



UNIVERSIDADE
ESTADUAL DE LONDRINA

LARA MARIE GUANAIS SANTOS

**TRATAMENTO E INTERPRETAÇÃO DE DADOS DE
PRODUTIVIDADE PARA AGRICULTURA DE PRECISÃO**

Londrina
2020

LARA MARIE GUANAIS SANTOS

**TRATAMENTO E INTERPRETAÇÃO DE DADOS DE
PRODUTIVIDADE PARA AGRICULTURA DE PRECISÃO**

Dissertação apresentada ao Departamento de Agronomia da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Mestre em Agronomia.

Orientador: Prof. Dr. Otávio Jorge Grigoli Abi Saab.

Coorientador: Dr. Hevandro Colanhese Delalibera – IDR Paraná

Londrina
2020

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

S237t Guanais Santos, Lara Marie .
Tratamento e interpretação de dados de produtividade para agricultura de precisão / Lara Marie Guanais Santos. - Londrina, 2020.
40 f. : il.

Orientador: Otávio Jorge Grigoli Abi Saab.
Coorientador: Hevandro Colonhese Delalibera.
Dissertação (Mestrado em Agronomia) - Universidade Estadual de Londrina, Centro de Ciências Agrárias, Programa de Pós-Graduação em Agronomia, 2020.
Inclui bibliografia.

1. Colheita de grãos - Tese. 2. Erros sistemáticos - Tese. 3. Álgebra de mapas - Tese. I. Abi Saab, Otávio Jorge Grigoli . II. Delalibera, Hevandro Colonhese. III. Universidade Estadual de Londrina. Centro de Ciências Agrárias. Programa de Pós-Graduação em Agronomia. IV. Título.

CDU 63

LARA MARIE GUANAIS SANTOS

**TRATAMENTO E INTERPRETAÇÃO DE DADOS DE
PRODUTIVIDADE PARA AGRICULTURA DE PRECISÃO**

Dissertação apresentada ao Departamento de Agronomia da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Mestre em Agronomia.

BANCA EXAMINADORA

Orientador: Prof. Dr. Otávio Jorge Grigoli Abi Saab
Universidade Estadual de Londrina - UEL

Dr. Gustavo Di Chiacchio Faulin
Faculdade de Tecnologia "Shunji Nishimura" de
Pompeia – FATEC Pompeia

Dr. Ricardo Ralisch
Universidade Estadual de Londrina – UEL

Londrina, 17 de Fevereiro de 2020.

Dedico este trabalho à minha família. Em especial, minha mãe Ivone Guanais, meu pai José Sebastião e meu noivo Caio Fernando.

AGRADECIMENTOS

À Deus, pilar em minha vida. Sem Ele nada seria possível.

Aos meus pais Ivone Guanais Mineiro Santos e José Sebastião dos Santos, que com muito esforço e dedicação venceram os momentos difíceis e sempre me proporcionaram o melhor.

Ao meu noivo Caio Fernando dos Santos Oliveira, por todo apoio e companheirismo.

Aos meus orientadores e mentores Otávio Jorge Grigoli Abi Saab, Hevandro Colonhese Delalibera e Ricardo Ralisch pela generosidade em contribuir para meu desenvolvimento intelectual, pessoal e de minha pesquisa.

À todos os docentes da Universidade Estadual de Londrina pelo conhecimento cedido a mim, em especial as professoras Maria de Fátima e Inês Fonceca pelo carinho e apoio.

Ao programa de pós graduação em Agronomia pela oportunidade de desenvolver minha pesquisa nesta universidade maravilhosa.

À CAPES, pela concessão da bolsa.

Ao IAPAR Londrina – PR, pela concessão do método analítico de análise de dados e álgebra de mapas.

Ao senhor Ademir Inocente pela concessão dos dados.

À empresa de Telemetria Agrícola Agritel, em especial ao Daniel Andrade e Fábio Okahara pelo repasse dos dados e toda ajuda e prontidão perante minhas dúvidas.

À Secretária Helenice, pela amizade, paciência e todo auxílio.

Aos técnicos do Laboratório de Solos Marcio Praxedes e João Machado, pela amizade e auxílio na realizações das análises.

Aos amigos que ingressaram no mestrado comigo Bruno Teixeira, Walter Junior, Janaina Matsumoto e Gabriel Shimizu pela amizade e apoio em momentos difíceis e de alegrias.

Ao Gabriel Shimizu, meu agradecimento por toda ajuda em estatística.

Aos amigos da pós Smaylla El Kadri, Jaqueline Felix, Victor Freitas, Natassia Armacolo, Alex Figueiredo e Alejandro Carrijo pelo apoio e auxílio sempre que precisei.

As estagiárias do laboratório de solo Gabriela Mores e Tamires Firmino

por todo auxílio.

À pós-doutoranda Adriana da Silva, por sempre me auxiliar em minhas análises, com a escrita e conselhos.

Aos meus amigos de longa data, Júlia Vicheti, Juliana Hara, Cinthia Nascimento, Tainá Santos, Bruna Garcia, Vinicius Favoni, Danilo Tedesco e Danilo Deo por todas as conversas e apoio.

Agradeço também a Júlia Vicheti formada em Designer gráfico, excelente profissional e minha amiga, por abrilhantar minha dissertação com a confecção da Figura 1.

Serei eternamente grata!

***A vida me ensinou a nunca desistir,
nem ganhar, nem perder mas procurar
evoluir.***

- Charlie Brown Jr.

SANTOS, Lara Marie Guanais. **Tratamento e interpretação de dados de produtividade para agricultura de precisão**. 2020. 44 f. Dissertação (Mestrado em Agronomia) - Universidade Estadual de Londrina, Londrina, 2020.

RESUMO

O objetivo do trabalho foi avaliar os dados obtidos em colhedoras instrumentadas e estabelecer um método de tratamento de dados por meio de procedimentos com base estatística como o teorema de Laplace e o teorema do limite central, visando eliminar dados discrepantes ou não pertencentes à população do conjunto estudado, sem proporcionar alterações na distribuição e no comportamento do fenômeno produtividade. Foi selecionada uma área agrícola na qual foram cultivados milho e soja em épocas distintas e colhidas por uma máquina dotada de monitor de colheita. Os dados obtidos foram convertidos em desvios padronizados e “limpos” utilizando o método estatístico *Jackknife*. Após este procedimento, foi realizada a álgebra de mapas das culturas agrícolas distintas, por meio dos mapas de índices de desvio obtidos na primeira fase de análise, a qual além de remover dados não representativos também proporciona um escalamento das variáveis (grandezas de massa de grãos distintas entre as culturas), obtendo uma saída de dados unificada e simplificada para tomada de decisão e a realização de amostragem de solos e sugestão de possíveis zonas de manejo. Concluiu-se que a aplicação do método testado para a filtragem de dados foi considerada simples e se apresenta como uma boa solução para adequar dados de colheita, tornando-os mais representativos e, também, proporcionar um escalamento dos dados em índices padrão compreendidos entre $\pm 3,09$. Foi possível realizar a álgebra de mapas com diferentes culturas de forma simples, obtendo-se um mapa final resumido com classes de índice de potencial produtivo da área.

Palavras-chave: Colheita de grãos. Erros sistemáticos. Álgebra de mapas.

SANTOS, Lara Marie Guanais. **Processing and interpretation of productivity data for precision agriculture**. 2020. 44 p. Dissertação (Mestrado em Agronomia) - Universidade Estadual de Londrina, Londrina, 2020.

ABSTRACT

The objective of the work was to evaluate the data obtained in instrumented harvesters and to establish a method of data treatment through statistical based procedures such as Laplace's theorem and the central limit theorem, aiming to eliminate discrepant data or not belonging to the population of the set studied, without providing changes in the distribution and behavior of the productivity phenomenon. An agricultural area was selected in which corn and soybeans were grown at different times and harvested by a machine equipped with a yield monitor. The data obtained were converted into standard deviations and "clean" using the Jackknife statistical method. After this procedure, the algebra of maps of different agricultural crops was carried out, using the maps of deviation indexes obtained in the first analysis phase, which in addition to removing non-representative data also provides a scaling of the variables grain mass quantities between cultures, obtaining a unified and simplified data output for decision making and the carrying out of soil sampling and suggestion of possible management zones. It was concluded that the application of the tested method for data filtering was considered simple and presents itself as a good solution for adapting harvest data, making them more representative and, also, for providing a scaling of the data in standard indexes understood between ± 3.09 , it was possible to perform the map algebra with different cultures in a simple way, obtaining a final map summarized with classes of index of productive potential of the area.

Key words: Corn. Soy beans. Data cleansing. Map algebra.

LISTA DE ILUSTRAÇÕES

Figura 1	Sensores necessários para obtenção de um mapa de produtividade.....	15
Figura 2	Fluxograma dos processos para identificação de zonas de manejo	17
Figura 3	Distribuição “z” com índices bilaterais	20
Figura 4	Histogramas dos dados de produtividade, diagrama de frequência ordenada de desvios padronizados e grau de ajuste da reta para massa de grão colhidos de milho e soja antes e depois da aplicação do método jackknife	27
Figura 5	Mapas de desvios padronizados para o rendimento das culturas, sem e com a aplicação do método jackknife e classificados em três zonas por mais ou menos um desvio padrão da média	30
Figura 6	Mapa de tendências de produtividade da área, resultante da álgebra entre os mapas dos índices de produtividade de milho e soja.....	31

LISTA DE TABELAS

Tabela 1	Área em hectare dos potenciais produtivos representados por cores e pelo Coeficiente de Variação (CV) dos mapas31
Tabela 2	Análise descritiva dos dados com e sem tratamento de dados Jackknife33

SUMÁRIO

1	INTRODUÇÃO	12
2	REVISÃO BIBLIOGRÁFICA	14
2.1	AGRICULTURA DE PRECISÃO	14
2.2	MAPAS DE PRODUTIVIDADE	15
2.3	INCONSISTÊNCIA DOS DADOS.....	17
2.4	MÉTODOS PARA TRATAMENTO DOS DADOS DE COLHEITA DE GRÃOS.....	18
2.5	MÉTODO JACKKNIFE	19
3	MATERIAL E MÉTODOS	23
4	RESULTADOS	26
5	CONSIDERAÇÕES FINAIS	35
6	CONCLUSÕES	36
	REFERÊNCIAS	37
	APÊNDICE	41
	APÊNDICE A Script para remoção de dados repetidos e para retirada de dados nulos e zerados	42

1 INTRODUÇÃO

Com a intensificação da produção agrícola e o elevado custo de produção, houve incorporação de tecnologias para otimizar os recursos e viabilizar a produção em larga escala com a finalidade de reduzir custos de mão de obra e auxiliar na gestão e tomada de decisão.

A mecanização tem sido a solução para maior capacidade operacional, seguida pela adoção da agricultura de precisão, que busca otimizar os recursos já aplicados com base na variabilidade temporal e espacial da área.

O mapa de produtividade gerado pelas colhedoras instrumentadas é considerado o primeiro passo para observação das diferentes zonas de manejo.

Devido ao grande número de dados coletados em um curto intervalo de tempo, pode haver inconsistências como: variação operacional da máquina (manobras, arremates e falha operacional) e falhas nos sensores de umidade, velocidade e produtividade.

A análise da sequência dos dados de produtividade obtidos pela máquina, associados a algumas características operacionais, podem ajustar estas informações, dando maior confiabilidade nos mapas.

Para o analista de dados, há uma grande dificuldade em saber quais pontos realmente devem ser removidos de um conjunto de dados e, até mesmo, um método eficiente para avaliar a variabilidade espacial de atributos que sejam relacionados ao ambiente de produção.

Devido a esta problemática, o presente trabalho teve por objetivo estudar um método de análise de dados, que permite remover, com embasamento estatístico, valores não representativos do volume total de dados de colheita, sem que proporcione alteração da distribuição do fenômeno avaliado, buscando torná-los mais representativos, gerando mapas mais fiéis.

E, através dos mapas resultantes dos filtros estatísticos, avaliar a possibilidade de se estudar conceitos baseados em álgebra de mapas com culturas distintas, visando gerar um mapa resumo que represente o potencial produtivo histórico de uma área, visando simplificar a informação de saída e permitir uma análise conjunta entre as diferentes culturas no tempo, diluindo

efeitos de fatores dos quais não se pode controlar, como clima, genética e espécies distintas, auxiliando na tomada de decisão sobre a gestão da atividade, bem como na determinação das zonas de manejo do solo.

2 REVISÃO BIBLIOGRÁFICA

2.1 AGRICULTURA DE PRECISÃO

A agricultura de precisão (AP) é um conjunto de princípios e tecnologias aplicados no manejo da variabilidade espacial e temporal associado à produção agrícola (PIERCE; NOWAK, 1999), e tem como objetivo o uso mais racional do solo e de insumos, proporcionando qualidade ambiental e melhores retornos econômicos. Para Lamparelli (2016), a AP caracteriza-se por um conjunto de técnicas que permite o gerenciamento localizado dos cultivos, prevê a otimização dos gastos da produção, utilizando técnicas que visam o melhor rendimento da cultura, maximizando diversos fatores como a localização, fertilidade do solo, entre outros, buscando através de ferramentas tratar especificamente cada local da propriedade agrícola, na verificação de suas particularidades.

A AP vai além da aquisição de maquinário e/ou ferramentas, compondo um sistema de manejo integrado de informações e tecnologias, fundamentado na premissa de que as variabilidades de espaço e tempo influenciam nos rendimentos dos cultivos. As tecnologias de AP têm como objetivo o gerenciamento detalhado do sistema de produção agropecuário como um todo, não somente das aplicações de insumos ou de mapeamentos diversos, mas de todo os processos envolvidos na produção (BERNARDI *et al.*, 2011).

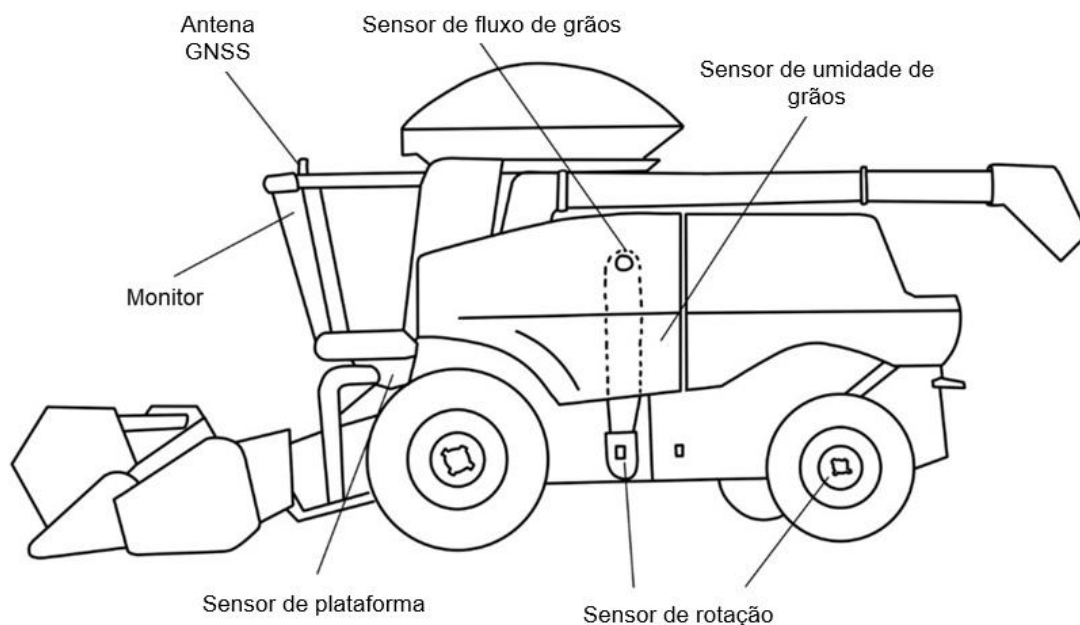
No final dos anos 80 surgiram as primeiras tentativas de medir o fluxo de grãos em colhedoras. Com isso, começaram a ser disponibilizados produtos no mercado mundial que se habilitam a gerar dados para a obtenção de mapas de produtividade (MOLIN, 2002).

Bernard *et al.* (2014), afirmam que o objetivo principal na adoção da AP é justamente facilitar a compreensão das causas de variabilidade, pois permite avaliar potencialidades do manejo de cada área específica. A escolha das ferramentas tecnológicas é uma escala gradual e customizada, tendo em vista a necessidade de cada propriedade e, tem por intuito aprimorar, em médio prazo, o manejo da variabilidade da lavoura com maior qualidade agrônômica.

2.2 MAPAS DE PRODUTIVIDADE

O emprego de monitores de desempenho em máquinas agrícolas, possibilita a extração de dados que são de grande utilidade para o gerenciamento de operações agrícolas específicas, como a colheita (COMPAGNON *et al.*, 2012). Com a inclusão de eletrônica para coleta e armazenamento de dados, sistemas de posicionamento global por satélite e transmissão de dados e sistemas para o tratamento de dados (Figura 1), torna-se possível o gerenciamento da operação e da variabilidade espacial (QUEIROZ *et al.*, 2000).

Figura 1 – Sensores necessários para obtenção de um mapa de produtividade.



Fonte: Próprio autor

Os mapas de produtividade consistem em um conjunto de pontos, onde cada um representa a medida da quantidade e da umidade dos grãos colhidos em uma área correspondente à largura da plataforma de colheita da colhedora pela distância no tempo percorrido entre duas leituras, referenciada pela coordenada geográfica (MOLIN, 2000). Esta informação é de grande valia, ajudando na tomada de decisões de gerenciamento e manejo da área (BRASE, 2006).

Os mapas de produtividade são frequentemente utilizados como ponto de partida por usuários e pesquisadores para visualizar a variabilidade espacial das lavouras, por materializar a resposta da cultura às condições da área, clima e ao manejo aplicado (MOLIN,2000).

Grisso *et al.* (1999) concluíram que, dentre as ferramentas como: fotografias aéreas, imagens de satélite, amostragem de solo em grade e condutividade elétrica do solo, o mapa de produtividade é a ferramenta que proporciona a informação mais completa, pois, possibilita a visualização da variabilidade espacial nas lavouras sobre a variável dependente resposta de interesse. As denominadas zonas de manejo têm por princípio a divisão da área total em áreas específicas, almejando o melhor desempenho destas (SPERANZA *et al.* 2016).

Os sistemas de monitoramento da produtividade durante a colheita, permitem configurar a frequência de aquisição dos dados, podendo ser de 250 a 1500 pontos por hectare, variando o tamanho em função da largura da plataforma. Os dados obtidos são interpolados através de *softwares* do tipo Sistema de Informação Geográfica (SIG) para assim serem convertidos em informação (MOLIN, 2000). Porém, nesta grande quantidade de dados, também estão presentes erros e dados inconsistentes (MENEGATTI *et al.*, 2004).

Lark *et al.* (1999) consideram que a análise de uma sequência de mapa de produtividade é uma forma de definir unidades de manejo de um talhão, associando às características do solo.

A análise da variabilidade regionalizada do solo é explorada para definição de unidades de manejo (KHAKURAL *et al.*, 1998), pois, pode influenciar o crescimento vegetal pelo fornecimento distinto de água, nutrientes e oxigênio (CAMPOS BERNARDI *et al.*, 2017). As propriedades do solo podem variar espacialmente dentro de um mesmo talhão até a escala regional em função de fatores intrínsecos, como formação do solo, e extrínsecos, como práticas de manejo, adubação e rotação de culturas (CAMBARDELLA; KARLEN, 1999).

O conhecimento da causa da variabilidade da produção é que subsidiará a decisão sobre o manejo do solo e da cultura (MILANI *et al.*, 2006; MOLIN, 2002).

Figura 2 – Fluxograma dos processos para identificação de zonas de manejo.



Fonte: Próprio autor.

2.3 INCONSISTÊNCIA DOS DADOS

Os dados advindos dos monitores de colheita apresentam inconsistências e erros. Portanto, é sempre necessária a realização de um tratamento preliminar dos dados antes de transformá-los em um mapa que vá servir para análise e tomada de decisão (MOLIN, 2002).

Com as limitações dos próprios equipamentos, os fabricantes dos monitores de produtividade, em geral, atribuem uma acurácia na ordem de 97 a 98% na totalização dos grãos colhidos. Grisso *et al.* (1999), acompanharam colhedoras equipadas com monitor de produtividade trabalhando em diferentes taxas de alimentação e observaram uma faixa de erro de aproximadamente 2%. Da mesma forma, Molin *et al.* (2000), encontraram diferença média de massa entre a leitura pelo monitor e a totalização de balança em cada tanque graneleiro da colhedora, da ordem de 0,5 e 1,06%, para dois anos seguidos em um mesmo talhão e com o mesmo equipamento.

Moore (1998), ao analisar dados de colheita de seis anos consecutivos, identificou os seguintes erros: suavização, calibração de volume, largura incorreta da plataforma de corte e tempo de enchimento da colhedora. Blackmore e Marshall (1996), também acrescentaram erros de posicionamento, erros devido à recirculação de grãos no mecanismo de trilha e retrilha da colhedora, perda de grãos pela colhedora e a acurácia do sensor de produtividade.

A umidade dos grãos é uma fonte de erros, pois, esta varia dentro do talhão em pequenos intervalos espaço, em função do estágio de maturação das plantas ao longo da lavoura (MENEGATTI; MOLIN, 2002).

De acordo com Moore (1998), é essencial para o operador da colhedora adequar a largura de corte informada ao monitor com a largura efetivamente utilizada. Qualquer desvio ocorrido entre a largura efetiva e a

largura informada ao sistema resulta em erros na medição de área e, conseqüentemente, em erros na medição da produtividade da cultura.

Outros fatores que devem ser observados, são os dados de velocidade zerados ou nulos. Também devem ser excluídos dados onde as velocidades são muito baixas, pois, afetam o fluxo de colheita.

2.4 MÉTODOS PARA TRATAMENTO DOS DADOS DE COLHEITA DE GRÃOS

Apesar dos cuidados para gerar bons mapas de produtividade, ainda existem desafios associados ao seu desenvolvimento. Por exemplo, como avaliar a qualidade de um mapa de produtividade (SOUZA *et al.*, 2016) ou qual o melhor método de tratamento de dados.

Spekken *et al.* (2013), desenvolveram um método genérico capaz de identificar e filtrar pontos de dados incorretos que são inconsistentes com seus pontos vizinhos. Este método propõe detectar todos os pontos localizados em uma faixa de raio em torno de um ponto; extraíndo o Coeficiente de Variação (CV) entre esses pontos. Se o valor CV encontrado para esses pontos for maior que o CV máximo, um peso será adicionado a todos eles; um próximo ponto após o ponto no conjunto de dados é selecionado e o processo se repete. Ao final, os autores concluíram que o ruído no mapa interpolado foi reduzido e houve melhora na suavização. Portanto, este método facilita a definição de zonas de gerenciamento, embora sejam necessários mais estudos para isso.

Leroux *et al.* (2018), propõem uma metodologia holística para filtrar observações defeituosas que possam estar presentes em conjuntos de dados de rendimento. A noção de vizinhança espacial foi aprimorada para incluir as características específicas de tais conjuntos de dados baseados em veículos em movimento. As observações são comparadas com sua vizinhança espacial recém-definida e as mais anormais são classificadas como observações defeituosas, com base em um algoritmo de agrupamento baseado em densidade. Apesar das melhorias significativas na distribuição e na estrutura espacial dos conjuntos de dados de rendimento, a avaliação do algoritmo ainda era subjetiva.

Outro método, proposto por Menegatti e Molin (2004) leva em consideração as características de cada erro presente em um mapa de produtividade, como posicionamento, produtividades muito altas ou baixas, largura de plataforma incorreta, umidade de grãos, distância entre pontos, intervalo de enchimento e limites estatísticos para dados discrepantes. A análise geoestatística dos conjuntos de dados mostrou que a aplicação do filtro diminuiu o efeito pepita e obteve melhores resultados qualitativos na validação cruzada por *krigagem*, elevando a qualidade final da informação.

2.5 MÉTODO *JACKKNIFE*

O método *Jackknife* (WU *et al.*, 1986) composto por duas análises de desvios, Desvios Padronizados (baseado no teorema de Laplace) e Desvios Studentizados (ou teorema do limite central), tem por intuito a remoção de amostras consideradas não pertencentes a população.

Este método é recomendado para casos onde não há ou não se considera tratamentos distintos (como tratamentos pode-se considerar anos agrícolas diferentes lembrando que são necessárias análises separadas para cada tipo de cultura, isto é, não é possível colocar dados de dentro de uma análise de resíduos de soja, devendo seus desvios ser em calculados em análises distintas).

A análise dos Desvios Padronizados (Equação 1) pode ser utilizada como um filtro de base estatística para exclusão de amostras consideradas não pertencentes à população analisada, utilizando como parâmetros os valores de distribuição da tabela “z” (HARTER, 1960).

Eq. 1

$$Z = \frac{x_i - \bar{x}_i}{\sigma}$$

Legenda:

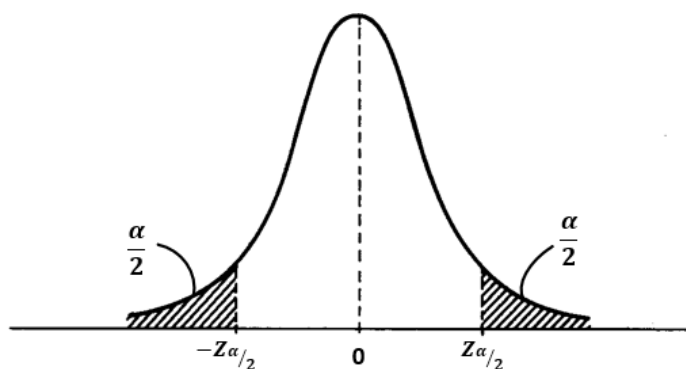
x_i – Amostra

\bar{x}_i – Média amostral ou do tratamento

σ – Desvio padrão

Esta análise dos Desvios Padronizados se refere ao valor crítico de -3,09 a 3,09 (bilateral), conforme a Figura 3, nos quais, dados que apresentam índice de Desvio Padronizado fora deste intervalo, são considerados como não pertencentes à população e devem ser excluídos.

Figura 3 - Distribuição “z” com índices bilaterais.



Fonte: Adaptada de Harter (1960).

A exclusão de dados dentro do intervalo citado deve ser criteriosa e seguir os parâmetros da estatística “z”, recomendada para σ conhecido, uma vez que podem alterar a distribuição dos dados e, assim, o comportamento da população estudada, ou seja, o conjunto de dados pode não mais representar a realidade do fenômeno (SNEDECOR; COCHRAN, 1980).

É importante que estes métodos de análise e remoção de dados sejam realizados em planilha eletrônica aplicando cálculos referenciados para que, após cada remoção de valor, os índices sejam recalculados, pois, com a remoção de dados as medidas de tendência central se alteram, bem como seus índices finais (DELALIBERA *et al.*, 2017).

Também é de suma importância iniciar sempre a remoção pelos índices mais extremos, pois, a cada remoção de valores discrepantes, os índices podem ser movidos para o centro ou para a extremidade da distribuição, podendo para os casos onde os valores se apresentem no limiar de corte, serem movidos para dentro para fora do intervalo crítico de corte escolhido (DELALIBERA *et al.*, 2017).

Ao final da aplicação do método *Jackknife* (WU *et al.*, 1986), é possível observar que os índices gerados são um escalamento de medidas dos

dados entre $\pm 3,09$ ou $\pm 1,96$, para desvio padronizado e studentizado respectivamente.

O termo Escalamento de dados, é oriundo da palavra *scale*, que significa a transformação de variáveis com unidades/grandezas distintas em algo comum, sendo utilizado em estatística multivariada. No caso dos dados de produtividade, a unidade comum é desvio.

Quando se tem fatores que podem ser considerados distintos, como tratamentos, recomenda-se utilizar o método dos Desvios Studentizados para a estimativa dos índices e a identificação de dados discrepantes. Este também é conhecido como “Teorema do limite central” (Equação 2), que utiliza a distribuição “t” de Student para inferência sobre a amostra com desvio padrão (σ) desconhecido, onde índices fora do intervalo $\pm 1,96$ (5% de significância para teste bilateral de valores críticos de t) são considerados discrepantes e podem ser eliminados do conjunto de dados, sem prejuízos à distribuição ou ao comportamento do fenômeno.

Eq.2

$$t = \frac{x_i - \bar{x}_i}{\sigma / \sqrt{n}}$$

Legenda:

x_i – Amostra

\bar{x}_i – Média amostral ou do tratamento

σ / \sqrt{n} – Erro padrão (tratamento)

Para este método, quando os valores apresentam índice dentro do intervalo $\pm 3,09$ e $\pm 1,96$, sua remoção fica a critério do analista, de acordo com o que se deseja com o ajuste da linearidade dos desvios dos dados e formatação da distribuição de frequências.

É válido ressaltar que, para a análise de desvios Studentizados ser considerada robusta, é necessário o levantamento de, no mínimo, trinta (30) valores para cada tratamento. Caso o número de amostras por tratamento seja menor, torna-se necessário verificar na tabela de valores críticos de “t” o intervalo

crítico adequado ao número de amostras (SNEDECOR; COCHRAN, 1980). Este método preserva o comportamento e a distribuição do fenômeno, propiciando a remoção de valores não pertencentes à população e, também, dados discrepantes, possibilitando a aplicação de métodos de análise paramétrica nos dados de saída.

3 MATERIAL E MÉTODOS

Para a execução deste trabalho foram utilizados dados referentes as culturas de soja (colhida em 07/02/2019) e milho (colhido em 28/07/2018) respectivamente de uma área agrícola em sistema plantio direto, localizada no município de Cambé, PR (23° 8'29,99" S, 51°20'48,51" O e 521 m de altitude). A área é constituída por Latossolo Vermelho (SANTOS *et al.*, 2018) e está sob clima Cfa - clima temperado úmido com verão quente (KÖPPEN, 1931).

Os dados foram obtidos por meio de uma colhedora da marca John Deere modelo STS 9750 equipada com sensor de produtividade tipo placa de impacto, sensor de umidade do tipo capacitância, sensor de velocidade magnético instalado nas rodas motrizes, sensor de inclinação tipo giroscópio e monitor modelo GS4 que apresenta e grava informações como: posicionamento global (coordenadas geográficas e convertidas posteriormente para Sistema Universal Transverso de Mercator (UTM)), produtividade (kg ha^{-1}) umidade (%), tempo percorrido para gerar um polígono (segundos) e velocidade média de deslocamento linear da colhedora (km h^{-1}) os quais foram exportados para tabela do *software* Excel (2010).

Os dados obtidos foram extraídos da colhedora e pré-processados por um sistema de telemetria e exportados para formato “xlsx”.

Por meio do *software* RStudio, foi desenvolvido um *script* (Apêndice) para limpeza de dados zerados ou nulos dos mapas de produtividade com o intuito de otimizar este processo que antes era feito manualmente. Os dados oriundos da colhedora devem estar em formato “xlsx” (Excel, 2010) contendo dados de produtividade, umidade do grão, coordenadas geográficas e velocidade de colheita.

Foram removidos valores zerados ou nulos em qualquer das variáveis, eliminando sempre a amostra inteira, ou seja, toda a linha da planilha, para que estas não perdessem sua referência de dados. Por se tratar de dados de colheita com uma grande densidade de dados estas retiradas não comprometem o resultado da análise e a interpolação dos dados. Após este

tratamento prévio, foi feita a correção da produtividade para 13% de umidade pelo método sugerido por Aguiar (1977), conforme a Equação 3.

Eq. 3

$$\text{Peso Convertido} = \frac{100 - \text{Umidade mensurada}}{100 - 13} * \text{Peso Inicial}$$

Ao final do script foi inserido uma função para gerar um arquivo no formato “csv”, formato utilizado pelo *software* de SIG para a geração dos mapas e um arquivo em Formato “xlsx” para remoção dos pontos discrepantes.

Os gráficos da frequência ordenada dos índices de desvio foram desenvolvidos no *software* RStudio com o intuito de auxiliar a visualização do comportamento da distribuição do fenômeno e facilitar a remoção dos dados discrepantes.

Com base neste arquivo “xlsx” gerado ao final do *script*, no Excel foi aplicado a análise de desvios pelo método dos Desvios Padronizados, visando estimar o índice de desvio pontual de cada leitura de rendimento, para posteriormente compará-la com a distribuição padrão “z”. Com o uso deste método os dados de produtividade (kg ha^{-1}) são convertidos para índice de Desvios (adimensional). Todos os valores abaixo de -3,09 ou acima de 3,09 foram considerados não pertencentes a população conforme a distribuição “z” e, portanto, foram removidos conforme sugere o método *Jackknife* (WU *et al.*, 1986). Ao remover os dados é necessário recalcular os índices (DELALIBERA *et al.*, 2017).

Após a remoção de pontos discrepantes foram gerados mapas com base nos índices de desvio por meio do *software* SIG QGIS 3.4.6. Os dados foram interpolados pelo método Redes de Triangulação Irregular (TIN), utilizou-se o modo Desvio Padrão com 1 desvio para mais e para menos para realizar a classificação do mapa em três zonas.

O método TIN foi utilizado pois não houve o interesse em estimar dados nos intervalos entre os pontos. Em dados de produtividade há uma grande densidade e uma proximidade dos pontos e o desejado foi obter uma informação fiel aos dados. Este método utiliza um menor número de pontos vizinhos para o

cálculo da interpolação mantendo o nível de acurácia de uma grade regular mais densa (MIKHAIL et al., 2001).

As áreas com índices menores que -1 desvio são consideradas de baixo índice de potencial produtivo, entre -0,999 a 0,999, são consideradas de potencial médio normal local e, acima de 1 desvio, são consideradas de alto índice de potencial produtivo.

Em uma distribuição normal perfeita, 68,26% das ocorrências se concentraram na área do gráfico demarcada por um desvio padrão à direita e um desvio padrão à esquerda da linha média. Não necessariamente essa porcentagem de representação ocorre na distribuição de um fenômeno medido, porém, para este caso, foi considerado que os dados compreendidos em cada intervalo definido (<-1; -1 a 1 e >1 desvio) correspondem a níveis de potencial produtivo distintos.

A partir da obtenção do mapa de índices de colheita corrigido (filtrado), foi realizada a álgebra entre os mapas de soja e milho, visando a obtenção de um único mapa que consiste na média dos índices das culturas. O intuito da realização dessa álgebra de mapas de índices de colheita foi proporcionar a diluição de efeitos de fatores não controlados e obter uma informação resumida para facilitar a interpretação dos resultados de forma conjunta e a tomada de decisão ser mais assertiva.

Após a confecção dos mapas foi extraído do *software* QGIS os Histogramas dos dados de produtividade informação gerada automaticamente nas propriedades da camada.

Para quantificar as áreas correspondentes aos potenciais produtivos, foi efetuado a conversão de raster para vetor com a ferramenta raster pixels to polygons, em seguida selecionada cada classe de interesse para quantificá-las em hectares por meio da ferramenta resumo estatístico.

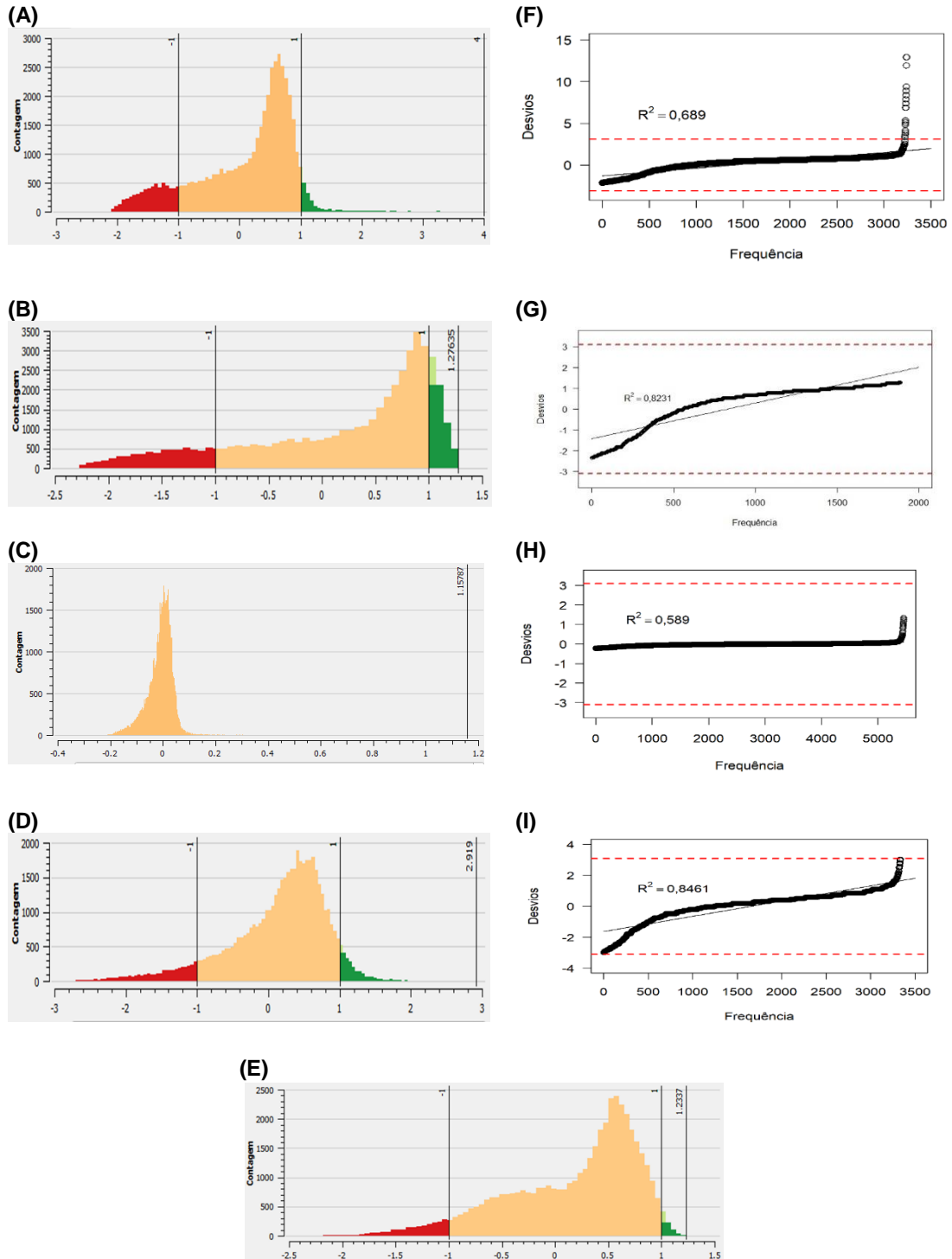
Para o cálculo do Coeficiente de Variação (CV), como foi trabalhado com os índices de desvio, que, em sua maioria são compreendidos por valores menores que 1, somou-se a constante 1 aos valores dos índices de desvio, visando proporcionar com que os valores do coeficiente de variação se tornem representativos, conforme recomendado por Hair *et al.* (2009), visto que esse procedimento não altera a distribuição e as medidas de tendência central dos dados.

4 RESULTADOS

Com a aplicação do método *jackknife* para remoção de dados discrepantes, foi possível observar que houve uma melhora no grau de linearidade dos índices de desvio, demonstrado pelo índice R^2 das Figuras de 4F a 4I, as quais também foram utilizadas como base para remoção de dados discrepantes extremos. Na Figura 4 também é possível observar que com a aplicação do método *Jackknife* junto ao ajuste de linearidade dos diagramas de frequência ordenada, houve uma melhora no comportamento dos dados observados nos histogramas de distribuição de frequência. Ainda se ressalva que a melhora no ajuste da linearidade do diagrama de frequência ordenada favorece o ajuste da normalidade e a representatividade das medidas de tendência central do fenômeno medido, que se trata da produtividade obtida na colheita mecanizada direta de grãos.

Este processo de verificação de dados é similar aos procedimentos matemáticos que alguns testes de normalidade utilizam na verificação da normalidade de conjuntos de dados, como Lilliefors e Anderson-Darling, porém estes métodos não podem ser aplicados visto que o tamanho do conjunto de dados não é adequado as regras dos testes citados.

Figura 4 – Histogramas dos índices de produtividade e diagrama de frequência ordenada dos índices de desvio padronizado e seus respectivos graus de ajuste da reta, referente ao fenômeno colheita mecanizada de grãos antes e após da aplicação do método *jackknife*.



De A à E - Corresponde aos histogramas dos mapas após a interpolação dos dados; De F à I - Corresponde aos diagramas de frequência ordenada de desvios padronizados das culturas; (A, F) – Dados de milho sem tratamento; (B,G)– Dados de milho com tratamento *Jackknife*; (C,H) – Dados de soja sem tratamento (D,I) – Dados de soja com tratamento *Jackknife*; (E) – Dados de soja e milho com tratamento *Jackknife* interpolados;

As Figuras 4F e 4H são referentes aos diagramas de frequência ordenada dos índices de desvio e demonstram a existência de pontos extremos discrepantes responsáveis pelas mudanças de direção nas extremidades da curva da frequência ordenada de índices. Mudanças estas que resultam em um ajuste ruim da linearidade dos dados.

Quando estes valores de desvios extrapolam os valores críticos da distribuição “Z” ($\pm 3,09$), podem ser excluídos, pois não pertencem a população avaliada (SNEDECOR; COCHRAN, 1980). Ainda, como apresenta a figura 4H, quando não há valores fora do intervalo crítico e o ajuste da linearidade é ruim, também se sugere a exclusão dos dados extremos, visando aumentar o grau de linearidade e melhorar a representatividade das medidas de tendência central.

Após a aplicação do método *Jackknife*, Figuras 4 G e I, que correspondem a dados de milho e soja respectivamente, conforme comentado anteriormente, foi observada uma melhor linearidade dos dados.

Os pontos fora do intervalo $\pm 3,09$ interferem diretamente na qualidade dos mapas. Estes pontos não pertencentes à população podem estar relacionados a alguns fatores como mão de obra, máquina, matéria-prima, meio ambiente, medição e método (SAMOHYL, 2005). Dentre os fatores correlatos a máquina pode-se citar colheita com meia plataforma ou arremate de áreas, falha em sensores, dentre outros que são normais no processo de colheita e não necessariamente devem ser excluídos pois fazem parte do colheita mecanizada e por não apresentarem repetibilidade ao longo das safras no mesmo local serão diluídos com a interpolação dos dados.

Em estudo desenvolvido por Delalibera *et al.*, 2017, que propõe e aplica métodos experimentais de análise de dados para avaliar perdas de grãos de feijão em plataforma durante a colheita mecanizada, concluíram que os experimentos realizados com a aplicação da metodologia *Jackknife* durante a análise preliminar dos dados, apresentaram diversas vantagens, como a qualidade, representatividade e a confiabilidade das análises.

Nas Figuras 4 de A a E correspondentes aos histogramas de frequência dos índices de desvio para a produtividade de milho e soja, após a utilização do método *Jackknife* verifica-se uma melhor distribuição dos dados ao se remover os pontos discrepantes. Com o tratamento de dados (Figuras 4B e

4D), constata-se um comportamento de assimetria a esquerda na curva refletindo na Figura 4E a qual se refere ao histograma de distribuição de frequência para a média entre os índices obtido através da álgebra entre mapas. Caso este comportamento perdure dentre os anos com culturas diferentes pode-se apontar como uma característica da área, que possui tendência, em sua maior parte, de ter índice de produtividade acima da média local (maior parte dos dados concentrados entre 0 e 1).

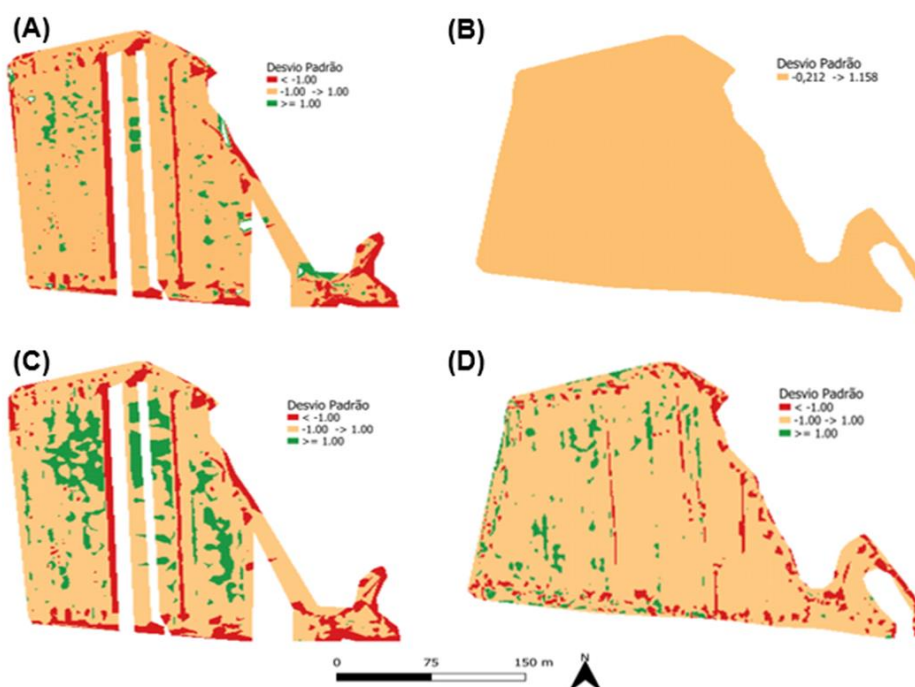
Por se tratar de dados de natureza contínua, se o histograma apresentar comportamento binomial aponta a existência de populações diferentes. Caso contrário, dados contínuos (Figura 4) apontam a existência de apenas uma população de dados.

Nota-se também melhor distribuição dos dados após a utilização do método *Jackknife* (Figuras 4B e 4D) ressaltando a influência dos dados discrepantes extremos sobre o comportamento destes. Também é possível observar essa melhora na distribuição, nos diagramas de frequência ordenada dos índices de desvio, visto o grau de ajuste das retas (R^2) para ambas colheitas.

Uma das formas de se observar a qualidade da distribuição do fenômeno é visualizar o grau de linearidade dos índices de desvio, o qual é obtido através da estimativa do R^2 do ajuste da reta nos diagramas da frequência ordenada dos índices de desvio, no qual, quanto mais próximo de 1 maior o grau de linearidade dos dados e, portanto, conforme prevê algumas análises de teste de normalidade, melhor é o ajuste da distribuição a normal, o que reflete na representatividade das medidas de tendência central dos dados.

A Figura 5 corresponde a distribuição espacial bidimensional da interpolação dos dados acima citados em forma gráfica. As Figuras 5A e 5C correspondem apenas a conversão dos dados de produtividade de milho e soja, respectivamente, em desvios padronizados para visualização do resultado entre os mapas com e sem a aplicação do método em uma mesma escala e comparando-as com as Figuras 5B e 5D, que representam os índices de produtividade das culturas de milho e soja, respectivamente, com a retirada dos pontos discrepantes pelo método *Jackknife* e verificação da linearidade dos índices de desvio.

Figura 5 – Mapas de desvios padronizados para o rendimento das culturas, sem e com a aplicação do método *jackknife*, classificados em três zonas de potencial produtivo por mais ou menos um desvio padrão da média.

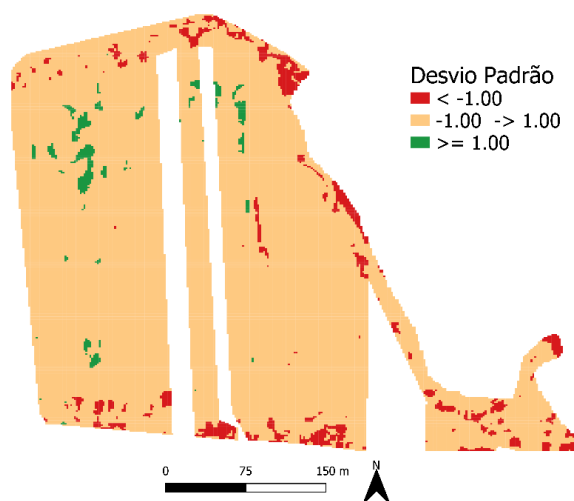


(A) – Dados de milho sem tratamento; (B) – Dados de soja sem tratamento; (C) – Dados de milho com tratamento *Jackknife*; (D) – Dados de soja com tratamento *Jackknife*;

Utilizando-se dos mapas de soja e milho após a aplicação do método *Jackknife*, foi elaborado um terceiro mapa, que consiste na álgebra entre os índices dos mapas de soja e milho (Figura 6), visando resumir a informação do potencial produtivo da área um único mapa, o qual foi classificado em três zonas de potencial produtivo a partir da metodologia dos intervalos de mais ou menos 1 desvio, que, em hipótese, podem se tornar zonas de manejo através do levantamento de variáveis de campo e proposição de manejos distintos para cada referência.

Para Blackmore *et al.* (2003) e Lark *et al.* (1999), um histórico de várias safras e de diferentes culturas é necessário para se identificar e propor zonas de manejo associando às características do solo.

Figura 6 – Mapa de tendências de produtividade da área, resultante da álgebra entre os mapas dos índices de produtividade de milho e soja.



No mapa médio de representação da tendência de produtividade da área (Figura 6), observou-se que o comportamento dos histogramas de distribuição de frequência entre as culturas distintas foi mantido. Já onde o fenômeno não se mantém, as zonas de manejo são diluídas, não sendo mantidas no mapa final.

Na Tabela 1 estão discriminadas as áreas em hectare correspondentes às classes de potencial produtivo de cada mapa, onde as cores amarelo, vermelho e verde correspondes aos potenciais produtivos na média, abaixo e acima da média local.

Tabela 1 – Área em hectare dos potenciais produtivos representados por cores e pelo Coeficiente de Variação (CV) dos mapas.

Mapas	Área (ha)					
	Vermelho ¹	(%)	Amarelo ²	(%)	Verde ³	(%)
Milho sem tratamento	1.50	13.16	9.34	82.10	0.54	4.74
Milho com tratamento	1.61	14.03	8.24	71.93	1.61	14.04
Soja sem tratamento			17.21	100.00		
Soja com tratamento	1.28	7.46	15.03	87.40	0.88	5.14
Milho+Soja com tratamento	0.58	5.20	10.34	93.24	0.17	1.57

¹valores abaixo da média; ² valores na média; ³ valores acima da média;

Por meio da Tabela 1 nota-se que, com a retirada de pontos discrepantes há um aumento das áreas de baixo e alto potencial produtivo

(vermelho e verde respectivamente) e redução na área de potencial médio normal (em amarelo do mapa da cultura do milho) após a aplicação do tratamento *Jackknife*, o que representa uma melhora na distribuição de frequências dos dados e melhor representatividade das medidas de tendência central. A distribuição destes dados observada na Figura 4F, demonstra existência de pontos extremos próximos a 15 desvios, onde o desejado é não conter índices maiores que 3,09. Também se observa uma concentração de dados próximo a 1 desvio padrão da média, o que indica que a maior parte da área apresenta rendimento acima da média local, ou seja, a área apresenta tendência de elevação gradativa do potencial produtivo, isto indica que para as safras futuras há tendência de aumento gradativo da média de produtividade.

Como mostra a Figura 5, a presença de pontos discrepantes muito distantes do centro da distribuição acaba por mascarar ou proporcionar diferenças que não são consistentes em uma área. Todos os pontos estão dispostos em apenas uma classe (Tabela 1). É possível explicar isto pela grande dispersão dos dados, ao observar Figura 4H nota-se que o Coeficiente Angular da reta estimada para a linearidade dos índices de desvios dos dados é próximo a 0. Portanto, mesmo não havendo dados acima de 3,09, para obter melhor distribuição dos dados foram excluídos via *jackknife* pontos da extremidade superior (Figura 4I), visando obter melhor linearidade dos dados.

Tabela 2 – Análise descritiva dos dados com e sem tratamento de dados *Jackknife*.

Análise descritiva	Milho			Soja		
	SJn (kg.ha ⁻¹)	SJn (IDP)	CJn (IDP)	SJn (kg.ha ⁻¹)	SJn (IDP)	CJn (IDP)
Média	6869.52	0.00	0.00	4097.58	0.00	0.00
Erro padrão	40.96	0.01	0.02	230.10	0.01	0.02
Mediana	7480.00	0.20	0.30	3937.00	-0.01	0.19
Modo	8819.00	0.62	0.89	4724.00	0.03	0.19
Desvio padrão	3124.72	1.00	1.00	18055.48	1.00	1.00
Variância da amostra	9763882.29	1.00	1.00	326000378.75	1.00	1.00
Curtose	19.64	19.64	-0.44	6024.11	6024.11	1.14
Assimetria	1.57	1.57	-0.82	77.21	77.21	-0.98
Intervalo	47123.00	15.08	3.66	1412770.00	78.25	6.02
Mínimo	121.00	-2.16	-2.38	143.00	-0.22	-2.98
Máximo	47244.00	12.92	1.28	1412913.00	78.03	3.04
Soma	39987453.00	0.00	0.00	25228816.00	0.00	0.00
Contagem	5821.00	5821.00	3442.00	6157.00	6157.00	3756.00
Coefficiente de Variação	45.49	87.89	86.04	440.64	107.78	102.03

SJn - Sem tratamentos de dados *Jackknife*; CJn - Com tratamento de dados *Jackknife*; IDP – Índice de desvio padronizado.

O CV representa a variabilidade do conjunto de dados. Na Tabela 2, ao retirar os pontos discrepantes observa-se uma redução do CV de 2% para cultura do milho e de 5% para cultura da soja.

Para cultura da Soja, onde antes do tratamento de dados os desvios estavam entre $\pm 3,09$, ao retirar os pontos na extremidade a média é ajustada a distribuição e conseqüentemente os valores dos desvios são movidos, apresentando uma leve redução do CV e melhor índice de ajuste a linearidade dos desvios, o qual é representado pelo R^2 da reta. Também pode-se dizer que o elevado CV para ambas as culturas é uma característica do fenômeno colheita e, portanto, deve-se conviver com esta grande variabilidade na aplicação de análises sequentes.

O uso de dados de diversos anos agrícolas e culturas distintas, para a criação do mapa histórico de índice de potencial produtivo, objetiva diluir efeitos não controlados de fatores como clima, genética e culturas distintas,

disponibilidade hídrica no solo, ressaltando assim os efeitos de variáveis de solo, as quais são de possível manejo.

A utilização deste método possibilita a exclusão de pontos utilizando um critério estatístico simples e de fácil aplicação. Por meio dele também é possível dar continuidade a análise utilizando dados de culturas diferentes para se obter uma informação final média e resumida em apenas um mapa simplificando a análise e tomada de decisão. Ao classificar os mapas em 3 níveis de potencial produtivo distintos, também buscou-se tornar a análise mais simples e fácil de ser interpretada e aplicada em campo tomando-se parâmetro de diferença um referencial estatístico, como o desvio da média.

Para a presente análise foram utilizados dados de duas safras agrícola, porém, se tem como hipótese de que com um conjunto maior de dados, ou seja, mais safras, a representatividade do mapa final de zonas de potencial produtivo (mapa médio, obtido pela álgebra dos mapas de cada safra) seja mais condizente com a realidade real da área. Portanto, sugere-se como continuidade deste trabalho, estudos direcionados ao levantamento de dados em campo, com base nas classes ou zonas de potencial produtivo formadas pelo mapa médio histórico final, visando identificar as variáveis de solo que se correlacionam com as zonas formadas e, assim, sugerir manejos localizados e mais adequados para cada zona obtendo-se por fim um mapa de tendências de produtividade.

5 CONSIDERAÇÕES FINAIS

O mapa de potencial produtivo obtido pode ser empregado para a definição de parâmetros de gestão da atividade agropecuária, como as zonas de manejo, reduzindo os custos de produção.

O comportamento de assimetria a esquerda da curva observado no histograma de distribuição de frequência, indica uma característica da área caso este fenômeno durante os anos com tendência a aumento de produtividade.

6 CONCLUSÕES

A aplicação dos métodos de análise de desvios sobre os dados é viável, pois, pode-se aplicar a remoção de dados discrepantes de forma fácil e com base em critérios estatísticos consagrados.

É possível realizar a álgebra de mapas para a obtenção de um único mapa médio do histórico dos índices de colheita, a partir de mapas de colheita de culturas distintas.

Dar mais confiabilidade nos dados obtidos das colhedoras e de outras operações instrumentadas.

REFERÊNCIAS

- AGUIAR, P. A. A. **Padronização da umidade dos grãos na experimentação agrícola**. In: Embrapa Semiárido-Resumo em anais de congresso (ALICE). In: EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Centro de Pesquisa Agropecuária do Trópico Semi-Árido. Petrolina. Resumo de atividades de pesquisa. Petrolina, 1977.
- BERNARDI, A. C. C.; FRAGALLE, E. P.; INAMASU, R. Y. **Inovação tecnológica em AP**. Livro AP: um novo olhar. São Carlos, SP: Embrapa Instrumentação, 2011. p. 297-302.
- Bernardi, A. C. D. C., de Souza, G. B., Polidoro, J. C., Paiva, P. R. P., & Monte, M. B. D. M. Yield, quality components, and nitrogen levels of silage corn fertilized with urea and zeolite. **Communications in soil science and plant analysis**, v. 42, n. 11, p. 1266-1275, 2011.
- BERNADI, A. C. DE C.; NAIME, J. DE M.; RESENDE, A. V. DE; BASSOLI, L. H.; INAMASU, R. Y. **Agricultura de precisão: Resultado de um novo olhar**. Brasília, DF. Embrapa, 2014.
- BRASE, T. A. **Precision agriculture** / Terry A. Brase., 1st ed., Clifton Park, NY: Thomson/Delmar Learning, 2006.
- BLACKMORE, S. **Consciousness in meme machines**. **Journal of Consciousness Studies**, v. 10, n. 4-5, p. 19-30, 2003.
- BLACKMORE, S.; MARSHALL, C. J. **Yield mapping; errors and algorithms**. In: International Conference on Precision Agriculture, 3, Minneapolis. Proceedings... Madison, WI: ASAE, p.403-415, 1996.
- BLACKMORE, S.; Moore, M. Remedial correction of yield map data. **Precision Agriculture**, v.1, n.1, p.51-66, 1999.
- CAMPOS BERNARDI, A. C.; BETTIOL, G. M.; GREGO, C. R.; ANDRADE, R. G.; RABELLO, L. M.; INAMASU, R. Y. **Ferramentas de agricultura de precisão como auxílio ao manejo da fertilidade do solo**. **Cadernos de Ciência & Tecnologia**, v. 32, n. 1/2, p. 211-227, 2017.
- CAMBARDELLA, C. A.; KARLEN, D. L. Spatial analysis of soil fertility parameters. **Precision Agriculture**, v. 1, n. 1, p.-14, 1999.
- CHIODEROLI, C. A.; SILVA, R. P.; NORONHA, R. H. F.; CASSIA, M. T.; SANTOS, E. P. **Perdas de grãos e distribuição de palha na colheita mecanizada de soja**. **Bragantia**, v.71, p.112-121, 2012.
- COMPAGNON, A. M.; SILVA, R. P.; CASSIA, M. T.; Graat, D.; Voltarelli, M. A. **Comparação entre métodos de perdas na colheita mecanizada de soja**. **Scientia Agropecuaria**, v.3, p.215-223, 2012.

DELALIBERA, H. C.; HIGASHIBARA, L. R.; RALISCH, R. **Estimativa de perdas na colheita mecanizada de grãos: Adequação da metodologia para plataformas universais modernas**. In: Anais do XLVI Congresso Brasileiro de Engenharia Agrícola – CONBEA, 2017.

GRISSE, R.; JASA, P.; SCHROEDER, M. A.; WILCOX, J. C. **Yield monitor accuracy: Successful Farming Magazine case study**. St. Joseph: ASAE, p.10 1999.

HARTER, H. L. Tables of range and studentized range. **The Annals of Mathematical Statistics**, p. 1122-1147, 1960.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. Bookman editora, 2009.

KHAKURAL, B. R.; ROBERT, P.C.; HUGGINS, D.R. **Variability of corn/soybean yield and soil/landscape properties across a southwestern Minnesota Landscape**. In: INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE, 4, 1998.St. Paul Proceedings... Madison: American Society of Agronomy, 1998. P. 573-9.

KÖPPEN, W. P.; **Grundriss der Klimakunde**. Berlim, Walter de Gruyter, 1931.

LAMPARELLI, R. A. C. AP. **Agência Embrapa de Informação Tecnológica**. 2016. Disponível em: <http://www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/arvore/CONTAG01_72_711200516719.html>.

LARK, R. M.; BOLAM, H. C.; MAUR, T.; BRADLEY, R. I.; BURTON, R. G. O.; DAMPNEY, P. .M P. Analysis of yield maps in support of field investigation. In: STAFFORD, J. **Precision Agriculture' 99**. Sheffield: Sheffield Academic Press, 1999. Part 1 Lentner and Bishop, 1993

LEROUX, C.; JONES, H.; CLENET, A.; DREUX, B.; BECU, M.; TISSEYRE, B. A general method to filter out defective spatial observations from yield mapping datasets. **Precision agriculture**, v.19, n.5, p.789-808, 2018.

MENEGATTI, L. A. A.; MOLIN, J. P. **Remoção de erros em mapas de produtividade via filtragem de dados brutos**. In: Revista Brasileira de Engenharia Agrícola e Ambiental, v.8, n.1, p.126-134, 2004.

MENEGATTI, L. A. A.; MOLIN, J. P. **Metodologia para identificação, caracterização e remoção de erros em mapas de produtividade**. Dissertação. Escola superior de Agricultura Luiz de Queiroz - USP. Piracicaba, 2002.

MIKHAIL, E. M.; BETHEL, J. S.; MCGLONE, J. C. **Introduction to modern Photogrammetry**. John Wiley & Sons, p. 479, 2001.

MILANI, L.; SOUZA, E. G. de; URIBE-OPAZO, M. A.; GABRIEL FILHO, A.; JOHANN, J. A.; PEREIRA, J. O. Unidades de manejo a partir de dados de produtividade. **Acta Scientiarum Agronomy**, v. 28, p. 591-598, 2006.

MOLIN, J. P. **Geração e Interpretação de Mapas de Produtividade para Agricultura de Precisão.** In: Borém, A.; Giúdice, M. P.; Queiroz, D. M.; Mantovani, E. C.; Ferreira, L. R.; Valle, F. X. R.; Gomide, R. L. Agricultura de Precisão. Viçosa: Universidade Federal de Viçosa, p. 237-258, 2000.

MOLIN, J. P.; CREMONINI, L. C. M.; MENEGATTI, L.; GIMENEZ L. **Acurácia de um monitor de produtividade com sensor de fluxo volumétrico.** In: CONGRESSO BRASILEIRO DE ENGENHARIA AGRÍCOLA, 2000, Fortaleza. Anais/CD-Rom Fortaleza: SBEA, 2000.

MOLIN, J. P. Definição de unidades de manejo a partir de mapas de produtividade. **Engenharia Agrícola**, v.22, n.1, p.83-92, 2002. <<http://www.leb.esalq.usp.br/download/APP%202002.02.PDF>>.

MOORE, M. **An investigation into the accuracy of yield maps and their subsequent use in crop management.** (Thesis Ph.D.) - Silsoe College. Silsoe, p. 379. 1998.

PIERCE, F. J.; NOWAK, P. **Aspects of precision agriculture.** Adv. Agronomy, v. 67, p.1-85, 1999.

QUEIROZ, D. M.; DIAS, G. P.; MANTOVANI, E. C. **Agricultura de precisão na produção de grãos.** In: Borém, A.B.; Giúdice, M.P.; Queiroz, D.M.; Mantovani, E.C.; Ferreira, Valle, F.X.R.; Gomide, R.T. Agricultura de Precisão. Viçosa: UFV, p.1-42. 2000.

SAMOHYL, R. W. **Controle estatístico de processo e ferramentas da qualidade.** In M. M. Carvalho, & E. P. Paladini (Eds.), Gestão da qualidade: Teoria e casos Rio de Janeiro: Elsevier. Cap. 9, p. 261-269, 2005.

SANTOS, H. G. et al. **Sistema brasileiro de classificação de solos.** Brasília, DF: Embrapa, 2018., 2018.

SOUZA, E. G.; BAZZI, C. L.; KHOSLA, R.; URIBE-OPAZO, M. A.; REICH, R. M. Interpolation type and data computation of crop yield maps is important for precision crop production. **Journal of Plant Nutrition**, v. 39, n. 4, p. 531-538. 2016.

SNEDECOR G. W., COCHRAN, W. G. Statistical Methods. Iowa State University Press. **Biometry**, ed.7, p.593, 1980.

SPEKKEN, M. A. R. K.; ANSELMINI, A. A.; MOLIN, J. P. A simple method for filtering spatial data. In: **Precision agriculture'13.** Wageningen Academic Publishers, Wageningen, p. 259-266, 2013.

SPERANZA, E. A., CIFERRI, R. R., & CIFERRI, C. D. A. **Clustering approaches and ensembles applied in the delineation of management classes in precision agriculture.** In Embrapa Informática Agropecuária-Artigo em anais de congresso (ALICE). In: BRAZILIAN SYMPOSIUM ON GEOINFORMATICS, 17., 2016, Campos do Jordão. Proceedings... São José dos Campos: INPE, 2016.

WU, C. J. *et al.* Jackknife, bootstrap and other resampling methods in regression analysis. **The Annals of Statistics**, v. 14, n. 4, p. 1261-1295, 1986.

APÊNDICE

APÊNDICE A Script para remoção de dados repetidos e para retirada de dados nulos e zerados

Pacotes necessários – xlsx, csv, sciplot e readxl.

1º - Informar onde estão os dados e qual arquivo deverá ser utilizado. Após inserir estas informações, é necessário acessar os dados.

```
setwd("C:/Users/...")
data.set=readxl::read_excel("Nome do arquivo.xlsx", sheet=1)
attach(data.set)
```

2º - Neste passo são eliminados dados de velocidade inferior a 1 km h⁻¹. Este comando pode ser aplicado aos demais dados como produtividade e umidade apenas modificando o nome do campo pelo desejado.

```
data.set2<-subset(data.set, Velocidade>1)
```

3º - Com os valores de umidade, torna-se possível a ajuste da produtividade à 13% de umidade do grão, umidade desejada para comercialização. Para isto é utilizado este comando com a fórmula.

```
data.set<-within(data.set4, Prod2<-c((100-Umidade)/(100-13)*Prod1))
attach(data.set)
```

4º - Por fim, este comando possibilita gerar um arquivo em formato csv e xlsx. O formato cvs é utilizado para posterior inserção da tabela final em um software SIG.

```
write.csv(resultado, "mapa1.csv")
library(xlsx)
write.xlsx(resultado, "mapa2.xlsx")
```

5º - Como função opcional, foram inclusas algumas análises para verificação dos dados.

Histograma

```

histograma <-
hist(resultado$Produtividade,
breaks=c(seq(0,10000,500)),
col = "green",
main = "Histograma",
ylab="Frequência",
las=1,
xlab="Produtividade",
ylim=c(0,1000))
xfit<-
seq(min(resultado$Produtividade),max(resultado$Produtivida
de))
yfit<-
dnorm(xfit,mean=mean(resultado$Produtividade),sd=sd(resulta
do$Produtivida))
yfit <-
yfit*diff(histograma$mids[1:2])*length(resultado$Produtivida
)
lines(xfit, yfit, col="red", lwd=1)
plot(resultado$Produtividade~
as.POSIXct(resultado$Data_Hora),
type="o",
ylab='Produtividade',
xlab="Tempo", las=1) # formato de series temporais
plot(resultado$Produtividade[order(resultado$Produtividade)],
ylab="Produtividade",
xlab="") # formato contínuo
duplicated(resultado$Produtividade)

```

Intervalo de confiança

```

attach(resultado)
sciplot::lineplot.CI(Data_Hora[order(Produtividade)],
  Produtividade[order(Produtividade)],
  ylab="Produtividade", las=1)
abline(h=mean(Produtividade), lty=1,col="red")
abline(h=mean(Produtividade)-sd(Produtividade),
  lty=2,col="red")
abline(h=mean(Produtividade)+sd(Produtividade),
  lty=2,col="red")
abline(h=max(Produtividade), lty=3,col="blue")
abline(h=min(Produtividade), lty=3,col="blue")

```

#Estatística Descritiva

```

attach(datadf)
mean(Prod) # Calcula a média de todos os valores
numéricos
median(Prod) # Calcula a mediana de todos os valores
numéricos
var(Prod) # Variância populacional
sd(Prod) # Desvio padrão populacional
max(Prod) # Valor máximo
min(Prod) #Valor mínimo
range(Prod) #amplitude
quantile(Prod) # Quantis_25% (Quartis)
length(Prod)# Número de observações

```

Obs – data.set é o nome dado ao conjunto de dados, pois à medida que os modifica, a cada alteração, é necessário renomear o arquivo (optou-se por incluir um número, exemplo data.set1) e este pode variar de acordo com cada necessidade.