IOP Conf. Series: Earth and Environmental Science **1426** (2024) 012004 doi:10.1088/1755-1315/1426/1/012004

Enhancing Electrical Conductivity and pH Sensor Measurements in Precision Hydroponics: Comparative Analysis of Six Data Processing Methods

Hamdan Sulaiman[1,2*], Ahmad Anas Yusof[3] and Mohd Khairi Mohamed Nor[3]

[1] Faculty of Mechanical Technology and Engineering, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia
[2] Agricultural Technology Research Initiative Group, Faculty of Plantation and Agrotechnology, UiTM Cawangan Melaka Kampus Jasin, 77300 Merlimau, Melaka, Malaysia
[3] Robotics and Industrial Automation Research Group, Faculty of Electrical Technology and Engineering, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia

*E-mail: hamdan91821@uitm.edu.my

Abstract. Precision hydroponics requires precise and accurate Electrical Conductivity (EC) and Potential of Hydrogen (pH) measurements. The analog signals from the EC and pH sensor are converted to voltage and then into either EC or pH readings depending on the type of the sensor used. This paper presents a comparative study of six data processing methods to increase the precision and accuracy of EC and pH sensor measurements. The data processing includes six data processing methods that comprise direct (Methods 1 and 2) and selective methods (Methods 3, 4, 5, and 6). The precision of each data processing method was measured by comparing the standard deviation (SD) and coefficient of variance (CV). The accuracy of the methods was calculated using Mean Absolute Error (MAE) and Root Mean Square Errors (RMSE). Outliers were identified through the interquartile range (IQR). The results show that Method 1 (SD = 4.935 mV; CV = 7.16; MAE = 3.090 mV; RMSE = 5.290 mV) is the unreliable method, whereas Method 5 (SD = 0.777 mV; CV = 1.15; MAE = 0.610 mV; RMSE = 0.866 mV) and Method 6 (SD = 0.669 mV; CV = 0.99; MAE = 0.640 mV; RMSE = 0.883 mV) demonstrate high precision and accurate measurements. This study offers reliable data processing methods in taking EC and pH measurements. Future research and hydroponics practitioners should use our suggested data processing methods.

Keywords: Data processing, Electrical conductivity, pH, Precision hydroponics, Smart Farming

1. Introduction

Hydroponics is the practice of growing plants without soil utilizing a closed-loop irrigation system that supplies essential nutrients for the crops [1]. Electrical Conductivity (EC) and the Potential

IOP Conf. Series: Earth and Environmental Science **1426** (2024) 012004 doi:10.1088/1755-1315/1426/1/012004

of Hydrogen (pH) have been established as salient parameters for monitoring Hydroponics Nutrient Solution (HNS). The EC represents the concentration of nutrients, while the pH indicates nutrient availability [2], [3]. These parameters are controlled and monitored to ensure plants receive balanced nutrients [4]. In precision hydroponics research and practices, the control of this setting necessitates a highly accurate and precise control of EC and PH values [5]. Commonly, researchers used analog EC and pH sensors to monitor these parameters. For instance, Hartanto et al., [6] employed an EC sensor integrated with an Arduino-based system. Another research was performed by Sihombing et al., [7], in which they utilized a pH sensor integrated with an ESP32 microcontroller to develop an automatic nutrition detection system. Both sensors used by those researchers are analog sensors. The analog sensors produce an analog signal output, which is then converted to voltage measurement. Next, the voltage is converted to either EC or pH measurement depending on the type of sensor used.

Different data processing methods were used for sensor-based measurements in hydroponics systems to improve the precision and accuracy of the measurements and limit the impact of sensor fluctuations. Shabana Parveen et al., [8] used time-based averaging and calculated the mean value, measured during ten-second intervals. High-volume sampling was used by Schuler et al., [9] who averaged 1000 voltage samples to calculate one measurement. The two methods are entirely different; instead, Abd Rahman et al., [10] used a selective average, which applied time-based averaging and calculated the average of the three most recent readings using an outlier detection strategy. The existing variety of methods used proves the significance of data processing methods. However, a gap remains in comparing the precision and accuracy of different data processing methods for enhancing EC and pH measurements. Therefore, this study aimed to compare six distinct data processing methods for voltage measurements in precision hydroponics systems. Furthermore, this study seeks to identify the most reliable data processing methods for measuring voltages. This research is of great significance in developing precision hydroponics. With newer and better techniques for measuring EC and pH values, nutrient control can become more exact, thus leading to improved crops with higher yields and more efficient use of resources in hydroponic growth systems such as this one. Therefore, future research is recommended to implement the findings of our study in establishing a reliable EC and pH measurement.

2. Materials and methods

This section describes the development of the test rig and experimental setup, ensuring replicability and clarity of the study's findings. It also explains the six data processing methods and analyses used to evaluate the precision and accuracy of voltage measurements used to measure EC or pH in hydroponic systems. Furthermore, the studies were conducted by utilizing Minitab Statistical Software.

2.1 Development of test rig

Temperature significantly influences the accuracy of both EC and pH measurements. To mitigate this effect, the HNS temperature was maintained at 25°C, widely accepted as the standard reference temperature for such measurements [11], [12]. The temperature is controlled by using a temperature control system as illustrated in Figure 1. It employs Thermoelectric Peltier Modules (TPM) as heating and cooling elements, following the methodology developed by Sulaiman et al. [13].

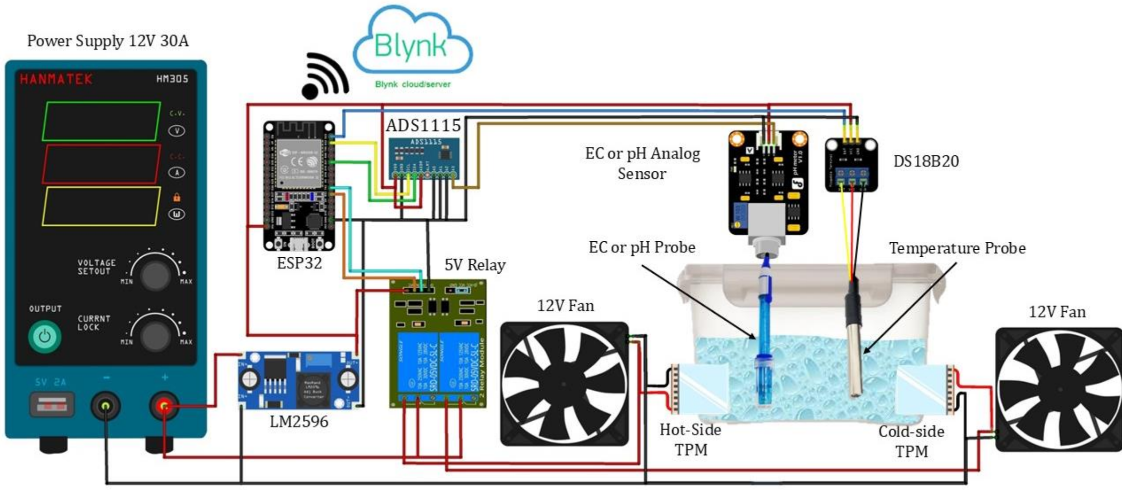

Figure 1. Schematic circuit of test rig with a temperature-controlled system.

This study employed an ESP32 microcontroller powered by a 12V 30A supply regulated at 5V. The setup incorporated two TPMs—one for cooling and one for heating—to regulate the HNS temperature precisely. Two units of 12V fans were included to enhance thermal transfer efficiency. Temperature feedback was provided by a DS18B20 sensor, which triggered the appropriate heating or cooling response via a 5V relay. To improve the reliability of analog sensor data acquisition, an ADS1115 analog-to-digital converter was employed for noise reduction [11], [14]. This experiment uses Gravity Analog Electrical Conductivity Sensor V2 (DFR0300) as the analog sensor. This comprehensive temperature control and data acquisition system ensures the reliability and reproducibility of EC and pH measurements across experiments. In addition, the test rig is integrated with the Internet of Things (IoT) through Blynk IoT as the data monitoring and acquisition platform as shown in Figure 2.

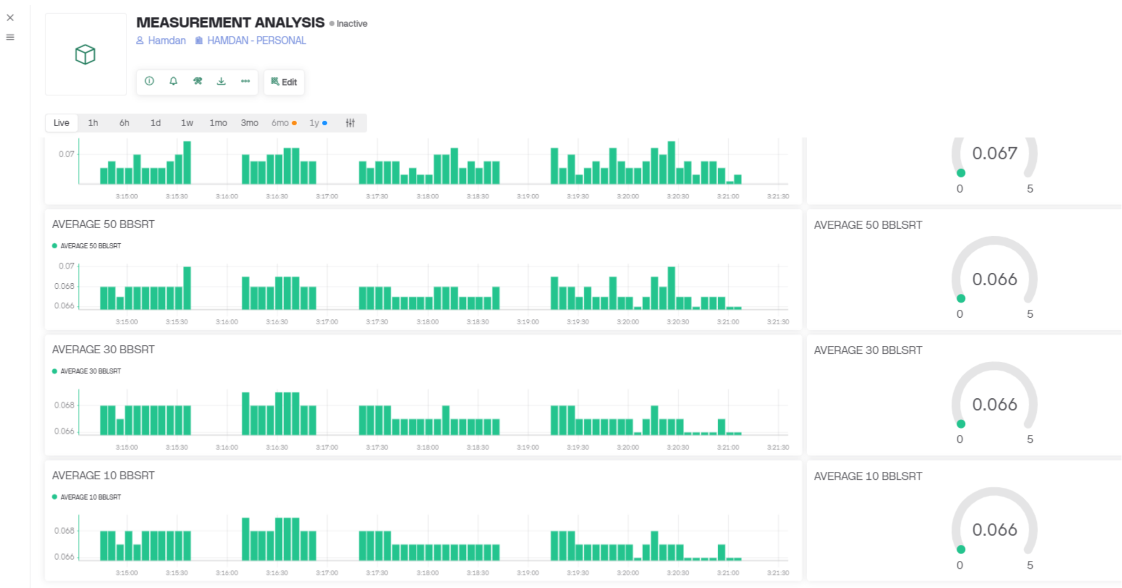

Figure 2. Data monitoring and acquisition through Blynk IoT platform.

IOP Conf. Series: Earth and Environmental Science **1426** (2024) 012004 doi:10.1088/1755-1315/1426/1/012004

2.2 Data collection and processing protocol

A rigorous protocol was implemented in the data collection and processing phase to ensure comprehensive and reliable sensor measurements in the HNS. Six distinct data processing methods were employed to analyze the collected voltage readings, as depicted in Figure 3. Method 1 involved directly taking 100 readings for 100 indices without any processing. For Method 2, each index was obtained by calculating the average of 100 readings. Methods 3 through 6 utilized a bubble-sorting algorithm to organize the data in ascending order, then averaged different subsets of the central readings.

In detail, Method 3 averaged 80 central readings, Method 4 averaged 50 central readings, Method 5 averaged 30 central readings, and Method 6 averaged ten central readings from the bubble-sorted data. Notably, a dataset of each methodology comprises 100 indices of voltage measurements obtained from each method. This systematic method allowed for a comprehensive comparison of data processing methods that ranged from raw data collection to increasingly selective averaging methods. Therefore, it enabled a thorough evaluation of their impact on measurement accuracy and precision in hydroponics systems. In addition, the EC of the HNS is regulated at 1.0 mS/cm by using Eutech CON 700 conductivity meter.

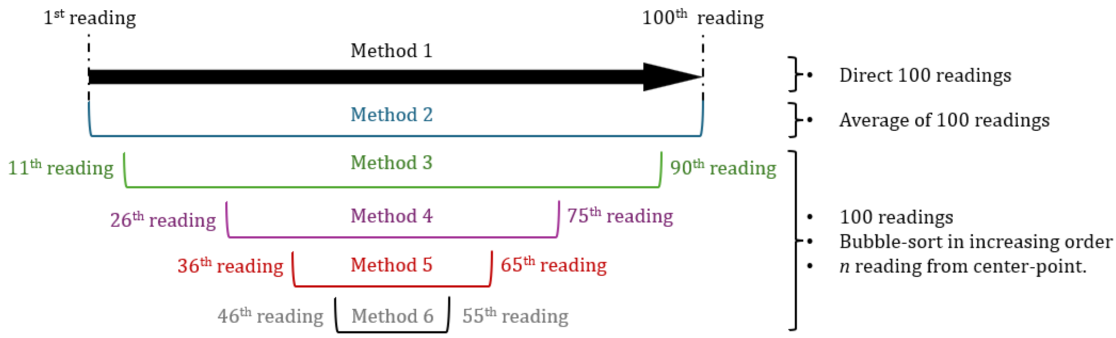

Figure 3. Data collection and processing protocol.

2.3 Precision and accuracy analysis

In this study, the precision analysis evaluates the consistency of voltage measurements, whereas the accuracy analysis measures the correctness of the measurements from the actual value. These analyses are performed through all six data processing methods. In addition, the outliers are also identified in each dataset using the interquartile range (IQR) method. This method establishes lower thresholds (LT) and upper thresholds (UT) [15]. Data points falling outside of these thresholds are considered outliers. The precision and accuracy analysis is run twice throughout the study: 1) Method 1 dataset and 2) Method 1 to 6 dataset.

In detail, the precision of each method is determined by identifying the number of outliers detected (OD) and calculating the coefficient of variance (CV) and the standard deviation (SD) [16]. This analysis compared precision across methods aiming to assess the effectiveness of data processing methods, specifically in minimizing variability and enhancing measurement reliability in hydroponic systems. On the other hand, the accuracy of each method was evaluated using the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) [17]. These calculations require an actual reference value, but this study has no reference voltage value. Therefore, we adapt the use of the median as a true value in contexts where data may be affected by anomalies or nonnormal distributions [18]. The robustness against outliers and skewed data makes the median a reliable tendency measure.

2.4 Statistical analysis
A one-way analysis of variance (ANOVA) is performed to determine if there are any statistically significant differences between the data processing methods for comparative analyses. This test checks if the variance of voltage measurement in differences between groups is large enough to conclude that there are actual differences. After conducting the ANOVA, post-hoc tests were conducted to determine whether significant differences occurred specifically between the data processing methods using Tukey's Honestly Significant Difference (HSD). Notably, the comparative nature of these statistical analyses afforded a strong structure on which to evaluate the effectiveness of each data processing method in relative terms and provide evidence-based support for how best to measure EC and pH outcomes within precision hydroponic systems.

3. Results and discussions

3.1 Outlier detection and inclusion between data processing methods
A dataset obtained from a direct voltage measurement was arranged in ascending order with the values of LT and UT of 0.0585 V and 0.0785 V, respectively. Figure 4 illustrates the scatter plot of the data set comprised of 100 direct measurements, with red markers indicating outlier measurements. 6 measurements were observed and identified as outliers as they fall outside these thresholds. In Method 1, the direct measurement of 100 readings included these six outliers. Similarly, Method 2, which averages 100 measurements, also encompassed all outliers. In contrast, Methods 3, 4, 5, and 6 effectively excluded outliers by averaging central readings, aligning with the approach by Abd Rahman et al., [10] where outliers are omitted during averaging; thus, these methods demonstrated higher precision than Methods 1 and 2.

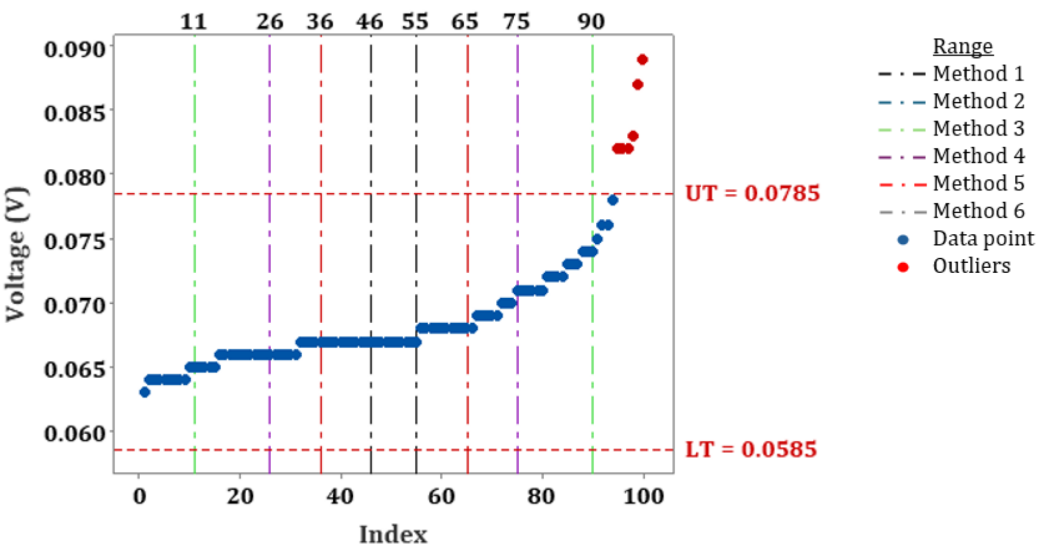

Figure 4. The scatter plotted of Method 1 measurements. The measurements were bubble sorted in ascending order.

3.2 Precision and accuracy analysis across data processing methods

Subsequently, a detailed precision analysis across datasets is performed, and the results are tabulated in Table 1 and illustrated in Figure 5. Method 1 has the lowest precision (SD = 0.004935; CV = 7.16%). Other data processing methods improved precision through the averaging process, which mitigated the impact of measurement fluctuation. Statistically, Methods 5 (SD = 0.000777; CV =1.15%) and 6 (SD=0.000669 V; CV=0.99%) offer the highest precision indicated by the lowest SD and CV values. This is consistent with the results of Dash et al., [19] who point out that defining an outlier and excluding it away can help to increase the precision of data. Our findings suggested that sorting and selective averaging methods are critical in reducing variability for EC and pH measurement to improve precision, which is indispensable for application in precision hydroponics.

Table 1. Summary of statistical parameters for voltage measurements across six methods

Method	LT (V)	Median (V)	UT (V)	OD	*n*	SD (V)	CV (%)	MAE (V)	RMSE (V)
Method 1	0.0585	0.067	0.0785	6	24	0.004935	7.16	0.00309	0.0052906
Method 2	0.0660	0.070	0.0740	0	34	0.001278	1.82	0.00095	0.0012923
Method 3	0.0650	0.069	0.0730	0	30	0.001361	1.97	0.00102	0.0013565
Method 4	0.0655	0.068	0.0695	5	47	0.000914	1.35	0.00063	0.0009220
Method 5	0.0655	0.068	0.0695	0	46	0.000777	1.15	0.00061	0.0008660
Method 6	0.0655	0.068	0.0695	0	62	0.000669	0.99	0.00064	0.0008832

The comparative analysis reveals varying accuracy levels, as indicated by the MAE and RMSE values. Method 1, which involved direct measurement of voltage readings, exhibited the highest MAE and RMSE values of 0.00309 V and 0.0052906 V, respectively. This suggests a relatively higher deviation from the assumed true value, likely due to including outliers, which increased variability. Method 2 significantly improved accuracy (MAE = 0.00095 V; RMSE = 0.0012923 V). The increase in accuracy is most likely due to the averaging process, which mitigates the impact of extreme voltage measurements. Methods 3 through 6, which employed bubble sorting and selective averaging, demonstrated further enhancements in accuracy. Method 5 achieved an impressive accuracy (MAE = 0.00061 V; RMSE = 0.0008660 V) as well as Method 6 (MAE = 0.00064 V; RMSE = 0.0008832 V). These excellent values are likely due to the exclusion of outliers and the selective averaging process, which focuses on central readings. Methods 5 and 6 are recommended for future research on EC and pH measurements in precision hydroponics. It offers more precise and accurate measurements by implementing selective averaging, reducing the influence of extreme voltage measurements.

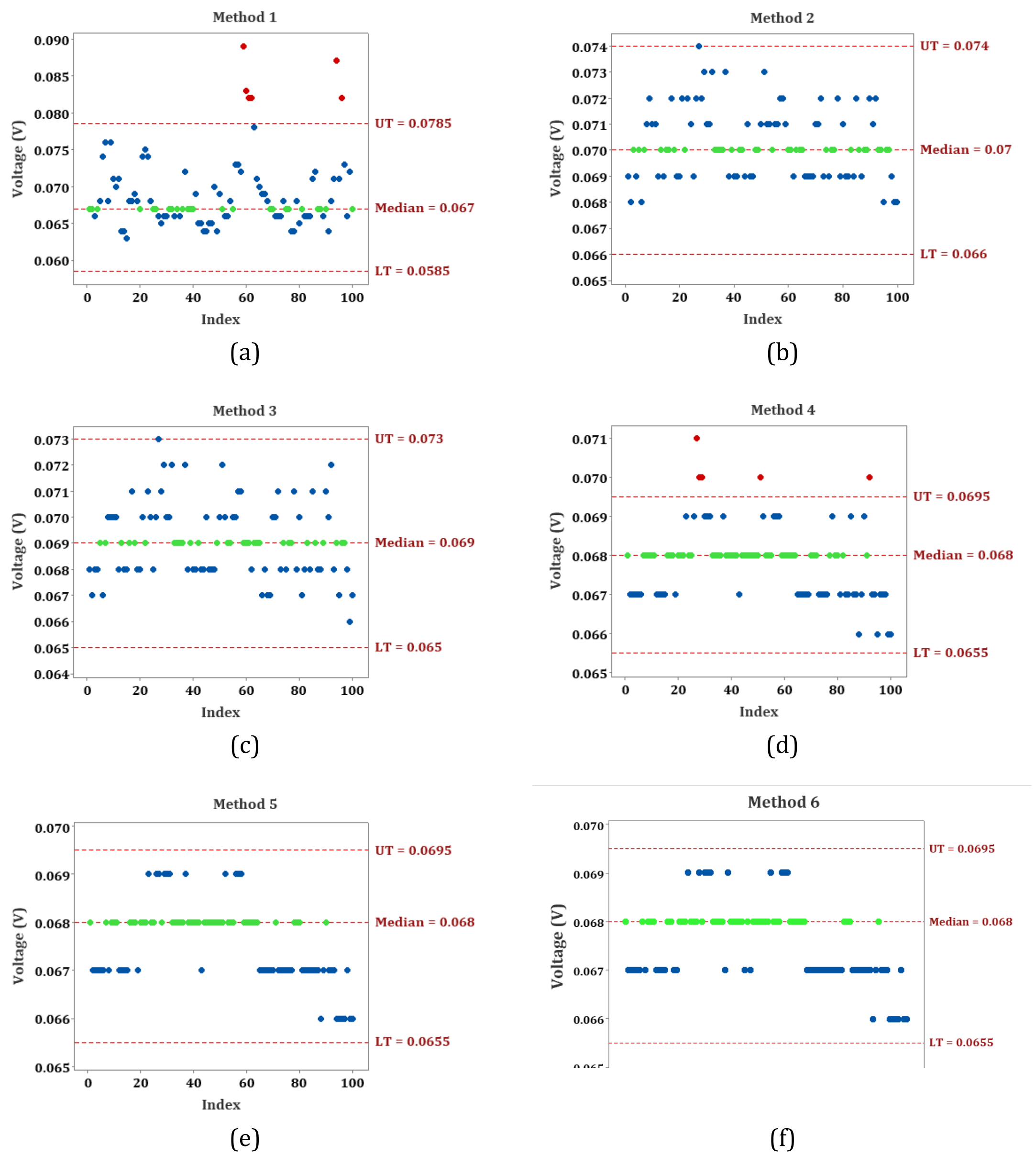

Figure 5. Distribution of voltage measurements across six datasets. This figure presents scatter plots of the data, illustrating the distribution and variability of voltage measurements. Each subplot (a-f) corresponds to a specific method: (a) Method 1, (b) Method 2, (c) Method 3, (d) Method 4, (e) Method 5, and (f) Method 6. The scatter plots provide a visual representation of the data, highlighting differences in measurement precision and accuracy of each method. The red marker indicates outlier whereas green marker represents the indices that are the same as the median value (true value).

3.3 ANOVA and Tukey's HSD

Based on the ANOVA results, as shown in Table 2, the analysis of variance was conducted to evaluate the differences in voltage measurements across six different methods. The ANOVA table indicates a significant F-value of 22.63 with a P-Value of 0.000, suggesting statistically significant differences among the means of the methods. This implies that at least one method differs significantly from the others regarding its voltage measurement precision and accuracy.

Table 2. Summary of ANOVA with a confidence level of 95%.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Method	5	0.000563	0.000113	22.63	0.000
Error	594	0.002957	0.000005		
Total	599	0.003521			

Further analysis using Tukey's HSD test reveals distinct groupings among the methods, as shown in Figure 6. Method 2, with a mean of 0.070230 V, stands out as significantly different from the others, forming its group A. Methods 3 and 1, with means of 0.069080 V and 0.068970 V, respectively, are grouped B, indicating no significant difference. Methods 4, 5, and 6, with means of 0.067850 V, 0.067610 V, and 0.067560 V, respectively, formed another distinct group C, suggesting that these methods are similar in their performance. Future research can use either Methods 4, 5, or 6 to measure voltage, as these methods do not differ significantly in precision and accuracy. However, Method 4 is rejected as it still results in outlier measurements even if it applies the selective averaging method.

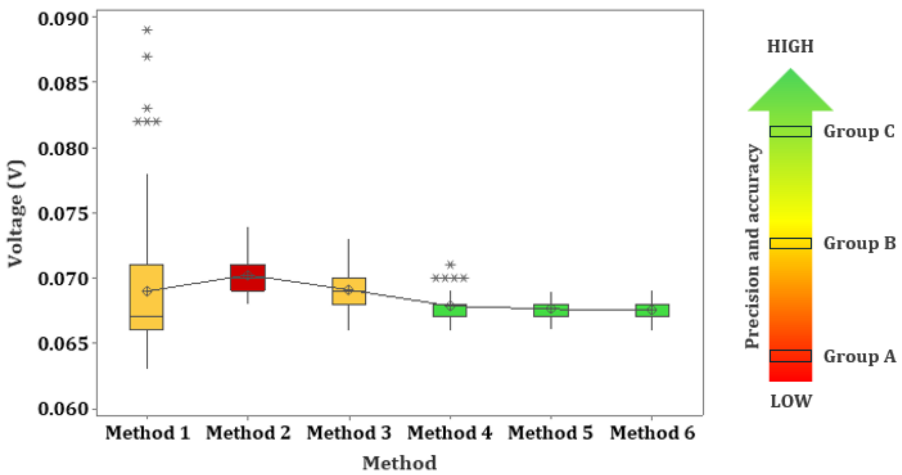

Figure 6. Box plot of Tukey's HSD test. The precision and accuracy of the methods within the same group color are not statistically different within the group, whereas different colors indicate statistical differences. Outlier measurements are represented by asterisks (*) in the data visualization

4. Conclusion

This comprehensive comparative study determined the precision and accuracy of six data processing methods for voltage measurements for EC and pH computation. Method 1 has the lowest accuracy and precision. Other methods have shown higher precision and accuracy due to the selective averaging process. In detail, Method 2 still includes the outlier in the averaging calculation. Method 4 was observed to have outliers within the dataset. Method 3 does not include the outlier in selective averaging methods but has slightly lower precision and accuracy. Therefore, by comparing the SD, CV, OD, MAE, and RMSE, Methods 5 and 6 offer the most precise and accurate data processing methods as they have the lowest value of the performance metrics. Finally, we recommend that future research and precision hydroponics applications utilize our suggested data processing methods (Methods 5 and 6), as the reliability of EC and pH measurements can help improve the quality and yield of hydroponic crops.

5. References

[1] Lenni L, Suhardiyanto H, Seminar K B and Setiawan R P A 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **542** 012067
[2] Velazquez-Gonzalez R S, Garcia-Garcia A L, Ventura-Zapata E, Barceinas-Sanchez J D O and Sosa-Savedra J C 2022 *Agriculture* **12(5)** 1–21 12050646
[3] Ali A M M 2023 101604
[4] Fathidarehnijeh E, Nadeem M, Cheema M, Thomas R, Krishnapillai M and Galagedara L 2024 *Canadian Journal of Plant Science* **104(2)** 88–102 0034
[5] Mishra P, Jimmy L, Ogunmola G A, Phu T V, Jayanthiladevi A and Latchoumi T P 2020 *J. Phys.: Conf. Ser.* **1712** 1–7 012040
[6] Hartanto A V., Kristiady F S, Atmadja W and Linggarjati J 2021 Automatic Hydroponic Nutrient Mixing for Hydroponic NFT and Fertigation *IOP Conf. Ser.: Earth Environ. Sci.* **794** 012128
[7] Sihombing P and Zarlis M 2019 *2019 4th International Conference on Informatics and Computing* 1-6 8985851
[8] Shabana Parveen M, Yashwanth D, Prasanth G, Keerthika M, Pooja R, Sandeep V and Gnanasudharsan A 2022 *2022 International Conference on Communication, Computing and Internet of Things* 1–5 9767724
[9] Schuler K, Logan J, Kim Y and Kendrick E 2015
[10] Abd Rahman M K I, Buyamin S, Abidin M S Z and Mokji M M 2018 *International Journal of Electrical and Computer Engineering* **8(3)** 1913–1919
[11] Pawar S, Tembe S and Khan S 2020 *2020 International Conference on Convergence to Digital World - Quo Vadis* 1–4 9318677
[12] Silva M G da, Soares T M, Gheyi H R, Oliveira M G B de and Santos C C dos 2020 *Engenharia Agricola* **40(6)** 674–683
[13] Sulaiman H, Yusof A A and Mohamed Nor M K 2023 *Advances in Intelligent Manufacturing and Mechatronics Lecture Notes in Electrical Engineering* **988** 285–292
[14] Wu Y M et al. 2023 *Comput Electron Agric.* 214 1–14 108266
[15] Vinutha H P, Poornima B and Sagar B M 2018 *Advances in Intelligent Systems and Computing* **701** 511–518
[16] Shechtman O 2013 *Methods of Clinical Epidemiology* 39–49
[17] Hodson T O 2022 *Geosci Model Dev* **15(14)** 5481–5487

[18] Tran K P 2017 *Qual Reliab Eng Int* **33(8)** 2437–2450
[19] Dash C S K, Behera A K, Dehuri S, Ghosh A 2023 *Decision Analytics Journal* **6** 1–8