GUI-Based Data Visualisation Tool for Option C IPMVP using Jupyter Notebook

*Ezmin Abdullah**, *Nabil M. Hidayat, Nik Muhammad Akram,*
Ahmad Haziq Ahmad Hisham, Nurain Izzati Shuhaimi
School of Electrical Engineering, College of Engineering,
Universiti Teknologi MARA, 40450, Shah Alam, MALAYSIA
**ezmin@uitm.edu.my*

ABSTRACT

IPMVP is an acronym for International Performance Measurement and Verification Protocol, used for calculating energy savings by taking some external factors into account to give an equal base for a linear comparison between two periods. However, determining the significant independent variable for routine adjustment causes inconvenient repeated action of regression analysis modeling, which leads to the exhaustive calculation and human error risk. These exasperating actions led to the creation of this project, which aims to provide a user-friendly, insightful, and accurate energy-saving calculation program tool. Thus, the main objective of this project is to develop an energy-saving computational tool for option C IPMVP using web-based data analytic tools to improve efficiency and accuracy in each project. This paper used a web-based mathematical computation, Jupyter Notebook, to determine the significant independent variable for routine adjustment of the adjusted baseline. The program tool was validated using V-model software verification and validation methodology based on IEEE 1012 standards. The study can be concluded that the exhaustive calculation to determine the energy saving based on IPMVP standards can be significantly reduced with precision values at 50%, 80%, 90%, and 95% confidence levels using the proposed program which is equipped with insightful graphs and a user-friendly interface.

Keywords: *IPMVP; Energy Saving; Energy Avoidance; GUI; Jupyter Notebook*

ISSN 1823-5514, eISSN 2550-164X

https://doi.org/10.24191/jmeche.v20i3.23898

Received for review: 2022-06-13
Accepted for publication: 2023-04-14
Published: 2023-09-15

Introduction

Measurement and Verification (M&V) of energy savings are important when new equipment investments or new operation management was made to adopt an energy efficiency operating plan [1]. It is the process of planning, measuring, collecting, and analysing data to verify and report the facility's energy savings resulting from the adoption of energy conservation measures (ECMs). The International Performance Measurement and Verification Protocol (IPMVP) was developed to measure and verify efficiency investment in energy efficiency engagement [2]. This protocol is owned by Efficiency Valuation Organization (EVO) as a result of a consensus approach. The IPMVP provides a complete framework to verify energy saving after making suitable adjustments for changes in condition [3]. With this proper framework, the process of M&V becomes more detailed and reliable.

However, the exhaustive calculations have made the process of M&V become intricate which leads to human error as well as time-consuming [4]. The main disadvantage of the process is that no proper calculation tool was provided for easy application. There are several studies that suggest application tools using Graphic User Interfaces (GUI). For instance, the GUI-based using Microsoft Visual Basic were introduced in [5]-[7] to calculate energy saving/avoidance for option A, B, and, C, respectively. The program successfully displayed the numerical results of energy saving. However, the selection of significant Independent Variables (IV) for routine adjustments was selected manually by modelling linear regression of IV separately. This program was not fully automatic and require improvement if more samples of independent variables were available to be tested.

There were also papers that developed a software tool for regression analysis but not based on IPMVP. This paper created software tools called "FuReA" for linear regression analysis problem-solving algorithms. This software executed a variety of tasks, including optimal solution search with the required level of dependability, graphic depiction of modelling, etc. [8]. The article proposed the ERA software to resolve issues by providing a user-friendly tool for computer-assisted regression analysis of kinetic experiments [9]. However, this software required subscription fees for a programmer to use.

It is now well established from a variety of studies that the Jupyter Notebook becomes a powerful programming platform for data analysis and visualisation [10]. It supports a variety of tools such as graphics, algorithms, and proofs, and has become the tool of choice for data scientists [11].

Software testing and validation account for a significant portion of development costs. Validation determines if the proposed program tool fits the criteria, whereas verification determines the system's compatibility with the user's needs and expectations as recommended by the IEEE 1012 standard [12]. Klebar et al. [13] approved that V-model is one of the comprehensive testing tools in their development. The model validated their automotive OEM embedded system. Software sequence verification activities allow reducing time for development since all unpredictable things can be detected in advance. A thorough process in verification flow helped in developing effective validation flow [14] especially to a visual analytic which involves a lot of data to be processed [15].

Therefore, this project aims to minimize the burden of exhaustive calculation of energy saving/avoidance that led to human error and time-consuming. This paper proposed to develop an energy-saving/avoidance program tool for option C that comply with the IPMVP standard methodology using web-based data analytic tools within four independent variables in each project producing four linear regression models and statistical precision analysis. The proposed tool provides insightful data visualisation and accurate data analysis using linear regression models and statistical precision analysis. The remaining of the article is organized as follows. Section 2 presents a theoretical background of the IPVMP framework. Next, Section 3 describes the methodology for the proposed GUI-based and data analysis while Section 4 presents the results and discussion. Finally, implications, contributions, and future recommendations are presented in the conclusion.

Theoretical Background

IPMVP framework

IPMVP offers four types of options to determine energy-saving or energy avoidance namely options A, B, C, and D. Each option represents each type of situation, and it is impossible to generalise on the best IPMVP option. To measure overall energy saving in a facility, option C is the best option compared to options A and B which only measure retrofit isolation. Option C qualifies energy saving for the whole facility to determine saving usually after multiple ECM has been done to the building. Energy data and independent variables (IV) data must be recorded continuously throughout the length of the twelve-month baseline period and reporting period. Based on the IPMVP framework, there are three components that need to be prepared to calculate savings. The components are, adjusted baseline energy, reporting period energy, and non-routine adjustment (if any) which relies upon the following expression:

$$\begin{aligned} &\text{Saving} \\ &= (A\text{djusted Baseline Energy} - \text{Reporting Period Energy}) \\ &\pm \text{Non routine Adjustment of Baseline Energ}y \end{aligned} \tag{1}$$

where, reporting period energy is data from the data collection after retrofit and adjusted baseline energy is data calculated by first developing a mathematical model of linear regression which correlates the actual baseline energy data, Y with suitable independent data, x.

Thus, in this case, multiple numbers of IV data must undergo linear regression analysis one by one to find the best correlation variables through the value of r-squared which equals or more than 0.75 [16]. Perhaps the most serious disadvantage of this step is the repetition of work which is prone to human error. Through this mathematical model, the value of conception, m and intercept, C are used to form the adjusted baseline energy, Y_{adj} by inserting the value of each reporting period IV, x' into the model [17]. The mathematical expression for baseline energy, Y and adjusted baseline energy, Y_{adj} are shown in Equations (2) and (3), respectively.

$$\text{Actual Baseline energy,} \qquad Y = mx + C \tag{2}$$

$$A\text{djusted Baseline energy,} \qquad Y_{adj} = mx' + C \tag{3}$$

Next, since there is always uncertainty in any measurement, precision, and confidence level are used to quantify how true the value is within that margin [18]. The true value that falls within the range at a given confidence level is established as follows:

$$\text{Range} = \text{saving} \pm \text{absolute or relative precision} \tag{4}$$

Absolute precision is calculated by multiplying standard error, SE with the t from the t-table, and relative precision is divided by the saving estimation in percentage.

$$\text{Absolute precision} = t \times SE \tag{5}$$

$$\text{Relative precision} = \frac{t \times SE}{\text{Savings}} \times 100\% \tag{6}$$

where, SE is calculated using, n sample size, p number of regression variables in case it is multiple regression analysis, modelled baseline energy, Y_i and, actual baseline energy, Y. The SE equation can be expressed as follows:

$$SE = \sqrt{\frac{\sum(Y_i - Y)^2}{n - p - 1}} \tag{7}$$

The modelled baseline energy is calculated for each of the IV samples.

$$\text{Modelled Baseline energy,} \qquad Y_i = mx + C \tag{8}$$

In this project, we developed the GUI-based data visualisation for option C as a pilot project to determine energy saving in a very efficient way.

Methodology

The development of this project was carried out based on a software verification and validation process as shown in Figure 1. In the verification process, requirement specification, overall design, detail design, program specification, and coding were carried out to produce an interactive energy-saving calculation and data visualisation tool. Next, there are some processes were carried out in the validation process such as unit testing, integration testing, system testing, and user acceptance.

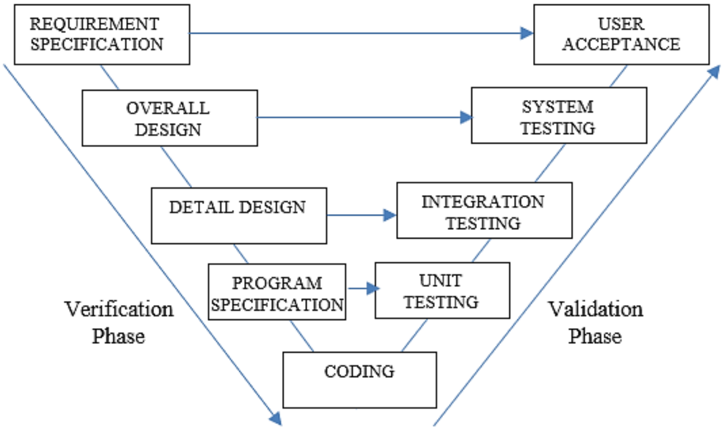

Figure 1: Software verification and validation methodology using V-Model [13]

Verification phase: developing energy-saving calculation and data visualisation tool

In this project, the main contribution relies upon the simplest steps for the users to use the tool. As the IPMVP framework shows exhausting calculations and steps, therefore, an efficient, and accurate tool is much needed. Therefore, we developed a Graphic User Interface (GUI) for users to calculate and visualise all the data only with a few clicks. Figure 2 shows the flow chart for the requirement specifications and overall design. Based on the requirement specification, the overall design flow was designed to suit the requirement. The overall design shows every function for the next step to detail the program specification.

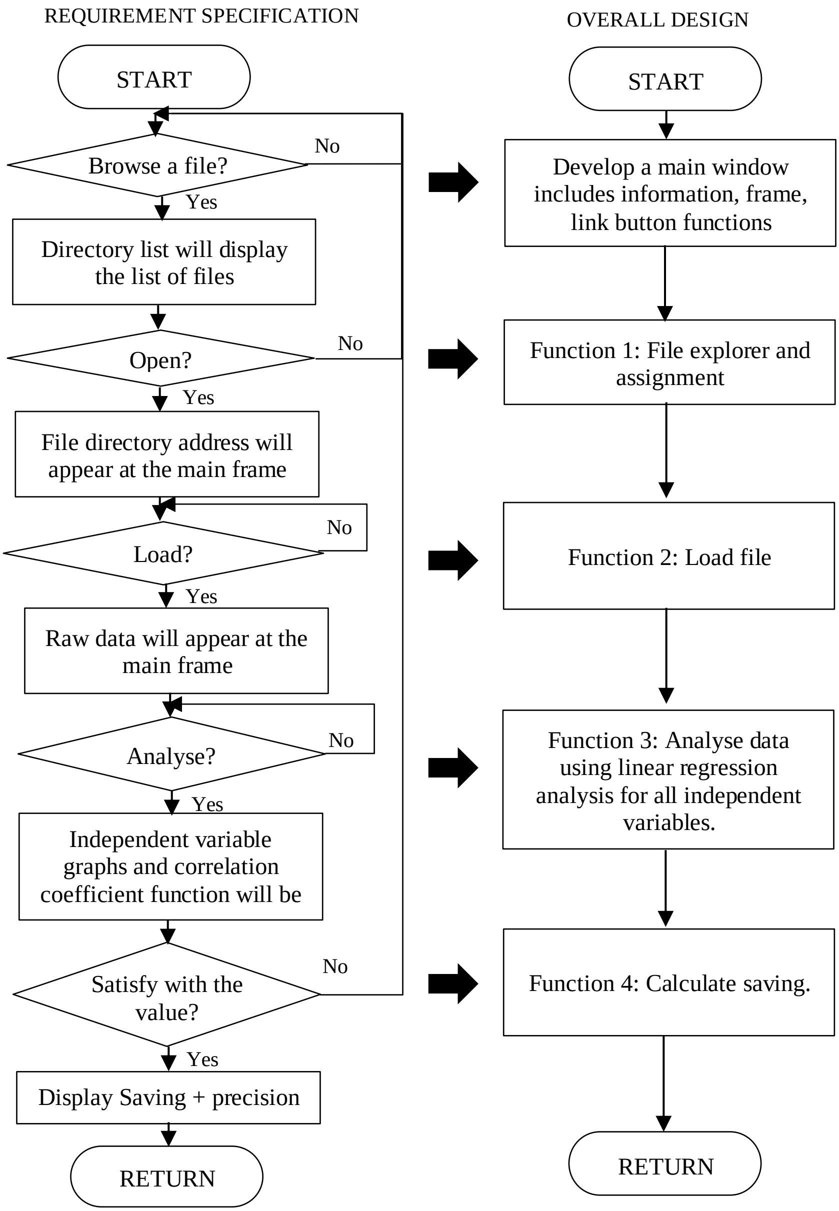

Figure 2: Flow chart of requirement specification and overall design

Table 1 shows the detailed design and program specification to list down all the necessary components to develop a full code. Components such as package tools, window interface, analysis items, and visualisation items are listed in the detailed design, followed by code construction specifications which are defined in the program specifications. Lastly, the development of the program codes can be referred to as the pseudocodes in Table 2.

Table 1: Detail design and program specification

Detail design	Program specifications
Package Tools	Import • Pandas • Matpotlib • Seaborn • NumPy • Scikit- learn • Tkinter
Interface: 1. Main Window 2. File explorer window 3. Load File window 4. Linear Regression window 5. Adjusted Baseline window 6. Energy Avoidance window	• Initialize Tkinter GUI • Setup link buttons for call functions
Load data	• Setup Dataset_path
Analysis: 1. Linear regression for 4 independent variables 2. *R*-squared 3. Adjusted baseline Precision and confidence level	• Construct scatter plot • Compute the coefficient(m), intercept(c), *R*-squared, standard error (SE)) reading. • Compute the precision based on confidence level
Visualisation: 1. Display the respective data 2. Linear Regression plots 3. Comparison between adjusted baseline with actual baseline 4. Energy Avoidance result	• Setup tree view widget for information display • Plot the linear regression • Plot the adjusted baseline • Print Linear regression line equation • Print Energy Avoidance

Table 2: Pseudocodes for the program codes development

```
Program start
Initialize Tkinter GUI:
Frame for data display
Buttons for link functions:
  Button 1: Call Function1 – new window
  Button 2: Call Function 2 – new window
  Button 3: Call Function 3 – new window
Tree view widgets for information display
Initialize Functions:
  Function 1: File explorer and assign the chosen
             file to label file Assign file location
  Function 2: Load file
              Read data from csv file
              Print the data in the frame
  Function 3: Analyze File
              Construct Scatter plots (independent
                variable1, 2, 3, 4)
              Analyze the plots -- Linear Regression
                Find out coefficient(m),
                intercept(c), R squared, standard
                error (SE))
              Repeat for all independent variables
              Determine which m and c has highest R^2.
                Highest R^2 = min (1 - R^2),
              Print scatter plots
              Print all Linear regression equation
                 and the highest R^2
              Initialize button 1: call Function 4 –
                                   new window
              Initialize button 2: close

  Function 4: Determine energy saving
                            Extract independent
                            variable data that has
                            high correlation to the
                            dependent data referring
                            to Function 3, R^2 value.
                            Calculate Adjusted
                            baseline, y' = mx + c,
```

```
                Calculate Energy saving =
                (Adjusted baseline -
                Reporting baseline)
                Calculate Absolute
                  precision (kWh)
                Calculate Relative
                  precision (%)
                Plot adjusted baseline and
                  reporting baseline
                Print Adjusted baseline
                  and reporting baseline
                  graphs
                Print Energy saving and
                  multiple precisions
                  based on multiple
                  confidence level (95%,
                  90%, 80%, 50%)
Program end
```

Validation phase: testing with sample data

In the validation phase, a set of datasets was prepared for program simulation to test the validity of the program tool.

Sample dataset

Prior to option C IPMVP, the usual baseline period is 12 months. Thus, the sample data for the Dependent Variable (DV) and IV can be considered as a full operating cycle before and after a retrofit. Table 3 shows the summarization of the type of data used for the validation of this project.

Table 3: Variable names and types

No.	Variable	Name of variable	Comment
1.	Dependent variable	Baseline energy	Energy data before a retrofit
2.	Dependent variable	Reporting period energy	Energy data after a retrofit
3.	Independent variables	Can be from multiple types of samples (eg: operating hours, number of working days, occupancy, weather, etc)	Data before a retrofit. To be used for routine adjustment.
4.	Independent variables	Can be from multiple types of samples (eg: operating hours, number of working days, occupancy, weather, etc)	Data after a retrofit. The data types must be the same as before the retrofit.

Data pre-processing

Once the sample data were collected, it was necessary to arrange the data in a specific format to allow the data to be saved in a table-structured format. In this project, CSV file type (*.csv) was used for an easier import to another storage database regardless of the specific software.

The first row of the table content was allocated for the labels. The columns are filled with months, dependent variables (DV) of the actual baseline and reporting baseline, and independent variables (IV) before and after retrofitting. Next, it is important to make sure the sample size of each variable must be 12 by following the one-year period of one cycle trend. In this example, the data was uploaded from January to December for each variable as shown in Figure 3.

	A	B	D	E	F	G	H	J	K	L	M
1	Month	Dependent	Independent1	Independent2	Independent3	Independent4	Dependent0	Independent01	Independent02	Independent03	Independent04
2	January	4850022	5271	21	100002	94731	3966405	3095	21	100800	97705
3	February	3838044	5775	15	71430	65655	5967247	3690	18	86400	82710
4	March	5578406	4990	22	104764	99774	6108354	2846	22	105600	102754
5	April	5959856	4670	20	95240	90570	5434476	2789	22	105600	102811
6	May	5021398	5038	20	95240	90202	4620803	2789	21	100800	98011
7	June	4444060	6654	20	95240	88586	5139372	3882	19	91200	87318
8	July	4863940	4665	22	104764	100099	4399953	1737	22	105600	103863
9	August	4395962	6129	21	100002	93873	3966405	3095	21	100800	97705
10	Septembe	4964390	3647	19	90478	86831	5967247	3690	18	86400	82710
11	October	6013058	4164	23	109526	105362	6108354	2846	22	105600	102754
12	Novembe	5231026	5649	20	95240	89591	5434476	2789	22	105600	102811
13	December	5212102	6060	20	95240	89180	4620803	2789	19	91200	88411

Figure 3: Data on energy consumption and four independent variables format in a CSV file

Significant independent variable

All Independent Variables (IV) were simulated using linear regression analysis to determine which IV has the best correlation to the DV. The correlation can be determined through the value of the *r*-squared generated for each IV. The *r*-squared indicates the level of the variance proportion to the DV and the acceptable benchmark defined by statistical terms is equal to or above 0.75. Then, the best-correlated IV was selected for the next step to calculate savings.

Routine adjustment

Using information extracted from the regression line of the best-correlated IV such as conception, *m*, and intercept, *C*, a routine adjustment can be made to account for differences between the baseline period and reporting period's condition. The routine adjustment is generated using Equation (3) and then was named adjusted baseline energy.

Energy saving precision

Four confidence levels (95%, 90%, 80%, and 50%) were set to quantify how true the value is within that margin. The precision for each confidence level stated above was generated using Equations (5), (6), (7), and (8).

Program tool validation

The program tool validation process was carried out based on the verification and validation using V-model [13]. The components such as window interface, analysis, and visualisation must follow and fulfill the verification design as referred to in Figure 1. Furthermore, all numerical results in this program must be corrected. The checklist for validation is summarized in Table 4.

Table 4: Validation checklist

<table>
<tr><th>Item</th><th>Unit testing</th><th>Integration testing</th><th>System testing</th><th>User testing</th></tr>
<tr><td colspan="5">Interface:</td></tr>
<tr><td>Main Window
File Explorer window
Load file window
Linear regression window
Adjusted baseline window</td><td>Every window must show all related features such as frames, buttons, and text</td><td>The content displayed in the window interface must be true and same with calculations and plots</td><td colspan="2">The flow of process must follow the overall design flow</td></tr>
<tr><td colspan="5">Load data</td></tr>
<tr><td>CSV file</td><td colspan="4">Call for CSV file must be fulfilled</td></tr>
<tr><td colspan="5">Analysis</td></tr>
<tr><td>Scatter Plot
R-squared
Adjusted baseline</td><td colspan="3">Every calculation is correct. Must compare with manual calculation.</td><td>All components reflect the IPMVP standard</td></tr>
<tr><td colspan="5">Visualisation</td></tr>
<tr><td>Data display
Graph comparison
Energy avoidance</td><td colspan="4">Every plot, graphs and display are correct. Compare with manual calculation</td></tr>
</table>

Results and Discussion

The data visualisation using this proposed tool is displayed using only 4 click functions. For user experience, first, we will discuss the GUI-based data visualisation and followed by data analysis in the next sub-section.

GUI based data visualisation

An interface of the GUI is displayed as shown in Figure 4. Three buttons are located on the main window and the flow of instructions is written for users to apply. Firstly, the user will upload their CSV file by clicking the “Browse A File” button and a “Select A File” window will appear as shown in Figure 5. As this GUI is running with a data analytic application, it is convenient to use CSV files for exchanging data between different applications [19]. Thus, in the “Select A File” window, only CSV files are shown in the user directory list. Users can select a CSV file to be analysed by highlighting the intended file and then clicking open.

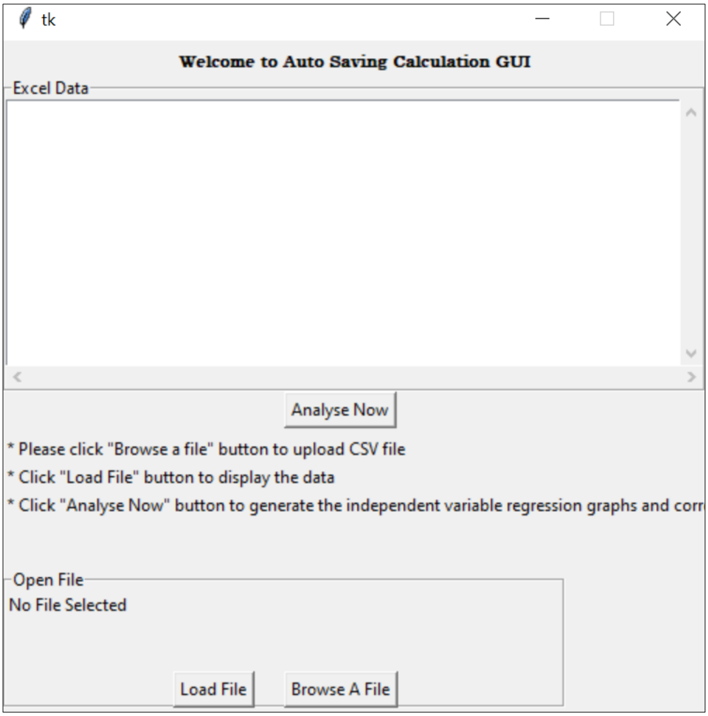

Figure 4: Main window of the GUI

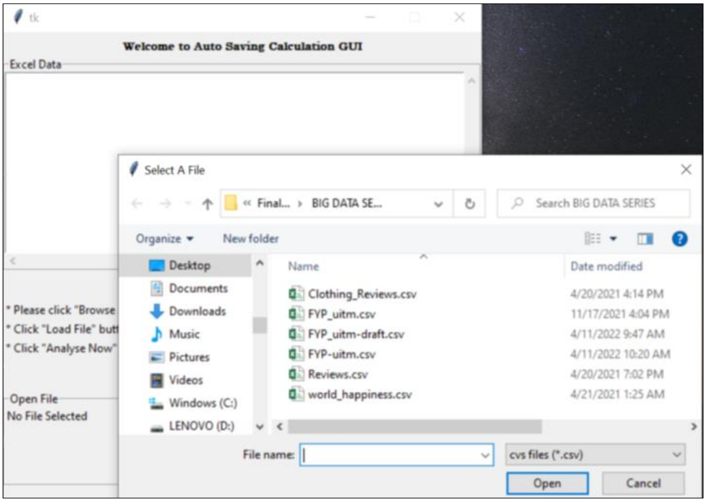

Figure 5: Window to browse and upload file

To make sure the selected file is properly uploaded to the program; the location of the file is displayed in the "Open File" frame at the bottom part of the GUI as shown in Figure 6. Next, the data can be observed by clicking the "Load File" button and the whole data from the CSV file will appear in the "Excel Data" frame. Users can scroll right or left to see other columns. In this project, the data that has been used to simulate this program was a sample from Universiti Teknologi MARA for two consecutive years.

The data on energy consumption were collected from the year 2018 and year 2019 and there were 4 IV were included in the data. Due to confidentiality reasons, all of these IVs were named independent 1, 2, 3, and 4, respectively. The dependent variable (DV) for this analysis was the Energy consumption data for 2018 (baseline energy consumption) and 2019 (reporting energy consumption). In the next sub-section, the data analysis of the data will be discussed in detail.

Data analysis

In this sub-section, data-driven energy analysis can be visualised using our proposed program. The correlation between multiple IVs can be obtained simultaneously with only one click on the "Analyse Now" button. This process reduces the risk of human error analysis and exhausting calculation steps. Information about the regression lines is depicted in the form of equations and *r*-squared for four IVs as shown in Figure 7. Furthermore, there are four graphs of linear regression lines are also shown in Figure 8 can be used as a fast method of visually depicting the relationship between the IV, x and the DV, Y.

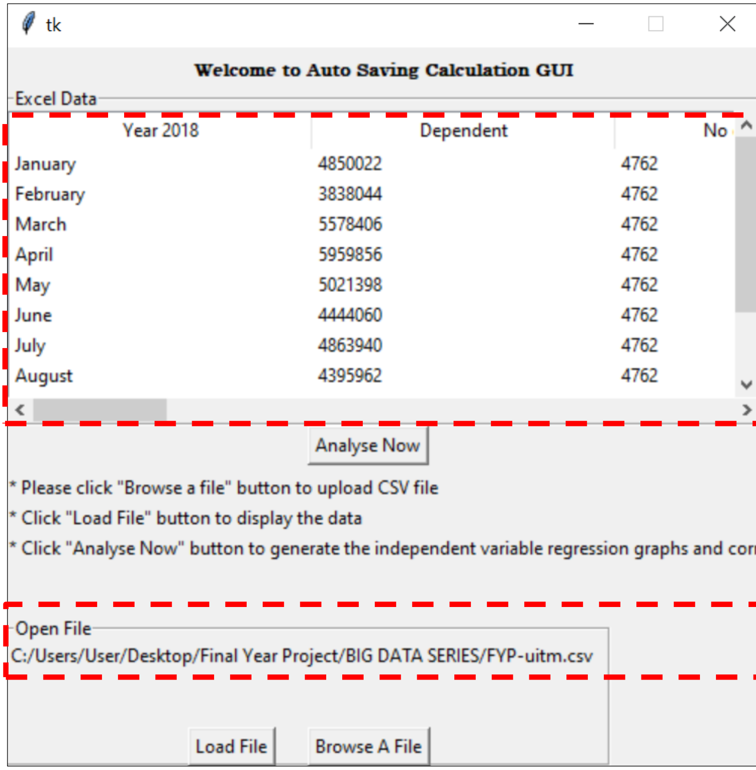

Figure 6: The data displayed on the main window

The information displayed in Figure 7 is not only important for the calculation of saving later on but also can be used to give some insights for proper actions for future energy saving measures (ESM) [20]. For instance, IV 1 shows the lowest correlation with the *r*-squared=0.2982 which can be interpreted as a very low correlation and has no significant effect on future ESM [21]. On the other hand, the best correlation shown in this simulation was the IV 4 with the *r*-squared=0.4466. With this information, users can choose whether they want to proceed with the calculation of energy saving or to close the program as the best correlation was not satisfied as it should be, which is, *r*-squared must be equal to or above 0.75. In some cases, the limitation of data collection makes the energy practitioners use the best data that they have [22]. Therefore, in this simulation, we proceeded to generate the value of energy-saving using IV 4 for validation purposes even though the sample data is inadequate.

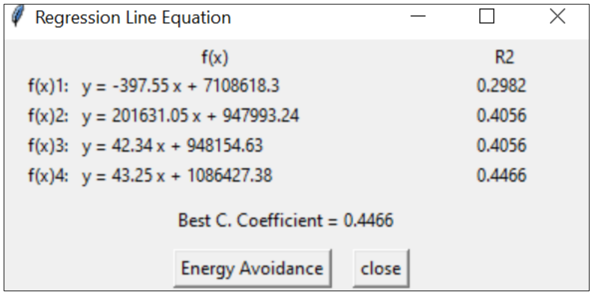

Figure 7: The display of regression lines equations and *r*-squared of the graphs

To obtain the energy avoidance/saving calculation, users can click the "energy avoidance" button and two windows will appear for the adjusted baseline graphs and energy-saving value with the relative precision according to the range of confidence levels as shown in Figure 9 and Figure 10, respectively. Referring to the graph in Figure 9, we can observe the trend of the one-year consumption for future actions as well as identify abnormal energy consumption [23]. The fluctuation of the lines was related to some events conducted within that year [24]. For instance, semester breaks were reported to be from the end of December to January 2019 and from the end of July to August for the reporting period. Hence, the low consumption during that time of the year. This graph can give some insights to the management for further investigation and evaluation of management and planning.

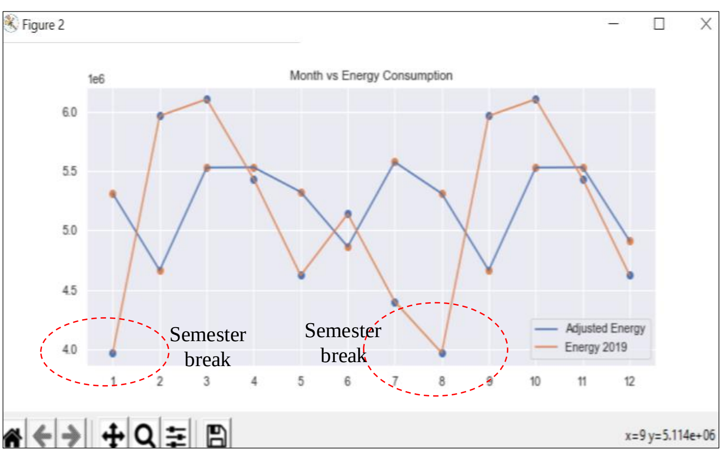

Figure 9: The comparison between the adjusted energy and actual energy consumption in 2019

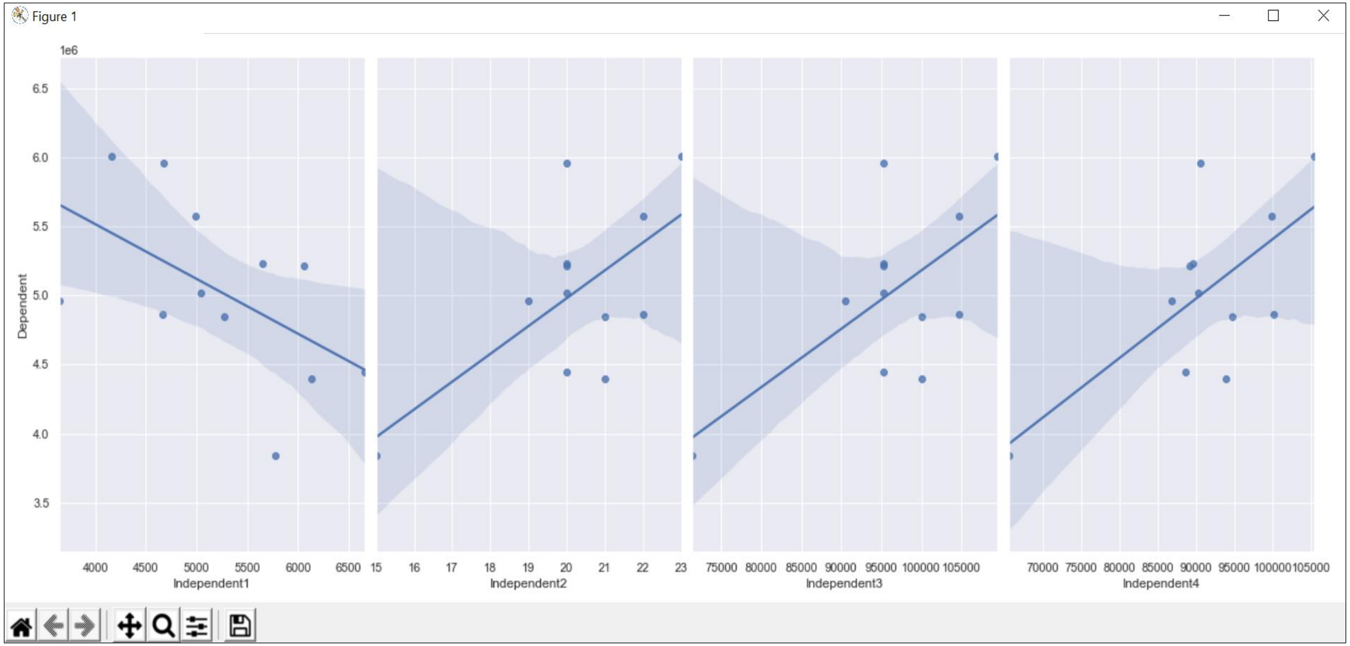

Figure 8: The graph of linear regression generated from the four independent variables

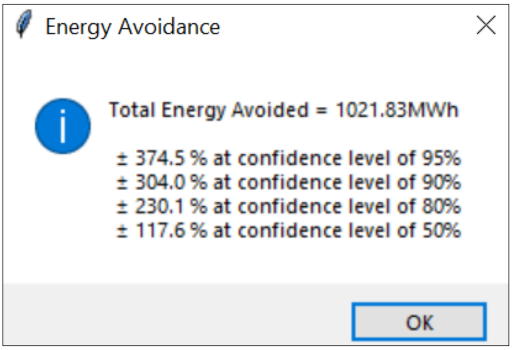

Figure 10: The total of energy avoided with precision value

Manual calculation

The numerical results in this program such as adjusted baseline, total energy avoidance, and precision analysis were compared with manual calculation as shown in Table 5, Table 6, and Equations (13), (15), (17), and (19). Lastly, the validation check of this program tool is summarized in Table 7. All items listed in Table 7 are successfully fulfilled.

Table 5: Manual calculation of adjusted baseline and energy avoided

Actual energy data			Adjusted baseline data			Energy savings
2019	Energy 2019	Actual staff presence 2019	Factors: Sensitivity 43.2451x	Factors: Constant 1086878.45	Adjusted energy baseline (kWh)	Energy saving avoidance (kWh)
Jan	3966405	97705	4225262.496	1086878.45	5312.136475	1345.731
Feb	5967247	82710	3576802.221	1086878.45	4663.676887	-1303.570
Mar	6108354	102754	4443607.005	1086878.45	5530.480754	-577.873
Apr	5434476	102811	4446071.976	1086878.45	5532.945722	98.470
May	4620803	98011	4238495.496	1086878.45	5325.369462	704.566
Jun	5139372	87318	3776075.642	1086878.45	4862.950097	-276.422
Jul	4399953	103863	4491565.821	1086878.45	5578.439519	1178.487
Aug	3966405	97705	4225262.496	1086878.45	5312.136475	1345.731
Sep	5967247	82710	3576802.221	1086878.45	4663.676887	-1303.570
Oct	6108354	102754	4443607.005	1086878.45	5530.480754	-577.873
Nov	5434476	102811	4446071.976	1086878.45	5532.945722	98.470
Dec	4620803	88411	3823342.536	1086878.45	4910.216941	289.414
Total						1021.56

Table 6: Manual calculation of standard error

Month	Independent variable	Consumption, Y	Modelled, Y_i	$(Y_i - Y)^2$
Jan	94731	4850022	5183543	111236344156
Feb	65655	3838044	3926006	7737336314
Mar	99774	5578406	5401653	31241665430
Apr	90570	5959856	5003580	914464017682
May	90202	5021398	4987664	1137990852
Jun	88586	4444060	4917772	224402945253
Jul	100099	4863940	5415709	304449172821
Aug	93873	4395962	5146435	563209168379
Sept	86831	4964390	4841868	15011608628
Oct	105362	6013058	5643334	136695924910
Nov	89591	5231026	4961238	72785494799
Dec	89180	5212102	4943462	72167245434
		Total		2454538914659

$$\mathrm{SE}_{(\mathrm{monthly})} = \sqrt{\frac{\sum(Y_i - Y)^2}{n-p-1}} = \sqrt{\frac{2454538914659}{12-1-1}} = 495433 \tag{9}$$

$$\mathrm{SE}_{(\mathrm{annually})} = \sqrt{12} \times 495433 = 1716230 \tag{10}$$

$$\text{Absolute precision} = t \times SE_{(annually)} \tag{11}$$

For confidence level 95%, DF=10, t=2.23,

$$\text{Absolute precision} = 2.23 \times SE_{(annually)} = 3827194 \tag{12}$$

$$\text{Relative precision} = \frac{3827194}{1021.56 \times 10^3} \times 100\% = 374.5\% \tag{13}$$

For confidence level 90%, DF=10, t=1.81,

$$\text{Absolute precision} = 1.81 \times SE_{(annually)} = 3106376.3 \tag{14}$$

$$\text{Relative precision} = \frac{3106376.3}{1021.56 \times 10^3} \times 100\% \\ = 304.0\% \tag{15}$$

For confidence level 80%, DF=10, t=1.37,

$$\text{Absolute precision} = 1.37 \times \ SE_{(annually)} \\ = 2351236 \tag{16}$$

$$\text{Relative precision} = \frac{2351236}{1021.56 \times 10^3} \times 100\% \\ = 230.1\% \tag{17}$$

For confidence level 50%, DF=10, t=0.7,

$$\text{Absolute precision} = 0.7 \times \ SE_{(annually)} \\ = 1201361 \tag{18}$$

$$\text{Relative precision} = \frac{1201361}{1021.56 \times 10^3} \times 100\% \\ = 117.6\% \tag{19}$$

Table 7: Validation checklist

Item	Unit testing	Integration testing	System testing	User testing
Interface:				
Main Window	√	√	√	√
File Explorer window	√	√	√	√
Load file window	√	√	√	√
Linear regression window	√	√	√	√
Adjusted baseline window	√	√	√	√
Load data				
CSV file	√	√	√	√
Analysis				
Scatter Plot	√	√	√	√
R-squared	√	√	√	√
Adjusted baseline	√	√	√	√
Visualisation				
Data display	√	√	√	√
Graph comparison	√	√	√	√
Energy avoidance	√	√	√	√

Conclusion

This project was undertaken to design a program of GUI-based data visualisation in determining the significant independent variables (IV) and energy-saving calculation that is compliant with option C IPMVP. This work contributes to existing knowledge of option C IPMVP by providing an efficient program tool for the energy practitioner to obtain insightful data visualisation and accurate results using web-based mathematical computation. This program also helps to avoid human error risk and is very easy to operate. The program was validated based on software verification and validation methodology using the V-model of IEEE 1012 standard.

Contributions of Authors

The authors confirm the equal contribution in each part of this work. All authors reviewed and approved the final version of this work.

Funding

This work received no specific grant from any funding agency.

Conflict of Interests

All authors declare that they have no conflicts of interest.

Acknowledgement

We would like to express our gratitude to Universiti Teknologi MARA Facility for providing the necessary materials that enabled us to conduct this research effectively. The access to their dataset was crucial in conducting analysis and validation of the software.

References

[1] Y. Lai, S. Papadopoulos, F. Fuerst, G. Pivo, J. Sagi, and C. E. Kontokosta, "Building retrofit hurdle rates and risk aversion in energy efficiency investments", *Applied Energy*, vol. 306, pp. 1-28, 2022. doi: 10.1016/j.apenergy.2021.118048.

[2] Efficiency Valuation Organization (EVO), “International Performance Measurement & Verification Protocol: Concepts and Options for Determining Energy and Water Savings”, *EVO* 10000-1:2012, vol. 1, pp, 1-86, 2012.

[3] X. Xia and J. Zhang, “Mathematical description for the measurement and verification of energy efficiency improvement”, *Applied Energy*, vol. 111, pp. 247–256, 2013. doi: 10.1016/j.apenergy.2013.04.063.

[4] Efficiency Valuation Organization (EVO), “M&V Issues and Examples Five-Year Statutory Protocol Review”, *DOE/EE -1287 -0286*, vol. 4.0, pp. 1-53, 2015.

[5] M. David and N. Y. Dahlan, “Development of visual basic based GUI for option a energy savings of IPMVP”, *Proceedings of the 2014 IEEE 4th International Conference on System Engineering and Technology, ICSET 2014*, pp. 1–6, 2014. doi: 10.1109/ICSEngT.2014.7111788.

[6] I. Ismail, N. H. Ahmad, N. Y. Dahlan, and H. Mohamad, “Development of GUI system using web application tool of microsoft visual studio for option b energy saving IPMVP”, *International Journal of Simulation: Systems, Science and Technology*, vol. 17, no. 41, pp. 46.1-46.4, 2017. doi: 10.5013/IJSSST.a.17.41.46.

[7] C. A. Cosmas and N. Y. Dahlan, “Development of Visual Basic GUI for Option C energy saving of IPMVP”, in *IET Seminar Digest*, vol. 2014, no. CP659, pp. 1–6, 2014. doi: 10.1049/cp.2014.1508.

[8] B. Izyumov and E. Kalinina, “Software tools for regression analysis of fuzzy data,” *Proceedings of 9th Zittau Fuzzy Colloquium*, pp. 221-229, 2001.

[9] P. Zamostny and Z. Belohlav, “A software for regression analysis of kinetic data”, *Computers & Chemistry*, vol. 23, no. 5, pp. 479–485, 1999. doi: 10.1016/S0097-8485(99)00024-8.

[10] L. Quaranta, F. Calefato, and F. Lanubile, “Eliciting Best Practices for Collaboration with Computational Notebooks”, *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW1, pp. 1–41, 2022. doi: 10.1145/3512934.

[11] B. Guo, J. Nagy, and E. Sekerinski, “Universal Design of Interactive Mathematical Notebooks on Programming”, *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2 (SIGCSE 2022), March 3â•ﬁ5, Providence, RI, USA*, 2022, vol. 1, no. 1, pp. 1132–1132, 2022. doi: 10.1145/3478432.3499102.

[12] IEEE, “IEEE Standard for System, Software, and Hardware Verification and Validation”, *IEEE Std 1012-2016* (Revision of *IEEE Std 1012-2012/ Incorporates IEEE Std 1012-2016*/Cor1-2017), pp.1-260, 2017. doi: 10.1109/IEEESTD.2017.8055462.

[13] K. N. Hodel, J. Reinaldo Da Silva, L. R. Yoshioka, J. F. Justo, and M. M. D. Santos, “FAT-AES: Systematic Methodology of Functional Testing

for Automotive Embedded Software", *IEEE Access*, vol. 10, pp. 74259–74279, 2022. doi: 10.1109/ACCESS.2021.3128431.

[14] S. Naumenko et al., "Formal Methods of FPGA Project Verification Flow," in *Proceedings of the 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2021*, vol. 2, pp. 1141–1146, 2021. doi: 10.1109/IDAACS53288.2021.9660906.

[15] E. Auer, W. Luther, and B. Weyers, "Reliable visual analytics, a prerequisite for outcome assessment of engineering systems", *Acta Cybernetica*, vol. 24, no. 3, pp. 287–314, 2020. doi: 10.14232/ACTACYB.24.3.2020.3.

[16] S. Touzani, J. Granderson, D. Jump, and D. Rebello, "Evaluation of methods to assess the uncertainty in estimated energy savings", *Energy and Buildings*, vol. 193, pp. 216–225, 2019. doi: 10.1016/j.enbuild.2019.03.041.

[17] P. Therkelsen, P. Rao, D. Sholes, B. Meffert, R.Green, S. Nimbalkar, and A. Mckane, "The value of regression models in determining industrial energy savings", *ECEEE Industrial Summer Study Proceedings Industrial Efficiency, Berlin, Germany*, pp. 389–399, 2016.

[18] L. A. Damiano, "Reduction of errors in ventilation rate determinations", *ASHRAE Transactions*, vol. 116, no. PART 2, pp. 54–69, 2010.

[19] E. Oanta, A. Raicu, and B. Menabil, "Applications of the numerical methods in mechanical engineering experimental studies", *IOP Conference Series: Materials Science and Engineering*, vol. 916, no. 1, pp. 1–9, 2020. doi: 10.1088/1757-899X/916/1/012074.

[20] H. K. Khozani, E. Esmaeili, M. N. Bisheh, S. A. Ayatollahi, and M. Gilanifar, "Using Regression Analysis to Predict the Demand Function of Electricity: A Case Study", *American Journal of Engineering and Applied Sciences*, vol. 13, no. 4, pp. 759–767, 2020. doi: 10.3844/ajeassp.2020.759.767.

[21] S. Zhang, P. Sun, and E. P. Sun, "Research on energy saving of small public building envelope system", in *Energy Reports*, vol. 8, pp. 559–565, 2022. doi: 10.1016/j.egyr.2022.03.146.

[22] A. Piccinini, M. Hajdukiewicz, and M. M. Keane, "A novel reduced order model technology framework to support the estimation of the energy savings in building retrofits", *Energy and Buildings*, vol. 244, p. 110896, 2021. doi: 10.1016/j.enbuild.2021.110896.

[23] M. Gaur, S. Makonin, I. V. Bajic, and A. Majumdar, "Performance Evaluation of Techniques for Identifying Abnormal Energy Consumption in Buildings", *IEEE Access*, vol. 7, pp. 62721–62733, 2019. doi: 10.1109/ACCESS.2019.2915641.

[24] Z. Miao, Y. Zhang, S. Liu, X. Chen, and T. Baležentis, "Heterogeneous strategy and performance decomposition: Energy-economy-environment nexus in the light of natural & managerial disposability", *Environmental*

Impact Assessment Review, vol. 95, p. 106777, 2022. doi: 10.1016/j.eiar.2022.106777.

Appendices

Table A1: t-table (IPMVP, vol. 1, 2012) [2]

DF	Confidence Level				DF	Confidence Level			
	95%	90%	80%	50%		95%	90%	80%	50%
1	12.71	6.31	3.08	1.00	16	2.12	1.75	1.34	0.69
2	4.30	2.92	1.89	0.82	17	2.11	1.74	1.33	0.69
3	3.18	2.35	1.64	0.76	18	2.10	1.73	1.33	0.69
4	2.78	2.13	1.53	0.74	19	2.09	1.73	1.33	0.69
5	2.57	2.02	1.48	0.73	21	2.08	1.72	1.32	0.69
6	2.45	1.94	1.44	0.72	23	2.07	1.71	1.32	0.69
7	2.36	1.89	1.41	0.71	25	2.06	1.71	1.32	0.68
8	2.31	1.86	1.40	0.71	27	2.05	1.70	1.31	0.68
9	2.26	1.83	1.38	0.70	31	2.04	1.70	1.31	0.68
10	2.23	1.81	1.37	0.70	35	2.03	1.69	1.31	0.68
11	2.20	1.80	1.36	0.70	41	2.02	1.68	1.30	0.68
12	2.18	1.78	1.36	0.70	49	2.01	1.68	1.30	0.68
13	2.16	1.77	1.35	0.69	60	2.00	1.67	1.30	0.68
14	2.14	1.76	1.35	0.69	120	1.98	1.66	1.29	0.68
15	2.13	1.75	1.34	0.69	∞	1.96	1.64	1.28	0.67