Multi-temporal yield pattern analysis method for deriving yield zones in crop production systems

Gerald Blasch, Zhenhai Li, James Taylor

▶ **To cite this version:**

Gerald Blasch, Zhenhai Li, James Taylor. Multi-temporal yield pattern analysis method for deriving yield zones in crop production systems. Precision Agriculture, inPress, 10.1007/s11119-020-09719-1 . hal-02942706

HAL Id: hal-02942706
https://hal.inrae.fr/hal-02942706

Submitted on 18 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Precision Agriculture
https://doi.org/10.1007/s11119-020-09719-1

Multi-temporal yield pattern analysis method for deriving yield zones in crop production systems

Gerald Blasch[1,2] · **Zhenhai Li**[3,4] · **James A. Taylor**[1,5]

© The Author(s) 2020

Abstract

Easy-to-use tools using modern data analysis techniques are needed to handle spatio-temporal agri-data. This research proposes a novel pattern recognition-based method, Multi-temporal Yield Pattern Analysis (MYPA), to reveal long-term (> 10 years) spatio-temporal variations in multi-temporal yield data. The specific objectives are: i) synthesis of information within multiple yield maps into a single understandable and interpretable layer that is indicative of the variability and stability in yield over a 10 + years period, and ii) evaluation of the hypothesis that the MYPA enhances multi-temporal yield interpretation compared to commonly-used statistical approaches. The MYPA method automatically identifies potential erroneous yield maps; detects yield patterns using principal component analysis; evaluates temporal yield pattern stability using a per-pixel analysis; and generates productivity-stability units based on *k*-means clustering and zonal statistics. The MYPA method was applied to two commercial cereal fields in Australian dryland systems and two commercial fields in a UK cool-climate system. To evaluate the MYPA, its output was compared to results from a classic, statistical yield analysis on the same data sets. The MYPA explained more of the variance in the yield data and generated larger and more coherent yield zones that are more amenable to site-specific management. Detected yield patterns were associated with varying production conditions, such as soil properties, precipitation patterns and management decisions. The MYPA was demonstrated as a robust approach that can be encoded into an easy-to-use tool to produce information layers from a time-series of yield data to support management.

Keywords Image analysis · Spatio-temporal variability · k-means clustering · Principal component analysis

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11119-020-09719-1) contains supplementary material, which is available to authorized users.

✉ Gerald Blasch
g.blasch@cgiar.org

Extended author information available on the last page of the article

Published online: 06 May 2020

Introduction

Yield data were one of the earliest forms of precision agriculture (PA) data available to growers and researchers, with Global Navigation Satellite System (GNSS) enabled grain yield monitors becoming commercially available in the early 1990s. The spatial variation displayed in these early grain maps was certainly a principal driver for the initial development of site-specific crop management. Grain yield mapping remains widely accessible to growers in mechanized production systems and is the only definitive method for spatially auditing the actual at-harvest yield. Despite this, the adoption of yield mapping has been relatively slow (Bramley 2009; Fountas et al. 2005; Kutter et al. 2011; Say et al. 2018; Schimmelpfennig and Ebel 2011), and a recent survey in Australia (Bramley and Ouzman 2018) has identified that yield maps are considered to be under-used by the industry and growers. Despite this, Bramley and Ouzman (2018) reported that the adoption and use of yield maps was highly "valuable as a lever to gaining 'buy-in' to sensing and PA more broadly".

There are several complementary reasons for yield data being under-utilised.

(i) Yield data can be inherently noisy and there are a variety of operational and calibration conditions that contribute to the noise in the data (Leroux et al. 2018a; Lyle et al. 2013). Noisy data are often difficult to interpret when mapped, and while many approaches for data-filtering and mapping have been proposed (e.g. Lyle et al. 2013; Robinson and Metternicht 2005; Sudduth and Drummond 2007; Sun et al. 2013 among others), these have not been well translated into industry services for growers to date (Leroux et al. 2018b).

(ii) Spatial yield patterns are also often affected by managerial as well as environmental effects. Local expertise is therefore often needed for correct interpretation, which requires a time investment from the grower and the service provider. Yield maps can also be potentially unstable, particularly in dryland systems, where differences in the amount and timing of precipitation often influences yield patterns (see Taylor et al. 2007 for examples).

(iii) Increasing time-series may make interpretation more difficult, not easier, in the absence of clear agri-diagnostic tools to simplify the multi-seasonal yield data. Even if data are correctly processed and local expertise is available to interrogate the yield maps, adding additional layers each year/season can make long-term visual and statistical interpretation more difficult.

All these issues mean that interpreting yield patterns, and therefore deriving spatial decisions from spatio-temporal yield data, is rarely a straightforward process. In the first instance, growers would like to know where yield is temporally consistent; consistently high, low or average, and where yield response is more variable over time, i.e. inconsistent.

From an early stage of yield monitor data analysis, it has been recognized that the inter-annual temporal variability in yield patterns may be as important as annual spatial yield variability (McBratney et al. 1997). Based on per-pixel calculations, multiple studies in the late 1990s and early 2000s investigated temporal variation using local yield means and variance statistics on relatively short time series of yield maps (2–6 years) (Blackmore and Moore 1999; Blackmore et al. 2003; Li et al. 2007; Ping and Dobermann 2005; Ping et al. 2005; Stafford et al. 1996). These statistical approaches assumed that crop yield response and its spatial distribution are mainly affected by permanent soil characteristics,

which leads to the same spatial yield pattern in each year (Blackmore et al. 2003). However, non-stable factors (e.g. climate patterns, crop type and rotation, soil water content, infiltration and crop water use) also have a significant influence on spatio-temporal yield variability and interact dynamically with spatially stable properties (e.g. soil texture, landscape position) (Basso et al. 2012). Such interactions impose a clear limitation on statistical approaches to analysing spatio-temporal yield variability. However, it should be noted that the analysis of the variability in multi-year yield data is not necessarily only limited to the coherence and stability of yield patterns. Spatio-temporal yield data has also been used to inform other decision models, such as drainage effects in irrigated corn (Delbecq et al. 2012), however, the models used in such approaches are not valid for yield pattern analysis.

Most previous research using yield data has focused on spatial analysis of yield and not on the temporal or spatio-temporal variability in yield data (Griffin et al. 2008; McKinion et al. 2010; Stafford et al. 1996). In part, this may be explained by a realization that the early temporal studies were data-limited and based on some assumptions of stability between crop types and rotations (Bakhsh et al. 2000; Kitchen et al. 2005; Ping and Dobermann 2005). These short time series were unlikely to capture the full range of potential crops, rotations and potential production climatic conditions. This limitation no longer exists and time series of yield data > 10 years and across multiple rotations are commonly available (e.g. Leroux et al. 2018b) when working with growers who were early adopters of grain yield monitoring in the 1990s and early 2000s. These longer time series form a more coherent and robust data set, which should facilitate a better understanding of temporal and spatio-temporal variation in production. It also permits temporal analyses to be targeted to specific conditions e.g. a specific crop type or specific climatic pattern (Leroux et al. 2018b. However, as noted above, longer time series pose a problem for visual interpretation. Interpreting 10–20 + individual yield maps is not an easy task, especially if the maps contain noise and artefacts.

To realize the potential of having long time series of high spatial resolution yield data, easy-to-use tools with modern data analysis techniques are needed to handle these data (Bramley and Ouzman 2018). Such tools must support high throughput computation and deliver interpretable and understandable results to producers and agronomists at a low computational cost (Bauckhage and Kersting 2013; Raj et al. 2015). Despite the recognized need, methods to simplify and interrogate observed (spatio-) temporal yield variation for site-specific decision-making are not available to growers. Recent papers have revisited this need: Leroux et al. (2018b) worked from a presumption that the yield data are cleaned but not interpolated, i.e. they exist as irregular points, and using a multivariate segmentation approach. Layton et al. (2019) developed a probability-based approach for identifying yield productivity zones, assuming that the yields corresponding to a given yield productivity zone behave similarly and therefore derive from the same probability distribution. Other studies have formed some assumption between crop vigour and yield and used multi-temporal satellite imagery to interrogate temporal variation in production (e.g. Boydell and McBratney 2002; Georgi et al. 2018; Tagarakis et al. 2013). These approaches tend to use image analysis techniques (or derivatives). When yield is correctly interpolated onto a regular grid, an image is generated. It is therefore possible to directly apply image analysis techniques to interpolated yield dat—either by assuming each map is a single-band image or each map is a band within a constructed multi-band image.

Generally, digital image processing and analysis in the remote sensing discipline uses simple or complex computer algorithms to reduce noise in data and prepare images for subsequent processing, information extraction and interpretation. Simple arithmetic pixel operations (e.g. data sums and differences, colour mixing, multispectral ratio and

vegetation indices) and more complex pixel operations (e.g. spectral transformations) allow for image enhancement, feature extraction and pattern recognition by exploiting the resulting (sometimes synthetic) images (Albertz 2009; Sabins 1996; Schowengerdt 2007). In a multi-temporal context (e.g. satellite time series analysis), these spectral operations are highly suitable for the evaluation of spatio-temporal pattern variability and change detection (e.g. Blasch et al. 2015a, b). They should equally be suitable to the analysis of yield map time-series; however, such an approach has yet to be tested.

To address the limitation of commonly-used purely statistical approaches, the need for easy-to-use tools capable of the efficient handling of high-resolution yield data and long yield map time series, and the demand of interpretable and understandable results for growers and agronomists, this study proposes a novel pattern recognition-based method—the Multi-temporal Yield Pattern Analysis (MYPA)—to reveal the long-term variation in yield, expressed in categorical productivity-stability units (PSUs). A PSU is a relatively homogenous unit in terms of its relative productivity (yield) each year and the temporal stability of this relative yield. The overall goal is to develop an alternative method to purely statistical approaches, based on modern image analysis techniques. The specific objectives are: to i) detect erroneous data sets within the yield map time series, ii) synthesis the information within multiple yield maps into a single understandable and interpretable layer that indicates the relative yield response and its stability over a 10 + years period by using principal component analysis (PCA), statistical per-pixel calculations and *k*-means clustering, and iii) evaluate the hypothesis that the novel MYPA approach enhances insights into spatio-temporal yield variation compared to existing statistical approaches. To evaluate the hypothesis, the MYPA output was compared and validated to a statistical approach (STAT), adopted from the methods proposed by Blackmore et al. (2003) and Ping and Dobermann (2005).

Materials and methods

Sites description

Four fields were chosen for validation of the MYPA. The fields were characterized by having a relatively long time series of available yield data (10 + years). Two relatively large fields (Field 1: 80 ha; Field 2: 132 ha) were available from the cropping belt near Gurley (− 9.8204, 150.0314) in North-West New South Wales, Australia, and two smaller fields (Field 3: 21 ha; Field 4: 25 ha) from a lowland arable farm near Warkworth (55.3597, − 1.6485) Northumberland, UK. These provided two contrasting production systems on which to test the methodology. The Australian fields, with predominantly heavy cracking clays, are dryland production systems in a humid, subtropical zone with warm to hot summers (average annual maximum temperature of 26.1 °C) and summer dominant rainfall. The annual average precipitation is ~ 650 mm. Winter cropping systems are short (~ 6 months) and reliant on stored moisture rather than in-season precipitation. Wheat production is strongly water-limited and typically 2–5 ton ha^{-1} on average. The UK fields mainly contained loamy and light clayey soils with slow hydraulic permeability. The fields are in a cool-climate production system (annual average maximum temperature is 12.2 °C) with regular monthly rainfall dispersed throughout the year (annual average precipitation of 690 mm). The growing season for winter crops is longer (~ 9 to 11 months) and wheat yields tend to be 7–10

ton ha^{-1} on average. Yield is strongly temperature/radiation limiting, and despite the small difference in annual precipitation with the Australian sites, waterlogging is often more problematic than water-deficit due to the much lower levels of evapotranspiration in northern England.

Pre-processing of yield data and yield map generation

Yield maps were obtained for a 10 + year period for all four fields. Not every year had available yield data, either due to crop failure (Australian fields), fallow rotations, alternative crops (particularly potatoes in the UK) or poor yield sensor operation (missing data). Yield data from the Australian fields was not available after 2006 due to a reorganization of the field boundaries. If the yield data for a year was missing, nonsensical or not relevant, then it was not included in the analysis. The decision on which years to include, or not include, was subjective and involved some local knowledge. Table 1 lists the yield data from grain and seed crops that were considered suitable after discussion with the growers for mapping and use in the analysis.

The following data pre-processing steps are based on the protocol of Taylor et al. (2007) for establishing management classes, developed for broadacre production systems at The University of Sydney. The yield data were filtered and cleaned to remove outliers and erroneous values before interpolation. Due to differences in mean yield between the Australian and UK fields and between canola and wheat in the UK, potential yield values were set to 0.1–10 ton ha^{-1} for all Australian fields and canola fields in the UK and to 1–20 ton ha^{-1} for UK wheat fields. Values outside these ranges were considered improbably low, as even bare areas in fields still record yield due to deconvolution within the combines (Whelan and McBratney 2000), or impossibly high based on known yield potentials and removed as a first step in data clean-up. Each field was then individually trimmed to values within $\pm 2.5\sigma$ (σ = standard deviation) of the field mean. This follows the simple filtering approach proposed by Taylor et al. (2007) that was derived using similar yield data sets. Interpolation onto a 5 m grid was done using block-kriging (20 m blocks) with a local variogram using the Vesper shareware (Minasny et al. 2005). Statistics of the filtered yield are shown in Table 1. It is recognized that the approach of Taylor et al. (2007) is simplistic and will not identify local spatial outliers and that other more sophisticated yield filtering approaches are now available (e.g. Leroux et al. 2018a; Vega et al. 2019). However, the use of local block kriging interpolation will provide additional smoothing of the data and has been shown to diminish the effect of local spatial outliers in the maps (Whelan and McBratney 1999). There are various ways that the yield data could be pre-processed, but the key points are that a well understood and robust method was used and that the same interpolated data will used as an input into the different methods of deriving yield patterns. The interpolation grid was kept constant for each year so that all yield predictions were co-located at the end of the process and the same yield maps were used for all comparisons. Georeferenced single-band raster data sets (saved in the GeoTIFF format) were generated for each year yield map using R (R Core Team 2018) and the R package "sp" (Bivand et al. 2013; Pebesma and Bivand 2005). These raster data sets were stacked to a multi-temporal layer stack for subsequent use as input data.

Table 1 All study fields—descriptive statistics of yield data collected from each year, including crop harvested. All data expressed as **ton ha^{-1}**

Yield (ton ha^{-1})	Mean	Standard deviation	Minimum	Maximum
Field 1—Australia				
Wheat 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
Field 2—Australia				
Sorghum 1998	4.41	1.18	0.27	7.76
Wheat 1999	4.28	1.03	0.46	6.66
Sorghum 2001	3.73	1.11	0.07	7.09
Wheat 2003	4.04	1.01	0.30	8.81
Wheat 2005	3.93	0.73	1.23	7.17
Wheat 2006	2.43	0.44	0.77	3.73
Field 3—UK				
Wheat 2002	8.24	0.77	5.85	10.44
Canola 2004	3.65	0.76	1.90	6.17
Canola 2007	4.15	0.20	3.15	4.79
Wheat 2008	10.24	0.97	6.51	12.01
Canola 2010	3.73	0.37	2.56	4.75
Wheat 2014	8.99	0.86	5.60	10.80
Canola 2016	2.71	0.23	1.77	3.30
Field 4—UK				
Wheat 2004	7.68	1.07	4.49	10.32
Wheat 2007	9.57	0.81	5.58	11.30
Canola 2011	5.23	0.33	3.39	6.30
Wheat 2012	9.94	1.12	5.61	13.12
Canola 2014	3.84	0.32	2.53	4.62
Wheat 2015	12.58	1.39	7.47	15.71

Multi-temporal yield pattern analysis workflow

The intent of the analysis is to synthesis the information within multiple yield maps into a single layer as a final output/map.

The MYPA process uses the multi-temporal layer stack for both pattern detection and pattern stability analysis before determining yield PSUs. The four processing steps: (1) Data Preparation, (2) Pattern Detection, (3) Pattern Stability Analysis and (4) Delineation of PSUs are described in the following sections. The MYPA final output is a map representing yield PSUs (Fig. 1a) that indicates the relative yield response (e.g. high, low or medium) and how stable that response is (i.e. the certainty associated with it). While the final output is a single image, intermediate outputs are also generated and can also be interrogated for relative yield only or stability only. The process is designed to be as automated as possible once a decision on suitable yield maps (and correct yield interpolation) has been

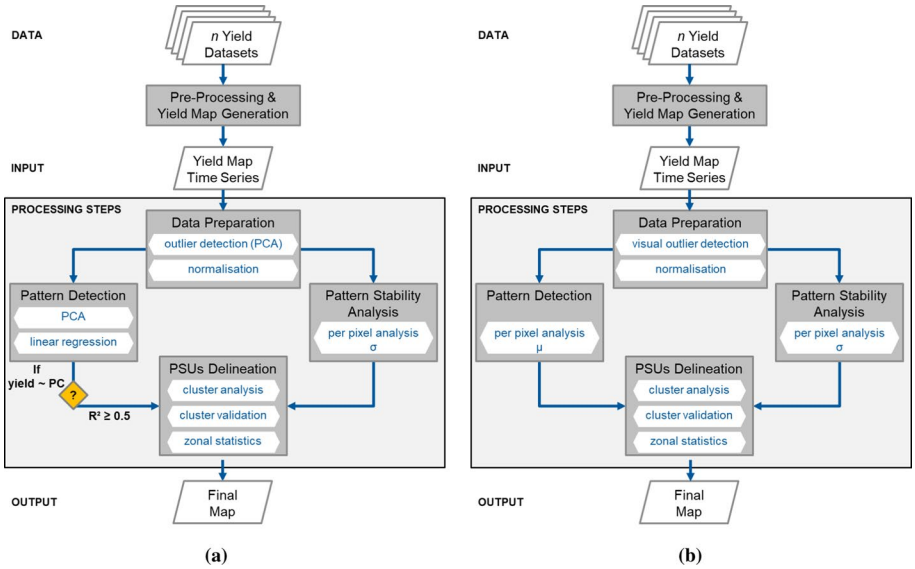

Fig. 1 Schematic illustrating the workflow and processing steps of **a** the proposed MYPA approach and **b** the conventional statistical-based STAT approach, including data preparation and key actions, to generate a final map

performed. Overall, the MYPA algorithm was implemented in R (R Core Team 2018)—a free software environment for statistical computing—using specific R packages for spatial and image analysis and raster/vector data such as "cluster" (Maechler et al. 2018), "gstat" (Pebesma 2004; Graeler et al. 2016), "raster" (Hijmans 2017), "rgdal" (Bivand et al. 2018), "RStoolbox" (Leutner et al. 2018) and "sp" (Bivand et al. 2013; Pebesma and Bivand 2005), and basic R packages for plotting ("ggfortify"; Tang et al. 2016; Horikoshi and Tang 2016) and statistical analysis ("stats"; R Core Team 2018). Maps were generated in QGIS (QGIS Development Team QGIS Development 2009) and all plots in R.

Data preparation

In order to analyse multi-temporal yield data from the same field, two main production constraints must be overcome: (a) the crop rotation effect and (b) the presence of erroneous data sets. The crop rotation effect refers to the varying absolute yield from year to year due to changing crop types within a specific crop rotation system or due to varying inter-annual weather conditions. Table 1 shows clearly that years with wheat or sorghum or canola have relatively high absolute yield values compared to years with chickpea (Australia: wheat: 0.3–8.8 ton ha^{-1}, sorghum: 0.1–7.8 ton ha^{-1}, chickpea: 0.1–2.6 ton ha^{-1}; UK: wheat: 4.5–15.7 ton ha^{-1}, canola: 1.8–6.3 ton ha^{-1}). Generally, the physical soil properties (e.g. organic matter content, soil texture) of a field are relatively stable over the analysis period of 10 + years, if no major erosion events or extensive land use changes occur. Secondly, unusual or outlying data sets might exist in the time series. If these are erroneous due to specific management effects (e.g. one-off incorrect or unusual management decisions) and/or sensor/calibration errors, they must be eliminated from the time series in order to avoid "contaminated" results. However, unusual data sets originating from soil, weather, site-specific conditions or variable weather patterns should remain in the time series.

To ensure a yield map time series free of erroneous yield map(s) for subsequent processing, data normalization and semi-automated map outlier detection are performed on the layer stack. This ensures that maps will exhibit the same range of yield response, without any influence of different crop types or crop rotation effects (Blackmore 2000). The two main goals of the data preparation step are;

a) Data normalization to permit comparison of yield from different crops with potentially very different absolute yield values (Li et al. 2007; Schenatto et al. 2017). For each layer (year), the yield maps were normalized to the maximum observed yield in that year (Table 1) using Eq. 1. This results in all maps having normalized yield values expressed between 1 and 2.

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

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 from the same yield map within data set. A constant of 1 was added to the equation so that the output is in the range [1,2] and suitable for any type of subsequent analysis, i.e. it cannot contain zero values. The natural variation in yield originating from soil properties, other site-specific conditions (e.g. relief, hydrology) or weather are still preserved in the data.

b) The identification and elimination of outlier yield maps from the yield map time series; years with unusual yield data should be identified and omitted before analysis in consultation with the grower/agronomist. However, this may not always be done properly. To validate the selected yield maps, a PCA was performed on the multi-temporal layer stack of all yield maps (see Yield Pattern Detection for PCA description).

Yield pattern detection

To detect spatio-temporal yield patterns, a transformation using PCA (R package "RStoolbox"; Leutner et al. 2018) was used. The normalized yield maps form an image where each yield map layer is one spectral band in the image and each pixel is considered as a discrete point with n years of yield data. The PCA transformation—a linear transformation of the (usually) correlated original spectral bands—is a common complex pixel operation in image analysis. It allows the analysis of highly correlated multi-dimensional data to remove redundant information, isolate noise components and reduce dimensionality without significant information loss (Panda et al. 2010). By rotating original spectral bands to a data variance maximum, the uncorrelated output bands (principal components—PCs) allow for the identification of patterns in the data, here yield data. The PCs highlight similarities and differences between bands (maps) to reveal the underlying dimensionality in the multi-band image. They also form a method of compressing the original image data (Abdel-Kader 2011; Mulla 2013; Smith 2002). This makes PCA a powerful tool for pattern recognition. Compared to the classic statistical approaches, as proposed by Blackmore et al. (2003), the PCA approach is free of any one or more factorial assumptions regarding inter-annual repetitiveness of yield response and its spatial distribution.

Following the PCA, plots of the first and second PCs (PC1 and PC2) and the eigenvectors associated with each layer (yield map) can be used to quickly and easily identify potential issues. If an outlying map is detected, the reasons for it being an outlier are

investigated. If it is considered to be erroneous (e.g. due to specific management effects, sensor/calibration errors), it is removed, usually after consultation with an expert with local field knowledge, such as the grower or agronomist. If data is removed, the PCA is repeated on the remaining layers until a stable solution (no clear outlying maps) is found. If a good selection process was employed in the pre-processing, then it is very plausible to have no outliers identified and no data removed at this step. This process is considered useful and necessary because it permits spatial differences in yield response to be simply expressed as a graph. The PC1 vs PC2 graph may help to identify issues not immediately obvious from a visual inspection of the maps. It also permits the user to err towards keeping, instead of omitting, maps in the initial selection of the yield data in the knowledge that truly unusual maps will be removed at this step.

Once a stable outcome from the PCA is achieved, the band (normalized yield) and PC values were extracted for each pixel, a linear regression performed for all combinations of years and PCs and the coefficient of determination (R^2) computed. PCs with at least a moderate relationship ($R^2 \geq 0.5$) with at least one year of data were selected for the final processing according to the rationale that if over half the variation in any one yield map was explained, then the PC was relevant for management and must be included.

Temporal pattern stability analysis

The temporal stability of the normalized yield response was determined by using statistical per-pixel analysis on the layer stack. For each pixel, the standard deviation (σ) of normalized yield was calculated across the multi-year stack. This generated a single final image of the σ of yield response, i.e. the relative stability of normalized yield response, across all analysed field-years.

Yield productivity-stability units delineation

The best performing PCs (§ Yield pattern detection) selected for yield productivity and the overall stability σ (§ Temporal pattern stability analysis) were subsequently stacked into one final layer stack. The *k*-means clustering algorithm (R package "cluster"; Maechler et al. 2018) was used to find groups (clusters) in the unlabelled data of the final stack. Briefly, the *k*-means clustering method aims to allocate the pixel values of desired data layers (here: various PCs and σ) into *k*-classes (clusters) in order to minimize the within-class variability and maximize the differences between the means of the *k*-classes (Hartigan and Wong 1979). Hard and soft (fuzzy) *k*-means clustering are widely used approaches for management class/zone delineation in PA (e.g. Guastaferro et al. 2010; Lark and Stafford 1997; Moral et al. 2010; Morari et al. 2009; Ortega and Santibáñez 2007; Pedroso et al. 2010; Ping and Dobermann 2005; Ping et al. 2005; Rodrigues Junior et al. 2011; Tagarakis et al. 2013; Taylor et al. 2007; Van Meirvenne et al. 2013). Easy-to-manage and coherent management units are desired for site-specific management (Tisseyre and McBratney 2008). Statistical approaches to calculate the optimum solution from fuzzy *k*-means classification algorithms exist (Boydell and McBratney 2002; Córdoba et al. 2013), but an agronomic solution for the optimum number of management units remains somewhat subjective. From the final layer stack, a 2-, 3-, 4-, 5-, 6- and 7-class solution from the *k*-means classification was performed to produce final PSUs options. In order to identify a preferred value for *k*, a validation step was performed using the zonal opportunity index derived from McBratney et al. (2000).

This index is based on the result of a multivariate analysis of variance (MANOVA) and the number and size of discrete PSUs generated from the 2–7-class solutions (details given in § Validation of final map quality and comparison of methods). Once the best solution was identified, PSU mean values of all yield data sets (excluding removed erroneous data sets), selected PCs, and the overall stability (σ) were calculated for each PSU. These PSU statistics were used to assign a descriptive label to each specific PSU indicating both productivity and stability, such as (a) *high productive and stable*, (b) *high productive and unstable*, (c) *low productive and stable* and (d) *low productive and unstable*. To assign the labels, all PSUs were first ranked according to their average yield response across all years and then secondly, the overall stability (high or low) of the two highest yield response clusters were analysed and described. For example, if for a specific cluster, the majority of yield data sets showed a high zonal mean yield response and a low σ, it was assigned to the *high productive and stable unit*, if the majority showed a high yield response and a higher σ than the neighbouring PSU, it is the *high productive and unstable unit*.

Statistical approach for multi-temporal yield data analysis

As the intent is to propose an alternative approach to multi-temporal yield analysis, the MYPA output was compared and validated against a purely statistical approach (STAT), derived from the method of Blackmore et al. (2003) and Ping and Dobermann (2005), applied to the same initial data set.

The STAT approach differs in two main areas, i) the visual outlier detection in the data preparation step, such that the final selected yield data stacks could differ between the two approaches and ii) the yield pattern detection step that does not use any form of data compression or translation (PCA) (Fig. 1b). It only uses a statistical simplification to calculate the pixel mean (μ) across the normalized yield maps. All other processing steps were kept equal to the MYPA approach in order to maintain comparability. Consequently, the σ layer is the same as the layer used for the MYPA approach (§ Temporal Pattern Stability Analysis). A two-layer stack, comprising the μ and σ layers, was generated to delineate PSUs. To make a direct comparison between both approaches possible, the PSUs were derived using the same *k*-means classification as the MYPA approach (and again with a range of 2–7 for *k*) and validated the same way (§ Validation of final map quality and comparison of methods) to assess the STAT PSU outputs. Linear regression analysis was used to reveal the relationships between the individual year yield data, individual PCs and μ.

Validation of final map quality and comparison of methods

To validate the quality of the PSU map outputs and statistically compare which approach delineated more feasible PSUs (PSU_{MYPA} vs. PSU_{STAT}), the following three methods were used: (a) MANOVA, (b) descriptive statistical analysis and (c) zonal opportunity index. The zonal opportunity index is also a key component of the cluster validation step during the PSUs delineation procedure. It allows for the identification of the best class solution (i.e. the most suitable number of PSUs) and the corresponding *k*-level (§ Yield productivity-stability units delineation).

Multivariate analysis of variance

A MANOVA analysis was performed with the yield data as the dependent variables and the derived PSUs from the two different approaches (PSU_{MYPA} and PSU_{STAT}) as the independent variables. MANOVA uses discriminant analysis to reduce the multi-temporal yield data to a single variable that is then related to the yield units of both approaches (Olson 1974). It has been shown to be effective in assessing the quality of zoning in agricultural fields (Ping et al. 2005; Taylor and Whelan 2011; Uribeetxebarria et al. 2018). Pillai's Trace, with a higher value indicating a better model fit, and Wilks' Lambda, with lower values indicating better model fit, were generated to determine how well the MANOVA performed. The MANOVA analysis was conducted in R.

For each field the MANOVA outputs for PSU_{MYPA} and PSU_{STAT} were compared and ranked between each level of k, with a score of 1 assigned to the method with the higher Pillai's value and the lower Wilks' lambda. The mean ranks for both statistics were derived across all k-levels for each individual field. This provided a summary of which method, PSU_{MYPA} or PSU_{STAT}, performed best according to the MANOVA output across all possible outcomes (k[2,7]).

Descriptive statistics of yield productivity-stability units

The final PSU raster maps across all levels of k for both approaches were converted in R into a polygon map and the area of each polygon recorded. Descriptive statistics across the PSUs generated by each method, for each field were calculated. This included the total number of discrete zones generated, the number of zones > 0.25 ha and the mean area of all zones and mean area of zones > 0.25 ha. The choice of the 0.25 ha threshold was made as zones smaller than this are likely to be unmanageable in these cereal systems given the speed and width (often 24 m) of typical, current input applications. However, this threshold is arbitrary, and could be easily altered to accommodate a change in cropping type, e.g. if MYPA is applied into narrow row perennial horticultural systems, or a change in operational resolution of future machinery. From these data, the percentage area of the field that was considered 'unmanageable' was calculated. While the MANOVA tests how well the PSUs account for yield variance, these statistics relate to the technical opportunity to respond to the yield PSU patterning (Tisseyre and McBratney 2008). The presence of many small units would increase the uncertainty associated with variable-rate applications.

Zonal Opportunity Index

To identify the preferred output for each field from the range of potential k-levels [2,7], an adaption of the zonal opportunity index (McBratney et al. 2000), based in turn on the Akaike Information Criterion (AIC) (Akaike 1974), was used

$$O_z = 2z - n_s * ln(1 - \Lambda) \tag{2}$$

where O_z is the Zonal Opportunity Index, z the number of zones generated by each level of k-classification (i.e. discrete units not the number of classes), Λ (Wilks' lambda) from MANOVA is a measure of the percent variance in dependent variables (yields) not explained by differences in levels of the independent variable (classes) thus subtracted from 1 to provide an estimation of 'goodness of fit' and, n_s the number of (independent) data (errors) in the data set. Due to spatial auto-correlation in the data, $n_s < n$ (and n is the

number of grid points in the interpolated data). For the comparison and ranking of k outputs within a field, n_s will be constant. For this analysis, n_s was set at 1000 for all fields, which is approximately 12% of the data in the smallest field (Field 3) and greater than the maximum z observed in the analyses (z = 755, Field 3, PSU_{MYPA}, $k = 7$).

The O_z allowed the k-level outputs for both approaches to be compared within a field. However, the use of a constant n_s value across the four fields restricted comparisons to within, and not between, fields. Approaches to estimating n_s exist (see discussion of Taylor and Bates 2012), but, as the intent here is to identify the preferred solution within a field (and not to try to identify if one field outperforms another), the use of a constant was considered sufficient. The use of the form of the AIC permits the quality of the fit to be corrected against the number of discrete units (zones) generated, with higher numbers of zones penalized as this makes management potentially more difficult (Pedroso et al. 2010). The value of z can be altered to reflect either the total number of zones (polygons) in the output or the total number of 'manageable' zones (here defined as zones > 0.25 ha). For this analysis, both were used in a two-step process. Firstly, the O_z was calculated using the total number of polygons in the output to identify the k-level associated with the best PSU_{STAT} solution based on the whole data set. Then, for this selected level of k, the O_z was recalculated and compared for the PSU_{STAT} and PSU_{MYPA} approach based on the number of 'manageable' zones (i.e. polygons > 0.25 ha in the output). This was termed zonal opportunity of manageable zones ($O_{z\text{-man}}$) and considered a more rigorous comparison as the intent is to identify PSUs that could be managed agronomically, and not to find a purely statistical solution. The percentage area associated with small, 'unmanageable' zones was also considered when comparing the outputs and, if it was > 10% of the total area, then the nearest k-level with an 'unmanageable' area < 10% was also considered in the analysis.

Results and discussion

Data preparation

The normalized interpolated yield maps for Field 1—Australia are shown in Fig. 2 (other normalized map series are presented in Supplementary Data S1a–c). Field 1—Australia has been chosen as an illustration as it contains maps that appear to have linear and unusual features (particularly 1997 and 2006). The field had seven years of yield data between 1997 and 2006 with wheat, sorghum and chick-pea crops (Table 1, Fig. 2). The plotting of PC1 vs. PC2 and the eigenvectors (Fig. 3a) indicated that the 1997 yield map (Fig. 2a) was an outlier in the yield map stack. In 1997, the northern part of the field had a very low yield response relative to the southern part that does not repeat in any other year. This was due to different management in the northern section creating lower-yielding conditions. As this management was a 'one-off', and not intended to be repeated, this information is not relevant to future management. The 1997 yield data was removed and the PCA repeated, which showed no remaining outlying years (Fig. 3b). However, there is some management effect in the Sorghum 2006 data (Fig. 2g). The 2006 yield map was similar to the Wheat 2004 map (Fig. 2e) in the attribute space (Fig. 3b). This map could have been removed based on local opinion and an apparent observable management effect (Fig. 2g). However, since statistically it was not mapped as an outlier, the decision was made to include it in the subsequent analysis. The PCA plots of the other three fields did not indicate any outlying yield maps (Fig. 4).

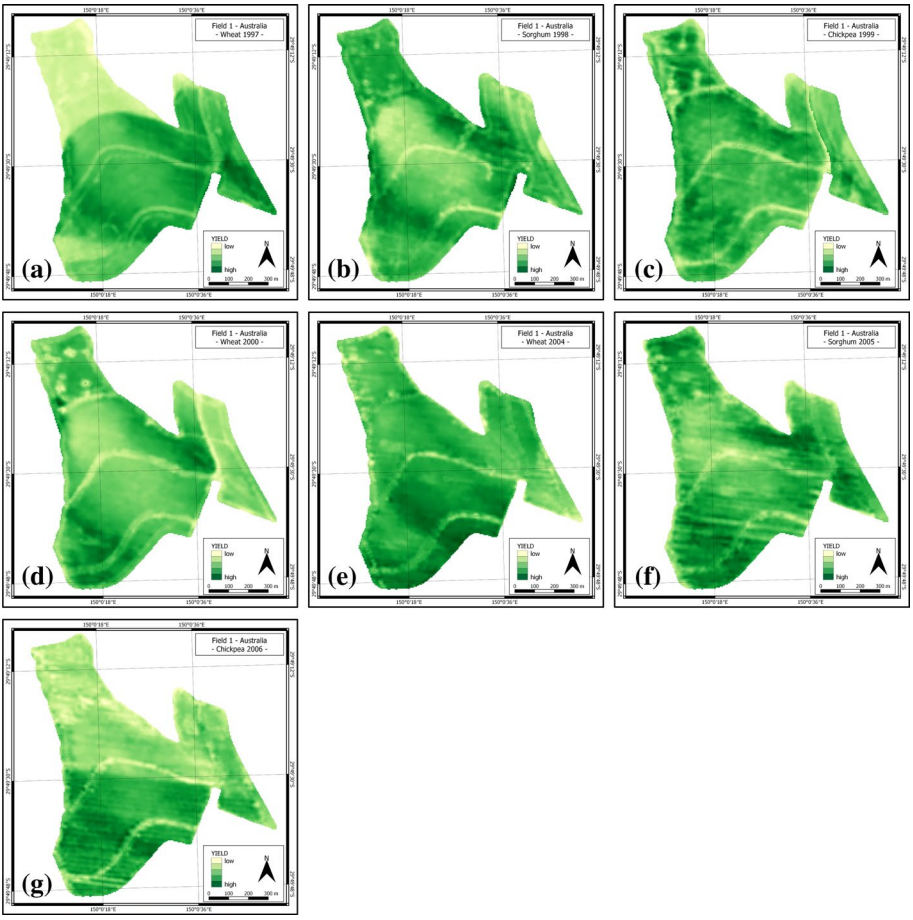

Fig. 2 Field 1—Australia: Normalized yield maps with varying crop types (Sorghum, Chickpea and Wheat) including the outlying yield map (wheat 1997)

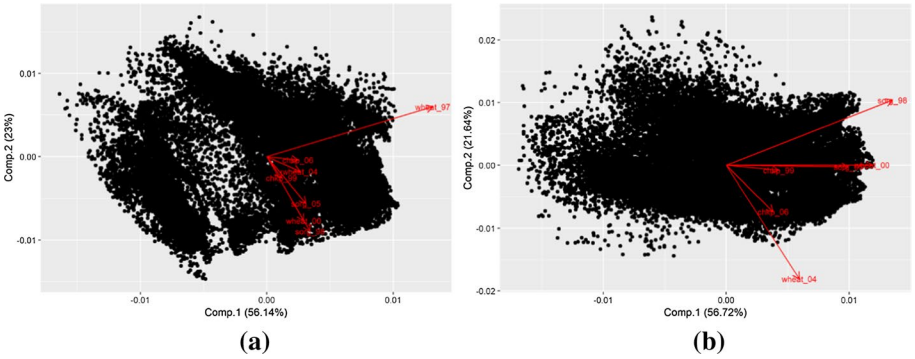

Fig. 3 Field 1—Australia: Eigenvectors (red arrows) and data values by PC1 and PC2 (black points)—**a** of all yield data sets after first iteration; **b** of remaining yield data sets after second iteration of the outlier detection process (Color figure online)

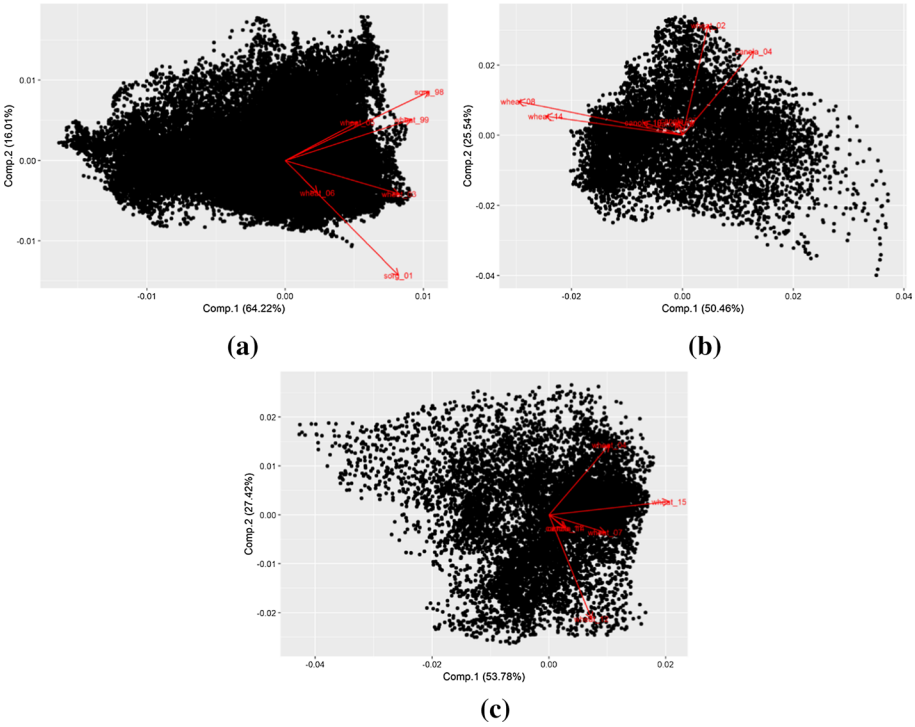

Fig. 4 Eigenvectors (red arrows) and data values by PC1 and PC2 (black points) of all yield data sets after first iteration: **a** field 2—Australia; **b** Field 3—UK; (c) Field 4—UK (Color figure online)

Yield pattern detection

From the PCA, the individual PCs were extracted and mapped. The first PC extracted a pattern that best describes the dominant pattern across *all years*. This pattern was then removed, and the second PC found the dominant pattern in the remaining data and so forth. Again, Field 1—Australia is used as an example to visualize the individual PCs arising from the PCA (Fig. 5) (other PC map series are presented in Supplementary Data S2a-c). Table 2 shows the R^2 values associated with a linear regression between the individual PCs and the individual yield maps for all 4 fields as well as the percentage of total variation explained by each PC. The R^2 values between the annual yields and the mean yield map (μ map) from the STAT method are also shown.

The PC maps tended to show an increasing trend in short-range variation as PC number increased. This is typical as the first PCs tend to extract the larger trends and patterns in the data, leaving the later PCs to explain more random effects and a much lower percentage of the overall variation (Table 2). The first two PCs explained 76-81% of the yield variation in each year (Table 2).

Using the PCA-based pattern recognition approach, relatively good relationships for individual yield data years were found with individual PCs, however, i) not all the PCs were found to have a good relationship with a yield year and ii) different fields had different PCs that exhibited a good relationship with individual year yield maps ($R^2 > 0.5$). This threshold value (0.5) for a good relationship could be altered to be more or less strict in the

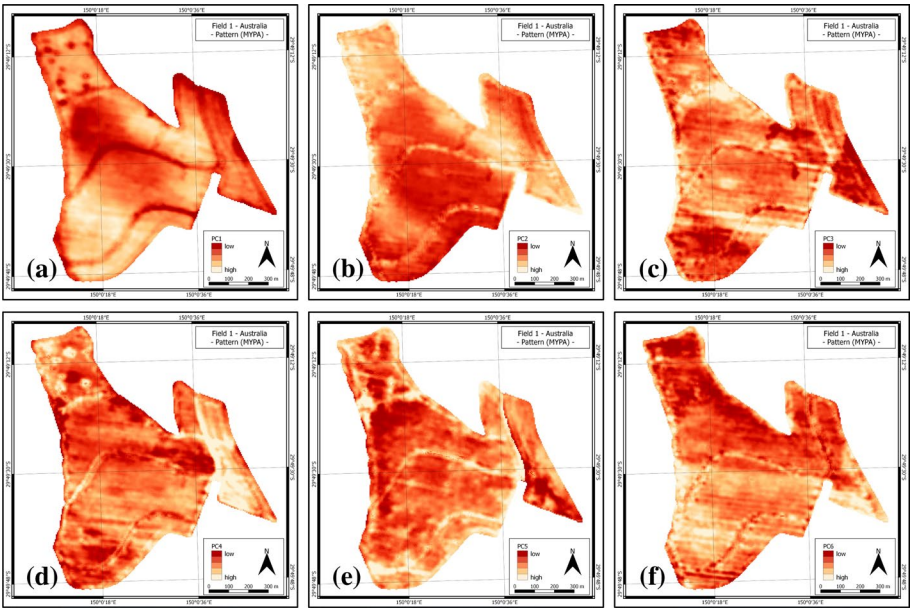

Fig. 5 Field 1—Australia: Maps of the first 6 Principal components showing the spatial patterns associated with each PC (**a**–**f** are respectively PCs 1–6)

selection of PCs. The rationale here was that if over half the variation in a yield map was explained, then the PC was relevant for management. It also meant that for any one yield year, only one PC could be selected as representative for that year. Using this threshold, there were between one and four PCs selected per field (Table 3). As expected, PC1 was always selected, and was related to multiple years in all fields. Fields where the yield maps differed considerably between years, tended to have PCs that related to individual years and more PCs chosen. Conversely, Field 2, with a very similar year-to-year pattern, only had PC1 selected (maps in Supplementary Data S2a).

The relationships between the μ map and the individual year yield maps (Table 2) were more varied than for the PCs. For example, Field 3—UK had 4 PCs selected, but the μ map did not have a $R^2 > 0.5$ with any year. In contrast, Field 2—Australia, had 5 years related to μ and only 4 years related to a PC. The μ maps for each field are shown in Fig. 6. The mean pattern tends to follow the PC1 pattern (Fig. 4 and in Supplementary Data S2a–c).

Temporal pattern stability analysis

To evaluate the temporal-spatial stability of yield, statistical pixel-wise analysis using σ over time as a measure was applied on the layer stack of the remaining normalized yield data. This pattern stability evaluation method enabled the identification and visualization of yield variability in space and time for all four fields (Fig. 7), where a high σ per pixel indicated high variability (unstable yield pattern) and low σ per pixel indicated low variability (stable yield pattern). Using Field 1—Australia again as an example (Fig. 7a), there was 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 central-eastern part

Table 2 Model fits (R^2) for the linear relationship between normalized yield (dependent variable) and the PCs/μ (independent variable)

Yield	PC1	PC2	PC3	PC4	PC5	PC6	PC7	μ
Field 1—Australia								
Sorghum 1998	**0.73**	0.16	0.09	0.02	<0.01	<0.01		0.48
Chickpea 1999	0.35	<0.01	<0.01	0.05	**0.59**	0.01		0.44
Wheat 2000	**0.78**	<0.01	0.02	0.18	0.01	<0.01		**0.67**
Wheat 2004	0.2	**0.72**	0.05	<0.01	<0.01	0.02		0.39
Sorghum 2005	**0.67**	<0.01	0.22	0.10	<0.01	0.02		**0.63**
Chickpea 2006	0.23	0.33	<0.01	0.06	0.01	0.36		0.49
Explained variation	56.7%	21.6%	7.8%	6.5%	3.9%	3.4%		
Field 2—Australia								
Sorghum 1998	**0.76**	0.13	0.05	0.01	0.04	<0.01		**0.65**
Wheat 1999	**0.76**	0.06	<0.01	0.16	0.01	<0.01		**0.71**
Sorghum 2001	**0.53**	0.40	0.07	<0.01	<0.01	<0.01		**0.57**
Wheat 2003	**0.65**	0.04	0.28	0.02	0.01	<0.01		**0.63**
Wheat 2005	0.49	0.10	<0.01	0.11	0.29	<0.01		**0.52**
Wheat 2006	0.27	0.20	<0.01	<0.01	0.08	0.45		0.44
Explained variation	64.2%	16.0%	8.4%	4.9%	4.7%	1.8%		
Field 3—UK								
Wheat 2002	0.03	**0.76**	0.18	0.03	<0.01	<0.01	<0.01	0.28
Canola 2004	0.26	0.45	0.23	0.06	<0.01	<0.01	<0.01	0.08
Canola 2007	0.01	0.21	0.05	0.01	0.07	0.17	0.48	0.41
Wheat 2008	**0.84**	0.05	0.06	0.06	<0.01	<0.01	<0.01	0.41
Canola 2010	0.33	0.04	<0.01	0.07	**0.54**	<0.01	<0.01	0.39
Wheat 2014	**0.75**	0.02	0.03	0.21	<0.01	<0.01	<0.01	0.28
Canola 2016	0.03	0.10	<0.01	<0.01	<0.01	**0.80**	0.07	0.32
Explained variation	50.5%	25.5%	10.1%	8.8%	2.7%	1.6%	0.1%	
Field 4—UK								
Wheat 2004	0.38	0.37	0.23	0.02	<0.01	<0.01		0.28
Wheat 2007	**0.57**	0.04	0.06	0.33	<0.01	<0.01		**0.67**
Canola 2011	0.25	0.14	<0.01	0.10	0.36	0.15		0.48
Wheat 2012	0.17	**0.74**	0.05	0.03	<0.01	<0.01		0.34
Canola 2014	0.32	0.15	<0.01	<0.01	0.34	0.19		**0.59**
Wheat 2015	**0.89**	<0.01	0.10	<0.01	<0.01	<0.01		**0.62**
Explained variation	53.8%	27.4%	10.9%	5.6%	1.5%	0.7%		

Relationships where more than 50% of the yield variance was explained by a PC or μ are indicated in bold font (Explained Variation = Percentage of explained yield variation across years)

Table 3 The Principal Components (PCs) selected in the yield pattern step to be used for clustering in the MYPA methodology (derived from Table 2)

Field ID	Location	PCs selected
1	Australia	1, 2 and 5
2	Australia	1
3	UK	1, 2, 5 and 6
4	UK	1 and 2

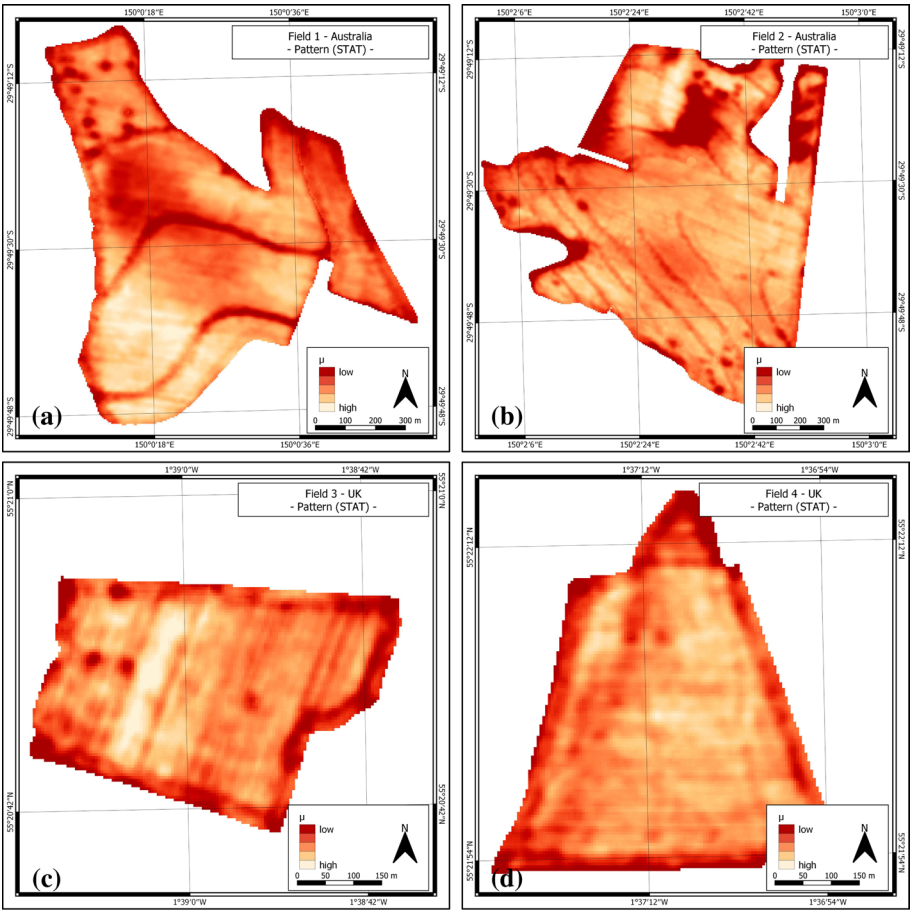

Fig. 6 Detected mean yield pattern by the STAT approach for each demonstration field: **a** Field 1—Australia, **b** Field 2—Australia, **c** Field 3—UK and **d** Field 4—UK

(denoted "**A**"). This area of instability was associated with a small gully and an area of heavier clay soils in the field. In drier years, this area had better soil moisture availability, generating higher yields. However, in 'wet' years, its landscape position and heavier soil type was susceptible to water-logging effects. It tended to be either an area of *relatively* higher or lower yield.

Yield productivity-stability units delineation

In order to produce information layers to support management decisions (especially risk management), the derived yield pattern layers (Fig. 5, Supplementary Data S2a-c and Fig. 6) were merged with the stability layer (Fig. 7). All possible solutions ($k=[2,7]$) for both the MYPA and STAT approaches were derived. An example, using the $k=4$ solution for Field 1, is shown in Fig. 8 for both the MYPA and STAT approaches. The outputs for

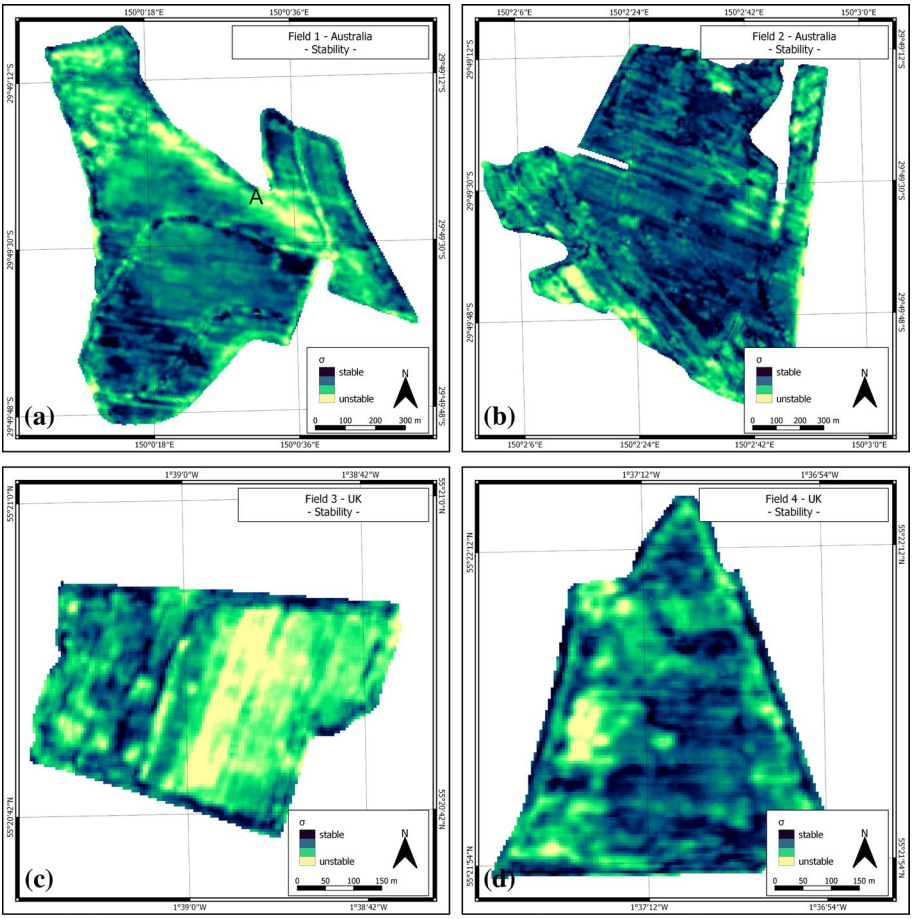

Fig. 7 Temporal stability of yield pattern as σ of all normalized yield maps, showing the spatial yield variability over time for each demonstration field: **a** Field 1—Australia, **b** Field 2—Australia, **c** Field 3—UK and **d** Field 4—UK

all fields and all levels of *k* were passed to the validation and comparison analyses to identify the optimal output.

Within Fig. 8, similar broad patterns were observed for the two methods, but clear differences were also noticeable. At this stage it is unclear which is preferred, but the key observation is that there are differences in the two outputs (see also Fig. 10).

Validation of final map quality and comparison of methods

Multivariate analysis of variance

Visually, the PSU_{MYPA} appears to be slightly more coherent, with larger more contiguous zones (Figs. 8 and 10). It may be expected that this increased coherence would come at a statistical price. However, the mean scores from the ranking of the MANOVA output

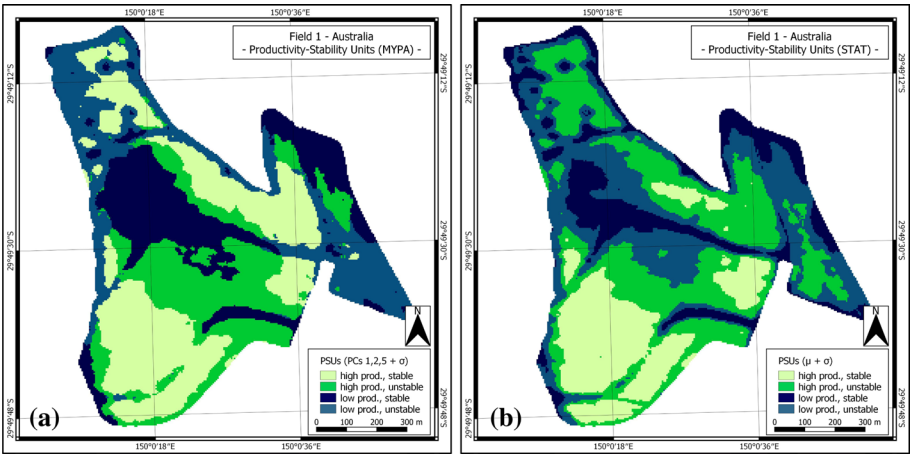

Fig. 8 Examples of the final PSU maps derived for Field 1 using k = 4 for both the MYPA (**a**) and STAT (**b**) approaches. Maps are shown in raster format. Comparisons between classes within a map were used to assign labels (high/low productivity; stable/unstable response). The mean annual yield, overall mean yield and mean PC response are given in Supplementary Data (S3a and S3b)

(Table 4 and Supplementary Data S4) indicated that the MYPA model was statistically better than the STAT (lower Wilks' Lambda and the higher Pillai's Trace) in three of the 4 fields (Fields 1, 3 and 4). In these fields, the MYPA approach outperformed STAT for all values of *k*. For Field 2, which only had PC1 selected for the MYPA yield stability analysis, Pillai's trace indicated that the STAT approach was optimal (but not for all *k*-levels) and there was no difference between the STAT and MYPA approach using Wilks' lambda (mean score of 0.5).

Descriptive statistics of clustering

Comparing both approaches, for a given *k*-level, the MYPA approach generated more discrete zones (polygons) at lower values of k (≤ 4) in the larger Australian fields (Fields 1–2) (Fig. 9a). With higher values of k (>4), the MYPA almost always produced fewer polygons than the STAT approach (Field 2, $k = 7$ is the only exception). Figure 9b shows the

Table 4 Mean scores for the performance of the MYPA vs. STAT calculated across all fields and all levels of *k* classes

Field ID	Mean Score Pillai's Trace	Mean Score Wilks' lambda
1	1.0	1.0
2	0.3	0.5
3	1.0	1.0
4	1.0	1.0

A value of 1 was assigned when the MYPA outperformed STAT (lower Wilks' Lambda or the higher Pillai Trace), and conversely a value of 0 if STAT was superior, before the calculation of the mean scores. As the mean score approaches 1, the more dominant the MYPA method was compared to the STAT method

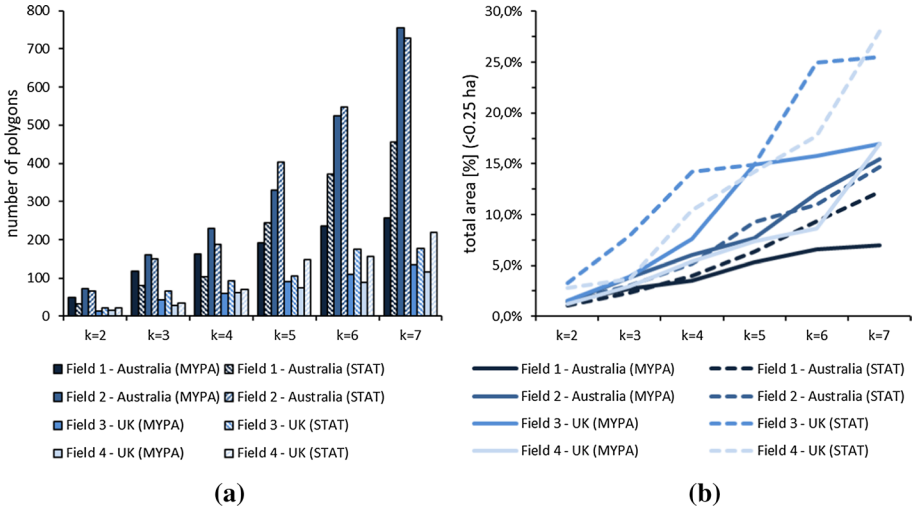

Fig. 9 **a** The total number of discrete polygons (zones) created by *k* [2,7] classes using the MYPA (solid bars) and STAT (hatched bars) methodologies in four different fields and, **b** the total percentage area of small ('unmanageable' polygons < 0.25 ha) in each field for both methodologies (MYPA solid lines; STAT dashed lines)

total area associated with small (considered 'unmanageable') polygons in each output. The smaller UK fields tended to have a more rapidly escalating percentage area associated with small polygons, but the trend for the MYPA to have less area associated with 'unmanageable' polygons was constant in Fields 3 and 4. At low *k*-levels, Field 1 is similar for the two approaches, but the MYPA had a consistently lower percentage area of 'unmanageable' zones at higher *k*-level*s*, while Field 2 showed little divergence between the two methods.

Zonal Opportunity Index

Results for the calculation of O_z for all values of *k*[2,7] for the two methods (MYPA and STAT) in the four fields are shown in Table 5 (full details of variables used in the calculation are given in Supplementary Data S5). For Field 1 and 2, the larger Australian fields, lower values of *k* indicated a better fit for the STAT method, while higher values showed a preference for MYPA. However, Field 2 was very similar in O_z responses at all levels of *k*. Fields 3 and 4 indicated that the MYPA was consistently the better performed method based on the O_z metric. The best performed *k*-level is indicated in bold in Table 5 for both STAT and MYPA. Although a small sample size (four fields), the optimum *k*-level for MYPA was equal to or less than the STAT optimum *k*-level in all cases. For the four fields, the *k*-level associated with the best performed STAT model was selected and used for a direct comparison between the MYPA and STAT output (Table 6).

When the zonal opportunity was recalculated using only zones of a manageable size, i.e. z = number of zones > 0.25 ha, a more pronounced outcome was achieved in favour of MYPA (Table 6). Despite selecting the *k*-level associated with the best performed STAT method in Table 5, the MYPA outperformed STAT in 3 of the 4 fields (lower $O_{z\text{-man}}$). Field 1 switched from having a lower STAT O_z to having a lower MYPA $O_{z\text{-man}}$. For Field 2, the STAT $O_{z\text{-man}}$ was lower, but the difference between the two methods was much closer, with similar $O_{z\text{-man}}$ and percentage area values. For the other three field (Fields 1, 3-4),

Table 5 Zonal Opportunity (O_z) calculated for 6 levels of classification (k[2,7]) using both the STAT and MYPA methods for generating Productivity-Stability Units (PSUs)

Number of Classes (k)	Field 1		Field 2		Field 3		Field 4	
	STAT	MYPA	STAT	MYPA	STAT	MYPA	STAT	MYPA
2	456.2	453.7	503.1	532.0	436.3	305.9	451.4	380.4
3	320.5	**374.2**	**461.0**	**484.0**	311.4	187.7	257.2	**181.4**
4	**297.2**	388.9	468.5	552.8	285.5	**183.1**	**252.3**	204.1
5	549.2	414.5	873.5	718.2	**270.0**	222.3	384.7	205.4
6	788.7	495.1	1144.6	1093.5	401.0	242.1	382.8	217.6
7	947.9	529.7	1499.7	1541.4	391.7	287.8	497.7	260.6

The best performed *k*-level (for both STAT and MYPA) are given in bold

O_z calculated using the total number of zones (polygons) in the PSU output. Lower values indicate a better result

Table 6 Zonal opportunity of manageable zones ($O_{z\text{-man}}$) calculated for the best performed *k*-level from the O_z STAT analysis in each field (see Table 5)

Field ID	*k*-level	STAT		MYPA	
		$O_{z\text{-man}}$	% Area (< 0.25 ha)	$O_{z\text{-man}}$	% Area (< 0.25 ha)
1	4	131.17	4.0	102.91	3.5
2	3	210.97	3.0	219.97	3.8
3	5	88.00	14.9	62.33	14.9
4	4	199.19	3.7	137.42	3.0
#3	3	193.44	7.9	115.69	3.9

Only the number of manageable zones was considered in the $O_{z\text{-man}}$ calculation (c.f. total number of zones in Table 5). The best solution in Field 3 (k = 5) had a total unmanageable area > 10% and was redone with *k* = 3, such that there was < 10% unmanageable area for both the STAT and MYPA approaches). Lower $O_{z\text{-man}}$ values indicate a better model fit

#Field 3 redone with k = 3 due to a large percentage of unmanageable area (< 0.25 ha) in the k = 5 solution

the lower MYPA $O_{z\text{-man}}$ was also associated with a smaller percentage of 'unmanageable' area, i.e. more of the field was manageable and the zoning was preferable using MYPA. For Field 3, the analysis was redone using *k* = 3 as the preferred result (*k* = 5) from Table 5 had 14.9% of the field area in zones < 0.25 ha for both methods. A maximum of 10% was set as an acceptable threshold, but this is arbitrary and should be adjusted to suit the target production system. It is noted that the levels of *k* selected in this process (3-4) reflect other work that has indicated that a 2-4 class solution tends to be optimal when classification is applied to spatial data sets (Pedroso et al. 2010).

Figures 8 and 10 illustrate that the two approaches generate similar patterns although there are some differences. In particular, the MYPA approach tended to generate larger and more coherent units, which translated into fewer and larger zones when converted into polygons and less area in small 'unmanageable zones. This is a desired result, as the more

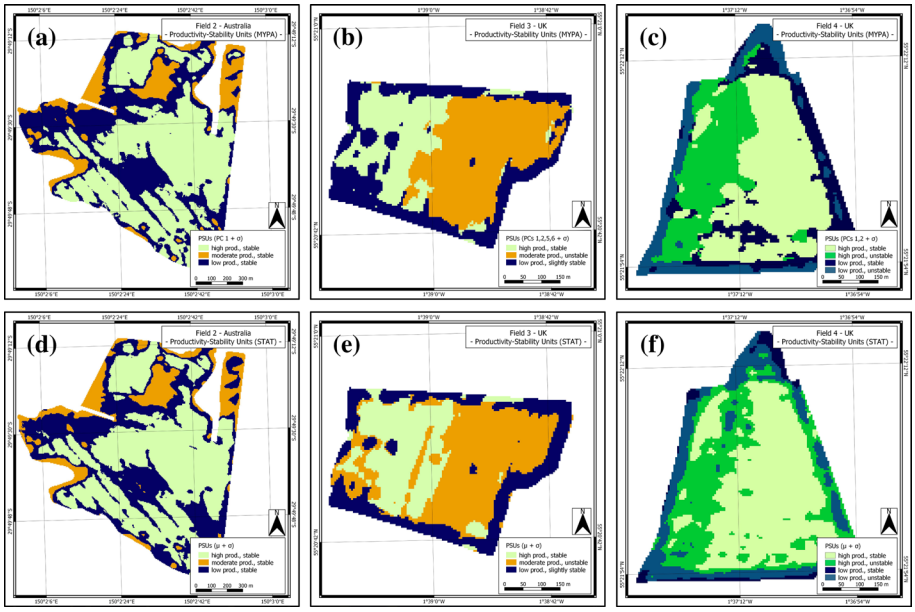

Fig. 10 Final PSU maps derived for Field 2, 3 and 4 using the optimal *k*-level solution for both the MYPA (**a**–**c**) and STAT (**d**–**f**) approaches. Maps are shown in raster format. Comparisons between classes within a map were used to assign labels (high/low productivity; stable/unstable response). The mean annual yield, overall mean yield and mean PC response are given in Supplementary Data S6ab–8ab

coherent the zoning the fewer the changes and decisions that need to be made when performing differential (variable-rate) operations. There is a cost and a level of uncertainty associated with any rate change (Tissyre and McBratney 2008). The smoother zoning is particularly evident in the UK fields (Field 3-4) and indicated 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.

Field 2 exhibited very little difference (visually and statistically) between the two approaches. This is the most temporally uniform field of the four. It had the highest percentage of variance explained by PC1, and PC1 was the only PC that was selected. This indicated a stable spatio-temporal pattern. In addition, 5 of the 6 individual years had a $R^2 > 0.5$ with the mean yield map (Table 2). Under these conditions there appears to be no advantage to the MYPA method. Using a straight statistical (μ and σ) approach to yield zoning is just as effective in temporally stable systems. This, however, is not usually the norm in dryland systems, as indicated by the other three fields where the MYPA approach was better.

Field 1 presented the biggest deviation between the two maps, particularly in the unstable area "**A**" identified in Fig. 7a that was classed as stable with the MYPA approach and unstable in the STAT approach. As described previously, this area was characterized by deeper, clay soils that tended to yield the highest in a normal year (e.g. 1998, 2000) but were prone to waterlogging in 'wet' years, with significant in-season rainfall (e.g. 2004), and had a relatively lower yield in those years. In the 'wet' years, the shallower, lower clay soils in other areas of the field tended to improve their yield. However, while the relative yield was lower in the "**A**" area in 2004, the actual yield tended 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 was stable in absolute terms, if not in relative terms between normal and 'wet' years.

Further discussion

Using the O_z statistic and the MANOVA outputs, the MYPA explained more of the yield variance and generated larger, more coherent zones than the established STAT approach. Furthermore, it offered an opportunity to assess the quality of the yield maps being used, unlike the STAT approach. However, while the O_z provides a metric for ranking and comparison, it is not intended to be an absolute statistic. It should be used with some degree of interpretation. In this study, it was used in conjunction with information relating to the percentage field area that was described by small, unmanageable zones (< 0.25 ha for these fields) and the patterns and mean area of the PSU zones. The MYPA was particularly more effective when only areas of the field that could be differentially managed were considered.

The MYPA is semi-automated; with the specification of fixed threshold values, associated with the selection of yield maps and of PCs, it could be fully automated. However, in the form presented here it invites the user to verify and engage in the PSU process. The STAT approach could equally be fully automated, but provides no options to vary the yield pattern analysis to cope with different types of yield data, e.g. arable cereal yield vs. field horticulture root crops vs. vineyard yields vs. tree crop yields etc. Different types of production have different effects and limitations with their yield data and, while the interpolation process can smooth some of these differences, it is not always intuitive to use a one-size fits all approach to analysing very different time-series of yield data.

The PSUs presented here are not potential management units (or zones), they are yield units. They could be used as management units; however, this is not recommended. The PSUs are intended to simplify and synthesize time-series of yield data. As such, the PSU maps could be used in the derivation of management units or for on-the-go differential management, but they should be used in conjunction with other relevant and available information, e.g. spatial soil data, maps of canopy vigour etc.…, when making agronomic decisions. The intent with this MYPA workflow is not to generate management units but to present a simpler, more coherent yield data layer that growers can interpret and work from (instead of a long time-series of yield maps) and to do this in a logical and quasi-automated fashion.

The MYPA primarily presents a new approach to assessing productivity patterns. Yield stability uses the same method, the σ over time at a point, as the STAT. The σ was considered a good method to assess stability. However, PSU delineation may benefit from alternative methods of assessing yield stability and is a potential area for further development.

Conclusions

In this study, a novel pattern recognition-based method—Multi-temporal Yield Pattern Analysis (MYPA)—was proposed for multi-year yield data analysis in arable cropping systems as an alternative approach to purely statistical approaches. Applying pattern recognition techniques to long yield time series offered a high potential to reveal long-term spatio-temporally yield production units at the local scale. The MYPA method was

designed free of assumptions of how both stable (e.g. permanent soil conditions) and non-stable factors (e.g. weather pattern, crop type and cycle) affect yield response and temporal variability. This approach synthesizes the information within multiple yield maps into a single layer that indicated the relative yield response and how stable that response was across a 10 + year period. Image analysis techniques, principal component analysis (PCA) and *k*-means clustering, were demonstrated as useful techniques, capable of handling high-resolution yield data and long yield map time series efficiently. For the detection of erroneous 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 such as double cropping). Compared to more commonly used purely statistical analysis approaches, the MYPA approach generated larger, more coherent units and with these four examples presented a better statistical solution. The hypothesis that *the novel MYPA approach enhances the insights into spatio-temporal yield variation* compared to purely statistical approaches was demonstrated by multivariate analysis of variance (MANOVA) and a zonal opportunity index. The advantages of the MYPA approach appear to be stronger when yield data do not exhibit strong spatio-temporal patterns, i.e. yield maps are changeable between years. Overall, the MYPA demonstrated its potential high value as a robust approach that can potentially be encoded to easily produce interpretable and understandable information layers to support management decisions, especially for risk management. Further work will examine if this holds true with other yield time series, such as satellite biomass time-series, and how it compares to other emerging approaches (e.g. Layton et al. 2019; Leroux et al. 2018b).

Acknowledgements The authors would like to acknowledge the support of Mr Michael Smith and Mr David Park for the use of their yield data and Dr Brett Whelan from the Precision Agriculture Laboratory, the University of Sydney for providing access to the Australian yield data. This research was part of Dr Gerald Blasch's Post-Doc studies at Newcastle University. It was funded by the UK-China Research and Innovation Partnership Fund (Newton Programme: PAFiC-Precision Agriculture for Family-farms in China project, National Natural Science Foundation of China (NSFC) (No. 61661136003) & Science and Technology Facilities Council (STFC) of United Kingdom (No. ST/N006801/1).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

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. https://doi.org/10.1016/j.ejrs.2011.04.001.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Albertz, J. (2009). *Einführung in die Fernerkundung—Grundlagen der Interpretation von Luft- und Satellitenbildern* (4th ed., p. 254). WBG: Darmstadt.

Bakhsh, A., Jaynes, D. B., Colvin, T. S., & Kanwar, R. S. (2000). Spatio-temporal analysis of yield variability for a corn-soybean field in Iowa. *Transactions of the American Society of Agricultural and Biological Engineers, 43,* 31–38. https://doi.org/10.13031/2013.2684.

Basso, B., Fiorentino, C., Cammarano, D., Cafiero, G., & Dardanelli, J. (2012). Analysis of rainfall distribution on spatial and temporal patterns of wheat yield in Mediterranean environment. *European Journal of Agronomy, 41,* 52–65. https://doi.org/10.1016/j.eja.2012.03.007.

Bauckhage, C., & Kersting, K. (2013). Data mining and pattern recognition in agriculture. *Künstliche Intelligenz, 27,* 313–324. https://doi.org/10.1007/s13218-013-0273-0.

Bivand, R. S., Keitt, T., & Rowlingson, B. (2018). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.3-4. Retrieved November 2, 2018, from http://CRAN.R-project.org/package=rgdal.

Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R* (2nd ed.). New York: Springer. https://doi.org/10.1007/978-1-4614-7618-4.

Blackmore, B. S. (2000). The interpretation of trends from multiple yield maps. *Computers and Electronics in Agriculture, 26*(1), 37–51.

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,* 455–466. https://doi.org/10.1016/S1537-5110(03)00038-2.

Blackmore, B. S., & Moore, M. (1999). Remedial correction of yield map data. *Precision Agriculture, 1,* 53–66. https://doi.org/10.1023/A:1009969601387.

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. https://doi.org/10.1016/j.compag.2015.01.012.

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. https://doi.org/10.3390/rs70911125.

Boydell, B., & McBratney, A. B. (2002). Identifying potential within-field management zones from cotton-yield estimates. *Precision Agriculture, 3,* 9–23. https://doi.org/10.1023/A:1013318002609.

Bramley, R. G. V. (2009). Lessons from nearly 20 years of precision agriculture research, development, and adoption as a guide to its appropriate application. *Crop and Pasture Science, 60,* 197–217. https://doi.org/10.1071/CP08304.

Bramley, R. G. V., & Ouzman, J. (2018). Farmer attitudes to the use of sensors and automation in fertilizer decision-making: Nitrogen fertilization in the Australian grains sector. *Precision Agriculture*. https://doi.org/10.1007/s11119-018-9589-y.

Córdoba, M., Bruno, C., Costa, J., & Balzarini, M. (2013). Subfield management class delineation using cluster analysis from spatial principal components of soil variables. *Computers and Electronics in Agriculture, 97,* 6–14. https://doi.org/10.1016/j.compag.2013.05.009.

Delbecq, B. A., Brown, J. P., Florax, R. J. G. M., Kladivko, E. J., Nistor, A. P., & Lowenberg-DeBoer, J. M. (2012). The impact of drainage water management technology on corn yields. *Agronomy Journal, 104,* 1100–1109. https://doi.org/10.2134/agronj2012.0003.

Fountas, S., Blackmore, S., Ess, D., Hawkins, S., Blumhoff, G., Lowenberg-DeBoer, J., et al. (2005). Farmer experience with Precision Agriculture in Denmark and the US Eastern Corn Belt. *Precision Agriculture, 6,* 121–141. https://doi.org/10.1007/s11119-004-1030-z.

Georgi, C., Spengler, D., Itzerott, S., & Kleinschmit, B. (2018). Automatic delineation algorithm for site-specific management zones based on satellite remote sensing data. *Precision Agriculture, 19,* 684–707. https://doi.org/10.1007/s11119-017-9549-y.

Graeler, B., Pebesma, E. J., & Heuvelink, G. (2016). Spatio-Temporal Interpolation using gstat. *The R Journal, 8*(1), 204–218.

Griffin, T. W., Dobbins, C. L., Vyn, T. J., Florax, R. J. G. M., & Lowenberg-DeBoer, J. M. (2008). Spatial analysis of yield monitor data: Case studies of on-farm trials and farm management decision making. *Precision Agriculture, 9,* 269–283. https://doi.org/10.1007/s11119-008-9072-2.

Guastaferro, F., Castrignanò, A., De Benedetto, D., Sollitto, D., Troccoli, A., & Cafarelli, B. (2010). A comparison of different algorithms for the delineation of management zones. *Precision Agriculture, 11,* 600–620. https://doi.org/10.1007/s11119-010-9183-4.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 28*(1), 100–108. https://doi.org/10.2307/2346830.

Hijmans, R. J. (2017). raster: Geographic data analysis and modeling. R package version 2.6-7. Retrieved November 2, 2018, from https://CRAN.R-project.org/package=raster.

Horikoshi, M., & Tang, Y. (2016). ggfortify: Data visualization tools for statistical analysis results. Retrieved November 2, 2018, from https://CRAN.R-project.org/package=ggfortify.

Kitchen, N. R., Sudduth, K. A., Myers, D. B., Massey, R. E., Sadler, E. J., Lerch, R. N., et al. (2005). Development of a conservation-oriented precision agriculture system: Crop production assessment and plan implementation. *Journal of Soil and Water Conservation, 60,* 421–430.

Kutter, T., Tiemann, S., Siebert, R., & Fountas, S. (2011). The role of communication and co-operation in the adoption of precision farming. *Precision Agriculture, 12,* 2–17. https://doi.org/10.1007/s11119-009-9150-0.

Lark, R. M., & Stafford, J. V. (1997). Classification as a first step in the interpretation of temporal and spatial variability of crop yield. *Annals of Applied Biology, 130,* 111–121. https://doi.org/10.1111/j.1744-7348.1997.tb05787.x.

Layton, A., Krogmeier, J. V., Ault, A., & Buckmaster, D. R. (2019). From yield history to productivity zone identification with hidden Markov random fields. *Precision Agriculture*. https://doi.org/10.1007/s11119-019-09694-2.

Leroux, C., Jones, H., Clenet, A., Dreux, B., Becu, M., & Tisseyre, B. (2018a). A general method to filter out defective spatial observations from yield mapping datasets. *Precision Agriculture, 19*(5), 789–808. https://doi.org/10.1007/s11119-017-9555-0.

Leroux, C., Jones, H., Taylor, J., Clenet, A., & Tisseyre, B. (2018b). A zone-based approach for processing and interpreting variability in multi-temporal yield data sets. *Computers and Electronics in Agriculture, 148,* 299–308. https://doi.org/10.1016/j.compag.2018.03.029.

Leutner, B., Horning, N., & Schwab-Willmann, J. (2018). RStoolbox: Tools for remote sensing data analysis. R package version 0.2.1. Retrieved November 2, 2018, from https://CRAN.R-project.org/package=RStoolbox.

Li, X., Pan, Y., Zhao, C., Wang, J., Bao, Y., & Wang, J. (2007). Delineation and scale effect of precision agriculture management zones using yield monitor data over four years. *Agricultural Sciences in China, 6,* 180–188. https://doi.org/10.1016/S1671-2927(07)60033-9.

Lyle, G., Bryan, B., & Ostendorf, B. (2013). Post-processing methods to eliminate erroneous grain yield measurements: Review and directions for future development. *Precision Agriculture, 15,* 377–402. https://doi.org/10.1007/s11119-013-9336-3.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2018). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1. Retrieved November 2, 2018, from https://cran.r-project.org/web/packages/cluster/index.html.

McBratney, A. B., Whelan, B. M., & Shatar, T. (1997). Variability and uncertainty in spatial, temporal and spatiotemporal crop-yield and related data. *Ciba Foundation Symposium, 210,* 141–160. https://doi.org/10.1002/9780470515419.ch9.

McBratney, A. B., Whelan, B. M., Taylor, J. A., & Pringle, M. J. (2000). A management Opportunity Index for Precision Agriculture. In: Proceedings of the 5th International Conference on Precision Agriculture, Bloomington, Minnesota, USA, 16–19 July, 2000.

McKinion, J. M., Willers, J. L., & Jenkins, J. N. (2010). Spatial analyses to evaluate multi-crop yield stability for a field. *Computers and Electronics in Agriculture, 70,* 187–198. https://doi.org/10.1016/j.compag.2009.10.005.

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. Retrieved April 27, 2018, from https://sydney.edu.au/agriculture/pal/software/vesper.shtml.

Moral, F., Terrón, J., & Silva, J. (2010). Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil and Tillage Research, 106,* 335–343. https://doi.org/10.1016/j.still.2009.12.002.

Morari, F., Castrignanò, A., & Pagliarin, C. (2009). Application of multivariate geostatistics in delineating management zones within a gravelly vineyard using geo-electrical sensors. *Computers and Electronics in Agriculture, 68,* 97–107. https://doi.org/10.1016/j.compag.2009.05.003.

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering, 114,* 358–371. https://doi.org/10.1016/j.biosystemseng.2012.08.009.

Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association, 69,* 894–908. https://doi.org/10.2307/2286159.

Ortega, R. A., & Santibáñez, O. A. (2007). Determination of management zones in corn (*Zea mays* L.) based on soil fertility. *Computers and Electronics in Agriculture, 58,* 49–59. https://doi.org/10.1016/j.compag.2006.12.011.

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. https://doi.org/10.3390/rs2081973.

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences, 30,* 683–691. https://doi.org/10.1016/j.cageo.2004.03.012.

Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2). Retrieved November 2, 2018, from https://cran.r-project.o.rg/doc/Rnews/.

Pedroso, M., Taylor, J., Tisseyre, B., Charnomordic, B., & Guillaume, S. (2010). A segmentation algorithm for the delineation of management zones. *Computer and Electronics in Agriculture, 70,* 199–208. https://doi.org/10.1016/j.compag.2009.10.007.

Ping, J. L., & Dobermann, A. (2005). Processing of yield map data. *Precision Agriculture, 6,* 193–212. https://doi.org/10.1007/s11119-005-1035-2.

Ping, J. L., Green, C. J., Bronson, K., Zartman, R. E., & Dobermann, A. (2005). Delineating potential management zones for cotton based on yields and soil properties. *Soil Science, 170,* 371–385. https://doi.org/10.1097/01.ss.0000169904.56743.75.

QGIS Development Team. (2009). QGIS geographic information system. Open Source Geospatial Foundation. Retrieved October 14, 2019, from http://qgis.org.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved November 2, 2018, from https://www.R-project.org/.

Raj, M. P., Swaminarayan, P. R., Saini, J. R., & Parmar, D. K. (2015). Applications of pattern recognition algorithms in agriculture: A review. *International Journal of Advanced Networking and Applications, 6,* 2495–2502.

Robinson, T. P., & Metternicht, G. (2005). Comparing the performance of techniques to improve the quality of yield maps. *Agricultural Systems, 85,* 19–41. https://doi.org/10.1016/j.agsy.2004.07.010.

Rodrigues Junior, F. A., Vleira, L. B., Queiroz, D. M., & Santos, N. T. (2011). Geração de zonas de manejo para cafeicultura empregando-se sensor SPAD e análise foliar. *Revista Brasileira de Engenharia Agrícola e Ambiental, 15,* 778–787. https://doi.org/10.1590/S1415-43662011000800003.

Sabins, F. F. (1996). *Remote sensing: Principles and interpretation* (3rd ed., p. 494). New York: W.H. Freeman and Company.

Say, S. M., Keskin, M., Sehri, M., & Sekerli, Y. E. (2018). Adoption of Precision Agriculture Technologies in Developed and Developing Countries. *Online Journal of Science and Technology, 8,* 7–15.

Schenatto, K., de Souza, E. G., Bazzi, C. L., et al. (2017). Normalization of data for delineating management zones. *Computers and Electronics in Agriculture, 143,* 238–248. https://doi.org/10.1016/j.compag.2017.10.017.

Schimmelpfennig, D., & Ebel, R. (2011). On the doorstep of the information age: Recent adoption of Precision Agriculture. Washington, DC: EIB-80, U.S. Department of Agriculture, Economic Research Service.

Schowengerdt, R. A. (2007). *Remote Sensing: Models and Methods for Image Processing* (3rd ed., p. 515). San Diego: Academic Press.

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

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. https://doi.org/10.1016/0168-1699(95)00042-9.

Sudduth, K., & Drummond, S. T. (2007). Yield editor: Software for removing errors from crop yield maps. *Agronomy Journal, 99,* 1471. https://doi.org/10.2134/agronj2006.0326.

Sun, W., Whelan, B. M., McBratney, A. B., & Minasny, B. (2013). An integrated framework for software to provide yield data cleaning and estimation of an opportunity index for site-specific crop management. *Precision Agriculture, 14,* 376–391. https://doi.org/10.1007/s11119-012-9300-7.

Tagarakis, A., Liakos, V., Fountas, S., Koundouras, S., & Gemtos, T. A. (2013). Management zones delineation using fuzzy clustering techniques in grapevines. *Precision Agriculture, 14,* 18–39. https://doi.org/10.1007/s11119-012-9275-4.

Tang, Y., Horikoshi, M., & Li, W. (2016). ggfortify: Unified interface to visualize statistical result of popular R packages. *The R Journal, 8*(2), 478–489.

Taylor, J. A., McBratney, A. B., & Whelan, B. M. (2007). Establishing management classes for broadacre agricultural production. *Agronomy Journal, 99,* 1366–1376. https://doi.org/10.2134/agronj2007.0070.

Taylor, J. A., & Whelan, B. M. (2011). Selection of ancillary data to derive production management units in sweetcorn (*Zea Mays* var. rugosa) using MANOVA and an information criterion. *Precision Agriculture, 12,* 519–533. https://doi.org/10.1007/s11119-010-9195-0.

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. https://doi.org/10.1007/s11119-008-9053-5.

Uribeetxebarria, A., Arnó, J., Escolà, A., & Martínez-Casasnovas, J. A. (2018). Apparent electrical conductivity and multivariate analysis of soil properties to assess soil constraints in orchards affected by previous parcelling. *Geoderma, 319,* 185–193. https://doi.org/10.1016/j.geoderma.2018.01.008.

Van Meirvenne, M., Islam, M. M., De Smedt, P., Meerschman, E., Van De Vijver, E., & Saey, T. (2013). Key variables for the identification of soil management classes in the aeolian landscapes of north–west Europe. *Geoderma, 199,* 99–105. https://doi.org/10.1016/j.geoderma.2012.07.017.

Vega, A., Córdoba, M., Castro-Franco, M., & Balzarini, M. (2019). Protocol for automating error removal from yield maps. *Precision Agriculture, 20*(5), 1030–1044. https://doi.org/10.1007/s11119-018-09632-8.

Whelan, B. M., & McBratney, A. B. (1998). Prediction Uncertainty and Implications for Digital Map Resolution. In: P. C., Robert, R. H., Rust & W.E., Larson (eds), Proceedings of the 4th International Conference on Precision Agriculture, ASA, Madison, USA. 1185–1196.

Whelan, B. M., & McBratney, A. B. (2000). An Approach to Deconvoluting Grain-Flow within a Conventional Combine Harvester using a Parametric Transfer Function. *Precision Agriculture, 2,* 389–398. https://doi.org/10.1023/A:1012356100172.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Gerald Blasch[1,2] · Zhenhai Li[3,4] · James A. Taylor[1,5]

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

[2] International Maize and Wheat Improvement Center - CIMMYT, 56237 Texcoco, Mexico

[3] School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, UK

[4] National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

[5] UMR ITAP, INRAE, University of Montpellier, 34196 Montpellier, France