



Remoção de erros em mapas de produtividade via filtragem de dados brutos¹



Leonardo A. A. Menegatti² & José P. Molin³

¹ Extraído da Dissertação de Mestrado do primeiro autor apresentada na ESALQ/USP, com apoio financeiro da FAPESP

² ESALQ/USP, Av Pádua Dias 11, CP 09, Piracicaba, SP. E-mail: laameneg@hotmail.com

³ ESALQ/USP. E-mail: jpmolin@esalq.usp.br

Protocolo 18 - 19/2/2002 - Aprovado em 16/12/2003

Resumo: A automação do processo de obtenção do mapa de produtividade introduz erros nos dados coletados. Processos de filtragem de dados podem eliminar parte dos erros e elevar a qualidade da informação. Um processo de filtragem foi desenvolvido para as características de cada erro presente em um mapa de produtividade. Tais erros podem ser de posicionamento, produtividades improvavelmente altas e baixas, largura de plataforma incorreta, umidade de grãos, distância entre pontos, intervalo de enchimento e limites estatísticos para dados discrepantes. O filtro foi aplicado a seis conjuntos de dados obtidos de seis diferentes monitores de produtividade comerciais. 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.

Palavras-chave: agricultura de precisão, monitor de produtividade, algoritmo

Removal of errors in yield maps through raw data filtering

Abstract: As yield data are collected automatically it allows for errors to be introduced into the data collected. Filtering processes might be able to eliminate part of the errors and improve the quality of yield data. The development of a filtering process may be set up based on the individual characteristics of each error. These errors may be related to incorrect position, too high or too low yield, swath width, grain moisture, null distance between points, filling interval and outliers. The filtering algorithm was implemented on six sets of data generated from six different commercial yield monitors. A geostatistical analysis of the data sets showed that the application of the filtering process decreased the nugget effect and improved the cross validation quality index.

Key words: precision agriculture, yield monitor, algorithm

INTRODUÇÃO

Diversas definições têm sido propostas para a agricultura de precisão e todas convergem para o gerenciamento agrícola, baseado na variabilidade espacial dos fatores de produção, visando lucro, sustentabilidade e redução do impacto ambiental. Nesse sistema baseado na variabilidade espacial, a coleta de informações é intensa. Estão inseridas as amostragens de solo, de folhas e os mapas de produtividade, que traduzem a resposta da cultura às práticas de manejo adotadas.

A medição da produtividade é um processo automático, a partir de sensores instalados em colhedoras, capazes de definir com relativa acurácia a quantidade de produto que está sendo colhida e a área onde foi produzida. Um receptor de GPS fornece o posicionamento georreferenciando a produção, para possibilitar seu mapeamento. Por ser automático, o processo permite que grande quantidade de dados seja coletada e como consequência, erros são introduzidos. Moore (1998), trabalhou

com dados de seis anos consecutivos de mapeamento de produtividade, identificando os seguintes erros: suavização, calibração de volume, largura incorreta da plataforma de corte e tempo de enchimento da colhedora. Blackmore & Marshall (1996) citaram ainda os erros de posicionamento, erros devidos à recirculação de grãos no mecanismo de trilha e retrilha da colhedora, perda de grãos pela colhedora e a acurácia do sensor de produtividade, indicando a necessidade do desenvolvimento de rotinas para redução dos erros.

Uma das primeiras propostas de redução de erros em mapas de produtividade foi sugerida por Blackmore & Marshall (1996) e denominada *Potential Mapping*. A técnica desenvolvida propunha que os monitores de produtividade gravassem massa colhida ao invés de massa por área. Os dados seriam então inseridos num sistema de informações geográficas (SIG) para a definição da área que representaria cada ponto, sendo que a produtividade seria derivada no próprio SIG. Esta técnica eliminaria erros de medição de área e de largura de plataforma,

uma vez que a área seria calculada após a colheita. Estes foram considerados, pelos autores, como os principais erros num mapa de produtividade. Posteriormente, o *Potential Mapping* foi abandonado por ter sido identificada significativa quantidade de erros nas regiões das bordas dos talhões (Blackmore & Moore, 1999).

Blackmore & Rands, citados por Blackmore & Moore (1999) desenvolveram um conjunto de regras em linguagem "C" para a filtragem de dados brutos provenientes de um monitor específico de mercado. O programa verificava os arquivos em busca de dados corrompidos e os eliminava do arquivo original. Num arquivo texto eram inseridos os valores máximos e mínimos para que o programa tivesse parâmetros para a remoção de produtividades extremas. A rotina identificava pontos fora dos talhões e os classificava como errôneos e pontos com pequenos erros de posicionamento eram reposicionados, tendo por base a rota da colhedora e a identidade do ponto. Outro fator considerado no desenvolvimento do filtro era a distância máxima entre pontos. A remoção do tempo de enchimento era feita removendo-se os primeiros 12 pontos de cada início de passada; também eram removidos pontos gravados nos arremates da colheita do talhão. Os autores concluíram que a aplicação do filtro poderia elevar significativamente a qualidade do mapa de produtividade.

Juerschik & Giebel (1999) trabalharam com mapas de produtividade obtidos num talhão de 37 ha durante as colheitas de trigo, nos anos de 1995 e 1996, buscando correlações entre os mapas produzidos em diferentes anos, para ratificar as práticas de agricultura de precisão baseadas em mapas de produtividade. Para a interpretação e análise de correlação dos mapas, identificaram a necessidade de filtragem de dados brutos. De modo bastante simples utilizaram, separadamente, três filtros, que classificavam como discrepantes os dados fora dos limites: média $\pm 1,64$ desvios padrão (90% dos dados); média ± 2 desvios padrão (95% dos dados); média ± 3 desvios padrão (99% dos dados). Os autores consideraram que a exclusão de valores fora dos limites foi favorável ao aumento da correlação, mas que, por esse método, a exclusão de pontos reais foi prejudicial e deteriorou a análise de correlação. Uma das conclusões do estudo foi que a utilização do limite média ± 2 desvios padrão foi favorável à análise de correlação quando aplicada aos dados completos, incluindo áreas de manobra e que, se forem analisados dados coletados fora das áreas de manobra e de cabeceiras, o limite de média ± 3 desvios padrão será satisfatório.

Molin & Gimenez (2000) propuseram um algoritmo de filtragem de dados brutos que avaliava cada ponto do arquivo produzido durante a colheita, com relação a um desvio da média e também quanto ao desvio com um ponto coletado 15 posições atrás, buscando-se eliminar pontos coletados durante o tempo de enchimento. Analisaram, ainda, a distância entre pontos consecutivos, sendo eliminados pontos reprovados em qualquer um dos testes. O algoritmo eliminou de 22 a 41% dos pontos de dois mapas de produtividade. Os autores concluíram que a aplicação do algoritmo é simples, mas que cuidados devem ser tomados na definição dos parâmetros a serem utilizados.

Haneklaus et al. (2000) elaboraram recomendações para a produção de mapas de produtividade com boa qualidade de dados, propondo a extração de dados problemáticos como uma das etapas obrigatórias. Consideraram natural a variação da umidade dos grãos de trigo de inverno dentro da faixa de 14 a 28%, sendo eliminados pontos fora desse intervalo. Eles trabalharam, também, com a eliminação de produtividades muito altas e muito baixas, tendo como parâmetros de definição pontos vizinhos destacados nos mapas de produtividade. Para a eliminação dos pontos, os autores selecionaram e retiraram manualmente os pontos destacados; também foram eliminados pontos localizados fora do talhão, considerados erro de localização. De acordo com os autores, o processo é bastante demorado e a aplicação de todas as etapas depende dos custos e da acurácia requerida.

Todos os monitores de produtividade geram pontos que, quando dispostos de maneira ordenada e baseados em suas coordenadas, formam o mapa de produtividade. Para combinar e comparar diferentes informações coletadas sobre a mesma área, todos os dados devem ser interpolados numa grade uniforme (Haneklaus et al., 2000). Em agricultura de precisão utilizam-se, basicamente, os algoritmos de interpolação, denominados inverso da distância a determinada potência e krigagem. De acordo com Isaaks & Srivastava (1989) o método do inverso da distância estipula pesos aos pontos que são inversamente proporcionais à distância elevada a uma potência que, normalmente, é quadrática. Desta forma, o interpolador dá maior importância a valores mais próximos do ponto a ser estimado. Quanto maior o valor da potência, maior o peso de pontos mais próximos e menor a influência de pontos mais longínquos. A krigagem é um método de interpolação mais refinado, que leva em consideração um modelo de dependência espacial para estimar dados em pontos não amostrados.

A dependência espacial é explorada utilizando-se, como ferramenta, o semivariograma. A construção do semivariograma foi discutida por Isaaks & Srivastava (1989) sendo que quatro parâmetros básicos podem ser estabelecidos. O alcance (*range*) é a distância em que a variância da amostragem se torna aleatória, a partir da qual não há mais dependência espacial. Patamar (*sill*) é a semivariância atingida pelas amostras no platô, e pode ser equivalente à variância das amostras. Efeito pepita (*nugget*) é a porção da variância aleatória e não explicada pela amostragem. O quarto parâmetro é o modelo de equação que foi ajustado aos dados no semivariograma. Esses parâmetros podem ser inseridos em um processo de interpolação por krigagem para a produção de superfícies contínuas de produtividade. A validação cruzada é uma técnica que permite a comparação de valores estimados pelo interpolador, baseado no modelo ajustado aos dados na análise de dependência espacial, com valores reais. Os resultados da validação cruzada podem ser avaliados com ferramentas da estatística como, por exemplo, a análise de correlação linear, como índice qualitativo.

A qualidade da informação torna-se fator primordial quando os mapas formam a base de dados que sustentará o processo de tomada de decisão. Os erros introduzidos pela automatização do processo podem levar a conclusões errôneas, apresentando

áreas com produtividades irreais e, sendo assim, este trabalho objetiva desenvolver uma rotina de filtragem de dados brutos, baseada nas características individuais de cada conjunto de dados, que elimine os erros passíveis de identificação inseridos por monitores de produtividade comerciais durante o mapeamento da produtividade.

MATERIAL E MÉTODOS

No processo de filtragem de dados, o material inicial é composto por dados brutos que provêm diretamente do monitor de produtividade, quando os arquivos são gravados em texto, ou do programa específico de cada monitor, quando é gerado em código, e então convertido em formato texto. Os arquivos brutos contêm erros grosseiros de posicionamento, de produtividade nula ou ausente, de interpretação de largura de plataforma, de umidade nula ou valor ausente de umidade, de distância nula entre pontos, de intervalo de enchimento, além de valores discrepantes de produtividade. Para a realização do trabalho utilizou-se um conjunto de dados obtidos com seis diferentes monitores de produtividade comerciais (Tabela 1). A Figura 1 mostra o fluxograma que representa as etapas de aplicação do processo de filtragem.

Tabela 1. Síntese dos dados obtidos com os monitores e utilizados para aplicação do processo de filtragem proposto

Monitores	Cultura	Área (ha)	Número de Pontos Coletados	Densidade de Pontos (pontos ha ⁻¹)
RDS [®] Ceres 2	Milho	22,0	12022	546
RDS [®] Pro Series 8000	Soja	22,0	9955	452
GreenStar [®]	Milho	17,7	29631	1674
FieldStar [®]	Soja	78,0	19309	247
AFS [®]	Soja	42,2	9047	214
New Holland [®]	Trigo	12,6	8356	663

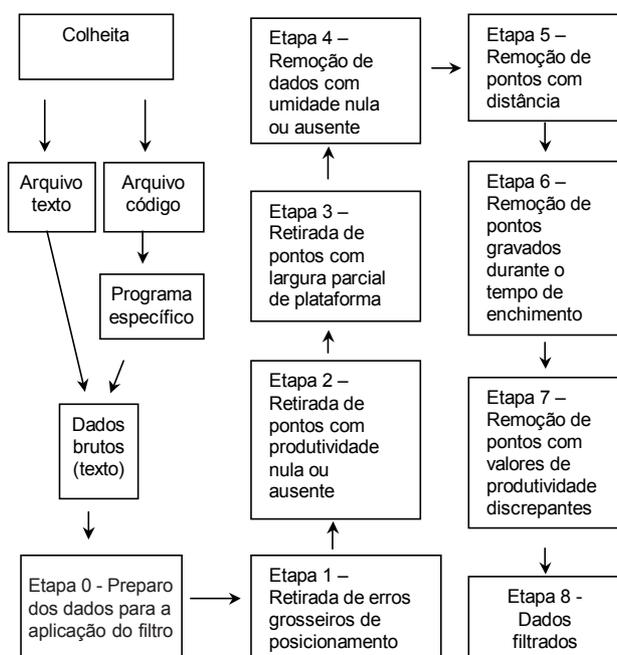


Figura 1. Fluxograma de etapas de aplicação do processo de filtragem de dados brutos

Os dados brutos no formato texto foram abertos em planilhas eletrônicas para a aplicação do filtro. No preparo dos dados, as coordenadas geodésicas foram convertidas em métricas pela projeção UTM 1983. Com base nessas coordenadas, calculou-se a distância entre os pontos, através do Teorema de Pitágoras. Para a determinação do sentido e direção de deslocamento da colhedora, foi calculado o índice de deslocamento. O índice foi calculado nas direções Norte (ID(N)) e Leste (ID(L)) de acordo com as Eq. (1) e (2) e a interpretação dos sinais pode ser vista na Tabela 2. Apenas um dos índices é suficiente para a caracterização do percurso da colhedora, sendo que a escolha depende da direção predominante no talhão; também foi inserida uma coluna contendo um índice de identificação de cada ponto. Ainda nesta etapa, o arquivo foi preparado para a aplicação do algoritmo, para eliminação de pontos registrados durante o intervalo de enchimento, sendo necessário que as duas últimas colunas do arquivo sejam, na ordem, índice de deslocamento (Norte ou Leste) e produtividade.

$$ID(N) = X(i) - X(i + 1) \quad (1)$$

$$ID(L) = Y(j) - Y(j + 1) \quad (2)$$

sendo:

- ID(N) - índice de deslocamento na direção Norte-Sul
- ID(L) - índice de deslocamento na direção Leste-Oeste
- X - coordenada Norte da colhedora
- i - ponto na direção N-S e sentido considerado
- Y - coordenada Leste da colhedora
- j - ponto na direção L-O e sentido considerado

Tabela 2. Interpretação do Índice de deslocamento (ID)

	Valores	Direções	Sentidos
ID(N)	negativo	N-S	Sul → Norte
	positivo	N-S	Norte → Sul
ID(L)	positivo	L-O	Leste → Oeste
	negativo	L-O	Oeste → Leste

Os dados foram então inseridos num sistema de informações geográficas (SIG) para a retirada de erros grosseiros de posicionamento, representados por pontos fora do talhão. Tais pontos foram selecionados e retirados do conjunto de dados (etapa 1). Na segunda etapa, foram selecionados e retirados do conjunto, dados com valores de produtividade nula ou ausente. A terceira etapa consistiu na seleção e retirada de dados coletados com largura de plataforma parcial, selecionando-se todos os pontos com largura de plataforma diferente da largura máxima informada ao sistema, incluindo-se pontos com largura de plataforma nula. Na quarta etapa foram selecionados e retirados pontos com umidade nula e a quinta etapa, realizada no SIG, foi a retirada de pontos com distância nula. Após esta etapa, os dados foram gravados em formato texto e abertos em planilha eletrônica. A sexta etapa compreendeu a caracterização e retirada de pontos gravados durante o intervalo de enchimento da colhedora. Para a caracterização do intervalo de enchimento foram escolhidas cinco passadas aleatórias de colheita. Foi calculada a produtividade média em função da

distância do início da passada, limitada em 50 m. A produtividade máxima ocorrida dentro do intervalo foi verificada e os valores de produtividade foram então convertidos em porcentagem da máxima produtividade observada. O espaço relativo ao intervalo de enchimento foi definido como o espaço necessário para que a produtividade média atinja 90% da produtividade máxima.

Considerando-se todos os pontos do arquivo, calculou-se a distância média entre pontos e o número de pontos a serem eliminados no início de cada passada de colheita; para tanto, foi utilizada a Eq. (3), sendo que o valor obtido foi sempre arredondado para mais, sem casas decimais, em caso de frações.

$$N = \frac{IE}{DM} \quad (3)$$

sendo:

- N - número de pontos a serem eliminados após o início da passada de colheita;
- IE - intervalo de enchimento, m
- DM - distância média entre pontos, m

O número de pontos coletados dentro do intervalo de enchimento (N) foi inserido num algoritmo para identificar e eliminar tal erro. O algoritmo pode ser dividido em duas partes estruturais, para facilitar a compreensão, em que a primeira parte do algoritmo testou os dados para identificar o início de passada. Foi adotada, como critério de estabilidade, a ocorrência de cinco pontos consecutivos com o mesmo sinal de ID, sendo que a alteração do sinal de qualquer ponto dentro deste intervalo resulta em valor falso, interpretado pelo algoritmo como início de passada. Em caso verdadeiro, o algoritmo repete o valor de produtividade e segue para o próximo ponto.

Se o algoritmo identificou início de passada, a primeira providência foi testar se o ponto analisado está dentro do intervalo de enchimento. Para tanto, foi checado o sinal do ID dos N pontos anteriores, onde N é o número de pontos a serem eliminados após o início da passada; se nos N pontos anteriores há mudança de sinal do ID, então o teste retorna falso e altera o valor de produtividade para valor nulo mas, se não há alteração de sinal dentro do intervalo verificado, então o ponto está fora do intervalo de enchimento e o algoritmo retorna verdadeiro, sendo mantido o valor da produtividade. Tal procedimento pode ser visualizado na Figura 2, que representa o fluxograma de atuação do algoritmo. O algoritmo é exemplificado na Eq. (4)

com critério de estabilidade de 5 pontos e eliminação de 4 pontos após o início da passada.

```

SE
  (N(i)<0 E N(i-1)<0 E N(i-2)<0 E N(i-3)<0 E N(i-4)<0)
  OU
  (N(i)>0 E N(i-1)>0 E N(i-2)>0 E N(i-3)>0 E N(i-4)>0)
ENTÃO
  M(i)
CASO CONTRÁRIO
  SE (N(i)>0 E N(i-1)>0 E N(i-2)>0 E N(i-3)>0; N(i-4)>0)
  OU
  (N(i)<0 E N(i-1)<0 E N(i-2)<0 E N(i-3)<0 E N(i-4)<0)
  ENTÃO
    M(i)
  CASO CONTRÁRIO
    0
  FIM
FIM

```

donde:

- N(i) - valores de ID
- M(i) - valores de produtividade

Após a aplicação deste algoritmo, os pontos com produtividade anulada foram eliminados do conjunto de dados. Os dados foram então reordenados com base na coluna de identificação dos pontos, para manter a seqüência original.

A sétima etapa consistiu na análise exploratória dos dados, em busca de dados discrepantes do conjunto. Foram calculadas as medidas de posição mediana, quartil superior, quartil inferior e amplitude interquartilica. Com base na metodologia proposta por Tukey (1977), os limites superior (Eq. 4) e inferior (Eq. 5) foram calculados, procedendo-se à eliminação de pontos cujos valores de produtividade estivessem fora dos limites estabelecidos. Os dados foram então reordenados pela identidade de cada ponto, findando o processo de filtragem.

$$LS = QS + 1,5AI \quad (4)$$

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

sendo:

- LS - limite superior
- LI - limite inferior
- QS - quartil superior

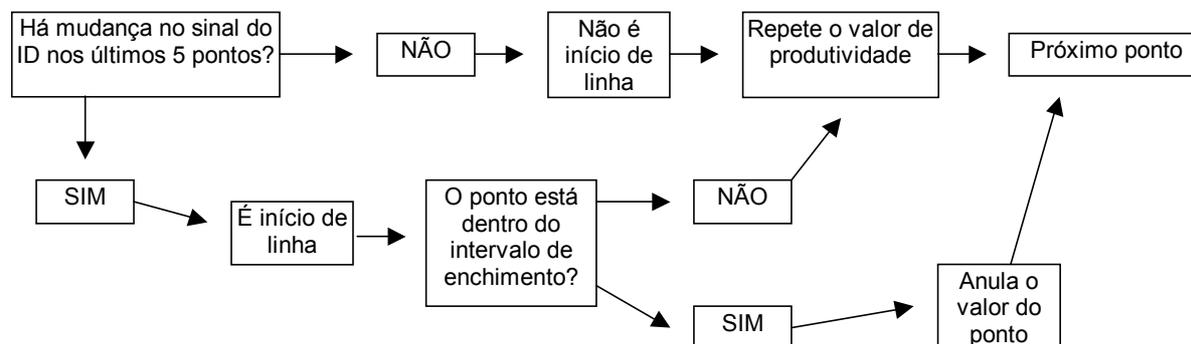


Figura 2. Fluxograma da atuação do algoritmo para eliminação do intervalo de enchimento

- QI - quartil inferior
 AI - amplitude interquartilica

A metodologia de filtragem foi aplicada ao conjunto de dados. Após a realização de cada etapa, realizou-se a contagem dos pontos restantes e totalização daqueles removidos. Foram construídos histogramas de distribuição de frequência de dados brutos e dados filtrados, além da análise de dependência espacial para os dados, antes e depois do processo de filtragem, sendo comparados os parâmetros de ajuste dos modelos com a utilização dos semivariogramas gerados pelo programa GS+ (Geostatistics for the Environmental Sciences®). Por motivos de limitação computacional, foram selecionados cerca de 1000 pontos de cada arquivo, localizados numa faixa diagonal ao talhão, de modo que todas as áreas representativas fossem amostradas. Os modelos ajustados aos dados brutos e filtrados foram comparados por meio da validação cruzada, utilizando-se correlação linear como medida qualitativa do ajuste.

RESULTADOS E DISCUSSÃO

A primeira etapa do processo foi referente à eliminação de pontos com erros grosseiros de posicionamento, causados por falhas no receptor de GPS ou mesmo falha do operador em baixar a plataforma da colhedora fora da área de colheita e com o monitor ligado. Os dados coletados com os monitores AFS®, New Holland®, RDS® Ceres 2, RDS® Pro Series 8000 e FieldStar®, apresentaram tal erro, tendo sido eliminados do conjunto percentagens que variaram de 0,1 a 7,8% dos dados. A segunda etapa do processo de filtragem retirou pontos com produtividade nula dos arquivos utilizados como exemplo. É importante salientar que erros de medida da produtividade ou inacurácia do sensor, não são detectados por esta metodologia, mas apenas pontos em que não houve leitura da produtividade. A mesma lógica pode ser aplicada à eliminação de pontos com umidade nula; o processo não identifica erros de leitura do sensor, mas apenas ausência de valor ou valores nulos. Por outro lado, a eliminação de erros devido à interpretação errônea da largura de plataforma de corte através da eliminação de pontos com largura de corte parcial, refere-se apenas aos pontos assinalados pelo operador como sendo de largura de corte

parcial. O processo não é capaz de identificar pontos que foram gerados com largura parcial de plataforma e que não foram informados ao sistema.

A Tabela 3 apresenta o resumo do processo de filtragem, indicando a retirada de pontos em função de cada etapa aplicada aos dados.

A retirada de pontos pelo processo de filtragem variou de 4,6 a 27,3% para os dados obtidos com os monitores AFS® e GreenStar®, respectivamente. A frequência de coleta de pontos de produtividade foi bastante diferente entre os monitores, resultando em densidades de amostragem diferentes. Os dados obtidos com o monitor GreenStar® apresentaram alta densidade e a retirada de 27,3% dos pontos reduziu esta densidade, de 1676 para 1218 pontos.ha⁻¹, enquanto para o monitor AFS® a densidade baixou de 214 para 205 pontos.ha⁻¹ (Tabela 4).

Tabela 4. Densidade de coleta de pontos antes e depois da aplicação do filtro, para dados coletados com os diferentes monitores

Densidades (pontos ha ⁻¹)	Monitores em que os Arquivos Foram Gerados					
	Green Star®	AFS®	New Holland®	RDS® Ceres 2	RDS® Pro Series 8000	FieldStar®
Iniciais	1676	214	661	546	453	247
Finais	1218	205	554	441	356	232

A alteração do arquivo original pelas diversas etapas do processo de filtragem é função da existência do erro no arquivo de dados brutos e, também, das interações existentes entre os erros apresentados pelos monitores. Em diversas ocasiões, o mesmo ponto apresentava dois ou mais tipos de erro, e a primeira etapa que o identificasse, eliminaria o ponto. Todas as etapas do processo de filtragem foram efetivamente utilizadas nos dados obtidos com os monitores New Holland® e RDS® Ceres 2. Para o monitor GreenStar®, apenas não foi efetiva a sexta etapa, sendo que o efeito do intervalo de enchimento foi retirado pela ação dos filtros anteriores, embora esse tipo de erro tenha sido identificado nos dados brutos. O conjunto de dados obtidos com o monitor AFS® não sofreu alterações com a aplicação das etapas 4, 5 e 6, que representam a retirada de pontos com problemas de leitura de umidade, distância nula entre pontos e de intervalo de enchimento. Durante o processo

Tabela 3. Síntese do processo de filtragem indicando o número e a porcentagem de pontos restantes de cada arquivo de dados brutos

Filtro Aplicado	Monitores em que os Arquivos Foram Gerados											
	GreenStar®		AFS®		New Holland®		RDS® Ceres 2		RDS® Pro Series 8000		FieldStar®	
	Pontos Restantes	(%)	Pontos Restantes	(%)	Pontos Restantes	(%)	Pontos Restantes	(%)	Pontos Restantes	(%)	Pontos Restantes	(%)
Dados originais	29631	100,0	9047	100,0	8356	100,0	12022	100,0	9955	100,0	19309	100,0
Erros grosseiros de posicionamento	29631	100,0	9028	99,8	8033	96,1	11918	99,1	9182	92,2	19299	99,9
Produtividade nula	28958	97,7	8998	99,5	7895	94,5	11883	98,8	9182	92,2	19299	99,9
Interpretação de largura de plataforma	27394	92,5	8891	98,3	7606	91,0	10412	86,6	8858	89,0	19299	99,9
Umidade nula	26201	88,4	8797	97,2	7577	90,7	10377	86,3	8500	85,4	19299	99,9
Distância nula	23408	79,0	8797	97,2	7573	90,6	10374	86,3	8500	85,4	19299	99,9
Intervalo de enchimento	23408	79,0	8797	97,2	7096	84,9	9991	83,1	8099	81,4	18548	96,1
Dados discrepantes	21537	72,7	8635	95,4	7006	83,8	9706	80,7	7830	78,7	18127	93,9
Pontos retirados (%)		27,3		4,6		16,2		19,3		21,3		6,1

de filtragem dos dados obtidos com o monitor RDS® Ceres 2, apenas o filtro de distância nula entre pontos não eliminou dados do conjunto. Para o arquivo obtido com o monitor RDS® Pro Series 8000, as etapas 2 e 5, representadas pelos filtros de produtividade e distância nulas, não modificaram o arquivo. A filtragem dos dados obtidos com o monitor FieldStar® foi efetiva apenas durante a aplicação das etapas 1, 6 e 7, sendo que as demais etapas não eliminaram pontos do arquivo. A inversão da ordem de aplicação dos filtros em qualquer um dos exemplos teria alterado a distribuição do número de pontos eliminados entre as etapas, mas não alteraria a quantidade total de pontos eliminados.

O processo de filtragem dos dados atuou principalmente nos pontos localizados nas extremidades da distribuição de frequência dos mapas de produtividade. Como exemplo, a Figura 3 mostra os histogramas de distribuição de frequência de ocorrência de produtividade para os dados obtidos com o monitor GreenStar® antes e depois do processo de filtragem, durante o qual foram retirados 27,3% dos pontos. A aplicação do filtro aumentou ligeiramente a frequência de ocorrência das

classes em função da eliminação de pontos, mas não alterou as tendências da distribuição, mantendo a alta ocorrência de produtividade nas classes de 6,5 a 10 t ha⁻¹.

Em todos os conjuntos de dados trabalhados, foram eliminados pontos nas extremidades da curva de distribuição, a partir da inflexão demonstrada pela distribuição dos dados brutos. A definição dos limites superior e inferior (Tabela 5) para dados discrepantes permitiu a eliminação de pontos improváveis, sendo que os limites calculados coincidiriam com as inflexões da distribuição de frequência.

O histograma de distribuição de frequência de ocorrência dos dados obtidos com o monitor AFS® pode ser visto na Figura 4 e exemplifica a distribuição de frequência de ocorrência numa situação de ação branda do filtro. Neste caso, apenas 4,6% dos pontos foram retirados (Tabela 3).

Nota-se grande semelhança entre ambas as situações, sendo que a ação do filtro ficou restrita aos pontos com produtividades maior ou menor que os limites de 5,26 e 1,18 t ha⁻¹ (Tabela 5).

A Tabela 6 mostra as características do erro de intervalo de enchimento para os arquivos utilizados para aplicação do filtro.

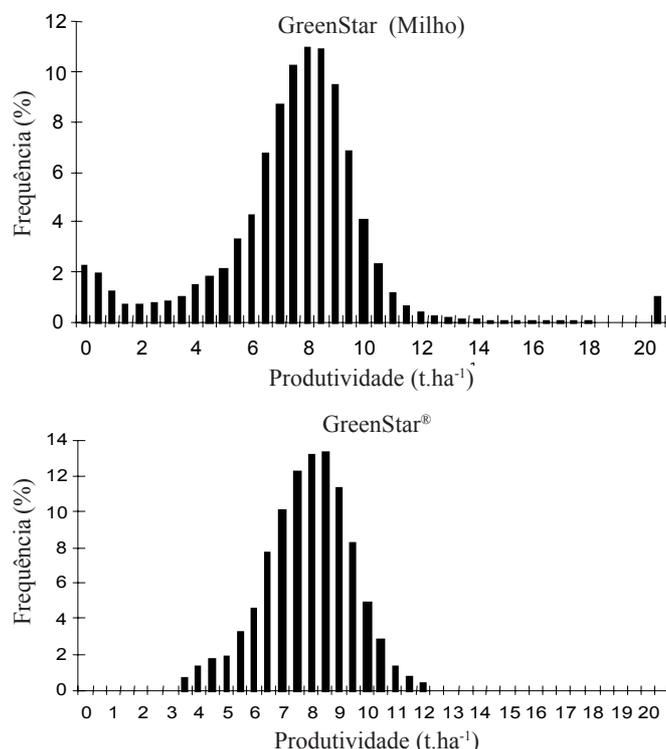


Figura 3. Distribuição de frequência de ocorrência de produtividade para o exemplo dos dados obtidos com o monitor GreenStar®, antes (A) e depois (B) do processo de filtragem

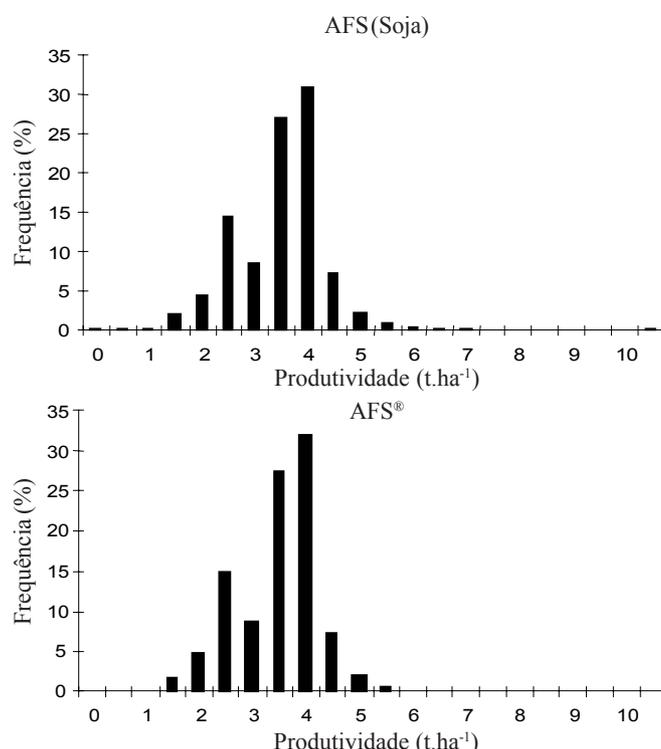


Figura 4. Distribuição de frequência de ocorrência de produtividade para os dados obtidos com o monitor AFS® antes (A) e depois do processo de filtragem (B)

Tabela 5. Limites estatísticos calculados para os dados durante o processo de filtragem

Fator	Monitor em que os Arquivos Foram Gerados					
	GreenStar®	AFS®	New Holland®	RDS® Ceres 2	RDS® pro Series 8000	FieldStar®
	Produtividade (t ha ⁻¹)					
Quartil Superior	8,68	3,73	4,53	5,04	5,34	3,36
Quartil Inferior	6,46	2,71	3,37	4,35	3,54	2,74
Amplitude interquartilica	2,22	1,02	1,15	0,69	1,80	0,62
Limite Superior	12,02	5,26	6,26	6,08	0,84	1,82
Limite Inferior	3,13	1,18	1,64	3,32	8,03	4,28

Tabela 6. Parâmetros levantados para caracterização e eliminação do erro de intervalo de enchimento para os arquivos analisados

	Monitor em que os Arquivos Foram Gerados						
	Green Star®	AFS®	New Holland®	RDS® Ceres 2	RDS® Pro Series 8000	Field Star®	
Intervalo de enchimento (m)	0	0	14	20	22	25	
Distância média entre pontos (m)	0	0	4	5	5,67	5,45	
N ¹ (Eq. 3)	0	0	4	4	4	5	

¹ Número de pontos a serem removidos no início de cada passada de colheita

Os dados obtidos com os monitores GreenStar® e AFS®, não apresentaram intervalo de enchimento, enquanto para os demais monitores o intervalo de enchimento variou de 14 a 25 m após o início da colheita. A eliminação de pontos coletados durante o intervalo de enchimento pode ser vista na Figura 5, onde a ação do algoritmo é exemplificada com o mapa de produtividade produzido com dados do monitor FieldStar®.

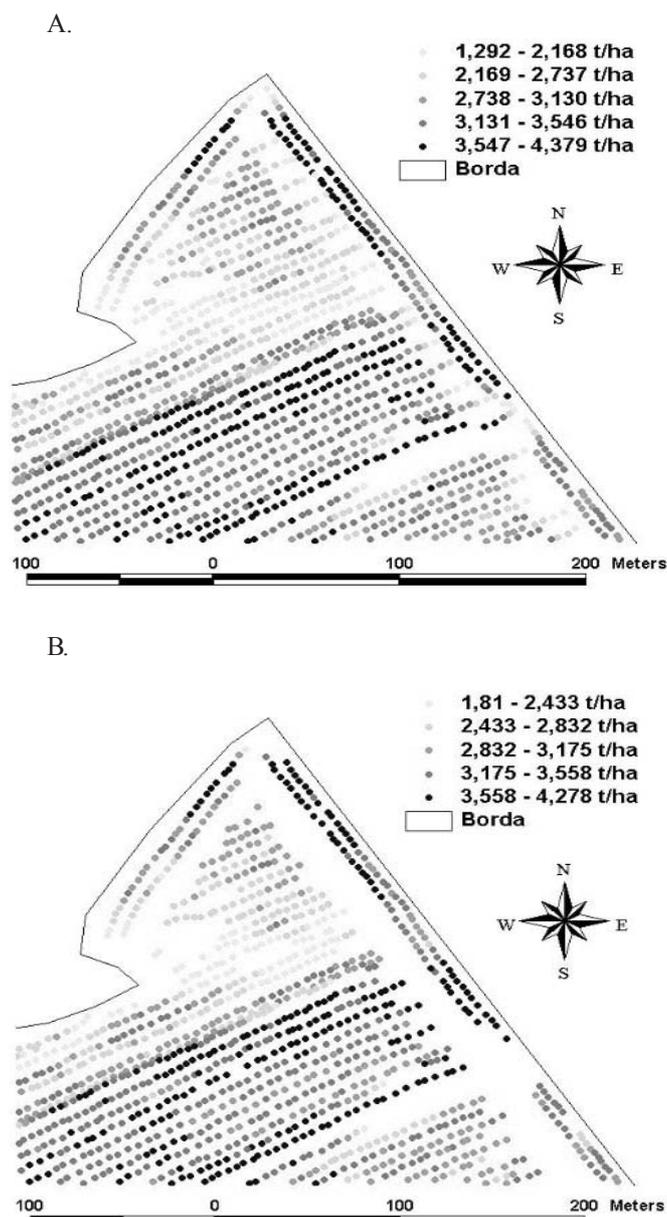


Figura 5. Resultados da aplicação da sexta etapa do processo de filtragem (B) que corresponde à eliminação do erro devido ao intervalo de enchimento, em comparação com o mapa produzido a partir de dados brutos (A)

O algoritmo atuou fortemente nas zonas de cabeceira dos talhões, eliminando pontos coletados durante o intervalo de enchimento. As passadas que tiveram pontos eliminados (Figura 5) correspondem às passadas em que a colhedora iniciou a colheita na área mostrada. Foram eliminados os cinco primeiros pontos de cada início delas, resultando na diminuição da quantidade de pontos cinza claro nas cabeceiras. A ausência de outros pontos no arquivo ocorre devido às ações das etapas anteriores à remoção do intervalo de enchimento.

A Figura 6 mostra um exemplo da seleção de dados utilizados na análise de dependência espacial. Os coeficientes ajustados aos semivariogramas produzidos com dados brutos e filtrados podem ser vistos na Tabela 7. Como exemplo, os semivariogramas

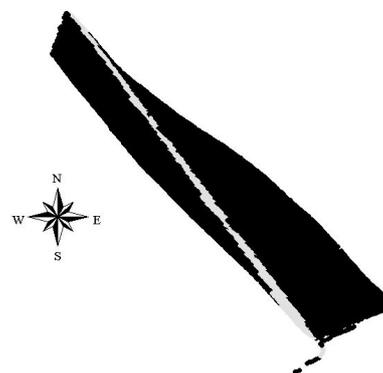


Figura 6. Exemplo de seleção de pontos para a realização da análise de dependência espacial; os pontos selecionados no arquivo obtido com o monitor RDS® Ceres 2, aparecem em cinza

Tabela 7. Parâmetros de ajuste do semivariograma para os dados brutos e filtrados

Monitor	Arquivo	Modelo*	Parâmetros			
			Pepita γ^{**}	Patamar γ	Alcance (m)	Validação Cruzada
AFS®	Bruto	Gau	0,32	1,32	550,00	0,56
	Filtrado	Gau	0,14	1,00	500,00	0,68
New Holland®	Bruto	Gau	1,20	3,20	850,00	0,38
	Filtrado	Esf	0,27	0,70	17,36	0,58
RDS® Ceres 2	Bruto	exp	16,04E ⁵	32,09E ⁵	171,00	0,47
	Filtrado	Exp	0,09	0,17	64,50	0,62
RDS® Pro Series 8000	Bruto	Pep	-	-	-	-
	Filtrado	Pep	-	-	-	-
Field Star®	Bruto	Exp	0,02	0,17	117,43	0,85
	Filtrado	Exp	0,00	0,09	28,05	0,83
Green Star®	Bruto	Esf	1,45	3,95	18,87	0,71
	Filtrado	Exp	0,58	2,02	42,90	0,75

* Exp – exponencial; Gau – gaussiano; Esf – esférico; Pep – pepita
 ** γ semivariância

gramas ajustados aos dados brutos e filtrados do monitor AFS®, podem ser vistos na Figura 7.

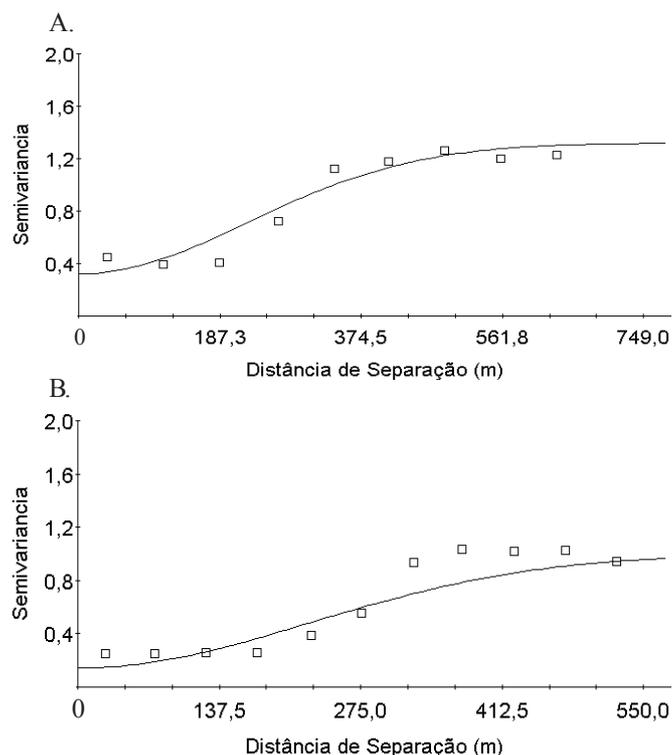


Figura 7. Semivariogramas ajustados aos dados coletados com o monitor de produtividade AFS®, antes (A) e depois do processo de filtragem (B)

O processo de filtragem contribuiu para a caracterização da dependência espacial, reduzindo a variabilidade não explicada pelos modelos ajustados aos dados; apenas os dados obtidos com o monitor RDS® Pro Series 8000 não apresentaram dependência espacial. Na Figura 7, observa-se a redução do efeito pepita de 0,32 para 0,14 γ , devido à atuação do processo de filtragem (Tabela 7). Especialmente com relação aos dados obtidos com o monitor RDS® Ceres 2, o processo de filtragem foi bastante benéfico, reduzindo o efeito pepita de 1604000 para 0,09 γ ; tamanha redução se dá pela presença de valores 9999,9 t ha⁻¹ no arquivo de dados brutos de produtividade. A presença desses valores, normalmente nas cabeceiras eleva, de forma considerável, a semivariância. O processo de filtragem melhorou a validação cruzada dos modelos ajustados, que representa a correlação entre valores reais e preditos pelo modelo, aos dados obtidos com os monitores RDS® Ceres 2, AFS®, New Holland® e GreenStar®, e reduziu de 0,85 para 0,83 para os dados obtidos com o monitor FieldStar®, sendo esta a menor alteração devido à aplicação do filtro.

Considerando-se que alguns erros são função da dinâmica da relação entre máquina, monitor e cultura, espera-se que as características de cada tipo de erro se mantenham constantes para uma mesma máquina, com um mesmo monitor trabalhando na mesma cultura, facilitando o trabalho de filtragem dos dados. Evidentemente, o processo de filtragem não elimina a necessidade de correta calibração do sensor de produtividade e das rotinas de pré-colheita que favoreçam a boa qualidade da coleta de dados.

A rotina de filtragem proposta foi baseada no princípio da remoção de pontos problemáticos cujos impasses foram previamente caracterizados. Outras soluções poderiam ser aplicadas para aumentar a qualidade final dos mapas de produtividade. Ao se lidar com pontos com problemas na leitura da umidade, ao invés de classificá-los como problemáticos e eliminá-los do conjunto, poder-se-ia utilizar os pontos vizinhos, elaborar uma média e atribuir a média dos vizinhos ao ponto cuja leitura era inválida; assim, evitar-se-ia a exclusão de pontos que, possivelmente, possuem informações úteis. O mesmo tipo de solução poderia ser aplicado a pontos com produtividade nula, buscando-se aproveitar as demais informações contidas na linha referente a cada ponto, sem que a sua presença no mapa causasse problemas de modificação das tendências.

A calibração incorreta do tempo de espera gera deslocamento dos pontos coletados na direção em que a colhedora se desloca e o sentido depende do valor do tempo de espera; se superdimensionado, o deslocamento ocorre no sentido do movimento e, se sub-dimensionado, no sentido contrário. Num processo mais elaborado, poder-se-ia calcular o deslocamento dos pontos a partir de medições na cabeceira dos talhões e, com esta informação associada ao índice de deslocamento e equações trigonométricas, todos os pontos dos mapas poderiam ser realocados em novo posicionamento, que deveria ser mais próximo do local onde os pontos foram realmente coletados. Pontos com pequenos erros de posicionamento também poderiam ter sua posição alterada para sua alocação mais provável. Desvios de trajeto de colheita e sobreposições de passadas poderiam ser alterados, tendo-se por base uma linha imaginária que simularia o percurso da colhedora.

O processo de filtragem desenvolvido é de fácil aplicação aos conjuntos de dados e o tempo de aplicação não sofre influência do número de pontos a serem trabalhados, sendo que o aumento da área de colheita e conseqüente aumento do número de pontos coletados, aumentariam o esforço computacional. Em virtude da tecnologia atualmente disponível aos usuários comuns e considerando-se a capacidade dos processadores utilizados em computadores domésticos, o esforço computacional durante a aplicação do processo é mínimo. Tratando-se de uma rotina integralmente aplicada em computadores, a criação de um programa baseado nas diretrizes do processo reduziria significativamente o tempo gasto na aplicação do filtro em cada arquivo que, atualmente, é de aproximadamente 40 min. A automação da caracterização dos erros e sua retirada baseada nas características de cada erro, podem ser consideradas como a evolução natural do processo.

CONCLUSÕES

1. O processo de filtragem desenvolvido é eficiente na eliminação de pontos problemáticos em pós-processamento de dados.
2. A eliminação de pontos errôneos favorece a descrição da dependência espacial e melhora o ajuste dos modelos aos dados.
3. A intensidade da remoção de pontos é dependente do equipamento utilizado na obtenção dos dados.

4. Outras soluções poderiam ser aplicadas aos dados provenientes de monitores de produtividade, visando não eliminar os pontos, mas a correção de erros e aproveitamento das demais informações contidas nas linhas que compõem o arquivo.

AGRADECIMENTOS

Os autores agradecem à FAPESP pelo apoio financeiro, e à Fundação ABC, AGCO, Case, John Deere e New Holland, pela colaboração na obtenção dos dados.

LITERATURA CITADA

- Blackmore, B.S.; Marshall, C.J. Yield mapping; errors and algorithms. In: International Conference on Precision Agriculture, 3, Minneapolis. Proceedings... Madison, WI: ASAE, 1996, p.403-415.
- Blackmore, S.; Moore, M. Remedial correction of yield map data. Precision Agriculture, Dordrecht, v.1, n.1, p.51-66, 1999.
- Haneklaus, S.; Lilienthal, H.; Schnug, E.; Panten, K.; Haverech, E. Routines for efficient yield mapping. In: International Conference on Precision Agriculture, 5, Madison. Proceedings... Minneapolis: ASAE, 2000. CD-Rom
- Isaaks, E.H.; Srivastava; R.M. Applied geostatistics. 1. ed. Oxford: Oxford University Press, 1989. 561p.
- Juerschik, P.; Giebel, A. Processing of point data from combine harvesters for precision farming In: European Conference on Precision Agriculture, n.2, 1999. Precision Agriculture'99, v.1, Odense, 1999, p. 297-307.
- Molin, J.P.; Gimenez, L.M. Desenvolvimento de um algoritmo para redução de erros em mapas de rendimento obtidos em agricultura de precisão. In: Congresso e Mostra de Agroinformática, 1, Ponta Grossa, Paraná, 2000. InfoAgro 2000. Ponta Grossa, 2000. CD-Rom
- Moore, M. An investigation into the accuracy of yield maps and their subsequent use in crop management, Cranfield: Cranfield University, 1998. p.379. PhD Thesis
- Tukey, J.W. Exploratory data analysis. 1 ed. Reading, Massachusetts, v.1, n.3, 1997.