ELKHA : Jurnal Teknik Elektro, Vol. 15 No.1, April 2023, pp. 32 - 40
ISSN: 1858-1463 (print), 2580-6807 (online)

Reliability Analysis of pH Measurement on TLC4502 with E201C Electrodes based on ATmega328P Microcontroller: Approach to Analysis of Variation with ANOVA

James Julian[1*)], Fitri Wahyuni[2)], and Faiz Daffa Ulhaq[3)]

[1, 2)] Department of Mechanical Engineering, University of Pembangunan Nasional Veteran Jakarta, Indonesia
[3)] Department of Electrical Engineering, University of Pembangunan Nasional Veteran Jakarta, Indonesia
Corresponding Email: [*)] zames@upnvj.ac.id

Abstract **– The development of the water management system has so far reached the stage of utilizing IoT technology in the monitoring and operation process. An essential factor in water that affects the quality of a substance is pH. The research aims to analyze and ensure that the devices have a small pH measurement error rate with TLC4502 & E201C. The calibration process was carried out using linear regression, and r^2 value of 0.99 was obtained. Analysis was carried out using one-way ANOVA and Tukey HSD methods, and it was found that all data pairs rejected the null hypothesis (H_0) and accepted the alternate hypothesis (H_1). This hypothesis indicated a significant difference in the relative measurement error of pH in each condition. The standard error value of each measurement after filtration was 0.00, with an uncertainty value ranging from 0.07 to 0.02. If the sensor can provide measurement results with low error and high accuracy, then the sensor can be widely circulated and used. Through this research, the feasibility of a measuring instrument was developed based on the perspective of errors and high accuracy. A quality measuring instrument certainly helpful in various fields from the fisheries, hydroponics, and environmental sectors.**

Keywords: ANOVA, Chauvenet Criterion, E201C, TLC4502.

I. Introduction

Indonesia is a country that has abundant water resources in various regions. Water is a substance that plays the most critical role as a source of life. Water is a medium for transporting nutrients, oxygen, and other life-supporting substances. One of the essential roles of water is as a medium for cultivating animals and plants [1]. Due to its transport of other compounds, water can also have negative impacts, such as environmental pollution, if not managed correctly. The development of water management systems has reached the stage of utilizing IoT technology in the monitoring and operating process [2], [3]. The sustainability of the surrounding environment can be affected by water quality. An essential ingredient in water is the potential of hydrogen. PH refers to the number of hydrogen ions present in the liquid. The amount of hydrogen concentration in water is directly proportional to the amount of acidic level. In this condition, the pH value of the water is lower. In Contrast, if the hydrogen is lower, the liquid is alkaline with the pH value of the water is higher [4].

The pH content in the water is the main parameter in determining whether or not the water is suitable for life. The influence of acidity levels on the development of organisms and microorganisms is dominant because, in fish and plant cultivation, the low pH level makes it difficult for beneficial microorganisms to grow [5]. The acidity level of water usually has a set point pH 3 – 7, to support the fish growth ecosystem [3]. For example, low pH levels make it difficult for fish to live in ponds, so only a tiny fish population can survive in farming. Harvest quality in fish farming can decrease due to the pH quality not being maintained at neutral levels with a value of 7. So, the pH level and its stability must be maintained to ensure that the development of the object of cultivation is not hampered [6].

The pH level in sweat solutions is also an indicator of health in medicine. The pH level measured with a flexible sensor integrated with a 3D PANI electrode accompanied by Ag/AgCl. Maximum sensor response time of 7.75 seconds to get readings with good accuracy. Measurements can still be trusted with a pH value susceptibility limit of 4 to 9. However, the sensor can present data that is not the same as the actual value or experience an error [7].

The error rate on the device can vary depending on the quality of the sensor, the data processing base, and how the device is used. Research on pH sensors compared values before and after calibration. PH V.1.1 (PH-4502C) sensors are used as measurement media with ATmega328P processing base. By using the Linear Trendline Approach method, the average percentage of absolute measurement error is 5%. Despite the average error decrease and irregular error distribution with a linear increase in pH-meter level during the measurement process, the data validity is still questionable. This phenomenon raises doubts about the success of the tool

Manuscript received 2023-02-14; revised 2023-03-14; accepted 2023-03-19

testing process and result. The purpose of this research is to carry out the refinement stage of the sensor analysis process [8].

This research aimed to develop a calibrated measurement technique that uses linear regression to compare the pH sensor readings on the microcontroller with the real-time pH value of the measuring device. The results of sensor measurement data were processed using the ANOVA test to assess measurement accuracy. The pH value of the measurement results of the device is also collected and then distributed to determine the standard deviation, which will later be used to find the uncertainty level from the sensor. The need for pH meters in the industry and the shortcomings left behind by devices in several articles are the main factors in this study discussing pH sensors.

II. METHODOLOGY

This study was designed with a quantitative research method using Linear Regression in the sensor calibration process. Linear Regression can help measurements to find the actual value of the sensor's measured value. The pH sensor calibration process uses pH Up and pH Down solutions as calibrators. pH measurements were taken one thousand times at each pH level and then averaged. In the next step, the data were analyzed using descriptive statistics. Chauvenet criterion process can remove outliers in data, and ANOVA was introduced to analyze the data. The analysis results are displayed as variations in distribution measurements obtained during the research process.

A. Linear Regression

The calibration process in the regression analysis involves using an optimization model to align the linearity between observed data from the independent variable, namely the value read by the sensor, and the dependent variable, namely the pH read from the measuring instrument [9]. A linear line is drawn based on the measurement results of the digital multimeter, which intersects with the sensor's analog value reading. Sensor measurements outside the distribution without a calibration process are more difficult to identify because the observed results produce higher deviations [10]. To determine linear regression using the following equation: [11]

$$y = \beta_0 + \beta_1 x \tag{1}$$

$$\beta_0 = \frac{[\sum y \sum x^2 - \sum x \sum xy]}{(n \sum x^2) - (\sum x)^2} \tag{2}$$

$$\beta_1 = \frac{[n \sum xy - (\sum x)(\sum y)]}{n \sum x^2 - (\sum x)^2} \tag{3}$$

$$y = \left[\frac{[\sum y \sum x^2 - \sum x \sum xy]}{(n \sum x^2) - (\sum x)^2}\right] + \left[\frac{[n \sum xy - (\sum x)(\sum y)]}{n \sum x^2 - (\sum x)^2}\right] x \tag{4}$$

where y is the dependent data value, β_0 is the linear line constant intercept variable (analog value), β_1 is the slope coefficient of the linear line (pH value), x is the independent data value, and n is the number of values.

The formulation of each line slope value uses equations (2) and (3). Then the equation (1) is embedded in the ATmega328P program as a basis for processing sensor data. The level of measurement error after the calibration process is close to the actual value compared to pH measurements using standard measuring instruments. However, it is necessary to conduct further analysis of the data to determine whether the measurement results from normal variations or has differences in the results of the distribution of data variations.

Before the ANOVA test is applied in the data analysis process, the decision to remove outliers is made to reduce noise in the data. The difference between the actual value on the measuring instrument and sensor measurements in the form of a percentage is a relative measurement error. The relative error is used to evaluate the accuracy of measurements or calculations. When the measurement deviation is too significant, the data cannot be classified into relative error. Due to the low probability of occurrence, the Chauvenet method was introduced to identify abnormal data in measurements.

B. Chauvenet Criterion

Chauvenet controls errors by using unpaired data types (abnormal data) [11], where the data deviates too far from other measurement results. That matter can be caused by noise from taking measurement data quickly and repeatedly. The data noise is included in the random error category. This reason categorizes noise data as invalid because the resulting values are not included in the measurement distribution results [12]. The outlier removal method is carried out to eliminate data outside the category. The Z-score on the chauvenet estimates the possibility of outliers (very different values) in the data, where the value must not exceed the data deviation. The process of filtering outliers can be done using the following equation: [11]

$$D_{max} \geq z_{score} \tag{5}$$

$$Q(P_z) \geq \frac{|x_{sus} - \bar{x}_{mean}|}{S_x} \tag{6}$$

where D_{max} is the maximum allowable deviation, z_{score} is the Z-score calculation, Q is the quantile distribution of the data, P_z is the representative of one of the probability distributions, x_{sus} is the expected outlier value, $\bar{x}_{mean}$ is the sample mean, S_x is the standard deviation sample.

Data is classified as an outlier when $D_{max} \geq z_{score}$ is rejected or inappropriate. $D_{max} > z_{score}$, so the data is classified as an outlier. If the conditions fulfilling the $D_{max} \geq z_{score}$ equation is accepted, the data will still enter the distribution. Looking for outliers one by one will undoubtedly take quite a long time. The calculation of D_{max} and z_{score} is done by programming to detect outlier data, so that the distribution results are obtained, which are ready to be used in the ANOVA test.

C. Uncertainty Level

Distribution adjustment by removing data that is not needed can help to find the standard deviation. The standard deviation can be used in determining the variable magnitude of the percentage level of uncertainty. After getting the standard deviation and average, the analysis continues to find the sensor measurement's uncertainty level. The level of uncertainty can be used as an indicator of the reliability aspect of measurement [13]. The 95% confidence level is determined by referring to several studies. The 95% confidence model with a statistical significance of 5% is often used for data representation of the level of uncertainty. To determine standard deviation using the following equation: [14]

$$\sigma^2 = \frac{\sum(x_i-\mu)^2}{n} \tag{7}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2} \tag{8}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{9}$$

where σ is the standard deviation, x_i is the individual value in the population, μ is the average value of the population, and $\sigma_{\bar{x}}$ is the standard error. Determining the standard deviation requires many distributions in the population collection of data variations. Equation (8) calculates the deviation value of all data collected. The data distribution will be better if the deviation value is closer to zero. Then, equation (9) can be used to determine the desired measurement margin error capacity (confidence interval) of 68.3%, 90%, 95%, 99%, or 99.99%.

D. ANOVA Test

ANOVA assumes that the data is usually distributed and that the compared groups have the same variation. The ANOVA results may be biased if the assumptions are not met [15]. The type of equation used in the calculation is one way ANOVA, with pH error data as a comparison of variance between groups. Then from these data, the distribution between groups can be seen whether there is a significant difference or not. The data on the ANOVA distribution chart has an upper and lower bound, with a continuous line representing excess distribution in the boxed region and little data in the striped portion. To calculate the ANOVA or F-ratio or coefficient of the data using the following equation: [15]

$$SSG = \sum_{i=1}^{k} n_i(\bar{x_i} - x)^2 \tag{10}$$

$$SSE = \sum_{i=1}^{k}(n_i - 1){s_n}^2 \tag{11}$$

$$F = \frac{SSG/(k-1)}{SSE/(N-k)} \tag{12}$$

$$F = \frac{MSG}{MSE} \tag{13}$$

where SSG is the sum of the squares between groups, k is the number of groups, $\bar{x_i}$ is the group mean of i. SSE is the sum of the squares in the group, s_n is the variance value of each group. F is the ratio of the comparison values between data variants, MSG is the average squared between groups, MSE is the average sum of squares within groups. Equation (13) represents the ANOVA test to find the data coefficient.

E. Tukey's HSD

The Tukey method was introduced as a companion to the ANOVA distribution plot. Tukey's Honestly Significant Difference (HSD) was applied to identify plots against the ANOVA distribution results mathematically [16]. This method is used to determine the hypothesis of the nine data variations by classifying the data as accepted or rejected. Equation for Tukey's HSD is shown: [16]

$$HSD = q\sqrt{\frac{MS_W}{n_k}} \tag{14}$$

where MS_W is the squared mean for within the group from ANOVA, q is the standardized statistics range, and n_k is the number in each category. Then, with H_0 = the average value of the data distribution does not have a significant difference, H_1 = the average value of the data distribution has a significant difference. H_0 and H_1 describe the condition of the data representation. A comparison is made between the data of each set of results, using boundaries with a p-value > 0.05.

F. Relative Error

Relative error is used to measure the accuracy of a measurement and allows comparisons between the results of different measurements or measurements with different measuring instruments. A small relative error value indicates that the measurement results are highly accurate, while a high relative error value indicates that the measurement results are not very accurate [17]. Relative percent error formula: [8]

$$\delta = \frac{|v_A - v_E|}{v_E} \times 100 \tag{15}$$

where δ is the resulting percentage error, v_A is the actual value of the measurement results, v_E is the expected value from the measurement or analog measurements.

G. Standard Deviation Relative Error

Uncertainty means deviation error is expressed as the standard deviation value of a series of measurements or as the deviation error of the result of a series of measurements. It measures how much variation there is in the measurement results from the mean value of the measurements and can be used to determine the degree of uncertainty of a measurement. A slight uncertainty value means deviation error indicates that the measurement results of the device are working correctly. In contrast, a significant uncertainty means deviation error, indicating that the measurement results of the device are not good enough. [14]

$$u_{error} = \sqrt{\frac{\sum|\delta|^2}{(n-1)}} \tag{16}$$

where u_{error} is the uncertainty mean deviation error, δ is the absolute value of the error, n is the number of numbers.

H. Q-Q Plot

Q-Q plots (quartile-quartile plots) can be used to assess the normality of the level of deviation of a data set by plotting the tested data against the normal distribution quartiles [9]. Data is sorted from the lowest to the highest value in making a Q-Q plot. The data are represented as points in the actual data, and these points will remain on the line formed by the z-score of each data. If the data is normally distributed, the points representing the data will follow a linear line and not deviate significantly. Conversely, if the data is distributed randomly, the data can be better in terms of order and distribution.

I. Circuit Schematic

The circuit has a 5 volt supply to the ATmega328P with the LM2596 regulator IC as the input voltage regulator to stabilize the input power. The goal is that the value of the sensor reading is more accurate with a stable voltage. ATmega328P is used as a data processing base from the value measured by the pH sensor. The PH-4502C sensor read the analog value that goes into the ATmega328P, which will be compared with a standard pH meter for calibration. Figure 1 shows, digital pH meter measuring instrument which has a measurement range 1 to 14 with accuracy $\pm$ 0.2.

Figure 1. Digital pH meter

TLC4502 and E201C is common pH sensor easy to find. Figure 2 shows a schematic circuit of a device. Since the TLC4502 and E201C can be found on various e-commerce sites, it is easier to experiment with measuring pH values. The challenge in using sensors is applying source code from the library, where several stages of measuring sensor reading values do not match the values of the measuring instruments used. With these unsatisfactory results, a calibration process was carried out independently to adjust for fluctuations in linear changes in sensor readings with a pH meter measuring device [8]. Then, the calibration results are analyzed. The sensor used is included in the criteria for the V.1.1 pH Module series. Probe E201C works with a temperature range of 10 to 50°C. The probe contains two electrodes immersed in the solution: the working electrode and the reference electrode. The pH value is measured from the potential difference between the two electrodes, which changes based on the number of hydrogen ions (H) [15]. TLC4502 is an electric signal amplifier that amplifies signals into measurable quantities. This device is usually used in measurement applications on two-foot electrodes, hotwire, and rogowski coil. The measurement results are stabilized and amplified into an electrical signal that can be measured using specific electronic circuits [15].

The engineering process of changing the pH value uses Potassium Hydroxide (KOH) compounds to increase pH levels, while Phosphoric Acid compounds (H_3PO_4) to lower pH levels. If the solution is alkaline, the probe electrode on the sensor will be negatively charged. Vice versa, if the solution is acidic, the electrode probe will be positively charged [18]. However, the work of the E-201 probe requires a reading signal amplifier circuit in the form of a transducer with the aim that the voltage reading can be read/appropriately entered into the microcontroller.

The transducer circuit for the pH sensor uses an amplifier with IC 4502C. This circuit is used to amplify a weak input signal from the output of the pH sensor so that it can be used to control the 0 – 5V reading voltage for managing the microcontroller Analog to Digital Converter. The microcontroller used is ATmega328P. This ATmega328P as a data processor, has a 10-bit ADC with a reference voltage of 5V. The 0 – 5V voltage from the pH sensor enters the analog pin. Then it is converted to digital by the Analog to Digital Converter (ADC) process to be scaled to a digital value of 0 – 1023. In collecting data from the pH sensor, variations in the pH levels are set to produce a range with a scale between 1-14 with the help of standard measuring instruments. For each variation in the pH level, a thousand data were taken and then averaged to calculate the standard deviation in linear regression.

J. Research Process Analysis

Figure 3 shows, the research process on sensor measurement results divided into two part. The first is pre-process, which includes sensor calibration with linear regression to find the digital value multiplier formula, then converted into a pH value. The calibrated microcontroller is then used to collect data from water with a certain pH level. The second is analyzing or managing data that has been measured or retrieved by sensors to enter a diagram representing data that has been processed. The research results are included in the analysis and discussion process, describing the data management process and determining the sensor's quality from the data result.

The study used primary data from the pH sensor with a data logger system using the ATmega328p microcontroller circuit. A search set of linear regression variables is performed to determine the calculation constants. After getting what is needed, the new ATmega328p can be combined with a pH sensor for the reading process in collecting analytical data. The data analysis in this study amounted to 5 sets, each amounting to 1000 data. The main variables used are the pH value (X), analog read (Y), measuring instrument results, and pH sensor readings. The ANOVA test method is applied to the differences between the 5 data sets, and the sensor results in the analyzed measurements are discussed to get conclusions in the final section.

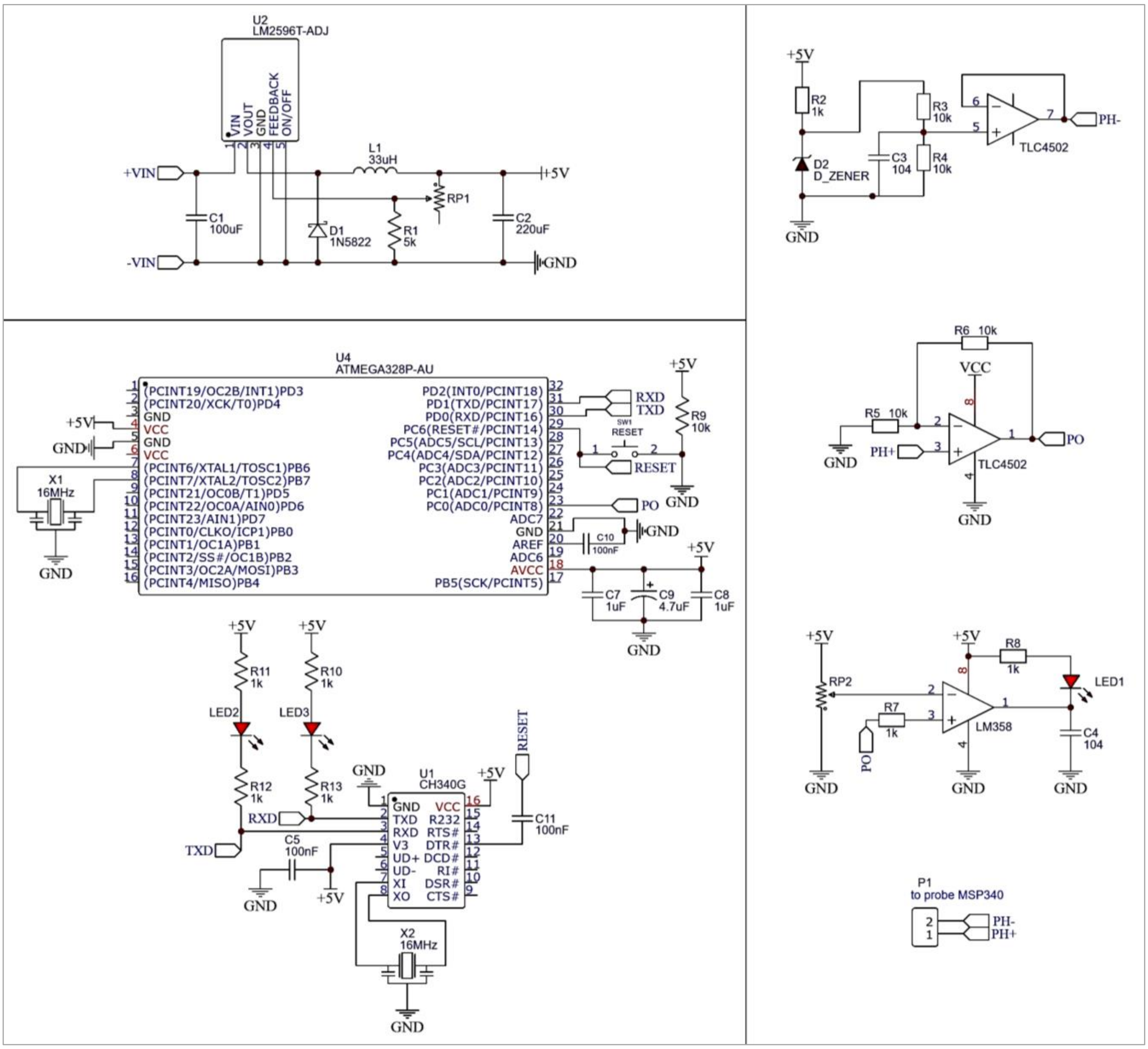


Figure 2. Schematic of pH measurement circuit

All research processes summarized in a flowchart have represented activities from the beginning to the end of the research. Where the final result shows an error in ANOVA chart and confidence level set with 95%.

III. Results and Discussion

A. Result

Data was collected using averaging filtering technique to minimize the noise effect of the ADC reading from the ATmega328P microcontroller. Then the process of reading the linear point from each measurement result is carried out as calibration. The calibration process is carried out by comparing the ADC readings with the pH measurement values using different sensor instruments, and the pH accuracy is ± 0.2.

Table 1. Linear Regression Results

β_0	β_1	r^2
23,19	0,031	0,99

Interpolation of pH data on ADC readings using linear regression in Figure 4 and the value of the coefficient of determination was 0.99, with the coefficient of each shown in Table 1. Accurate data collection from sensors is essential to get accurate sensor measurement results. In this case, data sampling takes pH data on a measuring instrument with 1000 samples. Variations in the intensity level of the pH given during the measurement process were 2.17, 5.75, 7.73, 10.36, and 13.15, with a description of the data in Table 2.

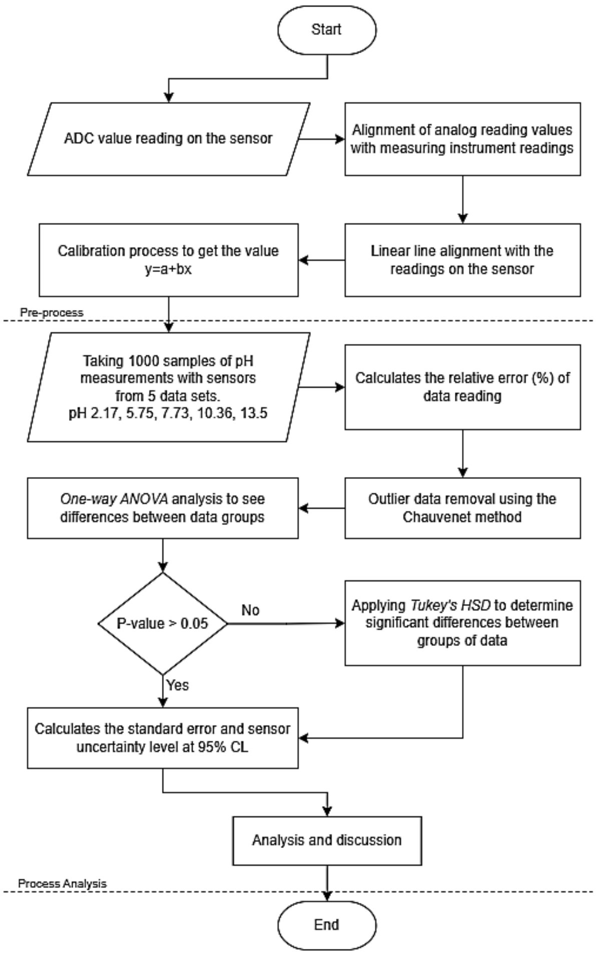


Figure 3. Research flowchart

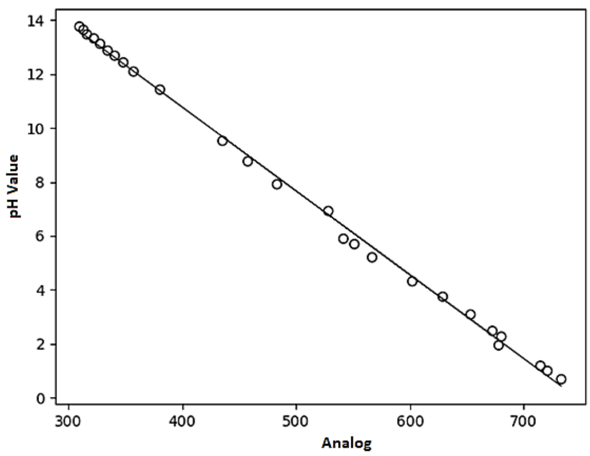


Figure 4. Linear regression of analog readings to a pH sensor

Figure 5 shows, Gaussian distribution with data on sensors. The distribution consists of a data set with a normal distribution curve with a mean representing the mea of the distribution. Each data is analyzed for distribution and variation to determine the characteristics of the sensor.

The data distribution is visualized using the boxplot in Figures 6(a) and 6(b). Figure 6(a) shows the distribution of relative error data for each pH measurement, which shows that the measurement at a pH value of 13.15 has a minor standard deviation and the smallest relative error distribution with other values. Figure 6(b) presents the relative error distribution of all pH measurements as a population.

Table 2. Result of Measurement Data

pH	2.17	5.75	7.73	10.36	13.15
Count	1000	1000	1000	1000	1000
Mean	2.177	5.761	7.726	10.372	13.152
Std	0.007	0.012	0.011	0.009	0.014
Min	2.154	5.727	7.687	10.345	13.107
25%	2.173	5.753	7.719	10.367	13.142
50%	2.177	5.762	7.726	10.372	13.152
75%	2.183	5.770	7.734	10.379	13.162
Max	2.205	5.794	7.763	10.399	13.200

The Chauvenet Criterion can eliminate data with a probability of less than 0.5%, which is considered an outlier and must be eliminated. In calculating probabilities, the first thing that needs to be done is to determine the mean and standard deviation of the data. Then, the probability of each data is calculated using the normal distribution formula. To remove the outlier, we can use equation (5).

In Table 4, the results are obtained from the calculation that the ANOVA test rejects the null hypothesis and accepts the alternate hypothesis. The value $\alpha = 0.05$ is used in determining the p-value limit of ANOVA. This result proves a significant difference between the data treatment groups on the dependent variable.

Tukey's HSD was used to calculate the difference in pH treatment between the groups. Tukey HSD compares applications by dividing them into different pairs. In this case, the total number is 10 data pair comparisons. Through Tukey's HSD test, all pairs accept the alternate hypothesis (H1) and reject the null hypothesis (H0), indicating significant differences between pairs. The probability of each pair coming from the same population is low. In Table 5 shows, results of the Tukey's HSD test.

Table 3. Acquisition of Iteration Data

	Iteration 1	Iteration 2	Iteration 3	Iteration 4
Mean	0.13	0.11	0.10	0.10
σ	0.25	0.21	0.20	0.20
Outlier	110	61	37	38

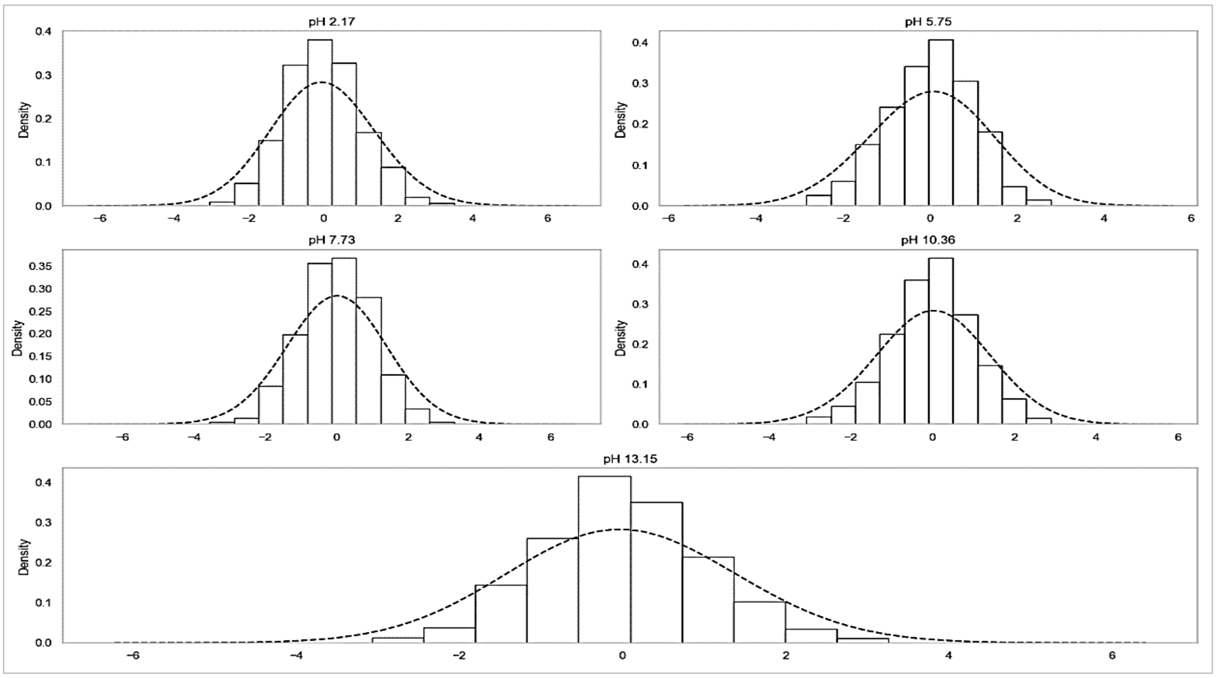


Figure 5. Standard deviation relative error (%) of each sample

Table 4. One Way ANOVA Results

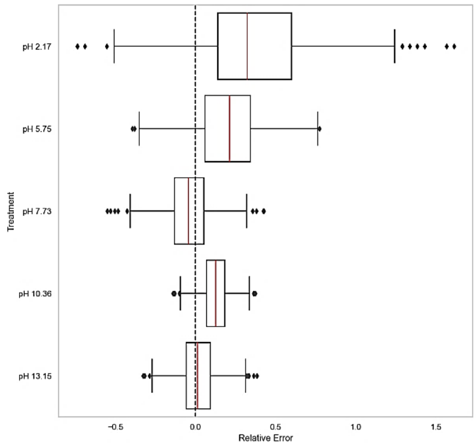
	Sum of Squares	Degrees of Freedom	F-Score	PR(>F)
Enactment	48.26	4	416.46	0
Residual	124.09	4746.00	-	-

Table 5. Tukey's HSD Results

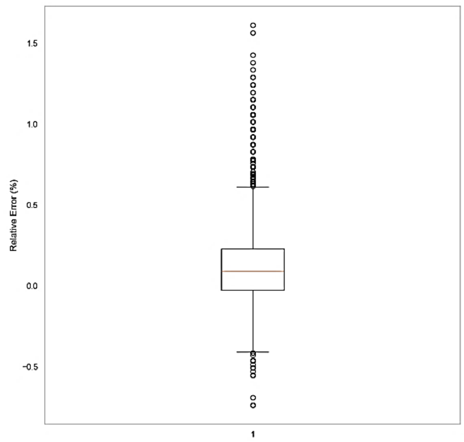
pH		Diff	L	U	V_q	V_p
2.17	5.75	0.04	0.01	0.06	6.48	0.001
2.17	7.73	0.27	0.25	0.29	49.04	0.001
2.17	10.36	0.10	0.08	0.12	18.90	0.001
2.17	13.15	0.21	0.19	0.23	38.34	0.001
5.75	7.73	0.23	0.21	0.25	45.53	0.001
5.75	10.36	0.07	0.05	0.09	13.27	0.001
5.75	13.15	0.17	0.15	0.19	34.08	0.001
7.73	10.36	0.17	0.15	0.18	32.37	0.001
7.73	13.15	0.06	0.04	0.08	11.53	0.001
10.36	13.15	0.11	0.09	0.13	20.86	0.001

*L = data with the lowest value, U = data with the highest value, V_q = Q-value from data, V_p = P-value from data.

The differential test results for each data show that the Vp value is ≤ 0.001. That indicates that there is no significant difference in the data. So the data collection process is going well. No analysis is being rejected.



(a)



(b)

Figure 6. (a) Relative error distribution for each sample, (b) Relative error population distribution

Table 6. Final Result

pH	2.17	5.75	7.73	10.36	13.15
Mean (pH)	2.18	5.76	7.73	10.37	13.15
Standard Deviation	0.01	0.01	0.01	0.01	0.01
Standard Error	0.00	0.00	0.00	0.00	0.00
Uncertainty	2.18 ± 0.2	5.76 ± 0.16	7.73 ± 0.11	10.37 ± 0.07	13.15 ± 0.09

B. *Discussion*

The variation analysis of data sensor measurements against 1000 sample population. Figure 6 shows, the standard deviation error for each sample has a density ≤ 0.45. The distribution is even, with the largest data distribution in the middle, where the deviation is close to zero. Figure 7(2) shows that the average movement error value of all measurements is between -0.1% to 0.25%. The highest error movement value ranges from -0.5% to 0.5%. Then the picture shows small dots as individual data, indicating outliers. Figure 7(1) shows the distribution of the 5 data sets in the ANOVA graph. The use of a mean value of 0 is done to see the data distribution. Distributions that still have outliers result in the error deviation having a value range that is outside of -0.75% to 1.5%. However, the distribution with a high error deviation only occurs at pH 2.17 measurements. The higher the base level in the liquid, the smaller the deviation error range. Even at pH 13.15, the error deviation value gets smaller, and the average error percentage approaches zero. Then the Cauvenet Criterion is performed to identify outliers on the ANOVA Figure 7(1) chart. Until the iteration 4 process, 246 data were identified as outliers. Figure 7 shows, the final result of ANOVA.

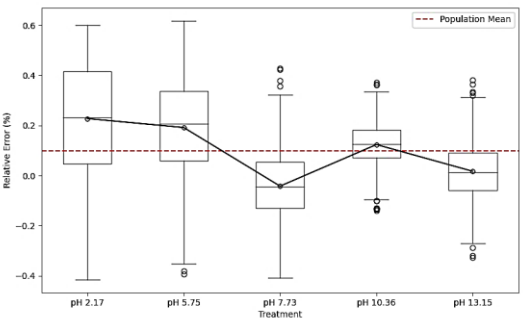
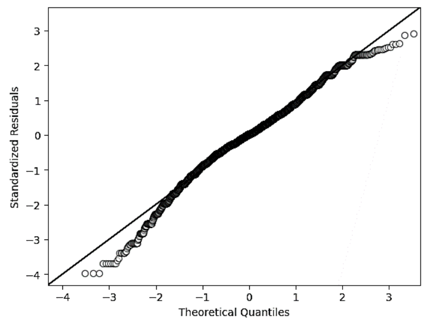


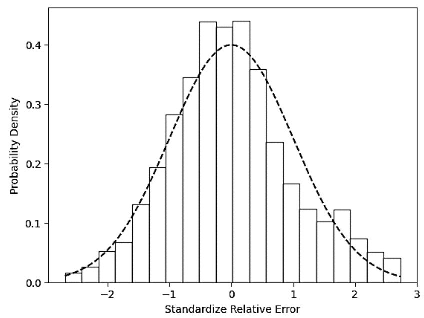
Figure 7. Final results relative error (%) of each sample

Of the 5 data sets in the graph, the median value of the entire distribution of data distribution lies at 0.1%. The distribution still has outliers, which are reduced from the previous chart. The process of removing outliers decreased the error's deviation range, ranging from -0.4% to 0.5% at pH 2.17. Just like the characteristics of the previous sensor, the error deviation range is getting smaller. The deviation value of the error ranged from -0.25% to 0.37% at pH of 13.15. This indicates that the Chauvenet criterion positively impacts the accuracy of data readings and can facilitate the ANOVA test in providing a more precise data analysis.

Then in Table 6, the final result of the uncertainty level of all measurements is less than ± 0.2 with a deviation of 0.01 for each data. This indicates that the measurement results or estimates are accurate and reliable. The more alkaline the liquid being measured, the better the degree of uncertainty of the measurement. The error value at the level of ± 0.5% proves that statistics can be used in acquiring data readings on the sensor to get results that are closer to perfect.



(a)



(b)

Figure 8. (a) Q-Q Plot, (b) Density Probability from Standard Relative Error

Analysis using a Q-Q plot was carried out to determine the distribution of data on the diagonal line. Figure 8(a) shows that the data is normally distributed through Q-Q plots and density probability graphs. With 249 wasted data, the resulting data is distributed on the quantile and residual lines from -2 to 2. This shows that all data is concentrated in that range. Figure 8(b) explains the overall distribution of measurement results better. The mean or deviation value of the whole is at zero, indicating that the data is normally distributed.

IV. CONCLUSION

Proof of self-calibration, performed without reference to the source code library, gives near-perfect results with experimental temperature conditions between 25 to 35 °C. The analysis of variance was carried out using one way ANOVA, Chauvenet Criterion, and Tukey HSD methods. It was found that all data pairs rejected the null hypothesis (H0) and accepted the alternative hypothesis (H1). After the filtering process, the standard error value for each measurement was 0.00, with the uncertainty value using the mean deviation method ranging from 0.07 to 0.02. Significant differences are seen in the relative errors in pH measurements under strongly acidic conditions. The higher the alkaline level of the liquid, the lower the measurement error deviation obtained. With the outlier filter, the average error value analysis results are less than + 0.5%. The measurements carried out in this study gave even better results than reference journals, where an average error value of 5%. In the future the research process or method can be useful in various aspects such as real-time reading, data logging, high-accuracy measurements, and others.

ACKNOWLEDGEMENT

The author would like to thank Base Engineering company for funding and facilitating the research process. With this support the research process can run smoothly to produce good quality articles.

REFERENCES

[1] S. Sopiandi, M. Marzuki, and B. D. H. Setyono, "Efektivitas sistem akuaponik untuk budidaya ikan Nilai (Oreochromis niloticus) dengan sistem resirkulasi," *Indones. J. Aquac. Mediu.*, vol. 2, no. 2, pp. 206–213, 2022.

[2] P. S. Micheev, K. A. Muraviev, E. V. Rezchikova, and K. V. Selivanov, "Designing smart pulse flow meters using diversion analysis," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 2, pp. 1338–1345, 2023.

[3] A. B. Pulungan, A. M. Putra, H. Hamdani, and H. Hastuti, "Sistem kendali kekeruhan dan pH air kolam budidaya ikan Nila," *Elkha*, vol. 12, no. 2, p. 99, 2020.

[4] B. Liu, H. Zhong, and D. Hu, "Construction of network-like cross-linked cellulose aerogel films with water-responsive properties for visualization of pH changes," *Colloids Surfaces A Physicochem. Eng. Asp.*, vol. 656, 2023.

[5] S. Zhang *et al.*, "Calibration of pH and oxygen sensors applied to aquaponic system," *J. Phys. Conf. Ser.*, vol. 656, no. 1, p. 120632, 2023.

[6] G. Gao, K. Xiao, and M. Chen, "An intelligent IoT-based control and traceability system to forecast and maintain water quality in freshwater fish farms," *Comput. Electron. Agric.*, vol. 166, 2019.

[7] Y. Zhao, Y. Yu, S. Zhao, R. Zhu, J. Zhao, and G. Cui, "Highly sensitive pH sensor based on flexible polyaniline matrix for synchronal sweat monitoring," *Microchem. J.*, p. 108092, 2022.

[8] D. Novianto, I. Setiyowati, and R. Ismawati, "Kalibrasi Sensor pH dengan Arduino Nano Sebagai Mikro Kontroler Menggunakan NaOH , HCl dan Aquades Sebagai Kalibrator," *J. Untidar*, vol. 3, no. 3, pp. 2–5, 2019.

[9] A. Maria, E. Suhartanto, and J. S. Fidari, "Validasi Data Curah Hujan Satelit dengan Data Stasiun Hujan di DAS Sadar, Kabupaten Mojokerto," *J. Teknol. dan Rekayasa Sumber Daya Air*, vol. 2, no. 2, pp. 367–375, 2022.

[10] G. Tancev and F. G. Toro, "Variational Bayesian calibration of low-cost gas sensor systems in air quality monitoring," *Meas. Sensors*, vol. 19, p. 100365, 2022.

[11] N. Ravishanker, Z. Chi, and D. K. Dey, *A First Course in Linear Model Theory. Boca Raton: Chapman and Hall*. Boca Raton: Chapman and Hall/CRC, 2021.

[12] J. L. Tabjula, S. Kanakambaran, S. Kalyani, P. Rajagopal, and B. Srinivasan, "Outlier analysis for defect detection using sparse sampling in guided wave structural health monitoring," *Struct. Control Heal. Monit.*, vol. 28, no. 3, 2021.

[13] V. V. Aroulanandam, Satyam, P. Sherubha, K. Lalitha, J. Hymavathi, and R. Thiagarajan, "Sensor data fusion for optimal robotic navigation using regression based on an IOT system," *Meas. Sensors*, vol. 24, 2022.

[14] H. Rappel, L. A. A. Beex, L. Noels, and S. P. A. Bordas, "Identifying elastoplastic parameters with Bayes' theorem considering output error, input error and model uncertainty," *Probabilistic Eng. Mech.*, vol. 55, pp. 28–41, 2019.

[15] A. Khaidir Hakam Gilang, "Sistem Kontrol Temperatur, pH, dan Kejernihan Air Kolam Ikan Berbasis Arduino Un," *Tek. Elektro*, vol. 02, no. 1, pp. 420–427, 2019.

[16] N. M. Oliveira, A. D. de Medeiros, M. de Lima Nogueira, V. Arthur, T. de Araújo Mastrangelo, and C. B. da Silva, "Hormetic effects of low-dose gamma rays in soybean seeds and seedlings: A detection technique using optical sensors," in *Computers and Electronics in Agriculture*, 2021, p. 187 (106251).

[17] Y. Jia, T. Wu, P. Dou, and M. Yu, "Temperature compensation strategy for ultrasonic-based measurement of oil film thickness," *Wear*, vol. 476, 2021.

[18] & I. G. A. P. R. A. Lindu P., Pratolo R., "Rancang Bangun Sistem Monitoring pada Hidroponik NFT (Nutrient Film Tehcnique) Berbasis Iot," *Spektrum*, vol. 8, no. 2, pp. 9–17, 2021.

[19] H. P. Ramadhan, C. Kartiko, and A. Prasetiadi, "Monitoring Kualitas Air Tambak Udang Dengan Metode Data Logging dan Algoritma KNN Berbasis Internet Of Things," *J. Inform.*, vol. 6, no. 1, pp. 1–16, 2021, [Online].