**Abstract**

**Introduction**

* **Demand of sensor measurement**

In recent years, the demand for sensor measurements has increased significantly due to the growth in various fields such as healthcare, industrial control, and environmental monitoring (Yurish et al., 2005). Sensors play a crucial role in capturing data and converting physical properties into electrical signals, which can then be processed to provide useful information. Furthermore, with advancements in technology, there is Growing demand for high accuracy and precision, such in self-driving cars that require high accuracy and precision in analyzing decision-making (Parekh et al., 2022).

* **Study about increasing the ability of measurement (calibration, filtering and digital signal processing)**

The study and approach to increasing sensor measurement quality is an important area of research that is aimed at improving the accuracy and precision of sensor measurements. In order to achieve high quality measurements, various methods and techniques are used to improve the performance of sensors and reduce the sources of error that can affect the measurement results. These methods can include the use of high-precision sensors by improving the quality of the material (Beaver et al., 2021), advanced digital signal processing techniques by implementing Kalman Filter (Ahmed et al., 2020) and Moving Average Filter (Redhyka et al., 2016), and the implementation of calibration procedures using polynomial regression (Maciej Serda et al., 2017). By combining these methods and techniques, researchers aim to provide reliable and accurate sensor measurements that can be used for a wide range of applications, from industrial process control to scientific research.

* **Purpose of this research**

There are several approaches to filtering data using digital signal processing methods in a microcontroller to avoid noise. An approach is to use digital signal processing techniques, such as averaging or median filtering, to remove noise from the data. Additionally, techniques such as Chauvenet Criterion can be used to estimate outlier in the data. hauvenet criterion over the Kalman filter is that it is computationally less intensive, making it well suited for use in microcontroller applications where processing power is limited. Furthermore, the Chauvenet criterion does not require any prior knowledge of the system being measured, whereas the other approach such as Kalman filter relies on having a model that accurately describes the system. This can make the Chauvenet criterion more flexible and adaptable to a wider range of data sets, without the need for extensive prior knowledge.

This research aims to implementing the Chauvenet criterion algorithm for data filtering using the ATmega328p microcontroller with a voltage sensor is a useful approach for ensuring data quality. By combining the Chauvenet criterion with a voltage sensor, the resulting system is able to produce high-quality data with high precission and accuracy, providing reliable information for further analysis and decision making.

**Literature Review**

* **Polynomial Regression**

Polynomial regression is a method of estimating data by calculating the relationship between dependent and independent data through the sum of the independent variables and the intercept value, increasing to the k-th order. Polynomials have smaller errors than linear regression because of their ability to estimate non-linear data by calculating large intercept coefficients and random errors with large residuals from the predicted and actual values. (Maulud & Mohsin Abdulazeez, 2020), with the equation of polynomial regression shown in equation (1).

|  |  |  |
| --- | --- | --- |
| 𝑥𝑘 | | (1) |
|  | = degree of the polynomial. | |
|  | = intercept of orde | |
|  | = random error | |

The residual value is calculated using the square of the residual value from the actual value and the predicted value after calculating the intercept coefficient value of each k-degree shown in equation (2). Partial derivatives are used to calculate the value of the optimum intercept coefficient for each order of k by finding the location of the global minima of the equation using the equation (3).

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | (2) |
|  | = Square of residuals | | |
|  | = Summation index of data | | |
|  | = Total number of data | | |
|  | = Actual value of data | | |
|  | | (3) | |
|  | = Square of residuals | | |
|  | = Intercept of order | | |
|  | = Total number of data | | |
|  | = Actual value of data | | |

The minimum residual global point of each intercept coefficient will have a partial derivative value close to zero when obtaining the optimum intercept value; the optimum intercept value is obtained through equation (5) by changing the form of equation (4) into a linear algebraic form, and find the value of each intercept using gauss elimination.

|  |  |  |
| --- | --- | --- |
| 0 |  | (4) |
|  |  |
|  |  | (5) |

The coefficient of determination () is used to determine the proportion of the variation between the dependent variables that can be predicted by the model, the value of () is obtained after knowing several parameters such as the average mean of the actual value and the intercept value in each order, which shown in equation (6).

|  |  |  |
| --- | --- | --- |
|  |  | (6) |
|  | = Coefficient of determination | |
|  | = Mean of actual value | |

**Chauvenet Criterion**

Chauvenet's criterion is a statistical method for identifying outliers in a dataset. It works by computing the probability that a given data point is an outlier based on its deviation from the mean of the data. If the probability is below a 50% data point is considered to be an outlier and is removed from the dataset (Maples et al., 2018). The equation for calculating probability of outier is shown in equation (7).

|  |  |  |
| --- | --- | --- |
|  | | (7) |
|  | = Standard deviation | |
|  | = Mean | |
|  | = Total number of data | |
|  | = Actual value of data | |

The complementary error function is used to detect the probability that the voltage value is in a normal distribution, while the error function is used to calculate the probability of detecting an outlier value in the data distribution. The error function is computed using the z-score of each value by calculating the mean and standard deviation of the data distribution.

The limitations of microcontrollers in performing complex calculations are limited by the amount of memory and processing capabilities they have so to get error values calculated, the error function needs to be converted into power series form by limiting the order size of the series so that the computational process does not consume processing power and decreasing memory usage. The equation used for calcuating error function using power series is shown in equation (8).

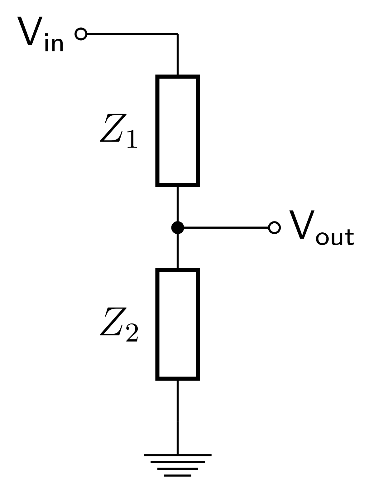
|  |  |  |
| --- | --- | --- |
|  |  | (8) |
|  |  | |
|  |  | |
| erf(z) | **=** Error function of each Z-Score value | |
| Z | **=** Z-score value | |

* **ATmega328p**

The ATmega328p is a microcontroller unit (MCU) from the AVR family of MCUs produced by Atmel (now Microchip). It is based on the RISC architecture and has a Harvard memory architecture. The ATmega328p has 32 KB of flash memory, 2 KB of SRAM, and 1 KB of EEPROM, allowing for both code and data storage. It has a clock speed of up to 20 MHz, making it capable of handling complex tasks. The ATmega328p also has a wide range of peripherals, including a 10-bit ADC, an SPI interface, and I2C communication (Barrett, 2013).

* **Voltage Divider**

A voltage divider is a electrical circuit that is used to reduce the voltage of an input signal. It consists of two resistors connected in series between the input voltage and ground. The output voltage is taken from the point between the two resistors and is proportional to the ratio of their values as shown in eqution (9). Based on the value of this resistor ratio, the measurement of high voltage values can be reduced to a voltage value that the microcontroller can read as Figure shown.



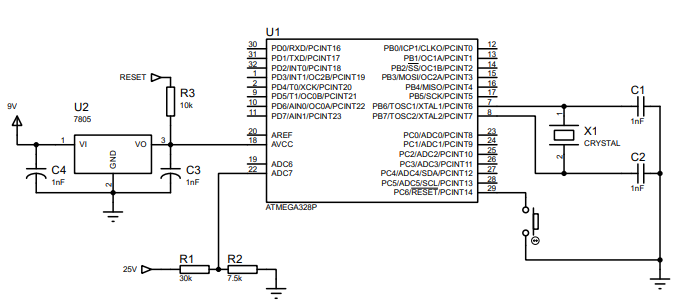
**Figure 1.** Voltage Divider Schematic

|  |  |  |
| --- | --- | --- |
|  | | (9) |
|  | = Voltage input (V) | |
|  | = Voltage output (V) | |
|  | = Resistor value one () | |
|  | = Resistor value two () | |

**Material and Method**

* **Circuit Design**

The research method begins with taking analog data using the ATmega328P microcontroller from a voltage source to carry out the calibration process using a regression polynomial. Analog data retrieval is performed using a Vref of 5v, representing 1024 data on the 10-bit Analog-to-Digital Converter (ADC) reading from the ATmega328P. Voltage data measurements calculate the potential difference between R2 and ground. Through the ratio of resistors R1 and resistor R2 with a value of R1 is 30kΩ and a value of R2 is 7.5kΩ using equation (9), obtain the reading value of voltage using ATmega328P is between 0 and 25V.

****

**Figure 2.** Schematic Design Circuit

* **Polynomial Regression**

The calibration process begins by measuring the reading of the ADC value on the ATmega328P; for the given voltage value, the ADC reading process is carried out repeatedly by providing a voltage value from 0 to 25V, to find out the relationship between the independent variables, namely the ADC value and the independent variables i.e., voltage value.

The data is then calculated using equation (5), with order k = 3, and the amount of data used for the calibration process is N = 83; after the value of each intercept is obtained, the value of is to determine the performance and variation of polynomial regression model that has been calculated using equation (6).

* **Pseudocode removing outlier**

Data collection was carried out by collecting stress data using the polynomial regression equation, which was obtained in the calibration process with a power supply source that has a voltage tolerance of 3%, this aims to determine the performance of the Chauvenet Criterion in carrying out the outlier data filtering process.

Data collection is carried out with many data values N = 50; the completion of this amount of data is caused to save on the amount of memory usage and reduces the processing process on the ATmega328P. The outlier probability of each data distribution is calculated using the average value and standard deviation with equation (7) to determine whether the probability of outlier data is more than 50%.

|  |
| --- |
| Algorithm 1 Chauvenet criterion Pseudocode |
| Input : (mean, standard deviation, and number of data)  Output :  *count\_outlier\_index* findOutlierIndex(*mean, standard deviation, N*)  While *count\_outlier\_index* not 0 Do:  For every number in data distribution Do:  *z-score* CalculateZscore(*mean, standard deviation*)  *probability* CalculateProbability(*mean, standard deviation, N*)  If *probability* < ½ Then  For every number in outlier Do:  *Remove\_data*  Else  *count\_outlier\_index* |

**Result and Discssion**

* **Polynomial Regression**

Proses kalibrasi dengan menggunakan polinomial regression di dapatkan hasil pengukuran dengan nilai sebesar 0.99 yang membuktikan bahw model yang telah di hitung mampu memiliki proporsi variasi yang kecil, menunjukkan bahwa pengukuran data tegangan terhadap pembacaan data ADC memiliki akurasi dan presisi yang tinggi, dengan nilai intercept dari setiap orde k di tunjukkan pada tabel 1.

**Tabel 1.** Polynomial regression result

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |

**Figure 3.** Polynomial Regression between ADC value and Voltage

* **Chauvenet**

Proses filterisasi menggunakan

|  |  |  |  |
| --- | --- | --- | --- |
| Number of iteration | Mean | Standard Deviation | Samples Size |
| Iteration-0 | 5.18 | 0.12 | 50 |
| Iteration-1 | 5.15 | 0.10 | 43 |
| Iteration-2 | 5.14 | 0.09 | 40 |
| Iteration-3 | 5.13 | 0.08 | 37 |
| Iteration-4 | 5.12 | 0.08 | 34 |
| Iteration-5 | 5.10 | 0.07 | 31 |
| Iteration-6 | 5.09 | 0.06 | 27 |
| Iteration-7 | 5.08 | 0.05 | 25 |
| Iteration-8 | 5.07 | 0.05 | 21 |
| Iteration-9 | 5.05 | 0.04 | 18 |
| Iteration-10 | 5.04 | 0.03 | 15 |
| Iteration-11 | 5.03 | 0.02 | 12 |
| Iteration-12 | 5.02 | 0.01 | 9 |
| Iteration-13 | 5.01 | 0.00 | 6 |
| Elapsed time: 10s 940ms | | | |

**Chart, scatter chart

Description automatically generated**

**Figure 3.** Polynomial Regression between ADC value and Voltage

**Conclusion**

**Reference**

Ahmed, M., Abdelrahem, M., Kennel, R., & Hackl, C. M. (2020). Maximum Power Point Tracking Based Model Predictive Control and Extended Kalman Filter Using Single Voltage Sensor for PV Systems. *IEEE International Symposium on Industrial Electronics*, *2020-June*, 1039–1044. https://doi.org/10.1109/ISIE45063.2020.9152256

Barrett, S. F. (2013). *Analog to Digital Conversion (ADC)*. 163–215. https://doi.org/10.1007/978-3-031-79864-1\_5

Beaver, K., Dantanarayana, A., & Minteer, S. D. (2021). Materials Approaches for Improving Electrochemical Sensor Performance. *Journal of Physical Chemistry B*, *125*(43), 11820–11834. https://doi.org/10.1021/ACS.JPCB.1C07063/ASSET/IMAGES/MEDIUM/JP1C07063\_0014.GIF

Maciej Serda, Becker, F. G., Cleary, M., Team, R. M., Holtermann, H., The, D., Agenda, N., Science, P., Sk, S. K., Hinnebusch, R., Hinnebusch A, R., Rabinovich, I., Olmert, Y., Uld, D. Q. G. L. Q., Ri, W. K. H. U., Lq, V., Frxqwu, W. K. H., Zklfk, E., Edvhg, L. v, … فاطمی, ح. (2017). Calibration of ZMPT101B voltage sensor module using polynomial regression for accurate load monitoring. *ARPN Journal of Engineering and Applied Sciences*, *12*(4), 1076–1084. https://doi.org/10.2/JQUERY.MIN.JS

Maples, M. P., Reichart, D. E., Konz, N. C., Berger, T. A., Trotter, A. S., Martin, J. R., Dutton, D. A., Paggen, M. L., Joyner, R. E., & Salemi, C. P. (2018). Robust Chauvenet Outlier Rejection. *The Astrophysical Journal Supplement Series*, *238*(1), 2. https://doi.org/10.3847/1538-4365/AAD23D

Maulud, D. H., & Mohsin Abdulazeez, A. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, *1*(4), 140–147. https://doi.org/10.38094/jastt1457

Parekh, D., Poddar, N., Rajpurkar, A., Chahal, M., Kumar, N., Joshi, G. P., & Cho, W. (2022). A Review on Autonomous Vehicles: Progress, Methods and Challenges. *Electronics 2022, Vol. 11, Page 2162*, *11*(14), 2162. https://doi.org/10.3390/ELECTRONICS11142162

Redhyka, G. G., Setiawan, D., & Soetraprawata, D. (2016). Embedded sensor fusion and moving-average filter for Inertial Measurement Unit (IMU) on the microcontroller-based stabilized platform. *Proceedings of the 2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology, ICACOMIT 2015*, 72–77. https://doi.org/10.1109/ICACOMIT.2015.7440178

Yurish, S., Kirianaki, N., & Myshkin, I. L. (2005). *World Sensors and MEMS Markets: Analysis and Trends*.