



UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS AGRÁRIAS
DEPARTAMENTO DE ENGENHARIA AGRÍCOLA

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

São Cristóvão – Sergipe

2025

Raphael Prazeres da Silva

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

Trabalho de Conclusão de Curso submetido ao Departamento de Engenharia Agrícola da Universidade Federal de Sergipe como requisito parcial à obtenção do título de bacharel em Engenharia Agrícola.

Orientador: Prof. DSc. Weligton Gonzaga do Vale

São Cristóvão – Sergipe

2025

Raphael Prazeres da Silva

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

Trabalho de Conclusão de Curso submetido ao Departamento de Engenharia Agrícola da Universidade Federal de Sergipe como requisito parcial à obtenção do título de bacharel em Engenharia Agrícola.

Trabalho aprovado. São Cristóvão – Sergipe, 14 de Abril de 2025:

Orientador

Prof. DSc. Weligton Gonzaga do Vale

Avaliador

Prof. DSc. Adilson Machado Enes

Avaliador

MSc. Valfran José Andrade

Avaliadora

Eng. Agrícola Janyelle do Nascimento Silva

São Cristóvão – Sergipe

2025

*Insista,
persista,
mas não desista.
(Alemão "in memoriam")*

Agradecimentos

Neste momento de significativa conquista, meu coração transborda gratidão. Primeiramente, elevo meus agradecimentos a Deus, força motriz que me impulsiona a cada desafio.

A vocês, meu amado pai, minha querida mãe e meu estimado irmão, minha eterna gratidão pelo amor incondicional, pelo apoio constante e por serem o alicerce da minha vida. Cada passo dado foi fortalecido pela presença de vocês.

Reconheço com carinho e apreço meus preciosos amigos Eduardo, Flavia, André, Emerson e Janyelle. A amizade sincera e os momentos de alegria e suporte foram essenciais para tornar esta caminhada mais leve e memorável.

Aos dedicados professores do curso de Engenharia Agrícola, minha profunda admiração pela paixão em compartilhar conhecimento, pela orientação valiosa e por moldarem meu desenvolvimento profissional e intelectual ao longo destes anos. Aqui deixo o meu agradecimento ao meu orientador Weligton Gonzaga do Vale.

Expressar minha gratidão à mãe da minha doce Melinda é fundamental. Seu incentivo em 2017 para realizar o ENEM e sua ajuda na escolha do curso foram decisivos para que eu chegasse até aqui.

Um agradecimento especial e cheio de amor à minha inspiradora filha, Melinda Reis. Seu nascimento durante esta graduação transformou minha perspectiva e me deu ainda mais motivação para seguir em frente. "Papai te ama!"

Também sou grato a todos os técnicos do departamento de Deagri e aos meus queridos colegas de curso. A troca de experiências, a colaboração nos estudos e o companheirismo foram inestimáveis.

Estendo meus sinceros agradecimentos à Agropecuária Crestani, empresa onde tive a oportunidade de realizar meu estágio acadêmico. A experiência prática e o aprendizado profissional foram um diferencial crucial na minha formação. Muito obrigado ao Alessandro por todos os ensinamentos.

A todos que, direta ou indiretamente, contribuíram para a realização deste sonho, o meu mais sincero reconhecimento. Gratidão por cada palavra de apoio, cada gesto de incentivo, importante nesta trajetória.

Minha jornada no curso de Engenharia Agrícola, iniciada em 2018 e concluída em 2025, é dedicada a cada um de vocês que tornaram este percurso possível e especial.

O meu muito obrigado.

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.

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.

Lista de ilustrações

Figura 1 – Mapa da textura do solo dos Talhões 1 (A) e 2 (B).	14
Figura 2 – Fluxograma do processo de filtragem dos dados brutos.	15
Figura 3 – Rastro das colhedoras trabalhando nos dois talhões.	17
Figura 4 – Parâmetros da interpolação IDW.	18
Figura 5 – Mapa de produtividade em pontos do Talhão 1 (dados brutos).	20
Figura 6 – Mapa de produtividade em pontos do Talhão 2 (dados brutos).	21
Figura 7 – Mapa de produtividade em pontos do Talhão 1 (dados tratados).	21
Figura 8 – Mapa de produtividade em pontos do Talhão 2 (dados tratados).	22
Figura 9 – Mapa de produtividade interpolado do Talhão 1.	23
Figura 10 – Mapa de produtividade interpolado do Talhão 2.	23

Lista de tabelas

Tabela 1 – Informações que devem conter no arquivo CSV.	15
Tabela 2 – Descrição dos códigos de estado dos equipamentos.	16
Tabela 3 – Estatísticas dos dados de rendimento da colheita.	19
Tabela 4 – Dados das colhedoras antes, durante e após tratamento (Talhão 1).	20
Tabela 5 – Dados das colhedoras antes, durante e após tratamento (Talhão 2).	20

Lista de abreviaturas e siglas

AP	Agricultura de Precisão
CSV	<i>Comma separated values</i>
IDW	Distância Inversa Ponderada
SIG	Sistema de Informação Geográfica
VS Code	<i>Visual Studio Code</i>

Sumário

1	INTRODUÇÃO	11
2	OBJETIVOS	11
2.1	Objetivo Geral	11
2.2	Objetivos Específicos	12
3	REVISÃO DE LITERATURA	12
4	MATERIAL E MÉTODOS	14
4.1	Estudo de caso	14
4.2	O <i>Script</i>	15
4.3	Visualização dos dados	18
5	RESULTADOS E DISCUSSÃO	18
6	CONCLUSÃO	23
	REFERÊNCIAS	25

1 INTRODUÇÃO

A agricultura de precisão (AP) tem se consolidado como uma abordagem essencial na modernização do agronegócio, integrando tecnologias avançadas para monitorar e otimizar as operações agrícolas. Sensores instalados em máquinas agrícolas, como colhedoras, são capazes de registrar uma vasta quantidade de informações por segundo, abrangendo desde dados operacionais até condições ambientais (Costa *et al.*, 2015). Essa capacidade de coleta massiva de dados permite um monitoramento detalhado da lavoura, possibilitando a identificação de variabilidades intra e entre talhões, o que é fundamental para a tomada de decisões mais assertivas (Silva; Silva-Mann, 2020).

As colhedoras modernas estão equipadas com sistemas automáticos de medição de produtividade, que integram sensores de fluxo de massa, umidade e posicionamento geográfico. Esses sensores coletam dados em tempo real durante a operação de colheita, permitindo o cálculo instantâneo da produtividade com base no volume de grãos colhido por unidade de área (Pereira; Molin, 2003; Li *et al.*, 2005). O sensor de fluxo, geralmente localizado no elevador de grãos limpos, estima o volume colhido por tempo, enquanto sensores de umidade garantem o ajuste para valores padrão. Combinados ao sinal de GPS, esses dados são registrados e armazenados automaticamente, formando a base dos mapas de produtividade.

Entretanto, a grande quantidade de dados gerados apresenta desafios significativos. Além do volume, a qualidade dos dados é uma preocupação central, uma vez que informações coletadas por sensores podem conter erros, ruídos e inconsistências, decorrentes de falhas nos dispositivos, interferências ambientais ou problemas operacionais (Fizza *et al.*, 2022). Essas imperfeições comprometem a análise da produtividade e podem levar a decisões equivocadas no manejo agrícola. Além disso, registros duplicados ou informações coletadas quando a colhedora não está em operação efetiva de colheita podem distorcer os cálculos de rendimento da lavoura. Para garantir a confiabilidade das informações, é essencial realizar um processamento e limpeza dos dados brutos, filtrando apenas aqueles que realmente representam a produtividade real do campo.

Diante desse contexto, este estudo tem como objetivo desenvolver e aplicar um *script* em Python para o tratamento de dados da produtividade agrícola, aplicando-o a um estudo de caso realizado em uma propriedade rural localizada no município de Brasnorte no estado do Mato Grosso.

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver e aplicar um *script* em Python para limpeza de dados de produtividade agrícola.

2.2 Objetivos Específicos

- Identificar possíveis problemas nos dados brutos;
- Aplicar técnicas de limpeza e filtragem;
- Comparar os dados antes e depois do tratamento.

3 REVISÃO DE LITERATURA

A agricultura de precisão (AP) tem se destacado no uso de tecnologias voltada para a coleta e processamento de dados agrícolas, possibilitando que os produtores adotem estratégias de manejo mais eficientes (Tschiedel; Ferreira, 2002). Essa metodologia se fundamenta na aplicação otimizada de insumos agrícolas, ajustando seu uso de acordo com a variabilidade espacial e temporal das lavouras (Silva; Silva-Mann, 2020). Para isso, a AP depende da captação, armazenamento e análise de grandes volumes de dados, proporcionando tomadas de decisão mais assertivas e ganhos significativos na produtividade (Basso *et al.*, 2020).

A obtenção e o processamento dessas informações são viabilizados por um conjunto de tecnologias, incluindo sensores embarcados em máquinas agrícolas, softwares de análise e equipamentos automatizados. Esses sistemas operam em diferentes níveis de automação, variando desde processos parcialmente manuais até soluções totalmente autônomas (Basso *et al.*, 2020). Nesse contexto, sensores instalados em máquinas agrícolas desempenham um papel essencial ao transmitir informações ambientais em tempo real, fornecendo dados valiosos para o produtor rural. Essa tecnologia tem se consolidado como uma tendência na otimização dos processos agrícolas, permitindo um controle mais preciso das operações no campo (Costa *et al.*, 2015).

Além dos sensores, os sistemas de monitoramento de produtividade instalados em colhedoras têm papel fundamental na coleta automática de dados durante a colheita. Esses sistemas são compostos por sensores de fluxo de massa, sensores de umidade e receptores GNSS, os quais permitem estimar a produtividade em tempo real, associando o volume colhido à posição geográfica da máquina (Chandel *et al.*, 2013). Quando devidamente calibrados, esses monitores oferecem informações confiáveis que subsidiam a geração de mapas de produtividade detalhados (Grisso *et al.*, 2002).

Apesar dos avanços tecnológicos, a qualidade dos dados agrícolas ainda enfrenta desafios significativos. Sensores podem apresentar problemas como ruídos, inconsistências e erros na coleta de dados, afetando a precisão das análises e, conseqüentemente, a tomada de decisões (Tschiedel; Ferreira, 2002). Segundo Menegatti e Molin (2004) os arquivos brutos podem conter erros de posicionamento; rendimento nulo, ausente ou discrepantes; largura da plataforma irregular; do intervalo de enchimento; umidade nula ou ausente e distância nula entre pontos. Além desses, Sudduth e Drummond (2007) também citaram o tempo de esvaziamento do

graneleiro; atraso de tempo do grão na colhedora e mudanças rápidas de velocidade - como erros comuns em dados brutos.

Para mitigar essas limitações, o pré-processamento de dados surge como uma etapa essencial, envolvendo técnicas como limpeza, normalização e transformação para garantir a qualidade das informações utilizadas (Maldaner; Molin; Spekken, 2022). No contexto da AP, são amplamente empregadas estratégias como a remoção de *outliers*, filtragem e ajustes de calibração para corrigir inconsistências nos dados coletados. A automação desse processo, por meio de *scripts*, tem demonstrado ser uma alternativa eficiente para melhorar a precisão das análises, além de reduzir significativamente o tempo necessário para a preparação dos dados (Damico, 2025). Entre as abordagens mais utilizadas na limpeza de dados agrícolas, destacam-se os filtros de máximo e mínimo rendimento da colheita, que, em algumas situações, são os únicos métodos aplicados (Sudduth; Drummond, 2007). Esses filtros são amplamente empregados em estudos voltados para a remoção de valores extremos e o aprimoramento da qualidade dos dados (Sudduth; Drummond, 2007; Gimenez; Molin, 2004; Vega *et al.*, 2019).

Gimenez e Molin (2004) aplicaram um filtro, com base na média, para remoção de dados discrepantes (*outliers*) dos sensores de produtividade e largura de corte da plataforma, $n \cdot média$ para valores maiores e $n/média$ para valores menores. Sudduth e Drummond (2007) sugerem um valor próximo de 0 (zero) para remover dados de produtividade baixa e o valor de máximo potencial da cultura para a remoção dos valores de alta produtividade.

Após o tratamento dos dados, autores como Santi *et al.* (2013) aplicam técnicas de interpolação para a geração de mapas contínuos de produtividade. Entre os algoritmos mais utilizados estão a Distância Inversa Ponderada (IDW) em que a saída é uma camada raster contendo valores em toda a extensão dos dados (Chandel *et al.*, 2013) e a Krigagem destacada por considerar a dependência espacial entre os pontos amostrados (Bottega *et al.*, 2012). A escolha do algoritmo impacta diretamente na precisão do mapa gerado, que servirá como base para decisões de manejo localizado.

A dependência espacial, nesse contexto, refere-se à correlação existente entre valores amostrados em posições geográficas próximas, e pode ser quantificada por meio de semivariogramas ou índice de Moran. Essas características são fundamentais para o uso da Krigagem e reforça a necessidade de entender o comportamento espacial da variável estudada (Igaz *et al.*, 2021).

Para avaliar a qualidade dos modelos gerados, a validação cruzada é frequentemente empregada. Essa técnica consiste em dividir o conjunto de dados em partes, utilizando algumas para calibrar o modelo e outras para testá-lo, permitindo mensurar o erro de predição e garantir maior confiabilidade na interpretação dos mapas gerados. Igaz *et al.* (2021) utilizaram a validação cruzada para definir o método mais confiável em sua pesquisa.

A qualidade da informação é essencial quando os mapas representam a base de dados para o processo de tomada de decisão. Erros decorrentes da automatização podem levar a interpretações equivocadas, apresentando áreas com produtividades irreais (Menegatti; Molin,

2004). Nesse cenário, a linguagem de programação Python tem se destacado como uma ferramenta poderosa na ciência de dados agrícolas devido à sua flexibilidade e robustez. Bibliotecas como Pandas e NumPy permitem a manipulação e análise de grandes volumes de dados, facilitando a automação de processos complexos (Sapre; Vartak, 2020). Estudos demonstram que a aplicação do Python, combinada a outras ferramentas, pode aprimorar significativamente a eficiência no processamento de dados agrícolas (Damico, 2025).

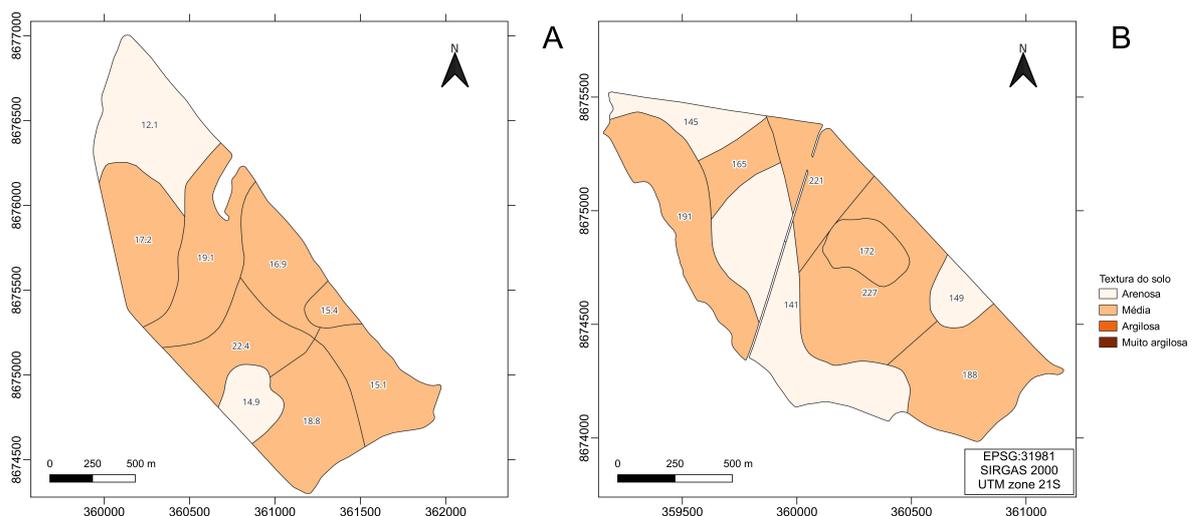
4 MATERIAL E MÉTODOS

4.1 Estudo de caso

Os dados utilizados neste estudo foram coletados durante a colheita na safra da soja 2024/25, utilizando seis colhedoras da mesma marca, todas com potência nominal de 378 cv e capacidade de armazenamento de 11.600 L. As máquinas estavam equipadas com balança por sensor de carga e plataforma *Draper* de 35 pés (10,66 m). A coleta foi realizada em dois talhões distintos: o Talhão 1 com uma área de 234,86 ha, e o Talhão 2 com 154,15 ha. As lavouras estão localizadas em uma fazenda no município de Brasnorte, Mato Grosso, cuja classificação climática, segundo Köppen (1936), é Aw (tropical com estação seca no inverno).

O mapa de textura do solo dos talhões, referente à camada de 0 a 10 cm (Figura 1), revela uma predominância de textura média, com teores de argila variando entre 15% e 35%. Em menor proporção, também são identificadas áreas com textura arenosa, caracterizadas por teores de argila inferiores a 15% (Santos *et al.*, 2018).

Figura 1 – Mapa da textura do solo dos Talhões 1 (A) e 2 (B).



As colhedoras são equipadas com um sistema de telemetria responsável pela aquisição dos dados brutos. As informações foram baixadas no formato *Comma separated values* (CSV). Os atributos utilizados do arquivo CSV podem ser visto na Tabela 1.

Tabela 1 – Informações que devem conter no arquivo CSV.

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

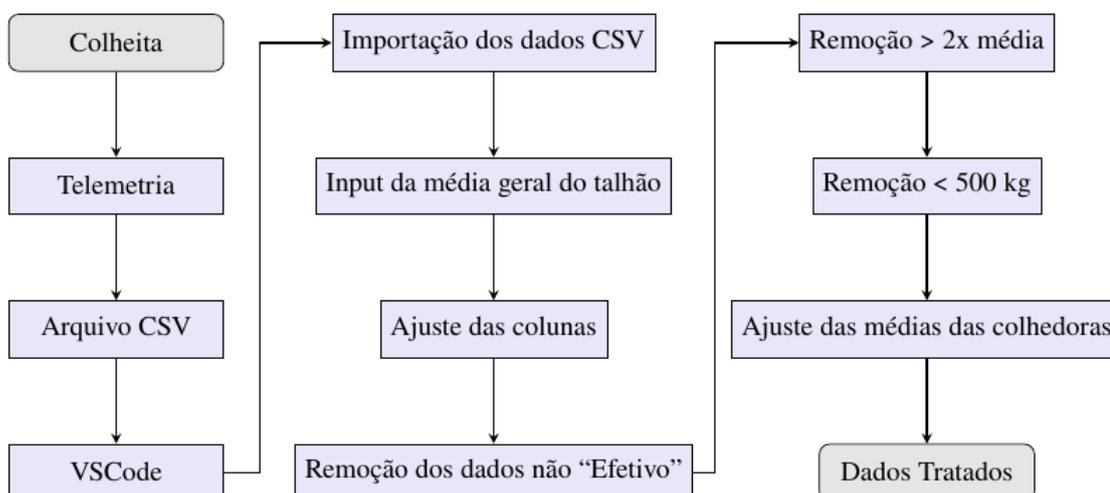
4.2 O Script

Para a realização do processamento dos dados, desenvolveu-se um *script* em Python (versão 3.9.16) utilizando o *software Visual Studio Code* (VS Code) (versão 1.98.2), um editor de código-fonte gratuito e de código aberto desenvolvido pela Microsoft. No VS Code, o usuário precisa apenas informar o caminho do arquivo contendo os dados brutos do talhão no formato CSV e a média da produtividade do mesmo, calculada com base no volume de grãos armazenado após a colheita.

O funcionamento do *script* depende da importação de bibliotecas essenciais para o processamento dos dados. A biblioteca Pandas foi utilizada para diversas operações, incluindo a importação de arquivos no formato CSV, organização, limpeza e análise estatística dos dados. As bibliotecas Matplotlib e Seaborn foram utilizadas para geração dos gráficos. Além disso, as bibliotecas Tkinter e OS, empregadas para desenvolver uma interface gráfica intuitiva, facilitou a interação do usuário e otimizou o processo de seleção dos arquivos diretamente no sistema operacional.

O *script* desenvolvido segue uma sequência estruturada de etapas (Figura 2). Os dados gerados durante a colheita são enviados ao sistema de telemetria, onde podem ser exportados no formato CSV. Implementou-se o código no editor de texto VS Code, ambiente o qual desenvolveu-se todo o processo de tratamento dos dados brutos.

Figura 2 – Fluxograma do processo de filtragem dos dados brutos.



No interior do *script*, inicialmente são importadas as bibliotecas necessárias para o processamento dos dados. Em seguida, realiza-se a importação do arquivo CSV contendo os dados do talhão, bem como a definição de uma variável destinada a armazenar a média geral de produtividade do talhão, a qual foi calculada com base nos dados obtidos no momento do armazenamento dos grãos. Na sequência, são realizados ajustes na formatação de colunas, incluindo a padronização do separador decimal, a reorganização da ordem das colunas e a exclusão daquelas consideradas irrelevantes para a análise.

Após essa etapa de preparação, o *script* prossegue para a fase de remoção de dados com base nas informações fornecidas pela telemetria. A filtragem inicial considera o campo "cd_estado", que indica o estado operacional da máquina no momento da aquisição de cada ponto. Durante a coleta, o sistema registra, dentre outros parâmetros, o tipo de atividade sendo realizada pela colhedora, como parada, manobra, deslocamento ou colheita propriamente dita. A Tabela 2 apresenta os códigos possíveis registrados durante esse processo.

Tabela 2 – Descrição dos códigos de estado dos equipamentos.

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

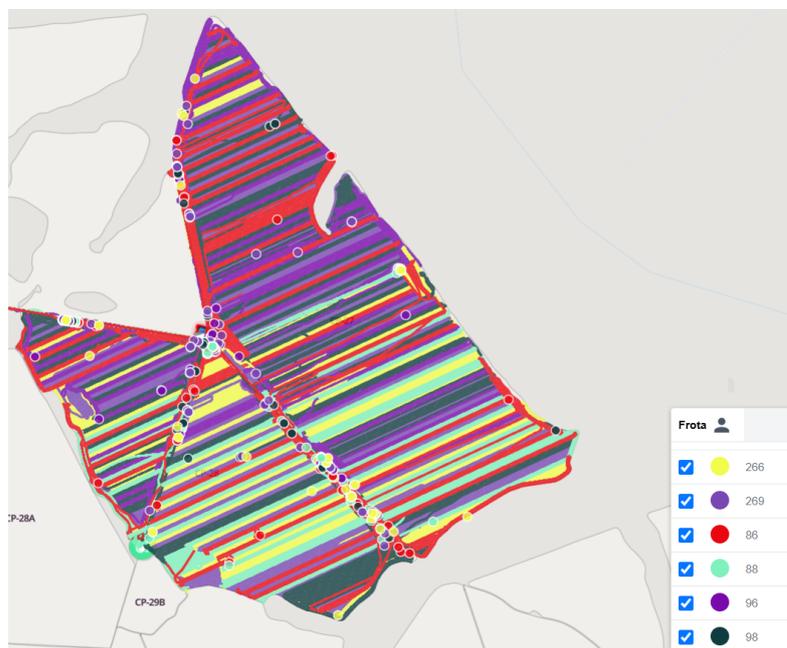
Neste trabalho, foram mantidas apenas as linhas cujo "cd_estado" era igual a "Efetivo", pois representam o momento em que a máquina está efetivamente realizando a colheita. Os demais registros foram excluídos, uma vez que não correspondem à operação de colheita e, portanto, podem introduzir ruídos e distorções nos dados de produtividade.

O passo seguinte consistiu na filtragem dos dados relacionados ao "vl_rendimento_colheita", que representa a quantidade de grãos colhidos em toneladas por hectare, um procedimento comumente adotado nesse tipo de análise (Sudduth; Drummond, 2007). Para isso, foram excluídos os registros com valores superiores ao dobro da média da coluna "vl_rendimento_colheita", bem como aqueles inferiores a 500 kg, visando eliminar possíveis inconsistências nos dados. O valor de 500 kg foi adotado de acordo com os gestores da fazenda.

A última etapa do pré-processamento dos dados foi o ajuste das médias. Para padronização dos dados de todas as colhedoras, foi aplicado um ajuste no "vl_rendimento_colheita" de cada máquina com base na média geral do talhão. O código executa um ajuste iterativo dos valores de rendimento da colheita para cada equipamento, garantindo a homogeneização dos dados em relação à média, dentro de uma margem de erro predefinida. Esse ajuste tem como objetivo corrigir possíveis variações nos valores registrados pelos sensores, considerando que, durante a colheita de um mesmo talhão (Figura 3), independentemente do seu tamanho, as máquinas

operam em proximidade.

Figura 3 – Rastro das colhedoras trabalhando nos dois talhões.



O detalhamento do código utilizado para ajuste das médias está descrito abaixo:

- **Iteração sobre equipamentos:** O código primeiro identifica todos os equipamentos em dados "cd_equipamento".
- **Ajuste iterativo:** Para cada equipamento, o código entra em um *loop while* que continua até que a média de rendimento do equipamento esteja dentro da tolerância em relação à média geral.
- **Cálculo do fator de ajuste:** Dentro do *loop*, o *script* calcula um fator de ajuste com base na comparação entre a média de rendimento do equipamento e a média geral. Se a média do equipamento for menor que a média geral, o fator aumenta os valores de rendimento. Se for maior, o fator diminui os valores. Se forem iguais, o fator é 1 (nenhuma mudança).
- **Aplicação do fator:** O código multiplica os valores de "vl_rendimento_colheita" do equipamento pelo fator calculado, ajustando assim o rendimento.
- **Verificação da tolerância:** Após o ajuste, o código verifica se a diferença absoluta entre a média de rendimento ajustada e a média geral está dentro da tolerância. Se estiver, o *loop* termina. Caso contrário, o processo de ajuste continua.

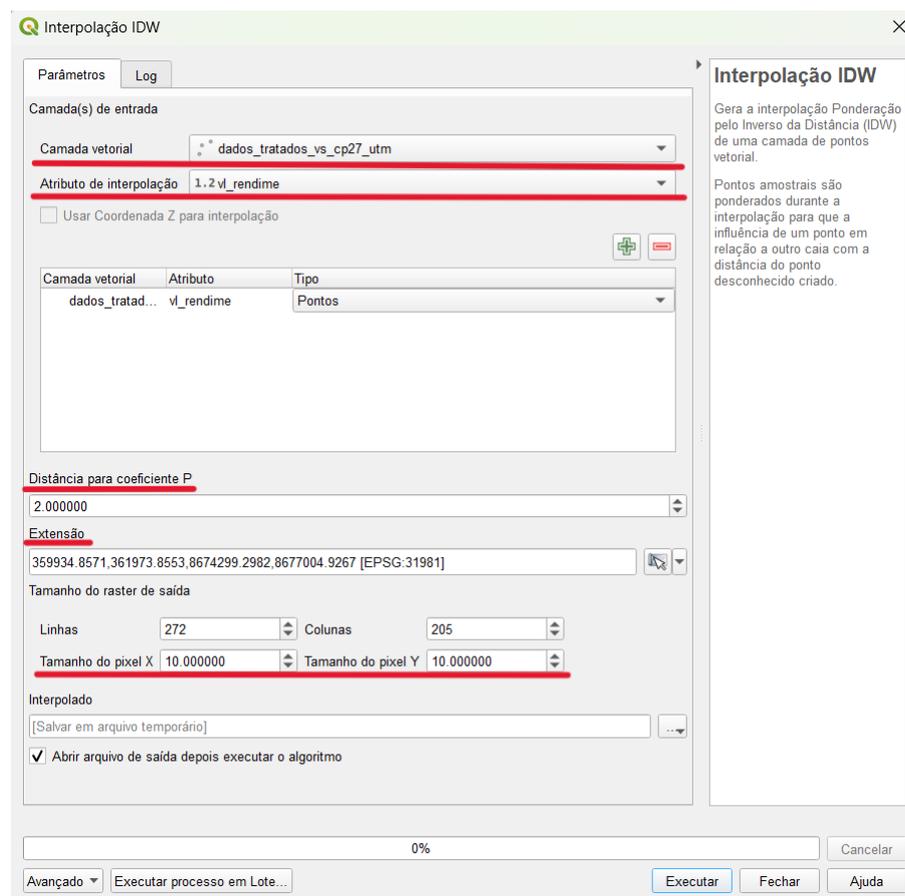
Por fim, o *script* salva os dados gerando um novo arquivo CSV que pode ser aberto em um Sistema de Informação Geográfica (SIG), utilizado para visualizar, editar e analisar dados geográficos.

4.3 Visualização dos dados

Para visualização/geração dos mapas de produtividade, foi utilizado o QGIS, um SIG gratuito e de código aberto. A interpolação dos dados foi feita pelo método IDW técnica usada para estimar valores em locais não amostrados com base em pontos de dados conhecidos, acessando a "Caixa de Ferramentas de Processamento" no QGIS e digitando no campo de busca "Interpolação IDW".

A Figura 4 mostra os parâmetros utilizados para a composição do mapa de produtividade. A camada vetorial, são os dados tratados (arquivo CSV) gerados pelo *script* e transformados em um arquivo *shapefile* em coordenadas UTM (*Universal Transverse Mercator*) pelo próprio QGIS; o atributo de interpolação é a coluna que representa o rendimento da colheita; a distância para o coeficiente P foi mantido o padrão do interpolador; a extensão foi um arquivo *shapefile* do limite do talhão e o tamanho do raster de saída foi com base na resolução dos pixels X e Y de 10 m.

Figura 4 – Parâmetros da interpolação IDW.



5 RESULTADOS E DISCUSSÃO

A Tabela 3 apresenta as estatísticas descritivas dos dados de rendimento nos dois talhões, antes e depois do processo de limpeza e ajuste das médias. Observou-se um aumento na média em ambos os talhões após o tratamento, o que indicou que os dados brutos continham uma

quantidade significativa de valores abaixo da média. Outros estudos também obtiveram aumento da média após a limpeza dos dados no *software* que eles desenvolveram (Sudduth; Drummond, 2007; Gimenez; Molin, 2004). A mediana zerada nos dois casos dos dados brutos aponta que, pelo menos, 50% dos registros de rendimento oriundos da telemetria apresentaram valor igual a zero. Além disso, houve uma redução no desvio padrão (St dev), o mesmo aconteceu nos estudos de Sudduth e Drummond (2007), Gimenez e Molin (2004), indicando maior uniformidade nos dados tratados. Os valores máximos caíram de 44,8 t/ha e 42,2 t/ha para 8,90 t/ha e 9,43 t/ha, respectivamente, evidenciando a remoção eficaz dos *outliers* e resultando em um conjunto de dados mais homogêneo e confiável para análises posteriores.

Tabela 3 – Estatísticas dos dados de rendimento da colheita.

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

Outra observação relevante refere-se à expressiva redução no número de dados após o processo de limpeza: aproximadamente 58,8% no Talhão 1 e 66,9% no Talhão 2 foram descartados. Para fins de comparação, no estudo conduzido por Gimenez e Molin (2004), as taxas de descarte foram de 41% e 21%, enquanto Sudduth e Drummond (2007) relataram a remoção de 13% a 27% dos dados em cinco talhões com áreas variando entre 11 ha e 48 ha, utilizando o *software* desenvolvido por eles.

A maior taxa de descarte observada neste trabalho pode estar relacionada à ocorrência de falhas em um dos sensores de uma das colhedoras, que resultou no registro de rendimento igual a zero em todos os pontos coletados por esse equipamento. Outro fato relevante foi a exclusão dos dados com o "c_estado" diferentes de efetivo, foram removidos 35,4% dos dados do Talhão 1 e 39,9% do Talhão 2, em alguns sistemas de telemetria esses dados nem aparecem nos dados brutos.

As Tabelas 4 e 5 apresentam a quantidade dos dados excluídos, após o processo de limpeza, variando de 42,23% a 78,27% para o Talhão 1 e de 46,49% a 80,08% para o Talhão 2, excluindo o equipamento 266 que foi perdido 100% dos dados. Com o resultado foi possível verificar que alguns sensores podem estar com problemas ou descalibrados. Outra informação são as médias, de rendimento e umidade, por máquina em cada etapa do processamento.

Tabela 4 – Dados das colhedoras antes, durante e após tratamento (Talhão 1).

cd_ equipa- mento	Dados Brutos			Dados Tratados sem ajuste da média			Dados Tratados com ajuste da média
	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	-	-	-	-

Tabela 5 – Dados das colhedoras antes, durante e após tratamento (Talhão 2).

cd_ equipa- mento	Dados Brutos			Dados Tratados sem ajuste da média			Dados Tratados com ajuste da média
	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	-	-	-	-

Os mapas de produtividade gerados a partir dos dados brutos, baixados diretamente do sistema de telemetria da fazenda, estão apresentados nas Figuras 5 e 6.

Figura 5 – Mapa de produtividade em pontos do Talhão 1 (dados brutos).

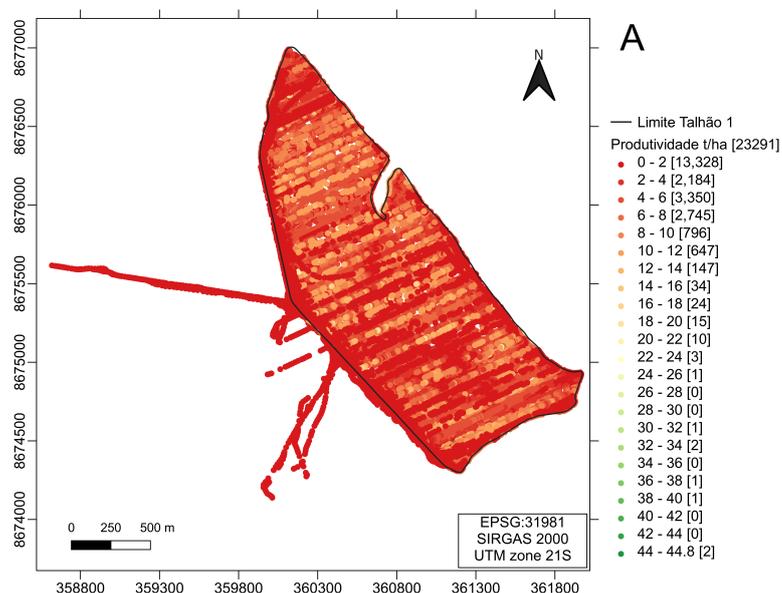
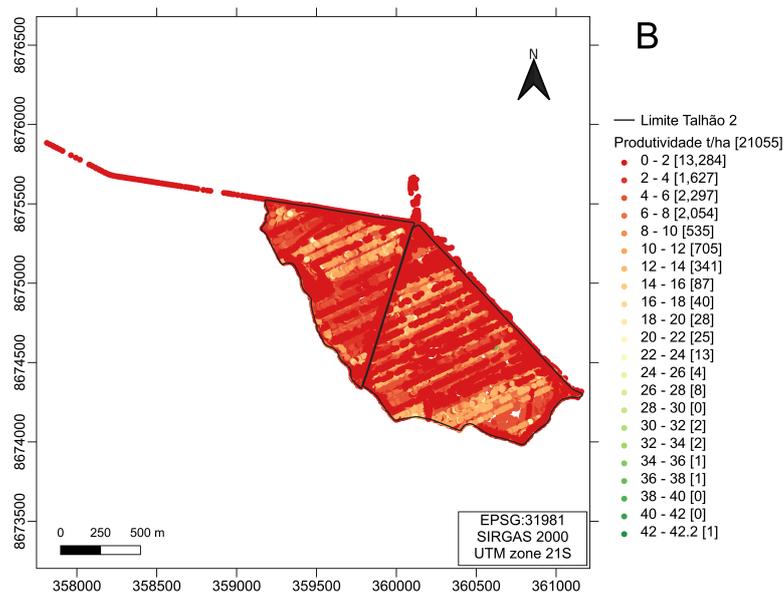


Figura 6 – Mapa de produtividade em pontos do Talhão 2 (dados brutos).



Observou-se, em ambos os talhões, a presença de pontos fora dos limites geográficos das áreas colhidas. No estudo conduzido por Menegatti e Molin (2004), erros de posicionamento representaram de 0,1% a 7,8% dos dados, evidenciando a recorrência desse tipo de inconsistência. Também há uma expressiva concentração de dados com valores inferiores a 2,0 t/ha, representando 57,22% dos pontos no Talhão 1 e 63,09% no Talhão 2. Como consequência dessas distorções, as médias de produtividade calculadas com base nos dados brutos foram de apenas 2,67 t/ha e 2,52 t/ha, respectivamente, valores significativamente inferiores às médias reais obtidas a partir da pesagem dos grãos no armazém, 3,68 t/ha e 3,84 t/ha, já ajustadas para 13% de umidade.

Já nas Figuras 7 e 8, são apresentados os mapas de produtividade gerados após o tratamento dos dados utilizando o *script* proposto.

Figura 7 – Mapa de produtividade em pontos do Talhão 1 (dados tratados).

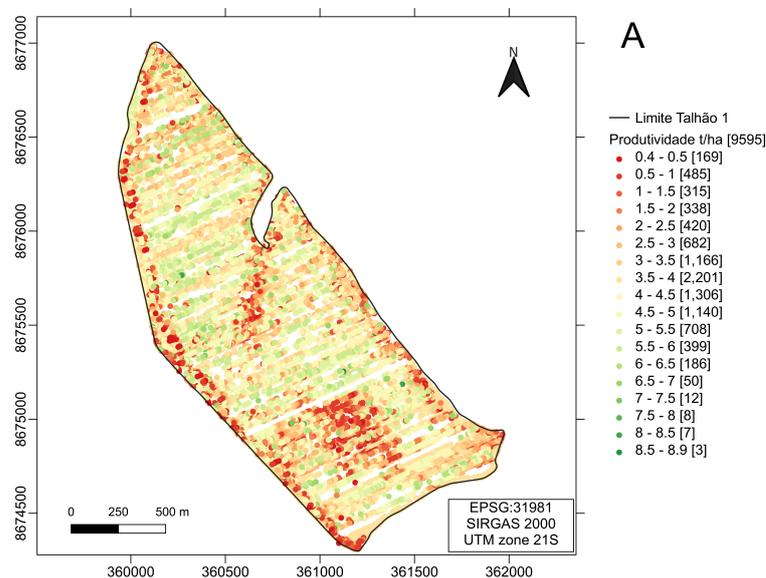
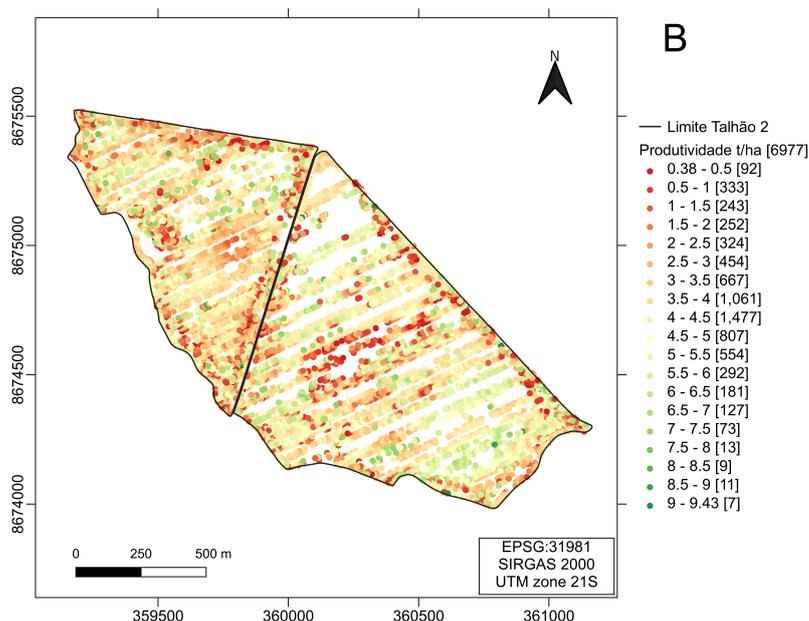


Figura 8 – Mapa de produtividade em pontos do Talhão 2 (dados tratados).



Os pontos localizados fora dos limites dos talhões foram excluídos, não por meio de um critério geográfico específico, mas por apresentarem valores abaixo do limite mínimo definido para rendimento.

Observou-se, especialmente nas bordaduras dos talhões, a presença de áreas com baixa produtividade. No Talhão 1, isso ocorre principalmente nas porções inferior e esquerda, enquanto no Talhão 2, observa-se na região superior — justamente na divisa entre os dois talhões, onde há uma estrada municipal. É provável que essa área funcione como zona de manobras e parada das máquinas ao final do expediente, o que pode ter provocado maior compactação do solo e, conseqüentemente, queda na produtividade.

Outro aspecto notável são as “falhas” ou lacunas visíveis em algumas regiões dos mapas. Isso se deve, além da natural redução de pontos ocasionada pela limpeza dos dados, ao fato de que uma das colhedoras registrou todos os valores de rendimento como zero, resultando em faixas sem dados válidos.

Mesmo antes da aplicação de técnicas de interpolação, os mapas já permitiam a identificação das zonas com menor produtividade, evidenciando o potencial do *script* para fornecer uma visualização inicial mais confiável da variabilidade espacial das lavouras.

Por fim foi gerado o mapa de produtividade após a realização da interpolação IDW (Figura 9 e 10), destacando as áreas de baixa em tons vermelhos e alta produtividade tons verdes. São esses mapas que são utilizados para a geração das zonas de manejo juntamente com outros.

Figura 9 – Mapa de produtividade interpolado do Talhão 1.

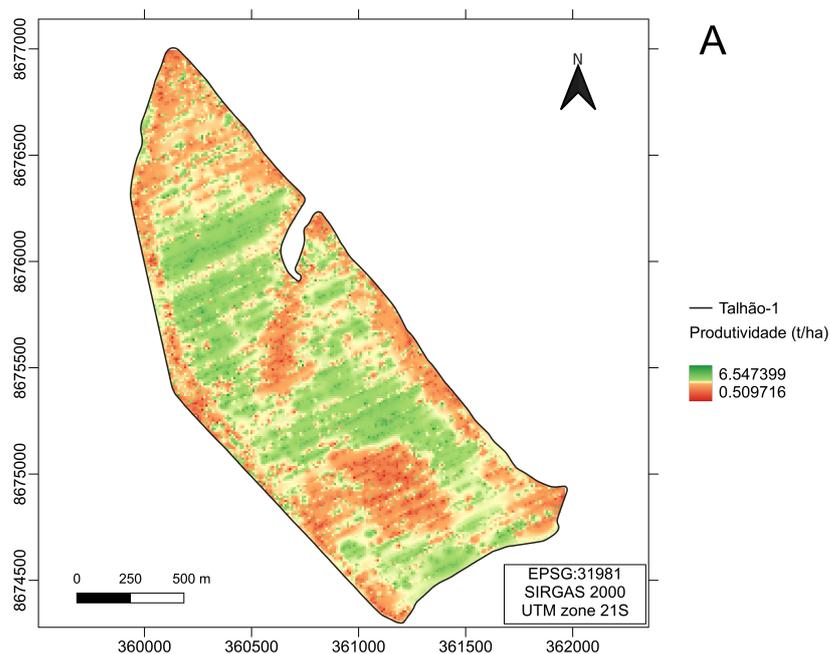
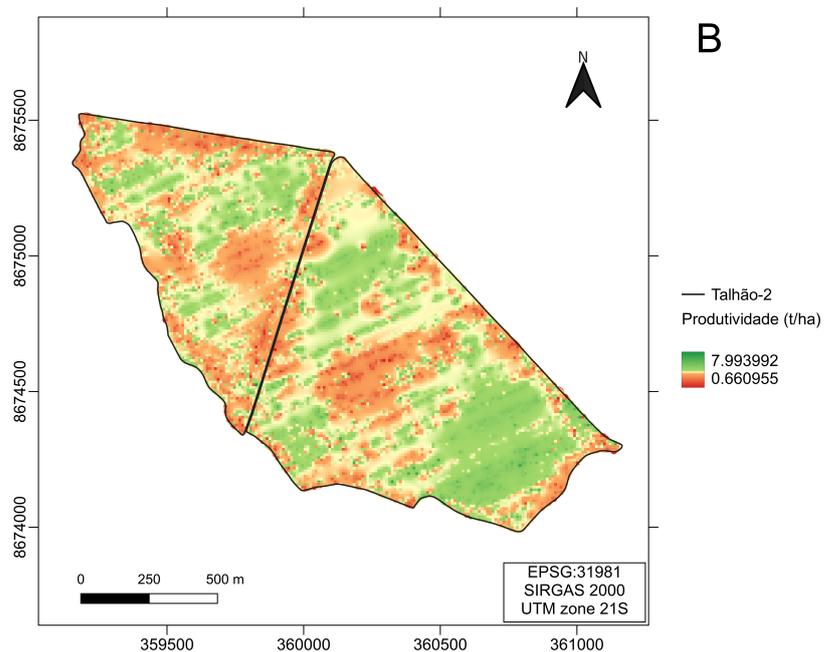


Figura 10 – Mapa de produtividade interpolado do Talhão 2.



6 CONCLUSÃO

O desenvolvimento e aplicação do *script* em Python para limpeza dos dados brutos de produtividade agrícola demonstraram-se eficazes no tratamento de informações coletadas via telemetria. A proposta contribuiu para a automação do processo de pré-processamento, reduzindo interferências causadas por erros nos dados brutos e aumentando a confiabilidade das análises.

A ferramenta se mostrou uma alternativa acessível e eficiente, permitindo maior controle e padronização na preparação dos dados para geração de mapas de produtividade. A aplicação desse tipo de solução é fundamental para garantir a qualidade da informação em projetos de AP, apoiando decisões agronômicas mais assertivas.

Comparado aos dados brutos, os dados tratados fornecem informações mais condizentes com a realidade dos talhões, permitindo uma melhor tomada de decisão para os gestores, possibilitando o uso do mapa interpolado à aplicações em taxa variável.

Referências

- BASSOI, L. H. *et al.* Agricultura de precisão e agricultura digital. **TECCOGS: Revista Digital de Tecnologias Cognitivas**, n. 20, maio 2020. ISSN 1984-3585. Disponível em: <<https://revistas.pucsp.br/teccogs/article/view/48542>>.
- BOTTEGA, E. *et al.* Uso de diferentes interpoladores na geração de modelo digital de elevação. In: CONGRESSO BRASILEIRO DE AGRICULTURA DE PRECISÃO - CONBAP. Ribeirão Preto - SP, 2012.
- CHANDEL, N. *et al.* Idw interpolation of soybean yield data acquired by automated yield monitor. **International Journal for Science and Emerging Technologies with Latest Trends**, v. 13, p. 36–45, 01 2013.
- COSTA, F. *et al.* Uma visão geral da aplicação de sensores em máquinas agrícolas. In: X CONGRESSO BRASILEIRO DE AGROINFORMATICA. Ponta Grossa, PR, 2015.
- DAMICO, J. Methodology for Cleaning, Quality, and Normalization of Vegetation Index Data Derived from Sentinel 2. jan. 2025. Disponível em: <https://www.academia.edu/126945828/Methodology_for_Cleaning_Quality_and_Normalization_of_Vegetation_Index_Data_Derived_from_Sentinel_2>.
- FIZZA, K. *et al.* Evaluating sensor data quality in internet of things smart agriculture applications. **IEEE Micro**, Institute of Electrical and Electronics Engineers (IEEE), v. 42, n. 1, p. 51–60, jan. 2022. ISSN 1937-4143. Disponível em: <<http://dx.doi.org/10.1109/MM.2021.3137401>>.
- GIMENEZ, L. M.; MOLIN, J. P. Algoritmo para redução de erros em mapas de produtividade para agricultura de precisão. **Revista Brasileira de Agrocomputação**, v. 2, n. 1, p. 5–10, 2004.
- GRISSE, R. D. *et al.* Yield monitor accuracy: Successful farming magazine case study. **Applied Engineering in Agriculture**, American Society of Agricultural and Biological Engineers, v. 18, n. 2, p. 147, 2002.
- IGAZ, D. *et al.* The evaluation of the accuracy of interpolation methods in crafting maps of physical and hydro-physical soil properties. **Water**, v. 13, p. 212, 01 2021.
- KÖPPEN, W. **Das geographische System der Klimate**. Berlin: Gebrüder Borntraeger, 1936. v. 1C. 1–44 p.
- LI, M. *et al.* Development of an Intelligent Yield Monitor for Grain Combine Harvester. In: LI, D.; WANG, B. (Ed.). **Artificial Intelligence Applications and Innovations**. New York: Springer-Verlag, 2005. v. 187, p. 663–670. ISBN 978-0-387-28318-0. Series Title: IFIP — The International Federation for Information Processing. Disponível em: <http://link.springer.com/10.1007/0-387-29295-0_72>.
- MALDANER, L. F.; MOLIN, J. P.; SPEKKEN, M. Methodology to filter out outliers in high spatial density data to improve maps reliability. **Scientia Agricola**, FapUNIFESP (SciELO), v. 79, n. 1, 2022. ISSN 1678-992X.

- MENEGATTI, L. A. A.; MOLIN, J. P. Remoção de erros em mapas de produtividade via filtragem de dados brutos. **Revista Brasileira de Engenharia Agrícola e Ambiental**, FapUNIFESP (SciELO), v. 8, n. 1, p. 126–134, abr. 2004. ISSN 1415-4366.
- PEREIRA, F. J.; MOLIN, J. P. Bancada de ensaio para avaliação de monitores de produtividade de grãos. **Engenharia Agrícola**, v. 23, n. 3, p. 568–578, 2003.
- SANTI, A. *et al.* Definição de zonas de produtividade em áreas manejadas com agricultura de precisão. **Revista Brasileira de Ciências Agrárias - Brazilian Journal of Agricultural Sciences**, v. 8, n. 3, p. 510–515, set. 2013. ISSN 19810997. Disponível em: <<http://www.agraria.pro.br/ojs32/index.php/RBCA/article/view/v8i3a2489>>.
- SANTOS, H. G. d. *et al.* **Sistema brasileiro de classificação de solos**. Brasília, DF: Embrapa, 2018. ISBN 978-85-7035-817-2.
- SAPRE, A.; VARTAK, S. Scientific Computing and Data Analysis using NumPy and Pandas. v. 07, n. 12, 2020.
- SILVA, W. d. V. R. d.; SILVA-MANN, R. Precision agriculture in Brazil: current situation, challenges and perspectives. **Research, Society and Development**, v. 9, n. 11, p. e1979119603, Nov. 2020. Disponível em: <<https://rsdjournal.org/index.php/rsd/article/view/9603>>.
- SUDDUTH, K. A.; DRUMMOND, S. T. Yield editor: Software for removing errors from crop yield maps. **Agronomy Journal**, v. 99, n. 6, p. 1471–1482, 2007. Disponível em: <<https://acsess.onlinelibrary.wiley.com/doi/abs/10.2134/agronj2006.0326>>.
- TSCHIEDEL, M.; FERREIRA, M. F. Introdução À agricultura de precisão: Conceitos e vantagens. **Ciência Rural**, Universidade Federal de Santa Maria, v. 32, n. 1, p. 159–163, Feb 2002. ISSN 0103-8478. Disponível em: <<https://doi.org/10.1590/S0103-84782002000100027>>.
- VEGA, A. *et al.* Protocol for automating error removal from yield maps. **Precision Agriculture**, v. 20, n. 5, p. 1030–1044, out. 2019. ISSN 1573-1618. Disponível em: <<https://doi.org/10.1007/s11119-018-09632-8>>.