

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE CIÊNCIAS AGRÁRIAS
CURSO DE AGRONOMIA

ARTHUR ALVES LEMOS

**VALIDAÇÃO DA PLANTABILIDADE EM SOJA COM USO DE VISÃO
COMPUTACIONAL**

UBERLÂNDIA – MG

2026

ARTHUR ALVES LEMOS

**VALIDAÇÃO DA PLANTABILIDADE EM SOJA COM USO DE VISÃO
COMPUTACIONAL**

Trabalho de Conclusão de Curso
apresentado ao Instituto de Ciências Agrárias
da Universidade Federal de Uberlândia, como
parte das exigências para obtenção do título de
Bacharel em Agronomia.

Orientador: Prof. Dr. Sandro Manuel
Carmelino Hurtado

UBERLÂNDIA – MG

2026

RESUMO

Este trabalho apresenta um fluxo de trabalho para aferição de plantabilidade em soja a partir de imagens pontuais obtidas por aeronave remotamente pilotada, integrando visão computacional, análise espacial e geoestatística. O estudo foi conduzido na Fazenda Mandaguari (Indianópolis-MG), pivô central 3-1 (talhão 1), com 14,13 ha. A aquisição de dados, realizada em 30 de outubro de 2025, seguiu duas etapas: (i) voo a aproximadamente 120 m AGL para geração de ortomosaico, MDE e MDT; e (ii) voo baixo por pontos, a aproximadamente 4 m AGL, com uma imagem por ponto no estádio VC. As imagens pontuais foram processadas por um modelo YOLOv8m de segmentação por instância treinado com tiles de 640×640 pixels, com divisão do conjunto de dados por ponto (foto/ponto). Na validação, foram obtidos $mAP_{50}(\text{box})=0,963$, $mAP_{50}(\text{mask})=0,960$, $mAP_{50-95}(\text{box})=0,782$, $mAP_{50-95}(\text{mask})=0,715$, $\text{precision}(\text{box})=0,918$ e $\text{recall}(\text{box})=0,938$. As máscaras segmentadas foram convertidas em centroides no QGIS, e um script em Python estimou a direção dominante de plantio, agrupou fileiras e calculou espaçamentos longitudinais entre plantas, classificando falhas e duplas de forma relativa ao espaçamento esperado em cada imagem. Como controle de qualidade do encadeamento entre centroides, pontos com $\text{drop_pct} \geq 25\%$ foram tratados como falha severa de linking e removidos, resultando em $n=39$ para as análises descritivas e espaciais. Após o filtro, observaram-se medianas de $\%CV=65,20\%$, $\%\text{duplas}=8,93\%$ e $\%\text{falhas}=18,68\%$. A krigagem ordinária permitiu gerar mapas temáticos e semivariogramas; entretanto, o I de Moran não indicou autocorrelação espacial significativa a 5% e a validação cruzada apresentou baixos R^2 , sugerindo estrutura espacial fraca e capacidade preditiva limitada nas condições avaliadas. Conclui-se que o método é reprodutível e adequado para quantificar indicadores pontuais de plantabilidade, enquanto a robustez dos produtos espaciais depende de maior densidade amostral e/ou incorporação de covariáveis.

Palavras-chave: Agricultura de precisão; Aeronave remotamente pilotada; Geoestatística; Krigagem ordinária.

ABSTRACT

This study presents a workflow for assessing soybean planting quality from point-based imagery acquired by a remotely piloted aircraft, integrating computer vision, spatial analysis, and geostatistics. The study was conducted at Mandaguari Farm (Indianópolis, Minas Gerais, Brazil), center pivot 3-1 (Field 1), covering 14.13 ha. Data acquisition, performed on October 30, 2025, followed two stages: (i) a flight at approximately 120 m AGL to generate an orthomosaic and digital elevation and terrain models; and (ii) a low-altitude point-based flight at approximately 4 m AGL, with one image per sampling point at the VC growth stage. Point images were processed with a YOLOv8m instance-segmentation model trained on 640×640 -pixel tiles using a point-level dataset split. Validation reached $mAP_{50}(\text{box})=0.963$, $mAP_{50}(\text{mask})=0.960$, $mAP_{50-95}(\text{box})=0.782$, $mAP_{50-95}(\text{mask})=0.715$, $\text{precision}(\text{box})=0.918$, and $\text{recall}(\text{box})=0.938$. Segmented masks were converted to centroids in QGIS, and a Python script estimated the dominant planting direction, grouped crop rows, and calculated within-row plant spacing, classifying skips and doubles relative to the expected spacing within each image. As a quality-control criterion for centroid linking, points with $\text{drop_pct} \geq 25\%$ were treated as severe linking failures and removed, yielding $n=39$ for descriptive and spatial analyses. After filtering, median values were $\%CV=65.20\%$, $\%\text{doubles}=8.93\%$, and $\%\text{skips}=18.68\%$. Ordinary kriging was used to generate thematic maps and variograms; however, Moran's I did not indicate significant spatial autocorrelation at the 5% level, and cross-validation yielded low R^2 values, indicating weak spatial structure and limited predictive performance under the evaluated conditions. The workflow was reproducible and suitable for quantifying point-level planting-quality indicators, whereas stronger spatial products will require higher sampling density and/or additional covariates.

Keywords: Precision agriculture; Remotely piloted aircraft; Geostatistics; Ordinary kriging.

LISTA DE ABREVIATURAS E SIGLAS

AGL	Above Ground Level - altura acima do nível do solo.
AP	Agricultura de precisão.
ARP	Aeronave remotamente pilotada.
CV	Coeficiente de variação.
DJI	Nome comercial do fabricante da aeronave e dos sensores utilizados.
GSD	Ground Sample Distance - distância de amostragem no terreno.
IoU	Intersection over Union - interseção sobre união.
KML	Keyhole Markup Language - formato de dados geoespaciais.
k-NN	k-Nearest Neighbors - k vizinhos mais próximos.
mAP	mean Average Precision - métrica agregada de precisão média.
MDE	Modelo Digital de Elevação.
MDT	Modelo Digital de Terreno.
PH	Pontos por hectare.
QGIS	Sistema de Informação Geográfica QGIS, software livre de geoprocessamento.
R²	Coeficiente de determinação.
RGB	Red, Green, Blue - sistema de cores vermelho, verde e azul.
RMSE	Root Mean Square Error - raiz do erro quadrático médio.
VC	Estádio VC da soja, caracterizado por folhas unifolioladas desenroladas.
YOLO	You Only Look Once - família de modelos de visão computacional.

SUMÁRIO

1. INTRODUÇÃO.....	2
2. MATERIAL E MÉTODOS.....	3
2.1. Área de estudo.....	3
2.2. Definição da grade amostral.....	3
2.3. Aquisição de imagens a partir de ARP.....	4
2.3.1. Geração de MDE, MDT e ortomosaico.....	4
2.3.2. Obtenção de imagens nos pontos da grade.....	5
2.4. Identificação de plântulas a partir de visão computacional.....	6
2.4.1. Divisão de dados, aumento e métricas.....	7
2.5. Extração de centroides e cálculo das métricas de plantabilidade.....	8
2.5.1. Algoritmo de direção, fileiras e espaçamento.....	9
2.5.2. Definição do espaçamento esperado e classificação de anomalias.....	10
2.6. Validação do processo (pipeline).....	11
2.7. Critério de criticidade do filtro leve.....	12
2.8. Análise estatística.....	12
3. RESULTADOS E DISCUSSÃO.....	14
3.1. Análise descritiva dos dados levantados.....	14
3.2. Eficiência da rotulagem assistida por auto-label.....	16
3.3. Desempenho do modelo de segmentação.....	16
3.4. Validação do pipeline de extração.....	17
3.5. Estatísticas descritivas dos indicadores por ponto.....	17
3.6. Interpolação por krigagem e semivariogramas.....	17
3.7. Observações adicionais e limitações.....	19
4. CONCLUSÕES.....	21
REFERÊNCIAS.....	22

1. INTRODUÇÃO

A plantabilidade pode ser compreendida como a qualidade com que a operação de semeadura e o estabelecimento inicial da cultura resultam em uma distribuição uniforme de plantas. Sua avaliação envolve, principalmente, a distribuição longitudinal ao longo das fileiras, com identificação de espaçamentos aceitáveis, falhas e duplas/múltiplos, além de medidas de dispersão, como o coeficiente de variação. A qualidade de implantação não depende exclusivamente do mecanismo dosador: profundidade de deposição, contato semente-solo, qualidade fisiológica e fluidez das sementes e condições ambientais durante a emergência também podem alterar o estande observado (WEIRICH NETO et al., 2015; NARDON; BOTTA, 2022).

Na cultura da soja, a desuniformidade entre plantas pode modificar a competição intraespecífica e a eficiência no uso de recursos. Pereyra et al. (2022) demonstraram que a uniformidade espacial entre plantas no início do ciclo pode influenciar a produtividade, sobretudo em ambientes de menor potencial produtivo e sob menores densidades de plantas. Assim, os indicadores de falhas, duplas e CV obtidos após a emergência devem ser interpretados como medidas diagnósticas do estande estabelecido, sem atribuir, isoladamente, a causa das anomalias à semeadora.

A agricultura de precisão permite espacializar indicadores de campo e investigar padrões de variabilidade dentro dos talhões, inclusive em sistemas de produção de soja (CORRÊA et al., 2022). Nesse contexto, aeronaves remotamente pilotadas possibilitam adquirir imagens de elevada resolução espacial em curto intervalo de tempo e têm sido amplamente empregadas em aplicações de sensoriamento remoto agrícola (MULLA, 2013; TSOUROS et al., 2019). A integração dessas imagens com técnicas de visão computacional possibilita detectar e segmentar plantas individualmente, convertendo informação visual em variáveis quantitativas passíveis de análise espacial.

O presente trabalho teve por objetivo descrever e validar um fluxo de trabalho para aferição da plantabilidade na cultura da soja a partir de imagens de alta resolução obtidas por ARP, integrando segmentação por visão computacional, processamento geoespacial e geoestatística para quantificar e espacializar indicadores de falhas, duplas e CV.

2. MATERIAL E MÉTODOS

2.1. Área de estudo

O estudo foi realizado na Fazenda Mandaguari, no município de Indianópolis–MG. O talhão analisado, correspondente ao pivô 3-1 (18°59'S e 47°48'O), possui 14,13 ha, altitude média de 977 m e solo classificado como Latossolo Vermelho (Figura 1). Foi definida uma grade amostral de 43 pontos para obtenção de indicadores de plantabilidade por meio de imagens aéreas adquiridas com ARP.

Figura 1. Mapa de localização do talhão com o ensaio de plantabilidade.

2.2. Definição da grade amostral

Com o objetivo de aprimorar a representatividade espacial, foi definida uma malha de 43 pontos amostrais (Figura 2), cujo espaçamento entre pontos foi calculado conforme as Fórmulas 1 e 2.

Fórmula 1:

$$PH = \frac{\text{Pontos Totais}}{\text{Área (ha)}}$$

Fórmula 2:

$$D = \sqrt{\frac{10.000}{PH}}$$

Onde: PH = Pontos por hectare
D = Distância entre pontos (m)

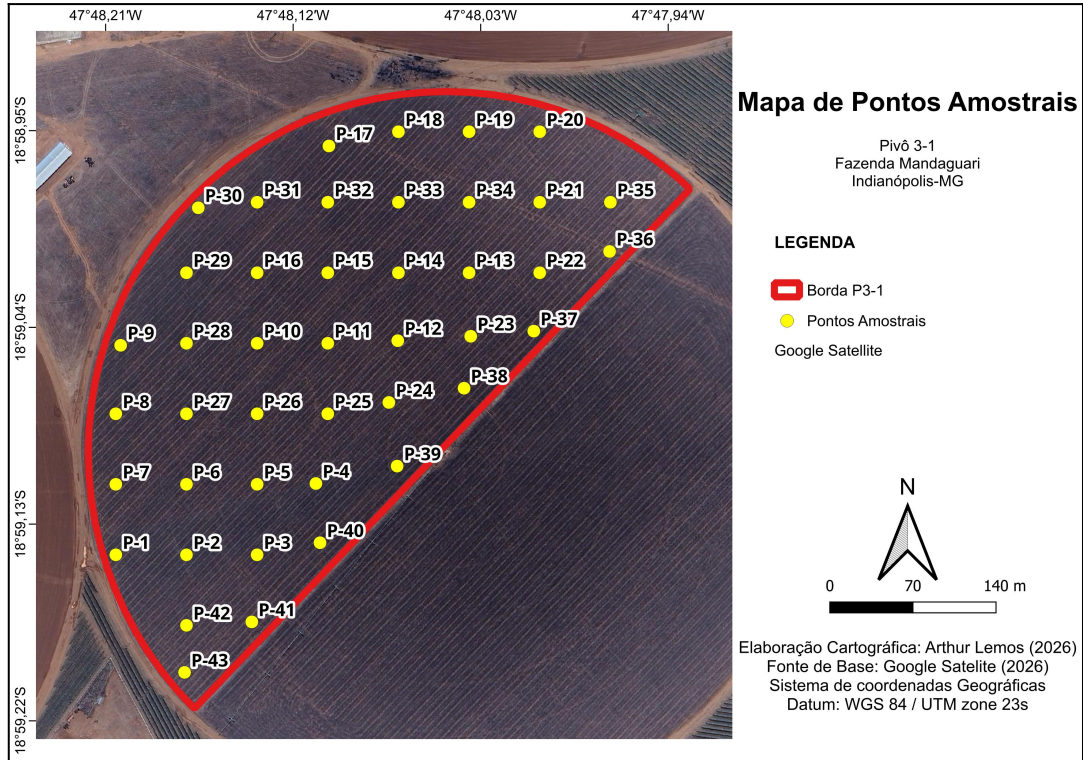


Figura 2. Mapa de pontos amostrais definidos para a área experimental.

2.3. Aquisição de imagens a partir de ARP

A aquisição de imagens foi realizada em 30 de outubro de 2025 e considerou duas etapas. A primeira teve como objetivo gerar o ortomosaico, o MDE e o MDT; a segunda foi destinada à captura de imagens RGB de alta resolução em cada ponto da grade amostral. Em ambas as etapas foi utilizada uma ARP DJI Mavic 3 Multispectral®.

2.3.1. Geração de MDE, MDT e ortomosaico

Para a geração do MDE, do MDT e do ortomosaico, foi executado um plano de voo no aplicativo DJI Pilot 2® com altitude de 120 m AGL, sobreposição frontal de 85% e lateral de 75% e velocidade de voo configurada em 20% abaixo do máximo recomendado pela plataforma. O voo ocorreu próximo das 10 h, buscando reduzir a interferência de sombras e variações abruptas de iluminação. As imagens foram processadas no WebODM, que permite a geração de ortomosaicos e modelos digitais derivados de imagens aéreas (WEBODM, 2026). O ortomosaico apresentou GSD aproximado de 3,2 cm/pixel. O MDE obtido para a área experimental é apresentado na Figura 3.

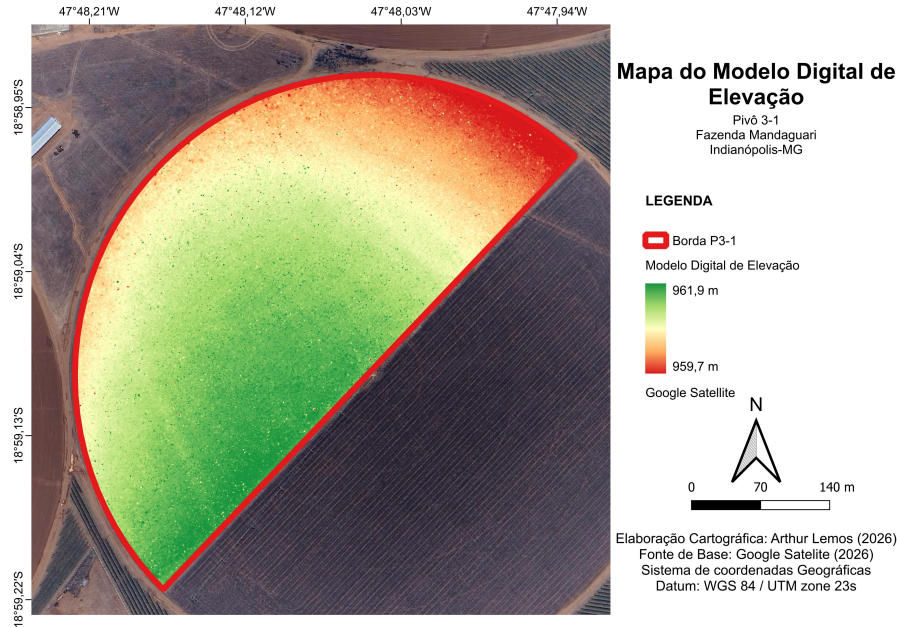


Figura 3. Mapa do Modelo Digital de Elevação (MDE) da área experimental.

2.3.2. Obtenção de imagens nos pontos da grade

A segunda etapa de aquisição de imagens foi realizada a 4 m AGL, com o gimbal em -90° (nadir), quando aproximadamente 50% das plantas se encontravam no estágio VC, caracterizado por folhas unifolioladas desenroladas (FEHR; CAVINESS, 1977). O plano de voo foi carregado no DJI Pilot 2® a partir de arquivo KML gerado por script em linguagem Python (PYTHON SOFTWARE FOUNDATION, 2026), ordenando os pontos para otimizar a rota. A variação altimétrica do terreno foi compensada com base no MDT, tomando como referência o ponto de decolagem (Fórmula 3), de modo que a altura AGL fosse mantida tão constante quanto possível.

Fórmula 3:

$$\text{Altura real} = (\text{Altitude do ponto de controle} - \text{Altitude do ponto}) + \text{Altura desejada}$$

Como os desníveis não exigiram a divisão do plano em múltiplas missões, o percurso foi realizado em uma única missão. A altitude de 4 m AGL resultou em GSD aproximado de 0,11 cm/pixel. A missão teve duração aproximada de 12 minutos, com parada programada em cada posição para a captura da imagem e consumo aproximado de 43% da bateria (Figura 4).



Figura 4. Exemplo de imagem capturada nos pontos amostrais a 4 m de altura.

Os parâmetros da câmera foram configurados em modo manual e mantidos fixos a partir da decolagem, após a calibração automática do sensor, com o objetivo de reduzir a variação de exposição entre os pontos da grade amostral.

2.4. Identificação de plântulas a partir de visão computacional

A identificação das plântulas foi realizada com o modelo médio de segmentação por instância YOLOv8m, disponibilizado pela Ultralytics® (ULTRALYTICS, 2026). As imagens obtidas no voo baixo foram divididas em tiles de 640×640 pixels para treinamento e inferência. A base inicial foi composta por 100 tiles anotados manualmente e complementada por 315 tiles gerados por rotulagem assistida (auto-label) na plataforma Roboflow® (ROBOFLOW, 2026; Figura 5). Cada auto-rotulagem foi submetida a auditoria visual, com correção de falsos positivos, falsos negativos e erros de delimitação das instâncias. O tempo médio observado para a anotação manual completa foi de aproximadamente 78 segundos por tile, enquanto a auditoria e correção de tiles auto-rotulados demandou aproximadamente 15 segundos por tile.

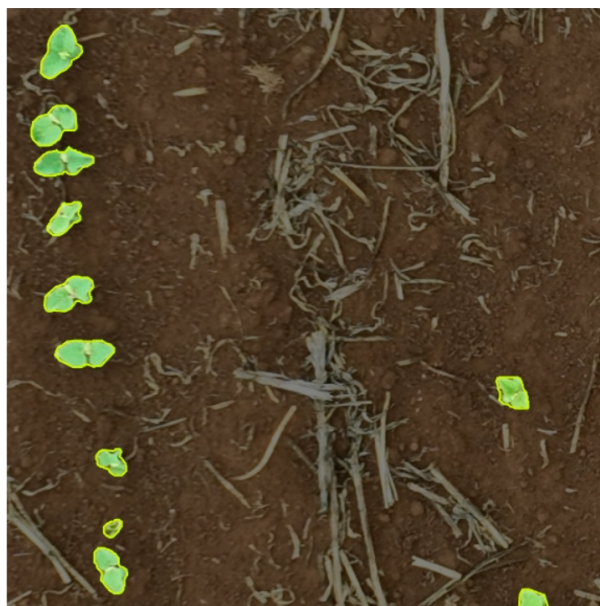


Figura 5. Exemplo de tile com anotações de segmentação no Roboflow®.

2.4.1. Divisão de dados, aumento e métricas

Para reduzir o risco de vazamento de informação (data leakage) entre conjuntos, a divisão do dataset foi realizada por ponto amostral (foto/ponto), e não por tiles individualmente. Dos 43 pontos, 38 foram utilizados para compor os conjuntos de treinamento (30), validação (4) e teste (4), enquanto 5 pontos foram mantidos como reserva. Procedimentos de aumento de dados (augmentation) foram aplicados somente ao conjunto de treinamento. Na configuração operacional foram utilizados limiar de confiança (confidence threshold) de 0,25 e limiar de IoU de 0,25 para filtragem das previsões. As métricas reportadas pelo Ultralytics Hub incluíram $mAP_{50}(box)=0,963$; $mAP_{50}(mask)=0,960$; $mAP_{50-95}(box)=0,782$; $mAP_{50-95}(mask)=0,715$; $precision(box)=0,918$; $precision(mask)=0,915$; $recall(box)=0,938$; e $recall(mask)=0,927$ (Figura 6).

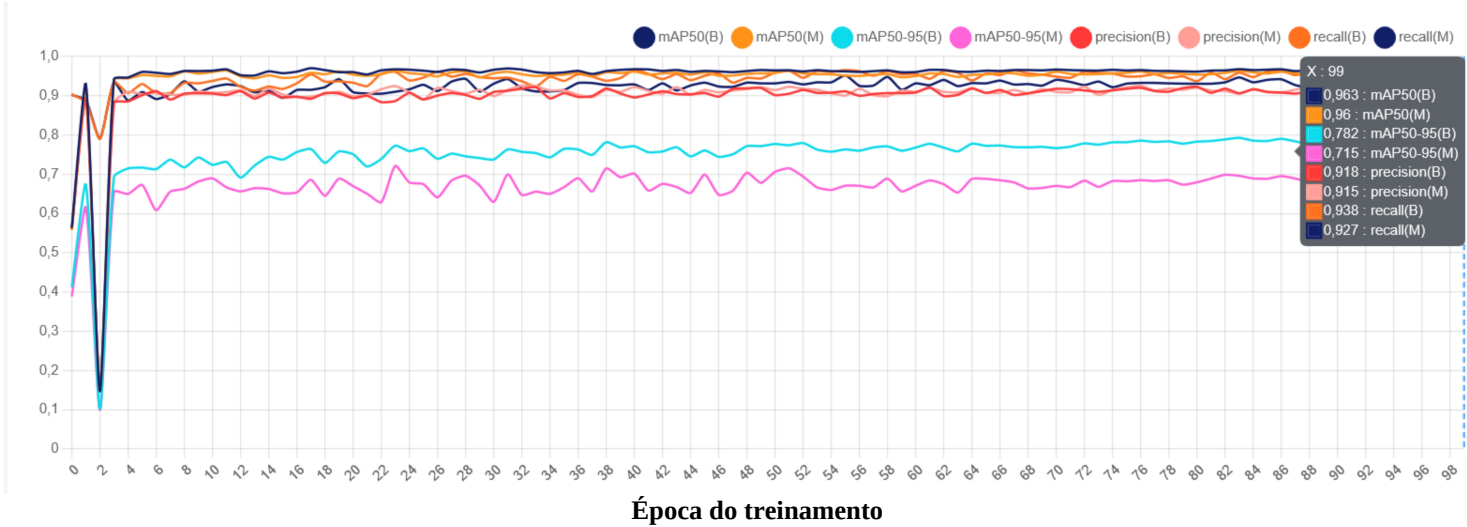


Figura 6. Métricas de validação do conjunto de dados obtidas pelo Ultralytics Hub.

2.5. Extração de centroides e cálculo das métricas de plantabilidade

Após a inferência, as máscaras segmentadas foram convertidas em feições vetoriais (polígonos), das quais foram extraídos os centroides correspondentes a cada plântula no QGIS 3.40 (QGIS DEVELOPMENT TEAM, 2024). Esses centroides foram utilizados para estimar a direção de plantio e os espaçamentos entre plantas em cada imagem (Figura 7). O sistema de cultivo adotou espaçamento de 0,50 m entre linhas.

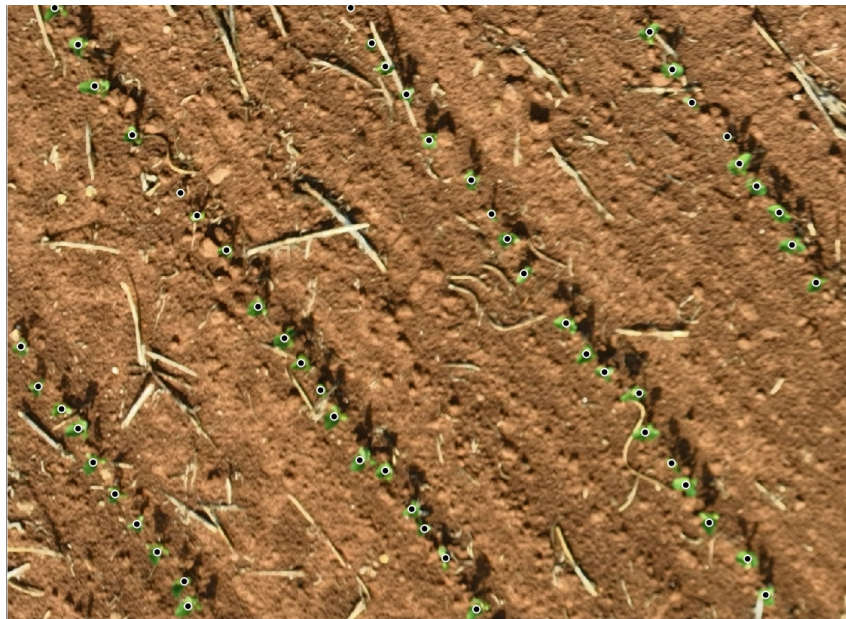


Figura 7. Centroides das plantas de soja extraídos no QGIS.

2.5.1. Algoritmo de direção, fileiras e espaçamento

O script em Python executado no QGIS recebe como entrada a camada de centroides das plântulas, obtida a partir das máscaras segmentadas, e produz, para cada imagem: (i) a direção dominante de plantio (θ); (ii) o espaçamento entre linhas (row_sp); (iii) os espaçamentos longitudinais intralinha (du); e (iv) a classificação das anomalias em duplas/múltiplos e falhas. O procedimento foi desenvolvido para ser robusto a rotações arbitrárias das imagens e baseia-se em estimação de orientação por vetores de vizinhança, rotação do sistema de coordenadas para um referencial alinhado às fileiras, agrupamento unidimensional das plantas em fileiras, encadeamento intralinha e filtros geométricos para reduzir ligações espúrias.

- Estimativa da direção (θ): para cada centroide i , são obtidos k vizinhos mais próximos (k -NN) e calculados vetores $\Delta x = x_j - x_i$ e $\Delta y = y_j - y_i$.
- Para cada vetor, obtém-se o ângulo $\alpha = \text{atan2}(\Delta y, \Delta x)$. Como a direção de plantio tem simetria de 180° (orientações opostas representam a mesma linha), os ângulos são tratados em módulo π . A direção dominante é então estimada via média circular com duplicação angular: $\theta = 0,5 \cdot \text{atan2}(\sum \sin(2\alpha), \sum \cos(2\alpha))$. A confiança (dir_conf) é derivada da concentração angular $R = \sqrt{(\sum \cos(2\alpha))^2 + (\sum \sin(2\alpha))^2} / N$, onde valores de R próximos de 1 indicam orientação bem definida e valores menores indicam maior dispersão.
- Estimativa do espaçamento entre linhas (row_sp) e agrupamento de fileiras: após a rotação, utiliza-se a coordenada v para identificar fileiras como clusters 1D. O script aplica um agrupamento guloso (greedy) por varredura em v : ordenam-se os centroides por v e inicia-se uma nova fileira quando o salto $|\Delta v|$ excede um limiar. Esse limiar é parametrizado por ROW_GAP_FACTOR (= 0,35) e é aplicado de forma conservadora para reduzir a fusão de fileiras adjacentes. Após a definição dos clusters, o espaçamento entre linhas row_sp é calculado como a mediana das diferenças entre as posições centrais (por exemplo, mediana de v) das fileiras adjacentes.
- Rotação do sistema de coordenadas: com θ , cada centroide (x, y) é projetado em um sistema rotacionado (u, v) para separar o eixo longitudinal (ao longo da sementeira) do eixo transversal (entre fileiras). Nesse sistema, plantas de uma mesma fileira tendem a apresentar valores semelhantes de v , enquanto a sequência ao longo da fileira pode ser ordenada por u .

- Construção de ligações intralinha e filtro anti-zig-zag: dentro de cada fileira, os centroides são ordenados por u . As ligações são estabelecidas entre vizinhos consecutivos ($u_i \rightarrow u_{i+1}$) e o espaçamento longitudinal é calculado como $du = u_{i+1} - u_i$. Para reduzir ligações espúrias que cruzam fileiras ou produzem zig-zague, aplica-se um filtro geométrico com base no desvio transversal: links com $|\Delta v|$ acima de $DV_MAX_FACTOR \cdot row_sp$ ($DV_MAX_FACTOR = 0,30$) são descartados. O algoritmo reporta a taxa de descarte ($drop_pct$), útil para diagnosticar casos de baixa separação entre fileiras ou baixa qualidade de detecção.
- Cálculo do espaçamento esperado e classificação: o espaçamento esperado ($expected$) é definido como a mediana dos valores de du após a filtragem. Neste estudo, adotaram-se como critérios operacionais próprios do pipeline: dupla/múltiplo quando $du < 0,55 \times expected$ e falha quando $du > 1,60 \times expected$, além do cálculo do CV (%) dos espaçamentos por imagem. Esses limites foram empregados como diretrizes deste estudo para aumentar a tolerância a pequenas variações do espaçamento estimado em imagens sem escala métrica absoluta; portanto, não devem ser interpretados como limites normativos universais.

2.5.2. Definição do espaçamento esperado e classificação de anomalias

Como as imagens pontuais não possuíam escala métrica absoluta vinculada a cada pixel, o espaçamento de referência foi estimado de forma relativa no interior de cada imagem. O valor $expected$ foi definido pela mediana dos espaçamentos du após filtragem. Em seguida, aplicaram-se os critérios operacionais definidos neste estudo: dupla/múltiplo quando $du < 0,55 \times expected$ e falha quando $du > 1,60 \times expected$ (Figura 8). O CV foi calculado sobre o conjunto de espaçamentos du de cada imagem. A classificação relativa permite comparar a regularidade interna entre pontos, mas sua interpretação deve permanecer vinculada às condições e à calibração do presente estudo.

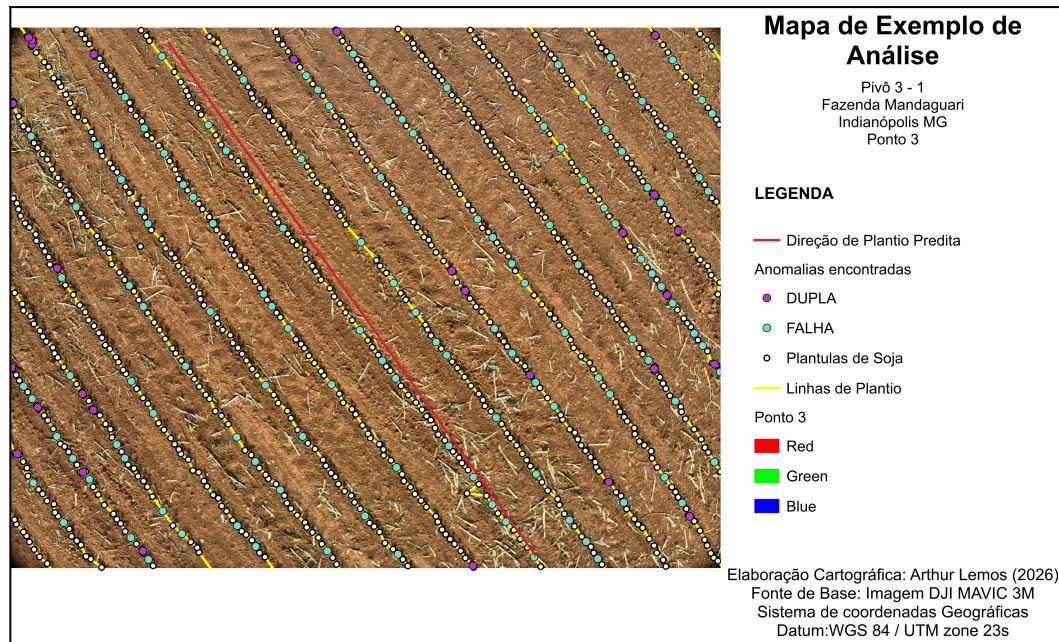


Figura 8. Exemplo de amostragem de pontos e ligações geradas pelo algoritmo em Python no QGIS.

2.6. Validação do processo (pipeline)

Para reduzir a subjetividade na avaliação do encadeamento intralinha, foi introduzido o indicador `drop_pct` como medida objetiva de consistência do linking. No passo de ligação entre centroides consecutivos em cada fileira, ordenados no eixo longitudinal u , o algoritmo calcula o salto transversal entre centroides consecutivos no eixo v e descarta ligações com potencial de cruzar fileiras adjacentes (filtro anti-zig-zag). Define-se:

Fórmula 4:

$$dv = |v(i+1) - v(i)|$$

Descartar link se: $dv > row_sp \times DV_MAX_FACTOR$

Onde `row_sp` é o espaçamento entre linhas estimado para a imagem e `DV_MAX_FACTOR` é um fator adimensional de tolerância. Para cada ponto, o script registra `links_attempt` (ligações tentadas), `links_kept` (ligações aceitas) e `links_drop` (ligações descartadas). O indicador `drop_pct` é então definido por:

Fórmula 5:

$$drop_pct = (links_drop / links_attempt) \times 100$$

Como interpretação, temos que valores baixos de `drop_pct` indicam que a maior parte das ligações consecutivas foi considerada consistente com a fileira estimada; valores elevados indicam maior fragmentação do encadeamento, situação que pode ocorrer em imagens com

baixa coerência direcional local, poucos pontos detectados, ruído, bordaduras ou interseções de plantio (Figura 9).



Figura 9. Ponto 30, descartado devido a condição operacional incompatível com a avaliação regular do espaçamento intralinha.

2.7. Critério de criticidade do filtro leve

Considerando $p = \text{drop_pct}/100$ como a probabilidade empírica de uma ligação ser descartada, o número esperado de ligações aceitas consecutivas antes de um descarte pode ser aproximado por $E[L] = (1 - p)/p$, em um modelo geométrico simplificado. Para $\text{drop_pct} = 25\%$ ($p = 0,25$), obtém-se $E[L]$ aproximadamente igual a 3 ligações aceitas, equivalente a cadeias curtas de cerca de quatro pontos conectados. Nessa condição, a estimativa do espaçamento longitudinal torna-se mais sensível a ruído e fragmentação. Por esse motivo, o valor de $\text{drop_pct} \geq 25\%$ foi definido, no presente estudo, como critério operacional de falha severa de linking, levando à exclusão do ponto das análises subsequentes.

2.8. Análise estatística

Como critério de controle de qualidade, foram descartados os pontos que apresentaram drop_pct igual ou superior a 25%. Assim, os pontos 28, 30, 39 e 43 foram desconsiderados nas análises estatísticas e espaciais. A inspeção dessas imagens mostrou situações inadequadas à avaliação convencional da distribuição intralinha, como interseções de linhas de plantio, terraços e rastros de pulverizador.

Para cada ponto da grade amostral foram analisados os percentuais de falhas e duplas e o CV dos espaçamentos, inicialmente por estatística descritiva. A análise espacial foi realizada por geoestatística, com ajuste de semivariogramas isotrópicos, interpolação em resolução de 3 m por krigagem ordinária e geração de mapas de superfície. Essa etapa foi executada no QGIS 3.40 com o plugin Smart-Map (PEREIRA et al., 2022). Não foram aplicadas transformações aos dados, priorizando a transparência do fluxo metodológico.

3. RESULTADOS E DISCUSSÃO

3.1. Análise descritiva dos dados levantados

Os indicadores calculados pelo pipeline para cada um dos 43 pontos amostrais são apresentados na Tabela 1. Essa síntese permite identificar, antes do filtro de criticidade, tanto a variabilidade dos indicadores de plantabilidade quanto a qualidade do encadeamento geométrico produzido pelo algoritmo.

Tabela 1. Indicadores calculados para os 43 pontos amostrais antes da aplicação do filtro de criticidade.

Ponto	Nº Plantas	Nº Linhas	theta_deg	dir_conf	row_sp	exp_sp	CV_pct	duplas_%	falhas_%	links_kept	links_drop	drop_pct
1	551	15	98,5	0,1	383,6	69,8	83,2	11,5	27,6	532,0	2,0	0,4
2	906	20	125,5	0,1	311,0	56,4	60,3	12,7	19,5	882,0	2,0	0,2
3	779	19	125,5	0,1	344,5	62,5	51,9	3,8	17,3	756,0	2,0	0,3
4	740	18	76,5	0,1	335,8	64,2	57,1	6,3	17,1	719,0	3,0	0,4
5	733	20	126,5	0,1	348,4	64,1	62,8	10,2	19,8	706,0	4,0	0,6
6	539	17	126,5	0,1	408,4	73,8	62,3	11,5	21,9	521,0	0,0	0,0
7	657	19	124,5	0,1	368,6	66,6	66,5	12,2	18,7	637,0	0,0	0,0
8	615	19	35,5	0,1	366,6	70,3	67,3	12,7	23,8	589,0	4,0	0,7
9	610	17	39,5	0,1	356,8	65,9	100,9	10,8	21,8	593,0	0,0	0,0
10	747	18	125,5	0,1	340,9	62,5	65,5	13,0	20,4	717,0	12,0	1,6
11	734	18	124,5	0,1	354,0	62,4	65,2	10,5	20,8	707,0	9,0	1,3
12	760	20	123,5	0,1	344,9	59,8	75,2	9,8	18,9	736,0	2,0	0,3
13	512	16	82,5	0,1	382,8	73,6	97,7	10,2	26,0	489,0	4,0	0,8
14	515	15	124,5	0,1	424,7	74,6	68,6	9,4	17,6	499,0	1,0	0,2
15	467	13	126,5	0,1	426,7	76,1	64,7	8,4	18,1	430,0	24,0	5,3
16	494	15	125,5	0,1	416,4	77,3	75,5	12,8	17,6	477,0	0,0	0,0
17	776	18	65,5	0,1	332,9	63,8	51,0	7,4	16,4	756,0	2,0	0,3
18	706	17	113,5	0,1	351,0	67,2	66,8	6,4	19,9	593,0	94,0	13,7
19	574	15	124,5	0,1	393,8	72,6	60,4	6,3	16,0	557,0	2,0	0,4
20	530	16	125,5	0,1	388,3	72,2	96,2	7,5	17,1	509,0	2,0	0,4
21	653	17	37,5	0,1	362,0	65,7	72,5	7,7	23,4	636,0	0,0	0,0
22	540	15	37,5	0,1	415,9	71,5	59,8	8,4	16,5	454,0	70,0	13,4
23	459	12	86,5	0,1	454,9	79,0	49,7	3,6	14,5	447,0	0,0	0,0
24	510	14	86,5	0,1	389,3	73,4	68,0	7,5	23,0	496,0	0,0	0,0
25	503	15	114,5	0,1	431,5	74,7	64,5	6,0	21,4	481,0	7,0	1,4
26	566	17	126,5	0,1	405,9	74,8	57,7	8,9	16,4	549,0	0,0	0,0
27	314	13	127,5	0,1	506,0	97,0	72,7	10,0	20,7	299,0	2,0	0,7
28	411	11	39,5	0,1	456,3	77,8	83,7	8,7	21,1	242,0	158,0	39,5
29	464	14	37,5	0,1	451,3	79,2	62,6	6,7	16,3	447,0	2,0	0,4
30	554	13	55,5	0,0	263,5	67,8	66,0	10,7	16,8	309,0	231,0	42,8
31	401	14	123,5	0,1	454,5	80,3	106,7	9,6	21,8	385,0	2,0	0,5
32	466	14	126,5	0,1	441,4	76,5	87,3	6,7	14,9	449,0	3,0	0,7
33	436	15	128,5	0,1	430,1	84,8	78,9	11,1	18,3	416,0	3,0	0,7
34	491	16	127,5	0,1	424,7	78,2	60,7	9,5	18,0	472,0	2,0	0,4
35	362	14	126,5	0,1	477,3	89,0	71,7	10,7	18,2	346,0	1,0	0,3
36	317	12	37,5	0,1	535,4	96,1	57,7	7,6	21,6	301,0	4,0	1,3
37	395	11	84,5	0,1	494,0	86,1	53,4	7,8	18,0	384,0	0,0	0,0
38	494	12	89,5	0,1	437,4	75,5	53,2	8,1	17,7	480,0	1,0	0,2
39	433	9	82,5	0,1	460,0	78,5	54,1	8,2	16,4	317,0	107,0	25,2
40	433	12	81,5	0,1	454,3	81,8	59,8	6,7	20,4	421,0	0,0	0,0
41	419	13	80,5	0,1	449,0	81,2	71,8	8,7	23,9	402,0	4,0	1,0
42	517	15	126,5	0,1	436,9	77,5	59,7	9,6	16,8	499,0	0,0	0,0
43	528	12	42,5	0,1	425,8	72,4	59,6	9,8	14,5	338,0	177,0	34,4

Nota: θ = direção estimada do plantio (graus); dir_conf = confiança da direção dominante; row_sp = espaçamento estimado entre linhas; exp_sp = espaçamento esperado intralinha; CV (%) = coeficiente de variação do espaçamento; Duplas (%) e Falhas (%) = percentuais classificados; links_kept = conexões mantidas; links_drop = conexões descartadas; drop_pct = percentual de conexões descartadas.

3.2. Eficiência da rotulagem assistida por auto-label

A etapa de auto-label correspondeu à preparação e expansão do conjunto de treinamento, e não à avaliação da acurácia do modelo final. A partir dos 100 tiles anotados manualmente, o recurso de rotulagem assistida foi utilizado para propor anotações adicionais, que passaram obrigatoriamente por auditoria e correção humana antes de integrar o dataset. O tempo médio caiu de aproximadamente 78 segundos por tile na anotação manual completa para 15 segundos por tile na revisão de auto-rótulos, redução operacional próxima de 81%. Esse ganho evidencia a utilidade da assistência automatizada para ampliar conjuntos de segmentação, mas não elimina a necessidade de controle de qualidade: falsos positivos, falsos negativos e contornos incorretos precisam ser corrigidos para evitar a propagação de erros ao treinamento (ROBOFLOW, 2026).

3.3. Desempenho do modelo de segmentação

Após a preparação do dataset, o desempenho do modelo YOLOv8m foi avaliado no conjunto de validação. O mAP50 atingiu 0,963 para boxes e 0,960 para máscaras, indicando elevada capacidade de localizar e segmentar plântulas quando adotado IoU de 0,50. Em um critério mais rigoroso, que agrega limiares de IoU de 0,50 a 0,95, o mAP50–95 foi de 0,782 para boxes e 0,715 para máscaras. A redução entre mAP50 e mAP50–95 é esperada quando pequenas diferenças no contorno das máscaras passam a ser penalizadas com maior intensidade (LIN et al., 2014; ULTRALYTICS, 2026). Os valores de $\text{precision}(\text{mask})=0,915$ e $\text{recall}(\text{mask})=0,927$ também se mostraram equilibrados, indicando simultaneamente baixa proporção de falsos positivos e boa recuperação das instâncias presentes. Apesar dos resultados favoráveis, a generalização deve ser interpretada com cautela, pois o conjunto de dados representa um único talhão, um estágio fenológico e condições específicas de iluminação, solo e palhada.

3.4. Validação do pipeline de extração

A aplicação do critério de criticidade identificou quatro pontos com $\text{drop_pct} \geq 25\%$ (28, 30, 39 e 43). A exclusão desses casos evitou que geometrias locais incompatíveis com a estrutura regular das fileiras influenciassem as estatísticas de espaçamento. Nos 39 pontos mantidos, a taxa de descarte de conexões foi baixa na maior parte das imagens, o que indica que o filtro anti-zig-zag precisou rejeitar apenas uma pequena fração das ligações para preservar o encadeamento intralinha.

3.5. Estatísticas descritivas dos indicadores por ponto

A partir da estatística descritiva dos metadados exportados após a aplicação do filtro leve (remoção apenas dos pontos críticos; $n = 39$), observou-se mediana de $\%CV = 65,20\%$ (49,72–106,69), mediana de $\%duplas = 8,93\%$ (3,58–12,97) e mediana de $\%falhas = 18,68\%$ (14,54–27,63) (Tabela 2). A confiança direcional (dir_conf) variou de 0,053 a 0,075 (mediana = 0,062). A taxa de descarte de conexões (drop_pct) apresentou mediana de 0,375% e máximo de 13,68%, indicando que, no conjunto de imagens considerado válido, o filtro anti-zig-zag descartou apenas uma pequena fração das ligações consecutivas, sem incluir os pontos críticos removidos.

Tabela 2. Estatística descritiva dos indicadores de plantabilidade calculados nos pontos válidos ($n = 39$).

Variável	Mínimo	Mediana	Máximo
% falhas	14,54	18,68	27,63
% duplas	3,58	8,93	12,97
CV (%)	49,72	65,20	106,69
dir_conf	0,05	0,06	0,08
drop_pct	0,00	0,37	13,68

Nota: CV = coeficiente de variação; dir_conf = confiança da direção dominante; drop_pct = percentual de conexões descartadas pelo algoritmo.

3.6. Interpolação por krigagem e semivariogramas

As métricas de autocorrelação espacial dos dados pontuais e a validação cruzada dos modelos de krigagem foram obtidas no QGIS com o plugin Smart-Map (PEREIRA et al., 2022). O I de Moran não indicou autocorrelação espacial significativa a 5% para $\%CV$ ($I = 0,598$; $p = 0,072$), $\%duplas$ ($I = 0,596$; $p = 0,071$) ou $\%falhas$ ($I = 0,462$; $p = 0,218$). A validação cruzada

também apresentou baixos R^2 (%CV = 0,186; %duplas = 0,225; %falhas = 0,010), com RMSE de 10,642, 2,044 e 2,948, respectivamente. Em conjunto, esses resultados indicam estrutura espacial fraca e capacidade preditiva limitada para a densidade amostral e as condições avaliadas.

A interpretação conjunta do I de Moran e da validação cruzada é particularmente importante neste estudo. Nowosad e Meyer (2025) demonstram que o I de Moran é sensível ao tamanho amostral e à proximidade espacial entre observações e, por isso, não deve ser empregado isoladamente como medida de qualidade de um modelo espacial. No presente conjunto, com $n = 39$ após o filtro, os valores relativamente elevados de I para %CV e %duplas, mas sem significância a 5%, não são suficientes para afirmar dependência espacial robusta. O baixo desempenho preditivo na validação cruzada reforça essa leitura e sustenta a interpretação cautelosa dos mapas como produtos exploratórios, e não como superfícies de alta precisão.

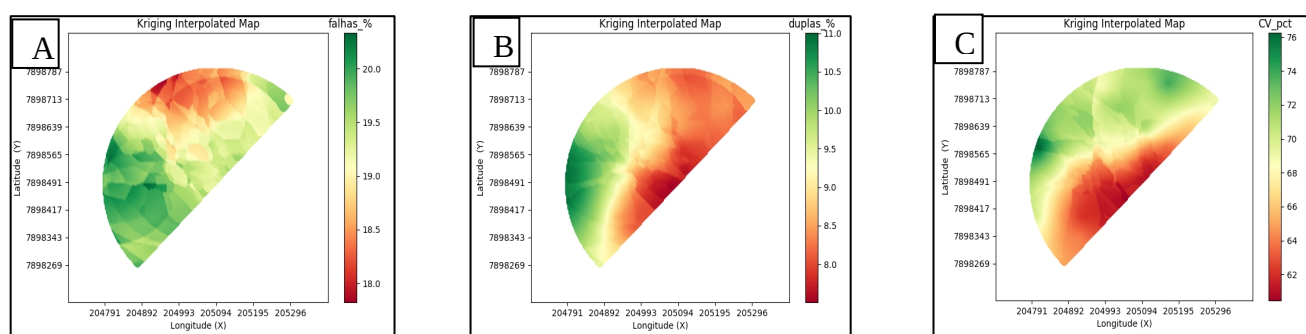
Do ponto de vista agrônômico, os indicadores extraídos (%falhas, %duplas e %CV) descrevem a uniformidade do estande, mas não identificam isoladamente a causa da anomalia. Valores elevados de falhas podem estar associados a problemas de deposição ou fechamento do sulco, contato semente-solo, qualidade de sementes ou emergência desuniforme; valores elevados de duplas/múltiplos podem refletir irregularidades de distribuição e deposição. Nardon e Botta (2022) discutem o uso de métricas de espaçamento intralinha para avaliação da semeadura, enquanto Pereyra et al. (2022) mostram que a uniformidade espacial entre plantas pode se relacionar à produtividade da soja em condições específicas. Assim, os mapas gerados aqui devem ser entendidos como ferramentas de diagnóstico e direcionamento de investigação em campo.

Os resultados de validação cruzada reforçam que a qualidade dos mapas krigados deve ser interpretada com cautela. R^2 baixos podem ocorrer quando a variabilidade local é elevada em relação ao sinal espacial capturado pela malha, quando o variograma ajustado representa de forma limitada a estrutura dos dados ou quando a variável apresenta comportamento não estacionário. Para estudos futuros, recomenda-se ampliar a densidade amostral, avaliar anisotropia, inspecionar resíduos e testar covariáveis de interesse, como atributos do relevo, índices de vegetação ou condutividade elétrica (DAMIAN et al., 2020).

Tabela 3. Síntese dos indicadores de autocorrelação espacial e validação cruzada por variável.

Variável	I de Moran	p-valor	RMSE (CV)	R ² (CV)
%CV	0,6	0,07	10,64	0,18
%duplas	0,6	0,07	2,04	0,22
%falhas	0,46	0,21	2,94	0,01

Visualmente, observou-se tendência de maiores concentrações de falhas e duplas em setores de maior altitude do pivô (Figura 10). O resultado é apresentado como observação exploratória e como exemplo de como o fluxo de trabalho pode revelar padrões relevantes para investigações futuras.

**Figura 10.** Mapas de superfície interpolados por krigagem para % de falhas (A), % de duplas (B) e %CV (C).

3.7. Observações adicionais e limitações

Como análise exploratória adicional, foi calculada a matriz de correlação de Pearson entre os indicadores de plantabilidade e o MDE (Tabela 4). Os coeficientes observados não foram utilizados para estabelecer relações causais, mas indicam hipóteses a serem testadas em estudos com maior número de observações e delineamento específico. A associação observada entre MDE e alguns indicadores reforça o interesse em avaliar, futuramente, o relevo e outras covariáveis ambientais como possíveis fatores relacionados à variabilidade da implantação da cultura.

Tabela 4. Matriz de correlação de Pearson entre os indicadores de plantabilidade e o MDE.

	Falhas	Duplas	CV	MDE
Falhas	1	0,77	0,19	0,11
Duplas	0,77	1	0,32	0,46
CV	0,19	0,32	1	0,35
MDE	0,11	0,46	0,35	1

As principais limitações do estudo incluem: (I) ausência de georreferenciamento e escala absoluta nas imagens pontuais, o que motivou o uso de regras relativas para definição do espaçamento esperado; (II) dependência de um único talhão e estágio fenológico, limitando a generalização; (III) sensibilidade potencial do modelo de visão a condições de campo não representadas no dataset; e (IV) escolhas relacionadas ao ajuste do variograma e ausência de transformação dos dados, que podem afetar o desempenho da krigagem em distribuições assimétricas. Apesar dessas limitações, o fluxo proposto mostrou-se consistente e replicável, fornecendo um caminho prático para transformar imagens em métricas e mapas.

4. CONCLUSÕES

Foi possível estruturar e validar um fluxo de trabalho para aferição de indicadores de plantabilidade em soja a partir de imagens pontuais obtidas por ARP, integrando segmentação por visão computacional, extração de centroides no QGIS e cálculo automatizado de métricas de espaçamento.

A divisão do dataset por ponto amostral reduziu o risco de vazamento de informação entre tiles e conjuntos, enquanto a rotulagem assistida por auto-label, seguida de auditoria humana, reduziu substancialmente o tempo necessário para ampliar o conjunto de anotações. O modelo de segmentação apresentou elevado desempenho no conjunto de validação, embora sua generalização para outros talhões e condições ainda dependa da incorporação de novas imagens originais.

Como controle de qualidade do pós-processamento, o indicador `drop_pct` permitiu identificar pontos críticos; o limiar operacional de 25%, definido especificamente para este estudo, resultou na exclusão dos pontos 28, 30, 39 e 43 e em $n = 39$ para as análises posteriores. Nos pontos válidos, a baixa taxa de descarte de conexões indicou estabilidade do encadeamento intralinha na maior parte das imagens.

A krigagem ordinária possibilitou a geração de semivariogramas e mapas temáticos a partir das observações pontuais. Contudo, a ausência de autocorrelação espacial significativa a 5% e os baixos R^2 da validação cruzada indicam estrutura espacial fraca e capacidade preditiva limitada nas condições avaliadas; por isso, os produtos cartográficos devem ser interpretados como exploratórios.

Para estudos futuros, recomenda-se: (I) ampliar a base de imagens originais rotuladas, contemplando outros talhões, estádios fenológicos e condições de solo, palhada e iluminação; (II) avaliar mapas de incerteza e sua utilidade no planejamento amostral; (III) incorporar georreferenciamento e/ou escala métrica robusta às imagens pontuais, permitindo expressar as métricas intralinha em unidades reais e realizar validações diretas; (IV) investigar relações entre atributos do relevo e a variabilidade da plantabilidade; e (V) aumentar a densidade de amostras para avaliar com maior robustez a presença de dependências e correlações espaciais.

REFERÊNCIAS

- CORRÊA, M. F.; GADOTTI, G. I.; PINHEIRO, R. M.; NADAL, A. P.; VERGARA, R. O.; NAVROSKI, R.; SCHUCH, L. O. B. Spatial variability of soil in soybean seed production and its yield components. *Diversitas Journal*, v. 7, n. 2, p. 542-554, 2022.
- DAMIAN, J. M.; PIAS, O. H. C.; CHERUBIN, M. R.; FONSECA, A. Z.; FORNARI, E. Z.; SANTI, A. L. Applying the NDVI from satellite images in delimiting management zones for annual crops. *Scientia Agrícola*, v. 77, n. 1, e20180055, 2020. DOI: 10.1590/1678-992X-2018-0055.
- FEHR, W. R.; CAVINESS, C. E. Stages of soybean development. Ames: Iowa State University, 1977. Special Report 80.
- LIN, T.-Y.; MAIRE, M.; BELONGIE, S.; HAYS, J.; PERONA, P.; RAMANAN, D.; DOLLÁR, P.; ZITNICK, C. L. Microsoft COCO: Common Objects in Context. In: EUROPEAN CONFERENCE ON COMPUTER VISION - ECCV, 2014. Proceedings [...]. Cham: Springer, 2014. p. 740-755. DOI: 10.1007/978-3-319-10602-1_48.
- MULLA, D. J. Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. *Biosystems Engineering*, v. 114, n. 4, p. 358-371, 2013. DOI: 10.1016/j.biosystemseng.2012.08.009.
- NARDON, G. F.; BOTTA, G. F. Prospective study of the technology for evaluating and measuring in-row seed spacing for precision planting: a review. *Spanish Journal of Agricultural Research*, v. 20, n. 4, e02R01, 2022. DOI: 10.5424/sjar/2022204-19269.
- NOWOSAD, J.; MEYER, H. Investigating Moran's I Properties for Spatial Machine Learning: A Preliminary Analysis. *AGILE GIScience Series*, v. 6, art. 40, 2025. DOI: 10.5194/agile-giss-6-40-2025.
- PEREIRA, G. W. et al. Smart-Map: An Open-Source QGIS Plugin for Digital Mapping Using Machine Learning Techniques and Ordinary Kriging. *Agronomy*, v. 12, n. 6, art. 1350, 2022. DOI: 10.3390/agronomy12061350.
- PEREYRA, V. M. et al. Early-season plant-to-plant spatial uniformity can affect soybean yields. *Scientific Reports*, v. 12, art. 17128, 2022. DOI: 10.1038/s41598-022-21385-z.
- PYTHON SOFTWARE FOUNDATION. Python 3 documentation. 2026. Disponível em: <https://docs.python.org/3/>. Acesso em: 10 ago. 2026.
- QGIS DEVELOPMENT TEAM. QGIS Geographic Information System: version 3.40 LTR. QGIS Association, 2024. Disponível em: <https://docs.qgis.org/3.40/>. Acesso em: 10 ago. 2026.
- ROBOFLOW. Roboflow Annotate documentation. 2026. Disponível em: <https://docs.roboflow.com/annotate>. Acesso em: 10 ago. 2026.
- TSOUROS, D. C.; BIBI, S.; SARIGIANNIDIS, P. G. A Review on UAV-Based Applications for Precision Agriculture. *Information*, v. 10, n. 11, art. 349, 2019. DOI: 10.3390/info10110349.
- ULTRALYTICS. Ultralytics YOLO documentation: instance segmentation and performance metrics. 2026. Disponível em: <https://docs.ultralytics.com/>. Acesso em: 10 ago. 2026.
- WEBODM. WebODM documentation. 2026. Disponível em: <https://docs.webodm.org/>. Acesso em: 10 ago. 2026.

WEIRICH NETO, P. H.; FORNARI, A. J.; JUSTINO, A.; GARCIA, L. C. Qualidade na semeadura do milho. Engenharia Agrícola, v. 35, n. 1, p. 171-179, 2015. DOI: 10.1590/1809-4430-Eng.Agric.v35n1p171-179/2015.