

DELIMITAÇÃO DE REGIÕES HOMOGÊNEAS DE PRODUTIVIDADE POR MEIO DE DOIS MÉTODOS DE ANÁLISE ESPAÇO-TEMPORAL

GISLAINE SILVA PEREIRA¹, LEANDRO MARIA GIMENEZ², MATHEUS GABRIEL ACORSI³, MATHEUS FONTANA WESTPHALEN³,

¹Doutoranda, Escola Superior de Agricultura “Luiz de Queiroz”, USP, Piracicaba, SP, gislainepereira@usp.br

²Professor Dr., Escola Superior de Agricultura “Luiz de Queiroz”, USP, Piracicaba, SP, lmgimenez@usp.br

³Mestrando, Escola Superior de Agricultura “Luiz de Queiroz”, USP, Piracicaba, SP, matheusacorsi@usp.br, matheus.westphalen@usp.br

Apresentado no
XLIX Congresso Brasileiro de Engenharia Agrícola - CONBEA 2020
23 a 25 de novembro de 2020 - Congresso On-line

RESUMO: A disponibilidade de mapas de produtividade oriundos das operações de colheita não garante o uso adequado dessas informações dada a presença de erros e da variabilidade temporal. Um processamento que permita eliminar erros e identificar as regiões que apresentam desempenho distinto e consistente deve ser adotado. Os métodos mais simples para o processamento são os supervisionados. Métodos não supervisionados vem sendo propostos e entre eles o denominado MYPA (*Multitemporal Yield Pattern Analysis*), que emprega a análise de componentes principais, regressões e agrupamento com uso de lógica fuzzy. O objetivo do trabalho foi avaliar o método. Para tal um banco de dados contendo oito mapas das culturas de milho e soja foi utilizado. Um método supervisionado foi considerado como referência. Como resultados a técnica MYPA apresentou menor variância relativa (0,12) comparado ao método de referência (0,68). Na avaliação da qualidade do agrupamento, realizada através de testes de análise multivariada, o método de referência obteve pior ajuste (Pillai = 0,23; Wilks' λ = 0,78) que a MYPA (Pillai = 1,14; Wilks' λ = 0,07). O uso da metodologia requer domínio de técnicas estatísticas trazendo, entretanto, vantagens na regionalização.

PALAVRAS-CHAVE: Mapa de colheita, unidades de produção, análise de componentes principais

MULTI-TEMPORAL ANALYSIS OF YIELD DATA FOR DERIVING YIELD ZONES

ABSTRACT: The existence of raw yield map collection does not ensure the appropriate diagnosis due to erroneous data and the temporal variability of yield. Data processing that can identify and remove outliers and establish zones that have different yield potential with temporal stability must be employed. Supervised methods are often a simplistic approach to delineate zones. Non-supervised methods are constantly being proposed, such as MYPA (*Multitemporal Yield Pattern Analysis*), which features principal component analysis, regression, and fuzzy logic classification. The objective of this work was to assess this method. A dataset containing eight yield maps from maize and soybean was used during the

tests. A supervised method was adopted as reference. The results showed that the MYPA method obtained lower values of relative variance (0.12) when compared to the reference method (0.68). Regarding clustering performance, based on multivariate analysis, the reference method delivered the worst adjustment coefficients (Pillai = 0.23; Wilks' λ = 0,78) in relation to MYPA (Pillai = 1.14; Wilks' λ = 0.07). The methodology requires experience with statistical analysis but offers advantages for zones delineation.

KEYWORDS: Yield map, management zones, principal components analysis.

INTRODUÇÃO: Com a expansão das tecnologias, se tornou possível monitorar o resultado da produção agrícola durante a etapa de colheita, através da espacialização de dados obtidos pela eletrônica embarcada nas colhedoras. Embora seja comum o uso desta informação como camada de resposta para manejar a variabilidade espacial (MCENTEE et al., 2019), a dificuldade dos usuários na interpretação da produção diante da variabilidade espaço-temporal de uma série de safras, torna o processo desafiador (LEROUX et al., 2018; BLASCH et al., 2020). Blasch et al. (2020), desenvolveram uma metodologia para Análise Multitemporal dos Padrões de Rendimento (MYPA), o que auxiliou na determinação de unidades de produtividade, minimizando a influência dos fatores que afetam a produção agrícola como clima e solo. Segundo Blackmore et al. (2003), são necessárias a aplicação de técnicas estatísticas para o tratamento de informações contidas em mapas de produtividade.

Para as condições brasileiras, estudos têm buscado soluções no processamento de mapas de colheita (SPEZIA et al., 2012). Também há contribuições no mapeamento e regionalização de parâmetros de importância para o gerenciamento das lavouras, como o zoneamento de variáveis de solo e clima (RICARDO et al., 2016), ou uso de métodos estatísticos (BAZZI et al., 2013). De acordo com Ping e Dobermann (2005), estudos em diferentes safras e com culturas distintas devem ser realizados para consolidação das técnicas utilizadas. Assim, o objetivo do estudo foi investigar um método para gerar grupos que incorporem informações de estabilidade da variabilidade espacial de uma série histórica de mapas de produtividade de grãos.

MATERIAL E MÉTODOS: Foram utilizados mapas de produtividade de um talhão de 20 hectares, localizado na região Noroeste do Paraná (23°23'S, 52°15'O, 485 m). A área é destinada a produção de soja em safra principal e milho segunda safra. Foram obtidos mapas de colheita das safras de 2015 a 2020, totalizando 8 safras, três de milho (2015, 2018 e 2019) e cinco de soja (2016, 2017, 2018, 2019 e 2020). Os dados brutos foram filtrados com uso do software Mapfilter® (SPEKKEN, ANSELMINI e MOLIN, 2013) para remoção de valores discrepantes. Posteriormente, os dados de produtividade foram normalizados. Superfícies foram produzidas por meio de interpolação dos valores de produtividade com o emprego de krigagem em bloco (10 x 10 m), utilizando o software Vesper 1.62® (MINASNY et al., 2005), configurado para uso de variograma local para geração superfície com pixel de resolução de 25 m² (5 m x 5 m) para cada uma das safras, possibilitando a comparação entre as superfícies.

Método MYPA - Com os mapas de produtividade normalizados, realizou-se uma análise de componentes principais (ACP) para verificar ocorrência de outliers entre os dados. Após a eliminação dos outliers, foram determinadas 7 camadas vetoriais para cada safra, resultando na obtenção de um raster multibanda (Figura 1).

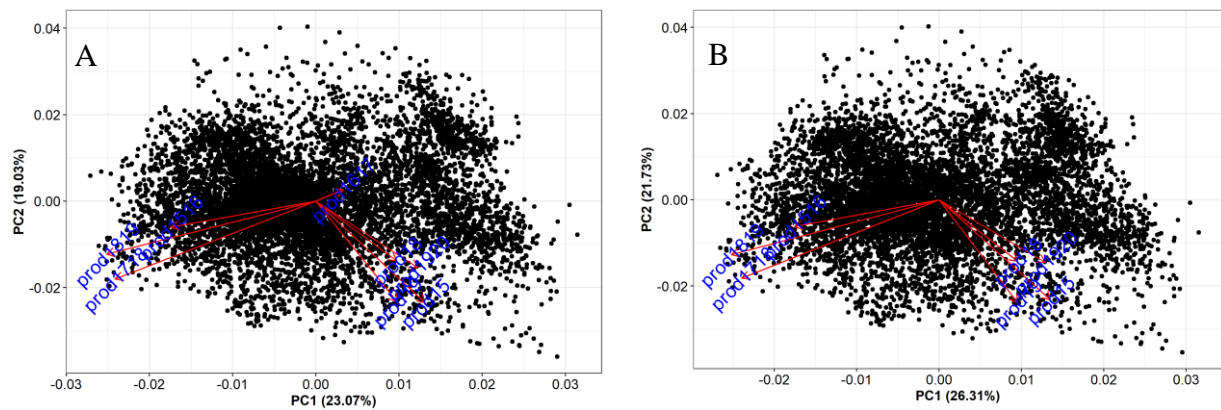


FIGURA 1. Análise de Componentes Principais (ACP) dos mapas de produtividade de grãos em 8 safras após primeira iteração (A) e de 7 safras após remoção de outliers e segunda iteração (B). **Principal Component Analysis (ACP) of grain yield maps in 8 seasons after the first iteration (A) and 7 seasons after removing outliers and the second iteration (B).**

Após a filtragem houve duas etapas de tratamento para obtenção das unidades de produtividade (Figura 2, MYPa e REF). A primeira etapa consistiu na análise de componentes principais do raster multibanda, sendo obtidos seis rasters de componentes principais (CP1 a CP6), utilizados para análise de regressão linear com cada raster de banda única da produtividade padronizada, obtendo assim o coeficiente de determinação corrigido. O processo possibilitou a seleção dos CP que possuíam uma relação de pelo menos 50% ($R^2 > 0,5$) com a produção em cada uma das safras. O agrupamento foi realizado através da técnica “k-means” utilizando o pacote cluster (MAECHLER et al., 2018), obtendo-se 4 classes, sendo: alta produtividade e estáveis (Alto + Est), alta produtividade, porém instáveis (Alto + Inst), baixa produtividade e estáveis (Baixo + Est) e de baixa produtividade e instáveis (Baixo + Inst).

Método de referência- Para a metodologia supervisionada, a regionalização da produtividade foi adaptada de Molin (2002), sendo utilizada a produtividade média (μ) e seu coeficiente de variação (CV%) oriundos de todos os pixels e todas as safras.

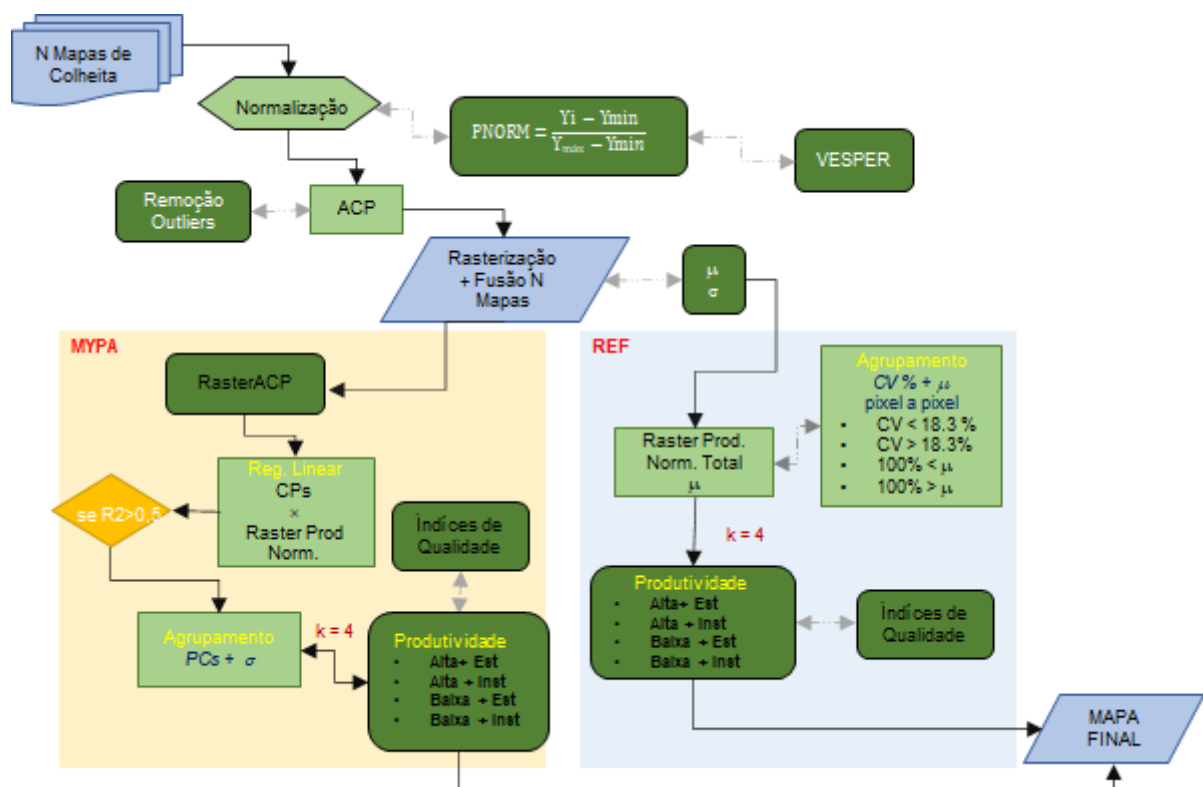


FIGURA 2. Fluxograma das etapas de processamento dos mapas de colheita até a obtenção das unidades de produtividade através de método não supervisionado (MYPA) e supervisionado (REF). **Flowchart for yield data processing to derive yield zones based on the non-supervised (MYPA) and the supervised method (REF).**

Para avaliar a qualidade da regionalização foi utilizado o índice de variação relativa e os testes de Pilai e Wilks. O índice de variação relativa (RV) determina a relação entre a variância cada unidade de produção e a variância total do talhão (PING e DOBERMANN, 2005; XIANG et al., 2007).

RESULTADOS E DISCUSSÃO: Os padrões espaciais da produtividade normalizada para as 8 safras podem ser visualizados na Figura 3. A safra 2016/17 foi considerada um outlier durante o processo de análise de componentes principais, sendo eliminada e realizada uma segunda iteração da análise com as demais safras. Através da espacialização da análise multivariada do raster multibanda da produtividade, foram observados comportamentos que possibilitaram a obtenção das unidades homogêneas de produção. O CP1 apresentou coeficiente de determinação (R2) acima de 0,5 para duas safras de soja (2017/18 e 2018/17), sendo selecionado como camada de determinação das unidades de produtividade. Para Blasch et al. (2020), se mais de 50% do mapa de produtividade for explicado pela variação do PC, a camada é considerada relevante para o agrupamento.

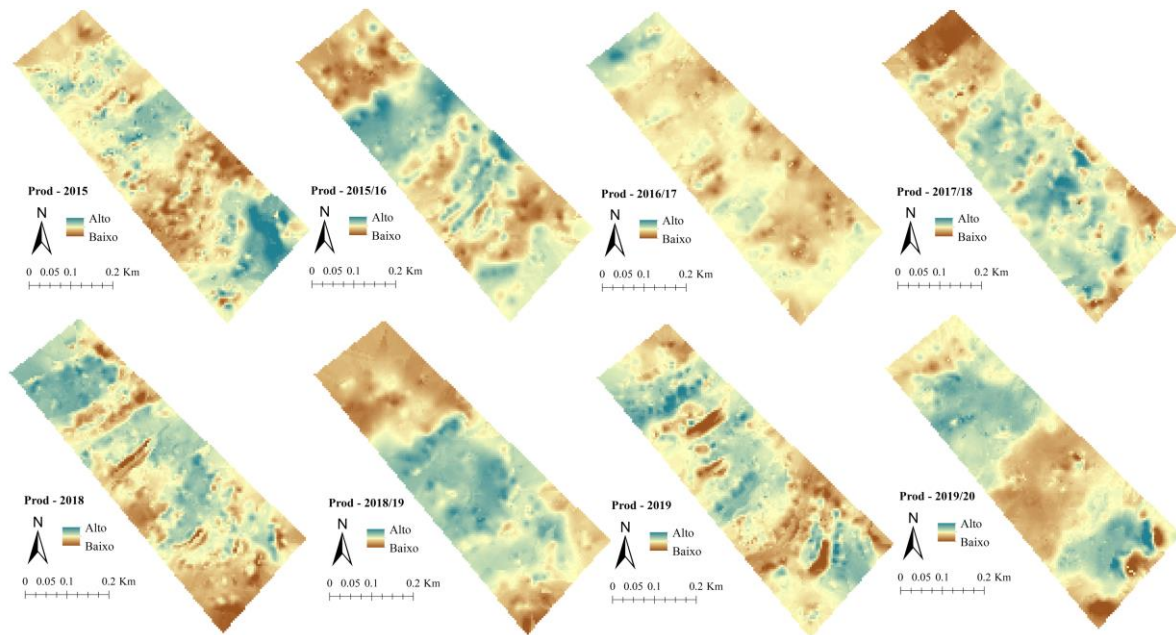


FIGURA 3. Mapas de colheita padronizados das culturas de soja (2015/16, 2016/17, 2017/18, 2018/19 e 2019/20) e milho (2015, 2018 e 2019). **Normalized yield maps of soybean (2015/16, 2016/17, 2018/19 and 2019/20) and maize (2015, 2018 and 2019).**

Semelhanças na regionalização das unidades de produtividade foram observadas nos dois métodos. No método MYPA (Figura 4A) as regiões foram contíguas e extensas, correspondendo a 58% de área estável no talhão. O agrupamento supervisionado (Figura 4B) apresentou regiões menores em continuidade e com 54% de estabilidade. De acordo com Ping e Dorbemann (2005), unidades de produtividade devem ser contíguas e extensas, similar ao que ocorreu em MYPA. Regiões com maior extensão de área possibilitam melhores ações de manejo, dada a finalidade da unidade, facilitando o emprego de ferramentas da agricultura de precisão, como aplicação em taxa variável e semeadura localizada (XIANG et al., 2007).

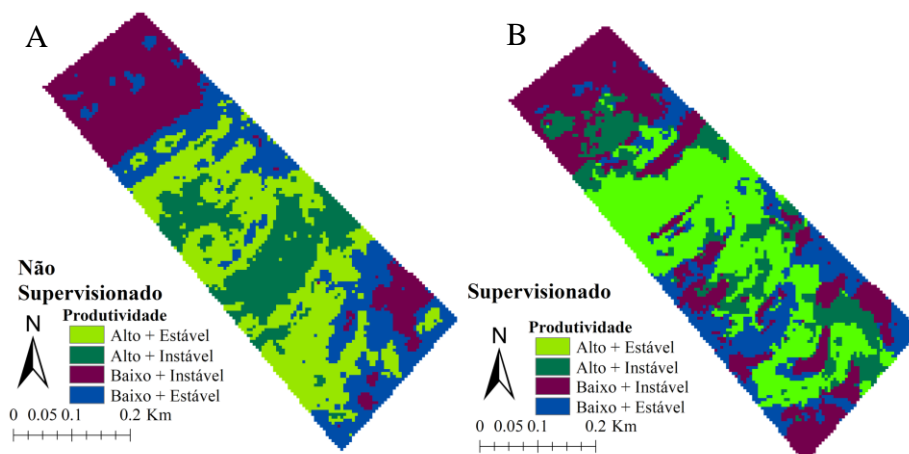


FIGURA 4. Unidades de produtividade obtidas baseadas na metodologia MYPA (A) e pelo método de referência (B). **Yield zones obtained with MYPA (A) and with the supervised technique based on reference method (B).**

O desvio padrão médio das unidades de produção nos dois agrupamentos foram semelhantes (Tabela 1). Ao avaliar RV, a metodologia MYPA apresentou um índice de 0,12, o que poderia indicar o melhor desempenho do agrupamento em comparação ao método de referência (RV = 0,64). Os índices de Pillai e Wilk's podem confirmar a qualidade de ajuste das unidades para MYPA. O valor de Pillai para a metodologia de referência (Pillai = 0,23) foi menor do que no método não supervisionado (Pillai = 1,14). Um maior índice de Pillai representa um melhor ajuste (BLASCH et al., 2020). Já o índice de Wilks' λ foi menor para MYPA com valor de 0,07 (Tabela 1), indicando alta qualidade de ajuste para o agrupamento não supervisionado.

TABELA 1. Avaliação do desempenho das metodologias de agrupamento MYPA e de Referência na geração de unidades de produtividade. **Coefficient of Variation (%) within and between yield zones from the different methods used.**

| Unidade | <i>Multitemporal Yield Pattern Analysis</i> | | | | | Método de Referência | | | | |
|--------------|---|-----------------------|-------|--------|------------------|----------------------|-----------------------|-------|--------|------------------|
| | Área % | $\sigma_{\text{méd}}$ | RV | Pillai | Wilks' λ | Área % | $\sigma_{\text{méd}}$ | RV | Pillai | Wilks' λ |
| Alta + Est | 33 | 0,079 | | | | 35 | 0,072 | | | |
| Alta + Inst | 19 | 0,093 | | | | 16 | 0,127 | | | |
| Baixa + Inst | 23 | 0,123 | 0,117 | 1,144 | 0,072 | 30 | 0,117 | 0,637 | 0,229 | 0,777 |
| Baixa + Est | 25 | 0,092 | | | | 19 | 0,074 | | | |

CONCLUSÕES: O método MYPA foi mais eficiente na obtenção das unidades de produtividade a partir de séries temporais de mapas de colheita, permitindo a identificação e eliminação de dados discrepantes e a sumarização da variabilidade em unidades mais consistentes que um método supervisionado.

AGRADECIMENTOS: Os autores agradecem a consultoria Farmer's na disponibilização de banco de dados utilizado para emprego das metodologias aqui descritas.

REFERÊNCIAS:

ALVARES, C.A.; STAPE, J.L.; SENTELHAS, P. C.; GONÇALVES, J.L. DE M.; SPAROVEK, G. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift* 22, 711–728.

BAZZI, C. et al. Management zones definition using soil chemical and physical attributes in a soybean area. *Engenharia Agrícola*, v. 33, n. 5, p. 952–964, out. 2013. DOI: 10.1590/s0100-69162013000500007. Disponível em: <http://dx.doi.org/10.1590/S0100-69162013000500007>.

BLASCH, G; LI, Z; TAYLOR. Multi-temporal yield pattern analysis method for deriving yield zones in crop production systems. *Precision Agriculture* [S. l], 2020, DOI: 10.1007/s11119-020-09719-1, Disponível em: <http://dx.doi.org/10.1007/s11119-020-09719-1>

BLACKMORE, S; GODWIN, R. J; FOUNTAS, S. The Analysis of Spatial and Temporal Trends in Yield Map Data over Six Years. *Biosystems Engineering* [S. l], v. 84, n. 4, p. 455–466, 2003, DOI: 10.1016/s1537-5110(03)00038-2

HARTIGAN, J. A.; WONG, M. A. 1979. “Algorithm AS 136: A K-means clustering algorithm.” *Applied Statistics*. Royal Statistical Society, 100–108.

LEROUX, C; JONES, H; CLENET, A; DREUX, B; BECU, M; TISSEYRE, B. A general method to filter out defective spatial observations from yield mapping datasets. **Precision Agriculture** [S. 1], v. 19, n. 5, p. 789–808, 2018, DOI: 10.1007/s11119-017-9555-0

LEUTNER, B., HORNING, N., SCHWAB-WILLMANN, J. **RStoolbox: Tools for remote sensing data analysis**. R package version 0.2.1. 2018. <https://CRAN.R-project.org/package=RStoolbox>.

LEVER, J.; KRZYWINSKI, M.; ALTMAN, N. Principal component analysis. **Nature Methods**, v. 14, n. 7, p. 641–642, jul. 2017. DOI 10.1038/nmeth.4346

MAECHLER, M., ROUSSEEUW, P., STRUYF, A., HUBERT, M., & HORNIK, K. **cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1**. 2018. <https://cran.r-project.org/web/packages/cluster/index.html>.

MINASNY, B., MCBRATNEY, A. B., & WHELAN, B. M. **VESPER version 1,62, Australian Centre for Precision Agriculture**, McMillan Building A05, The University of Sydney, NSW 2005, Acesso em jun 2020

MOLIN, J.P. Definição de unidades de manejo a partir de mapas de produtividade. **Engenharia Agrícola**, v.22, p.83-92, 2002.

MORISSETTE, L.; CHARTIER, S. The k-means clustering technique: General considerations and implementation in Mathematica. **Tutorials in Quantitative Methods for Psychology**, v. 9, n. 1, p. 15–24, 1 fev. 2013. DOI 10.20982/tqmp.09.1.p015.

PING, J.L., DOBERMANN, A. Processing of Yield Map Data. **Precision Agric** 6, 193–212 (2005). <https://doi.org/10.1007/s11119-005-1035-2>

R Core Team. (2018). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria.

RICARDO, S. et al. Redundant variables and the quality of management zones. **Engenharia Agrícola**, v. 36, n. 1, p. 78–93, fev. 2016. DOI 10.1590/1809-4430-eng.agric.v36n1p78-93/2016.

SPEZIA, G. R. et al. Model to estimate the sampling density for establishment of yield mapping. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 16, n. 4, p. 449–457, abr. 2012. DOI 10.1590/s1415-43662012000400016.

SPEKKEN, M.; ANSELMINI, A. A.; MOLIN, J. P. **A simple method for filtering spatial data**. In: STAFFORD, J. V (Ed.). *Precision agriculture '13*. Wageningen: Wageningen Academic Publishers, 2013. p. 259–266.

XIANG, LI; YU-CHUN, PAN; ZHONG-QIANG, GE; CHUN-JIANG, ZHAO, Delineation and Scale Effect of Precision Agriculture Management Zones Using Yield Monitor Data Over Four Years, **Agricultural Sciences in China**, [S. 1], v. 6, n. 2, p. 180–188, 2007. DOI: 10.1016/s1671-2927(07)60033-9.