YIELD ANALYSIS IN SUGARCANE HARVESTERS USING DESIGN OF EXPERIMENTS (DOE) METHODOLOGY

[1,4]M.L. Silva, [2]J.J.A. Lima, [3]J. P. Molin, [1]A. Balbinot.

1. Laboratory of Electro-electronic Instrumentation and Artificial Intelligence (EEI-AI), Federal University of Rio Grande do Sul (UFRGS), 90035-190, Porto Alegre, Rio Grande do Sul, Brazil

2. AGCO Corporation, Hesston, Kansas, United States

3. Laboratory of Precision Agriculture (LAP), Department of Biosystems Engineering, "Luiz de Queiroz" College of Agriculture (ESALQ), University of São Paulo (USP), 13418900 Piracicaba, São Paulo, Brazil

4. AGCO Corporation, Canoas, Rio Grande do Sul, Brazil

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.

The sugarcane crop is highlighted in Brazil as the world's largest producer and the prospection of specialists is of strong growth for the next years. However, technological interventions through precision agriculture must be implemented to increase productivity and sustainability. Among them, the management of inputs guided by yield spatial variability for optimizing production and income is a promising strategy. This project approaches the implementation of the methodology of analysis of experiments to propose a new sugarcane flow measurement system on the harvester. The research sought to identify the correlation and influence of displacements of the feed rolls and hydraulic pressures to which the machine is exposed during the harvesting process with the amount of mass harvested from the crop. After the implementation of DoE (design of experiments), using a significance level of 95%, the model showed that elevator pressure, the interaction between pressure in the chopper and elevator pressure, the interaction between cutting pressure, chopping pressure, and elevator pressure (bottom), showed influence (dependence) on the measurement of sugarcane flow.

Keywords.
DoE; Sensor correlation; Yield maps; Sugarcane.

M.L.Silva, J.J.A.Lima, J. P. Molin & A. Balbinot (2024). Yield Analysis in Sugarcane Harvester Using Design of Experiments (DoE) Methodology. In Proceedings of the 16th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

Introduction

Sugarcane is crucial for renewable energy, especially in Brazil, the world's largest producer. São Paulo state leads the country's production, accounting for 53%. In the 2023/2024 harvest, Brazil produced around 713.2 million tons [1]. The total national area was about 8,333.9 thousand hectares, with an average yield of 85,580 kg ha^{-1}. Globally, sugarcane covers 26.1 million hectares and is Brazil's second-largest energy source, with ethanol playing a key role in reducing greenhouse gas emissions [1].

The sugarcane industry plays a prominent role in national agribusiness and is increasingly driven by the growing need for renewable energy sources to meet local and global demands. The factors fueling production growth and the expansion of cultivated areas have nearly doubled in value in recent years [2] and it's foreseeable that there could be a rise in sugarcane production in the upcoming years. Yet, to realize this potential, technological interventions must be integrated across the production process. These interventions would assess potential efficiency gains and resource input reductions, leveraging precision agriculture principles. At present, about 98% of production employs mechanized harvesting systems [2], consolidating this production system on Brazil [7].

A useful approach involves replacing inputs based on yield, which is represented spatially in a map. This map helps producers identify and address areas with lower yield while highlighting high-production zones and factors that contribute to them. To create accurate yield maps, robust systems are needed to evaluate the flow of sugarcane harvesters using sensors onboard [9]. These sensors measure various factors affecting crop yield, and the system integrates this data with GNSS information to generate instant yield maps. These maps show crop variability, enabling precise fertilization recommendations for future cycles based on previous yield and nutrient levels. Yield maps are essential for managing spatial variations in crops and analyzing them over time reveals patterns in crop behavior, aiding in the characterization of consistent regions [10].

Most current instant yield systems consist of scales located in the elevator region [14], as shown in Figure 2. However, the weighing system presents issues such as tilt compensation, vibration, and tare change caused by mineral material, sediment, and debris. [11] developed a measurement system based on hydraulic pressure from the chopper and elevator drives of the harvester to determine material flow, [6] and [14] studied variations for scale use in the transfer, [15] developed a measurement system based on the harvester's feed rollers, [16] described the development of a yield monitor using optical sensors for sugarcane harvesting. The measurement errors obtained in the reported tests are about 70% higher than those obtained in monitors installed in grain and forage harvesters. Furthermore, there are only reports of brief tests, usually with a calibration adjusted for a full transfer load. There are no reports of continuous measurement, so the actual behavior of the sensors can be evaluated.

Based on the research conducted, the main studies relating sugarcane yield crop to experimental sensing in sugarcane harvesters were compared. Table 1 presents the results of the studies.

Table 1 - Systems evaluated for measuring productivity in sugarcane harvesters.[19]

Authors	Measure concept	Sensor Position	Test performed	Average Error [%]	R^2 [%]
Cox et al. 1998	Pressure	Chopper and elevator	NA	10	95
Benjamin et al. 2001	Scale	Elevator base	118 tests	11	96,6
Molin e Menegatti 2004	Scale	Elevator base	4 steps 17 Transshipment	3,5	NA
Cerri e Magalhães 2005	Scale	Elevator base	1 Field test 11 Transshipment	1,0	66
Price et al. 2011	Optic sensor	Elevator base	1 Field Test 50 Transshipment	7,5	97
Hernandez et al. 2003, 2005	Displacement sensor	Roller opening	NA	NA	NA

*NA – Not Available; AE – Average error

The term "experiment" is used very precisely to indicate an investigation where the system under study is under the control of the researcher, certain traits may fall outside the investigator's control. [11]. According to [12], an experiment can be seen as a test, or a series of tests, where changes are applied to the input variables of a process or system, and then observations are made to identify the changes that occur in the output variables. Additionally, the design of the experiment refers to the process of planning experiments so that appropriate data can be analyzed using statistical methods, resulting in valid and objective values for assertive conclusions. According to [13], one of the main objectives of experimental design is to estimate how changes in input factors affect the results or responses of the experiment.

DOE as scientific method was most popular in scientific areas of medicine, engineering, biochemistry, physics, and computer science. Its application in these areas counts about 50% compared to all other scientific areas. Agricultural application just represents 4% of the application [20].

Based on Montgomery's guide to the design and analysis of experiments [12], the seven steps for planning and executing a project and experiment analysis were addressed. Initially, the problem was characterized, including the research conducted in the area up to the present. The influencing factors and levels were chosen, the response variables were selected, an experimental design model was determined, the experiment was conducted, and finally, the data were analyzed, and conclusions drawn.

In this study, we aimed to comprehend how controllable factors influence the response variable and propose a new measurement system. Previous research showed that analyzing individual machine systems wasn't sufficient to predict yield accurately. To improve evaluation, we needed more instrumentation and data collection, analyzing additional variables from the harvesting system.

To achieve this, we implemented the Design of Experiments (DoE) to understand how controllable factors in the harvesting machine system correlate and influence the sugarcane flow. This involved installing sensors, collecting data, and conducting a thorough analysis to gain deeper insights into the mechanical dynamics during harvesting in the field.

2. Methodology

Based on the pre-design and experiment master guide, the first step is to clarify the problem clearly and precisely [13]. The capability to measure the instantaneous sugarcane flow, allows us to generate confidence yield maps useful to growers.

For a better understanding of the problem and analysis of the factors that interfere with the measurement of instantaneous flow of sugarcane in sugarcane harvesters, an Ishikawa diagram was implemented to assist in identifying the real causes of the problem.

The Fishbone diagram (also called the Ishikawa diagram) is a tool for identifying the root causes of problems [17]. The Fishbone diagram is an analysis tool that provides a systematic way to look at the effects and causes that create or contribute to these effects. Due to its function in the experiment, the Fishbone diagram can also be related to a cause-and-effect diagram [18]. It primarily represents a model that suggests correlations between an event and its multiple causes. The structure provided by the diagram, helps stakeholders to think systematically, facilitating group dynamics. Among the benefits of constructing a Fishbone diagram is the ease of determining the root causes of a problem or characteristic, using a structured approach, encouraging group participation, and utilizing the group's knowledge.

Due to the limited research related to the area, the formulation of the root causes of the problem and the selection of influencing factors were developed together with experts who supported the elaboration of this experiment. The main factor identified for measurement errors in current systems was vibrations in the elevator due to the activation of the hydraulic system, leading to instability in the system. Variables such as soil type, terrain slope, crop variety, harvesting speed, and others, were also identified as possible influencers for measurement errors. Based on these factors and the research conducted, an Ishikawa diagram was developed to facilitate understanding and elaboration of the next steps. The diagram can be seen in Figure 1.

Based on the discussion related to the factors influencing the low accuracy of current models of instantaneous yield measurement, points to be instrumented on the sugarcane harvester were selected to correlate with sugarcane flow. The points were selected by mapping the possible problems presented in the Ishikawa diagram. The sensors used collect data on pressure, displacement, acceleration, and mechanical tension, to examine their correlation with sugarcane flow to implement the new measurement system.

Thus, the influencing factors for the experiment are mapped through the physical quantities collected by the sensors, including the acceleration of the elevator base, mechanical tension in the elevator and transfer, roller displacement, and hydraulic system pressures. By collecting this data, we can broaden our comprehension of the most significant variables associated with sugarcane flow, facilitating the establishment of a more precise measurement system.

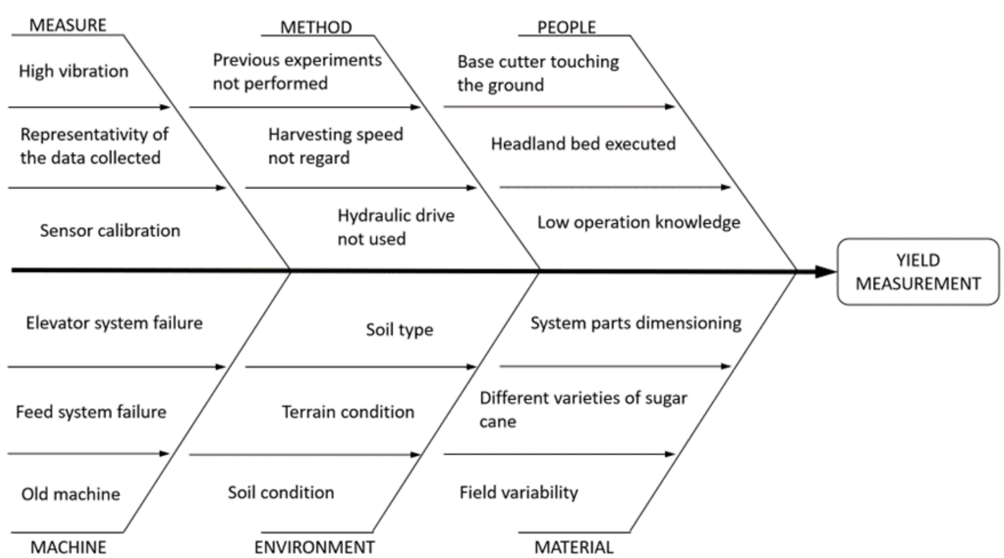

Figure 1 – Sugarcane yield measurement fishbone diagram.

Due to the large number of controllable factors and levels for analysis, four controllable factors were examined and their levels (Table 2). Several interactions are necessary, and up to now, 79 measurements have been taken, involving Pressure on elevator 1 (Sensor F), Pressure on chopper (Sensor B), pressure on elevator 2 (Pressure E), and roller displacement (Disp. E2). The choice of levels for each controlled factor was based on previous studies conducted to enhance the understanding of the system through certain pressure and displacement values.

Table 2 – Levels and controlled factors

Factor Information

Factor	Levels	Values
SensorB	2	72; 80
SensorE	2	15; 16
SensorF	3	41; 42; 43
Disp_E2	2	1,0; 1,1

The response variable is the measurement of instantaneous sugarcane flow through the harvester. To measure the response variable, we used the setup [19] of measurement system with good accuracy (uncertainty of +/- 1%).

With the prior definition of the influencing factors and the response variable of the problem, a set of questions were set (Figure 2), with the innermost level correlated with the other levels, up to the outermost level. For the current experiment, questions related to roller displacement, hydraulic pressures, and their interactions will be addressed.

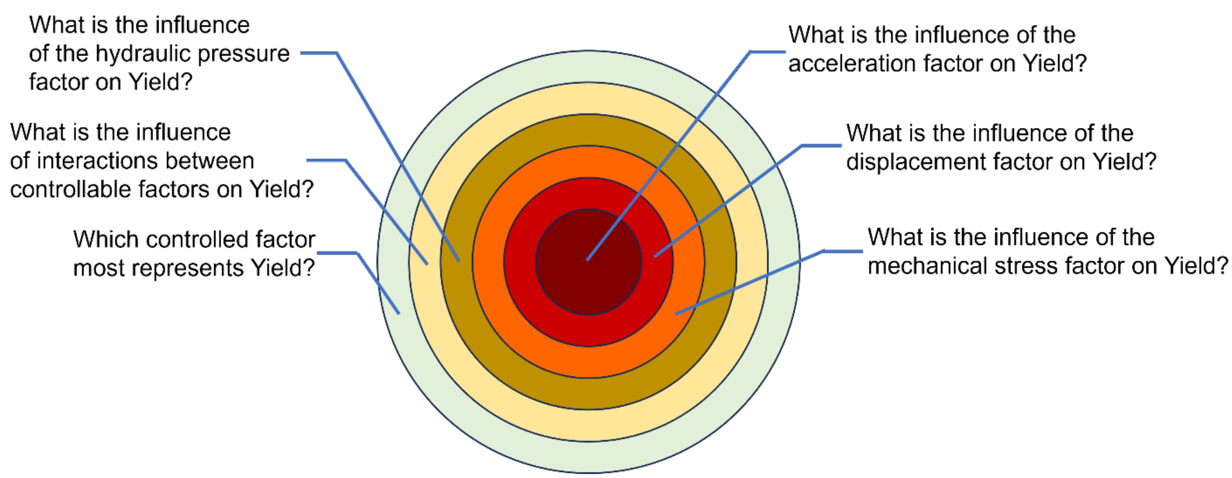

Figure 2 – Hierarchical relationships of the study objective.

The instrumentation points were selected based on the mechanisms present in the sugarcane harvester directly linked to the amount of harvested mass. In total, the system comprised 15 data collection channels [19], and for the current study, random measurements were taken, containing 4 data collection channels.

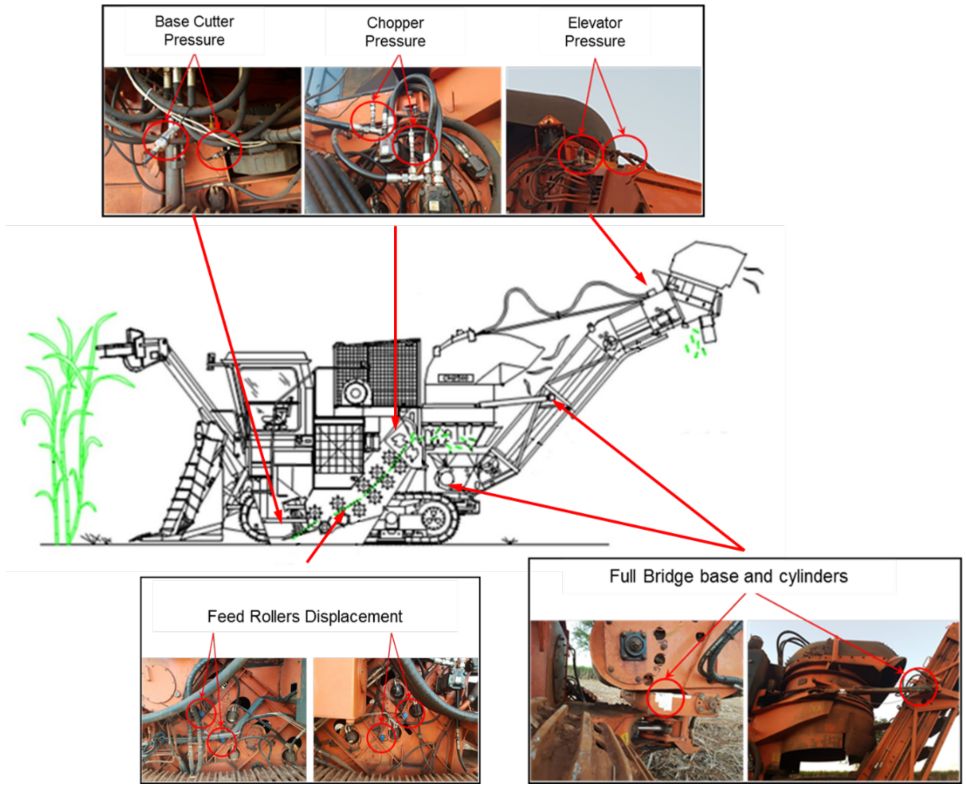

Figure 3 – Sensors positioning - [19]

Data collection was carried out in different locations with varied productivity, and the locations where data acquisition was conducted, and the type of cutting are shown in Figure 4 and Table 3. After data acquisition, preprocessing steps were performed to ensure data quality. The data were acquired at a rate of 400 Hz and subsequently evaluated in their spectrum to find significant frequencies for sugarcane cutting operation. Preprocessing steps involved filtering, resampling of the sample, spectral analysis, and mathematical operations to assess the loads.

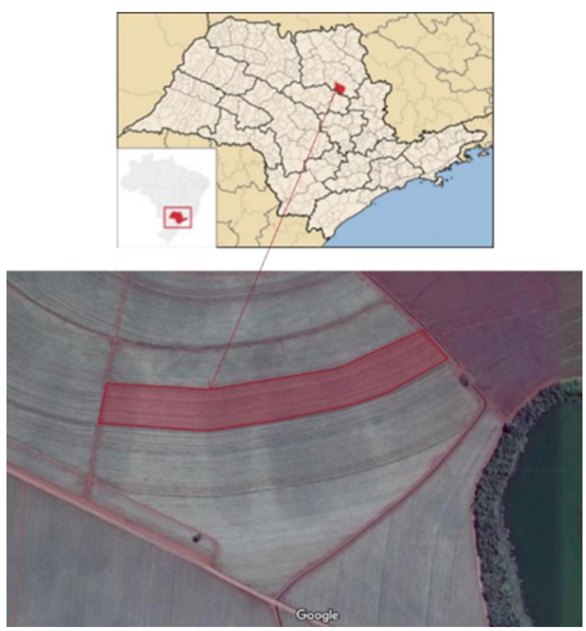

Figure 4 – Field localization: [19]

Table 3 – Data related to crop cutting and variety - Fonte: [19]

Place	Luiz Antônio – SP
Mill	Moreno
Average Yield [t ha^{-1}]	90
Season	2º
Variety	RB98769
Harvester	Valtra BE1035 (AGCO, Ribeirão Preto)
Harvesting Area	10 rows – 16 ha

After collecting data related to the sugarcane yield as a function of the controllable factors present in sensor B (72 bar and 80 bar), sensor E (15 bar and 16 bar), sensor F (41 bar, 42 bar, and 43 bar), and sensor Disp E2 (1.0m and 1.1m), the database was evaluated. The Minitab computational tool version 20.4.0 was used to perform the DoE assessment. The correlation was evaluated using the Pearson coefficient, with a total number of 79 data rows.

3. Results and discussion

3.1. Data Analyses

The correlation performed between the controllable factors and the control variable (Figure 5)

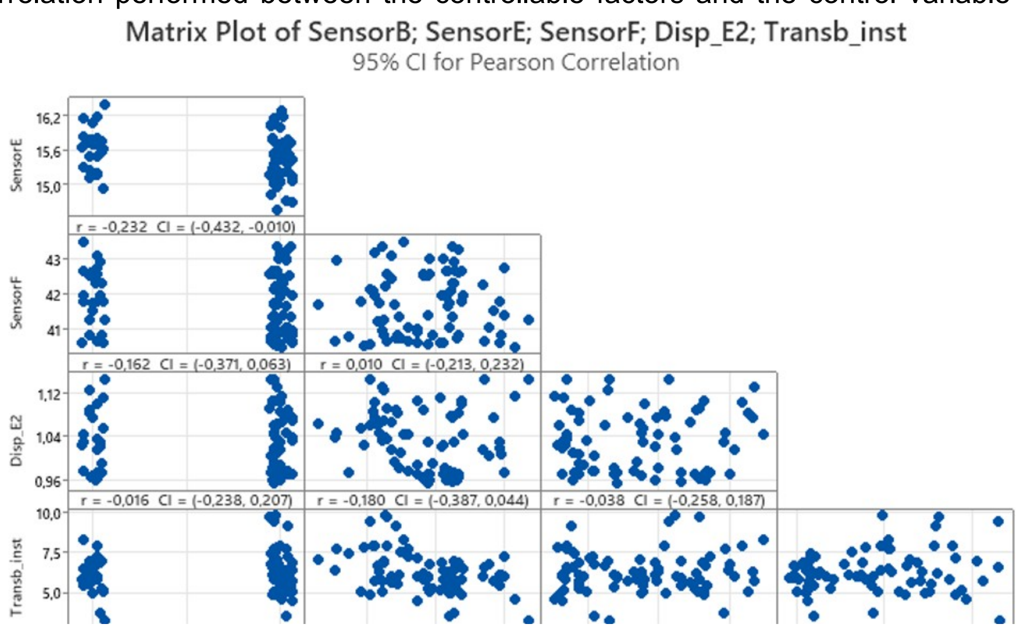

Figure 5 – Correlation between experiment data.

help to understand the behavior of hydraulic pressures and roller displacements. It is possible to identify that the most significant correlation between the controllable factors and sugarcane flow occurs between Sensor E and instantaneous overflow. We can highlight the correlation between them where the coefficient was -0.594, showing a significant interaction between these variables. However, due to the negative coefficient, the correlation between them is inverse.

Given the study's interest in developing a new yield measurement system for sugarcane harvesters based on variables directly measured in the harvesting system, linear regression calculations were also performed, relating the control variables to the variable of interest (Figure 6). It can be inferred that the linear regression model analyzing the factors separately does not accurately represent the variable of interest in the study.

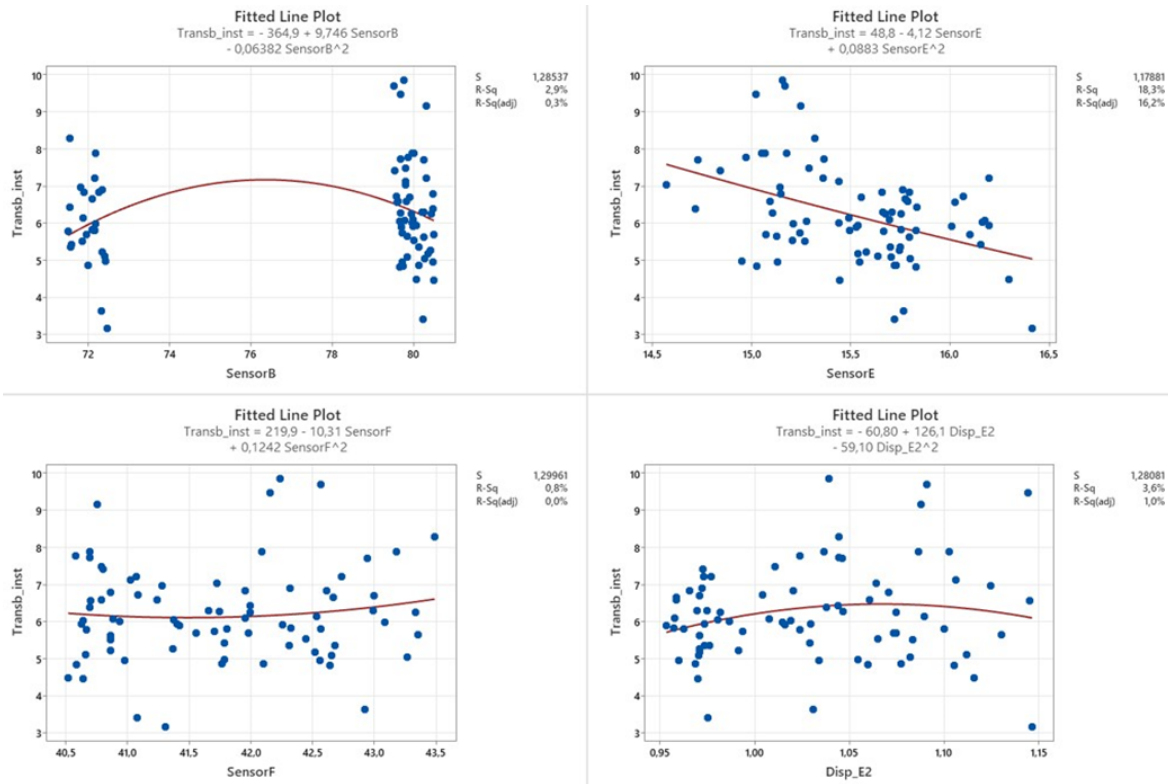

Figure 6 - Linear regression.

The controllable factor of pressure on the chopper (sensor B) with 2 levels, namely 72 and 80 bar, the pressure on elevator 2 (sensor F) which was divided into 3 levels, namely 41, 42, and 43 bar, the pressure on elevator 1 (sensor E) which was divided into 2 levels, namely 15 and 16 bar, and finally the displacement on the rollers (Disp E2) which was divided into 2 levels, namely 1.0 and 1.1 m. It can be observed through the Adj MS (Adjusted mean squares) column that for instantaneous sugarcane flow, the most significant controllable factor (individually) is sensor E, followed by sensor B, albeit with less representativeness. Regarding interactions between variables, the interaction between sensor B and sensor E stands out. The values and significances are displayed in Table 4.

Table 4 - Variance analysis

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	23	59,468	2,5856	1,96	0,019
Linear	5	26,763	5,3526	4,07	0,003
SensorB	1	2,870	2,8697	2,18	0,145
SensorE	1	18,202	18,2020	13,83	0,000
SensorF	2	5,685	2,8424	2,16	0,124
Disp_E2	1	0,506	0,5063	0,38	0,537
2-Way Interactions	9	18,011	2,0012	1,52	0,161
SensorB*SensorE	1	6,406	6,4062	4,87	0,031
SensorB*SensorF	2	2,085	1,0427	0,79	0,457
SensorB*Disp_E2	1	0,043	0,0430	0,03	0,857
SensorE*SensorF	2	4,488	2,2439	1,70	0,190
SensorE*Disp_E2	1	0,590	0,5898	0,45	0,506
SensorF*Disp_E2	2	4,455	2,2275	1,69	0,192
3-Way Interactions	7	17,094	2,4420	1,86	0,093
SensorB*SensorE*SensorF	2	12,158	6,0790	4,62	0,014
SensorB*SensorE*Disp_E2	1	0,229	0,2291	0,17	0,678
SensorB*SensorF*Disp_E2	2	5,994	2,9968	2,28	0,111
SensorE*SensorF*Disp_E2	2	0,224	0,1121	0,09	0,918
4-Way Interactions	2	0,092	0,0462	0,04	0,966
SensorB*SensorE*SensorF*Disp_E2	2	0,092	0,0462	0,04	0,966
Error	62	81,611	1,3163		
Total	85	141,079			

Analyzing the error distribution through the graphs shown in Figure 7, we can say that the error tends to be normally distributed, meeting the criteria of a factorial study. We can identify a distribution of points slightly different from a straight line, possibly due to the unbalanced database and the data being collected over time.

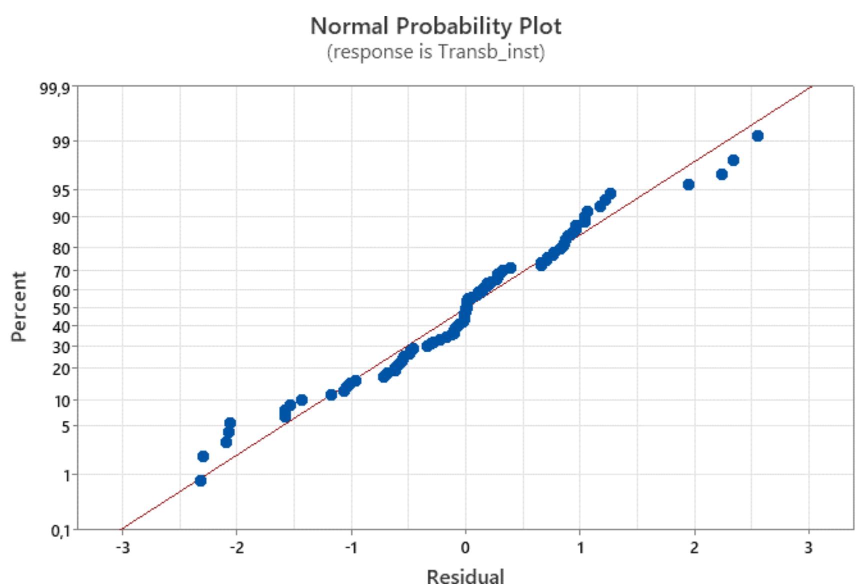

Figure 7 – Error distribution.

The analysis of residuals was performed through a histogram shown in Figure 8. The database was composed of a total of 79 data rows. It can be observed that for the residuals, the distributions follow a normal distribution.

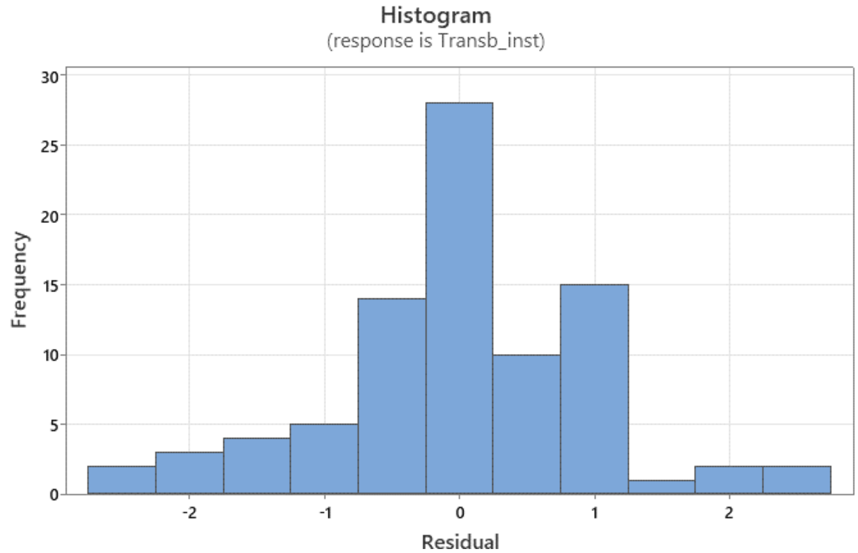

Figure 8 – Normal distribution.

The data were also assessed for the assumption of constant variance, where the residuals should appear randomly dispersed around the line 0. The results are displayed in Figure 9, with no evident clustering of data.

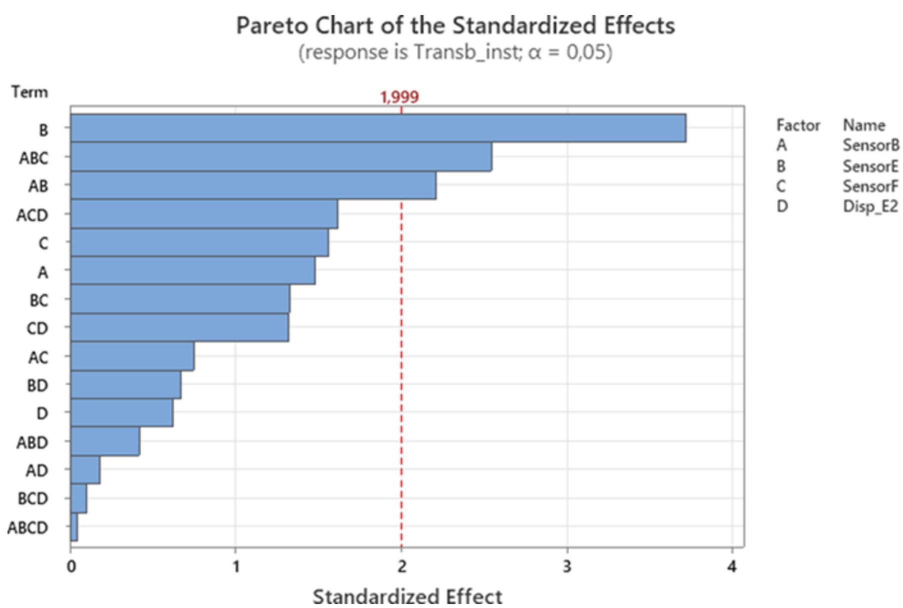

Figure 9 - Constant variance,

Through the graph of the residuals versus the order of observations, it was possible to notice that the model exhibits a random behavior, indicating that the samples are independent of each other, a relevant factor for implementing the DoE (Figure 10).

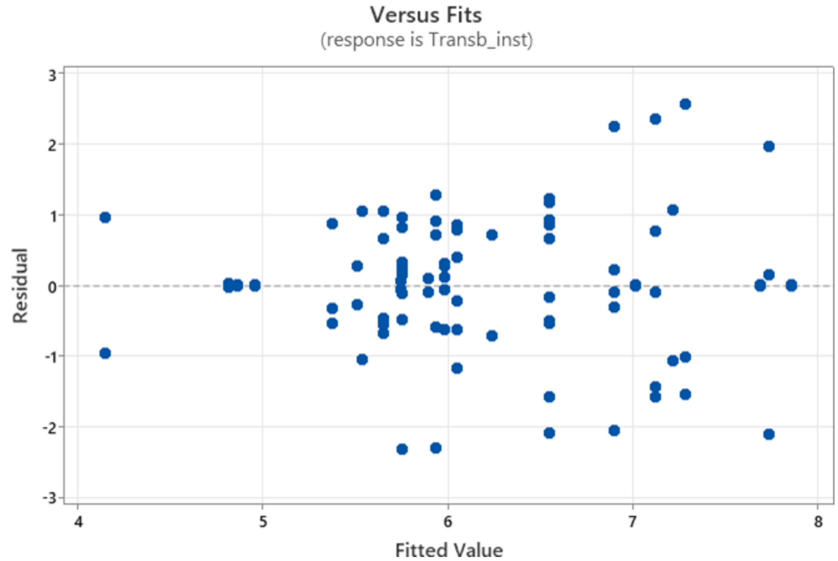

Figure 10 - Graph of residuals versus order of observations

Using the Pareto chart, which aids in visualizing the significant controllable factors related to the variable of interest (Figure 11), it was possible to answer questions about the influence of factors on the measured variable. Using a significance level of 95%, the model indicated that Sensor E, the interaction between Sensor B and Sensor E, and the interaction among sensors E, B, and F had influence (dependency) on sugarcane flow measurement. The greatest significance was represented by Sensor E, corresponding to the pressure of elevator 1.

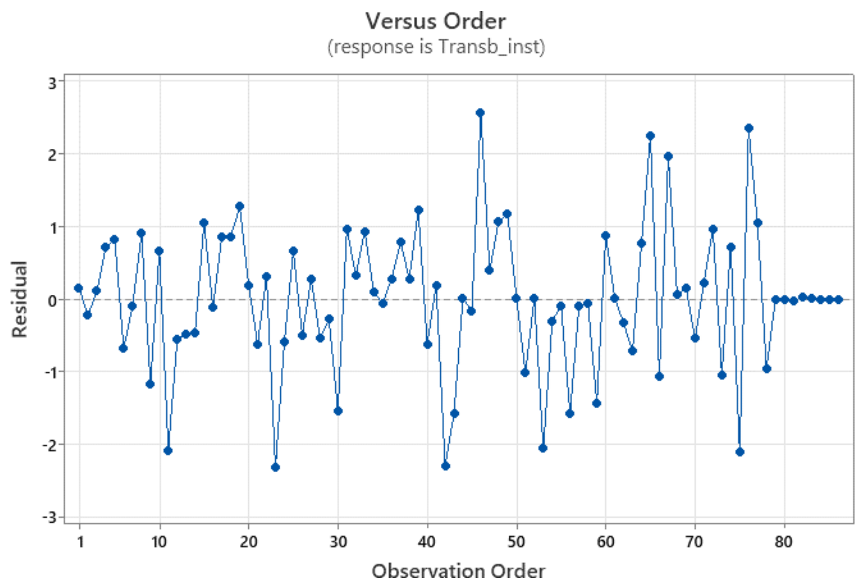

Figure 11 - Effects of factors on the variable of interest

The results obtained in this study demonstrate a significant correlation between controllable factors and the control variable, with Sensor E and instantaneous overflow standing out with a coefficient of -0.594. This inverse correlation reveals an important interaction between these variables, highlighting the relevance of Sensor E in measuring sugarcane flow. The linear regression analyses indicate that models analyzing factors separately do not accurately represent the variable of interest, suggesting the need for more complex models to capture the dynamics of the harvesting system.

Comparing these results with previous studies, such as Cox et al. (1998), who also investigated pressure in choppers and elevators, and Benjamin et al. (2001), who used scales at the base of the elevators, it is observed that the findings align with measurement trends in the field. However, the approach of these studies using single variables to measure crop productivity may be less accurate compared to identifying significant interactions between multiple sensors, such as sensors B and E, which provide a more detailed and integrated view of the system's behavior. Furthermore, the use of the Pareto chart and residual analysis confirms the validity and robustness of the proposed model.

Thus, this study makes a significant contribution to the field by presenting a new methodology for measuring yield in sugarcane harvesters using variables directly measured in the harvesting system. The results indicate that to improve the accuracy of sugarcane flow measurements, it is important to consider both individual factors and their interactions. This paves the way for future research and technological improvements in harvesters, aiming to optimize the harvesting process and increase operational efficiency.

4. Conclusion

After analyzing the data employing the DoE methodology, it was possible to enhance the understanding regarding the harvesting system of the sugarcane harvester and make correlations with sugarcane flow. The Ishikawa diagram provided a better understanding of the possible root causes of the low quality of the currently studied and/or implemented productivity measurement systems, which aided in the experiment design.

Through the linear regression calculations applied to the controllable factors, it was inferred that the experimental equations developed considering only one controllable factor do not accurately represent sugarcane flow measurement with the necessary precision, and that a possible combination of sensors should be used for better characterization of the response variable.

Using error distribution analyses, residual distribution, constant variance assumption, and the residual plot versus order of observations, it was verified that the experiment conducted so far follows a normal distribution, displays random behavior, does not exhibit any clustering in measurements, and shows a normally distributed error. These factors are extremely important for the implementation of the DoE.

The implementation of DoE for analyzing the responses of the controllable factors on the variable of interest made it clear that the null hypothesis H0 (the response variable, instantaneous sugarcane flow, is independent of the controllable factors) was rejected, and that with a significance level of 95%, Sensor E, the interaction between Sensor B and Sensor E, and the interaction among sensors E, B, and F showed influence (dependency) on sugarcane flow measurement.

The study contributes to the implementation of a new measurement system, where only variables that have dependency on sugarcane flow should be considered. For the continuation of the study, more measurements related to other controllable factors of the system should be implemented, as well as increasing the number of trials to increase the degrees of freedom of the model. At the end of the experiment, the controllable factors that show greater significance with instantaneous sugarcane flow will be used for the implementation of a new measurement system with acceptable errors, comparing error metrics used in the study's area of interest.

References

1. CONAB. Acompanhamento de safra brasileira: cana-de-açúcar. Brasília 2023, 17 (1).
2. UNICA. União da Indústria de cana-de-açúcar-Brasil. 2020, 10 (1).
3. Dinardo, M., Vasconcelos, A. & Landell, M. Cana-de-açúcar Campinas. Instituto Agronômico 2010, 6 (882).
4. Faostat. Food and Agriculture Organization. 2018.
5. MINISTÉRIO DA AGRICULTURA, P. E. A. Usinas e destilarias cadastradas. 2015, 6 (1).

6. Molin, J. Agricultura de precisão em cana-de-açúcar é mais do que uma realidade. Coplana 2008 (80).
7. Silva, R. P. e. a. Controle estatístico aplicado ao processo de colheita mecanizada de cana-de-açúcar. Engenharia Agrícola 2008, 28 (292-304).
8. Ripoli, T. C. C. & Ripoli, M. L. C. Biomassa de cana-de-açúcar: colheita, energia e ambiente Piracicaba. Escola Superior de Agricultura “Luiz de Queiroz” 2009, 1 (333).
9. Balastreire, L. A. A experiência com pesquisas em Agricultura de Precisão na ESALQ. Feira para usuários de geoprocessamento na américa latina 1998, 4.
10. Fleming, L. & Westfall, G. Evaluating Farmer Defined Management Zone Maps for Variable Rate Fertilizer Application. Precision Agriculture 2000, 2 (201–215).
11. Cox, D. & N., R. The Theory of the Design of Experiments. Chapman and Hall CRC 2000 (8-20).
12. Montgomery, D. C. Design and analysis of experiments (John wiley e sons, 2017).
13. Coleman, D. E. & Montgomery, D. C. A systematic approach to planning for a designed industrial experiment. Technometrics 1993, 35 (1-12).
14. Mailander, M., Benjamin, C., Price, R. & Hall, S. Sugar cane yield monitoring system. Applied Engineering in Agriculture 2010, 26 (965).
15. Fernandz, F. et al. Yield mapping sugarcane in Australia. Proceedings of the 6th European Confe- rence on Precision Agriculture 2007, 1 (3-7).
16. Mailander, M., Benjamin, C., Price, R. & Hall, S. Sugar cane yield monitoring system. Applied Engineering in Agriculture 2010, 6 (965).
17. Juran, J. & Godfrey, A. B. Quality handbook. Republished McGraw-Hill 1999, 34–51 (173).
18. Watson, G. The Legacy Of Ishikawa. Quality Progress 2004, 37 (54-47).
19. Lima, J. d. J. A. d., Maldaner, L. F. & Molin, J. P. Sensor Fusion with NARX Neural Network to Predict the Mass Flow in a Sugarcane Harvester. Sensors 2021, 21 (14).
20. Durakovic, B. Design of experiments application, concepts, examples: State of the art. Available, Periodicals of Engineering and Natural Scinces 2017, Vol 5, No 3, December 2017, pp. 421–439. DOI: 10.21533/pen.v5i3.145.