

REMOÇÃO DE ERROS EM MAPAS DE PRODUTIVIDADE VIA FILTRAGEM DE DADOS BRUTOS¹

Escrito para apresentação no
XXXI Congresso Brasileiro de Engenharia Agrícola – CONBEA 2002
Salvador – Bahia, 31 de julho a 03 de agosto de 2002

José P. Molin²; Leonardo A. A. Menegatti³

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. O processo de filtragem desenvolvido trabalha com as características de cada erro presente no mapa de produtividade, apresentados em artigo anterior, que podem ser erros 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

YIELD MAP ERRORS REMOTION BY RAW DATA FILTERING

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

² Eng Agrícola, Prof. Dr. Depto. de Eng. Rural, ESALQ/USP. Av Pádua Dias, 11, CP09, Piracicaba, SP e-mail: jpmolin@esalq.usp.br

³ Eng. Agrônomo MSc., Depto de Eng. Rural, ESALQ/USP, e-mail: laameneg@hotmail.com

SUMMARY: As yield data is collected automatically it allows for errors to be gathered together with good data. Filtering processes might be able to eliminate part of the errors and improve the quality of yield data. The develop of a filtering process may be set up based on the individual characteristic of each error, as presented in a previous paper. Those errors may be related to wrong position, too high or too low yield, swath width error, grain moisture error, null distance between points, filling time error and outliers. The filtering algorithm was implemented on six fields with 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's quality index.

KEYWORDS: precision farming; yield monitor; algorithm

INTRODUÇÃO

A medição da produtividade é um processo automático, a partir de sensores instalados em colhedoras, que são 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 georeferenciando a produção para possibilitar seu mapeamento. Por ser automático, o processo permite que grande quantidade de dados seja coletada, porém, permitindo que erros sejam introduzidos. Desde o início da obtenção de mapas de produtividade diversos pesquisadores têm reportado a existência de erros nos mapas. MOORE (1998), trabalhou com dados de seis anos consecutivos de mapeamento de produtividade, identificou 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 devido à recirculação de grãos no mecanismo de trilha

e retilha da colhedora, perdas 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 de *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).

JUERSCHIK & GIEBEL (1999) trabalharam com mapas de produtividade obtidos num talhão de 37 hectares 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 se 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 era 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 com relação ao desvio com um ponto coletado 15 posições atrás, buscando eliminar pontos coletados durante o tempo de enchimento. Analisaram também a distância entre pontos consecutivos, sendo eliminados pontos reprovados em qualquer 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.

Este trabalho é a continuação de trabalho anterior que descreveu e caracterizou erros possíveis de serem identificados em mapas de produtividade após a produção do mapa e 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 foi utilizado 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 exemplo obtidos com os monitores utilizados para aplicação do processo de filtragem proposto.

Monitores	Culturas	Áreas(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

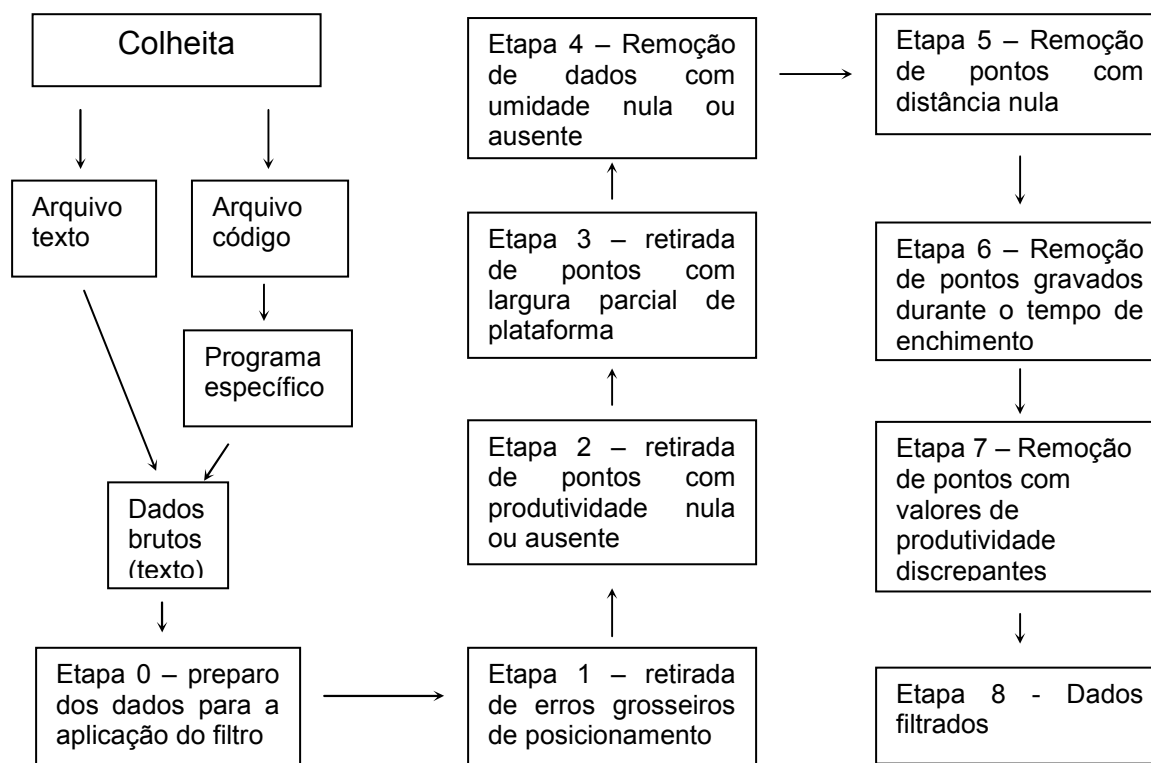


Figura 1 – Fluxograma de etapas de aplicação do processo de filtragem de dados brutos.

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, foi calculada 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 equações (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 é preparado para a aplicação do algoritmo para eliminação do intervalo de enchimento, sendo necessário que as duas últimas colunas do arquivo sejam, na ordem, índice de deslocamento (Norte ou Leste) e produtividade.

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

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

em que,

ID(N) é o índice de deslocamento na direção Norte-Sul;

ID(L) é o índice de deslocamento na direção Leste-Oeste;

X é a coordenada Norte da colhedora;

i é o ponto na direção N-S e sentido considerado;

Y é a coordenada Leste da colhedora;

j é o 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, representado 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 consiste 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. A quinta etapa, realizada no SIG, é a retirada de pontos com distância nula entre si. Após esta etapa, os dados foram gravados em formato texto e abertos em planilha eletrônica.

A sexta etapa compreende 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 percentagem 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, foi calculada 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 equação (3), sendo que o valor obtido foi sempre arredondado para cima sem casas decimais, em caso de frações.

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

em que:

N é o número de pontos a ser eliminado após o início da passada de colheita;

IE é o intervalo de enchimento, em metros;

DM é a distância média entre pontos.

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. A primeira parte do algoritmo testa os dados para identificar início de passada. Foi adotado, 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 é testar se o ponto analisado está dentro do intervalo de enchimento. Para tanto é 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. 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 equação (4), com critério de estabilidade de 5 pontos e eliminação de 4 pontos após o início da passada.

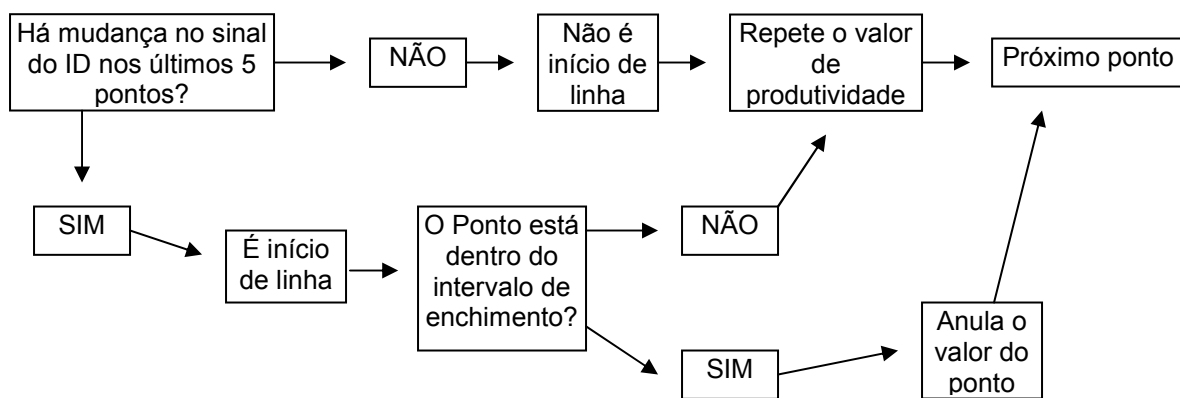


Figura 2 – Fluxograma da atuação do algoritmo para eliminação do intervalo de enchimento.

SE

$(N(i) < 0 \text{ E } N(i-1) < 0 \text{ E } N(i-2) < 0 \text{ E } N(i-3) < 0 \text{ E } N(i-4) < 0)$

OU

$(N(i) > 0 \text{ E } N(i-1) > 0 \text{ E } N(i-2) > 0 \text{ E } N(i-3) > 0 \text{ E } N(i-4) > 0)$

(4)

ENTÃO

$M(i)$

CASO CONTRÁRIO

SE $(N(i) > 0 \text{ E } N(i-1) > 0 \text{ E } N(i-2) > 0 \text{ E } N(i-3) > 0; N(i-4) > 0)$

OU

$(N(i) < 0 \text{ E } N(i-1) < 0 \text{ E } N(i-2) < 0 \text{ E } N(i-3) < 0 \text{ E } N(i-4) < 0)$

ENTÃO

$M(i)$

CASO CONTRÁRIO

0

FIM

FIM

em que:

$N(i)$ são os valores de ID e

$M(i)$ são valores de produtividade.

Após a aplicação deste algoritmo, os valores 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-se 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 interquartílica. Com base na metodologia proposta por Tukey (1977), os limites superior (equação 5) e inferior (equação 6) foram calculados, procedendo-se a eliminação de pontos cujos valores de produtividade estão fora dos limites estabelecidos. Os dados foram então reordenados pela identidade de cada ponto, findando o processo de filtragem.

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

$$LI = QI - 1,5.AI \quad (6)$$

em que:

LS é o limite superior;

LI é o limite inferior;

QS é o quartil superior;

QI é o quartil inferior e

AI é amplitude interquartílica.

A metodologia de filtragem foi aplicada ao conjunto de dados exemplo. Após a realização de cada etapa, foi realizada a contagem dos pontos restantes e totalização daqueles retirados do mapa. Foram construídos histogramas de distribuição de freqüência de dados brutos e dados filtrados. Também foi feita análise de dependência espacial para os dados, antes e depois do processo de filtragem, sendo que os parâmetros de ajuste dos modelos foram comparados. 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. 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 essa metodologia, 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, 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 parcial de corte, elimina apenas os pontos assinalados pelo operador com largura 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.

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

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 essa densidade de 1676 para 1218 pontos.ha⁻¹, enquanto que para o monitor AFS® a densidade abaixou de 214 para 205 pontos.ha⁻¹ (Tabela 4).

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

Densidades (pontos.ha ⁻¹)	Monitores em que os arquivos foram gerados					
	GreenStar [®]	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 erros, 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 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 durante a aplicação das etapas 1, 6 e 7 apenas, sendo que as demais etapas não eliminaram pontos do arquivo. A inversão da ordem de aplicação dos filtros em qualquer 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.

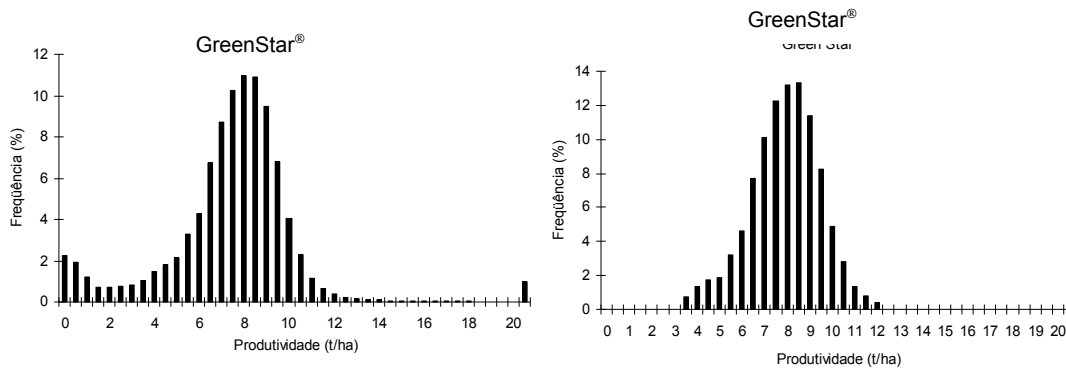


Figura 3 – Distribuição de frequência de ocorrência de produtividade para o exemplo dos dados obtidos com o monitor GreenStar[®] antes (esquerda) e depois (direita) do processo de filtragem.

Em todos os conjuntos de dados exemplos 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.

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

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).

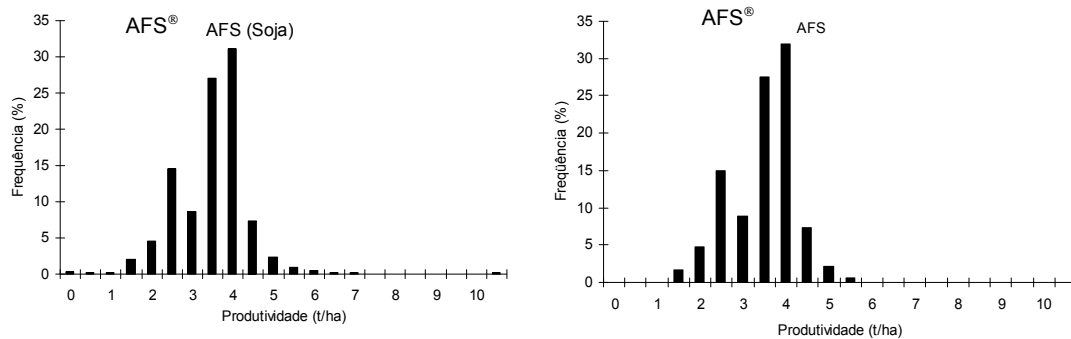


Figura 4 – Distribuição de frequência de ocorrência de produtividade para os dados obtidos com o monitor AFS® antes (esquerda) e depois do processo de filtragem (direita).

Pode-se notar grande semelhança entre ambas situações, sendo que a ação do filtro ficou restrita aos pontos com produtividade maiores ou menores 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. Os dados obtidos com os monitores GreenStar® e AFS® não apresentaram intervalo de enchimento, enquanto que 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 4, onde a ação do algoritmo é exemplificada com o mapa de produtividade produzido com dados do monitor FieldStar®.

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

	Monitor em que os arquivos foram gerados					
	GreenStar®	AFS®	New Holland®	RDS® Ceres 2	RDS® Pro Series 8000	FieldStar®
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

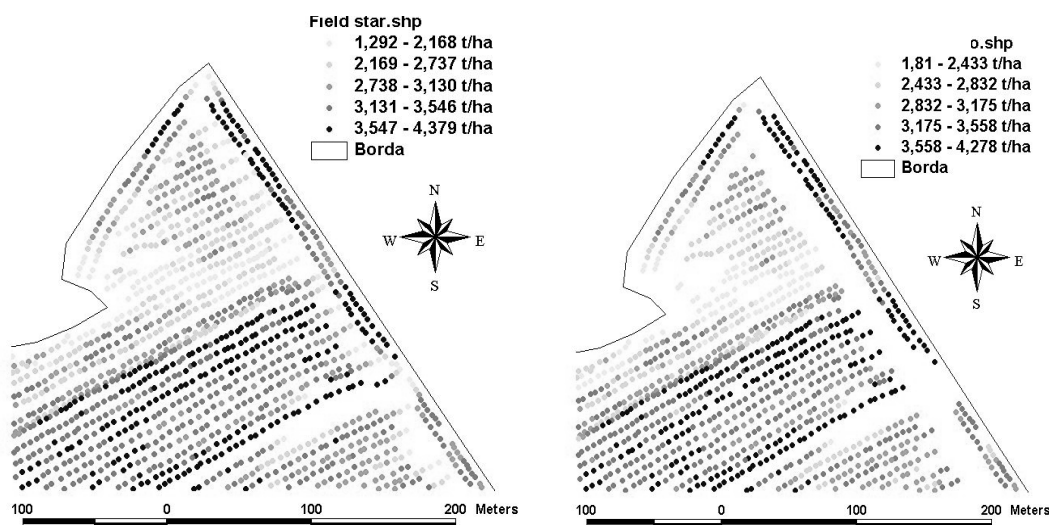


Figura 5 – Resultados da aplicação da sexta etapa do processo de filtragem (direita) que corresponde à

eliminação do erro devido ao intervalo de enchimento, em comparação com o mapa produzido a partir de dados brutos (esquerda).

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 5 primeiros pontos de cada início delas, resultando na diminuição da quantidade de pontos cinza claro nas cabeceiras. A ausência de outros pontos no arquivo é 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. A título de exemplo, os semivariogramas ajustados aos dados brutos e filtrados do monitor AFS[®] podem ser vistos na Figura 7.

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

Monitor	Arquivo	Modelo	Parâmetros			Validação cruzada
			Pepita γ^*	Patamar γ	Alcance (m)	
AFS [®]	Bruto	Gaussiano	0,32	1,32	550,00	0,56
	Filtrado	Gaussiano	0,14	1,00	500,00	0,68
New Holland [®]	Bruto	Gaussiano	1,20	3,20	850,00	0,38
	Filtrado	Esférico	0,27	0,70	17,36	0,58
RDS [®] Ceres 2	Bruto	exponencial	1604000,00	3209000,00	171,00	0,47
	Filtrado	exponencial	0,09	0,17	64,50	0,62
RDS [®] Pro Series 8000	Bruto	Pepita	-	-	-	-
	Filtrado	Pepita	-	-	-	-
FieldStar [®]	Bruto	exponencial	0,02	0,17	117,43	0,85
	Filtrado	exponencial	0,00	0,09	28,05	0,83
GreenStar [®]	Bruto	Esférico	1,45	3,95	18,87	0,71
	Filtrado	exponencial	0,58	2,02	42,90	0,75

γ é a semivariância

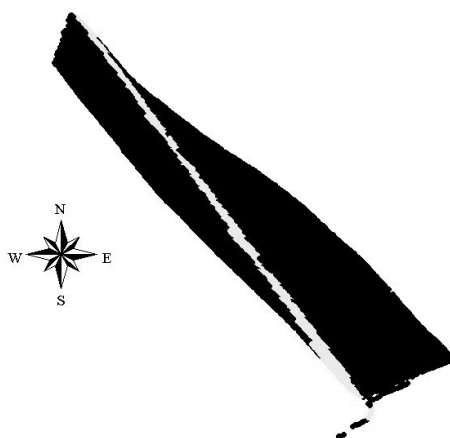


Figura 6 – Exemplo de seleção de pontos para a realização da análise de dependência espacial; os pontos selecionados no arquivo obtido com o monitor RDS[®] Ceres 2 aparecem em cinza.

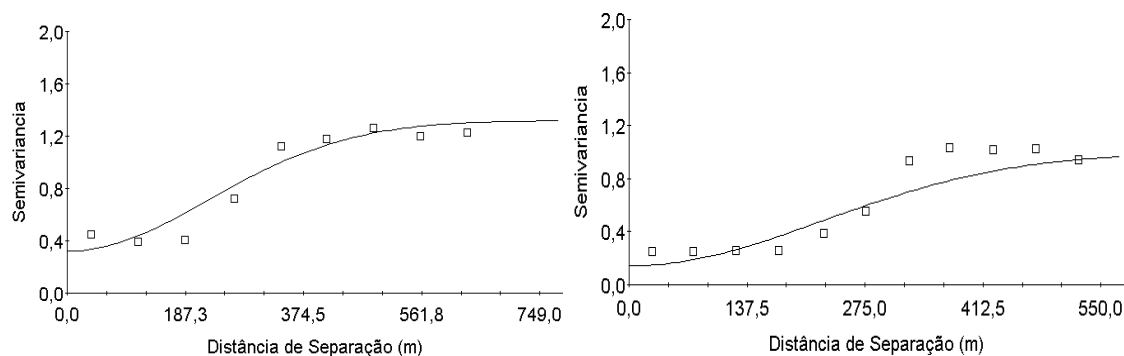


Figura 7 – Semivariogramas ajustados aos dados coletados com o monitor de produtividade AFS[®] antes (esquerda) e depois do processo de filtragem (direita).

De maneira geral, 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 pode-se observar a redução do efeito pepita de 0,32 para 0,14 Υ 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. Tamanho

redução é devido à presença de valores 9999,9 t.ha⁻¹ no arquivo de dados brutos de produtividade. A presença destes valores, normalmente nas cabeceiras, eleva consideravelmente 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 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 problemas foram previamente caracterizados. Outras soluções poderiam ser aplicadas para aumentar a qualidade final dos mapas de produtividade. Ao lidar-se com pontos com problemas na leitura da umidade, ao invés de classifica-los como problemáticos e elimina-los do conjunto, poderia utilizar-se os pontos vizinhos, elaborar uma média e atribuir a média dos vizinhos ao ponto cuja leitura era inválida. Assim, evitaria-se a exclusão de pontos que possivelmente possuem informações valiosas. O mesmo tipo de solução poderia ser aplicado a pontos com produtividade nula, buscando-se aproveitar as outras informações contidas na linha que compõe o arquivo sem que a presença do ponto cause problemas de modificação das tendências.

A calibração incorreta do tempo de espera gera um deslocamento dos pontos coletados na direção do deslocamento e o sentido depende do valor do tempo de espera. Se superdimensionado, o deslocamento ocorre no sentido do movimento e se sub-dimensionado, o deslocamento ocorre em 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, sobreposição de passadas poderiam ser alteradas tendo 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 aumentaria o esforço computacional. Em virtude da tecnologia atualmente disponível aos usuários comuns, levando em consideração a capacidade dos processadores utilizados em computadores domésticos, o esforço computacional é irrisório durante a aplicação do processo. Tratando-se de uma rotina integralmente aplicada em computadores, a criação de um programa baseado nas diretrizes do processo reduziria grandemente o tempo gasto na aplicação do filtro em cada arquivo que, atualmente é de aproximadamente 40 minutos. A automação da caracterização dos erros e sua retirada baseada nas características de cada erro pode ser considerado como a evolução natural do processo.

CONCLUSÃO

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

Outras soluções poderiam ser aplicadas aos dados provenientes de monitores de produtividade, visando não a eliminação de pontos, mas a correção de erros e aproveitamento das demais informações contidas nas linhas que compõe 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.

REFERÊNCIAS BIBLIOGRÁFICAS

BLACKMORE, B.S.; MARSHALL, C.J. Yield mapping; errors and algorithms. In: International Conference On Precision Agriculture, n.3, Minneapolis,1996. Proceedings of the 3rd International Conference of Precision Agriculture, Madison, WI, 1996, p.403-415.

BLACKMORE, S.; MOORE, M. Remedial correction of yield map data. *Precision agriculture*, v.1, n.1,p. 51-66, 1999.

JUERSCHIK, P.; GIEBEL, A. Processing of point data from combine harvesters for precision farming In: European Conference On Precision Agriculture, n.2, 1997. Precision Agriculture'99, v.1, Denmark, 1999, p. 297-307.

MOLIN, J.P.; GIMENEZ, L.M. Desenvolvimento e um algoritmo para redução de erros em mapas de rendimento obtidos em agricultura de precisão. In: Congresso e Mostra de Agroinformática, n.1, Ponta Grossa, Paraná, 2000. InfoAgro 2000. Ponta Grossa, CDROM, 2000.

MOORE, M. *An investigation into the accuracy of yield maps and their subsequent use in crop management*. 1994. 379 f. Thesys (Ph.D) - Silsoe College, Inglaterra, 1998.