Multi-thematic delineation of ‘natural zones’ of arable fields and their correspondence to spatial yield variation

D. Ekanayake*[1], J. Owens[2], A. Werner[1], A. Holmes[3]

1Lincoln Agritech Ltd., PO Box 69 133, Lincoln, Canterbury 7640, New Zealand.

2Agriculture and Life Sciences Faculty, PO Box 84, Lincoln University, Lincoln, NZ

3Foundation for Arable Research, 133C, Ruakura Road, Ruakura, Hamilton, New Zealand.

Abstract

Properties such as soil apparent electric conductivity (EC_a), topography and other site-related data (e.g. canopy reflectance from aerial images) vary across field. The agronomic effects of such variability can sometimes be seen in the spatial variations of crop yield on that field. However, yield maps do not always represent the natural boundaries based on site characteristics. Identification of these boundaries as “management zones” (MZ) can be beneficial in crop management and improving crop input use efficiency. A simple methodology is required to delineate such zones. This research presents an effective methodology to delineate MZ in an irrigated and a non-irrigated (rain-fed) arable maize field in New Zealand. Elevation data for the sites were acquired from Google Earth images and a soil survey. Soil EC_a was collected from a soil survey with an electromagnetic device. Yield values (t/ha) were obtained from combine harvesters equipped with yield monitor and Global Positioning System (GPS), over the course of four years for the irrigated site, and two years for the non-irrigated site. The yield data was quality controlled using a filtering system to remove outliers and technically non-plausible data. The data sources were combined in Geographic Information Systems (GIS) and three MZ were delineated for each field through standard clustering methods. The maize yields were aggregated per derived MZ to compare yields between different MZ-classes. The results showed that there was some consistency in yields related to the MZ, derived without yield data. In both the non-irrigated and irrigated fields, the lowest yield consistently occurred in the same class each year, however, the MZ-class with the highest yield varied year to year. The results show that it is possible for the studied type of fields to delineate ‘natural’ clusters or zones of site properties that can be used as MZ-classes as they represent different yield levels. The required inputs are freely available and easily obtained data.

Background

Site properties of the arable fields vary, affecting the crop performance and crop input use efficiency. When the extent of variability is identified, crop performance and crop input use efficiency can be improved through managing the field by localised conditions (Gotway Crawford et al., 1997). This is known as site specific crop management (SSCM). For SSCM, derivation of sub-field regions of similar quantitative characteristics (i.e. topography, yield, soil nutrients, elevation, etc.) is necessary. Sub-field regions delineated by boundaries also known as “management zones” (MZ). Areas with the same MZ-class often have similar production potential (Fridgen et al., 2004). As the MZ represent varying potential yield levels having derived MZ allows crop inputs to be adjusted to the yield dependent crop demands through variable rate application of fertiliser or water (Zhang et al., 2010).

Typically, various types of spatial data have been used to delineate MZ. These include grid soil surveys, digital elevation data, aerial photographs, soil apparent electrical conductivity (EC_a), soil physical properties, topography (Fridgen et al., 2004), and crop yield data (Blackmore et al., 2003).

Yield maps generated during harvesting are increasingly available to farmers and are a useful tool in SSCM. These maps can be used alone or in combination with other spatial data for quantifying spatial variability. Patterns shown in the yield maps reflect the combined effects of spatial variables (e.g. soil properties, topographical attributes, weather, diseases) and management practices (e.g. fertiliser application) (Souza et al., 2016). However, within-field spatial patterns of yields represented in yield maps have been found to lack consistency year to year, likely due to the influence of complex interactions between factors such as rain-fall distribution, soil type, and soil water holding capacity (Blackmore et al., 2003). Alone, yield maps may have limited value for delineating MZ to use in SSCM.

There are many procedures and Geographic Information Systems (GIS) packages developed to analyse geo-spatial data, but adaptation of SSCM is not common among end users. The low popularity may be due to the complexity of the methods, requirement of specific skills to operate the software (Fridgen et al., 2004), and unavailability of geo-referenced data. Therefore, objective of this paper is to develop simple methods to derive 'natural' clusters or zones of site properties using readily available geo-data sources including EC_a maps, elevation maps, and satellite images (e.g. Google Earth) for farmers.

Methods

Data collection

The data for the study was obtained from two separate fields in Cambridge, Hamilton (-37.941026°, 175.484264°) and Bulls, Palmerston North (-40.227790°, 175.292542°), North Island, New Zealand. The field in Cambridge is 18.93 ha of flat terrain. The field is not irrigated and relies entirely on rain for water inputs. The field rotates between maize, either grain or silage, and grazed pasture. The soil is classified as a well-drained loam over sandy loam Typic Orthic Allophanic soil (Hewitt, 2010). The field in Bulls is irrigated with a centre pivot-irrigator, which extends to cover the 57.05 ha area. The terrain is sloped in a North to South direction. Bulls field had a crop rotation of maize crop and several other crops such as pasture, barley, and peas, and the field was sown with mixture of crops in some years. The soil is classified as a poorly drained Typic Orthic Gley silty loam soil (Hewitt, 2010). At the time of the yield measurements presented in this paper, both sites were growing maize.

Maize grain at 20% moisture was harvested with a John Deere combine equipped with an AgLeader yield monitor and an AgLeader differential global positioning system (GPS) receiver at the Cambridge field in 2013 and 2014. At Bulls field, a CLASS combine equipped with a Trimble yield monitor and a GPS receiver was used when harvesting maize grain at 14% mositure.

Soil EC_a data from 0 to 125 cm depth, and elevation data for the Cambridge field were acquired through an electromagnetic soil survey conducted with a Trimble real time kinematic (RTK) GPS (Trimble Navigation Ltd., Sunnyvale, California, USA) enabled Geonics EM38 (Geonics Ltd., Mississauga, Ontario, Canada) soil survey in 2014. A similar soil survey was carried out at Bulls field in 2011.

Satellite images available for the Cambridge and the Bulls field were retrieved from Google Earth Pro (v. 7.1.7.2602, Google Inc., California, USA) historic imagery archive. Images available on 30/07/2013 and 23/10/2015 for the Bulls field were used to extract the elevation data for the analysis. In addition, other images available in the archive for the Cambridge field (10/01/2008, 09/07/2010, 05/02/2013, 11/03/2013, 28/05/2013, 02/09/2013, 21/07/2014, 20/07/2015, and 11/02/2016) and for the Bulls field (08/09/2004, 28/02/2005, 04/03/2007, 05/05/2010, 27/03/2013, 10/04/2013, 04/04/2014, and 19/02/2015) were referred to observe any field variability (e.g. water channels, infrastructure of the fields, cropping systems, etc.) shown in the images.

Data Analysis

Yield monitor data obtained during the harvest was imported into ArcMap, ArcGIS Desktop 10.5 (Environmental Systems Research Institute, Redlands, CA, USA), and the data was screened for erroneous yield values using methods adopted from Kleinjan et al. (2002) and Ping and Dobermann (2005). This involved first detecting frequency distribution outliers for distance travelled, grain flow, and grain moisture based on the global means and standard deviations (SD). The SD value was decided from 1, 1.5, 2, 2.5 or 3 SD for each variable according to the histogram distribution of data, compared to its mean value (large range – higher SD). Then the values lying outside the mean ± selected SD (e.g. 1, 1.5, 2, 2.5, or 3) range were discarded. In the next step, dry matter (DM) maize grain yield data was filtered for cutting swath width and all data that was less than a defined cutting width of the combine was deleted. Then, DM maize grain yields were filtered for their defined lower and upper yield limits. The lower yield limit selected was 0.5 t/ha. The upper yield limit was decided based on the maximal yields recorded for the maize grain crop in the harvested year for the region of the selected fields. Finally, yield data recorded at double-planting rows, headland turns, and traffic areas which remained unfiltered were cleaned manually as required.

Soil EC_a and elevation data was filtered for erroneous data by setting the lower and higher threshold values according to the histogram distributions of global mean and SD values. The data lying outside the mean ± 2.5 SD or 3 SD were eliminated before the data sets were manually cleaned for erroneous data.

Cleaned EC_a and elevation data were interpolated in to a common grid of 1 m x 1 m using block kriging in VESPER (Variogram Estimation and Spatial Prediction with ERror) software (Australian Centre for Precision Agriculture, University of Sydney, Sydney, Australia). Interpolated data was normalised by dividing the point data value by the global mean (Fridgen et al., 2004). These normalised data were used to derive three clusters for combined variables of elevation and EC_a in Management Zone Analyst (v. 1.0.0) (MZA) software (Agricultural Research Service, USDA, USA). Outputs from MZA were imported into ArcGIS to develop the MZ maps for the Cambridge and Bulls field. Derived MZ maps were classified into three MZ-classes (Classes A, B and C) for practical purposes. Quality controlled maize yield data were overlayed on field MZ maps and the yield data lying in each class were extracted to calculate yield averages per class. Recurring differences in the average yield values of the MZ-classes validate the relevance of the derived MZ representing groups of yield variability. A 5 m buffer area from each boundary of the classes in the map was allocated when extracting yield data.

There were four years (2005, 2007, 2009 and 2011) and two years (2013 and 2014) of yield data for Bulls and Cambridge field, respectively, available for the analysis. Extracted yield data from separate classes for each year are presented as means with error bars as SD.

Summary statistics are presented by year, and include the mean, SD of the mean, median, minimum, maximum, number of samples and skewness. Data analysis was competed in R Studio (v. 0.99.887). A one-way ANOVA with a Fischer's LSD was completed to detect differences ($P \leq 0.05$) between the means from years, and to compare means from MZ-classes from the same year. The ANOVA used Type I sum of squares to account for the unbalanced sample sizes. The LSD was completed using the R package "agricolae".

Results

Data Quality Control Using Filtering

For Cambridge field, 28.4% and 27.1% of the original yield data were removed from 2013 and 2014, respectively, after filtering. The percentages of original yield data removal after filtering for the Bulls field were 37.0% (2005), 8.3% (2007), 15.2% (2009) and 17.9% (2011) (Table 1). There were large differences between minimum and maximum DM maize grain yields of the raw data recorded in 2013 and 2014 in Cambridge field. Filtering did not result in large changes to the statistical characteristics such as mean, median, SD and skewness of the yield data sets (Table 1).

Yield Classes Based on MZ Map

Irrigated, Bulls field (2005-2011)

Mean DM maize grain yields at 14% moisture from 2005 to 2011 increased over time, and on average, were highest in 2011. During the previous years, the lowest detected yields ranged from 6.8 to 8.6 t ha^{-1}, and reached maximums ranging from ≈14 to ≈22.4 t ha^{-1} (Table 1).

There were significant differences between classes during each of the years measured (Table 2). In all years measured except 2007, mean dry matter yields differed between each of the three classes (Fig. 1). Class B had the lowest average yields during all four years. Classes A and C had similar yields in all years, except in 2007. In 2007, mean dry matter maize yield in Class C was significantly higher than in Classes A and B.

Non-irrigated, Cambridge field (2013-2014)
Mean DM maize grain yields at 20% moisture were lowest in 2013 (Table 1). For the data from Cambridge field, there were significant differences between classes from both 2013 and 2014. The lowest yield was derived from Class A in both years while Class B recorded medium yield and Class C recorded the highest yield (Fig. 2).

Table 1. The summary statistics for each year before and after the data was filtered. SD represents the standard deviation, count represents the number sample, skew represents the skewness, and cleaned data % is the percentage of data removed during filtering from the initial dataset. The "after filtering" data was used for analyses.

	Field	Year	Min	Max	Mean	Median	SD	Count	Skew	Cleaned data %
Before filtering	Bulls	2005	6.813	14.081	10.576	10.688	1.180	32074	-0.095	
		2007	8.281	13.832	11.022	11.265	0.941	56230	-0.258	
		2009	8.553	16.503	12.954	13.140	1.170	6138	-0.159	
		2011	7.856	22.437	15.430	15.576	2.358	18473	-0.062	
	Cambridge	2013	0.000	90.733	8.318	9.326	7.383	22094	-0.137	
		2014	0.000	87.548	8.701	10.281	5.655	22564	-0.279	
After filtering	Bulls	2005	6.813	14.081	10.584	10.702	1.129	20212	-0.105	36.98
		2007	8.281	13.832	11.268	11.315	0.879	51546	-0.054	8.33
		2009	8.616	16.503	13.084	13.183	0.970	5305	-0.102	15.20
		2011	7.993	22.423	15.702	15.922	2.360	15167	-0.093	17.90
	Cambridge	2013	0.521	19.932	10.194	10.673	2.938	15811	-0.163	28.44
		2014	0.500	19.874	11.328	12.069	3.546	16453	-0.209	27.08

Table 2. The mean and standard deviation for all grids from each year, along with a *P* value and degrees of freedom obtained from the ANOVA.

Location	Year	Mean	Standard deviation	Degrees of freedom	*P* value
Bulls (irrigated)					
	2005	10.583	1.133	17885	2.96×10^{-11}
	2007	14.395	2.940	54769	2.20×10^{-16}
	2009	13.083	0.956	4647	1.02×10^{-07}
	2011	15.754	2.301	12956	0.008305
Cambridge (non-irrigated)					
	2013	10.299	2.800	11311	2.20×10^{-16}
	2014	11.295	3.519	11808	2.20×10^{-16}

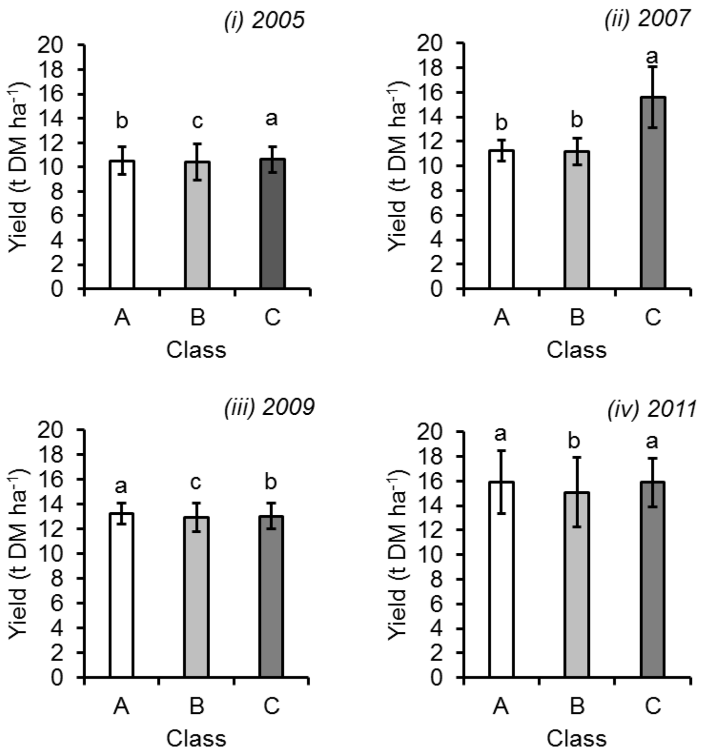

Figure 1. Means (± SD) of DM from irrigated maize grain yield (t ha-1) in Bulls field comparing each 'Class'. The Class represents with-in field management zone from the MZ map and 'n' represents the number of yield points harvested.
i) 2005 (Class A, n=4057; Class B, n=1992; Class C, n=11838), ii) 2007 (Class A, n=9971; Class B, n=5124; Class C, n=39676), iii) 2009 (Class A, n=2923; Class B, n=597; Class C, n=1129), and iv) 2011 (Class A, n=3118; Class B, n=2103; Class C, n=7737). Means that share the same letter are not significantly different (*P* ≤ 0.05).

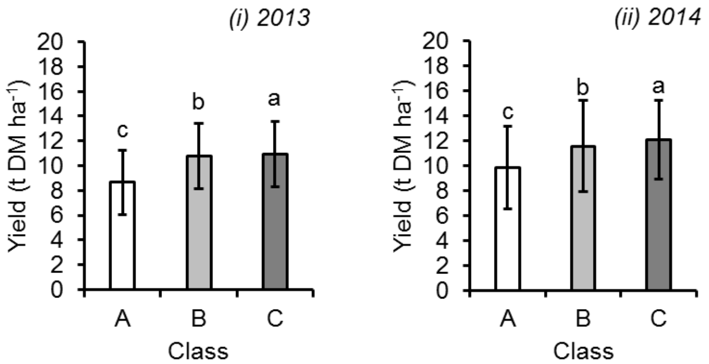

Figure 2. Means (± SD) of DM from non-irrigated maize grain yield (t ha-1) in the Cambridge field comparing each 'Class'. The Class represents with-in field management zone from the MZ map and 'n' represents the number of yield points harvested. i) 2013 (Class A, n=2781; Class B, n=5221; Class C, n=3311), and ii) 2014 (Class A, n=2812; Class B, n=5529; Class C, n=3469). Means that share the same letter are not significantly different (*P* ≤ 0.05).

Discussion

Evaluation of Data Quality Control and Filtering Methods

Yield data collected by the combine yield monitors inevitably contains systematic and random erroneous data. Cleaning of data is required to remove the erroneous data. This is necessary for statistical comparisons of yield data with other data layers in a decision support system (Kleinjan et al., 2002) and for precise interpretation of yield patterns in SSCM (Ping and Dobermann, 2003). The approach used in the current paper did not change the frequency distribution of cleaned yield data, which was only slightly negatively skewed from normal. Standard methods developed to filter error yield data are not able to capture all the erroneous data (Ping and Dobermann, 2005), so continued development of improved filtering methods is required to avoid manual cleaning.

The percentage of yield data removed from each year is within the range reported by others in similar conditions and plant types, however, the use of different filtering methods will result in the removal of different amounts of data due the different algorithms they employ. Filtering algorithms used by Simbahan et al. (2004) removed 13 to 20 % of erroneous data in irrigated and rain-fed maize whereas Ping and Dobermann (2005) removed 16.4% of the original data in an irrigated maize system. Removal of erroneous yield data has been noted to be up to 50% depending on the data and filtering approach used (Sudduth and Drummond, 2007).

Evaluation of Managements Zone Delineation based on Filtering and Data Choice

In most cases, there was little absolute difference in the mean crop yields between MZ-classes from the irrigated maize fields measured in Bulls (Fig. 1). The low variability of the yield of this field is confirmed by the low variability of optical reflectance of the crop canopy, as visually assessed on aerial images of these crops (Google Earth). However, the differences detected between those yields per MZ were statically significant supported by the high number of samples (yield points). The greater absolute differences in the mean yields between the classes during both 2013 and 2014 reflect the natural, rain-fed conditions of the system in Cambridge field (Fig. 2). Better yield consistency with irrigation is due to reduced plant stress with water application to the crops when rain is insufficient also on the lower yielding MZ (Trost et al., 2013). Under the assumption that all classes in the rain-fed Cambridge field received roughly the same amount of water during a precipitation event, the small in-field differences observed between the classes (Fig. 2) must be attributed to other factors.

In most instances, MZ created by jointly clustering soil EC_a and elevation were able to represent grouped yield variability. Properties that influence soil water storage and root growth, including landscape features like slope and aspect, or soil texture, will influence within field variability in crop yields (Kitchen et al., 1999). Higher EC_a values have previously been associated with foot-slope or alluvial areas, where excess water from runoff can accumulate (Fraisse et al., 1999). In non-saline soils, EC_a correlates well with organic matter (Jaynes et al., 1994) and cation exchange capacity (McBride et al., 1990), both of which effect nutrient and water-holding capacity in soil, and therefore influence crop yield (Jaynes et al., 1995). Electrical conductivity may be a suitable surrogate for other soil chemical properties that are not as easily measured, and provide a way to distinguish within field differences associated with root-zone suitability for crops and yield (Kitchen et al., 1999). Others have noted that MZ maps created based on the soil physical properties and topographic attributes such as soil EC_a, and elevation are effective in predicting crop production potential within the field (Fridgen et al., 2004). Future studies may focus on further evidence of suitability of the non-yield based MZ delineation efforts presented in this paper, e.g. on further sites and with more site variables (e.g. patterns of canopy reflectance in aerial imagery).

Conclusion

The method in this paper used easily accessible and free data sources to derive MZ maps to be used in SSCM. Potential MZ derived using a combined clustering of soil EC_a and elevation data were validated with recorded yield data. The benefits of the data used in this study (topography and soil EC_a) are that the data is easily obtained without collecting yield maps with yield monitors over a few years. Soil EC_a surveys are available as an affordable service to consultants and farmers. Geo-referenced elevation data can be obtained through soil surveys or free data sources (i.e. council websites, free aerial images, etc). As not every farmer has access to yield monitoring (Zhang et al., 2010), adoption of the method presented in this paper allows to create MZ that represent growth conditions and thus varying yield potentials. Also farmers that start using precision agriculture technologies could easily create MZ without first collecting yield maps over a few seasons. The suggested method for delineating MZ works better in non-irrigated cropping systems than irrigated systems as the rain fed situation lead to stronger with-in field yield variations.

Acknowledgments

This work was completed as part of *Optimum N – Nitrogen Sensing and Management*, (contract No.: CONT-29854-BITR_LVL) project with funding from The New Zealand Ministry of Business Innovation and Employment, and from Ballance Agri-Nutrients Ltd., New Zealand. We would like to express our special thanks to John Austin Ltd., Te Awamutu and to Waiatapia Station, Bulls for providing data.

References

Blackmore S, Godwin RJ, Fountas S 2003. The analysis of spatial and temporal trends in yield map data over six years. Biosystems Engineering 84(4): 455–466.

Fraisse CW, Sudduth KA, Kitchener NR, Fridgen JJ 1999. Use of supervised clustering algorithms for delineating within-field management zones. ASAE Paper 993043, St Joseph, MI.

Fridgen JJ, Kitchen NR, Sudduth KA, Drummond WJW, Fraisse CW 2004. Management Zone Analyst (MZA): Software for subfield management zone delineation. Agronomy Journal 96: 100–108.

Gotway Crawford CA, Bullock DG, Pierce FJ, Stroup WW, Hergert GW, Eskridge KM 1997. Experimental design and statistical evaluation techniques for site-specific management. *In:* Pierce FJ, Sadler EJ eds. The State of Site-Specific Management for Agriculture, ASA, CSSA, SSSA, Madison, WI. Pp. 301–335.

Hewitt AE 2010. New Zealand soil classification. Landcare Research Science Series No. 1. 3rd edition. Manaaki-Whenua Press, Lincoln, New Zealand.

Jaynes DB, Colvin TS, Ambuel J 1995. Yield mapping by electromagnetic induction. Proceedings of the 2nd International Conference on Site-Specific Management for Agricultural Systems. Pp. 383–394.

Jaynes DB, Novak JM, Moorman TB, Cambardella CA 1994. Estimating herbicide partition coefficients from electromagnetic induction measurements. Journal of Environmental Quality 24: 36–41.

Kitchen NR, Sudduth KA, Drummond ST 1999. Soil electrical conductivity as a crop productivity measure for claypan soils. Journal of Production Agriculture 12(4): 607–617.

Kleinjan J, Chang J, Wilson J, Humburg D, Carlson G, Clay D, Long D 2002. Cleaning yield data. retrieved from http://static1.1.sqspcdn.com/static/f/891472/12244161/1305571958573/Kleinjan+et+al.+2002+Cleaning+yield+data.pdf?token=2U847zuSMlRWaiGHQcluIS5yeNY%3D.

McBride RA, Gordon AM, Shrive SC 1990. Estimating forest soil quality from terrain measurements of apparent electrical conductivity. Soil Science Society of America Journal 54: 290–293.

Ping JL, Dobermann A 2003. Creating spatially contiguous yield classes for site specific management. Agronomy Journal 95(5): 1121–1131. DOI: 10.2134/agronj2003.1121.

Ping JL, Dobermann A 2005. Processing of yield map data. Precision Agriculture 6(2): 193–212. DOI: 10.1007/s11119-005-1035-2.

Simbahan GC, Dobermann A, Ping JL 2004. Screening yield monitor data improves grain yield maps. Agronomy Journal 96: 1091–1102. DOI: 10.2134/agronj2006.0326.

Souza EG, Bazzi CL, Khosla R, Uribe-Opazo MA, Reich RM 2016. Interpolation type and data computation of crop yield maps is important for precision crop production. Journal of Plant Nutrition, 39(4): 531–538. DOI: 10.1080/01904167.2015.1124893.

Sudduth KA, Drummond ST 2007. Yield editior software for removing errors from crop yield maps. Agronomy Journal 99: 1471–1482.

Trost B, Prochnow A, Drastig K, Meyer-Aurich A, Ellmer F, Baumecker M 2013. Irrigation, soil organic carbon and N2O emissions. A review. Agronomy for Sustainable Development (33): 733–749.

Zhang X, Shi L, Jia X, Seielstad G, Helgason C 2010. Zone mapping application for precision-farming: a decision support tool for variable rate application. Precision Agriculture 11(2): 103–114.