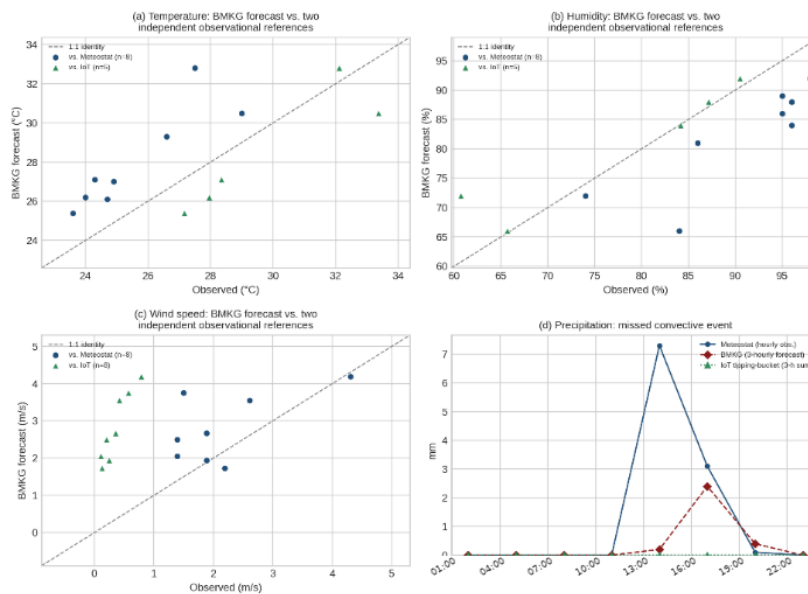


Fig. 6 BMKG forecast vs. two independent observational references, 4 May 2026: (a) temperature, (b) humidity, and (c) wind-speed scatter against Meteostat and the IoT station; (d) precipitation time series showing the missed 13:00 WIB convective event.

Table 5. BMKG vs. Meteostat, basic metrics, $n = 8$, rounded to 2 decimal places

Variable	n	Mean Met. ($\bar{O}$)	Mean BMKG ($\bar{P}$)	r [95% CI]	MAE [95% CI]	nMAE%	MBE [95% CI]
Temperature ($^{\circ}\text{C}$)	8	25.58	28.05	0.886 [0.482, 0.979]	2.48 [1.81, 3.36]	9.68	+2.48 [1.43, 3.52]
Humidity (%)	8	90.50	82.25	0.841 [0.334, 0.971]	8.25 [5.38, 11.75]	9.12	-8.25 [-12.36, -4.14]
Wind speed (m/s)	8	2.15	2.80	0.597 [-0.186, 0.916]	0.80 [0.39, 1.29]	37.22	+0.65 [-0.06, 1.36]
Precipitation (mm)	8	1.31	0.38	0.325 [-0.493, 0.838]	1.01 [0.04, 2.79]	77.14	-0.94 [-3.03, 1.16]

Table 6. BMKG vs. Meteostat, regression, Bland–Altman, significance, and distribution diagnostics, $n = 8$, rounded to 2 decimal places. $**p < .01$.

Variable	OLS (b, a) [95% CI]	BA bias [95% LoA] ($\pm$ LoA 95% CI)	Paired test	Cohen's dz	Skew/Kurt
Temperature ($^{\circ}\text{C}$)	1.19 [0.57, 1.81], -2.37 [-18.34, 13.59]	+2.48 [0.03, 4.92] (± 1.50)	$t=5.61$, $df=7$, $p<.001^{**}$	1.98	1.57 / 1.44
Humidity (%)	0.90 [0.32, 1.48], 0.95 [-51.51, 53.40]	-8.25 [-17.90, 1.40] (± 5.91)	$t=-4.74$, $df=7$, $p=.002^{**}$	-1.68	-0.87 / 0.04
Wind speed (m/s)	0.57 [-0.19, 1.33], 1.58 [-0.20, 3.36]	+0.65 [-1.02, 2.32] (± 1.02)	$t=2.17$, $df=7$, $p=.067$	0.77	0.52 / -0.31
Precipitation (mm)	0.10 [-0.19, 0.40], 0.24 [-0.59, 1.07]	-0.94 [-5.85, 3.97] (± 3.01)	$t=-1.06$, $df=7$, $p=.325$	-0.37	-2.21 / 3.00

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

DISCUSSIONS

Revisiting the Four Pitfalls

H1 (the granularity pitfall) receives mixed support. Against Open-Meteo's 15-minute cadence, extra granularity bought little for temperature, pressure, and humidity (0.7 – 1.4%) but much for wind speed (53.9%). Against BMKG's 3-hourly cadence it mattered for every variable, most for wind speed (67.1%), least for temperature (17.4%). This is a property of the PWS-product-variable pairing, not the PWS alone (Table 1).

H2 (the hardware/siting-bias pitfall) is supported: once corrected, Open-Meteo's bias against the IoT station remained large for temperature (-3.74°C), humidity ($+11.75\text{ pp}$), and wind speed ($+0.40\text{ m/s}$), and present but not significant for pressure. Wind speed, most sensitive to micro-siting on this obstructed rooftop, showed the least stable agreement; pressure showed the tightest, a predictable bias signature for an unshielded station co-located with electronics.

H3 (the single-reference pitfall) is only partly supported. Triangulating BMKG against two independent references found no single consistent bias direction: signs flip for temperature and humidity but stay consistently positive for wind speed. The simplest explanation is that the IoT station's own hardware bias, already evident from IoT vs. Open-Meteo, reverses BMKG's much smaller true bias.

H4 (the spatial-representativeness pitfall) receives suggestive but inconclusive support. BMKG's forecast missed the day's only substantial rainfall event (0.2 mm forecast vs. 7.3 mm observed at Meteostat), while the IoT gauge recorded 0.0 mm. This pilot cannot distinguish a missed regional event from one that never reached the IoT station's coordinates, a reminder that a single station is a spatial sample of one.

Fitness-for-Purpose, Reframed Around Three Preliminary Comparisons

Because all three comparisons above are single-day and small-sample ($n = 61 - 116$ for IoT–Open-Meteo; $n = 5 - 8$ for BMKG-involving pairs), none can adjudicate the undersampling-versus-adequacy dispute alone. BMKG's 3-hourly forecast tracked trends well against Meteostat (r up to 0.886) but missed a short-lived convective event. The Granularity Value test adds a variable-specific answer: a 15-minute product is adequate for temperature, pressure, and humidity, but not for wind speed.

Derived Algorithmic Mitigation

Based on the empirical fault signatures identified, the fault-aware quality-control (FA-QC) procedure is formalized as an explicit, replicable algorithm (Algorithm 1) screening raw telemetry through six sequential fault checks before escalating to automatic stream isolation.

Algorithm 1 Fault-Aware Quality Control (FA-QC)Input: 30-s sample $x_t = (T, P, RH, WS)$ from BME280 + anemometer

Output: valid | flagged(type), with automatic stream isolation

```
1: if I2C_status( $x_t$ )  $\in$  {NULL, 0xFF, NaN} then
2:   flag  $\leftarrow$  BUS_FAULT
3: else if  $T \notin [10, 50]^{\circ}\text{C}$  or  $P \notin [950, 1050]\text{ hPa}$  or  $RH \notin [20, 100]\%$  or  $WS \notin [0, 50]\text{ m/s}$  then
4:   flag  $\leftarrow$  RANGE_FAULT
5: else if  $|T_t - T_{t-1}| > 2.5^{\circ}\text{C}$  or  $|P_t - P_{t-1}| > 3.0\text{ hPa}$  then
6:   flag  $\leftarrow$  DISCONTINUITY_FAULT
7: else if  $|WS_t - WS_{t-1}| / \Delta t > 0.5\text{ m/s}^2$  then
8:   flag  $\leftarrow$  RATE_FAULT
9:  $W \leftarrow$  last 10 samples (5-min sliding window)
10: if  $|\text{unique}(W_T)| \leq 2$  or  $\text{Var}(W_P) = 0$  (static/alternating) then
11:   flag  $\leftarrow$  FROZEN_FAULT
12: else if  $\text{Var}(W_{RH}) = 0$  and  $\text{Var}(W_T) > 0.5$  then
13:   flag  $\leftarrow$  DESYNC_FAULT  $\triangleright$  cross-variable check
14: fault_counter  $\leftarrow$  fault_counter + 1 if flag else 0
15: if fault_counter  $\geq 10$  ( $\equiv 300\text{ s}$  persistent) then
16:   isolate_stream(BME280)  $\triangleright$  fault window (see Temporal Alignment & FA-QC)
17: return valid if flag =  $\emptyset$ , else flagged(flag)
```

Threshold values reflect the site's physical context rather than dataset tuning. The range filter (10 – 50°C, 950 – 1050 hPa, 20 – 100% RH, 0 – 50 m/s) brackets plausible tropical conditions while catching the fault's corrupted values (such as 182.23°C, negative hPa). The discontinuity thresholds (2.5°C, 3.0 hPa per 30 seconds) sit above the BME280's noise floor but below any plausible step for slowly varying fields. The wind-speed rate cap (0.5 m/s²) reflects anemometer inertia rather than genuine gusts, and the 5-minute variance window with 300-second isolation delay balances fault detection against false positives. The frozen-fault check (≤ 2 distinct values within the window) catches sensors stuck alternating between fixed values, as observed here (22.52°C and 182.23°C), which could otherwise pass a pure variance threshold despite being non-physical.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Limitations

This entire study remains a preliminary case study. Coverage is a single day, partial for IoT temperature, pressure, and humidity, so statistics should not be assumed stable across days or seasons. Spatial coverage is a single site and cannot be generalized elsewhere in Indonesia. The IoT sensors were uncalibrated, unshielded, and sited below recommended heights, so IoT-involving biases reflect combined instrument and model error the design cannot separate. The results demonstrate the framework's mechanics, not general accuracy for any source.

CONCLUSION

This paper set out to document, as a preliminary case study, the practical pitfalls encountered when a single affordable Personal Weather Station is designed, deployed, and evaluated at one tropical site. The first pitfall, whether high-frequency sampling is worth its cost, proved variable-dependent, mattering little for temperature, pressure, and humidity against a 15-minute product but much against a 3-hourly one, and for wind speed against both. The second pitfall, uncorrected hardware and siting bias, was large throughout, attributable to the station's own unshielded, obstructed-rooftop construction. The third, relying on only one reference instrument, was exposed directly: extending the framework to three sources showed BMKG's apparent bias against Meteostat reversing sign against the IoT station, best explained by the IoT sensor's own heat bias. The fourth, treating a single station as spatially representative, was raised but not resolved. This pilot offers a tested, replicable protocol for the multi-day, multi-site confirmatory test H1–H4 would require.

REFERENCES

- Afyouni, S., Smith, S. M., & Nichols, T. E. (2019). Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation. *NeuroImage*, 199, 609–625. <https://doi.org/10.1016/j.neuroimage.2019.05.011>
- Al Kalaany, C. M., Kimaita, H. N., Abdelmoneim, A. A., Khadra, R., Derardja, B., & Dragonetti, G. (2025). The potential of low-cost IoT-enabled agrometeorological stations: A systematic review. *Sensors*, 25(19), Article 6020. <https://doi.org/10.3390/s25196020>
- Beele, E., Reyniers, M., Aerts, R., & Somers, B. (2022). Quality control and correction method for air temperature data from a citizen science weather station network in Leuven, Belgium. *Earth System Science Data*, 14(10), 4681–4717. <https://doi.org/10.5194/essd-14-4681-2022>
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476), 307–310. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)
- Brousse, O., Simpson, C. H., Poorthuis, A., & Heaviside, C. (2024). Unequal distributions of crowdsourced weather data in England and Wales. *Nature Communications*, 15, Article 4828. <https://doi.org/10.1038/s41467-024-49276-z>
- Cesana, B. M., & Antonelli, P. (2024). Bland and Altman agreement method: To plot differences against means or differences against standard? An endless tale? *Clinical Chemistry and Laboratory Medicine*, 62(2), 262–269. <https://doi.org/10.1515/ccbm-2023-0306>
- Chen, J., Saunders, K., & Whan, K. (2021). Quality control and bias adjustment of crowdsourced wind speed observations. *Quarterly Journal of the Royal Meteorological Society*, 147(740), 3647–3664. <https://doi.org/10.1002/qj.4146>
- Dadras Javan, F., Pineda Solis, J. L., Perotti, S., Rinaldi, F., & Najafi, B. (2025). Implementation of predictive modelling-empowered HVAC-driven PV self-consumption enhancement strategies in a warehouse. *Journal of Physics: Conference Series*, 3140, 052013. <https://doi.org/10.1088/1742-6596/3140/5/052013>
- DeFord, L., Stoll, R., & Pardyjak, E. R. (2025). Development of a low-cost, 3D-printed, aspirated air temperature measurement radiation shield. *Journal of Atmospheric and Oceanic Technology*, 42(2), 155–170. <https://doi.org/10.1175/JTECH-D-24-0006.1>
- Garcia-Marti, I., Overeem, A., Noteboom, J. W., de Vos, L., de Haij, M., & Whan, K. (2026). Incorporating crowdsourced data into operational products: A perspective from a national meteorological service. *International Journal of Climatology*, Article e70408. <https://doi.org/10.1002/joc.70408>
- Herlambang, D. (2023, October 18). *Prakiraan sering meleset, ternyata data aplikasi cuaca di HP bukan berasal dari BMKG* [Press release]. Badan Meteorologi, Klimatologi, dan Geofisika. [BMKG](https://www.bmkg.go.id/)
- Idris, M. I. M., Rahman, B., & Said, M. S. (2025). Android-based weather data monitoring information system to support agriculture in South Konawe Regency. *Jurnal Sistem Informasi dan Teknik Komputer*, 10(2), 45–52.

- Jankauskas, M., Serackis, A., Paulauskas, N., Pomarnacki, R., & Hyunh, V. K. (2024). The impact of the weather forecast model on improving AI-based power generation predictions through BiLSTM networks. *Electronics*, 13(17), Article 3472. <https://doi.org/10.3390/electronics13173472>
- Katsumata, M., Mori, S., Hamada, J.-I., Hattori, M., Syamsudin, F., & Yamanaka, M. D. (2018). Diurnal cycle over a coastal area of the Maritime Continent as derived by special networked soundings over Jakarta during HARIMAU2010. *Progress in Earth and Planetary Science*, 5, Article 32. <https://doi.org/10.1186/s40645-018-0216-3>
- Kim, B., Hwang, S., Lee, Y., Shin, S., & Kim, K. (2022). Comparative analysis of environmental standards to install a rooftop temperature monitoring station. *Scientific Reports*, 12, Article 22468. <https://doi.org/10.1038/s41598-022-27070-5>
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3), 1217–1241. <https://doi.org/10.1214/aos/1176347265>
- Mazzoglio, P., Moccia, B., & Ghaffaripour, E. (2026). Personal weather stations in hydrology and meteorology: A review of applications, challenges and future directions. *Hydrological Sciences Journal*, 71(10), 2111–2127. <https://doi.org/10.1080/02626667.2026.2665735>
- Mustafa, J. M., Matthews, A. J., Hall, R. A., Heywood, K. J., & Azaneu, M. V. C. (2024). Characterisation of the observed diurnal cycle of precipitation over the Maritime Continent. *Quarterly Journal of the Royal Meteorological Society*, 150(762), 2602–2624. <https://doi.org/10.1002/qj.4725>
- Nan, F., Zeng, C., Shen, H., & Lin, L. (2025). Calibration of integrated low-cost environmental sensors for urban air temperature based on machine learning. *Sensors*, 25(11), Article 3398. <https://doi.org/10.3390/s25113398>
- Pavithran, M. S., Sreeram, B., Pillai, A. V., & Jothi, R. (2025). Multi-scale weather forecasting using deep learning architectures with Chennai climate data. *IEEE Access*, 13, 207303–207319. <https://doi.org/10.1109/ACCESS.2025.3640667>
- Ramadhan, R., Marzuki, M., Suryanto, W., Sholihun, S., Yusnaini, H., & Muharsyah, R. (2024). Validating IMERG data for diurnal rainfall analysis across the Indonesian maritime continent using gauge observations. *Remote Sensing Applications: Society and Environment*, 34, Article 101186. <https://doi.org/10.1016/j.rsase.2024.101186>
- Romero Rodríguez, L., Guerrero Delgado, M. C., Castro Medina, D., Sánchez Ramos, J., & Álvarez Domínguez, S. (2024). Forecasting urban temperatures through crowdsourced data from citizen weather stations. *Urban Climate*, 56, Article 102021. <https://doi.org/10.1016/j.uclim.2024.102021>
- Smith, E., Grant, N., Luo, X., & Diaz, D. (2025). Evaluating the ability of gridded climate datasets to capture temperature and precipitation trends and extremes. *Scientific Reports*, 15, Article 12607. <https://doi.org/10.1038/s41598-025-97570-7>
- Wang, Y.-H., Chen, W.-T., & Wu, C.-M. (2024). Novel perspectives on multiple-peak diurnal convection over a tropical mountainous island from idealized large-eddy simulations. *npj Climate and Atmospheric Science*, 7, Article 296. <https://doi.org/10.1038/s41612-024-00884-y>
- World Meteorological Organization. (2017). *WMO guidelines on the calculation of climate normals* (WMO-No. 1203). World Meteorological Organization.
- Zippenfenig, P. (2023, April 12). *Open-Meteo.com weather API* [Data set]. Open-Meteo Substack. [Open-Meteo](https://open-meteo.com)