

Remoção de erros em mapas de produtividade de cana-de-açúcar

Removal of errors on maps of productivity of sugarcane

Leonardo Felipe Maldaner¹, Mark Spekken¹, Mateus Tonini Eitelwein¹, José Paulo Molin²

¹ Pós-graduando em Engenharia de Sistemas Agrícolas, Departamento de Engenharia de Biossistemas, ESALQ/USP, Piracicaba, São Paulo, Brasil, leonardofm@usp.br, mspekken@usp.br, mateus_eitelwein@usp.br

² Departamento de Engenharia de Biossistemas, ESALQ/USP, Piracicaba, São Paulo, Brasil, jpmolin@usp.br

RESUMO

O processo de automação de colhedoras para monitoramento da produtividade permite que uma grande quantidade de dados seja coletada, sendo comum a ocorrência de erros na sua coleta. Devido à presença destes, metodologias de filtragem de dados foram desenvolvidas para obtenção de mapas representativos da realidade. O objetivo deste trabalho é implementar modelos de filtragem e submetê-los a dados de produtividade de cana-de-açúcar, com características distintas, para avaliar seu desempenho. O algoritmo proposto realiza dois procedimentos de filtragem distintos para dados armazenados em sequência: 1) filtragem local; 2) filtragem global. Uma filtragem combinada foi realizada aplicando-se, respectivamente, os filtros local e global sobre os dados. Todos os processos de filtragem aplicados foram capazes de efetuar a remoção de valores discrepantes, porém com características distintas. A filtragem local mostrou-se mais apropriada para observar a real produtividade, enquanto o processo global de limpeza de dados mostrou-se mais apropriado para sinalizar as regiões por seu real potencial produtivo.

PALAVRAS-CHAVE: Filtragem de dados espaciais, Mapeamento de produtividade, Cana-de-açúcar, Agricultura de precisão.

ABSTRACT

The automation process of harvesters for monitoring productivity allows a large amount of data to be collected, and common errors in its collection. Due to the presence of errors, many data filtering methodologies were developed for obtaining grain yield maps. This work proposes the implementation and assessment of filtering procedures to be submitted to sugarcane yield recorded data. An implemented algorithm submits spatial data to two distinct filtering procedures for data stored in a machine pass sequence: 1) a local filter; 2) a global

filtering with the values within a machine pass. A third filtering procedure applied both procedures in the respective order mentioned. All the filtering processes applied were capable to remove outlying data, yet with distinct characteristics. The local filtering process showed itself more appropriate to keep the real yield after filtering, while the global data cleaning process showed itself more appropriate to address the regions for their respective yield potential.

KEYWORDS: Spatial data filtering, Yield mapping, Sugarcane, Precision agriculture.

INTRODUÇÃO

A utilização de ferramentas de agricultura de precisão no setor sucroenergético tem aumentado consideravelmente nos últimos anos visando à otimização do uso de recursos e aumento da produtividade. A agricultura de precisão (AP) é um sistema de gerenciamento do sistema solo-planta-atmosfera baseado nos princípios da caracterização e gestão da variabilidade espacial que engloba fatores de produção e produtividade das culturas (MONTANARI et al., 2012).

Segundo Molin (2001) as atividades relacionadas à AP compõem um ciclo de procedimentos que se inicia com o mapeamento da produtividade das culturas. Adotar o monitoramento da produtividade como ponto de partida das atividades de AP se justifica facilmente. Os mapas de produtividade expressam a resposta da planta a um complexo conjunto de práticas de manejo adotadas durante o seu ciclo, permitindo observar a variabilidade e investigar os fatores que a afetam. No caso da cana-de-açúcar em que a colheita é realizada em uma ou duas fileiras, os dados de produtividade podem indicar problemas pontuais ocasionados por falha no plantio e danos na soqueira, podendo auxiliar nas práticas de manejo localizado, uma vez que o canavial é explorado por aproximadamente cinco anos antes da reforma.

O processo de automação de colhedoras para monitoramento da produtividade permite que uma grande quantidade de dados seja coletada. Segundo Molin (2000) é preciso ter cautela na utilização dos dados, pois nem todos os pontos refletem a produtividade real, sendo comum a ocorrência de erros na sua coleta. Thylén et al. (1997) identificaram diversos erros em mapas de produtividade de grãos e os classificaram como relacionados às características físicas da colhedora, do sistema de medição da produtividade, relacionados às variações dentro do talhão, ao operador e devido ao procedimento de obtenção do mapa.

Segundo Gimenez e Molin (2004) a presença de erros nos dados utilizados para a elaboração dos mapas de produtividade é muito prejudicial à qualidade do mapa gerado,

podendo inclusive levar a interpretações errôneas mesmo que o volume de erros em relação ao total de dados coletados seja pequeno. Desta forma, é extremamente importante que esses erros sejam removidos antes de qualquer análise dos mapas. Nesse sentido Menegatti e Molin (2003) desenvolveram uma metodologia para caracterização dos erros em mapas de produtividade de grãos. Posteriormente os mesmos autores (MENEGATTI e MOLIN, 2004) desenvolveram um modelo de filtragem dos dados estruturado em sete etapas compostas na seguinte ordem: 1) remoção dos erros grosseiros de posicionamento; 2) remoção dos pontos com produtividade nula ou ausente; 3) remoção dos pontos com largura parcial de plataforma; 4) remoção dos dados com umidade nula ou ausente; 5) remoção dos pontos com distância nula; 6) remoção dos pontos gravados durante o tempo de enchimento; 7) remoção de pontos com produtividade discrepante.

Como opção alternativa para filtragem de dados, procedimentos foram propostos para identificar a adequação de um dado espacial por sua consistência de valor com dados vizinhos. Tais procedimentos foram também implementados em modelos para identificação e eliminação de erros (GIMENEZ e MOLIN, 2004; SPEKKEN et al., 2013).

Para o sistema de colheita mecanizado na cana-de-açúcar há trabalhos focando o mapeamento de produtividade, tais como Cox et al. (1997), Mailander et al. (2002), Pagnano (2001) e Cerri (2005), sendo que os autores relatam a existência de erros nos dados de produtividade. Muitos desses erros em dados de produtividade em cana-de-açúcar são em função da dinâmica da relação entre máquina, monitor e cultura.

A despeito da presença de erros, muitas das metodologias de filtragem de dados foram desenvolvidas para dados de produtividade de grãos. Neste sentido, o objetivo deste trabalho é implementar modelos de filtragem e submetê-los a dados de produtividade de cana-de-açúcar, com características distintas, para avaliar sua performance.

MATERIAL E MÉTODOS

Os dados necessários para os procedimentos aqui utilizados para filtragem resumem-se à localização dos registros de produtividade, provido em forma de coordenadas geográficas e o próprio registro da produtividade, para os quais os procedimentos implementados independem da unidade em que estes foram medidos (Mg ha^{-1} ou kg ha^{-1}). A Tabela 1 fornece um exemplo de disposição dos registros de entrada do modelo.

Coordenadas geográficas são convertidas em coordenadas métricas para formação de um plano cartesiano de plotagem de pontos. Essa conversão transforma as coordenadas geográficas decimais em coordenadas UTM (Universal Transversa de Mercator). Como esta

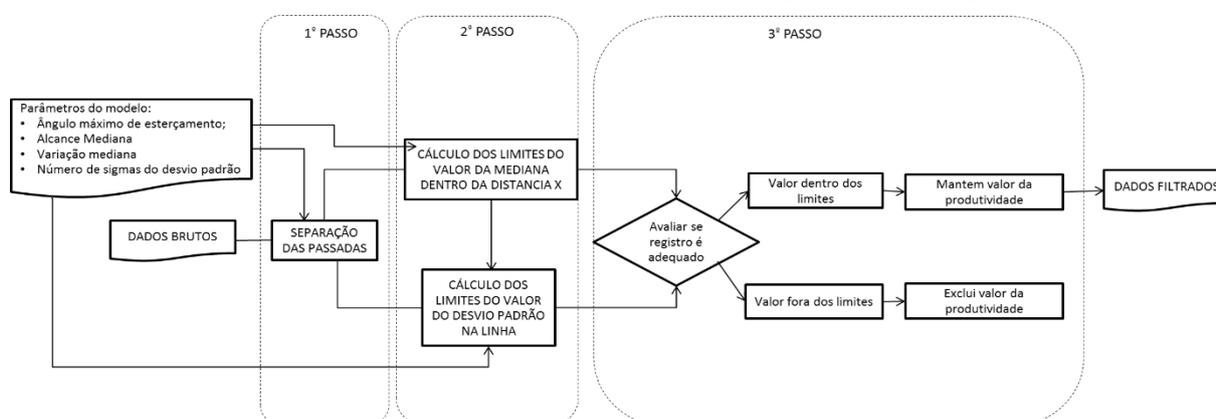
conversão limita conjuntos de coordenadas dentro de blocos/zonas (com alcance de 1000 x 1000 km), o modelo não tolera coordenadas que se dividam em blocos (zonas UTM) diferentes.

Tabela 1. Exemplo de um arquivo utilizado na filtragem.

Latitude	Longitude	Prod.
-21.3792	-47.9121	191.66
-21.3792	-47.9121	210.16
-21.3792	-47.9121	109.96
-21.3792	-47.9122	14.01
-21.3792	-47.9122	14.01
-21.3792	-47.9122	14.01
-21.3792	-47.9122	56.15

As coordenadas e os dados de produtividade compõem os dados brutos necessários para filtragem, cujo modelo conceitual é disposto na Figura 1.

Figura 1. Fluxograma do modelo conceitual do processo de filtragem de dados de produtividade.



Pela premissa de que máquinas percorrem áreas em passadas paralelas alternadas (*boustrophedon paths*), a separação das passadas é feita pelo método modificado de Menegatti e Molin (2004) por meio da identificação dos extremos de uma passada. Porém, neste caso, o modelo implementado percorre ponto a ponto os registros armazenados extraindo a direção de trabalho da máquina (na forma de um ângulo em graus). Este ângulo é calculado para dois pontos localizados à frente de um ponto inicial (*look ahead*). Com ângulos obtidos para todos os pontos, quando a diferença do ângulo de um ponto-registro do próximo ponto-registro supera o Ângulo Máximo de Esterçamento (*AngME*) identifica-se o fim de uma passada e o início da próxima passada da colhedora.

A separação dos dados por fileira colhida antes da filtragem se justifica pelo fato de que os algoritmos de filtragem de dados de produtividade de grãos propostos por outros autores (MOLIN e GIMENEZ, 2000; MENEGATTI e MOLIN, 2004; PING e DOBERMANN, 2005;

SIMBAHANET et al., 2004; ARSLAN e COLVIN 2002; BLACKMORE e MOORE, 1999; SPEKKEN; ANSEMI e MOLIN, 2013) podem eliminar dados importantes para a identificação de falhas da cultura na linha de plantio.

Filtragem local

Após a separação das passadas da colhedora o modelo aplica uma filtragem local para os dados utilizando uma modificação do método proposto por Spekken et al. (2013). Neste, conjuntos de dados vizinhos são identificados uma distância constante, e um parâmetro estatístico é extraído entre estes. Neste caso, o parâmetro estatístico extraído é a mediana, e todos os dados vizinhos que se desviarem (acima ou abaixo) de um dado percentual desta mediana tem um peso de desclassificação atribuído a eles (adiciona-se um peso de '1'). Este processo é repetido para todos os pontos existentes, continuamente atribuindo pesos a todos os pontos cujos valores se sobressaem de seus vizinhos. Ao final, pontos com pesos iguais ao número de vizinhos são pontos com valores discrepantes e são identificados para eliminação. Neste trabalho, a identificação de pontos vizinhos é feito apenas dentro da linha de colheita (análise geoestatística anisotrópica). O percentual de desvio da mediana é um parâmetro fornecido pelo usuário.

Filtragem global da passada de máquina

Neste procedimento, a filtragem faz uso da metodologia modificada de Sudduth et al. (2012) que propõe filtrações globais dos dados de produtividade baseadas na remoção de dados extremos de uma distribuição normal de valores. Nesta filtragem, os dados da passada localizados em extremos de n sigmas (dados com valores distantes de n vezes o desvio padrão da média de produtividade) são marcados para exclusão.

O modelo foi implementado para que ambos os filtros (local e global) pudessem ser aplicados distintamente ou em conjunto (aplicando-se primeiramente o filtro local). Na última etapa é determinada a exclusão do dado a partir dos parâmetros determinados pelo usuário e a seleção do processo de filtragem.

Após serem submetidos ao algoritmo, os dados de coordenadas e valor da produtividade de cada ponto considerado pelo algoritmo, são gravados para um novo arquivo, podendo depois disso serem importados novamente pelo SIG para a geração dos mapas.

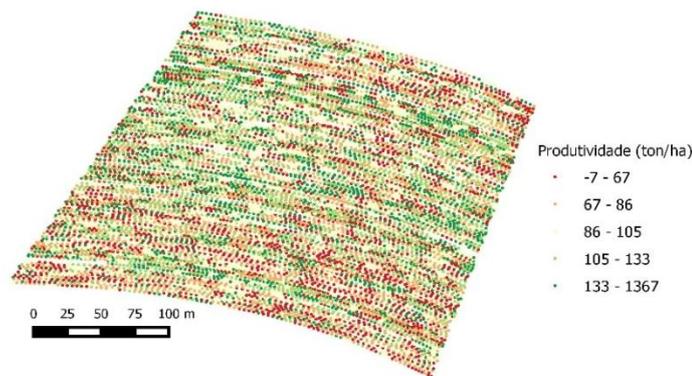
RESULTADOS E DISCUSSÃO

Para o teste do algoritmo utilizou-se um conjunto de 13.255 pontos de produtividade coletados por uma colhedora CASE 7700 equipada com o monitor de produtividade.

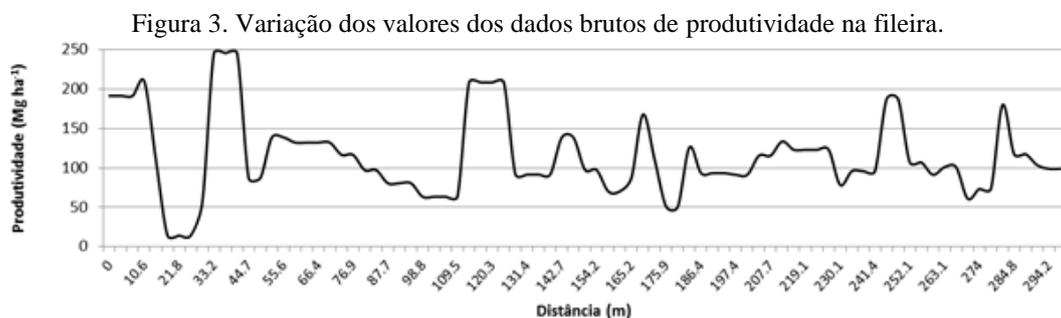
SIMPROCANA[®] - (ENALTA Soluções Tecnológicas). O sensor de registro de produtividade opera registrando a massa por meio de células de carga que mensuram a massa de cana que passa pela esteira antes de serem lançados ao veículo de transbordo. A colhedora opera cortando uma fileira por passada totalizando uma superfície de 8,01 ha.

Observando os dados brutos de produtividade percebe-se a ausência de manchas e uma alta quantidade de ruído nos pontos de produtividade, dado a ocorrência de produtividades elevadas e muito baixas em sequência (Figura 2).

Figura 2. Dados brutos da produtividade no talhão.



A intensa variação da produtividade dentro do talhão pode ser visualizada com maior clareza na Figura 3, onde são demonstrados os valores de produtividade ao longo de uma linha de colheita. Pode-se observar que há pontos de produtividade discrepantes (muito altos ou muito baixos) em uma pequena distância, o que mostra incoerência com a dependência espacial esperada da produtividade.



Foi realizada uma filtragem local permitindo-se uma variância dos dados de até 50% da mediana obtida localmente, considerando uma dependência espacial de produtividade de 10 m para determinação de pontos vizinhos. A filtragem global foi aplicada estabelecendo-se os limites superior e inferior a 1,2 sigmas dentro da passada da colhedora. Uma terceira filtragem foi realizada aplicando-se o filtro local seguido pelo global.

Os dados considerados discrepantes foram excluídos do conjunto original e gerou-se três arquivos de dados filtrados. Da estatística descritiva dos dados (Tabela 3), foi possível

verificar que após a remoção dos pontos discrepantes o valor mínimo de produtividade na filtragem local permaneceu inalterado, mas houve redução no valor máximo e no coeficiente de variação de todos os dados filtrados, mostrando a eficácia do filtro.

Tabela 3. Estatística descritiva dos dados brutos de produtividade.

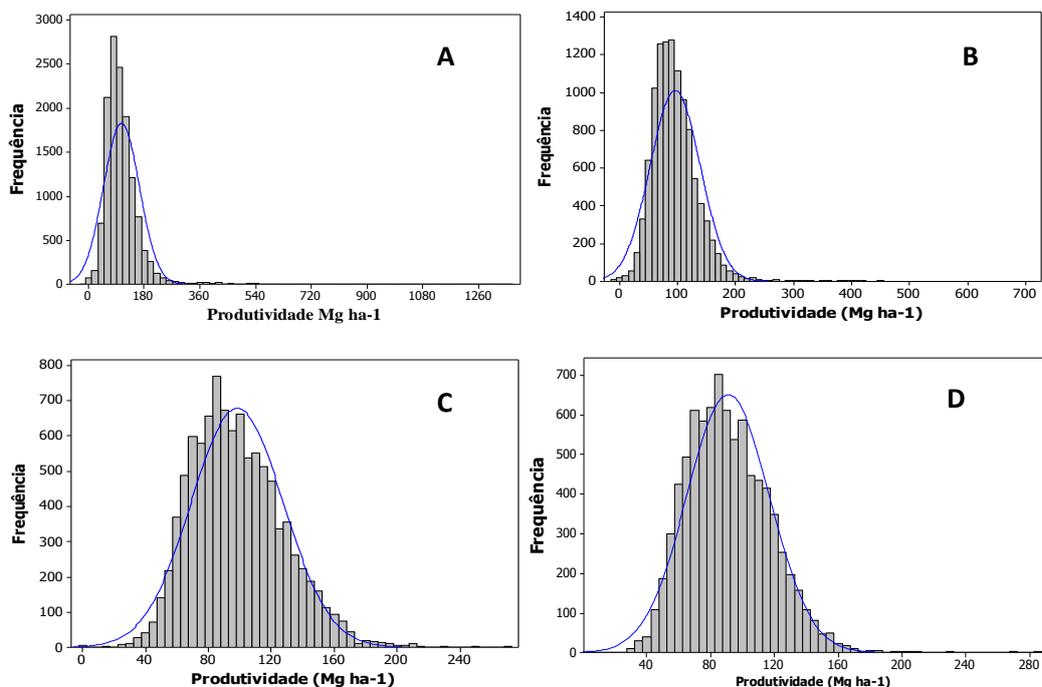
Parâmetros	Valores			
	Dados brutos	Filtragem 1	Filtragem 2	Filtragem 3
Nº pontos	13255	10931	9959	8458
Mínimo (Mg ha ⁻¹)	-11,76	-11,76	-4,50	6,14
Máximo (Mg ha ⁻¹)	1366,80	712,56	305,98	283,81
Média (Mg ha ⁻¹)	105,78	95,98	97,45	90,99
Mediana (Mg ha ⁻¹)	95,95	89,94	93,76	88,47
Simetria	3,98	2,70	0,58	0,66
Curtose	39,26	19,26	0,71	0,66
Variância	3363,77	1862,07	858,56	671,68
Desvio padrão	57,99	43,15	29,30	25,91
Coeficiente de variação (%)	54,83	44,96	29,83	28,48

A presença de dados de valor negativo (produtividade mínima na Tabela 3) se deve por características do sensor e sua respectiva calibração. A ocorrência de valores negativos acumulados em uma região (ou até mesmo em passadas curtas de máquina) levou a permanência de tais dados nos processos de filtragem.

Na filtragem global, como esperado, houve diminuição da amplitude e variância dos dados, sendo também bem-sucedido na eliminação dos dados discrepantes. Um número considerável de dados (36,2% da quantidade total) foi excluído mediante a aplicação de ambas as filtragens.

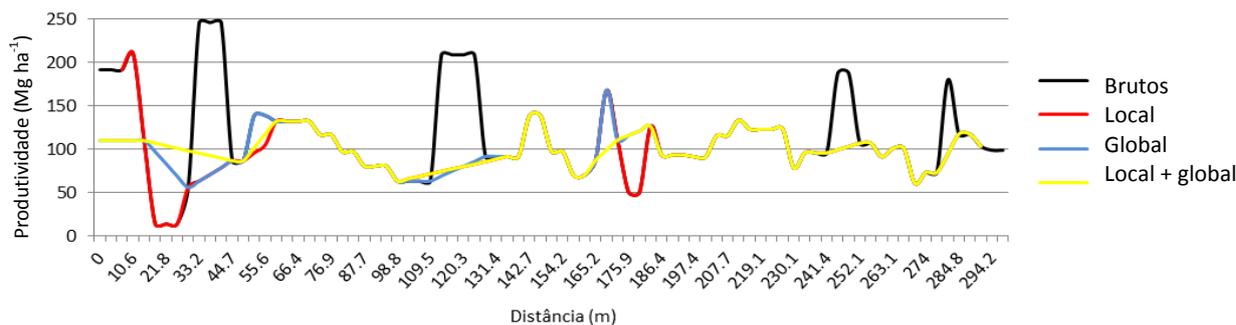
No histograma dos dados brutos na Figura 4a, observa-se uma distribuição anormal com assimetria positiva (3,98). Com a eliminação de dados pelas filtragens realizadas há uma diminuição dos valores de assimetria e curtose. Como esperado, a filtragem global induziu os dados a uma distribuição normal, devido a premissa de desvio de sigmas da média (da qual uma distribuição normal é esperada).

Figura 4. Histograma dos dados de produtividade originais (A), dados remanescentes da filtragem local (B), dados remanescentes da filtragem global (C) e dados remanescentes da aplicação de ambas as filtragens (D).



A variação da produtividade dentro do talhão depois da filtragem dos dados pode ser visualizada na Figura 5, onde um gráfico dos valores remanescentes ao longo de uma passada são exibidos.

Figura 5. Comparação da variação da produtividade em relação ao três métodos de filtragem de dados.

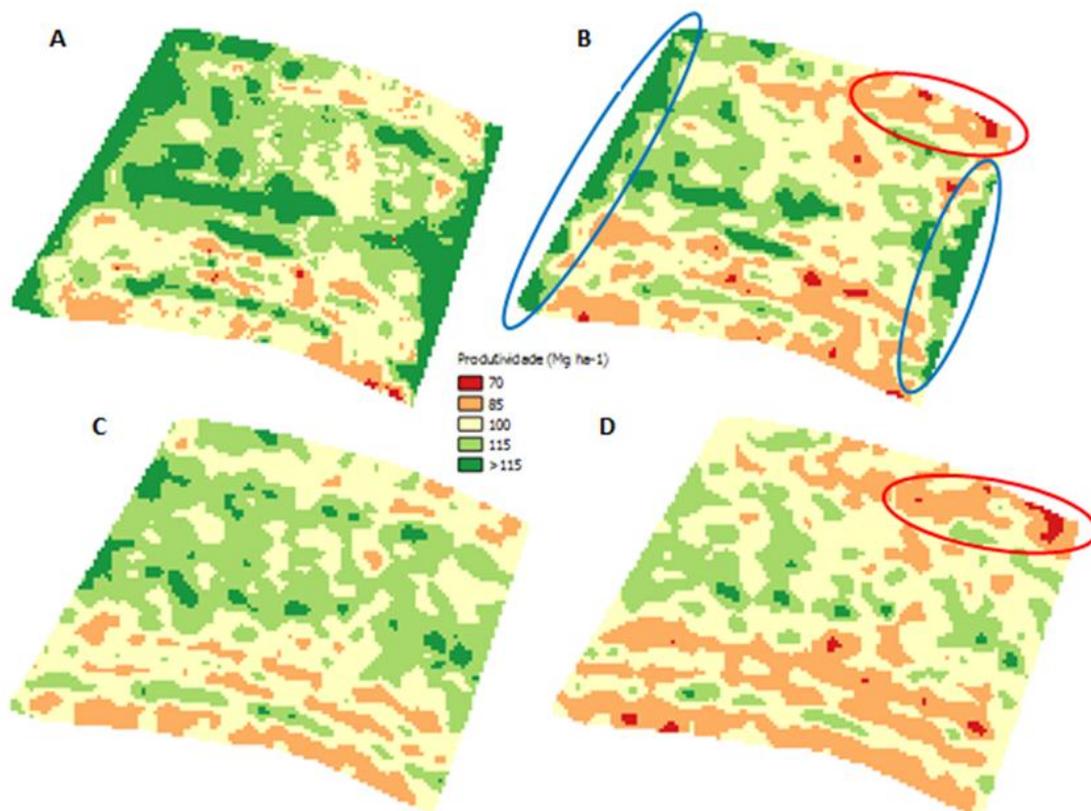


Diferentemente de dados de produção de grãos, na cana-de-açúcar podem-se observar falhas que ocorrem na implantação (com tendência crescente no decorrer das colheitas), onde produtividade nula (ou quase nula) ocorre. A filtragem local encontrou grupos de pontos com baixa produtividade e os manteve, enquanto a filtragem global delimitou os limites de produtividade (linhas azuis e amarelas no gráfico), eliminando-os. Dados discrepantes de alta produtividade foram efetivamente excluídos por ambos os métodos.

A eliminação de dados discrepantes pela filtragem local auxiliou na visualização de valores de baixa produtividade (circunferência vermelha, Figura 6B) antes não detectada com os dados brutos (Figura 6A). A filtragem global diminuiu a amplitude dos valores dos pontos remanescentes após a filtragem, eliminando os dados com valores de alta e baixa

produtividade (Figura 6C), entretanto realizando a filtragem global após a filtragem local manteve-se as regiões com valores de baixa produtividade (Figura 6D). A eliminação de dados com valores de alta produtividade pela filtragem local mudou os valores dos limites utilizados na filtragem global.

Figura 6. Mapas interpolados para o estudo de caso da cana: dados brutos (A), dados após a filtragem local (B), dados após a filtragem global (C) e dados após da aplicação de ambas as filtrações (D).



Observando a Figura 6 nota-se que os dados após a filtragem local (em 'B') mostram uma variabilidade mais evidente da produtividade dentro do talhão, identificando valores locais de produtividade baixa (elipse vermelha) e alta (como nas cabeceiras nas elipses azuis), enquanto os dados resultantes da filtragem global demonstram a variabilidade do potencial produtivo do talhão, pois eliminam flutuações de pequeno alcance. A aplicação de ambos os filtros (local + global) demonstrou grande eficiência na identificação de valores locais de baixa produtividade, porém houve eliminação de grande parte dos valores com alta produtividade.

Entretanto, os parâmetros do modelo no algoritmo de filtragem devem ser ajustados de acordo com cada área e com as necessidades do usuário, não havendo parâmetros padrão a serem utilizadas para a filtragem de dados de produtividade.

CONCLUSÕES

1. A cultura da cana-de-açúcar demonstra a necessidade da filtragem do conjunto de dados de produtividades devido a presença de erros gerados pelos sensores dentro de uma curta distância.
2. Os métodos de filtragem propostos são capazes de identificar e eliminar dados de produtividade de cana de açúcar com variação intensa em distâncias próximas.

REFERÊNCIAS

- ARSLAN, S., COLVIN, T. S. (2002). Grain yield mapping: yield sensing, yield reconstruction, and errors. *Precision Agriculture* 3, 135–154.
- BLACKMORE, B. S., MOORE, M. 1999. Remedial correction of yield map data. *Precision Agriculture* 1, 53–66.
- CERRI, D.G.P. Agricultura de precisão em cana-de-açúcar: instrumentação de uma colhedora, mapeamento da produtividade e de atributos do solo. 156p. Tese (Doutorado em Engenharia Agrícola) - Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, SP, 2005.
- COX, G.; HARRIS, H.; PAC, R. Development and testing of prototype yield mapping system. In: PROCEEDINGS OF AUSTRALIAN SOCIETY OF SUGAR CANE TECHNOLOGISTS, 1997.
- 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.
- MENEGATTI, L. A. A.; MOLIN, J. P. Remoção de erros em mapas de produtividade via filtragem de dados brutos. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.8, p.126-134, 2004.
- MOLIN, J.P. Agricultura de precisão: o gerenciamento da variabilidade. Piracicaba: O Autor, 2001. 83p
- MONTANARI, R.; SOUZA G.S.A.; PEREIRA, G.T.; MARQUES Jr. J.; SIQUEIRA, D.S.; SIQUEIRA, G.M. The use of scaled semivariograms to plan soil sampling in sugarcane fields. *Precision Agriculture* v. 13, n. 5, p. 542-552, 2012.
- PAGNANO, N.B. Sistema de medição de fluxo de cana-de-açúcar em colhedoras para a geração de mapas de produtividade. 88p. Tese (Doutorado em Engenharia Agrícola) - Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, 2001.
- PING, J. L., DOBERMANN, A. (2005). Processing of Yield Map Data. *Precision Agriculture*, 6, 193–212.

SPEKKEN, M. ; ANSEMI, A. A. ; MOLIN, J. P. . A simple method for filtering spatial data. In: 9th European conference on precision agriculture, 2013, Lleida. Precision Agriculture`13. Wageningen: Wageningen Academic Publishers, 2013. v. 1. p. 1-823.

THYLÉN, L.; JURSCHIK, P.; MURPHY, D.L.P. Improving the quality of yield data, In: EUROPEAN CONFERENCE ON PRECISION AGRICULTURE, 1., Warwick, 1997. Precision agriculture'97, Warwick: BIOS Scientific Publishers, 1997. v.1, p. 743-750.