IMPACTO CIENTÍFICO EDITORA

PROCESSAMENTO E LIMPEZA DE DADOS DE PRODUTIVIDADE AGRÍCOLA: APLICAÇÃO DE UM SCRIPT EM PYTHON

PROCESSING AND CLEANING OF AGRICULTURAL PRODUCTIVITY DATA: APPLICATION OF A PYTHON SCRIPT

PROCESAMIENTO Y LIMPIEZA DE DATOS DE PRODUCTIVIDAD AGRÍCOLA: APLICACIÓN DE UN SCRIPT DE PYTHON

10.56238/edimpacto2025.015-011

Raphael Prazeres da Silva[1], Welington Gonzaga do Vale[2], Janyelle do Nascimento Silva[3], Valfran José Santos Andrade[4], Patricia de Azevedo Castelo Branco do Vale[5], Adilson Machado Enes[6], Diego Andrade Pereira[7]

ABSTRACT

A script written in Python was developed to process and clean agricultural yield data from grain harvesters on a farm located in Brasnorte (MT), aiming to improve data reliability in Precision Agriculture. The code, using the Pandas library, followed three main steps: (1) filtering by machine operation status (retaining only “Effective” records), (2) removal of outliers (values <500 kg/ha or >twice the average), and (3) iterative adjustment of machine-specific yield values to match the field average. The cleaned data were interpolated in QGIS using the IDW method. The results showed that 58.8% of the raw data were discarded in Field 1 and 66.9% in Field 2, mainly due to failures or zeroed sensors. Yield averages increased from 2.67 t/ha to 3.67 t/ha (Field 1) and from 2.52 t/ha to 3.82 t/ha (Field 2), with the elimination of extreme values. The generated maps highlighted critical zones near field edges and data gaps. The results suggest that the tool efficiently automates data cleaning, though future studies should consider including cross-validation to reinforce the reliability of the results.

Keywords: Precision agriculture. Data analysis. Yield map.

[1] Undergraduate student in Agricultural Engineering — Federal University of Sergipe
E-mail: rprazeress@gmail.com
[2] Dr in Plant Production — Federal University of Sergipe
Email: valewg@gmail.com
[3] Master's student in Intellectual Property Science — Federal University of Sergipe
E-mail: janyelle.engagricola@gmail.com
[4] Master in Water Resources — Federal University of Sergipe
Email: valfranjose40@gmail.com
[5] Dr in Animal Science — Federal University of Sergipe
Email: patriciavale78@gmail.com
[6] Dr in Agricultural Engineering — Federal University of Sergipe
E-mail: adilsonenes@gmail.com
[7] Mechanical Engineer — Federal University of Sergipe
Email: diegoandrade_senai@yahoo.com.br

RESUMO

Desenvolveu-se um *script* em Python para processar e limpar dados de produtividade agrícola de colhedoras em uma fazenda localizada em Brasnorte (MT), visando melhorar a confiabilidade em Agricultura de Precisão. O código, utilizando a biblioteca Pandas, aplicou três etapas: 1) filtragem por estado operacional (apenas registros "Efetivo"); 2) remoção de *outliers* (valores <500 kg/ha ou >dobro da média) e 3) ajuste iterativo das médias por equipamento. Os dados tratados foram interpolados no QGIS utilizando o método IDW. Os resultados mostraram que 58,8% dos dados brutos foram descartados no Talhão 1 e 66,9% no Talhão 2, principalmente devido a falhas ou sensores zerados. As médias de produtividade aumentaram de 2,67 t/ha para 3,67 t/ha (Talhão 1) e 2,52 t/ha para 3,82 t/ha (Talhão 2), com a eliminação de valores extremos. Os mapas gerados revelaram áreas críticas nas bordas e regiões com falhas. Conclui-se que a ferramenta é eficaz na automação da limpeza dos dados de produtividade, no entanto estudos futuros devem considerar a inclusão de validação cruzada para reforçar a confiabilidade dos resultados.

Palavras-chave: Agricultura de precisão. Análise de dados. Mapa de produtividade.

RESUMEN

Se desarrolló un script en Python para procesar y depurar datos de productividad agrícola de las cosechadoras en una finca ubicada en Brasnorte (MT), con el objetivo de mejorar la confiabilidad en la agricultura de precisión. El código, utilizando la biblioteca Pandas, aplicó tres pasos: 1) filtrado por estado operativo (solo registros "Efectivos"); 2) eliminación de valores atípicos (valores <500 kg/ha o >2 veces el promedio); y 3) ajuste iterativo de promedios por equipo. Los datos procesados se interpolaron en QGIS mediante el método IDW. Los resultados mostraron que el 58,8% de los datos brutos se descartaron en la Parcela 1 y el 66,9% en la Parcela 2, principalmente debido a fallas o sensores puestos a cero. Los promedios de productividad aumentaron de 2,67 t/ha a 3,67 t/ha (Parcela 1) y de 2,52 t/ha a 3,82 t/ha (Parcela 2), con la eliminación de valores extremos. Los mapas generados revelaron áreas críticas en los bordes y regiones con fallas. Se concluye que la herramienta es eficaz para automatizar la limpieza de datos de productividad; sin embargo, estudios futuros deberían considerar la inclusión de validación cruzada para reforzar la fiabilidad de los resultados.

Palabras clave: Agricultura de precisión. Análisis de datos. Mapa de productividad.

INTRODUCTION

Precision agriculture (PA) has been consolidated as an essential approach in the modernization of agribusiness, integrating advanced technologies to monitor and optimize agricultural operations. Sensors installed in agricultural machinery, such as harvesters, are capable of recording a vast amount of information per second, ranging from operational data to environmental conditions (Costa et al., 2015). This capacity for massive data collection allows detailed monitoring of the crop, enabling the identification of intra- and inter-field variabilities, which is essential for making more assertive decisions (Silva; Silva-Mann, 2020).

Modern harvesters are equipped with automatic productivity measurement systems, which integrate mass flow, humidity and geographic positioning sensors. These sensors collect data in real time during the harvesting operation, allowing the instant calculation of productivity based on the volume of grains harvested per unit area (Pereira; Molin, 2003; Li et al., 2005). The flow sensor, usually located in the clean grain elevator, estimates the harvested volume per time, while moisture sensors ensure adjustment to standard values. Combined with the GPS signal, this data is automatically recorded and stored, forming the basis of productivity maps.

However, the large amount of data generated presents significant challenges. In addition to volume, data quality is a central concern, since information collected by sensors may contain errors, noise, and inconsistencies, resulting from device failures, environmental interference, or operational problems (Fizza et al., 2022). These imperfections compromise the analysis of productivity and can lead to wrong decisions in agricultural management. In addition, duplicate records or information collected when the harvester is not in effective harvesting operation can distort crop yield calculations. To ensure the reliability of the information, it is essential to process and clean the raw data, filtering only those that really represent the real productivity of the field.

In this context, this study aims to develop and apply a Python script for the treatment of agricultural productivity data, applying it to a case study carried out in a rural property located in the municipality of Brasnorte in the state of Mato Grosso.

OBJECTIVES

GENERAL OBJECTIVE

Develop and apply a Python script for cleaning agricultural productivity data.

SPECIFIC OBJECTIVES

- Identify possible problems in the raw data;
- Apply cleaning and filtering techniques;
- Compare data before and after treatment.

LITERATURE REVIEW

Precision agriculture (PA) has stood out in the use of technologies aimed at collecting and processing agricultural data, enabling producers to adopt more efficient management strategies (Tschiedel; Ferreira, 2002). This methodology is based on the optimized application of agricultural inputs, adjusting their use according to the spatial and temporal variability of the crops (Silva; Silva-Mann, 2020). To this end, PA depends on capturing, storing, and analyzing large volumes of data, providing more assertive decision-making and significant gains in productivity (Bassoi et al., 2020).

The collection and processing of this information is made possible by a set of technologies, including sensors embedded in agricultural machinery, analysis software and automated equipment. These systems operate at different levels of automation, ranging from partially manual processes to fully autonomous solutions (Bassoi et al., 2020). In this context, sensors installed in agricultural machinery play an essential role in transmitting environmental information in real time, providing valuable data to the rural producer. This technology has been consolidated as a trend in the optimization of agricultural processes, allowing a more precise control of operations in the field (Costa et al., 2015).

In addition to sensors, productivity monitoring systems installed in harvesters play a key role in the automatic collection of data during harvesting. These systems are composed of mass flow sensors, moisture sensors, and GNSS receivers, which allow estimating productivity in real time, associating the harvested volume with the geographic position of the machine (Chandel et al., 2013). When properly calibrated, these monitors provide reliable information that supports the generation of detailed productivity maps (Grisso et al., 2002).

Despite technological advancements, the quality of agricultural data still faces significant challenges. Sensors can present problems such as noise, inconsistencies and errors in data collection, affecting the accuracy of analyses and, consequently, decision-making (Tschiedel; Ferreira, 2002). According to Menegatti and Molin (2004), raw files may contain positioning errors; zero, absent or discrepant income; irregular platform width; the filling interval; zero or no moisture and zero distance between points. In addition to these, Sudduth and Drummond (2007) also mentioned the emptying time of the bulk carrier; grain time delay in the harvester and rapid speed changes – as common errors in raw data.

To mitigate these limitations, data pre-processing emerges as an essential step, involving techniques such as cleaning, normalization, and transformation to ensure the quality of the information used (Maldaner; Molin; Spekken, 2022). In the context of PA, strategies such as the removal of outliers, filtering, and calibration adjustments are widely employed to correct inconsistencies in the data collected. The automation of this process, through scripts, has proven to be an efficient alternative to improve the accuracy of analyses, in addition to significantly reducing the time required for data preparation (Damico, 2025). Among the most used approaches in the cleaning of agricultural data, the filters of maximum and minimum harvest yield stand out, which, in some situations, are the only methods applied (Sudduth; Drummond, 2007). These filters are widely used in studies aimed at removing extreme values and improving data quality (Sudduth; Drummond, 2007; Gimenez; Molin, 2004; Vega et al., 2019).

Gimenez and Molin (2004) applied a filter, based on the average, to remove outliers from the productivity and cutting width sensors of the platform, n · average for higher values and n/me for smaller values. Sudduth and Drummond (2007) suggest a value close to 0 (zero) to remove data of low productivity and the value of maximum potential of the crop for the removal of high productivity values.

After data treatment, authors such as Santi et al. (2013) apply interpolation techniques to generate continuous productivity maps. Among the most used algorithms are the Inverse-Weighted Distance (IDW) in which the output is a raster layer containing values throughout the data (Chandel et al., 2013) and the Kriging highlighted for considering the spatial dependence between the sampled points (Bottega et al., 2012). The choice of the algorithm directly impacts the accuracy of the generated map, which will serve as a basis for localized management decisions.

Spatial dependence, in this context, refers to the correlation between values sampled in nearby geographic positions, and can be quantified by means of semivariograms or Moran's index. These characteristics are fundamental for the use of kriging and reinforce the need to understand the spatial behavior of the variable studied (Igaz et al., 2021).

To assess the quality of the generated models, cross-validation is often employed. This technique consists of dividing the data set into parts, using some to calibrate the model and others to test it, allowing the measurement of the prediction error and ensuring greater reliability in the interpretation of the generated maps. Igaz et al. (2021) used cross-validation to define the most reliable method in their research.

The quality of information is essential when maps represent the basis for the decision-making process. Errors resulting from automation can lead to misinterpretations,

presenting areas with unrealistic productivity (Menegatti; Molin, 2004). In this scenario, the Python programming language has stood out as a powerful tool in agricultural data science due to its flexibility and robustness. Libraries such as Pandas and NumPy allow the manipulation and analysis of large volumes of data, facilitating the automation of complex processes (Sapre; Vartak, 2020). Studies show that the application of Python, combined with other tools, can significantly improve efficiency in agricultural data processing (Damico, 2025).

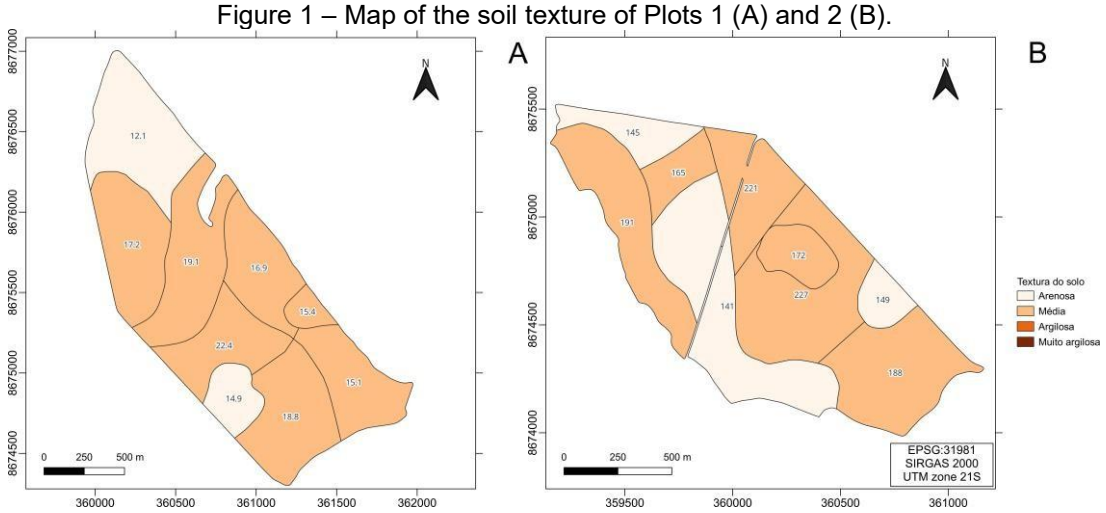
MATERIAL AND METHODS

CASE STUDY

The data used in this study were collected during the harvest in the 2024/25 soybean harvest, using six harvesters of the same brand, all with a nominal power of 378 hp and a storage capacity of 11,600 L. The machines were equipped with a load sensor scale and a 35-foot (10.66 m) Draper platform. The collection was carried out in two distinct plots: Plot 1 with an area of 234.86 ha, and Plot 2 with 154.15 ha. The crops are located on a farm in the municipality of Brasnorte, Mato Grosso, whose climatic classification, according to Köppen (1936), is Aw (tropical with a dry season in winter).

The soil texture map of the plots, referring to the layer from 0 to 10 cm (Figure 1), reveals a predominance of medium texture, with clay contents ranging between 15% and 35%. To a lesser extent, areas with a sandy texture are also identified, characterized by clay levels of less than 15% (Santos et al., 2018).

Figure 1 – Map of the soil texture of Plots 1 (A) and 2 (B).



The harvesters are equipped with a telemetry system responsible for the acquisition of raw data. The information was downloaded in Comma separated values (CSV) format. The attributes used in the CSV file can be seen in Table 1.

Table 1 – Information that must be contained in the CSV file.

Atributo	Descrição
cd_equipamento	Código do equipamento
cd_estado	Código de estado dos equipamentos
vl_rendimento_colheita	Valor de rendimento da colheita

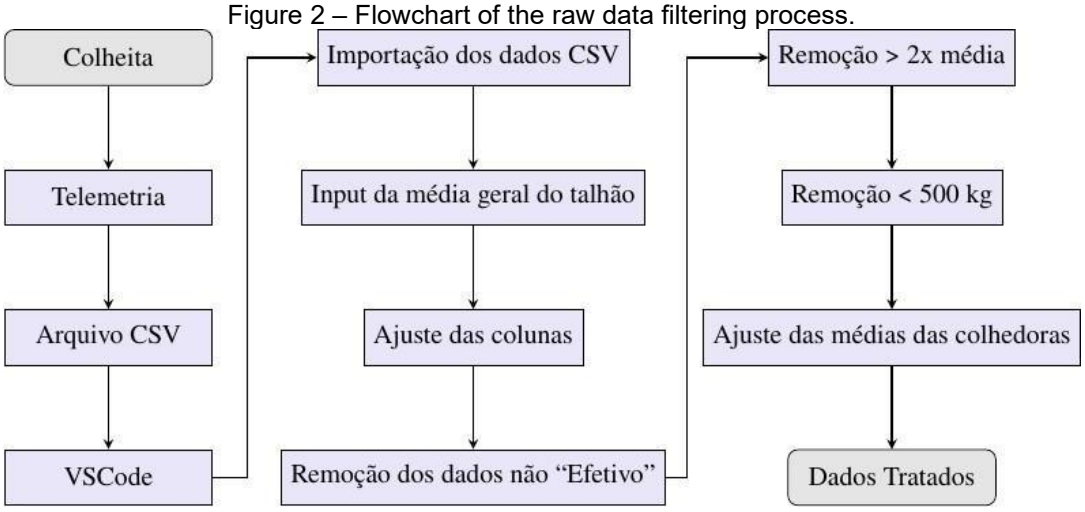
THE SCRIPT

To perform the data processing, a Python script (version 3.9.16) was developed using the Visual Studio Code (VS Code) software (version 1.98.2), a free and open source code editor developed by Microsoft. In VS Code, the user only needs to enter the path of the file containing the raw data of the field in CSV format and the average of its productivity, calculated based on the volume of grains stored after harvest.

The operation of the script depends on the import of libraries essential for processing the data. The Pandas library was used for several operations, including importing files in CSV format, organizing, cleaning, and statistical analysis of the data. The Matplotlib and Seaborn libraries were used to generate the graphs. In addition, the Tkinter and OS libraries, employed to develop an intuitive graphical interface, facilitated user interaction and optimized the process of selecting files directly in the operating system.

The developed script follows a structured sequence of steps (Figure 2). The data generated during harvesting is sent to the telemetry system, where it can be exported in CSV format. The code was implemented in the text editor VS Code, an environment in which the entire process of processing the raw data was developed.

Figure 2 – Flowchart of the raw data filtering process.



Inside the script, the libraries necessary for processing the data are initially imported. Then, the CSV file containing the field data is imported, as well as the definition of a variable intended to store the overall average productivity of the field, which was calculated based on the data obtained at the time of grain storage. Subsequently, adjustments are made to the formatting of columns, including the standardization of the decimal separator, the reorganization of the order of the columns, and the exclusion of those considered irrelevant to the analysis.

After this preparation step, the script proceeds to the data removal phase based on the information provided by the telemetry. The initial filtering considers the "cd_estado" field, which indicates the operational state of the machine at the time of acquisition of each point. During the collection, the system records, among other parameters, the type of activity being carried out by the harvester, such as stopping, maneuvering, displacement or harvesting itself. Table 2 presents the possible codes recorded during this process.

Table 2 – Description of the status codes of the equipment.

cd_estado	Descrição
B	Descarregamento
C	Desloc p/Descarregamento
D	Deslocamento
E	Efetivo
F	Parada
M	Manobra

In this work, only the lines whose "cd_estado" was equal to "Effective" were maintained, as they represent the moment when the machine is effectively harvesting. The other records were excluded, since they do not correspond to the harvesting operation and, therefore, may introduce noise and distortions in the productivity data.

The next step consisted of filtering the data related to the "vl_rendimento_co-lheita", which represents the amount of grains harvested in tons per hectare, a procedure commonly adopted in this type of analysis (Sudduth; Drummond, 2007). For this, records with values greater than twice the mean of the "vl_rendimento_colheita" column, as well as those less than 500 kg, were excluded, in order to eliminate possible inconsistencies in the data. The value of 500 kg was adopted in agreement with the farm managers.

The last step of the data pre-processing was the adjustment of the means. To standardize the data of all harvesters, an adjustment was applied to the "vl_rendimento_colheita" of each machine based on the overall average of the field. The code performs an iterative adjustment of the crop yield values for each piece of equipment,

ensuring homogenization of the data against the average, within a predefined margin of error. This adjustment aims to correct possible variations in the values recorded by the sensors, considering that, during the harvest of the same field (Figure 3), regardless of its size, the machines operate in close proximity.

Figure 3 – Trace of the harvesters working in the two fields.

The details of the code used to adjust the averages are described below:

- Iteration over equipment: The code first identifies all equipment in "cd_equipamento" data.
- Iterative fitting: For each piece of equipment, the code enters a while loop that continues until the average throughput of the equipment is within tolerance with respect to the overall average.
- Adjustment Factor Calculation: Within the loop, the script calculates an adjustment factor based on the comparison between the equipment's average performance and the overall average. If the average of the equipment is lower than the general average, the factor increases the performance values. If it is higher, the factor decreases the values. If they are equal, the factor is 1 (no change).
- Factor application: The code multiplies the "vl_rendimento_colheita" values of the equipment by the calculated factor, thus adjusting the yield.
- Tolerance check: After adjustment, the code checks whether the absolute difference between the adjusted yield average and the overall average is within tolerance. If it is, the loop ends. Otherwise, the adjustment process continues.

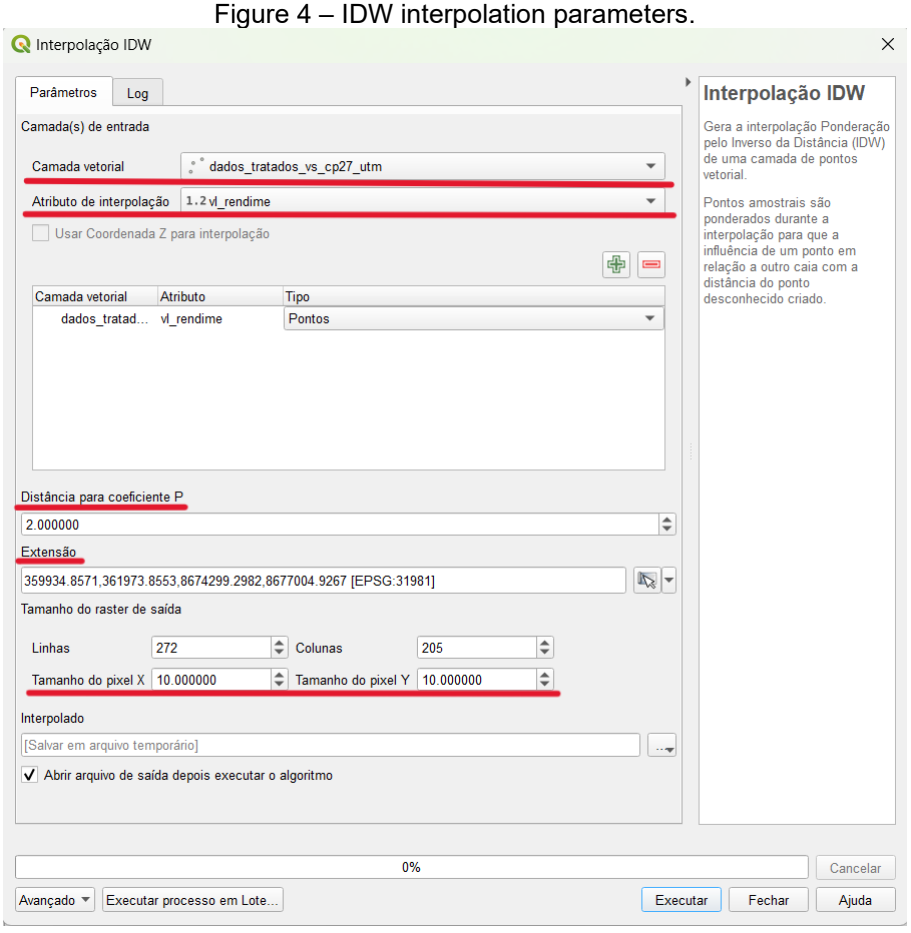
Finally, scrypt saves the data by generating a new CSV file that can be opened in a Geographic Information System (GIS), used to view, edit, and analyze geographic data.

DATA VISUALIZATION

For visualization/generation of the productivity maps, QGIS, a free and open source GIS, was used. The interpolation of the data was done by the IDW method, a technique used to estimate values in unsampled locations based on known data points, by accessing the "Processing Toolbox" in QGIS and typing in the search field "IDW Interpolation".

Figure 4 shows the parameters used to compose the productivity map. The vector layer is the processed data (CSV file) generated by the script and transformed into a shapefile in UTM (Universal Transverse Mercator) coordinates by QGIS itself; the interpolation attribute is the column that represents the crop yield; the distance to the coefficient P was maintained the interpolator standard; the extension was a shapefile of the field boundary and the size of the output raster was based on the resolution of the X and Y pixels of 10 m.

Figure 4 – IDW interpolation parameters.



RESULTS AND DISCUSSION

Table 3 presents the descriptive statistics of the yield data in the two plots, before and after the cleaning process and adjustment of the averages. An increase in the mean was observed in both plots after treatment, which indicated that the raw data contained a significant amount of values below the mean. Other studies also obtained an increase in the mean after cleaning the data in the software they developed (Sudduth; Drummond, 2007; Gimenez; Molin, 2004). The median of zero in both cases of the raw data indicates that at least 50% of the performance records from telemetry had a value equal to zero. In addition, there was a reduction in the standard deviation (St dev), the same happened in the studies by Sudduth and Drummond (2007), Gimenez and Molin (2004), indicating greater uniformity in the treated data. The maximum values fell from 44.8 t/ha and 42.2 t/ha to 8.90 t/ha and 9.43 t/ha, respectively, evidencing the effective removal of outliers and resulting in a more homogeneous and reliable dataset for further analysis.

Table 3 – Crop yield data statistics.

Estatística	Dados Brutos	Dados Tratados	Dados Brutos	Dados Tratados
	Talhão 1		Talhão 2	
Contagem	23291	9595	21055	6977
Média	2,674	3,672	2,520	3,823
Mediana	0,000	3,985	0,000	4,033
St dev	3,469	1,373	3,810	1,484
Mínimo	0,000	0,397	0,000	0,381
Máximo	44,800	8,899	42,200	9,432
Q1	0,000	2,989	0,000	3,068
Q3	5,200	4,544	4,900	4,717
IQR	5,200	1,556	4,900	1,649

Another relevant observation refers to the significant reduction in the number of data after the cleaning process: approximately 58.8% in Plot 1 and 66.9% in Plot 2 were discarded. For comparison purposes, in the study conducted by Gimenez and Molin (2004), the discard rates were 41% and 21%, while Sudduth and Drummond (2007) reported the removal of 13% to 27% of the data in five plots with areas ranging from 11 ha to 48 ha, using the software developed by them.

The highest disposal rate observed in this work may be related to the occurrence of failures in one of the sensors of one of the harvesters, which resulted in the registration of yield equal to zero at all points collected by this equipment. Another relevant fact was the exclusion of data with the "c_estado" other than the number of personnel, 35.4% of the data from Plot 1 and 39.9% from Plot 2 were removed, in some telemetry systems this data does not even appear in the raw data.

Tables 4 and 5 show the amount of data excluded after the cleaning process, ranging from 42.23% to 78.27% for Plot 1 and from 46.49% to 80.08% for Plot 2, excluding equipment 266 that was lost 100% of the data. With the result, it was possible to verify that some sensors may be having problems or out of calibration. Another piece of information is the averages, yield and moisture, per machine at each stage of processing.

Table 4 – Data from the harvesters before, during and after treatment (Plot 1).

	Dados Brutos			Dados Tratados sem ajuste da média			Dados Tratados com ajuste da média
cd_ equipa-mento	Contagem	vl _rendimento _colheita	vl _umidade _graos	Contagem	vl _rendimento _colheita	vl _umidade _graos	vl _rendimento _colheita
86	5288	3,152	13,16	2366	6,649	17,64	3,680
88	2214	2,473	10,82	1271	4,231	14,79	3,682
96	4284	2,647	15,16	2475	4,450	19,33	3,677
98	4842	1,890	5,54	2628	3,370	6,89	3,657
269	3935	4,991	14,63	855	4,616	18,99	3,669
266	2728	0,000	13,87	-	-	-	-

Table 5 – Data from harvesters before, during and after treatment (Plot 2).

	Dados Brutos			Dados Tratados sem ajuste da média			Dados Tratados com ajuste da média
cd_ equipa-mento	Contagem	vl _rendimento _colheita	vl _umidade _graos	Contagem	vl _rendimento _colheita	vl _umidade _graos	vl _rendimento _colheita
86	3515	2,970	17,76	1410	6,766	22,13	3,790
88	3394	2,550	11,04	1801	4,631	15,59	3,832
96	2947	1,569	7,60	903	4,997	19,46	3,808
98	3835	1,899	5,56	2052	3,378	6,93	3,839
269	4071	5,418	16,23	811	4,367	20,09	3,838
266	3293	0,000	16,02	-	-	-	-

The productivity maps generated from the raw data, downloaded directly from the farm's telemetry system, are presented in Figures 5 and 6.

Figure 5 – Productivity map at points in Plot 1 (raw data).

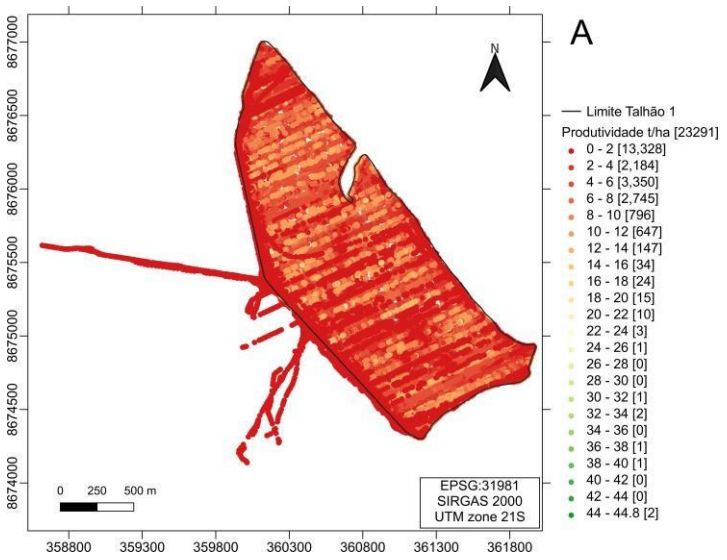
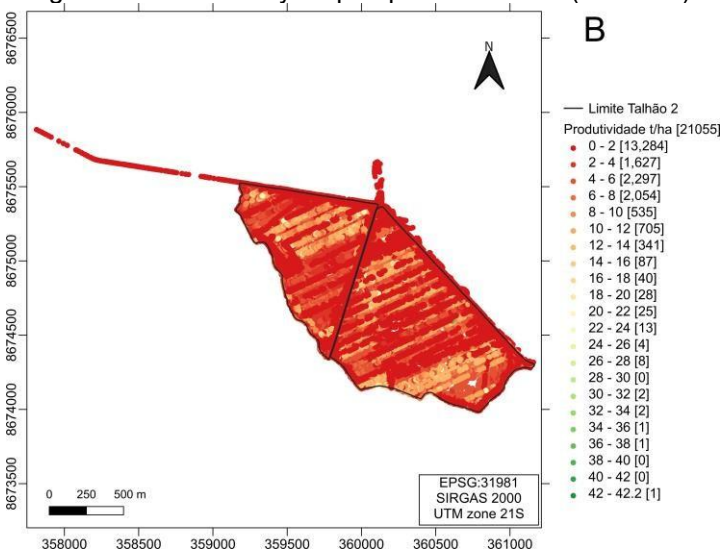


Figure 6 – Productivity map at points in Plot 2 (raw data).



It was observed, in both plots, the presence of points outside the geographical limits of the harvested areas. In the study conducted by Menegatti and Molin (2004), positioning errors represented 0.1% to 7.8% of the data, evidencing the recurrence of this type of inconsistency. There is also a significant concentration of data with values below 2.0 t/ha, representing 57.22% of the points in Plot 1 and 63.09% in Plot 2. As a consequence of these distortions, the average yields calculated based on the raw data were only 2.67 t/ha and 2.52 t/ha, respectively, values significantly lower than the actual averages obtained from the weighing of the grains in the warehouse, 3.68 t/ha and 3.84 t/ha, already adjusted to 13% moisture.

Figures 7 and 8 present the productivity maps generated after the data treatment using the proposed script.

Figure 7 – Productivity map at points in Plot 1 (data treated).

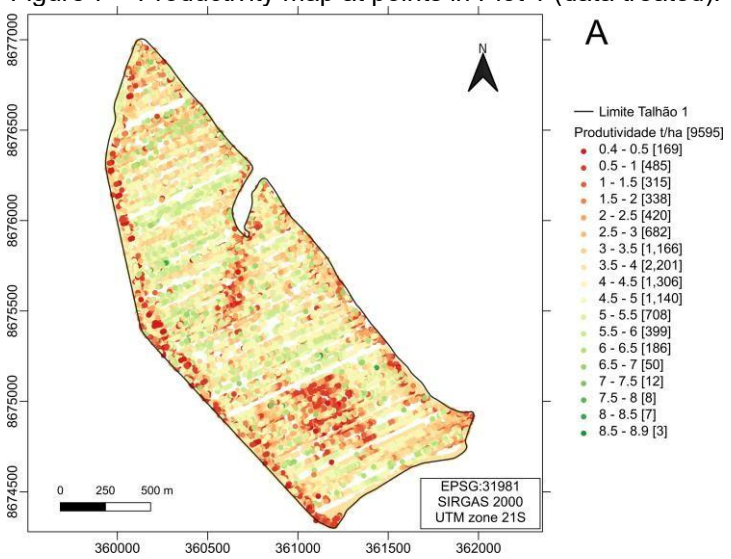
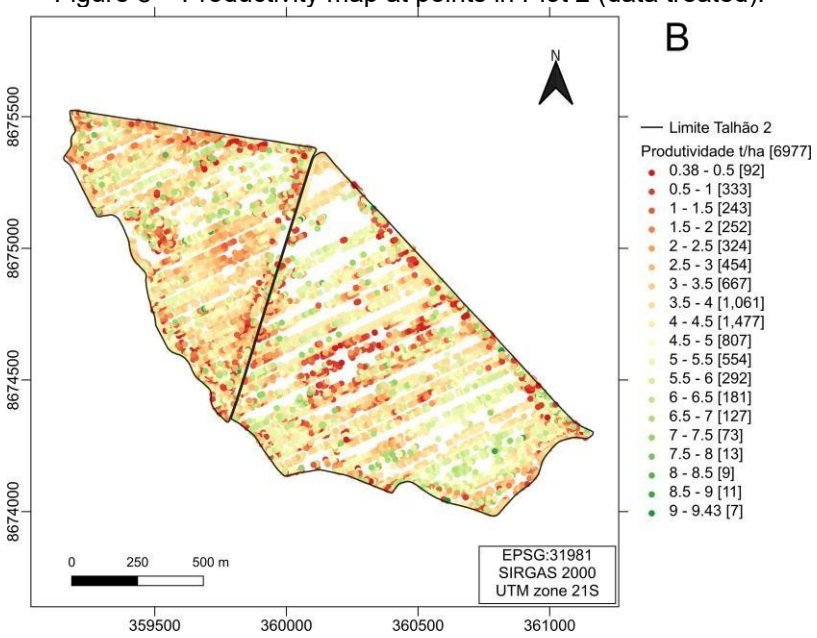


Figure 8 – Productivity map at points in Plot 2 (data treated).



Points located outside the boundaries of the plots were excluded, not by means of a specific geographical criterion, but because they presented values below the minimum limit defined for yield.

It was observed, especially in the edges of the plots, the presence of areas with low productivity. In Plot 1, this occurs mainly in the lower and left portions, while in Plot 2, it is observed in the upper region — precisely on the border between the two plots, where there is a municipal road. It is likely that this area functions as a maneuvering and stopping area for machines at the end of the day, which may have caused greater soil compaction and, consequently, a drop in productivity.

Another notable aspect is the "flaws" or gaps visible in some regions of the maps. This is due, in addition to the natural reduction of points caused by the cleaning of the data, to the fact that one of the harvesters recorded all yield values as zero, resulting in ranges without valid data.

Even before the application of interpolation techniques, the maps already allowed the identification of the zones with lower productivity, highlighting the potential of the script to provide a more reliable initial visualization of the spatial variability of crops.

Finally, the productivity map was generated after performing the IDW interpolation (Figures 9 and 10), highlighting the low areas in red tones and high productivity green tones. These maps are used to generate the management zones along with others.

Figure 9 – Interpolated productivity map of Plot 1.

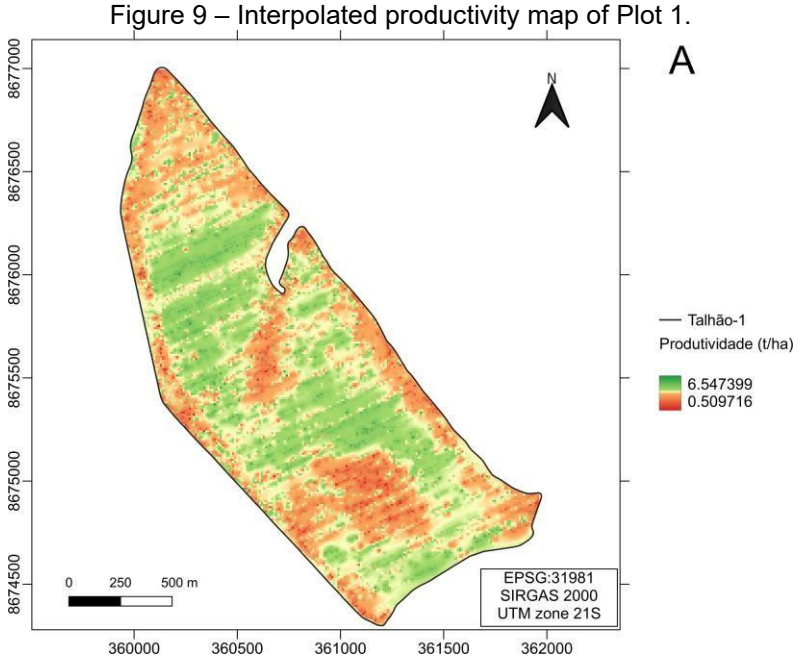
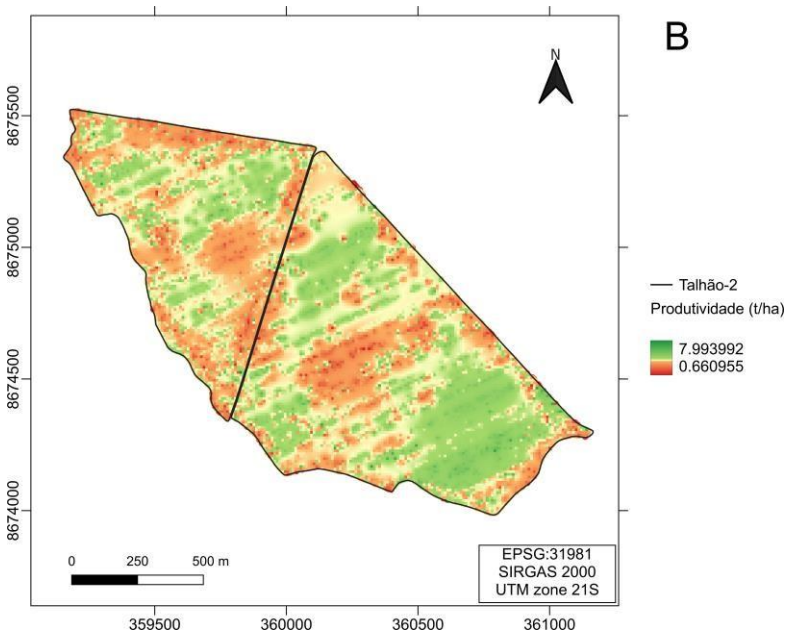


Figure 10 – Interpolated productivity map of Plot 2.



CONCLUSION

The development and application of the Python script to clean the raw data of agricultural productivity proved to be effective in the treatment of information collected via telemetry. The proposal contributed to the automation of the pre-processing process, reducing interferences caused by errors in the raw data and increasing the reliability of the analyses.

The tool proved to be an accessible and efficient alternative, allowing greater control and standardization in the preparation of data for the generation of productivity maps. The application of this type of solution is essential to ensure the quality of information in PA projects, supporting more assertive agronomic decisions.

Compared to the raw data, the processed data provide information more consistent with the reality of the fields, allowing better decision-making for managers, enabling the use of the interpolated map for variable-rate applications.

REFERENCES

1. Basso, L. H., & et al. (2020). Precision agriculture and digital agriculture. *TECCOGS: Digital Journal of Cognitive Technologies, 20*. Retrieved July 3, 2025, from https://revistas.pucsp.br/teccogs/article/view/48542
2. Bottega, E., & et al. (2012). Use of different interpolators in the generation of digital elevation model. In *Brazilian Congress of Precision Agriculture - CONBAP, 2012, Ribeirão Preto* (pp. [page range if available]). Ribeirão Preto: [Publisher if known].
3. Chandel, N., & et al. (2013). IDW interpolation of soybean yield data acquired by automated yield monitor. *International Journal for Science and Emerging Technologies with Latest Trends, 13*, 36–45.
4. Costa, F., & et al. (2015). An overview of the application of sensors in agricultural machinery. In *Congresso Brasileiro de Agroinformática, 10, 2015, Ponta Grossa* (pp. [page range if available]). Ponta Grossa: [Publisher if known].
5. Damico, J. (2025). *Methodology for cleaning, quality, and normalization of vegetation index data derived from Sentinel 2*. Retrieved July 3, 2025, from https://www.academia.edu/126945828/Methodology_for_Cleaning_Quality_and_Normalization_of_Vegetation_Index_Data_Derived_from_Sentinel_2
6. Fizza, K., & et al. (2022). Evaluating sensor data quality in internet of things smart agriculture applications. *IEEE Micro, 42*(1), 51–60. https://doi.org/10.1109/MM.2021.3137401
7. Gimenez, L. M., & Molin, J. P. (2004). Algorithm for error reduction in productivity maps for precision agriculture. *Revista Brasileira de Agrocomputação, 2*(1), 5–10.
8. Grisso, R. D., & et al. (2002). Yield monitor accuracy: Successful farming magazine case study. *Applied Engineering in Agriculture, 18*(2), 147.
9. Igaz, D., & et al. (2021). The evaluation of the accuracy of interpolation methods in crafting maps of physical and hydro-physical soil properties. *Water, 13*, 212.
10. Köppen, W. (1936). *Das geographische System der Klimate* (Vol. 1C, pp. 1–44). Berlin: Gebrüder Borntraeger.
11. Li, M., & et al. (2005). Development of an intelligent yield monitor for grain combine harvester. In D. Li & B. Wang (Eds.), *Artificial intelligence applications and innovations* (pp. 663–670). New York: Springer-Verlag. https://doi.org/10.1007/0-387-29295-0_72
12. Maldaner, L. F., Molin, J. P., & Spekken, M. (2022). Methodology to filter out outliers in high spatial density data to improve maps reliability. *Scientia Agricola, 79*(1). Retrieved July 3, 2025, from https://www.scielo.br/j/sa/a/
13. Menegatti, L. A. A., & Molin, J. P. (2004). Error removal in productivity maps via raw data filtering. *Brazilian Journal of Agricultural and Environmental Engineering, 8*(1), 126–134. Retrieved July 3, 2025, from https://www.scielo.br/j/rbeaa/a/
14. Pereira, F. J., & Molin, J. P. (2003). Test bench for evaluation of grain yield monitors. *Engenharia Agrícola, 23*(3), 568–578.
15. Santi, A., & et al. (2013). Definition of productivity zones in areas managed with precision agriculture. *Revista Brasileira de Ciências Agrárias, 8*(3), 510–515. Retrieved July 3, 2025, from http://www.agraria.pro.br/ojs32/index.php/RBCA/article/view/v8i3a2489
16. Santos, H. G., & et al. (2018). *Brazilian soil classification system*. Brasília: Embrapa.
17. Sapre, A., & Vartak, S. (2020). Scientific computing and data analysis using NumPy and Pandas. *[Journal Name if known], 7*(12).
18. Silva, W. V. R., & Silva-Mann, R. (2020). Precision agriculture in Brazil: Current situation, challenges and perspectives. *Research, Society and Development, 9*(11), e1979119603. Retrieved July 3, 2025, from https://rsdjournal.org/index.php/rsd/article/view/9603

19. Sudduth, K. A., & Drummond, S. T. (2007). Yield editor: Software for removing errors from crop yield maps. *Agronomy Journal, 99*(6), 1471–1482. https://doi.org/10.2134/agronj2006.0326
20. Tschiedel, M., & Ferreira, M. F. (2002). Introduction to precision agriculture: Concepts and advantages. *Ciência Rural, 32*(1), 159–163. https://doi.org/10.1590/S0103-84782002000100027
21. Vega, A., & et al. (2019). Protocol for automating error removal from yield maps. *Precision Agriculture, 20*(5), 1030–1044. https://doi.org/10.1007/s11119-018-09632-8

.