

Comparative Evaluation of Moving Average Filtering and Linear Regression Calibration in an ESP32-Based Water pH Sensor

Mukhammad Indra Ukasya¹, Farid Baskoro², Rifqi Firmansyah³, Lilik Anifah⁴, and Lusia Rakhmawati⁵

Universitas Negeri Surabaya, Surabaya, Indonesia¹⁻⁵

E-mail Corresponding: mukhammad.23017@mhs.unesa.ac.id

Received: March 12, 2026

Revised: August 16, 2026

Accepted: September 20, 2026

Abstract

This study compared Moving Average filtering and Linear Regression calibration in an ESP32-based low-cost pH measurement system using nominal buffers of pH 4.01, 6.86, and 9.18. One ESP32-PH-4502C sensor unit was evaluated through 60 repeated measurement cycles per buffer under approximately 25 °C room conditions. Performance was assessed using MAE, MSE, RMSE, bias, standard deviation, 95% confidence interval, drift slope, and lag-1 autocorrelation. Moving Average filtering did not reduce measurement error, and MAE generally increased with larger window sizes. Linear Regression reduced MAE by 25.2% at pH 4.01 but increased MAE by 158.9% at pH 6.86 and 42.4% at pH 9.18. The recorded Linear Regression condition also showed lower dispersion, drift slopes closer to zero, and lower lag-1 autocorrelation. However, because the processing conditions were not evaluated as paired transformations of the same underlying voltage sequence, these stability differences cannot be attributed solely to the regression transformation. The findings show that smoothing did not improve accuracy, while regression performance varied across buffer values. The results are device-specific and do not establish field performance or generalization across sensor units.

Keywords: ESP32; pH sensor; Moving Average Filter; Linear Regression; sensor calibration.

INTRODUCTION

The main challenge in developing a low-cost ESP32-based pH measurement system is not merely acquiring and displaying sensor data, but ensuring that the resulting values remain reliable. A glass pH electrode produces an analog potential whose relationship with pH depends on calibration and measurement temperature (Wang, 2023). At the same time, the built-in analog-to-digital converter of the ESP32 may exhibit noise, nonlinearity, imprecise reference voltage, and dead-zone effects that can influence precision analog measurements (Espinosa-Gavira et al., 2024). These limitations create an important distinction between signal smoothing, which reduces short-term fluctuation, and calibration, which adjusts the relationship between sensor output and reference pH values.

The development of low-cost electronics, wireless communication, and Internet of Things (IoT) technology has accelerated the adoption of smart water quality monitoring systems. Wireless sensor networks and IoT-based platforms allow water quality data to be collected, processed, and transmitted in near real time with lower implementation cost and broader spatial coverage than conventional laboratory-centered monitoring (Jan et al., 2021; Singh & Walingo, 2024). Similarly, Lakshmikantha et al. (2021) demonstrated that IoT-based water quality monitoring systems can integrate multiple sensor parameters to support continuous observation and early detection of water condition changes. This makes low-cost sensor systems a practical alternative for distributed monitoring, particularly in environments with limited instrumentation resources (de Camargo et al., 2023).

Although low-cost sensor systems may be introduced in science laboratories, the present

study was designed primarily as a technical validation study rather than an investigation of teaching or learning outcomes. Laboratory inquiry can involve the use of instruments to collect, analyze, and interpret experimental evidence, while sensor-based activities may make quantitative investigations more accessible (de Jong et al., 2013; Hofstein & Lunetta, 2004; Wijayanti et al., 2022). Therefore, the educational relevance of the developed device is limited to its potential use for illustrating measurement error, calibration, signal processing, and data reliability. Classroom usability, student responses, and learning outcomes were not evaluated in this study.

Despite these advantages, the main challenge of low-cost sensing is not only data acquisition but also data reliability. Previous studies have emphasized that accuracy, precision, and long-term reliability remain critical concerns when low-cost environmental sensors are used without strict calibration and validation procedures (Nalakurthi et al., 2024; Slongo et al., 2024). Low-cost sensors are vulnerable to systematic bias, environmental disturbance, and inconsistency between deployments; therefore, calibration is a prerequisite before the measurement output to be dependable for practical monitoring or learning-oriented laboratory use.

Among water quality parameters, pH is one of the most important indicators because it directly represents the acid-base condition of water and influences chemical equilibrium, biological processes, and the suitability of water for environmental or industrial use (Singh & Walingo, 2024). For this reason, pH sensors are commonly integrated into Internet of Things (IoT) based water monitoring architectures together with other parameters such as temperature, turbidity, conductivity, and dissolved oxygen (de Camargo et al., 2023; Slongo et al., 2024). Previous studies in Indonesia have also applied pH and temperature parameters in water quality control systems for aquaculture environments, indicating that pH remains a practical and relevant variable in low-cost water monitoring applications (Ardyansyah et al., 2022). However, although low-cost pH sensors are widely available, their practical performance depends on correct calibration, stable signal acquisition, and suitable processing of the measured output.

In low-cost embedded implementation, ESP32 is frequently selected as the main controller because it is affordable, compact, equipped with integrated Wi-Fi, and compatible with IoT applications (Olanubi et al., 2024). Nevertheless, the analog acquisition stage of the ESP32 has its own limitations. Espinosa-Gavira et al. (2024) showed that the built-in ADC of the ESP32 may display nonideal behavior for precision analog measurements due to noise, nonlinearity, imprecise reference voltage, and dead-zone effects. These characteristics are relevant for pH sensing because pH sensor output is sensitive to small voltage changes and therefore requires careful signal processing and calibration.

Previous studies have addressed low-cost sensor reliability primarily through calibration or system-level evaluation. Regression-based calibration has been shown to improve the accuracy of low-cost water-quality sensors when the calibration model is validated against professional reference instruments (Mohd Jais et al., 2024). In contrast, Moving Average filtering is conventionally used to suppress short-term signal fluctuations rather than to align sensor output with a reference measurement scale (Oppenheim & Schaffer, 2010). However, systematic reviews indicate that many low-cost water-quality studies emphasize system development, communication, and real-time display rather than detailed sensor-performance evaluation. Only a limited proportion of the reviewed studies compared low-cost sensors with reference probes, and gaps remain in the reporting of calibration procedures and statistical evaluation (de Camargo et al., 2023; Nalakurthi et al., 2024).

Accordingly, the contribution of the present study is not the introduction of a new filtering or calibration algorithm, but a descriptive, device-specific comparison of the recorded performance characteristics of the baseline output, three Moving Average window sizes, and Linear Regression output within an ESP32-based pH measurement system. The processing conditions were evaluated using accuracy, bias, dispersion, drift, and sequential-dependence

indicators. This comparison was intended to determine whether a visually smoother signal was associated with lower measurement error and whether the recorded Linear Regression condition showed consistent accuracy across the three tested buffer values.

Based on this research gap, the present study compared the recorded performance characteristics of Moving Average filtering and Linear Regression calibration in an ESP32-based pH measurement system. Measurements were conducted using nominal buffer solutions of pH 4.01, 6.86, and 9.18 under laboratory room-temperature conditions estimated at approximately 25 °C. The baseline output, Moving Average windows of 5, 10, and 20 samples, and Linear Regression output were evaluated using MAE, MSE, RMSE, bias, standard deviation, descriptive 95% confidence interval, drift slope, and lag-1 autocorrelation. The study aimed to compare the recorded accuracy and sequential characteristics of the processing conditions without assuming that the observed differences were caused solely by the processing algorithms or that one method would be universally superior.

METHODS

Research design

This study used a quantitative experimental design to compare the recorded baseline output, three Moving Average conditions, and Linear Regression output in an ESP32-based pH measurement system. The study focused on technical sensor-performance evaluation; therefore, the units of analysis were the recorded sensor outputs, nominal buffer values, and the resulting performance metrics.

The experiment compared five recorded processing conditions: baseline output, Moving Average windows of 5, 10, and 20 samples, and Linear Regression output. The same sensor unit, nominal buffer values, and general measurement procedure were used across the tested conditions. However, the condition-specific outputs were not retained as paired transformations of an identical underlying voltage sequence. Therefore, comparisons among the processing conditions were interpreted descriptively and were not used to isolate the causal effect of an algorithm from possible session-level variation.

Hardware configuration and measurement system

The measurement system consisted of an ESP32 development board, a PH-4502C analog pH sensor interface module, a glass pH electrode probe, and a 16 x 2 I2C liquid-crystal display (LCD) for local readout. The glass electrode responds to hydrogen-ion activity and produces an electrical potential that is conditioned by the analog interface module before being acquired by the microcontroller (Wang, 2023). Based on the firmware configuration used in this study, the analog output of the PH-4502C module was connected to GPIO 34 of the ESP32. The analog-to-digital converter (ADC) was configured at 12-bit resolution with 11 dB attenuation, using a reference voltage of 3.3 V and a maximum digital count of 4095. For voltage conversion in the firmware, the ADC count was mapped to an assumed range of 0–3.3 V. The manufacturer-specified output range of the PH-4502C module and glass electrode probe was not recorded in the available experimental documentation. Each sensor output was obtained by averaging 100 consecutive ADC readings collected at intervals of 5 ms to reduce short-term fluctuations. The separate manufacturer and model specification of the glass electrode probe were not recorded in the available experimental documentation.

The unfiltered voltage was converted into the baseline pH output using $pH_{baseline} = -5.276V + 21.25$. In this study, the term baseline output refers to the pre-existing unfiltered pH conversion implemented in the original firmware. The coefficients -5.276 and 21.25 were already present in the original firmware. The available experimental documentation did not record the calibration procedure from which these coefficients were obtained. Therefore, the baseline equation was treated as the pre-existing firmware conversion rather than as an independently validated baseline measurement.

For the Moving Average Filter condition, the filter was applied to the voltage data, and the

filtered voltage was converted using $pHMA = -5.276VMA + 21.25$. The Linear Regression output was calculated from the unfiltered voltage using $pHLR = -5.5528V + 22.2189$. These equations were implemented directly in the ESP32 firmware so that the baseline, filtered, and regression-calibrated outputs could be generated during data acquisition. A concise fragment of the firmware implementation is presented in Listing 1.

```
#define ADC_RES 4095.0
#define VREF 3.3

float slope_basic = -5.276;
float offset_basic = 21.25;
float slope_reg = -5.5528;
float offset_reg = 22.2189;

analogReadResolution(12);
analogSetAttenuation(ADC_11db);

long adc = readADC(100);
float voltage = adc * VREF / ADC_RES;
float voltageMA = sumMA / MA_WINDOW;

float pH_raw = slope_basic * voltage + offset_basic;
float pH_ma = slope_basic * voltageMA + offset_basic;
float pH_lr = slope_reg * voltage + offset_reg;
```

Listing 1. Core firmware configuration and equations for voltage and pH conversion.
Source: research firmware, 2026.

The firmware configuration described above was integrated into the hardware arrangement shown in Figure 1. The figure identifies the main components involved in signal acquisition, processing, and local output display, including the pH electrode probe, PH-4502C analog interface module, ESP32 development board, and LCD display.

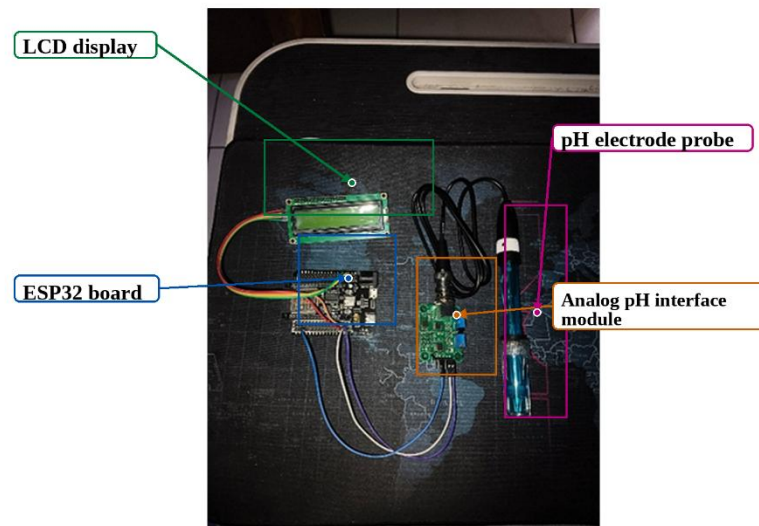


Figure 1. Hardware configuration of the ESP32-based pH sensor used in this study.

Figure 1 presents the physical arrangement of the ESP32-based pH measurement system. The pH electrode probe was connected to the PH-4502C analog interface module, which conditioned the electrode signal before it was acquired through GPIO 34 of the ESP32. The ESP32 processed the signal using the firmware equations described in Listing 1, while the LCD provided a local display of the resulting pH output.

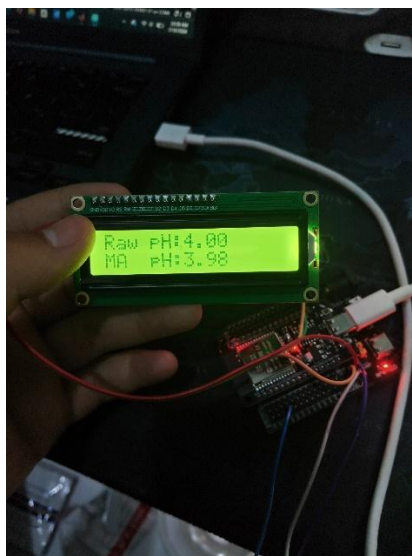


Figure 2. Operational testing of the developed pH sensor system during data acquisition

Figure 2 shows the measurement system during operation, where the baseline pH and moving average output were displayed locally while the ESP32 executed the embedded measurement program.

Reference solutions and data acquisition

The experiment was conducted using three standard buffer solutions with nominal values of pH 4.01, 6.86, and 9.18 under laboratory room-temperature conditions, estimated at approximately 25 °C. These nominal buffer values were used as references for evaluating the sensor output. The use of standard buffer points is consistent with common calibration and evaluation procedures for low-cost pH sensors and water-quality monitoring devices (Mohd Jais et al., 2024). A laboratory-grade pH meter was not used as an independent reference instrument; therefore, the evaluation represents the closeness of the sensor readings to the nominal buffer values rather than validation against a reference-grade measurement device.

For each buffer solution, 60 repeated measurement cycles were conducted using the same sensor unit. In each cycle, the pH probe was immersed in the buffer solution for approximately 60 seconds before one reading was retained. No numerical stability threshold was applied; the fixed stabilization period and visual observation of the displayed reading were used before recording the measurement. The probe was then removed, rinsed with distilled water, gently dried using lint-free tissue without rubbing the electrode surface, and allowed to rest for approximately 60 seconds before the next measurement cycle. This procedure was repeated separately for the pH 4.01, 6.86, and 9.18 buffer solutions, resulting in 180 observations for each tested processing condition. Therefore, the 60 observations for each buffer were obtained from separate immersion cycles rather than from one continuous immersion. However, because all measurements were obtained using one sensor unit under the same laboratory setting, they were treated as repeated measurements rather than independent replications across different sensor units.

The buffer solutions were tested in the fixed order of pH 4.01, followed by pH 6.86 and pH 9.18. The firmware generated one output approximately every 1.5 seconds. However, only one reading obtained after the approximately 60-second stabilization period was retained for each measurement cycle. Temperature was not actively controlled or continuously recorded using a calibrated thermometer. Therefore, the reported temperature represents an approximate laboratory room condition rather than a precisely controlled experimental temperature.

For each measurement cycle, the recorded output was compared with the nominal value of the corresponding buffer solution. The repeated measurements were used to characterize measurement variation and compare the performance of the tested processing conditions. The evaluation included mean absolute error (MAE), mean squared error (MSE), root mean squared

error (RMSE), bias, mean, standard deviation, 95% confidence interval, drift slope, and lag-1 autocorrelation.

Moving Average processing

A trailing moving average filter was implemented directly in the ESP32 firmware to reduce short-term fluctuations in the sensor voltage. The filter was applied to the voltage data before conversion into pH values. Three window sizes were evaluated separately, namely 5, 10, and 20 samples. For each experimental session, only one window size was activated by changing the MA_WINDOW parameter in the firmware, while the sensor unit, hardware configuration, and measurement procedure were maintained under the same conditions. In discrete-time signal processing, a moving average filter calculates the arithmetic mean of adjacent observations to suppress rapid signal variations (Oppenheim & Schafer, 2010; Ranjan et al., 2024).

The moving-average voltage at sample t was calculated using Equation (1):

$$V_{MA,t} = \frac{1}{N} \sum_{i=0}^{N-1} V_{t-i} \quad (1)$$

where $V_{MA,t}$ is the filtered voltage at sample t , V_{t-i} is the current or preceding sensor-voltage observation, and N is the selected window size of 5, 10, or 20 samples. Because the equation uses the current observation and preceding observations, the implemented method was a trailing moving average rather than a centered moving average.

After the filtered voltage was obtained, it was converted into the moving-average pH output using Equation (2):

$$pH_{MA} = -5.276V_{MA} + 21.25 \quad (2)$$

During data preparation, the first $N-1$ outputs generated while the moving-average buffer was being filled were excluded from the exported dataset and statistical analysis to avoid the start-up effect associated with incomplete window initialization. Accordingly, MA5 analysis began after five voltage observations were available, MA10 analysis began after ten observations, and MA20 analysis began after twenty observations. The statistical and graphical analyses used the numerical values exported by the ESP32 firmware without additional rounding; rounding was applied only when the final results were presented in tables.

A concise fragment of the firmware implementation of the trailing moving average filter is presented in Listing 2.

```
sumMA -= bufferMA[indexMA];
bufferMA[indexMA] = voltage;
sumMA += bufferMA[indexMA];

indexMA++;
if (indexMA >= MA_WINDOW) {
    indexMA = 0;
}

float voltageMA = sumMA / MA_WINDOW;
float pH_ma = slope_basic * voltageMA + offset_basic;
```

Listing 2. Firmware implementation of the trailing moving average filter

Linear Regression correction

Linear Regression was applied as a calibration-based method to convert the sensor voltage into the corresponding pH value. This method was selected because it provides a simple mathematical relationship between measured sensor output and known reference values and can be implemented directly in the ESP32 firmware for real-time correction (Mohd Jais et al., 2024; Nalakurthi et al., 2024). A linear calibration model is appropriate when the calibration range is represented by adequate reference points and the relationship between sensor voltage and pH is approximately linear (Vogel et al., 2025).

Sensor calibration was conducted on 18 July 2026 using standard buffer solutions of pH 4.01, 6.86, and 9.18. The calibration dataset consisted of 60 voltage observations, with 20

observations recorded for each nominal buffer value. Voltage values recorded in millivolts were converted into volts before analysis. The measured voltage was used as the predictor variable, while the nominal pH value was used as the response variable. The calibration coefficients were estimated using the least-squares method based on the linear model shown in Equation (3):

$$y = aV + b \quad (3)$$

where V is the measured sensor voltage in volts, y is the predicted pH value, a is the slope coefficient, and b is the intercept. Table 1 summarizes the calibration data and the prediction obtained at the mean voltage of each buffer group.

Table 1. Summary of Linear Regression calibration fitting results

Nominal pH	n	Mean calibration voltage (V)	Predicted pH	Residual*
4.01	20	3.27945	4.00894	+0.00106
6.86	20	2.76555	6.86250	-0.00250
9.18	20	2.34845	9.17856	+0.00144

*Residual = nominal pH - predicted pH.

Source: processed calibration data, 2026.

Least-squares fitting using all 60 calibration observations produced a slope of -5.5527618 and an intercept of 22.2189428 . The coefficient of determination was $R^2 = 0.999976$, indicating a strong linear relationship within the tested calibration dataset. The predicted pH values at the mean calibration voltages were 4.00894, 6.86250, and 9.17856 for the nominal pH values of 4.01, 6.86, and 9.18, respectively. The corresponding residuals were small, ranging from -0.00250 to $+0.00144$ pH units. For implementation in the firmware, the regression coefficients were rounded to four decimal places, resulting in Equation (4):

$$pH = -5.5528V + 22.2189 \quad (4)$$

The slope and intercept obtained from the calibration dataset were implemented in the ESP32 firmware on 18 July 2026. Validation was conducted separately on 19 July 2026 using the updated firmware, and the validation observations were not included in the regression fitting process. For each buffer solution, 60 repeated measurement cycles were performed, resulting in 180 validation observations for each tested processing condition. Because the calibration and validation procedures used one sensor unit under the same laboratory setting, the resulting model is interpreted as a device-specific calibration within the tested buffer range rather than as evidence of broad performance across different sensor units or field conditions.

Performance metrics and data analysis

The performance of each processing condition was evaluated using mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), bias, mean, standard deviation (SD), 95% confidence interval (CI), drift slope, and lag-1 autocorrelation. Multiple indicators were used because measurement performance involves both closeness to the nominal reference value and variation among repeated measurements. The error metrics were calculated using Equations (5)-(7):

$$MAE = \frac{\sum_{i=1}^n |\hat{p}_i - p_{ref,i}|}{n} \quad (5)$$

$$MSE = \frac{\sum_{i=1}^n (\hat{p}_i - p_{ref,i})^2}{n} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{p}_i - p_{ref,i})^2}{n}} \quad (7)$$

where $\hat{p}_i$ is the pH value produced by the tested processing condition, $p_{ref,i}$ is the nominal value of the corresponding buffer solution, and n is the number of retained observations. MAE represents the average absolute deviation from the nominal buffer value, MSE gives greater weight to larger deviations, and RMSE expresses the measurement error in the same pH unit as the sensor output.

Bias was calculated as the mean signed difference between the processed sensor output and the nominal buffer value. A positive bias indicates that the sensor output tended to exceed the nominal reference, whereas a negative bias indicates underestimation. The mean and SD were used to describe the central value and dispersion of the measurements. The 95% CI was calculated as a descriptive interval for the mean output of each processing condition. Because the recorded measurements showed serial dependence, the interval was not interpreted as a formal measurement-uncertainty estimate and was not adjusted for autocorrelation.

Drift slope was estimated by fitting a simple linear trend between the recorded pH output and the measurement-cycle index. A slope close to zero indicates relatively stable output across the measurement sequence, whereas a positive or negative slope indicates an increasing or decreasing trend. Lag-1 autocorrelation was used to evaluate the relationship between each observation and the immediately preceding observation. Because all repeated measurements were obtained using one sensor unit, the CI and autocorrelation results were interpreted descriptively and were not used to claim general performance across different sensor units.

The analysis compared the baseline output, MA5, MA10, MA20, and Linear Regression conditions separately for each buffer solution. Lower MAE, MSE, and RMSE values indicated closer agreement with the nominal buffer value, while bias was used to identify the direction of systematic deviation. SD was interpreted as a descriptive measure of output dispersion, while the 95% CI was treated only as a descriptive interval for the mean output and was not adjusted for serial dependence. Drift slope and lag-1 autocorrelation were used to characterize sequential behavior rather than to establish statistical independence. The results were therefore interpreted as a comparative, device-specific performance evaluation under the experimental conditions used in this study.

RESULT AND DISCUSSION

Comparison of evaluation metrics across methods

The performance of the baseline, MA5, MA10, MA20, and Linear Regression outputs was evaluated separately for the nominal buffer values of pH 4.01, 6.86, and 9.18. The evaluation included mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), bias, mean, standard deviation (SD), 95% confidence interval (CI), drift slope, and lag-1 autocorrelation. Tables 2–7 present the results obtained from 60 retained measurement cycles for each buffer and processing condition. The results were interpreted descriptively because all measurements were obtained using one sensor unit under the same laboratory setting.

Overall, the recorded performance of Linear Regression differed across the tested buffer conditions. Linear Regression produced the lowest error for the pH 4.01 buffer, whereas the baseline output produced the lowest MAE, MSE, and RMSE for the pH 6.86 and 9.18 buffers. The recorded Linear Regression condition showed lower standard deviation, drift slopes closer to zero, and lower lag-1 autocorrelation than the baseline and Moving Average conditions. However, because the condition-specific outputs were not retained as paired transformations of the same underlying voltage sequence, these differences are interpreted descriptively and cannot be attributed solely to the Linear Regression transformation. The accuracy performance of Linear Regression varied across the tested buffer values.

Table 2. Accuracy and dispersion metrics for the pH 4.01 buffer solution (reference = 4.01; n = 60).

Method	Mean	MAE	MSE	RMSE	Bias	SD	Descriptive 95% CI of mean
Raw	4.0297	0.0537	0.00482	0.0694	+0.0197	0.0671	4.0123-4.0470
MA5	4.0367	0.0553	0.00517	0.0719	+0.0267	0.0673	4.0193-4.0541
MA10	4.0433	0.0587	0.00579	0.0761	+0.0333	0.0690	4.0255-4.0612
MA20	4.0573	0.0650	0.00669	0.0818	+0.0473	0.0673	4.0400-4.0747
Linear Regression	4.0502	0.0402	0.00207	0.0455	+0.0402	0.0216	4.0446-4.0557

Sequential behavior was evaluated separately to avoid combining accuracy indicators with time-ordered diagnostics in one table. Drift slope was estimated against the measurement-cycle index, whereas lag-1 autocorrelation quantified the association between each observation and the immediately preceding observation. The results are presented in Table 3.

Table 3. Sequential metrics for the pH 4.01 buffer solution (n = 60)

Method	Drift slope (pH cycle-1)	Lag-1 autocorrelation
Raw	-0.003448	0.923
MA5	-0.003615	0.983
MA10	-0.003776	0.994
MA20	-0.003762	0.995
Linear Regression	-0.000035	0.304

For the pH 4.01 buffer, Linear Regression produced the lowest MAE (0.0402), MSE (0.00207), and RMSE (0.0455), corresponding to an approximately 25.2% reduction in MAE relative to the baseline output. The recorded Linear Regression condition also had the lowest SD (0.0216), a drift slope close to zero (−0.000035 pH per cycle), and a lower lag-1 autocorrelation (0.304). These sequential characteristics are reported descriptively and are not interpreted as effects caused solely by the regression transformation. Nevertheless, its mean remained above the nominal buffer value, producing a positive bias of 0.0402 pH units. The Moving Average conditions did not reduce error relative to the baseline output, and larger windows increased error, positive bias, and sequential dependence.

The same evaluation was applied to the nominal pH 6.86 buffer using 60 retained measurement cycles. Table 4 summarizes accuracy, bias, dispersion, and the descriptive interval of the mean output. Bias was calculated as the mean output minus the nominal buffer value, while the 95% confidence interval (CI) of the mean was calculated using the t distribution (df = 59). Because the observations showed serial dependence and were obtained using one sensor unit, the CI was interpreted descriptively and was not adjusted for autocorrelation.

Table 4. Accuracy and dispersion metrics for the pH 6.86 buffer solution (reference = 6.86; n = 60).

Method	Mean	MAE	MSE	RMSE	Bias	SD	Descriptive 95% CI of mean
Raw	6.8798	0.0532	0.00460	0.0678	+0.0198	0.0654	6.8629-6.8967
MA5	6.8853	0.0567	0.00514	0.0717	+0.0253	0.0676	6.8679-6.9028
MA10	6.8917	0.0600	0.00563	0.0751	+0.0317	0.0686	6.8739-6.9094
MA20	6.9058	0.0665	0.00656	0.0810	+0.0458	0.0673	6.8884-6.9232
Linear Regression	6.9977	0.1377	0.02009	0.1418	+0.1377	0.0341	6.9889-7.0065

Sequential behavior was evaluated separately to avoid combining accuracy indicators with time-ordered diagnostics in one table. Drift slope was estimated against the measurement-cycle index, whereas lag-1 autocorrelation quantified the association between each observation and the immediately preceding observation. The results are presented in Table 5.

Table 5. Sequential metrics for the pH 6.86 buffer solution (n = 60).

Method	Drift slope (pH cycle-1)	Lag-1 autocorrelation
Raw	-0.003330	0.934
MA5	-0.003540	0.985
MA10	-0.003692	0.993
MA20	-0.003748	0.995

Method	Drift slope (pH cycle-1)	Lag-1 autocorrelation
Linear Regression	-0.000122	-0.085

For the pH 6.86 buffer, the baseline output produced the lowest MAE (0.0532), MSE (0.00460), and RMSE (0.0678). The Moving Average conditions did not reduce the measurement error relative to the baseline output, and the error increased as the window size became larger. Linear Regression produced the lowest SD (0.0341), an almost zero drift slope (-0.000122 pH per cycle), and negligible lag-1 autocorrelation (-0.085). However, its mean output remained substantially above the nominal buffer value, producing the largest positive bias (0.1377 pH units) and an MAE approximately 158.9% higher than that of the baseline output. Thus, for this buffer condition, the recorded Linear Regression output showed lower sequential variation but substantially poorer agreement with the nominal buffer value than the baseline output.

The same evaluation was applied to the nominal pH 9.18 buffer using 60 retained measurement cycles. Table 6 summarizes accuracy, bias, dispersion, and the descriptive interval of the mean output. Bias was calculated as the mean output minus the nominal buffer value, while the 95% confidence interval (CI) of the mean was calculated using the t distribution (df = 59). Because the observations showed serial dependence and were obtained using one sensor unit, the CI was interpreted descriptively and was not adjusted for autocorrelation.

Table 6. Accuracy and dispersion metrics for the pH 9.18 buffer solution (reference = 9.18; n = 60).

Method	Mean	MAE	MSE	RMSE	Bias	SD	Descriptive 95% CI of mean
Raw	9.2007	0.0507	0.00460	0.0678	+0.0207	0.0652	9.1838-9.2175
MA5	9.2060	0.0523	0.0494	0.0703	+0.0260	0.0658	9.1890-9.2230
MA10	9.2132	0.0575	0.0557	0.0746	+0.0332	0.0674	9.1958-9.2306
MA20	9.2277	0.0643	0.0664	0.0815	+0.0477	0.0666	9.2104-9.2449
Linear Regression	9.2518	0.0722	0.0640	0.0800	+0.0718	0.0356	9.2426-9.2610

Sequential behavior was evaluated separately to distinguish accuracy indicators from time-ordered diagnostics. Drift slope was estimated against the measurement-cycle index, whereas lag-1 autocorrelation quantified the association between each observation and the immediately preceding observation. The results are presented in Table 7.

Table 7. Sequential metrics for the pH 9.18 buffer solution (n = 60).

Method	Drift slope (pH cycle-1)	Lag-1 autocorrelation
Raw	-0.003313	0.918
MA5	-0.003463	0.989
MA10	-0.003655	0.993
MA20	-0.003724	0.996
Linear Regression	-0.000095	0.375

For the pH 9.18 buffer, the baseline output produced the lowest MAE (0.0507), MSE (0.00460), and RMSE (0.0678). The Moving Average conditions did not reduce the measurement error relative to the baseline output, and larger windows increased the positive bias. Linear Regression produced the lowest SD (0.0356), a near-zero drift slope (-0.000095 pH per cycle),

and a lower lag-1 autocorrelation (0.375) than the baseline and Moving Average outputs. However, its mean remained above the nominal buffer value, resulting in a positive bias of 0.0718 pH units and an MAE approximately 42.4% higher than that of the baseline output. Thus, for this buffer condition, the recorded Linear Regression output showed lower dispersion and sequential dependence but did not improve accuracy relative to the baseline output.

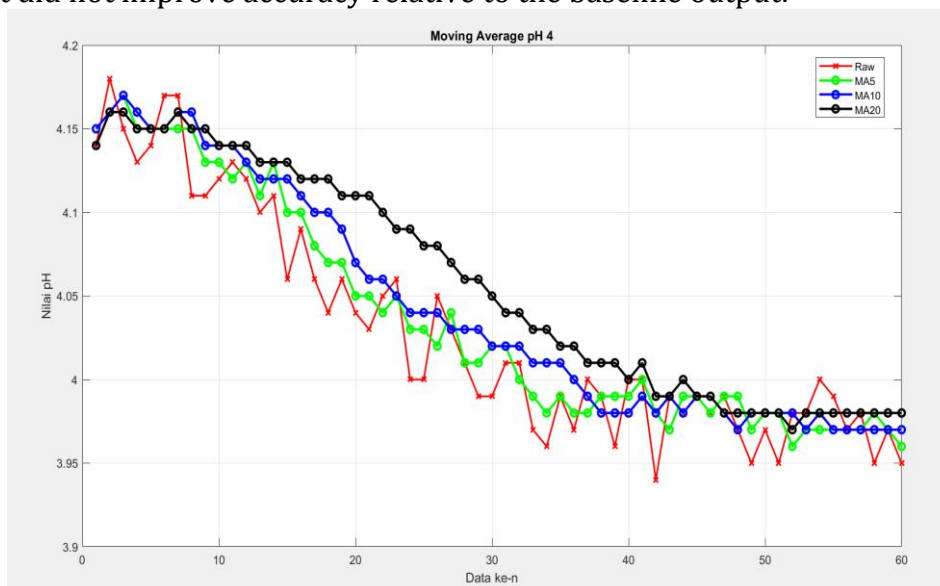


Figure 3. Comparison of baseline output and Moving Average outputs for the pH 4.01 buffer solution.

Figure 3 shows that larger Moving Average windows produced smoother curves while retaining the downward time-ordered trend of the baseline output. The larger-window conditions also showed greater positive bias and higher lag-1 autocorrelation. The visual pattern supports the numerical result that smoothing alone does not automatically improve accuracy.

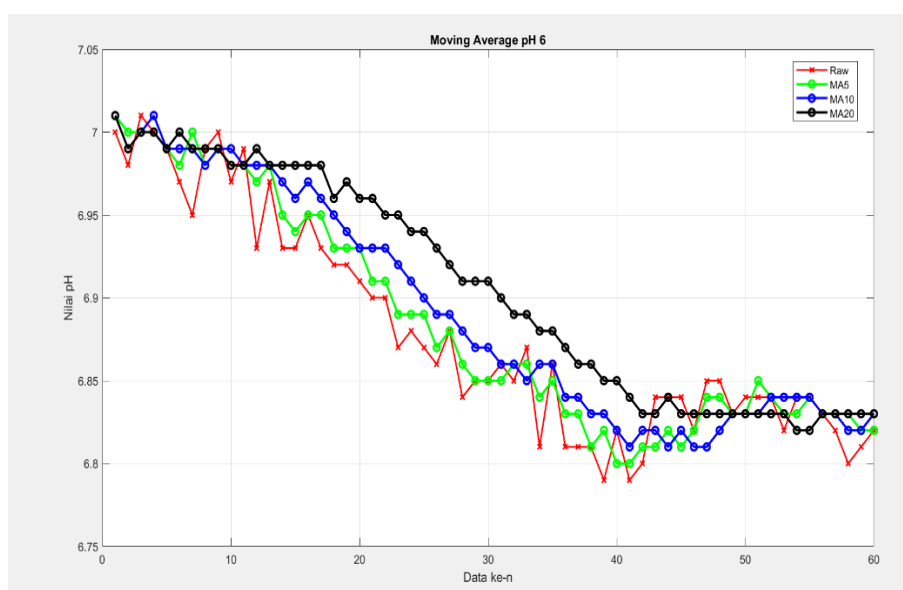


Figure 4. Comparison of baseline output and Moving Average outputs for the pH 6.86 buffer solution.

Figure 4 shows a similar pattern for the pH 6.86 buffer. Increasing the window size made the signal visually smoother, but the filtered output remained positively shifted from the nominal reference. The larger-window conditions also showed higher positive bias and lag-1 autocorrelation.

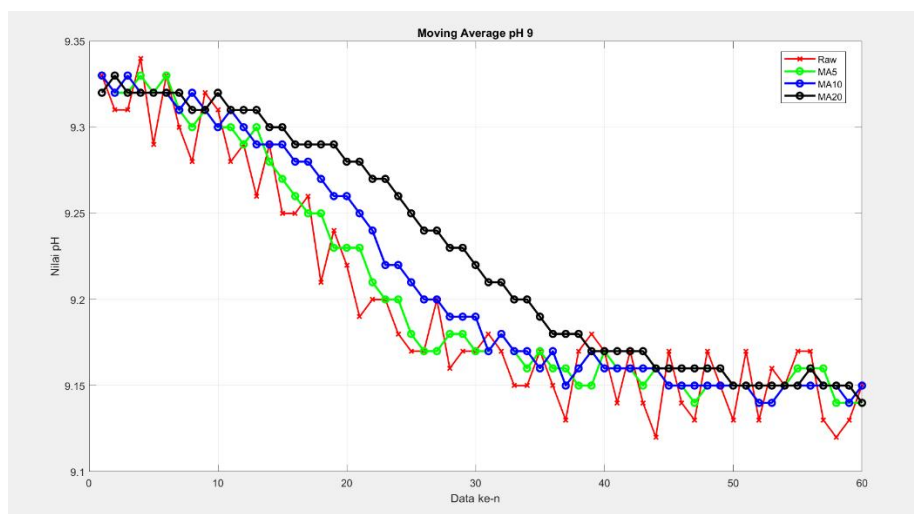


Figure 5. Comparison of baseline output and Moving Average outputs for the pH 9.18 buffer solution.

Figure 5 shows a similar recorded pattern across the tested pH range. The larger window produced stronger smoothing but was also associated with greater positive bias and higher lag-1 autocorrelation. Therefore, the smoother appearance did not correspond to lower measurement error.

Performance of Moving Average Filtering

The results across the three buffer conditions show a consistent pattern: as the Moving Average window increased from MA5 to MA10 and MA20, the MAE, MSE, and RMSE values increased rather than decreased. For the pH 4.01 buffer, the MAE of baseline output was 0.0537 and increased to 0.0553 for MA5, 0.0587 for MA10, and 0.0650 for MA20. Similar trends were observed for the pH 6.86 and pH 9.18 buffers, where MA20 consistently produced the highest error among the tested Moving Average conditions.

This result indicates that the Moving Average Filter did not improve measurement accuracy relative to the nominal buffer values. Instead, error increased gradually as the window size became larger. From a signal-processing perspective, a trailing Moving Average primarily suppresses short-term fluctuation by combining the current observation with preceding samples; it does not use a reference value and therefore cannot directly correct measurement bias (Oppenheim & Schaffer, 2010).

In the recorded sequences, the baseline output showed a gradual downward trend, while the Moving Average conditions exhibited high lag-1 autocorrelation values. Under this condition, earlier and relatively higher observations remained in the trailing average, shifting the filtered estimate upward relative to later observations. This interpretation is consistent with the increasing positive bias and the progressively larger MAE observed as the window size increased. Thus, the smoother appearance of the larger-window outputs did not indicate closer agreement with the nominal buffer values.

From the perspective of practical measurement and science laboratory use, this finding is important because a visually stable reading may create the impression of better measurement quality even when its accuracy is not improved. Therefore, Moving Average should be interpreted as a display-stabilization method or noise-reduction aid, not as a substitute for calibration when the main objective is to obtain values close to reference pH standards.

Performance of Linear Regression as a calibration method

The performance of Linear Regression varied across the tested buffer conditions. For the pH 4.01 buffer, Linear Regression produced the lowest MAE, MSE, and RMSE values of 0.0402, 0.00207, and 0.0455, respectively, representing an approximately 25.2% reduction in MAE relative to the baseline output. However, for the pH 6.86 and 9.18 buffers, the baseline output produced lower error values than Linear Regression. The MAE values of Linear Regression were 0.1377 at pH

6.86 and 0.0722 at pH 9.18, compared with baseline-data MAE values of 0.0532 and 0.0507, respectively. Therefore, the regression model improved measurement accuracy at pH 4.01, but the same improvement was not observed at the other two tested buffer values.

Regression-based calibration has been reported to improve low-cost water-quality sensor accuracy when the model is calibrated and validated against professional reference instruments (Mohd Jais et al., 2024). However, calibration performance depends on environmental conditions, sensor characteristics, the validation instrument, the calibration range, and the selected regression model (Nalakurthi et al., 2024). Vogel et al. (2025) further showed that sample size and the spread of reference pH values can affect calibration performance and generalizability. In the present study, The performance of Linear Regression was not uniform across the three tested buffers; it improved accuracy at pH 4.01 but produced larger positive bias and MAE at pH 6.86 and 9.18. Therefore, the model should be interpreted as a device-specific calibration whose performance depends on the representativeness of the calibration data and the tested range.

The recorded Linear Regression condition showed lower standard deviation, drift slopes closer to zero, and lower lag-1 autocorrelation than the baseline and Moving Average conditions. However, because the condition-specific outputs were obtained from separate measurement sessions and were not retained as paired transformations of the same underlying voltage sequence, these differences cannot be attributed solely to the regression transformation. The stability indicators are therefore reported as descriptive characteristics of the recorded Linear Regression condition.

Lower dispersion did not consistently correspond to better measurement accuracy. At pH 6.86 and 9.18, the Linear Regression condition produced lower variability but larger positive bias and MAE than the baseline output. Accordingly, the regression model cannot be described as simultaneously improving accuracy and stability across the tested pH range.

Comparative analysis between Moving Average Filter and Linear Regression

Moving Average Filter and Linear Regression serve different processing purposes. Moving Average reduces short-term fluctuations by averaging sequential samples, but it does not use reference values and therefore does not directly correct measurement bias. Linear Regression uses calibration data to map sensor voltage onto the reference pH scale. In principle, this approach can correct offset and scale differences, although its effectiveness depends on the calibration dataset and the measurement range being evaluated.

The observed results indicate that the measurement error was not attributable solely to short-term random fluctuation. The baseline and Moving Average outputs exhibited positive bias, negative drift slopes, and high lag-1 autocorrelation, indicating directional deviation and temporal dependence within the recorded sequences. However, these descriptive indicators are not sufficient to determine whether systematic error was the dominant source of measurement error. The Moving Average Filter did not reduce bias or overall error, whereas the effectiveness of Linear Regression varied across the tested buffer values. Therefore, calibration remains important, but its performance should be evaluated across the intended pH range rather than assumed to be uniformly effective.

Table 8 summarizes the percentage change in mean absolute error (MAE) relative to the baseline output for each buffer condition. A negative value indicates a reduction in MAE, whereas a positive value indicates an increase in measurement error.

Table 8. Percentage change in MAE relative to the baseline output (negative values indicate improvement).

Method	pH 4.01 (%)	pH 6.86 (%)	pH 9.18 (%)
Moving Average 5	+3.1	+6.6	+3.3
Moving Average 10	+9.3	+12.9	+13.5
Moving Average 20	+21.1	+25.1	+27.0
Linear Regression	-25.2	+158.9	+42.4

As shown in Table 8, all Moving Average window sizes increased MAE relative to the baseline output. Linear Regression reduced MAE by 25.2% at pH 4.01 but increased it by 158.9% at pH 6.86 and 42.4% at pH 9.18. Therefore, Linear Regression did not provide uniform accuracy improvement across the tested pH range. The recorded Linear Regression condition showed lower variability and sequential dependence, but these differences are interpreted descriptively because the processing conditions were obtained from separate measurement sessions.

Laboratory inquiry can involve collecting, analyzing, and interpreting experimental evidence, while physical and virtual instruments can support investigation of scientific phenomena (de Jong et al., 2013; Hofstein & Lunetta, 2004). Sensor-based applications may also make quantitative experiments and graphical data analysis more accessible (Wijayanti et al., 2022). In this context, the distinction between signal smoothing and calibration may be useful for illustrating that a visually stable sensor output is not necessarily more accurate. However, the present study did not evaluate classroom implementation, student responses, usability, or learning outcomes; therefore, educational relevance is presented only as a potential application requiring separate empirical investigation.

Observed drift and possible measurement-system influences

The time-ordered analysis showed negative drift slopes in the baseline and Moving Average outputs across the three buffer conditions. The baseline-output drift slopes were -0.003448 , -0.003330 , and -0.003313 pH units per cycle at pH 4.01, 6.86, and 9.18, respectively. The Moving Average outputs exhibited high lag-1 autocorrelation values ranging from 0.983 to 0.996. The recorded Linear Regression condition had drift slopes closer to zero, ranging from -0.000035 to -0.000122 pH units per cycle, and lower lag-1 autocorrelation values ranging from -0.085 to 0.375. Because the condition-specific data were collected in separate sessions, these differences are descriptive and do not establish that regression alone caused the lower temporal variation. They nevertheless show that output stability and agreement with nominal pH were not equivalent in the recorded data.

Several measurement-system factors may have contributed to the observed drift and bias. The ESP32 internal ADC can exhibit noise, nonlinearity, imprecise reference voltage, and dead-zone effects (Espinosa-Gavira et al., 2024). The electrical response of a glass pH electrode is also temperature-dependent, and calibration should ideally be performed at the same temperature as measurement (Wang, 2023). Fernandes et al. (2025) demonstrated an ESP32-based multiparametric probe design incorporating temperature compensation, traceable buffer solutions, repeated measurements, and explicit uncertainty estimation, illustrating the controls required for more defensible measurement performance. These factors were not tested separately in the present experiment; therefore, the dominant source of error cannot be identified conclusively. Further work should use a laboratory-grade reference instrument, controlled temperature, multiple sensor units, and an external ADC.

Discussion

The present study demonstrates that signal smoothing and calibration should not be treated as interchangeable strategies for improving the performance of low-cost pH measurement systems. Across the three nominal buffer conditions, Moving Average filtering consistently failed to improve measurement accuracy relative to the baseline output. Instead, MAE, MSE, and RMSE generally increased as the window size expanded from MA5 to MA10 and MA20. By contrast, Linear Regression produced heterogeneous results across the tested pH range. At pH 4.01, it reduced MAE from 0.0537 to 0.0402, corresponding to an improvement of approximately 25.2%; however, at pH 6.86 and 9.18, regression increased MAE by approximately 158.9% and 42.4%, respectively. Mukhammad Indra Ukasya 2193 The Linear Regression condition nevertheless exhibited lower dispersion, drift slopes closer to zero, and lower lag-1 autocorrelation than the baseline and Moving Average conditions. Importantly, these stability differences cannot be interpreted causally because the outputs were obtained from separate measurement sessions rather than as paired transformations of the same underlying voltage sequence. Mukhammad

Indra Ukasya 2193 Collectively, these findings reveal a central measurement principle: a smoother or more stable signal is not necessarily a more accurate signal, and an apparently excellent calibration fit does not guarantee equivalent predictive performance throughout the intended measurement range.

Link to Global Studies, 2020–2026

The findings are consistent with recent international research emphasizing that the primary challenge in low-cost environmental sensing is not data acquisition itself but the reliability, calibration, and validation of the resulting measurements. De Camargo et al. (2023), in a systematic review of low-cost water-quality sensors, found that relatively few published systems had been rigorously compared with reference instruments and identified calibration, precision, reliability, and field validation as persistent limitations. Zhu et al. (2023) reached a similar conclusion from a metrological perspective, arguing that inexpensive water-monitoring sensors require explicit calibration and validation procedures before their output can be considered suitable for scientific monitoring. These observations directly reinforce the present finding that neither visual smoothness nor a simple calibration equation is sufficient evidence of measurement validity. [PMC](#)

The variable performance of Linear Regression is particularly informative. Mohd Jais et al. (2024) reported substantially improved performance of low-cost water-quality sensors after calibration against professional reference instruments, including an analog pH sensor calibrated with standard buffers. In contrast, the current regression model improved accuracy at only one of the three tested buffer values. This discrepancy is technically plausible rather than contradictory. Calibration effectiveness depends on sensor-specific characteristics, calibration-data coverage, reference quality, environmental conditions, and the relationship between calibration and validation data. Liang and Daniels (2022) similarly showed that low-cost sensor calibration is sensitive to sample size, predictor selection, and the representativeness of calibration observations, while Vogel et al. (2025) demonstrated that calibration performance may vary substantially with the number and distribution of reference values. More recent calibration research has therefore increasingly adopted adaptive or machine-learning approaches capable of accounting for nonlinearities, drift, and environmental interactions rather than relying exclusively on a single static regression equation (Mahajan & Helbing, 2025; Nan et al., 2025; Qian et al., 2026). [DOI](#)

The current experiment also helps clarify why Moving Average filtering did not reduce measurement error. A trailing Moving Average is designed primarily to attenuate high-frequency fluctuation by combining current and preceding observations; it does not incorporate an external reference and therefore cannot intrinsically correct calibration bias. In the present sequences, the baseline output exhibited negative drift while Moving Average conditions showed very high lag-1 autocorrelation. Consequently, earlier and relatively higher measurements remained within the moving window as the underlying series declined, producing an upward displacement that became more pronounced with larger windows. The article's figures for pH 4.01, 6.86, and 9.18 visually show this combination of increased smoothness and persistent positive displacement. Mukhammad Indra Ukasya 2193 This result complements Ranjan et al. (2024), who showed that the usefulness of moving-average and other filters depends strongly on signal dynamics and the performance criterion being optimized. Their findings, together with the present results, suggest that filtering parameters cannot be selected simply on the assumption that greater smoothing corresponds to greater accuracy. [MDPI](#)

The broader low-cost water-monitoring literature further supports the need to distinguish sensor architecture from metrological performance. Prapti et al. (2022) identified pH as one of the most frequently monitored parameters in IoT-based aquaculture but also noted the need for longer-term experimental validation. Chen et al. (2022) similarly integrated pH and other water parameters into an IoT fish-farm monitoring architecture while explicitly recognizing sensor maintenance and calibration requirements. Recent developments have gone further by incorporating temperature compensation, reference-grade validation, multiple sensor units, and

long-term stability assessment. Zeta et al. (2025), for example, developed a low-cost pH sensor incorporating temperature compensation and machine-learning-based electronic calibration, whereas Naloufi et al. (2024) evaluated the long-term reliability of multiparameter low-cost systems against controlled laboratory and field conditions. Fernandes et al. (2025) likewise emphasized traceable buffers, temperature considerations, repeated validation, and explicit uncertainty assessment in a cost-effective aquatic probe. These studies indicate that future improvement of the present ESP32–PH-4502C platform should focus less on choosing between a simple Moving Average and a single regression equation and more on strengthening the complete calibration-validation chain. [Wiley Online Library](#)

Link to Indonesian and Local Studies, 2020–2026

The Indonesian evidence provides an especially useful comparison. Fiddariani and Sumanto (2024) evaluated Kalman, Moving Average, and Hanning filters for fish-pond monitoring and reported that Moving Average filtering slightly reduced MSE for a PH-4502C pH sensor, whereas Kalman filtering considerably reduced output dispersion. Their result differs from the present study, where every tested Moving Average window increased MAE relative to the baseline. [Jurnal Universitas Gadjah Mada](#) Rather than indicating an inconsistency, the contrast shows that filter performance depends on the temporal structure of the raw signal, filter implementation, window length, calibration function, experimental sequence, and reference procedure. The present data contained pronounced time-ordered dependence and downward drift; under such conditions, a trailing window can preserve obsolete higher measurements and increase bias. Thus, an algorithm that improves MSE in one hardware and measurement configuration cannot be assumed to produce equivalent improvement in another.

Other Indonesian studies demonstrate the increasing use of ESP32-based and IoT-enabled water-quality monitoring while simultaneously underscoring the importance of calibration. Ardyansyah et al. (2022) incorporated pH and temperature into an aquaculture water-quality control system, whereas more recent Indonesian systems have combined ESP32 hardware with pH, dissolved oxygen, temperature, TDS, conductivity, and Water Quality Index calculations. These developments indicate that Indonesia is moving from proof-of-concept IoT monitoring toward explicit attention to calibration and data-quality assurance. The present study contributes to this trajectory by demonstrating empirically that processing algorithms themselves must also be validated rather than assumed to improve sensor quality.

Theoretical, Pedagogical, and Policy Implications

Theoretically, the results strengthen the distinction among **accuracy, precision, stability, and calibration** in measurement science: Moving Average filtering primarily modifies temporal variability, whereas regression calibration modifies the mathematical mapping between sensor voltage and the reference scale; consequently, neither reduced variation nor a high calibration $\backslash(R^2\backslash)$ is sufficient evidence of lower measurement error outside the calibration sample. This distinction refines the interpretation of low-cost sensing by showing that filtering and calibration should be evaluated against separate performance criteria. Pedagogically, the system has potential value for inquiry-based science and engineering education because students can directly investigate the difference between raw data, filtered data, calibration equations, bias, drift, MAE, RMSE, and autocorrelation. However, the study did not involve students and therefore does **not** provide evidence of effects on student engagement, emotional well-being, media literacy, online behavior, or identity formation. Digital-learning studies such as Hariandini et al. (2026) and science-media research such as Mahfud et al. (2026) support purposeful integration of digital tools into inquiry and evidence interpretation, but classroom effects of this sensor must be tested separately. [WISE Edu](#) From a policy perspective, the findings are more relevant to environmental-monitoring and laboratory-quality policies than to social-media or mental-health regulation. Procurement and deployment standards for low-cost monitoring systems should require transparent calibration equations, traceable reference measurements, temperature recording or compensation, validation across the operating range, multiple-device replication, drift assessment, and explicit reporting of uncertainty. This aligns with broader sustainability-

governance arguments that environmental decision-making should be supported by scientifically defensible monitoring evidence rather than technological adoption alone (Febriani et al., 2026). [WISE Edu](#)

Novelty and Contribution

The novelty of this study lies not in introducing Moving Average filtering or Linear Regression as new algorithms but in demonstrating, within the same low-cost ESP32-PH-4502C measurement architecture, that **signal smoothness, output stability, and accuracy can move in different directions**. Many IoT sensor-development studies report successful system integration or calibration but provide limited comparative examination of MAE, MSE, RMSE, bias, dispersion, drift, and serial dependence simultaneously. The present research contributes a more discriminating evaluation by comparing three Moving Average window sizes, a baseline conversion, and a regression-calibrated condition across acidic, near-neutral, and alkaline buffers. Particularly important is the finding that the regression model achieved an extremely strong calibration fit yet did not generalize uniformly to the validation measurements, while Moving Average filtering visually stabilized the output without improving reference agreement. This evidence challenges the common engineering assumption that smoother signals necessarily constitute better measurements and demonstrates the importance of separating signal-processing objectives from calibration objectives in resource-constrained sensing systems.

Limitations and Future Research

The findings must nevertheless be interpreted within substantial experimental constraints. Only one ESP32-PH-4502C sensor unit was tested; measurements were conducted using nominal buffers rather than simultaneous readings from a laboratory-grade reference pH meter; buffer order was fixed; temperature was estimated at approximately 25 °C rather than actively controlled and continuously measured; and condition-specific outputs were obtained in separate sessions rather than generated as paired transformations of exactly the same raw voltage series. Mukhammad Indra Ukasya 2193 These limitations restrict construct, internal, and external validity and prevent causal attribution of the lower drift or dispersion observed in the Linear Regression condition. Mukhammad Indra Ukasya 2193 Future studies should therefore use multiple independently manufactured sensor units, certified reference instruments, random or counterbalanced buffer orders, continuously recorded temperature, external precision ADCs, and identical raw-voltage sequences processed offline through every competing algorithm. Calibration should include more reference points distributed across the intended pH range and should be evaluated using independent validation or cross-validation. Future comparisons should also include Kalman, robust regression, piecewise calibration, polynomial or spline models, and appropriately constrained machine-learning approaches. Finally, laboratory validation should be followed by prolonged testing in real water matrices because ionic strength, fouling, electrical interference, temperature variation, and sensor ageing may produce performance characteristics that cannot be reproduced using standard buffers alone.

CONCLUSION

This study compared the recorded performance of Moving Average filtering and Linear Regression calibration in an ESP32-based pH sensor using nominal buffers of pH 4.01, 6.86, and 9.18. Moving Average filtering did not improve accuracy, and MAE increased as the window size increased. Linear Regression reduced MAE by 25.2% at pH 4.01 but increased it by 158.9% at pH 6.86 and 42.4% at pH 9.18. The recorded Linear Regression condition showed lower dispersion, drift slopes closer to zero, and lower lag-1 autocorrelation; however, because the conditions were obtained from separate measurement sessions, these differences cannot be attributed solely to regression. Thus, neither method was uniformly superior: Moving Average primarily altered signal smoothness, whereas regression accuracy depended on the calibration data and tested buffer value. Because the study used one sensor unit, nominal buffers, a fixed buffer order, approximate room-temperature conditions, and no reference-grade pH meter, the findings represent a device-specific laboratory evaluation rather than field validation. Further validation

should address the reference, replication, temperature-control, and paired-data limitations identified in this study.

ACKNOWLEDGMENT

The authors acknowledge Universitas Negeri Surabaya for supporting the preparation of this research manuscript and the experimental documentation used in the development of the ESP32-based pH sensor system.

REFERENCE

- Ardyansyah, H., Pradana, R., & Puspasari Handayani, I. (2022). Fuzzy logic dalam sistem kualitas air kolam *Betta splendens* berdasarkan pH dan suhu. *Jurnal Ticom: Technology of Information and Communication*, 10(2), 78–86. <https://doi.org/10.70309/ticom.v10i2.21>
- Chen, C.-H., Wu, Y.-C., Zhang, J.-X., & Chen, Y.-H. (2022). IoT-based fish farm water quality monitoring system. *Sensors*, 22(17), Article 6700. <https://doi.org/10.3390/s22176700>
- de Camargo, E. T., Spanhol, F. A., Slongo, J. S., da Silva, M. V. R., Pazinato, J., de Lima Lobo, A. V., Coutinho, F. R., Pfrimer, F. W. D., Lindino, C. A., Oyamada, M. S., & Martins, L. D. (2023). Low-cost water quality sensors for IoT: A systematic review. *Sensors*, 23(9), Article 4424. <https://doi.org/10.3390/s23094424>
- de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science*, 340(6130), 305–308. <https://doi.org/10.1126/science.1230579>
- Espinosa-Gavira, M. J., Agüera-Pérez, A., Palomares-Salas, J. C., Sierra-Fernandez, J. M., Remigio-Carmona, P., & González de-la-Rosa, J. J. (2024). Characterization and performance evaluation of ESP32 for real-time synchronized sensor networks. *Procedia Computer Science*, 237, 261–268. <https://doi.org/10.1016/j.procs.2024.05.104>
- Febriani, E., Woods, R., Madkur, A., & Hasanah, U. (2026). Implementation of the Sustainable Development Goals (SDGs) in the education and environmental sectors: A comparative study of Indonesia and Australia. *Gender, Islamic Law, and Sustainability*, 1(2), 102–116. <https://doi.org/10.70211/gils.v1i2.547>
- Fernandes, S., Fialho, A., Santos, J. M., Ferreira, T., & Filipe, A. F. (2025). Development of a cost-effective multiparametric probe for continuous real-time monitoring of aquatic environments. *Sensors*, 25(23), Article 7110. <https://doi.org/10.3390/s25237110>
- Fiddariani, A. S., & Sumanto, B. (2024). Studi penerapan filter digital pada sistem pemantau parameter di kolam ikan. *Jurnal Listrik, Instrumentasi, dan Elektronika Terapan*, 5(1), 30–37. <https://doi.org/10.22146/juliet.v5i1.87377>
- Hariandini, Y., Yuda, I. B., & Ahmed, W. (2026). Implementing digital integrated learning in schools: Evidence from classroom practices. *Digital Learning, Social Science, and Life-course Studies*, 2(1), 46–56. <https://doi.org/10.70211/disolife.v2i1.348>
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, 88(1), 28–54. <https://doi.org/10.1002/sce.10106>
- Jan, F., Min-Allah, N., & Düşteğör, D. (2021). IoT based smart water quality monitoring: Recent techniques, trends and challenges for domestic applications. *Water*, 13(13), Article 1729. <https://doi.org/10.3390/w13131729>
- Lakshmikantha, V., Hiriyanagowda, A., Manjunath, A., Patted, A., Basavaiah, J., & Anthony, A. A. (2021). IoT based smart water quality monitoring system. *Global Transitions Proceedings*, 2(2), 181–186. <https://doi.org/10.1016/j.gltp.2021.08.062>
- Liang, L., & Daniels, J. (2022). What influences low-cost sensor data calibration? A systematic assessment of algorithms, duration, and predictor selection. *Aerosol and Air Quality Research*, 22, Article 220076. <https://doi.org/10.4209/aaqr.220076>
- Mahajan, S., & Helbing, D. (2025). Dynamic calibration of low-cost PM2.5 sensors using trust-based consensus mechanisms. *npj Climate and Atmospheric Science*, 8, Article 257. <https://doi.org/10.1038/s41612-025-01145-2>
- Mahfud, N. B. A., Indana, S., & Isnawati. (2026). Development of a problem based learning science e-module integrating Ledre local wisdom to foster junior high school students' critical thinking skills. *JPI: Jurnal Pustaka Indonesia*, 6(2), 452–472. <https://doi.org/10.62159/jpi.v6i2.2586>
- Mohd Jais, N. A., Abdullah, A. F., Mohd Kassim, M. S., Abd Karim, M. M., Abdulsalam, M., & Muhadi, N. 'A. (2024). Improved accuracy in IoT-based water quality monitoring for aquaculture tanks using low-

- cost sensors: Asian seabass fish farming. *Heliyon*, 10(8), Article e29022. <https://doi.org/10.1016/j.heliyon.2024.e29022>
- Nalakurthi, N. V. S. R., Abimbola, I., Ahmed, T., Anton, I., Riaz, K., Ibrahim, Q., Banerjee, A., Tiwari, A., & Gharbia, S. (2024). Challenges and opportunities in calibrating low-cost environmental sensors. *Sensors*, 24(11), Article 3650. <https://doi.org/10.3390/s24113650>
- Naloufi, M., Abreu, T., Souihi, S., Therial, C., Rodrigues, N. A. d. P., Le Goff, A. G., Saad, M., Vinçon-Leite, B., Dubois, P., Delarbre, M., Kennouche, P., & Lucas, F. S. (2024). Long-term stability of low-cost IoT system for monitoring water quality in urban rivers. *Water*, 16(12), Article 1708. <https://doi.org/10.3390/w16121708>
- 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>
- Olanubi, O. O., Akano, T. T., & Asaolu, O. S. (2024). Design and development of an IoT-based intelligent water quality management system for aquaculture. *Journal of Electrical Systems and Information Technology*, 11(1), Article 15. <https://doi.org/10.1186/s43067-024-00139-z>
- Oppenheim, A. V., & Schaffer, R. W. (2010). *Discrete-time signal processing* (3rd ed.). Pearson.
- Prapti, D. R., Mohamed Shariff, A. R., Che Man, H., Ramli, N. M., Perumal, T., & Shariff, M. (2022). Internet of Things (IoT)-based aquaculture: An overview of IoT application on water quality monitoring. *Reviews in Aquaculture*, 14(2), 979–992. <https://doi.org/10.1111/raq.12637>
- Qian, J., Wynn, T., Liu, B., Shan, Y., Bartington, S. E., Pope, F. D., Dai, Y., & Shi, Z. (2026). Enhancing accuracy of indoor air quality sensors via automated machine learning calibration. *Atmospheric Measurement Techniques*, 19, 603–615. <https://doi.org/10.5194/amt-19-603-2026>
- Ranjan, R., Shin, D., Jung, Y., Kim, S., Yun, J.-H., Kim, C.-H., Lee, S., & Kye, J. (2024). Comparative analysis of integrated filtering methods using UWB localization in indoor environment. *Sensors*, 24(4), Article 1052. <https://doi.org/10.3390/s24041052>
- Singh, Y., & Walingo, T. (2024). Smart water quality monitoring with IoT wireless sensor networks. *Sensors*, 24(9), Article 2871. <https://doi.org/10.3390/s24092871>
- Slongo, J., Lindino, C., Martins, L. D., Spanhol, F. A., Carneiro, E., & Camargo, E. T. (2024). Evaluation of low-cost sensors to integrate in a water quality monitor for real-time measurements. *Environmental Monitoring and Assessment*, 196(8), Article 716. <https://doi.org/10.1007/s10661-024-12884-9>
- Vogel, S., Gebbers, M., Schröter, I., Schwanghart, W., Bönecke, E., Rühlmann, J., Kramer, E., & Gebbers, R. (2025). Towards site-independent calibration of in situ soil pH sensor data: Relevance of spatial and temporal proximity, sample size and data spread for calibration model performance. *Geoderma*, 456, Article 117261. <https://doi.org/10.1016/j.geoderma.2025.117261>
- Wang, J. (2023). *Analytical electrochemistry* (4th ed.). John Wiley & Sons.
- Wijayanti, A., Marwoto, P., Wiyanto, W., Ridlo, S., & Parmin, P. (2022). Contextual science learning using smartphone audio sensor application (Phyphox) on restitution coefficient measurement. *Jurnal Penelitian dan Pembelajaran IPA*, 8(2), 256. <https://doi.org/10.30870/jppi.v8i2.13400>
- Zeta, B. M. A., Alam, S. U., Rahman, G. M. A. E. U., & Ahmed, K. I. U. (2025). A low-cost pH sensor for real-time monitoring of aquaculture systems in a multi-layer wireless sensor network. *Sensors*, 25(9), Article 2824. <https://doi.org/10.3390/s25092824>
- Zhu, Q., Cherqui, F., & Bertrand-Krajewski, J.-L. (2023). End-user perspective of low-cost sensors for urban stormwater monitoring: A review. *Water Science & Technology*, 87(11), 2648–2684. <https://doi.org/10.2166/wst.2023.142>