

ALGORITMO PARA REDUÇÃO DE ERROS EM MAPAS DE PRODUTIVIDADE PARA AGRICULTURA DE PRECISÃO¹

Leandro M. Gimenez², José Paulo Molin³

Recebido para publicação em 22/03/2002

Aprovado para publicação em 25/03/2004

RESUMO: *Os mapas de produtividade gerados durante a colheita são constituídos por um grande número de pontos georreferenciados. Nem sempre estes pontos refletem a produtividade de forma correta, em função de erros que ocorrem durante o processo de coleta dos dados. A quantidade e os tipos de erros variam em função de uma série de fatores como o tipo de monitor de produtividade, máquina colhedora, tamanho e formato dos talhões, qualidade do sinal de GPS e outros. A ocorrência de erros nos dados utilizados na geração dos mapas leva a interpretações errôneas da variabilidade presente no campo, reduzindo a qualidade do mapa gerado. Neste trabalho foram propostas e utilizadas funções de lógica na composição de um algoritmo para realizar a eliminação dos erros presentes nos dados coletados, detectando e eliminando os erros mais importantes e mais freqüentes. O algoritmo foi testado em dados gerados em diferentes áreas, sendo apresentados resultados de duas dessas áreas onde é demonstrada substancial redução no volume de erros e aumento da qualidade dos mapas de produtividade.*

Palavras-chave: *monitor de produtividade, filtro, GPS*

ALGORITHM FOR REMOVING ERRORS ON YIELD MAPS DATA FOR PRECISION AGRICULTURE

ABSTRACT: *Yield maps generated during the harvesting operation are made from a great number of georeferenced points. Each point shows the yield of a limited area around it and they all may be interpolated to generate a representation of surface. Not always the points represent the correct yield because of errors during data collecting process. The amount and type of errors depend on yield monitor and combine model, size and shape of fields, GPS signal quality and others. Errors on yield mapping data will cause missinterpretation on field variability, reducing map quality. Errors may be eliminated by using spreadsheets and in this work were proposed and used logic functions for an algorithm that detect and eliminate most frequent and important errors on raw data. The algorithm was tested with data from several fields and here is presented the example of two of those fields showing the improvement of yield maps quality.*

Keywords: *yield monitor, filtering, GPS*

¹ Trabalho desenvolvido com dados obtidos em projeto financiado pela FAPESP

² Eng^o Agrônomo, Pesquisador, Fundação ABC, C.P. 1003Rod. PR 151 – km 155,5, 84166-990, Castro, PR. E-mail: mecaniza@fundacaoabc.com.br

³ Eng^o Agrícola, Professor Doutor, ESALQ/USP, Bolsista CNPq, Av. Pádua Dias, 11, 13418-900, Piracicaba, SP. E-mail: jpmolin@esalq.usp.br

1 INTRODUÇÃO

A geração de mapas de produtividade ou rendimento já é uma prática bastante comum em países mais desenvolvidos, onde o monitoramento da produtividade é considerado o primeiro passo na implantação da Agricultura de Precisão. No Brasil, em algumas áreas onde a tecnologia vem sendo estudada, já se tem mapas de produtividade há alguns anos.

Os mapas gerados durante a colheita, pela própria colhedora equipada com sensores e um sistema para posicionamento em campo (GPS - Global Positioning System), são constituídos de um grande número de pontos onde se tem a produtividade georreferenciada. Nem todos pontos refletem a produtividade adequadamente, sendo comum a ocorrência de erros na coleta dos dados (Molin, 2000).

Alguns desses erros são eliminados por programas computacionais geradores de mapas. No entanto, parte dos erros é de identificação e caracterização relativamente complexa. Moore (1998), trabalhando com dados de seis anos consecutivos de mapeamento de produtividade, identificou os seguintes erros: suavização, calibração de volume, largura incorreta da plataforma e tempo de enchimento da colhedora. Blackmore & Marshall (1996) listaram várias fontes de erros em mapas de colheita. Incluíram erros oriundos da retrilha de grãos e das perdas de colheita. Ensaios conduzidos por Larscheid et al. (1997) mostraram que quando se utiliza sensores volumétricos, a densidade dos grãos deve ser aferida pelo menos 5 ou 6 vezes por dia, para reduzir erros sistemáticos. Thylén et al. (1997) identificaram erros de tempo de enchimento da colhedora em mapas de produtividade.

Também existem limitações dos próprios equipamentos. Os fabricantes dos monitores de produtividade em geral atribuem uma acurácia na ordem de 2 a 3% na totalização dos grãos colhidos. Grisso et al. (1999) acompanharam colhedoras equipadas com monitor de produtividade trabalhando a diferentes taxas de alimentação. A faixa de erro observada foi da ordem de 2%. Da mesma forma, Molin et al. (2000) encontraram diferença média de massa entre aquela lida pelo monitor e a totalização de balança em cada tanque graneleiro da colhedora, da ordem de 0,5% e 1,06% para dois anos seguidos em um mesmo talhão e com o mesmo equipamento. Kormann et al. (1998) realizaram ensaios com sensores de fluxo volumétrico em condição controlada de laboratório, obtendo erros em um intervalo de 6,85% a mais e 7,02% a menos que o peso real para dados pontuais.

O número e os tipos de erros variam em função de uma

série de fatores como o modelo do monitor de produtividade e da colhedora, sistema utilizado para o posicionamento, tamanho e formato dos talhões, entre outros. A presença de erros nos dados utilizados para a confecção dos mapas de produtividade é bastante prejudicial à qualidade do mapa gerado, podendo inclusive levar a interpretações errôneas mesmo que seu volume em relação ao total de dados coletados seja pequeno. É importante que esses erros sejam removidos para melhorar a qualidade da informação contida nos mapas. Este trabalho tem por objetivo gerar um filtro para reduzir os erros em mapas de produtividade a partir da eliminação dos dados discrepantes dentre aqueles coletados durante a colheita.

2 MATERIAL E MÉTODOS

Foram utilizados dados coletados durante a colheita de duas lavouras com diferentes culturas, colhedoras e monitores de colheita. As lavouras são identificadas como área 1, colheita de soja na safra 1999 em Castro, PR e área 2, colheita de milho em 1998 em Campos Novos Paulista, SP. Os monitores, juntamente com o sistema de posicionamento, geram um arquivo texto contendo os dados de posicionamento (latitude e longitude) e de quantidade de grão que entra na máquina, além de outros fatores como a largura de corte da plataforma, teor de umidade dos grãos e possíveis fatores encontrados no campo e que sejam georreferenciados durante a colheita pelo operador da máquina.

A Tabela 1 representa o tipo de informação gerada pelos monitores de produtividade. Nas primeiras colunas (latitude e longitude) têm-se as coordenadas que permitem o posicionamento no campo. Na terceira coluna encontram-se os dados de quantidade de grão que passam pelo sensor no momento em que a máquina se encontra naquela posição e na última coluna são apresentados os dados de umidade do grão.

Tabela 1 - Exemplo de um arquivo gerado pelo monitor de produtividade

LONGITUDE	LATITUDE	FLOW	TIME	DISTANCE	SWATH	MOISTURE
.....
-49,933493	-24,857763	349,27	7119068	594	4775	31,40
-49,933490	-24,857411	353,80	7119069	594	4775	31,40
-49,933490	-24,857058	385,55	7119070	594	4775	31,40
-49,933490	-24,856705	385,55	7119071	620	4775	31,40
-49,933487	-24,856352	381,02	7119072	645	4775	31,40
-49,933487	-24,855999	471,74	7119073	645	4775	31,40
-49,933487	-24,855788	430,91	7119074	645	4775	31,40
.....

Os dados foram importados para o software SStolbox®, Sistema de Informações Geográficas (SIG), próprio para aplicações em agricultura de precisão, onde um algoritmo converte as informações em produtividade para cada ponto. Na tabela 2 é apresentado o mesmo conjunto de dados após o cálculo da produtividade de grãos.

Tabela 2 - Exemplo de arquivo gerado pelo monitor de produtividade após o cálculo da produtividade, expressa neste caso como “dry yield”

LONGITUDE	LATITUDE	FLOW	TIME	DISTANCE	SWATH	MOISTURE	DRY_YIELD
-49,933493	-24,857763	349,27	7119068	594	4775	31,40	11,1978
-49,933490	-24,857411	353,80	7119069	594	4775	31,40	11,3430
-49,933490	-24,857058	385,55	7119070	594	4775	31,40	12,3609
-49,933490	-24,856705	385,55	7119071	620	4775	31,40	11,8426
-49,933487	-24,856352	381,02	7119072	645	4775	31,40	11,2498
-49,933487	-24,855999	471,74	7119073	645	4775	31,40	13,9283
-49,933487	-24,855788	430,91	7119074	645	4775	31,40	12,7228

Após o cálculo, os dados podem ser utilizados na geração do mapa de produtividade. Este mapa normalmente apresenta erros, uma vez que o SIG apenas converte os dados gerados pelo monitor de produtividade. Cada linha do arquivo representa um ponto com uma série de atributos, sendo a produtividade um deles. Os erros nos valores de produtividade se expressam como valores muito altos, muito baixos, nulos ou até mesmo pela falta de valores.

A eliminação dos dados discrepantes foi realizada com a utilização da planilha eletrônica, Microsoft Excel® e suas funções de lógica. O algoritmo trabalha os dados de produtividade um a um, comparando-os à média de todos os dados, ao valor de um dado coletado antes do dado avaliado e também à distância em metros entre o ponto analisado e o ponto imediatamente anterior. O valor do ponto analisado é classificado pelo algoritmo como VERDADEIRO, quando está dentro das exigências impostas pelo algoritmo, ou FALSO em caso contrário.

Quando o valor do ponto é considerado VERDADEIRO, é atribuído a ele o valor ZERO uma vez que o mesmo deve ser descartado; quando o mesmo é considerado FALSO, seu valor é mantido.

Desta forma é possível eliminar pontos que estejam com valores 2, 3, 4, ..., n vezes superiores e inferiores à média de todos os pontos, corrigindo possíveis falhas grosseiras do sensor de produtividade e de largura de corte da plataforma inadequada. Através da comparação do dado avaliado com um dado coletado cerca de 15, 16, 17, ..., n pontos atrás e também da distância entre os pontos coletados, é possível eliminar boa parte dos erros pre-

sentes nos finais dos talhões, onde são realizadas manobras com a colhedora e também erros no tempo de enchimento da colhedora.

A eliminação de pontos com valores maiores ou menores que a média é realizada pela condição: $SE(OU(X > n * MÉDIA); (X < MÉDIA/n))$.

O valor do ponto X avaliado é considerado VERDADEIRO quando o mesmo é n vezes maior ou menor que a média de todos os dados, recebendo o valor zero. O valor do ponto X será FALSO quando se encontra no intervalo compreendido entre n vezes o valor da média de todos os dados e o valor dessa mesma média, dividido por n, sendo mantido então o seu próprio valor.

A eliminação de pontos com valores discrepantes, principalmente devido a problemas com coleta de dados durante as manobras e também com o tempo de enchimento da colhedora, é realizada com a seguinte condição: $SE(OU((X_n > (X_{n-15}) * n); (X_n < (X_{n-15})/n)))$. O valor do ponto X_n avaliado é considerado VERDADEIRO quando o mesmo é n vezes maior ou menor que o valor do ponto X_{n-15} , recebendo o valor zero. Será FALSO quando X_n considerado se encontra entre $n * (X_{n-15})$ e $(X_{n-15})/n$, sendo mantido o seu valor. O valor 15 foi adotado pois este é o período (15 segundos) aproximado para o enchimento da máquina e estabelecimento do fluxo de grãos podendo ser alterado caso necessário.

De modo complementar, a distância entre os pontos coletados pode também ser utilizada para reduzir os erros de manobras e de falhas no sistema de posicionamento com a seguinte condição: $SE(OU(DISTÂNCIA < Y_{min}); (DISTÂNCIA > Y_{máx}))$. DISTÂNCIA é o valor calculado, em metros, entre o ponto X_n e o ponto X_{n-1} a partir das coordenadas geográficas dos pontos, previamente convertidas para coordenadas métricas. Y_{min} e $Y_{máx}$ são os valores limites admitidos para a menor e a maior distância entre pontos, de acordo com a colhedora e monitor de produtividade utilizados. Os valores de distância entre a coleta dos pontos são bastante regulares para uma mesma colhedora colhendo o mesmo tipo de grão, o que permite obter para cada colhedora valores bastante seguros de Y_{min} e $Y_{máx}$.

Após a junção das condições acima estabelecidas chega-se ao algoritmo:

$$SE(OU(X > n * MÉDIA); (X < MÉDIA/n); (X_n > (X_{n-15}) * n); (X_n < (X_{n-15})/n)); (DISTÂNCIA < Y_m); (DISTÂNCIA > Y_m))$$

O valor do ponto avaliado será ZERO caso o resultado da expressão for VERDADEIRO e será mantido o valor

da produtividade indicada caso o resultado da expressão for FALSO.

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

3 RESULTADOS E DISCUSSÃO

Foram estudados dados de duas áreas colhidas com diferentes culturas, máquinas e monitores de produtividade. Os mapas de produtividade com dados originais referentes às duas áreas estudadas são apresentados na figura 1. Na figura 2 são apresentados os mesmos mapas após o tratamento dos dados com o algoritmo proposto.

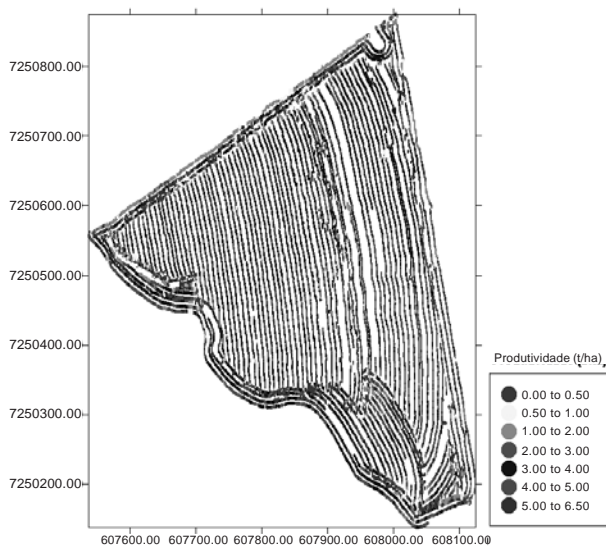
Houve significativa melhora na qualidade dos dados sendo

reduzidos a variância e o desvio padrão dos dados nas duas áreas. O número de dados eliminados pelo filtro foi bastante diferente em termos proporcionais, sendo eliminados 41% dos dados da área 1 (Castro) e 21% dos dados da área 2 (Campos Novos Paulista). A qualidade da informação pode ter sido em parte prejudicada pela eliminação de um grande número de pontos. Entretanto, a densidade de pontos é bastante elevada, mesmo após a remoção de todos os pontos considerados problemáticos pelo algoritmo, sendo possível identificar com bastante segurança as zonas de maior e menor produtividade.

No mapa de produtividade da figura 1, são destacadas regiões onde se tem uma série de erros causados durante a manobra e pelo tempo de enchimento da máquina.

Após a aplicação do filtro pode-se notar uma significativa redução nos erros de manobras e tempo de enchimento da máquina, ou seja, os pontos que possuíam valores discrepantes, próximos às regiões de manobra nas cabeceiras das áreas, foram eliminados.

Área 1 (Castro, PR)
Mapa de produtividade em pontos – soja, 1999



Área 2 (Campos Novos Paulista, SP)
Mapa de produtividade em pontos – milho, 1998

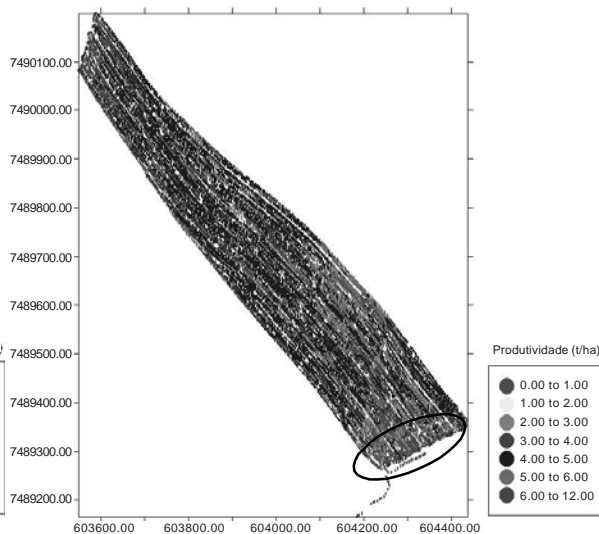


Figura 1 - Mapas de produtividade das duas áreas de estudo com os dados brutos provenientes dos monitores de produtividade: a) soja, 1999, área 1 (Castro); b) milho, 1998, área 2 (Campos Novos Paulista).

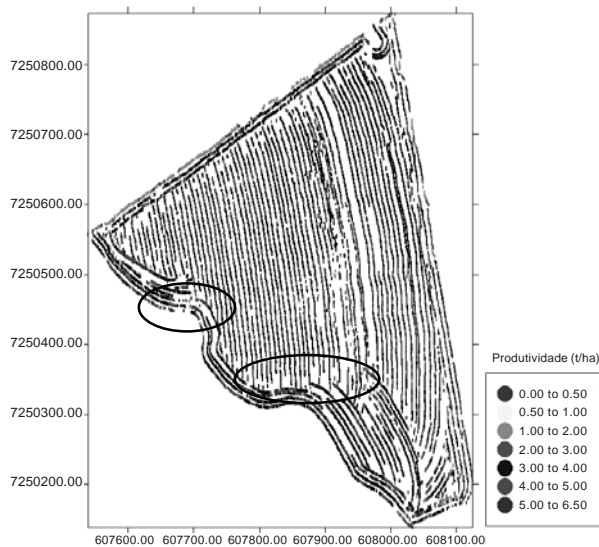
Tabela 3 - Caracterização dos dados antes e após a utilização do algoritmo de filtragem.

	Área 1, Soja, 1999		Área 2, Milho, 1998	
	Dados Brutos	Dados Filtrados	Dados Brutos	Dados Filtrados
Média (t/ha ⁻¹)	2,97	3,17	4,25	4,68
Variância (t/ha ⁻¹)	1,01	0,41	2,51	0,36
Desvio Padrão (t/ha ⁻¹)	1,01	0,64	1,58	0,6
Nº de Total de Pontos	27.177	15.811	12.022	9.396
Nº de pontos/ha	1.151	669	505	394

Na tabela 3 é possível perceber que ocorre redução do desvio padrão e da variância dos dados, devido à eliminação de pontos com valores discrepantes. A elevação da média se deve à eliminação de um grande número de

pontos onde os valores de produtividade são bastante baixos, característica marcante dos erros em manobras de cabeceira da colhedora e tempo de enchimento da máquina

Área 1 (Castro, PR)
Mapa de produtividade em pontos – soja, 1999



Área 2 (Campos Novos Paulista, SP)
Mapa de produtividade em pontos – milho, 1998

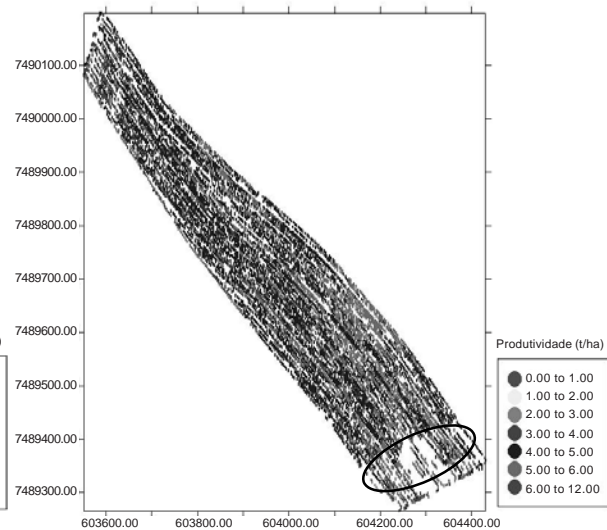


Figura 2 - Mapas de produtividade após a aplicação do algoritmo: a) soja, 1999, área 1 (Castro); b) milho, 1998, área 2 (Campos Novos Paulista).

Na área 2, o comportamento do algoritmo de filtragem foi bastante similar ao da área 1, havendo entretanto uma menor redução no número de pontos eliminados, devido principalmente à menor frequência de coleta de pontos pelo sistema durante a colheita e ao menor número de manobras. Na região inferior da figura 2 referente à área 2, é bastante significativo o número de pontos retirados. A observação do mapa produzido com os dados antes da aplicação do filtro permite afirmar, entretanto, que os pontos retirados representavam erros, o que pode se notar

pela produtividade que variou de 0 – 0,5t.ha⁻¹ a 5 – 6 t.ha⁻¹ em um espaço muito reduzido, o que não ocorre normalmente. Outro tipo de erro que foi bastante reduzido é o erro de largura de corte da plataforma da colhedora, que é caracterizado pela presença de linhas contínuas com produtividade bastante diferente das linhas próximas.

Os valores adotados como parâmetros para o algoritmo variaram para as duas áreas em virtude das diferenças intrínsecas a cada uma, sendo necessário obter estes valores para o filtro apresentar resultados satisfatórios.

No entanto, observa-se que a aplicação de filtros para a remoção de pontos onde os valores coletados pelo monitor de produtividade são discrepantes é importante para a obtenção de um mapa de produtividade de qualidade, evitando interpretações errôneas. A criação de um algoritmo para eliminação dos erros mostrou ser simples, sendo necessário, entretanto, o conhecimento dos principais tipos de erros presentes nos mapas, bem como o tipo de monitor e colhedora utilizados. Por outro lado, a aplicação do algoritmo no grande volume de dados referentes a uma dada área é bastante trabalhosa quando se utiliza uma planilha eletrônica, além de exigir grande espaço para o armazenamento das extensas planilhas geradas.

Uma próxima etapa seria o desenvolvimento de um programa computacional que contemple os passos estabelecidos neste algoritmo e processe os dados brutos provenientes dos monitores de produtividade. Tal programa deve contemplar a agilidade e funcionalidade do processo de limpeza dos dados originais.

5 CONCLUSÃO

O desenvolvimento do algoritmo de eliminação de erros em arquivos de dados para mapas de produtividade mostrou ser simples, embora seja trabalhosa a sua aplicação, além de exigir grande espaço para o armazenamento das extensas planilhas geradas, em função do grande volume de dados dos arquivos. O algoritmo criado permitiu remover boa parte dos erros presentes nos dados estudados, podendo ser utilizado com a restrição de que sejam fornecidos adequadamente os parâmetros requeridos pelo mesmo.

A implementação de um programa computacional com este tipo de algoritmo parece ser o próximo passo na redução dos erros presentes nos dados gerados pelos

monitores de produtividade e pelo sistema de posicionamento.

5 REFERÊNCIAS BIBLIOGRÁFICAS

BLACKMORE B.S., MARSHALL, C.J. Yield mapping; erros and algorithms. In: PRECISION AGRICULTURE INTERNATIONAL CONFERENCE, 3, 1996, Minneapolis. **Proceedings...** Minneapolis: ASA-CSSA-SSSA, 1996. p.403-415.

GRISSO R.D., JASAP J., SCHOROEDER M.A., WILCOX J.C. Yield Monitor Accuracy: Successful Farming Magazine Case Study. ASEA paper, n. 991047, Toronto, 1999.

KORMANN G., DEMMEL M., AUERNHAMMER H. Testing Stand Yield measurement systems in combine harvesters. ASEA paper, n. 983102, Orlando, 1998.

LARSCHEID G., BLACKMORE B.S., MOORE M. Management decisions based on yield maps. In: EUROPEAN CONFERENCE ON PRECISION AGRICULTURE, 1, 1997, Warwick. **Precision Agriculture'97...** Oxford: BIOS Scientific Publishers Ltd, 1997. p. 895-902.

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

MOORE M. **An investigation into the accuracy of yield maps and their subsequent use in crop management.** 1998. 379f. Tese (Doutorado em Engenharia Agrícola) Cranfield University, Silsoe.

THYLÉN L., JURSCHIK P., MURPHY D.L.P. Improving the quality of yield data. In: EUROPEAN CONFERENCE ON PRECISION AGRICULTURE, 1, 1997, Warwick. **Precision Agriculture'97...** Oxford: BIOS Scientific Publishers Ltd, 1997. p. 895-902.