

REVISITANDO O TEMA DOS ERROS EM DADOS OBTIDOS POR MONITORES DE PRODUTIVIDADE DE MILHO E SOJA

**LUIZ GUSTAVO DE GÓES STERLE¹; MARCELO CAMPONEZ DO BRASIL CARDINALI^{1,2};
LEONARDO VEDOVATO GIMENEZ¹; VINICIUS CAMPANA BENASSI¹; JOSÉ PAULO MOLIN¹**

¹Departamento de Engenharia de Biosistemas, Escola Superior de Agricultura Luiz de Queiroz (ESALQ/USP). Av. Pádua Dias, 11, Piracicaba – SP, Brasil. gustavosterle@usp.br / (18) 9 9623-5962;

²Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, IFSP – Câmpus Piracicaba, Av. Diácono Jair de Oliveira, 1005 - Santa Rosa, Piracicaba – SP, Brasil.

Apresentado no
Congresso Brasileiro de Agricultura de Precisão e Digital - ConBAP 2024
Porto Alegre, RS, 02 a 04 de julho de 2024

RESUMO: Uma das ferramentas mais difundidas na agricultura de precisão são os mapas de produtividade. Considerados essenciais para a gestão agrícola espacializada, sua disseminação enfrenta vários desafios. Dentre estes, a obtenção dos dados requer um processo de pré-processamento para remover erros. O objetivo deste estudo consiste em analisar, quantificar e caracterizar dados errôneos presentes em conjuntos atuais de dados provenientes de lavouras comerciais e abordar aspectos de evolução da qualidade dos dados, avaliando simultaneamente dois métodos de filtragem, estatística e MapFilter 2.0. Na filtragem estatística, realizou-se uma análise exploratória dos dados. A partir dos quartis superior e inferior, foram estabelecidos os limites para a filtragem dos valores de produtividade. Comparativamente, utilizou-se o software MapFilter 2.0, que permite realizar a filtragem estatística e posteriormente a filtragem localizada, por meio da avaliação comparativa de valores de pontos vizinhos. O processo de filtragem estatística removeu entre 1,8 e 14,9%, o software Mapfilter entre 6,8 e 38,9% de dados errôneos e discrepantes. Apesar de ocorrerem em menor frequência nos dados de produtividade atuais, alguns erros persistem, principalmente relacionado aos valores decorrentes de leituras de produtividade nula, indicando que a etapa de filtragem de dados segue sendo fundamental para a obtenção de mapas de produtividade confiáveis.

PALAVRAS CHAVES: Mapas de Produtividade, Sensores, Filtragem de dados.

REVISITING THE THEME OF ERRORS IN DATA OBTAINED BY CORN AND SOYBEAN YIELD MONITORS

ABSTRACT: Yield mapping is one of the most widely used tools in precision agriculture and is essential for spatial agricultural management. However, its dissemination faces several challenges, particularly in data acquisition, which requires preprocessing to remove errors. This study aims to analyze, quantify, and characterize erroneous data in current datasets from commercial crops and to address the evolution of data quality by evaluating two filtering methods: statistical filtering and MapFilter 2.0. For statistical filtering, exploratory data analysis was conducted, and upper and lower quartiles were used to set limits for filtering yield values. The MapFilter 2.0 software was also used, which allows for statistical filtering followed by localized filtering through comparative evaluation of neighboring point values. The statistical filtering process removed between 1.8% and 14.9% of erroneous and discrepant data, while the MapFilter software removed between 6.8% and 38.9%. Although errors occur less frequently in current productivity data, some persist, especially those related to zero yield readings, indicating that data filtering remains essential for obtaining reliable yield maps.

KEYWORDS: Yield Maps, Sensors, Data Filtering.

INTRODUÇÃO

Os monitores de produtividade em colhedoras de grãos foram introduzidos há mais de três décadas, mas sua adoção na agricultura brasileira tem pouco mais de duas décadas. A disseminação dos mapas de produtividade, considerados essenciais para a gestão agrícola especializada, enfrenta vários desafios, principalmente no que diz respeito à confiabilidade dos dados. Esses monitores combinam dados de sensores, como posicionamento por sistemas globais de navegação por satélite (GNSS), velocidade da colhedora, largura da plataforma, fluxo de grãos e umidade, para calcular a produtividade. A precisão desses sistemas comerciais continua a evoluir e a variedade de opções disponíveis está aumentando (FULTON et al. 2018). Os dados do monitor de produtividade contêm erros de medição associados ao próprio processo de mapeamento da produtividade (PING e DOBERMANN, 2005). Menegatti e Molin (2003) apontaram que o erro de produtividade e umidade nula, à época, era decorrente da falha na leitura pelos sensores e com isso tem-se pontos com valores nulos. O erro da distância nula ocorre quando a distância entre os pontos coletados pelo GNSS é nula. Este erro pode ser decorrente da alta taxa de coleta de pontos associada a pequenos erros de posicionamento do GNSS. Outro erro que pode estar presente nos dados coletados é o erro da largura de corte, decorrente da divergência entre o valor da largura informada no monitor de colheita e a largura parcial efetiva na plataforma de corte. A obtenção dos dados requer um processo de pré-processamento para remover erros (BLACKMORE e MARSHALL, 1996; STAFFORD et al., 1996; LARK et al., 1997; MOORE, 1998; BLACKMORE e MOORE, 1999; THYLÉN et al., 2000; ARSLAN e COLVIN, 2002b). Diferentes sistemas de filtragem de dados têm sido utilizados para melhorar a qualidade da informação obtida pelos mapas. Gimenez e Molin (2004) propuseram uma metodologia para remoção de erros na largura de corte na plataforma, ao avaliar valores médios de pontos coletados, e compará-los com a distância ao ponto imediatamente anterior. Simbahan et al. (2004) analisaram os dados do monitor de produtividade para eliminar erros comuns, aplicando um algoritmo de filtragem sequencial, que rastreia e exclui seis tipos de valores errôneos. Zanella et al. (2018) realizaram a remoção de erros em três etapas principais: primeiro a remoção de dados fora da realidade utilizando limiares superiores e inferiores de produtividade, na segunda etapa a remoção de dados discrepantes globais estabelecendo limiares baseados na estatística descritiva, utilizando o desvio padrão e por fim considerando os pontos vizinhos em um determinado raio. Programas desenvolvidos para essa finalidade são ferramentas utilizadas, a exemplo de Maldaner et al. (2021), que construíram um aplicativo-algoritmo denominado MapFilter 2.0 com a metodologia de remoção de outliers espaciais. Portanto, existem métodos baseados em estatística, em distância, em densidade e em clusters (SMITI, 2020). O objetivo deste estudo consiste em analisar, quantificar e caracterizar dados errôneos presentes em conjuntos atuais de dados provenientes de lavouras comerciais e abordar aspectos de evolução da qualidade dos dados. Simultaneamente, são avaliados e comparados a filtragem estatística e utilizando o MapFilter 2.0, como estratégia para lidar com esses erros.

MATERIAL E MÉTODOS

Caracterização dos dados

Foram obtidos dados de produtividade de 12 lavouras de milho e soja, provenientes de colhedoras comerciais amplamente utilizadas no Brasil (TAB. 1). Sete conjuntos de dados são de produtividade da cultura do milho e cinco da cultura da soja. Como as áreas não foram completamente mapeadas, a área efetivamente colhida foi obtida multiplicando-se a distância percorrida, pela largura de corte da colhedora. As áreas colhidas possuem entre 27,4 e 226,4 ha, oriundos de propriedades dos estados do Paraná, Mato Grosso e Piauí, entre os anos de 2013 e 2016. A densidade de coleta de dados variou de 398 a 648 pontos ha⁻¹, dependendo da configuração da frequência de coleta do monitor de produtividade e da largura de corte. Em média, a densidade de pontos foi superior na cultura de soja (ID 8 a 12) em comparação com a cultura de milho (ID 1 a 7), devido à maior frequência de coleta dos dados em algumas colhedoras de soja, nos casos analisados neste trabalho. Também são apresentados os dados das máquinas, com informações referentes à largura de corte nominal (plataforma) e frequência de coleta de dados.

TABELA 1. Informações gerais das máquinas e dos dados obtidos nos 12 talhões.

ID	Estado	Ano de colheita	Cultura	Var.	Nº de Pontos	Área Colhida (ha)	Densidade de Pontos (ha)	Máquina	Largura de Corte (Plataforma)	Frequência de coleta (Hz)
1	MT	2016	Milho	MG 652	81224	168,5	482,0	JD	7,5	1
2	MT	2016	Milho	MG 652	21765	46,2	471,1	JD	7,5	1

3	MT	2016	Milho	MG 652	29406	58,6	501,8	JD	7,5	1
4	MT	2015	Milho	6030	81476	150,9	539,9	JD	7,5	1
5	MT	2016	Milho	MG 652	72710	145,6	499,4	JD	7,5	1
6	MT	2016	Milho	MG 652	56687	123,1	460,5	JD	7,5	1
7	MT	2016	Milho	MG 652	106035	226,4	468,4	JD	7,5	1
8	MT	2013	Soja	-	64852	124,7	520,1	CASE	11,8	0,2
9	MT	2013	Soja	-	16050	27,4	585,8	CASE	11,8	1
10	PI	2016	Soja	19342	31485	79,1	398,0	JD	10,25	1
11	MT	2014	Soja	-	106641	222,2	479,9	CASE	10,65	0,33
12	PR	2015	Soja	-	35145	54,2	648,4	CASE	9,14	0,2

Os dados foram exportados da plataforma Operations Center (John Deere), em formato shapefile, para serem trabalhados em ambiente SIG, com o software QGIS, versão 3.22.9 Bialowieza (QGIS, 2023). Os arquivos não continham a coluna de informação das coordenadas, as quais foram inseridas utilizando o complemento Lat Lon Tools versão 3.6.12 (HAMILTON, 2023), utilizando o sistema de projeção de coordenadas métricas, UTM, e os fusos das respectivas propriedades. A camada vetorial de pontos foi exportada no formato CSV para análise exploratória dos dados, classificação dos tipos de erro e realização das filtrações estatísticas. Foi desenvolvido um script em linguagem Python para realização da análise exploratória, classificação dos tipos de erro, pré-filtragem e filtragem estatística.

Pré-filtragem

Na pré-filtragem dos pontos coletados de cada talhão, foram identificadas as frequências de erros de produtividade nula (PN), umidade nula (UN) e distância nula (DN). Os erros de largura de corte (LC) não puderam ser contabilizados devido à ausência de dados. Foram eliminados todos os pontos nessas condições por meio da filtragem estatística. A distância (D) entre dois pontos consecutivos foi obtida por

$$D = \sqrt{IDN^2 + IDL^2} \quad (1)$$

em que: IDN e IDL são as distâncias norte-sul e leste-oeste entre dois pontos consecutivos, respectivamente. Para valores de D inferiores a 0,01 m, os pontos foram considerados como erro de posicionamento e foram eliminados.

Filtragem estatística

A filtragem estatística, classificada segundo Smiti (2020), como um método estatístico, não paramétrico, que leva em consideração intervalos interquartílicos para detecção de outliers. Esta etapa foi realizada com base na metodologia proposta por Tukey (1997). Realizou-se uma análise exploratória dos dados, incluindo o cálculo da média, mediana, quartil superior (QS) e quartil inferior (QI), desvio padrão e coeficiente de variação. A partir dos quartis superior e inferior, foram estabelecidos os limites superiores (LS) e inferiores (LI) para a filtragem dos valores de produtividade, dados por:

$$LS = QS + 1,5(QS - QI) \quad (2)$$

$$LI = QI - 1,5(QS - QI) \quad (3)$$

MapFilter 2.0

O software permite realizar a filtragem estatística para remoção de erros grosseiros e posteriormente a filtragem espacial, por meio da avaliação comparativa de valores de produtividade de pontos próximos. É, portanto, uma filtragem primeiramente estatística e posteriormente localizada dos dados, baseado em distância, que considera os respectivos valores da vizinhança, sendo esta última isotrópica e anisotrópica. A partir da mediana, foram estabelecidos os limites superiores (*LimS*) e inferiores (*LimI*) para a filtragem dos valores de produtividade, dados por:

$$LimS = MK + MK \cdot v \quad (4)$$

$$LimI = MK - MK \cdot v \quad (5)$$

em que: *LimS* é o limite superior; *LimI* é o limite inferior; *MK* é a mediana de todos os valores localizados no conjunto de dados; *v* é a variação máxima aceita para a mediana.

Um outlier estatístico no conjunto de dados é um ponto com um valor maior ou menor que os limites de corte superior e inferior, respectivamente. O filtro local ocorre em duas etapas: filtros locais anisotrópicos e isotrópicos. O filtro anisotrópico desenvolvido por Maldaner e Molin (2020) foi originalmente usado para filtrar dados de produtividade na colheita de cana-de-açúcar, evitando comparativo com vizinhos laterais, que podem conter ausência de plantas (falhas). O filtro detectou todos os pontos localizados dentro de um raio (*R*) em torno de um ponto x_i dentro de uma única direção. O ponto x_i é comparado com *k* vizinhos anteriores e subsequentes, sendo *k*, o número de vizinhos cuja distância euclidiana é menor ou igual ao raio *R* (linha azul na Figura 1). A mediana desses *k* vizinhos foi calculada e as Eq. 4 e 5 foram aplicadas ao ponto x_i . Se o valor do ponto x_i foi maior ou menor do que os limites de corte superior e inferior, é considerado um outlier local e então excluídos do conjunto de dados. No filtro isotrópico, metodologia criada por Spekken et al. (2013), foi adaptada para identificar outliers em um conjunto de dados. O filtro isotrópico, ao contrário, detectou todos os *k* pontos vizinhos localizados em um raio *R* em torno de um ponto x_i em qualquer direção (Figura 1). Então, a mediana desses *k* vizinhos foi calculada e a Eq. 4 e Eq. 5 foram aplicadas ao ponto x_i . Enquanto o proposto por Spekken et al. (2013) adiciona peso para os pontos com valores fora do limite de corte, a metodologia do MapFilter 2.0 exclui o ponto x_i com um valor maior ou menor que o corte superior e inferior.

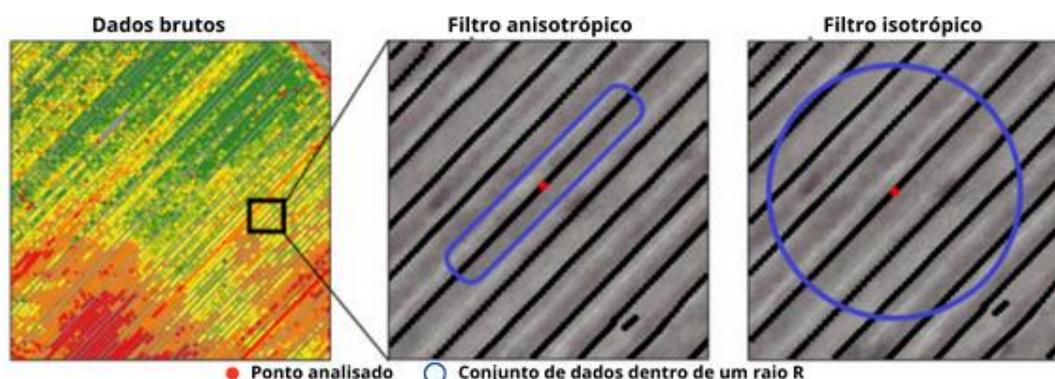


FIGURA 1. Filtragem local dos dados, isotrópica e anisotrópica (MALDANER et al., 2021).

Parâmetros de entrada para filtragem no MapFilter foram: 50% como variação máxima aceita para a mediana (*v*) para a filtragem estatística e 75% para a filtragem local do software, essas variações aceitas foram as mesmas para todos os talhões. O parâmetro que diferiu foi a dependência espacial, utilizando a largura média da plataforma de corte (Tabela 1) multiplicado por dez.

RESULTADOS E DISCUSSÃO

A Tabela 2 apresenta a frequência de erros obtida com o processo de pré-filtragem dos dados. A frequência de erros de produtividade nula (PN) apresentou valores que variaram entre 0,1 e 10,1%, enquanto as ocorrências de dados de umidade nula (UN) e distância nula (DN) foram muito baixas ou ausentes. Menegatti e Molin (2003), estudando seis diferentes monitores de produtividade, encontraram os valores variando entre 0,0 e 4,6% para (PN), 1,0 e 9,5% para (UN) e 0,0 e 11,5% para (DN).

TABELA 2. Frequência dos erros, dados de produtividade brutos e após filtragem estatística e local com o aplicativo MapFilter 2.0.

Frequência de erros (%)					Dados brutos			Dados após filtragem estatística				Dados filtrados no MapFilter 2.0			
ID	PN	UN	DN	LC	N	\bar{x}	DP	N	\bar{x}	DP	DR	N	\bar{x}	DP	DR
						-- ton/ha --			-- ton/ha --		%		-- ton/ha --		%
1*	1,55	0,00	0,21	-	81224	5,82	1,74	75467	6,00	1,21	7,08	68103	6,10	1,11	16,15
2*	0,24	0,00	0,10	-	21765	7,37	1,85	19889	7,53	1,23	8,62	19322	7,49	1,21	11,22
3*	0,70	0,00	0,00	-	29406	7,71	1,86	26364	7,74	0,90	10,34	25878	7,75	0,87	12,00
4*	2,22	0,00	0,00	-	81476	6,72	2,15	70011	7,24	0,95	14,07	67819	7,23	0,91	16,76
5*	0,34	0,00	0,00	-	72710	8,33	1,52	64600	8,31	0,58	11,15	66653	8,29	0,70	8,33
6*	0,72	0,00	0,00	-	56687	7,64	1,95	48919	7,95	0,90	13,70	49186	7,88	1,01	13,23
7*	1,22	0,00	0,00	-	106035	4,22	1,26	96408	4,34	0,79	9,08	89591	4,35	0,70	15,51
8**	2,05	0,00	0,00	-	64852	1,77	2,78	60325	1,76	0,24	6,98	59344	1,76	0,23	8,49
9**	10,11	0,00	0,00	-	16050	8,05	80,14	13665	7,05	1,63	14,86	12258	7,35	1,13	23,63
10**	0,28	0,00	0,00	-	31485	1,80	0,37	29872	1,83	0,27	5,12	29349	1,82	0,27	6,78
11**	1,39	0,00	0,00	-	106641	1,61	0,81	104672	1,62	0,71	1,85	66462	1,58	0,44	37,68
12**	0,11	0,00	0,00	-	35145	4,09	7,01	32697	3,63	1,24	6,97	21479	3,77	0,69	38,88

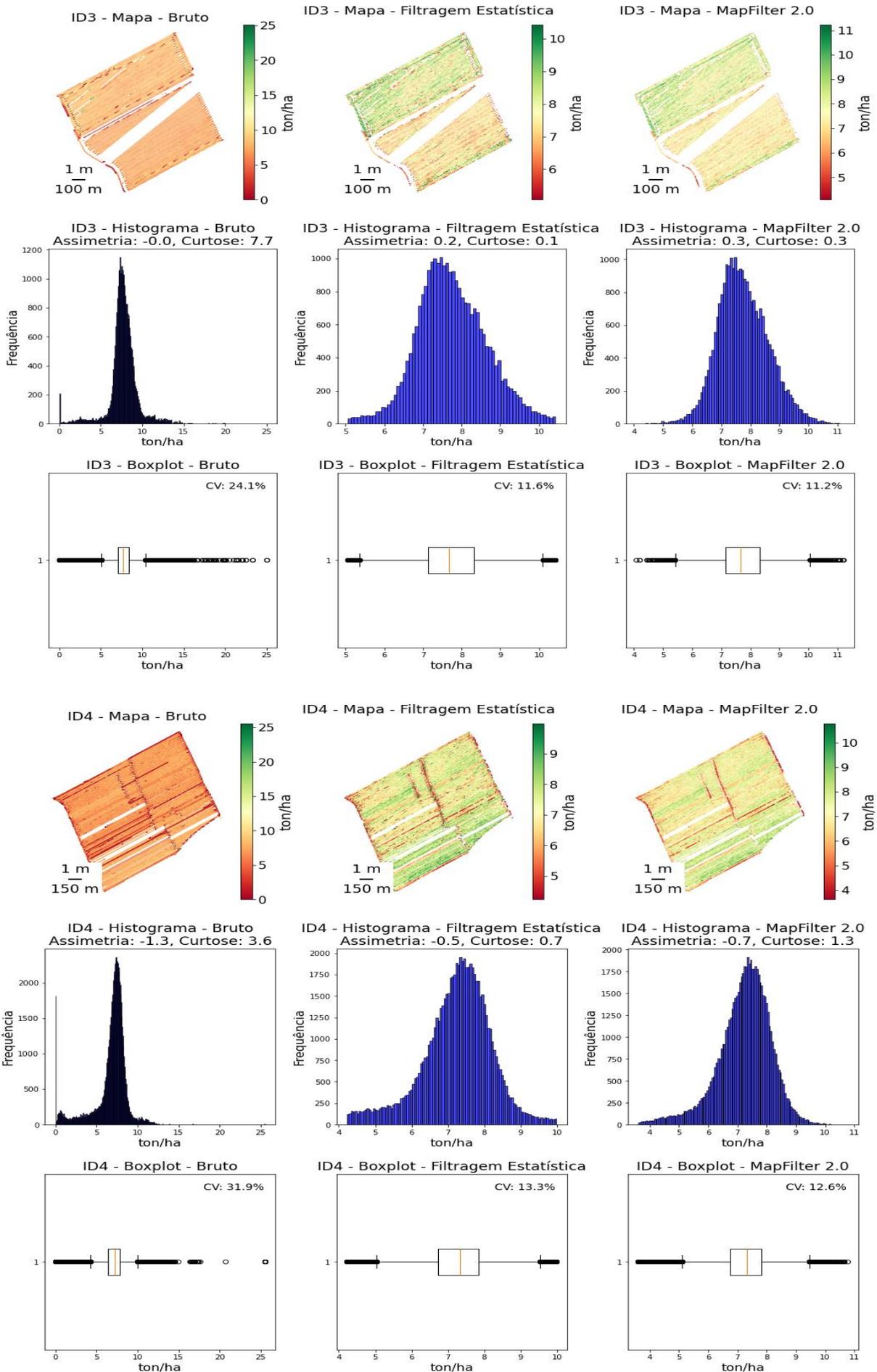
* milho; ** soja; PN: Produtividade Nula; UN: Umidade Nula; DN: Distância Nula; LC: Largura de Corte; N: Número de Pontos; \bar{x} : Média; DP: Desvio Padrão; DR: Dados Removidos

Com exceção do talhão 11, os demais tiveram redução de mais de 5% do total de pontos de produtividade, com valores variando entre 1,8 e 14,9% de dados removidos após a filtragem estatística. Com o uso do software MapFilter 2.0, com exceção dos talhões 5, 8 e 10, os demais tiveram redução de mais de 10% do total de pontos de produtividade, com valores variando entre 6,8 e 38,9% de dados removidos após a filtragem. A filtragem com o MapFilter 2.0 removeu, na média, 8,3% mais pontos, em comparação à filtragem estatística. Menegatti e Molin (2003), após realizarem filtragem dos dados de diferentes monitores, obtiveram resultados de remoção de pontos de produtividade de milho e soja na ordem de 4,6 a 27,3%. Gimenez e Molin (2004) relataram que o número de dados eliminados pelo filtro, utilizando funções de lógica, para realizar a eliminação dos erros foi bastante diferente em termos proporcionais, sendo eliminados 41% dos dados de soja e 21% dos dados de milho. Thylén et al. (2000), usando uma técnica de três etapas, removeram 10, 25 e 50% dos dados. O algoritmo de filtragem proposto por Simbahan et al. (2004) removeu entre 13 e 20% dos dados originais do monitor de produtividade de milho e soja, baseado em uma ordem lógica e sequencial de filtragem de dados, com a maioria da remoção ocorrendo durante a triagem primária de erros técnicos no processo de monitoramento da produtividade, resultados muito próximos aos encontrados por Sun et al. (2013), de 13,1 a 19,6%. De acordo com Zanella et al. (2018), os pontos removidos nos dados de produtividade de milho representaram 9,3 % do total de pontos. A maioria dos estudos sugere que a remoção de dados variou de 0,4 a 32% (SHEARER et al., 1997; BLACKMORE e MOORE, 1999; BECK et al., 2001; ROBINSON e METTERNICHT, 2005).

O aumento no valor da média de produtividade dos dados filtrados foi notado para a maioria dos talhões analisados, a qual se deve à eliminação de muitos pontos onde os valores de produtividade são bastante baixos, dado que corrobora com os estudos de Lyle et al. (2014).

Em todos os casos da filtragem estatística houve uma redução do desvio padrão dos dados, com a menor redução, de aproximadamente 12%, no talhão 12 e maior redução, de 98%, no talhão 9, com média de 53%. O mesmo fenômeno ocorre com a filtragem no MapFilter 2.0, com a menor redução, de 27% no talhão 10 e a maior redução, de 99% no talhão 9, com média de 57,4%. Essas reduções ocorrem por conta de pontos de produtividade muito elevadas ou muito baixas com relação à média. Sun et al. (2013) também observaram a redução no valor do desvio padrão do conjunto de dados filtrados em relação aos dados não filtrados. Portanto, também ocorre redução do coeficiente de variação, devido aumento da média e redução do desvio padrão.

A Figura 2 apresenta alguns dos mapas de produtividade brutos e após filtragem estatística e local no MapFilter 2.0, onde é possível observar locais onde ocorrem maior remoção de pontos, identificados como de baixa produtividade, nas bordas dos talhões. Também apresenta os histogramas e os boxplots dos dados brutos, após filtragem estatística e após a filtragem local (MapFilter 2.0).



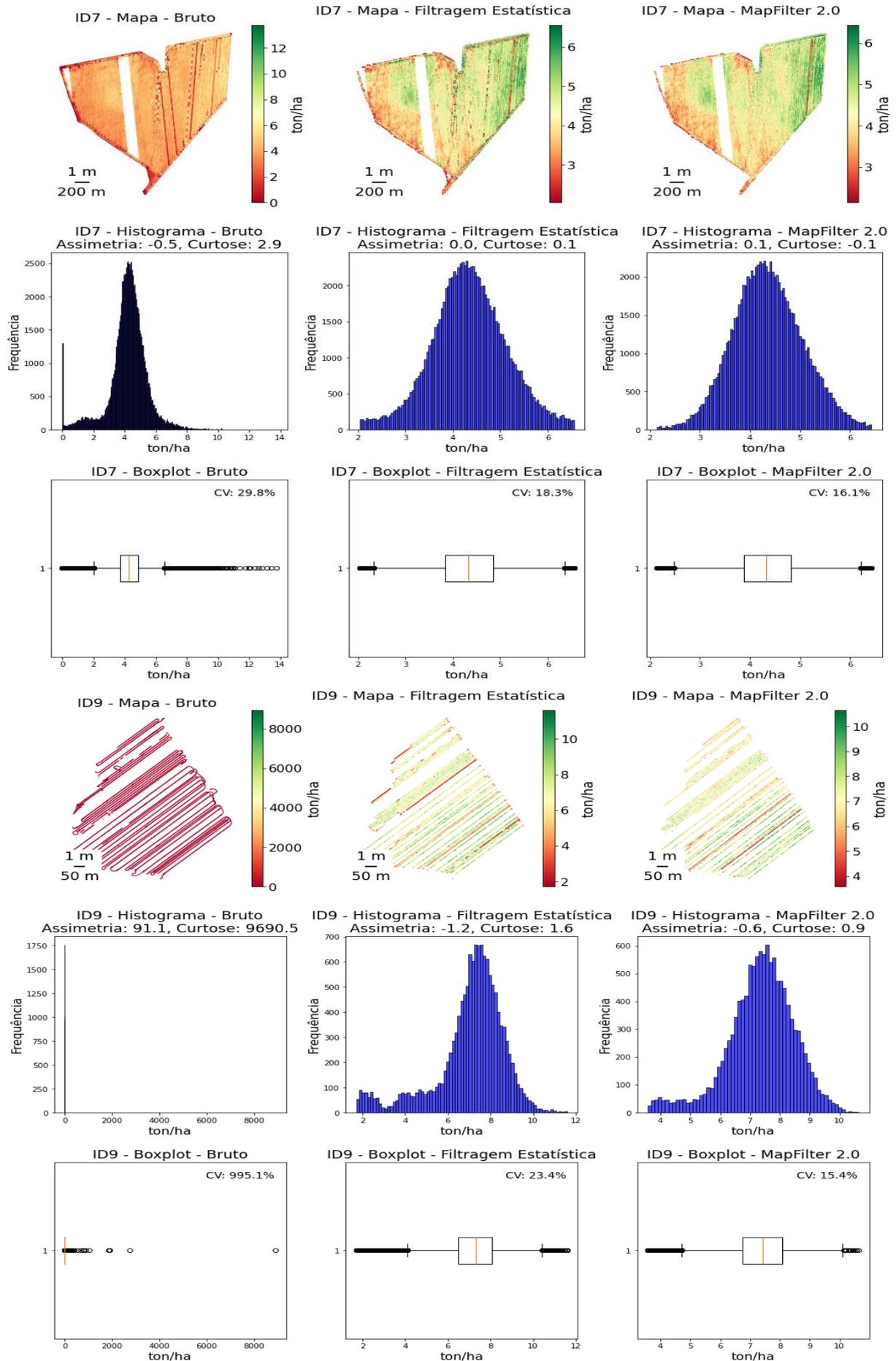


FIGURA 2. Mapas de produtividade, Histograma e boxplot dos dados brutos e após filtragem estatística e local no MapFilter 2.0 dos talhões 3, 4, 7 e 9.

Um desafio importante para avaliar o benefício da filtragem é a seleção de critérios de validação precisos (PING e DOBERMANN, 2005). Outro desafio é a potencial remoção de grandes quantidades de dados. Isso varia com base na abrangência dos algoritmos implementados e nos valores de limiar escolhidos. Diferentes filtragens de dados de produtividade provavelmente resultarão em diferentes quantidades e localizações de dados removidos. A remoção parcial de pontos de produtividade errôneos pode não melhorar efetivamente a precisão do mapa, enquanto a remoção excessiva pode causar uma perda significativa de dados úteis de produtividade. Métodos para avaliar a filtragem podem ser utilizados, no presente trabalho utilizou-se parâmetros estatísticos, como distribuição de frequência, assimetria, curtose, média, desvio padrão, porcentagem de dados removidos, coeficiente de variação. No entanto existem outras ferramentas importantes para avaliar algoritmos de filtragem, e podem ser usadas para ajudar na seleção de parâmetros apropriados no processo de remoção de dados errôneos, como mapas de dados removidos, semivariogramas antes e depois da filtragem, e validação cruzada (jackknife) (SIMBAHAN et al., 2004).

A forma como os histogramas dos dados brutos se apresentam afasta do formato da distribuição normal, a qual é uma característica intrínseca da variabilidade espacial da produtividade no talhão, podendo ser resultado de registros imprecisos do monitor de produtividade da colhedora, que resultam em dados discrepantes. Após as filtragens o comportamento dos histogramas se altera, tomando uma forma que se aproxima à distribuição normal. Características estatísticas dos dados brutos de produtividade de grãos, sugeridas por Blackmore (2003), seria que os histogramas de rendimento derivados de conjuntos de dados de produtividade devem ser normalmente distribuídos. Isso é assumido porque, com a coleta de grandes números de observações de amostragem, a distribuição amostral se aproximará de uma forma normal, como previsto pelo Teorema do Limite Central (BETHEA et al., 1985). Utilizando dados obtidos de produtividade, Shearer et al. (1997, 2005) demonstraram que os dados brutos do monitor de produtividade pareciam ser normalmente distribuídos. Estudos adicionais usando análise de histograma encontraram dados de rendimento como não normais, existindo certa assimetria, com distribuições tanto positivas (NOACK et al., 2003b) quanto negativas (BECK et al., 2001). Lyle et al. (2014) apresentaram resultados de 183 campos com mapas de produtividade da Austrália Ocidental e mostraram que apenas 11% dos arquivos tinham uma distribuição normal, enquanto a maioria apresentava altos valores de assimetria positiva ou negativa.

Dado que a filtragem local leva em consideração a mediana local para realizar a filtragem, nota-se a presença de uma quantidade maior de outliers, ou seja, uma amplitude maior na dispersão dos dados mostrados pelos boxplot, quando comparado ao que representa o método estatístico. Essa é a grande vantagem da filtragem local com relação à filtragem estatística, pois ela leva em consideração as diferenças locais do talhão, não permitindo remover pontos que podem ser relevantes localmente.

CONCLUSÕES

A caracterização dos erros encontrados durante a pré-filtragem dos dados é um aspecto crucial. Essa caracterização e quantificação fornecem uma base para filtrar os pontos considerados como errôneos ou discrepantes, permitindo a melhoria da qualidade da informação contida nos mapas de produtividade. O processo de filtragem estatística removeu entre 1,8 e 14,9%, o software Mapfilter 2.0 entre 6,8 e 38,9% de dados errôneos e discrepantes, valores dentro dos limites que a maioria dos estudos sugere. Apesar dos erros ocorrerem em menor frequência nos monitores de produtividade atuais, alguns destes erros ainda persistem, principalmente relacionados aos valores decorrentes de leituras de produtividade nula. A etapa de filtragem de dados segue sendo fundamental para a obtenção de mapas de produtividade confiáveis.

REFERÊNCIAS

- ARSLAN, S.; COLVIN, T. S. Mapeamento de rendimento de grãos: sensoriamento de rendimento, reconstrução de rendimento e erros. **Precision Agriculture**, 3, 135–154, 2002b.
- BECK, A. D.; SEARCY, S. W.; ROADES, J. P. Yield data filtering techniques for improved map accuracy. **Applied Engineering in Agriculture**, 17(4), 432. 2001.
- BETHEA, R. M.; DURAN, B. S.; BOULLION, T. L. **Statistical methods for engineers and scientists** (2nd ed.). New York: M. Dekker. 1985.

- BLACKMORE, B. S.; MOORE, M. Remedial correction of yield map data. **Precis. Agric.** 1:53–66. 1999.
- BLACKMORE, B.S. **The role of yield maps in precision farming.** National Soil Resources Institute. PhD.thesis. Cranfield University, Great Britain. 170 pp. 2003.
- FULTON, J.; HAWKINS, E.; TAYLOR, R.; FRANZEN, A. Yield Monitoring and Mapping. In: **Precision Agriculture Basics**, p. 63-77, 2018.
- GIMENEZ, L. M.; MOLIN, J. P. Algoritmo para redução de erros em mapas de produtividade para agricultura de precisão. *Revista Brasileira de Agrocomputação*, 2(1), 5- 10, 2004.
- HAMILTON, C. J. Lat Lon Tools. 2023. Disponível em: < <https://plugins.qgis.org/plugins/latlontools/> >. Acesso em: 20 jul. 2023.
- LARK, R. M.; BOLAM, H. C.; STAFFORD, J. V. Limites para a resolução espacial de sistemas de mapeamento de rendimento para culturas combináveis. **Journal of Agricultural Engineering Research**, 66, 183–193, 1997.
- LYLE, G.; BRYAN, B. A.; OSTENDORF, B. Métodos de pós-processamento para eliminar medições errôneas de rendimento de grãos: revisão e direções para futuros desenvolvimentos. **Precision Agriculture**, v. 15, n. 4, p. 377–402, 2014.
- MALDANER, L. F.; MOLIN, J. P. Data processing within rows for sugarcane yield mapping. **Scientia Agricola** 77. 2020.
- MALDANER, L. F.; MOLIN, J. P., SPEKKEN, M. Map Filter 2.0. 2021. Disponível em: < <https://www.agriculturadeprecisao.org.br/software/> >. Acesso em: 22 jul. 2023.
- MENEGATTI, L. A. A.; MOLIN, J. P. Metodologia para identificação e caracterização de erros em mapas de produtividade. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.7, n.2, p.367-374, 2003.
- NOACK, P. H.; MUHR, T.; DEMMEL, M. Relative accuracy of different yield mapping systems installed on a single combine harvester. In J.V.Stafford and A. Werner (eds.), *Proceedings of the 4th European conference on precision agriculture* (pp. 445–451). Wageningen: Wageningen Academic Publishers. 2003b.
- PING, J. L.; DOBERMANN, A. Processamento de dados de mapas de rendimento. **Precision Agriculture**, v. 6, p. 193–212, 2005.
- QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation Project. 2023. Disponível em: < <http://qgis.osgeo.org> >. Acesso em: 10 jul. 2023.
- ROBINSON, T. P.; METTERNICHT, G. Comparing the performance of techniques to improve the quality of yield maps. **Agricultural Systems**, 85, 19–41. 2005.
- SHEARER, S. A.; HIGGINS, S. G.; MCNEILL, S. G.; WATKINS, G. A.; BARNHISEL, R. I.; DOYLE, J. C.; LEACH, J. H.; FULTON, J. P. Post-Processing Correction to Improve the Accuracy of Yield Monitor Data in Grain Crops. In D. J. Mulla (ed.), *Proceedings of the seventh international conference on precision agriculture* (pp. 173–186). MN, USA: Precision Agriculture Center, University of Minnesota. 2005.
- SHEARER, S.A.; HIGGINS, S.G.; MCNEILL, S.G.; WATKINS, G.A.; BARNHISEL, R. I.; DOYLE, J. C.; LEACH, J. H.; FULTON, J. P. **Data filtering and correction techniques for generating yield maps from multiple combine systems.** Paper No. 971034. ASAE, St Joseph, MI, USA. 1997.
- SIMBAHAN, G. C.; DOBERMANN, A. PING, J. L. Screening yield monitor data improves grain yield maps. **Agronomy Journal** 96, 1091–1102. 2004.
- SMITI, A. A critical overview of outlier detection methods. **Computer Science Review**, Volume 38, 2020.

SPEKKEN, M.; ANSEMI, A. A.; MOLIN, J. P. A simple method for filtering spatial data. p. 259-266. In: Stafford, J.V., ed. Precision agriculture. Wageningen Academic Publishers, Wageningen, The Netherlands. 2013.

STAFFORD, J. V.; AMBLER, B.; LARK, R. M.; CATT, J. Mapping and interpreting the yield variation in cereal crops. **Computers and Electronics in Agriculture**, Volume 14, Issues 2–3, pp. 101-119, 1996.

SUN, W.; WHELAN, B.; MCBRATNEY, A.B. An integrated framework for software to provide yield data cleaning and estimation of an opportunity index for site-specific crop management. **Precision Agric** 14, 376–391. 2013.

THYLÉN, L.; ALGERBO, P. A.; GIEBEL, A. An expert filter removing erroneous yield data. In P.C. Robert et al. (ed.) Proc. 5th Int. Conf. on Precision Agriculture, Bloomington, MN [CD-ROM]. 16–19 July 2000. ASA, CSSA, and SSSA, Madison, WI. 2000.

TUKEY, J.W. **Exploratory data analysis**. 1 ed. v. 1. Reading, Massachusetts, 1977.

ZANELLA, M. A.; MARTINS, R. N.; GOMES, E. Q.; LIMA, A. N.; SANTOS, J. A. de S.; GONÇALVES, L. R. Remoção de erros em mapas de produtividade de milho via filtragem de dados brutos. In: SIMPÓSIO MATO-GROSSENSE DE MECANIZAÇÃO AGRÍCOLA E AGRICULTURA DE PRECISÃO – SIMAP, 1., 2018, Sinop. **Anais [...]**. Sinop: UFMT, Campus Universitário de Sinop, 2018.