Multi-temporal Yield Pattern Analysis – Adaption of Pattern Recognition to Agronomic Data

Gerald Blasch[1], James A. Taylor[1]

[1] Newcastle University, School of Natural and Environmental Sciences, Agriculture Building, Newcastle upon Tyne, Tyne and Wear, NE1 7RU, UK

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract. *In precision agriculture, the understanding of yield variability, both spatial and temporal, can deliver essential information for the decision making of site-specific crop management. Since commercial yield mapping started in the early 1990s, most research studies have focused on spatial variance or short-term temporal variance analyzed statistically in order to produce trend maps. Nowadays, longer records of high-quality yield data are available offering a new potential to evaluate yield variability over time by using alternative (to the traditionally statistical approach) analysis methods, for example pattern recognition. The research idea of Multi-temporal Yield Pattern Analysis (MYPA) was inspired by the digital soil mapping method Multitemporal Soil Pattern Analysis (MSPA). In order to produce soil property maps, the MSPA method extracts stable soil reflectance pattern from satellite time series using pattern recognition combined with statistical pattern stability analysis. The MYPA approach is the adaption of image analysis techniques of the remote sensing discipline (here: pattern recognition) to agronomic data (here: yield data). The current state of the MYPA method will be presented that makes it possible to i) select outlier yield maps from yield map time series, ii) detect spatially homogenous yield pattern, and iii) evaluate their spatiotemporal variability. This method enables the generation of site-specific crop management zones considering both the productivity and stability of yield over space and time. The MYPA method consists basically of following steps: (1) identification and elimination of outlier yield maps, (2) yield pattern detection using principal component analysis; (3) evaluation of spatiotemporal yield pattern stability using statistical per-pixel analysis; and (4) management zones delineation based on k means clustering. Results from one demonstration field are presented and contrasted (with favourable outcomes) with the more traditional statistical mean approach to multi-temporal yield pattern delineation.*

Keywords. *Yield maps, Spatio-temporal yield variability, k-means clustering, Principal component analysis.*

1 Introduction

The use of yield sensors at harvest is the only definitive way that final production data can be gathered in crop production systems, particularly cereal systems. Yield prediction approaches exist (Basso et al. 2013; Meroni et al. 2013; Rembold et al. 2013; Shanahan et al. 2001) that can provide information on expected yields. However, predictions are always subject to potential in-season effects, particularly mid-season (e.g. frost) that affect grain set and late-season that may affect grain fill (e.g. water limitation or extreme weather). For this reason, yield sensors linked to Global Navigation Satellite Systems (GNSS) are the only way of site-specifically auditing yield.

Yield sensing and mapping has been performed since the beginning of site-specific crop management, and was one of the key technologies that lead to the development of this domain. It became clear from an early stage that the temporal analysis of yield patterns was likely to be as important as the spatial analysis (McBratney et al. 1997). Temporal analysis has been slower to be addressed, in part because it takes several years to generate a suitable data set. Several early attempts were made in the late 1990s and early 2000s to address temporal variability issues using short time-series of yield maps (Blackmore and Moore 1999; Blackmore et al. 2003; Pringle et al. 2003; Stafford et al. 1996). These mainly addressed the average yield and temporal variance statistically and employed trend maps.

Since 2005 there has been relatively little work published on the analysis of the temporal variance in yield maps despite over this period larger temporal yield data sets being generated by growers. However, the importance of the temporal variance has not diminished, nor the need to manage this, particularly in relation to the spatio-temporal patterns that can be generated from weather-crop interactions (e.g. Taylor et al. 2007). A recent paper (Leroux et al. 2018) has revisited this issue and applied a segmentation approach to the analysis of multi-year spatial yield monitoring data. Given the availability within the industry of these historical datasets, there is a clear need for methods of data processing to provide growers with information on the temporal variability within their system. Manually assessing 5-15 years of yield data in the form of maps, often with error measurement errors included, is difficult even for the manager of the field.

In this paper, an approach based on pattern recognition (associated with image analysis) is proposed as a method for integrating multi-year yield data into several information layers to assist with management, particular risk management, in cereal production systems. These information layers assess at the local scale both the relative yield response over time as well as the stability or certainty of the yield response in any given year. The approach is derived from a recently published methodology for assessing soil patterns within multi-temporal bare soil images (Blasch et al. 2015a, 2015b).

2 Materials and Methods

2.1 Site Description

The approach is demonstrated on an 80-ha field (-29°49'47"S, 150°00'21"E, 324 m above sea level) in a mixed cropping enterprise. The field (AOI1 Moree) is located ~43 km southeast of Moree, New South Wales, Australia. The topography is undulating with the predominant soil type on the cropping areas (flat and gently rolling terrain) being Vertisols (cracking clays) with shrink-swell properties upon drying and wetting (ASRIS 2011; FAO et al. 1998; Northcote et al. 1960-1968). The climate is humid subtropical (Köppen-Geiger climate: Cfa; Peel et al. 2007) with a summer dominant rainfall and high evaporation rates in summer. The long-term (1995-2016) average minimum and maximum temperature are 12.5°C and 26.7°C and mean rainfall is 589.3 mm/year (Moree Aeroport weather station; Australian Bureau of Meteorology 2016).

2.2 Pre-processing of Yield Data and Yield Map Generation

To analyze the multi-temporally complexities in yield patterns, a seven-year yield map time series within the period 1997 to 2006 was obtained. Yield data (1996-2004) were recorded at harvest using a Case IH combine harvester equipped with an AgLeader Yield Monitor. In 2005-2006 a John Deer harvester with the Greenstar yield monitoring system was used. No yield data were collected in 2001, which was a planned fallow season and in 2002 due to crop failure associated with drought. The field is managed in a three year crop rotation (nominally wheat (*Triticum aestivum*), wheat, break crop), with the break crop typically being a legume (e.g. Chickpea (*Cicer arietinum*)). The production system is rain-fed only. In years with dry conditions leading into a winter crop, the winter crop may be abandoned before sowing and the field left fallow leading into a summer cereal (Sorghum (*Sorghum bicolor*)).

Yield data were cleaned and removed of outliers (values > $\pm 2.5\sigma$) following the protocol of Taylor et al. (2007). Data were then interpolated onto a common 5 m grid using block-kriging (10 m blocks) with a local variogram using the Vesper shareware (Minasny et al. 2005). As a common grid was used, data from all years were interpolated to the same grid nodes, generating co-located data. The raw data is irregularly spaced and not co-located, due to differences in harvest operations each year. Interpolated data was stored in a geo-database and geo-tiffs were generated for each year using ArcMap v10.3 (ESRI, Redlands, CA). The descriptive statistics of the filtered yield data are shown in Table 1.

Table 1: Descriptive statistics of yield data collected from each year, including crop harvested. All data expressed as t/ha.

Yield (t/ha)	Mean	Std. Deviation	Minimum	Maximum
Sorghum 1997	3.45	1.28	0.40	6.27
Sorghum 1998	5.41	0.71	3.05	7.21
Chickpea 1999	1.31	0.32	0.06	2.25
Wheat 2000	2.55	0.60	0.78	4.30
Wheat 2004	4.85	0.59	2.52	6.22
Sorghum 2005	3.53	0.54	1.65	4.80
Chickpea 2006	1.70	0.35	0.87	2.58

2.3 Multi-temporal Yield Pattern Analysis (MYPA) Workflow

The Multi-temporal Yield Pattern Analysis (MYPA) approach presented here is composed of four processing steps: (1) data mining and preparation, (2) yield pattern detection, (3) spatiotemporal yield pattern stability analysis, and (4) yield productivity-stability zones delineation (Figure 1).

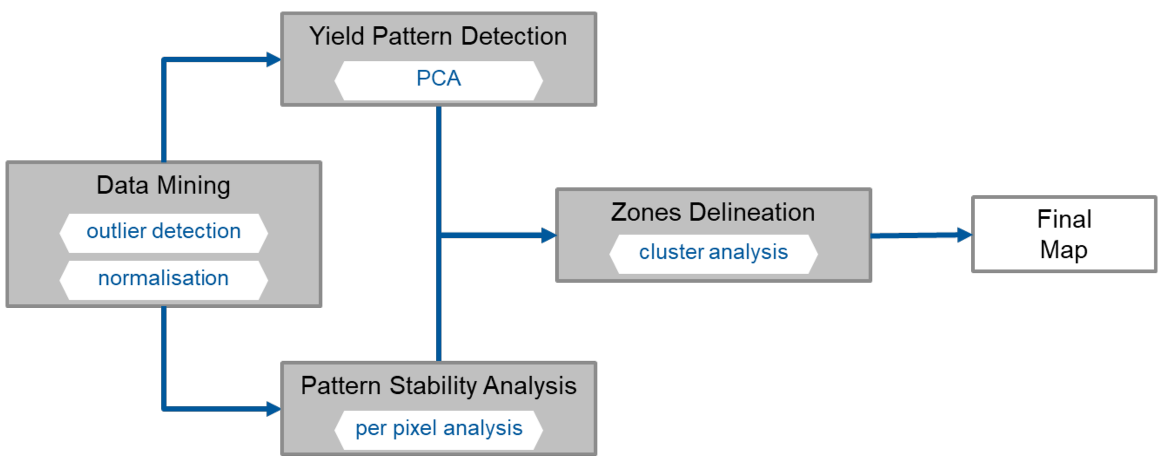


Figure 1: Schematic illustrating the processing steps of the MYPA workflow, including key actions, to generate a final map.

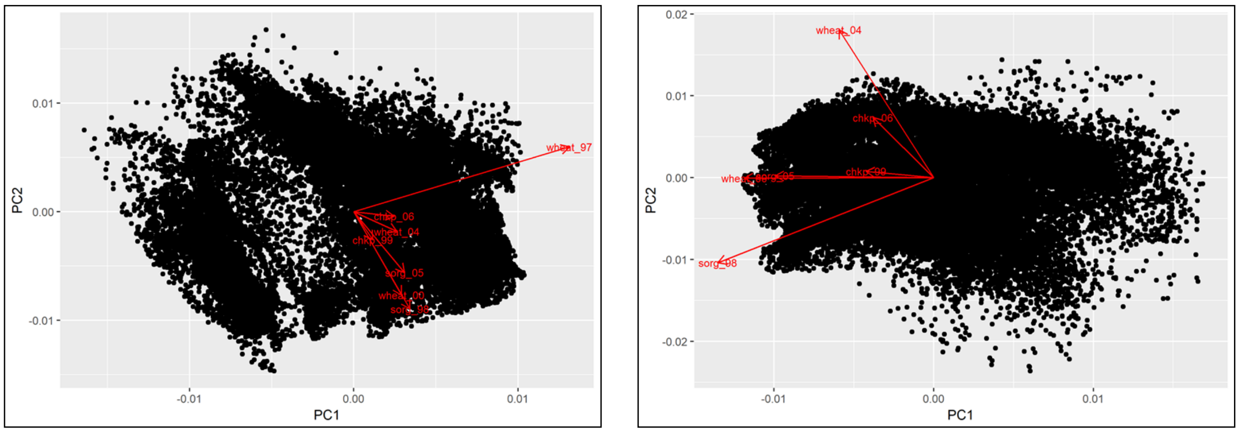
2.3.1 Data Mining and Preparation

The two main goals of the data mining step are;

a) the identification and elimination of outlier yield maps from the yield map time series and,

b) data normalization to permit comparison of yield from different crops with potentially very different yield values.

Important data features such as outliers and departures from a multi-normal distribution can be identified using PCA. Accordingly, PCA was applied to a multi-temporal layer stack of all yield data sets, where one layer represents one yield map, i.e. one year of yield data. The visualization of resulting eigenvectors in plotted data points by PC1 and PC2 (Figure 2a) shows clearly that the wheat yield map from 1997 (wheat_97) is deviating considerably from the norm of all yield data sets. Consequently, this yield data set was assumed as outlier and removed from the layer stack after visual validation and discussion with the grower. This outlier detection procedure was iterated as long as no other outlier data set was identified. Here, after the second iteration, no other outlying data set was detected (Figure 2b) although various years did appear to group along different vectors.



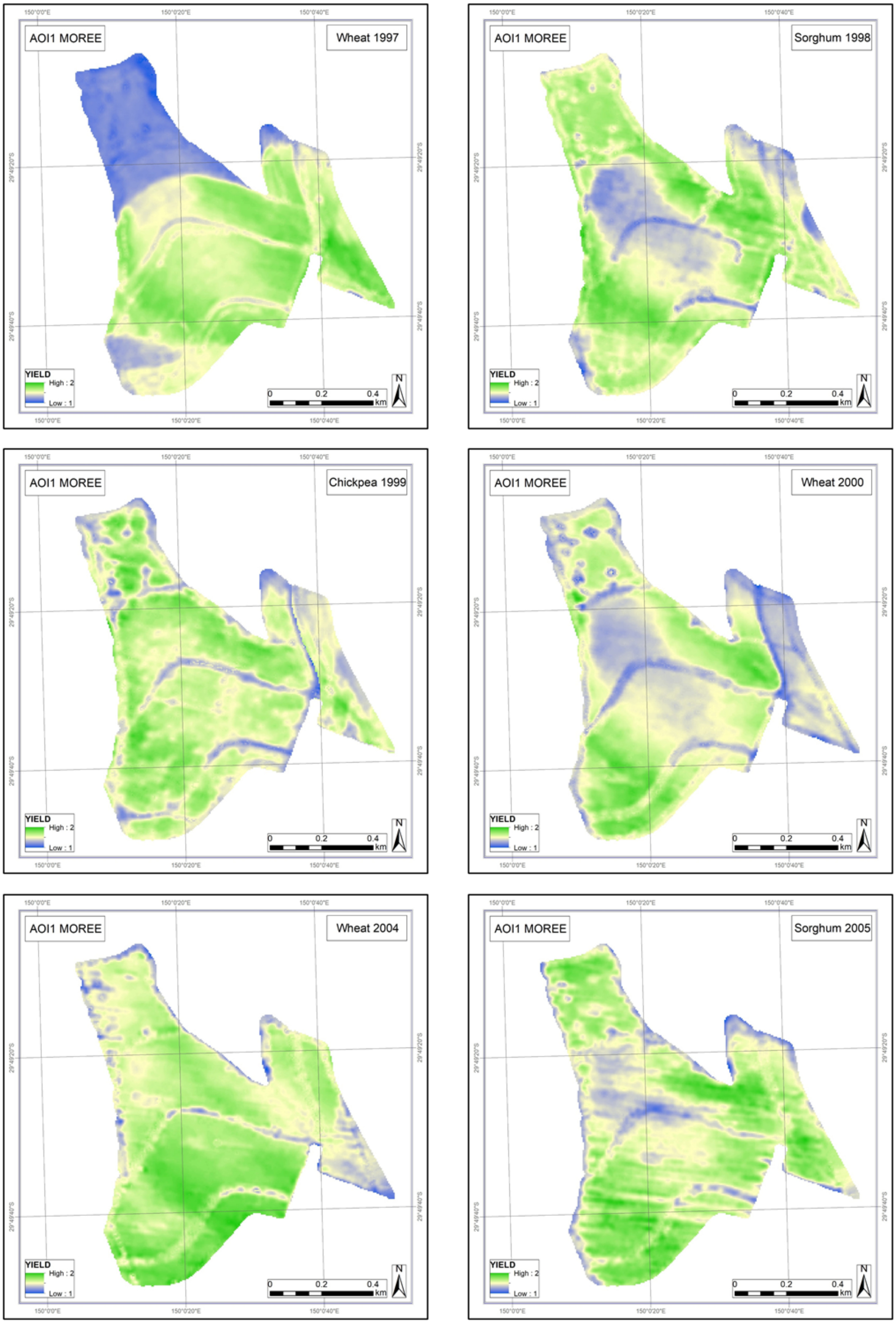
(**a**) (**b**)

Figure 2: Eigenvectors (red arrows) and data values by PC1 and PC2 (black points) – (a) of all yield data sets after first iteration; (b) of remaining yield data sets after second iteration of the outlier detection process.

Normalization of yield data to the mean or maximum is a common method to compare different crops for the same field. Here, the remaining yield maps were normalized to the maximum by calculating normalized yield pixel values (Y_{norm}) for each single yield map data set using equation 1:

$$Y_{norm} = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} + 1 \quad (1)$$

where Y_{norm} is the normalized yield pixel value, Y the original yield pixel value, and Y_{min} and Y_{max} the minimum and maximum pixel value of each yield map data set. Thus, the yield data of each year were scaled between 1 and 2, and the potential disproportionate varying influence from the magnitude and variation of different crop types was eliminated. Figure 3 shows the normalized yield maps, including the previously removed outlier yield map from 1997 (wheat 1997). For consecutive MYPA processing steps, a layer stack of the six remaining normalized yield data was generated.


AOI1 MOREE
Wheat 1997
AOI1 MOREE
Sorghum 1998
AOI1 MOREE
Chickpea 1999
AOI1 MOREE
Wheat 2000
AOI1 MOREE
Wheat 2004
AOI1 MOREE
Sorghum 2005
YIELD
High : 2
Low : 1
0
0.2
0.4
km
N

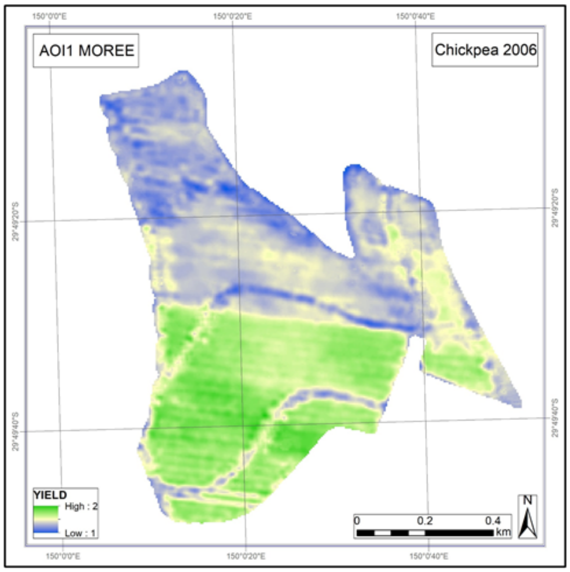


Figure 3: Normalised yield maps with varying crop types (Sorghum, Chickpea, and Wheat) including the outlier yield map (wheat 1997).

2.3.2 Yield Pattern Detection

In the remote sensing discipline, digital image processing is used to reduce disturbance impacts on data, prepare images for consecutive processing, as well as extract and interpret acquired information. Complex spectral pixel operations such as Principal Component Analysis (PCA) are highly suitable for pattern recognition, change detection, and stability analysis using satellite time series or other multi-temporal data sets. PC transformation – a linear transformation of correlated original spectral bands – is one of the most common complex pixel operations applied to analyze highly correlated multi-dimensional data without harming the band information, remove redundant information, isolate noise components, and reduce dimensionality without significant information loss (Panda et al. 2010). The produced uncorrelated output bands (principal components - PCs), which are rotated to a data variance maximum, allow for the identification of patterns in the data describing the spectral variance, highlighting their similarities and differences, to expose the underlying dimensionality of multivariate data, and to compress image data (Abdel-Kader 2011; Mulla 2013; Smith 2002).

To detect spatio-temporal yield patterns, PC transformation (PCA) (software R: package “RStoolbox”; Leutner et al. 2018; R Core Team 2018) was applied to multiple normalized yield maps (images) from different years and varying crop types, whereby the layer stack of normalized yield maps (here: 6 layers) acts as a spectral image and each yield map layer as one spectral band. In order to identify those PCs which are most responsible for yield pattern over time, the relationships between normalized yield and PCs were analyzed using linear regression analysis, and as correlation measure the coefficient of determination (R^2) was computed. Therefore, the location-specific normalized yield value from each yield map, and the corresponding PC values were extracted at every pixel. PCs with at least moderate correlation ($R^2 \geq 0.5$) were selected for the final processing step to generate yield productivity-stability zones (section 2.3.4).

2.3.3 Spatiotemporal Pattern Stability Analysis

The multi-temporal approach makes it possible to evaluate the spatial variability over time. Taking advantage of that, the spatio-temporal stability of yield responses were analyzed using statistical per-pixel analysis on the layer stack of the normalized yield data. For each pixel, the standard deviation (SD) of normalized yield was calculated across this multi-temporal stack. This generated a final image indicating the variance of response, or rather the relative stability of yield response, over the 6 field-years.

2.3.4 Yield Productivity-Stability Zones Delineation

The best performing PCs (section 2.3.2) and the overall stability (SD) (section 2.3.3) were selected as the desired data layers for the delineation of yield productivity-stability zones and subsequently stacked into one final layer stack. A *k*-means clustering approach (software R: package “cluster”; Maechler et al. 2018; R Core Team 2018) was performed to generate yield zones using PCs for yield productivity and the SD layer for the stability of yield variability (respectively stability). Briefly, the *k*-means clustering method aims to allocate the pixel values of desired data layers (here: PC1, PC2, PC5, and SD) into *k*-classes (clusters) in order to minimize the within-cluster variability and maximize the differences between the means of the *k*-classes.

2.4 Statistical Approach

In order to compare the MYPA approach with the “traditional” statistical approach, the average normalized yield using the arithmetic mean (MEAN) was calculated for each pixel across the normalized yield maps. Subsequently, a final two layer stack composed of MEAN and SD (section 2.3.3) was generated and regression analysis preformed. This two layer stack was also used for *k*-means clustering to obtain yield productivity-stability zones (according to section 2.3.4).

2.5 Zonal Statistics/MANOVA

To statistically compare the yield zones generated by the MYPA and the traditional MEAN approaches, a MANOVA analysis was performed using the all 6 years of yield data as the dependent variables and the derived zones from the two different approaches (MYPA with PC1, PC2 and PC5, and MEAN) as the independent variables. MANOVA uses discriminant analysis to ‘simplify’ the multi-temporal yield data into a single variable that is then related to the zoning approaches. The quality of the zoning models is assessed using Pillai’s Trace, with a higher value indicating a better model fit. Wilks’ Lambda, an alternative measure, is also presented (with lower values indicating better model fit).

3 Results and Discussion

To demonstrate the high potential of pattern recognition based on PCA, it is essential to identify those PCs that explain yield patterns for subsequent steps in the MYPA processing. The obtained best PCs and computed overall stability of yield pattern allow the delineation of yield productivity-stability zones. These results were compared with the traditional statistical approach (here: MEAN).

3.1 Data Mining and Preparation

The PCA of the yield maps indicated that the 1997 yield map was an outlier in the yield map stack (Figure 2a). It can be clearly seen that the northern part of the field in this year (Figure 3) has a very low yield response. This was due to double cropping in that part of the field causing a large water deficit in the winter wheat production. It is clear that the pattern here is dominated by a “one-off” management decisions and that this information should be removed. The subsequent PCA did not highlight any outliers (Figure 2b). However it is clear that there is some management effect in the Sorghum 2006 data (a clear horizontal delineation between the north and south parts of the field). This yield map tracks similar to the Wheat 2004 map in the attribute space (Figure 2b). It has higher yields on the central part of the field that typically has a lighter texture and shallow soil depth (see Taylor et al. 2007 for full details). This map could have been removed based on local opinion and an observable management effect, however since statistically it was not mapped as an outlier the decision was made to include it in the subsequent analysis.

3.2 Yield Pattern Detection

The PCA applied to the yield maps produced good results for yield pattern detection with little noise. Based on the covariance matrix, six principal components (PC 1-6) were calculated from the transformation of the 6 layers of the normalized yield map layer stack. Maps of the 6 PCs generated are shown in Figure 4. The linear regression relationship between the PCs and normalized yield are given in Table 2.

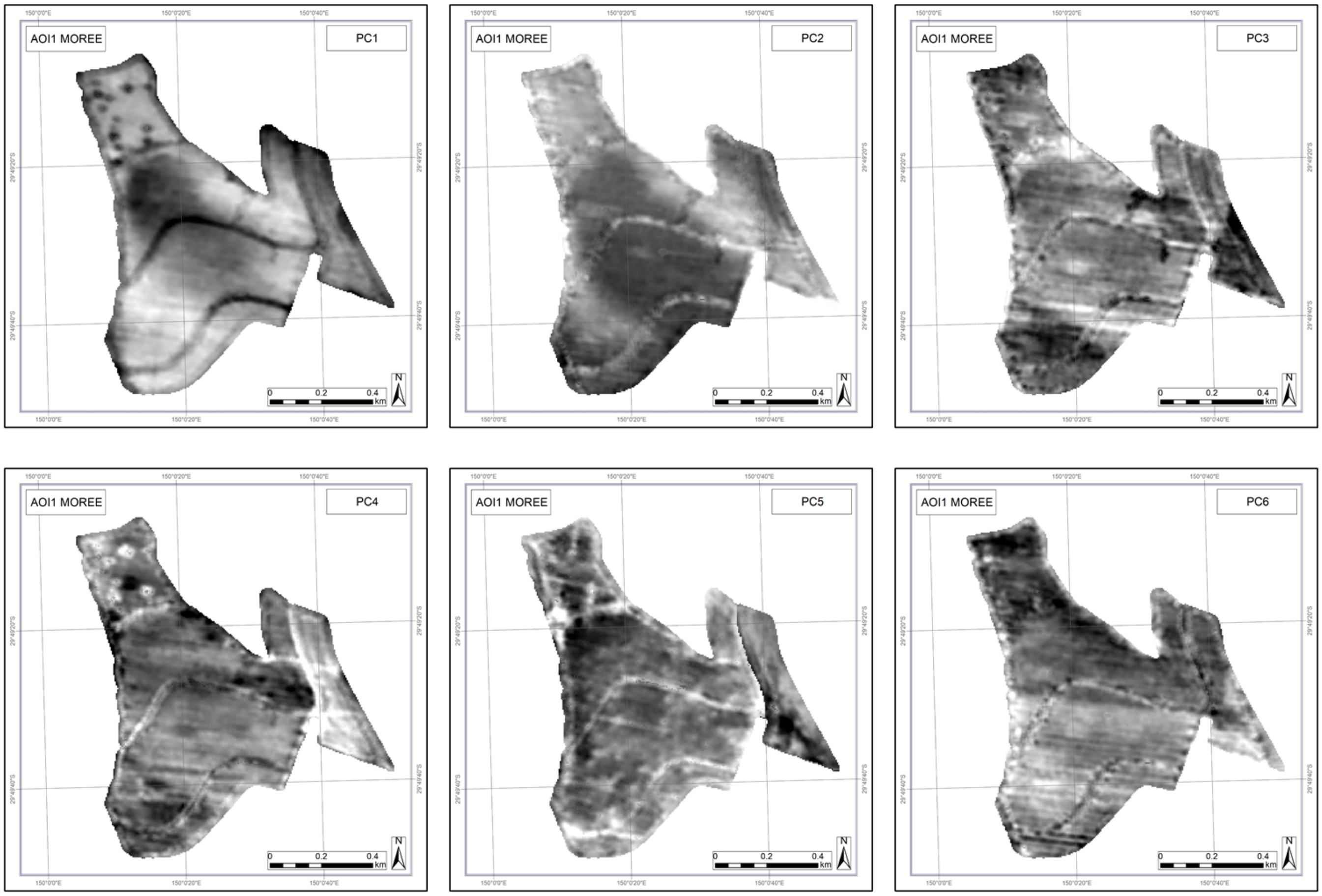


Figure 4: Detected pattern by the PCA approach (PC 1-6).

Table 2: R^2 for the relationship between normalised yield (dependent variable) and the PCs/MEAN (independent variable). Relationships where more than 50% of the yield variance was explained by a PC or MEAN are indicated in bold font.

Yield	PC1	PC2	PC3	PC4	PC5	PC6	MEAN
Sorghum 1998	**0.73**	0.16	0.09	0.016	0.00034	1.7e-05	0.48
Chickpea 1999	0.35	0.0044	7.7e-05	0.05	**0.59**	0.011	0.44
Wheat 2000	**0.78**	7e-06	0.024	0.18	0.012	0.0037	**0.67**
Wheat 2004	0.2	**0.72**	0.049	0.00062	0.0032	0.02	0.39
Sorghum 2005	**0.67**	0.00014	0.22	0.097	3.5e-05	0.015	**0.63**
Chickpea 2006	0.23	0.33	0.0015	0.063	0.011	0.36	0.49

Each obtained uncorrelated output band (PC) contain a certain percentage of the total variance of the yield map time series: PC1 56.7 %, PC2 21.6 %, PC3 7.8 %, PC4 6.5 %, PC5 3.9 %, and PC6 3.4 %.

By the pattern recognition approach, relatively good correlations for yield data were found with PC1 (R^2: 0.78-0.67; Wheat 2000, Sorghum 1998, and Sorghum 2005), PC2 (R^2 = 0.72; Wheat 2004), and PC5 (R^2 = 0.59; Chickpea 1999). PC1 picks up the 'normal' yield patterns that can be related to patterns in soil depth and moisture holding conditions in a typically water limiting production system. PC2 picks up a flip-flop effect in years where in-season rainfall permits stronger growth (and less water-logging) on lighter soils. PC5 describes an unusual year yield map that is fairly very even (Chickpea 1999). The management affected Chickpea 2006 yield map, which was not discarded in the preprocessing, has no dominant interaction with a PC but picks up weak relationships (R^2: 0.23 – 0.36) with PCs 1, 2 and 6. It is likely due to the fact the

2006 yield contains elements of various PCs that it was not identified as an outlier in preprocessing. The management effect is evident in the PC6 map (Figure 4).

In contrast, the statistical approach (MEAN) revealed relatively good relationships with just two yield map data sets (R^2: 0.67-0.63; Wheat 2000 and Sorghum 2005 – Table 2). This demonstrates clearly that PCA is highly suitable for extracting yield patterns associated with differing production conditions and proves an improved potential for spatiotemporal yield analysis compared to the statistical approach, which melds and smooths conditions across years. Figure 5 shows the detected pattern using the statistical MEAN approach, which reveal similarities with the PC1-pattern.

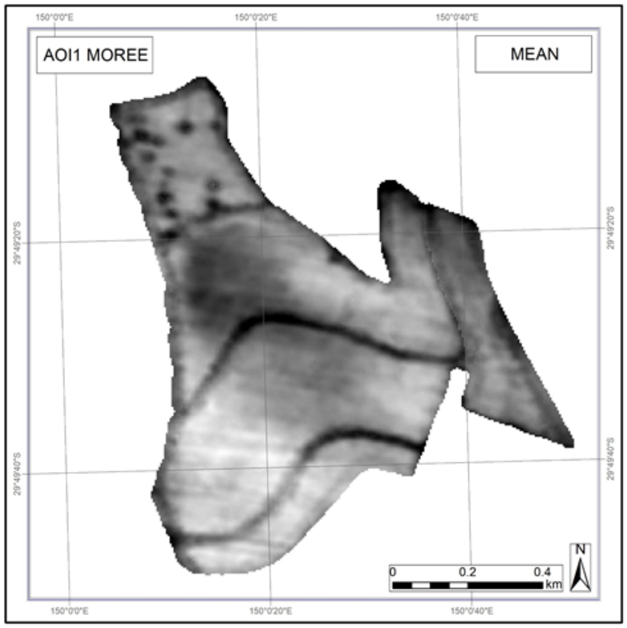


Figure 5: Detected pattern by the statistical MEAN approach.

3.3 Spatiotemporal Pattern Stability Analysis

To evaluate the temporal-spatial stability of yield, statistical pixel-wise analysis using standard deviation (SD) over time as a measure was applied on the layer stack of remaining normalized yield data. This pattern stability evaluation method enabled the identification and visualization of yield variability in space and time, where high SD per pixel indicates high variability (unstable yield pattern) and low SD per pixel low variability (stable yield pattern) (Figure 6). Figure 6 shows a highly stable area in the south-western part of the field and moderate to highly unstable areas along the field boundary, especially in the northern and eastern part. This area of instability is associated with a small gully and an area of heavier clay soils in the field. In dry years, this area has better soil moisture availability, generating higher yields, however in ‘wet’ years its landscape position and heavier soil type is susceptible to water-logging effects. It tends to be either an area of higher or lower yields.

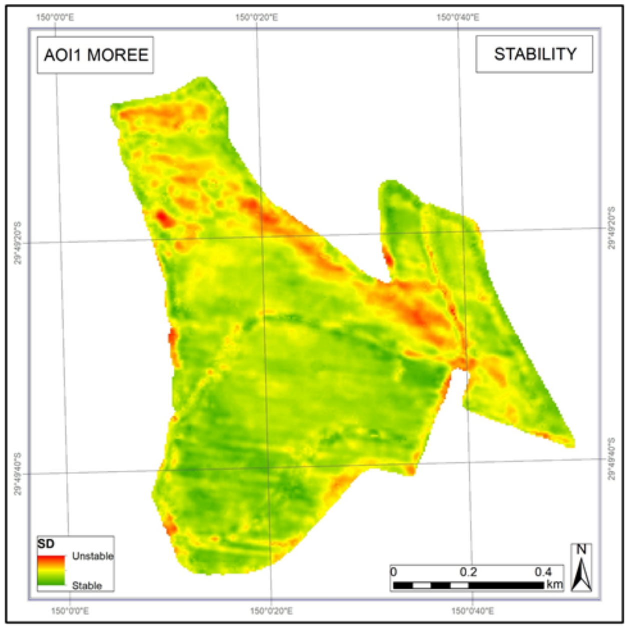
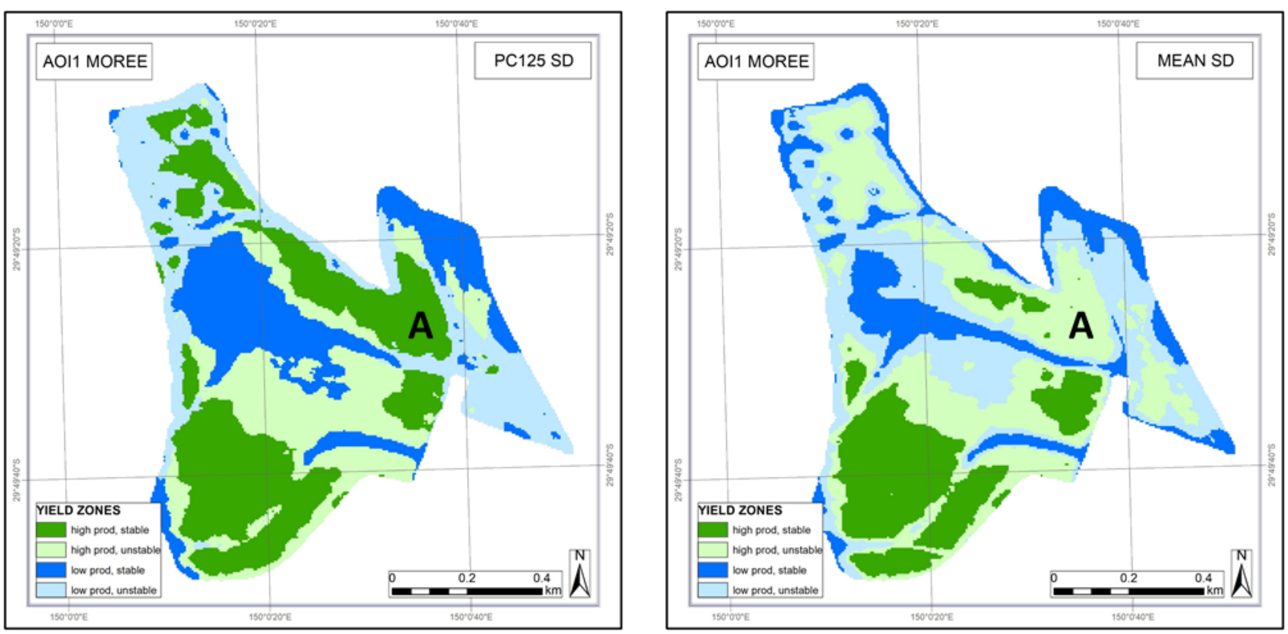


Figure 6: Spatiotemporal stability of yield pattern as SD of all normalised yield maps, showing the spatial yield variability over time.

3.4 Yield Productivity-Stability Zones Delineation

In order to produce information layers to support management decision (especially risk management) in cereal production systems, the MYPA approach using pattern recognition and image analysis techniques on agronomic yield data was developed and compared to a "traditional" statistical approach based on per-pixel calculation of the average yield response. For the demonstration field AOI1 Moree, site-specific management zones could be derived from both approaches, considering in both the productivity and stability of yield over space and time (Figure 7).



(**a**) (**b**)

Figure 7: Yield productivity-stability zones – (a) based on PC1, PC2, PC5, and SD; (b) based on MEAN and SD.

Similar size, shape, and distribution of yield zones across all four productivity-stability classes were found in the southern part of Moree field and considerable differences in the center as well as in the northern and western parts. With the MYPA approach, larger stable areas were mapped at the field center with low productivity and at the western and northern parts with higher productivity. Overall, the MYPA approach produced more concise and manageable yield zones than the statistical approach. This would indicate that the MYPA pattern is more amenable to site-specific management and technical constraints (Tisseyre and McBratney 2008) although a formal analysis of this has not yet been done.

There is one large area of discrepancy between the two zoned outcomes, with the area labelled '**A**' in Figure 7 being classed as stable with the MYPA approach and unstable In the MEAN approach. This area is characterized by deeper, clay soils that tend to yield the highest in a normal year (e.g. 1998, 2000) but are prone to waterlogging in 'wet' years, with significant in-season rainfall (e.g. 2004), and have a relatively lower yield in those years. In the 'wet' years, the shallower, lower clay soils in other areas of the field tend to improve their yield. However, while the relative yield is lower in the '**A**' area in 2004, the actual yield tends to be fairly stable, and it was actually slightly higher in this area in 2004 compared to 2000. The MYPA has identified that this area is stable in absolute terms, if not in relative terms between normal and 'wet' years.

Tables 3 and 4 depict the average (mean) of yearly yield production according to crop type, yield variability over time (SD), and pattern detection variables (PC1, PC2, PC5, MEAN) for each derived yield zone based on zonal statistic calculation of all pixel values within the corresponding k-means cluster. The quality measures of MANOVA – Pillai's Trace and Wilks' Lambda – (Table 5) reveal that the MYPA model is statistically the better model for yield data than the statistical approach. Generally, the lower Wilks' Lambda and the higher Pillai Trace are the better fits the model to the analyzed data. Accordingly, the MYPA approach surpasses the MEAN approach (Table 5).

Table 3: Zonal statistics of yield data [t/ha] within the k means clusters based on MEAN and SD.

Cluster	Interpretation	Sorghum 1998	Chickpea 1999	Wheat 2000	Wheat 2004	Sorghum 2005	Chickpea 2006	MEAN	SD
Dark green	high; stable	5.99	1.56	3.23	5.36	4.07	2.14	1.73	0.102
Light green	high; unstable	5.69	1.42	2.78	5.00	3.74	1.72	1.61	0.154
Dark blue	low; stable	4.62	0.95	1.87	4.26	2.84	1.37	1.37	0.135
Light blue	low; unstable	5.16	1.22	2.24	4.68	3.33	1.58	1.50	0.140

Table 4: Zonal statistics of yield data [t/ha] within the k means clusters based on PC1, PC2, PC5, and SD.

Cluster	Interpretation	Sorghum 1998	Chickpea 1999	Wheat 2000	Wheat 2004	Sorghum 2005	Chickpea 2006	PC1	PC2	PC5	SD
Dark green	high; stable	5.22	1.30	2.51	5.35	3.54	1.92	0.03	-0.57	0.02	0.129
Light green	high; unstable	6.08	1.55	3.23	5.07	4.01	1.83	1.12	0.08	-0.01	0.134
Dark blue	low; stable	4.43	1.11	1.93	4.67	2.95	1.52	-1.31	-0.24	-0.02	0.131
Light blue	low; unstable	5.63	1.19	2.29	4.21	3.43	1.48	-0.29	0.71	0.01	0.152

Table 5: MANOVA of clusters based on PC1, PC2, PC5, and SD compared to clusters based on MEAN and SD.

Model	Wilks' Lambda	Pillai Trace
MEAN-SD	0.39	0.69
PC125-SD	0.29	0.89

Visually the two approaches to yield zoning presented in Figure 7 show similarities, however the MYPA approach (Figure 7a) has larger, more coherent zones than the MEAN approach (Figure 7b). The MYPA approach, with this example at least, appears to be presenting a better statistical and technical option than the historic statistical approach. Further work will examine if this holds true with other yield time-series and how it compares to other emerging approaches (e.g. LeRoux et al. 2018).

4 Conclusions

This study has shown that multitemporal analysis of agronomic data (here: yield data) using pattern recognition techniques offers high potential to reveal spatiotemporally precise and stable or unstable yield production zones at the local scale. For the detection of outlier yield data sets from multi-year yield maps, PCA was demonstrated as an efficient approach to identify years with management effects (e.g., “one-off” management decisions, double cropping). Compared to common statistical analysis approaches, the MYPA method surpass clearly the delineation of coherent and practical management zones, as demonstrated by MANOVA. Overall, derived information layer might be of great value to assist decision making in cereal production systems.

Acknowledgements

The authors would like to acknowledge the support of Mr Michael Smith for use of his yield data and Dr Brett Whelan from the Precision Agriculture Laboratory, the University of Sydney for providing the yield data. This work was performed as part of a UK Newton Fund Project to look at methods for improving Precision Agriculture in Family-sized Farms with an emphasis on Chinese systems.

References

Abdel-Kader, F. H. (2011). Digital soil mapping at pilot sites in the northwest coast of Egypt: A multinomial logistic regression approach. *The Egyptian Journal of Remote Sensing and Space Science*, 14, 29-40.

ASRIS (2011). ASRIS - Australian Soil Resource Information System. Available at http://www.asris.csiro.au (verified 27 April 2018).

Australian Bureau of Meteorology (2016). Climate statistics for Australian locations. Summary statistics MOREE AERO. Available at http://www.bom.gov.au/climate/averages/tables/cw_053115.shtml (verified 27 April 2018).

Basso, B., Cammarano, D., Carfagna, E. (2013). *Review of crop yield forecasting methods and early warning systems. Improving methods for crops estimates.* Rome, Italy: FAO Publication.

Blackmore, B. S., Godwin, R. J., Fountas, S. (2003). The Analysis of Spatial and Temporal Trends in Yield Map Data over Six Years. *Biosystems Engineering*, 84(4), 455-466.

Blackmore, B. S. & Moore, M. (1999). Remedial correction of yield map data. *Precision Agriculture*, 1, 53-66.

Blasch, G., Spengler, D., Hohmann, C., Neumann, C., Itzerott, S., Kaufmann, H. (2015a). Multitemporal soil pattern analysis with multispectral remote sensing data at the field-scale. *Computers and Electronics in Agriculture*, 113, 1-13.

Blasch, G., Spengler, D., Itzerott, S., Wessolek, G. (2015b). Organic matter modelling at the landscape scale based on multitemporal soil pattern analysis using RapidEye data. *Remote Sensing*, 7, 11125-11150.

FAO, ISRIC, ISSS (1998). *World Reference Base for Soil Resources. World Soil Resources Report No.84.* Rome, Italy: FAO Publication.

Leroux, C., Jones, H., Taylor, J. A., Tisseyre, B., Clenet, A. (2018). A zone-based approach for processing and interpreting variability in multi-temporal yield data sets. *Computers and Electronics in Agriculture*, 148, 299-308.

Leutner, B., Horning, N., Schwab-Willmann, J. (2018). RStoolbox: Tools for Remote Sensing Data Analysis. R package version 0.2.1. Available at https:// CRAN.R-project.org/package=RStoolbox (verified 27 April 2018).

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2018). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1. Available at https://cran.r-project.org/web/packages/cluster/index.html (verified 27 April 2018).

McBratney, A. B., Whelan, B., Shatar, T. (1997). Variability and Uncertainty in Spatial, Temporal and Spatiotemporal Crop-Yield and Related Data. Ciba Foundation symposium. 210. 141-60. 10.1002/9780470515419.ch9.

Meroni, M., Marinho, E., Sghaier, N., Verstrate, M. M., Leo, O. (2013). Remote Sensing Based Yield Estimation in a Stochastic Framework - Case Study of Durum Wheat in Tunisia. *Remote Sensing*, 5, 539-557.

Minasny, B., McBratney, A. B., Whelan, B. M. (2005). VESPER version 1.62.Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney, NSW 2006. Available at https://sydney.edu.au/agriculture/pal/software/vesper.shtml (verified 27 April 2018).

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114, 358-371.

Northcote, K. H., Beckmann, G. G., Bettenay, E., Churchward, H. M., Van Dijk, D. C., Dimmock, G. M., Hubble, G. D., Isbell, R. F., McArthur, W. M., Murtha, G. G., Nicolls, K. D., Paton, T. R., Thompson, C. H., Webb, A. A., Wright, M. J. (1960-1968). *Atlas of Australian Soils, Sheets 1 to 10. With explanatory data.* Melbourne, Australia: CSIRO Aust. and Melbourne University Press.

Panda, S. S., Hoogenboom, G., Paz, J. O. (2010). Remote Sensing and Geospatial Technological Applications for Site-specific Management of Fruit and Nut Crops: A Review. *Remote Sensing*, 2, 1973-1997.

Peel, M. C., Finlayson, B. L., McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences Discussions*, 11, 1633-1644.

Pringle, M. J., McBratney, A. B., Whelan, B. M., Taylor, J. A. (2003). A preliminary approach to assessing the opportunity for site-specific crop management in a field, using yield monitor data. *Agricultural Systems*, 76, 273-292.

R Core Team (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/ (verified 27 April 2018).

Rembold, F., Atzberger, C., Savin, I., Rojas, O. (2013). Using Low Resolution Satellite Imagery for Yield Prediction and Yield Anomaly Detection. *Remote Sensing*, 5, 1704-1733.

Shanahan, J. F., Schepers, J. S., Francis, D. D., Varvel, G. E., Wilhelm, W. W., Tringe, J. M., Schlemmer, M. R., Major, D. J. (2001). Use of remote-sensing imagery to estimate corn grain yield. *Agronomy Journal*, 93, 583-589.

Stafford, J. V., Ambler, B., Lark, R. M., Catt, J. (1996). Mapping and interpreting the yield variation in cereal crops. *Computers and Electronics in Agriculture*, 14, 101-119.

Smith L. I. (2002). A tutorial on principal components analysis. Cornell University, USA. p. 27.

Taylor, J. A., McBratney, A. B., Whelan, B. M. (2007). Establishing Management Classes for Broadacre Agricultural Production. *Agronomy Journal*, 99, 1366-1376.

Tisseyre, B., McBratney, A. B. (2008). A technical opportunity index based on mathematical morphology for site-specific management: An application to viticulture. *Precision Agriculture*, 9, 101-113. 10.1007/s11119-008-9053-5.