27th INTERNATIONAL CONGRESS OF MECHANICAL ENGENEERING
COBEM 2023
ABCM Brazilian Society of Mechanical Sciences and Engineering
27th ABCM International Congress of Mechanical Engineering
December 04-08, 2023, Florianópolis-SC, Brazil

COB-2023-0524
EXPERIMENTAL DATA REDUCTION USING THE STATISTICAL COMPUTING FOR ANALYSIS OF THE BATTERY INTERCONNECTION TUBES CIRCUIT OF AN ELECTRIC VEHICLE

Guilherme Augusto Souza Plácido
Bruno Ferreira Rossanês
São Paulo State University, School of Engineering, Department of Mechanical Engineering, 15385-000, Ilha Solteira, SP, Brazil
guilherme.placido@unesp.br, bf.rossanes@unesp.br

Aluisio Viais Pantaleão
São Paulo State University, School of Engineering, Department of Mechanical Engineering, 15385-000, Ilha Solteira, SP, Brazil
aluisio.pantaleao@unesp.br

Leandro Oliveira Salviano
São Paulo State University, School of Engineering, Department of Mechanical Engineering, 15385-000, Ilha Solteira, SP, Brazil
leandro.salviano@unesp.br

Abstract. *The present work proposes the development of a data reduction methodology using descriptive and bivariate statistics for the battery interconnection ducts in the thermal management system of an electric bus operating in high - temperature environments. However, these vehicles are designed at European headquarters and may not be well-suited to the distinctive Brazilian and Latin America climate conditions. Therefore, data verification and validation are essential for operations in higher temperatures and solar radiation, which pose performance risks. The R language was used to analyze the extensive experimental data from road tests and to identify battery life-influencing parameters through Pearson's and Spearman's correlation. Preliminary estimates indicate that the presence of an insulator around the tubes, although it minimizes heat exchange through convection, can influence the heating of the coolant temperature due to solar radiation. In addition, the analysis of fluid dynamics and heat transfer data established boundary conditions for Computational Fluid Dynamics (CFD) simulations. Finally, this strategy's coherence enables the development of virtualization tools on increasingly challenging problems, replication in similar applications, and reduces testing costs, mitigating environmental impact.*

Keywords: *data reduction, electric bus, thermal management, R language, Pearson's and Spearman's correlation.*

1. INTRODUCTION

The development of technological innovations is accompanied by social and environmental demands from consumers and regulatory agencies, sustainable manufacturing processes, as well as financial returns for shareholders. In this context, one of the main challenges of current and long-term vehicle projects is to find alternative propulsion to conventional energy sources, such as oil and its derivatives, thus moving research centers and academies worldwide (Sanguesa *et al.*, 2021) and (Bufalo *et al.*, 2017).

In line with this demand, the development of electric vehicles emerges as an attractive alternative considering financial and technological aspects. However, automotive companies in Brazil face challenges due to the concentration of electrification projects in headquarters in Europe, North America, and Asia. Vehicles marketed in Latin America undergo verifications and validations to meet specific operational needs, particularly related to climatic conditions. Therefore, it is important to address operational issues, such as thermal management of battery systems.

In electric vehicles for public transportation, batteries are often installed far from the cooling system and on the vehicle's roof, connected by ducts exposed to various thermal sources, such as solar radiation, recirculation of hot air in the compartment, convective air currents, and residual heat dissipation from the batteries and chiller. This poses a challenge in designing the thermal management systems for battery packs, as the coolant temperature is crucial. Tete *et al.* (2021) provide an important compilation of various techniques that have been investigated and applied in the development of these systems and subsystems. Karimi and Li (2013) analyzed the relationship between battery thermal behavior and system design parameters. The use of computational tools that analyze flow dynamics and heat transfer data to design duct circuits that are part of the thermal management system of the electric vehicle battery pack is of utmost relevance in the highly competitive electric vehicle industry.

The reduction of experimental data is a fundamental step in various fields of science and engineering. Often, it is necessary to deal with a large volume of information during data collection, which can be difficult to analyze and interpret efficiently. Consequently, data reduction methods are used to extract relevant information in a clear and concise manner.

In this regard, the present work proposes the development of a robust methodology for reducing and statistically treating experimental data from tests on an electric bus, utilizing the open-source software R as a mathematical computational tool. Lafaye de Micheaux *et al.* (2013), Farcomeni and Greco (2016), Wickham and Grolemund (2016) and Wickham (2016) deal with several data reduction methodologies using the R software, as well as highlighting the existence of specific packages available for data processing, which implies greater agility in the data manipulation process that were applicable to this theme.

Therefore, data analysis aims to assist in making more accurate decisions regarding the construction and testing of a prototype circuit of tubes that connect the cooling system to the batteries through Computational Fluid Dynamics. In this way, it is possible to identify patterns, relationships, and behaviors of the electric vehicle in different conditions of use, aiding in decision-making to improve energy efficiency, autonomy, and vehicle performance. Finally, the development of thermal management systems and subsystems for the battery pack must be carefully evaluated, under the risk of performance drop, undesirable factors that impact the autonomy of the electric vehicle, especially in stressful operational and climate conditions.

2. EVALUATION OF THERMAL BEHAVIOR AND CORRELATION MODELS

Experimental data are made available in several separate files and are not able to be analyzed directly. As an alternative to this limitation, the R software was used due to the fact that its ability to investigate and process increasingly complex and larger data (Wickham, 2015) and (Vuong *et al.*, 2020). Hence, with the use of a coherent computational statistical tool, was possible to evaluate the heat transfer in the battery circuit tubes of the electric vehicle, for two different insulation configurations, with and without insulator, through the behavior of the inlet and output of temperatures, solar radiation and fluxes, coupled with a robust experimental technique of data reduction based on statistical-mathematical models and correlations.

Correlation analysis is an important statistical method that provides information about the relationship between two variables and can be used to investigate the correlation between variables. The strength of the correlation is determined by the correlation coefficient, which ranges from -1 to +1, in which the statistical analysis can be used to make a statement about the strength and direction of the correlation. In the present work, the index performance of two correlation models was evaluated to address challenging phenomena from the perspective of the relationship of different factors that the circuit of interconnecting tubes to the battery pack would be susceptible. These correlation models correspond to Pearson's and Spearman's correlation coefficient.

2.1 Pearson's Correlation

One of the primary applications of the bivariate normal correlation model is to examine the relationship between two variables (Neter *et al.*, 1996). Pearson's correlation model is implemented in the R language due to the fact that its ability to measure the linear relationship between variables in a large data set, making it possible to infer the presence or absence of significant correlation. Pearson's correlation model is suitable for continuous variables. However, this particular correlation model assumes a linear relationship and data normality.

The covariance between the variables is used in the calculation of this coefficient. Nevertheless, the covariance is not standardized, and can assume values from minus infinity to plus infinity. This lack of standardization makes it difficult to compare the magnitude of relationships between different variables. For this reason, the correlation coefficient, also known as the product-moment correlation, is calculated. This coefficient (r) is obtained by normalizing the covariance, using the variances of the two variables involved. The correlation coefficient is calculated according to Eq. (1).

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}}, \tag{1}$$

where x_i corresponds to each observation of the first variable, $\bar{x}$, to the average of these observations, y_i, to the observation of the second variable, and $\bar{y}$, the average of those observation. Values of r near 1, on one hand, indicate a strong positive (direct) linear association between the variables, values of r near -1, on the other hand, indicate a strong negative (indirect) linear association, whereas values of r near 0 indicate little or no linear association.

2.2 Spearman's Correlation

On the other hand, other models or statistical techniques may be more appropriate for analyzing large volumes of data. For example, in situations where the relationship between the variables is non-linear or the data distributions are non-normal, the use of non-parametric methods or other correlation models, such as the Spearman's correlation (ρ), may be considered.

Unlike the previous one, the calculation of this correlation model is based on the classification system of the data series. In this sense, the values are not used directly in the calculation, requiring the transformation of these values into rankings. Under these circumstances, the test is performed using the classifications expressed in Eq. (2).

$$\rho = \frac{1 - 6 \cdot \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}, \tag{2}$$

where d_i corresponds to the differences between the rankings of the two variables and n is the number of observations with non-missing values in the two variables. Similarly to Pearson's correlation model, the values of ρ also vary between -1 and +1, in which a value lower than zero ($\rho < 0$) indicates a negative linear correlation between the variables, values greater than zero ($\rho > 0$) indicate a positive linear correlation, and a value equal to zero ($\rho = 0$) indicates the absence of a relationship between the variables. Spearman's correlation makes no assumptions about the distribution of the data, so it is less sensitive to outliers than Pearson's correlation.

3. METHODOLOGY

For the development of the research work on the reduction and statistical treatment of experimental data, the workflow shown in Fig. 1 was drafted to guide the steps taken to reduce data from vehicle tests.

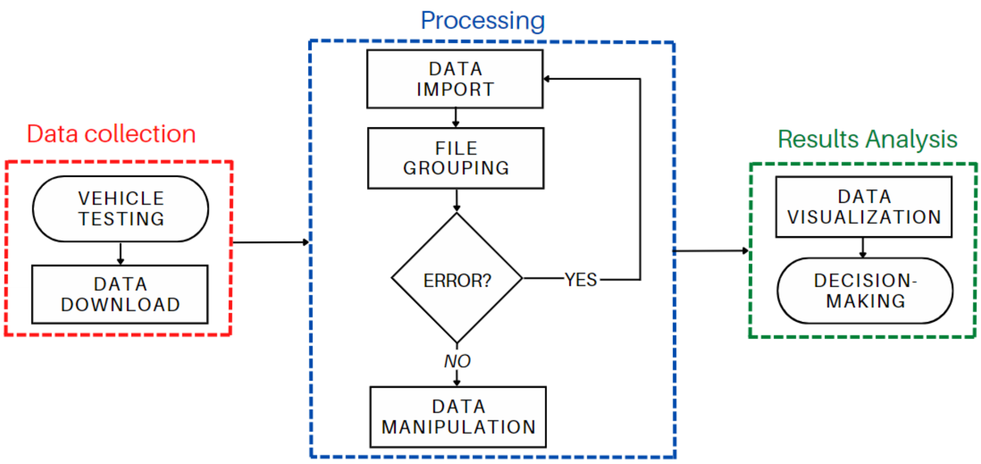

Figure 1. Workflow for data processing.

3.1 Data collection

This first stage comprises the vehicle's instrumentation, the positioning of measurement sensors, the definition of the vehicle's traffic route, and the transmission of measurement data to a folder. The data are generated by a 4G connection through the router, which communicates with the vehicle's Controller Area Network (CAN) data logger.

The measured data are available in the *.mat file extension format, which is a file format used by MATLAB software. These files, representing the start and finish of each test, were downloaded. Since each test lasts approximately two hours, this pattern generates around 90 files, with a new file created every 80 seconds. This results in a data set of approximately 16 GB and about 7000 observations for each variable. Therefore, due to the high variety, speed, and volume of data, there is a need for big data processing and the use of a computational statistical language to reduce this experimental data (Pandey and Dhoundiyal, 2015).

3.2 Processing

Since the data set is not fully structured for direct implementation in algorithms, the R software was chosen as it is a potential computational tool for statistical treatments and convenient for data analysis (Gareth *et al.*, 2013). The files were organized into folders named according to test days and periods (morning and afternoon). Because of it, methods for reading and storing the files were developed and studied (Wickham, 2015), and, after implementing the code, it became evident that this data import process requires significant computational resources to reduce processing time due to the large volume of data. The data was grouped in list format, which allows various types of objects to be stored in the computer's memory. This format enables the selection and concatenation of data for the desired parameter, even if it is spread across different files.

Finally, the data were manipulated and organized according to the TidyData pattern in R (Wickham, 2014). In this pattern, each column corresponds to a variable, each row to an observation, and each cell to a single data point. This format has the potential to simplify data visualization and make the code more understandable to a wider audience. As a result, a database was created for implementing data reduction procedures or for exporting data to *.xlsx spreadsheets, which could then be uploaded to a graphical interface. Although this step may add some time to the data reduction process, it is essential for data analysis, as well-organized data allow for quicker reading and visualization (Chang, 2018).

3.3 Results Analysis

In this step, a graphical interface was developed and executed using the flexdashboard package based on RMarkdown. It provided a clearer and more accessible way to visualize the results. This approach, aligned with the context of Business Intelligence (BI), employed graphs and tables, allowing for interactive visualizations and flexible design elements that can be explored to analyze the data more intuitively (Haymond, 2022). Similar to the previous step, the large volume of data was manipulated and loaded into the interface to generate useful information that supports and facilitates decision-making.

For this purpose, sessions were created within the graphical interface, categorizing the information into tabs that cover various aspects, from data upload to the visualization of graphs. These tabs incorporate both Pearson's and Spearman's correlations, as well as the behavior of each variable during the test (Neter *et al.*, 1996) and (Johnson *et al.*, 2002). Therefore, this setup allows for the identification of potential patterns, trends and provides valuable insights to support strategic and operational decision-making.

4. RESULTS AND DISCUSSION

4.1 Experimental Data Reduction

Data reduction plays a crucial role in scientific analysis, enabling for a clearer and more concise understanding of data sets (Farcomeni and Greco, 2016). Using statistical summary methods, data reading, and removal, the goal is to extract relevant information, eliminate redundancies and reduce the number of dimensions. Furthermore, for effective data summarizing, a correct and meticulous reading of the data is essential in the data reduction process. This entails comprehending the structure and significance of the variables, as well as considering the validity and reliability of the measured values. However, since the electric bus is equipped with numerous measurement sensors that collect data beyond the intended scope, in order to investigate the potential influence of thermal sources on the interconnecting tube circuit of the battery system, resulting in a total of 11478 parameters, it becomes imperative to reduce this data volume and focus on analyzing only the pertinent factors.

That way, an algorithm was developed to reduce experimental data. Figure 2 illustrates the main points implemented, ensuring that it does not demand excessive computational resources while effectively reducing the data sets size. The data collection step involved defining the working directory, loading the necessary packages to read the files, and storing the data in the form of a list. The processing stage includes delimiting the reading of the measurement data set, reviewing and refining it in case of any inconsistencies throughout the procedure, and creating a function to concatenate and organize the parameters. Finally, with the data loaded into the graphical interface, it is possible to visualize temperature behavior graphs during the test and draw inferences from the results obtained.

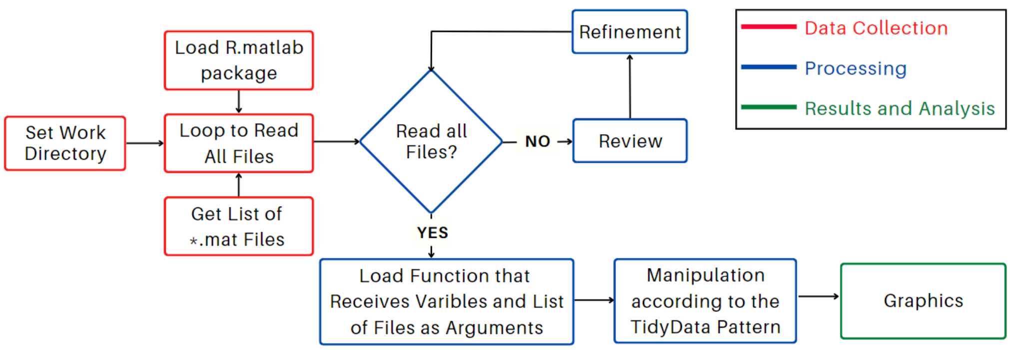

Figure 2. Diagram of the method used to elaboration the algorithm.

By appropriately organizing the data, redundant or irrelevant information can be identified and discarded for analysis. The exported data was then imported into the flexdashboard interface developed using RMarkdown, as shown in Fig. 3.

The application of the data reduction methodology presented consistent results and maintained the integrity and coherence. Careful selection of relevant variables and visualization of key parameters ensured data completeness, representativeness, and quality.

The use of specialized packages like R.Matlab, tidyverse, and dplyr optimized the data reduction process, enabling more efficient analysis and secure storage without information loss. Therefore, this confirms the consistency of the computational data reduction methodology and enhances the reliability and validity of the obtained results in data analysis.

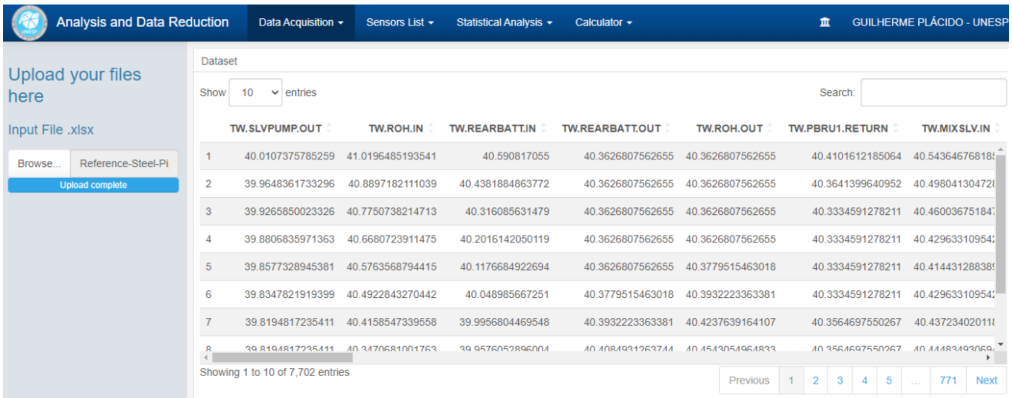

Figure 3. Data input.

4.2 Verification and Validation of Data Set

To ensure that the data divergences are only from variations in data volume characteristics, the data distribution was checked for accuracy and consistency. This involved reviewing the data for errors, outliers, or any other anomaly that might compromise the integrity of the results.

Data verification was carried out by applying statistical and mathematical techniques, such as calculating the mean, standard deviation, and coefficient of variation. These techniques of descriptive statistics aim to simplify data representation while preserving its essence. Figure 4 illustrates how this data consistency can be interactively analyzed in the graphical interface for the configuration of tubes without thermal insulator, referred to as Reference Steel Pipes.

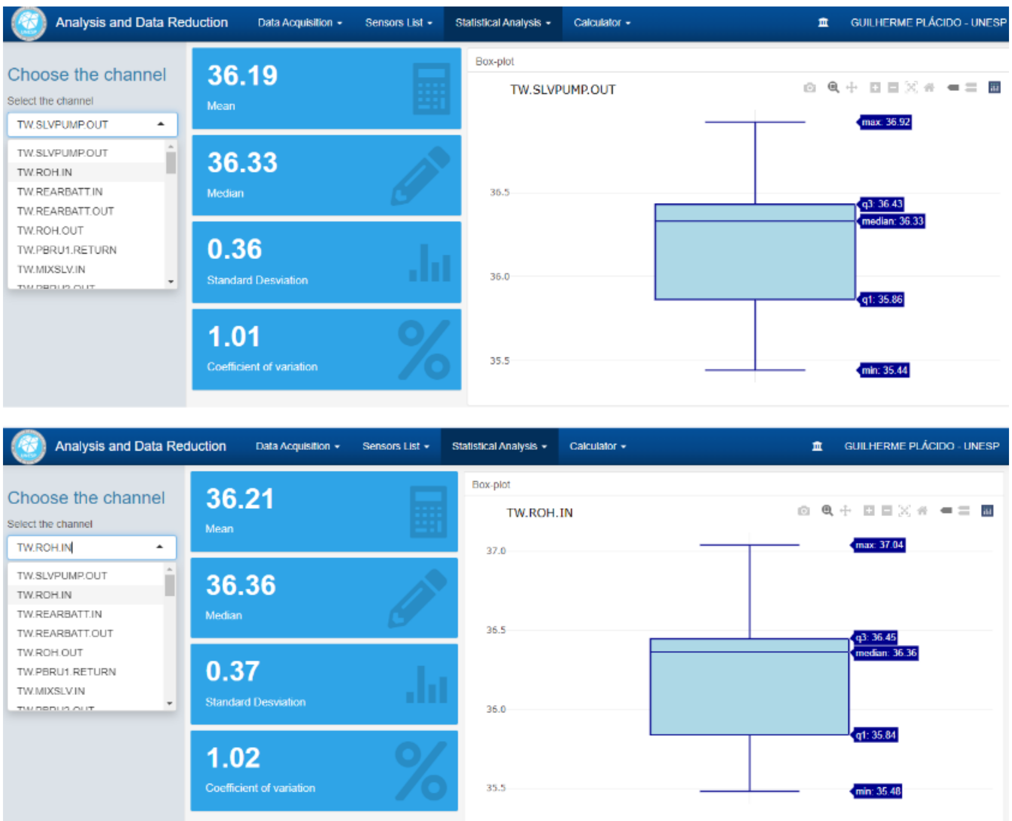

Figure 4. Interactive panel for descriptive statistical of temperatures from Reference Steel Pipes.

Thus, it is possible to observe that the analysis of the data collected through descriptive statistics and box-plots provides a comprehensive view of the data distribution and consistency. This analysis was conducted to verify the quality and reliability of the data used in this study. Descriptive statistical analysis and box-plots revealed that the data temperature have a symmetrical distribution without significant outliers. The average inlet temperature (TW.SLVPUMP.OUT) and outlet temperature (TW.ROH.IN) were 36.19 ^{o}C and 36.21 ^{o}C, respectively, with standard deviations of 0.36 and 0.37, respectively, and coefficients of variation of 1.01% and 1.02%, which indicates a low dispersion and data homogeneity.

Theoretical expectations for the studied phenomenon involve a coolant temperature at the outlet higher than in the inlet of the ducts. Since the interconnection ducts are installed on the roof of the vehicle, in a region susceptible to solar radiation, heat exchange from convection and recirculation of hot air, the dynamics of the flow and the heat transfer in this compartment can likely result in the formation of heat islands.

The results indicate that for the configuration without a thermal insulator, there is a greater exchange of heat with the external environment, and coolant temperatures at the inlet are lower than at the outlet of the ducts that interconnect the battery pack, although the difference between them was practically insignificant. Data verification through descriptive statistics and box-plots provided a basis for the continuity of this study, ensuring the reliability of data collection and representativeness of the sample used in subsequent analyses.

Analogously, for the configuration with 3 mm thick thermal insulator made of elastomer around the tubes, the verification and validation of the results was performed. Figure 5 displays the collected values for these ducts with an insulator, referred to as Thin Insulation.

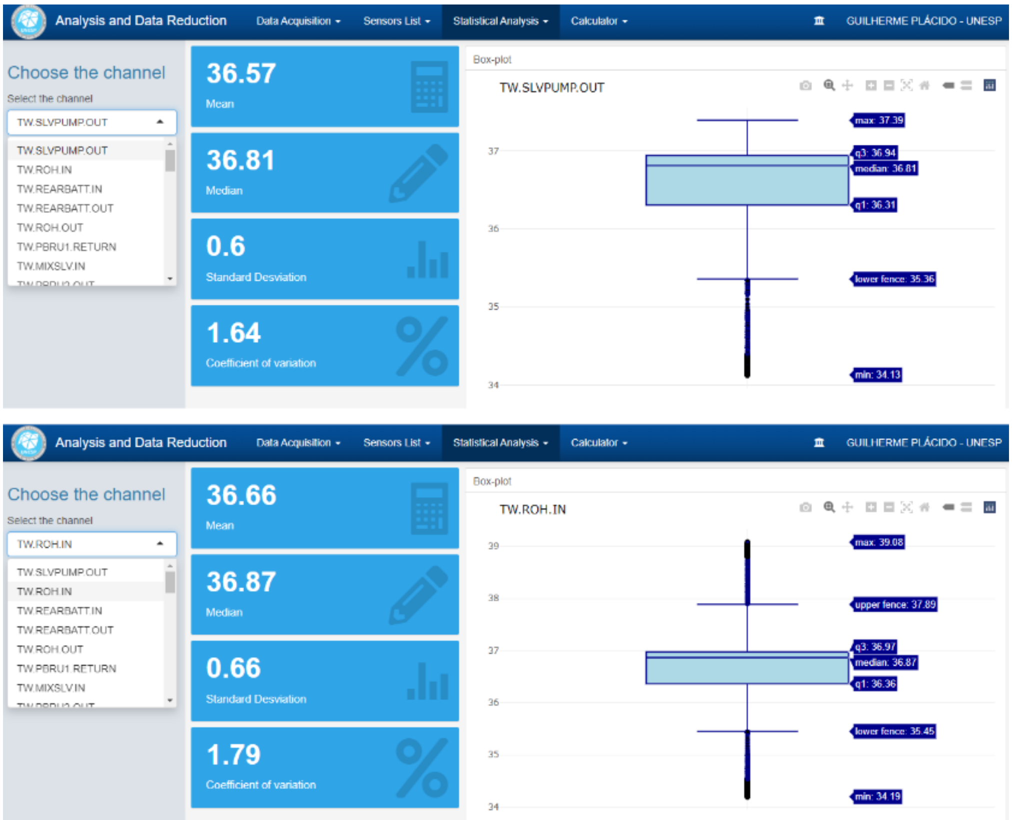

Figure 5. Interactive panel for descriptive statistical of temperatures from Thin Insulation.

Although there are outliers for values below the lower limit of 35.36 ^{o}C for the coolant temperature at the inlet of the duct, and a approximately symmetrical distribution of outliers above the upper limit of 37.89 ^{o}C and below the lower limit of 35.45 ^{o}C for the coolant temperature at the outlet in the pipeline, the average values with their respective deviations are very close. This observation may indicate that the presence of thermal insulator minimizes heat exchange with the environment through convection while increasing heat exchange through radiation. This, in turn, leads to minor discrepancies in inlet and outlet temperatures, even with larger temperature differences compared to the previous configuration.

Furthermore, although the data showed low dispersion and a homogeneous distribution, both coefficients of variation resulted in 1.64% for the inlet temperature and 1.79% for the outlet temperature. This phenomenon may indicate that the presence of a thermal insulator makes it difficult the heating of the fluid directed to the battery pack, which is, in principle, desirable. On the other hand, it does not result in a significant decrease in temperature compared to the Reference Steel Pipes configuration.

At last, as energy efficiency and the minimization of heat losses are essential aspects in the operation of heat transfer systems, Fig. 6 presents the inlet and outlet temperature profiles for each insulator configuration. Descriptive statistics data alone are not sufficient for decision making. The blue line represents the inlet coolant temperature in the pipeline, while the red line represents the outlet coolant temperature.

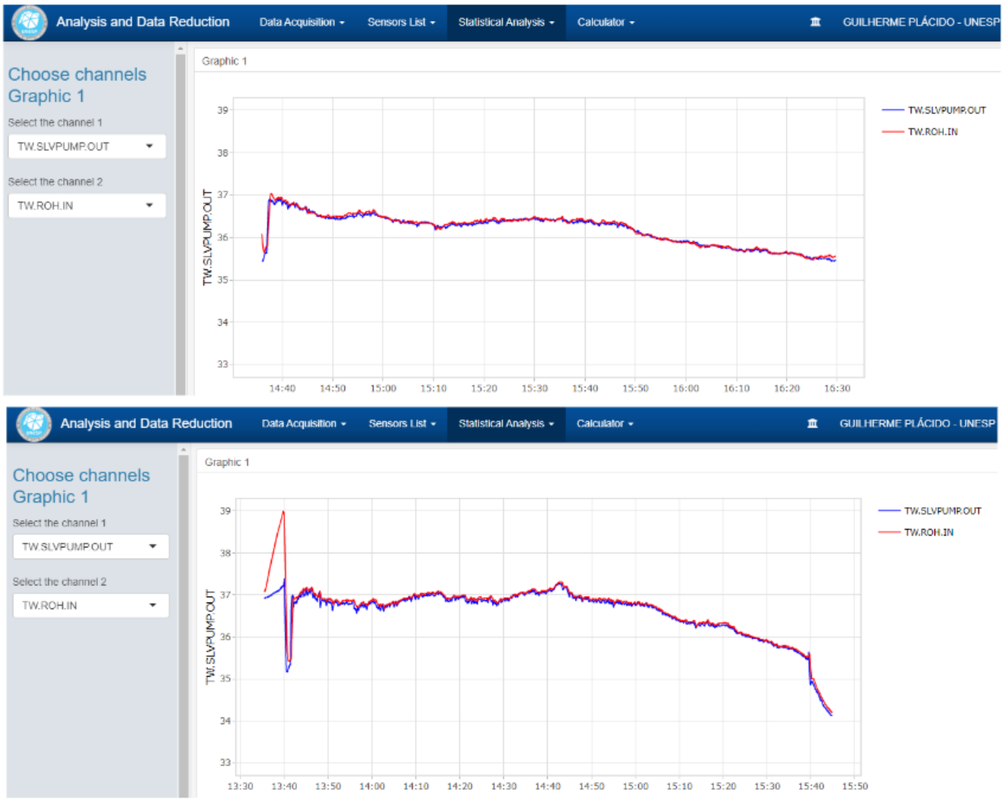

Figure 6. Behavior of inlet and outlet temperatures.

Therefore, the difference between the two isolation configurations was observed. The Reference Steel Pipes configuration, represented by the first figure, exhibits a significant reduction in outlet temperature compared to the Thin Insulation configuration shown in the bottom figure.

The temperature behavior indicates greater heat loss along the pipeline in the configuration without a thermal insulator and better heat retention in the configuration with the insulator. In addition, it was observed that the Reference Steel Pipes configuration presents a more uniform temperature profile along the pipeline with less variation between inlet and outlet temperature. This likely indicates better temperature distribution and maintenance throughout the test.

In this context, after excluding the initial temperature behavior that presented an abnormal peak at the beginning of the test, which can be explained by the lag of the vehicle's CAN network measurement module in starting to record the data, both the inlet and outlet temperature data were collected at the same moment. For the Reference Steel Pipes configuration, this occurred around 15:00, while for the Thin Insulation configuration, it happened around 14:40, and the respective flows at this moment were recorded, providing initial boundary conditions for numerical simulations by CFD (Porter *et al.*, 2002).

4.3 Application of Correlation Models

Correlation coefficients are statistical methods used to measure relationships between variables and what they represent. In this work, the application of Pearson and Spearman's correlation models was foreseen. The first is a parametric test, while the second is a non-parametric bivariate statistical test.

For the application of Pearson's correlation, two assumptions were verified for applying the model: a linear relationship between inlet and outlet temperatures and the normal distribution of the residuals. These assumptions were assessed based on the graphs in Fig. 7.

The graph on the left shows the distribution of residuals, which represent the difference between predicted values and collected values, concerning the adjusted values of the model (Kim, 2015). Observing this graph, it is evident that the red line is nearly coincident with the dashed line, indicating a linear relationship. The graph on the right demonstrates that the points are well distributed along the dashed line, confirming the behavior of the residuals. Furthermore, as observed in the graph on the right, the residuals exhibit a normal distribution, meeting both criteria for applying the model to the configuration of pipes without insulation.

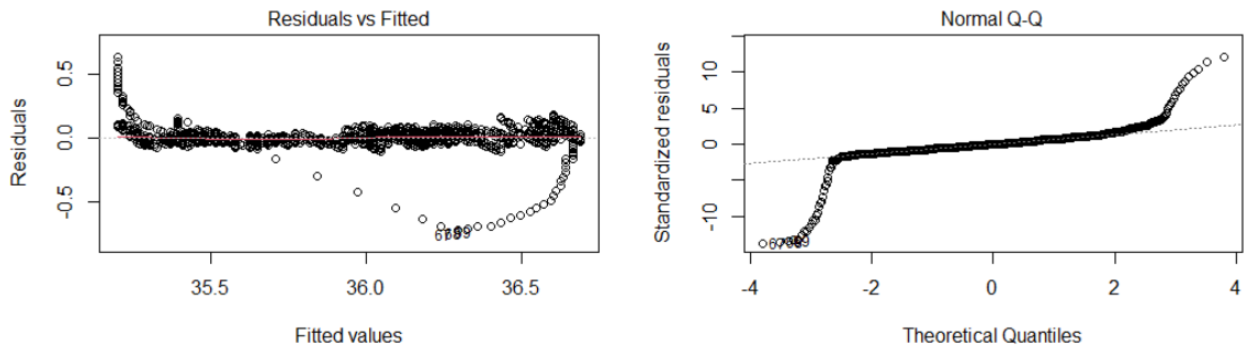

Figure 7. Assumptions for Pearson's Correlation - Reference Steel Pipes.

The same procedure was repeated, for the analysis of the tubes with insulator, in which the behaviors shown in Fig. 8 were obtained.

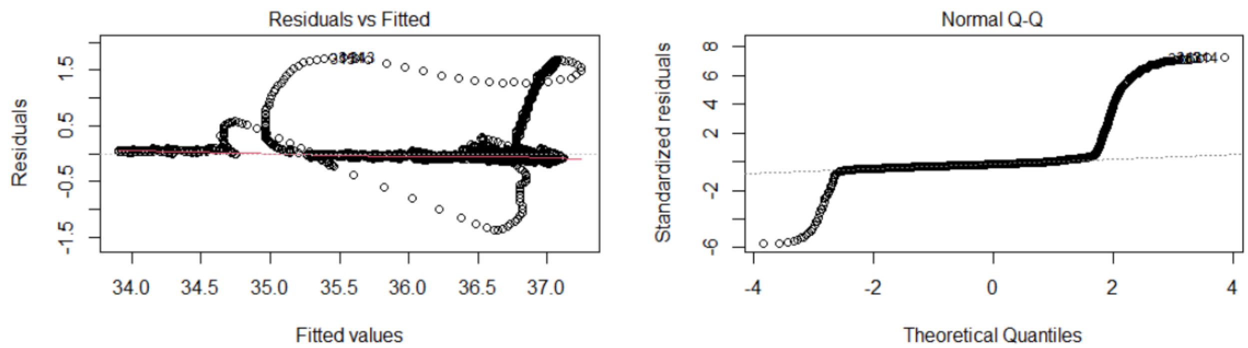

Figure 8. Assumptions for Pearson's Correlation - Thin Insulation.

On the other hand, it is evident from the graph on the right that the assumption of the normal distribution is not satisfied. Consequently, Spearman's correlation can be applied.

Table 1 shows the correlation model applied for each tube configuration and their respective associated correlation coefficient.

Table 1. Experimental results for each configuration tubes.

Configuration of tubes	**Correlation Model**	**Correlation Coefficient**
Reference Steel Pipes	Pearson	0.99
Thin Insulation	Spearman	0.96

The calculation of the correlation coefficient (r) using the graphical interface indicated that $r = 0.99$ with a p-value different from zero. Therefore, by analyzing these two values together, can be inferred, with a confidence interval of 95%, that the alternative hypothesis is accepted, indicating that the correlation coefficient is statistically different from zero and with $r \approx 1$, as one variable increases, the other follows the same pattern, and the opposite also occurs.

The calculation of the correlation coefficient (ρ) using the interface indicated $\rho = 0.96$, and the p-value also different from zero. Therefore, by analyzing these two quantities together, also for a confidence interval of 95%, the coefficient is statistically different from zero too.

Therefore, although the application of correlation models is important for analyzing the relationship between variables in a large volume of data, it is important to highlight that correlation coefficients do not imply causality. Thus, the correlation indices only provide information if a variable tends to a increase or decrease along with another, without implying that they are influencing each other.

4.4 Computational Tool Applied to Thermal System

The heat transfer in the tube of each configurations was calculated within the same graphic interface. This was made possible by monitoring the behavior of both inlet and outlet temperatures, as well as the coolant flow. The working fluid was ethylene glycol with a 50/50% volume mixture with water. The thermal properties of the fluid, such as density and specific heat capacity, were obtained from a supplier's data sheet. Using a get point tool and a script developed in R, the properties were obtained through iterative interpolation.

In both cases, heat transfer mainly occurs due to solar radiation and convection. However, the specific characteristics of the configurations can influence this energy balance. In the configuration without an insulator, the tubes are more exposed to the environment and direct solar radiation, leading to heat transfer primarily through solar radiation. Figure 9 illustrates the calculation of heat transfer at each moment of the test for each configuration.

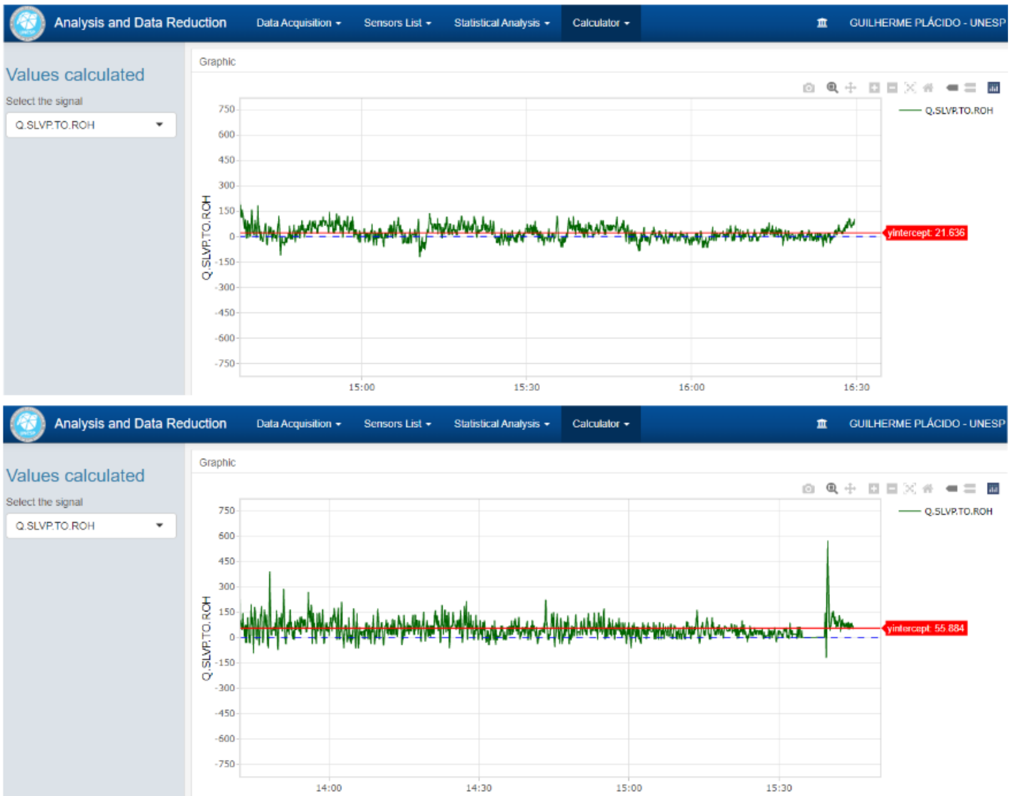

Figure 9. Heat transfer: Reference Steel Pipes and Thin Insulation.

However, the average value of 21.636 W suggests that the fluid heating rate is relatively lower than in the insulator configuration. The configuration with elastomer material around the tubes reduces direct exposure to solar radiation, which may appear counterintuitive. On the other hand, the average value of 55.884 W indicates that the fluid heating rate is considerably higher than in the reference configuration. This can be explained by the Thin Insulation configuration, which can minimize convective heat losses and retain more heat in the system.

Therefore, the configuration without an insulator can help cool the circuit of tubes connecting the batteries to the thermal management system more efficiently than the configuration with an insulator, even with more direct exposure to solar radiation. This can be important for the overall efficiency of the system and may impact the performance of the electric vehicle in different weather conditions.

5. CONCLUSION

This work has demonstrated that the usage of a robust and customized statistical computational tool is imperative within the highly competitive electric vehicle industry. The primary purpose of this tool is to streamline extensive data sets pertaining to the flow dynamics and heat transfer processes occurring within the tube circuits that constitute the thermal management system of the electric bus battery pack. To ensure data reliability, an exhaustive assessment was carried out involving the verification of data consistency and integrity, which included employing descriptive statistics and thermal signal behavior graphs derived from the experimental data. Consequently, it was verified that the data set accurately represented the phenomenon under investigation. However, it is important to note that when applying correlation methods, certain assumptions must be met to maintain consistency with the subject of study.

Therefore, for each of the configurations, based on the analysis of the data referring to the dynamics of the flow and the heat transfer process, it was possible to provide boundary conditions for numerical simulations by Computational Fluid Dynamics (CFD) and compose a database that will be used for numerical validations. This approach, along with the knowledge acquired in this complete process of developing a data reduction and the methodology applied to the ducts circuit connecting the batteries to the thermal management system of the electric vehicle, can be extrapolated to other similar vehicle applications.

6. ACKNOWLEDGEMENTS

The authors are grateful to the Mechanical Engineering Departament of the São Paulo State University and the Computational Support Center of São Paulo State University campus of Ilha Solteira, which provided the infrastructure and computational resources to perform this research.

This work has been possible in function of Teaching, Research and Extension Foundation of Ilha Solteira (FEPISA) grants, process nº 007/2022 (Scientific Initiation).

7. REFERENCES

Bufalo, L., Goneli, G. and Baumgartner, L., 2017. "Battery thermal management in electric vehicles: the combined liquid system". *Blucher Engineering Proceedings*, Vol. 4, pp. 435 – 447.

Chang, W., 2018. *R graphics cookbook: practical recipes for visualizing data*. O'Reilly Media.

Farcomeni, A. and Greco, L., 2016. *Robust methods for data reduction*. CRC press.

Gareth, J., Daniela, W., Trevor, H. and Robert, T., 2013. *An introduction to statistical learning: with applications in R*. Spinger.

Haymond, S., 2022. "Create laboratory business intelligence dashboards for free using r: A tutorial using the flexdashboard package". *Journal of Mass Spectrometry and Advances in the Clinical Lab*, Vol. 23, pp. 39 – 43.

Johnson, R.A., Wichern, D.W. *et al.*, 2002. *Applied multivariate statistical analysis*. Pearson.

Karimi, G. and Li, X., 2013. "Thermal management of lithium-ion batteries for electric vehicles". *International Journal of Energy Research*, Vol. 37.

Kim, B., 2015. "Understanding diagnostic plots for linear regression analysis". University of Virginia Library, Research Data Services + Science.

Lafaye de Micheaux, P., Drouilhet, R. and Liquet, B., 2013. *The R Software: Fundamentals of programming and statistical analysis*. Springer.

Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W. *et al.*, 1996. *Applied linear statistical models*. McGraw-Hill/Irwin.

Pandey, R. and Dhoundiyal, M., 2015. "Quantitative evaluation of big data categorical variables through r". *Procedia Computer Science*, pp. 582 – 588.

Porter, J., Agarwal, R. and Azad, R., 2002. *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*. Reston-VA.

Sanguesa, J.A., Torres-Sanz, V., Garrido, P., Martinez, F.J. and Marquez-Barja, J.M., 2021. "A review on electric vehicles: Technologies and challenges". *Smart Cities*, Vol. 4, pp. 372 – 404.

Tete, P., Gupta, M. and Joshi, S., 2021. "Developments in battery thermal management systems for electric vehicles: A technical review". *Journal of Energy Storage*, Vol. 35.

Vuong, Q.H., La, V.P., Nguyen, M.H., Ho, M.T. and Mantello, P., 2020. "Improving bayesian statistics understanding in the age of big data with the bayesvl r package". *Software Impacts*, Vol. 4, p. 100016.

Wickham, H., 2014. "Tidy data". *Journal of Statistical Software*, Vol. 59, pp. 1 – 23.

Wickham, H., 2015. *R packages: organize, test, document, and share your code*. O'Reilly Media.

Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Wickham, H. and Grolemund, G., 2016. *R for data science: import, tidy, transform, visualize, and model data*. O'Reilly Media.

8. RESPONSIBILITY NOTICE

The authors are solely responsible for the printed material included in this paper.