# A zone-based approach for processing and interpreting variability in multi-temporal yield data sets

**Article** *in* Computers and Electronics in Agriculture · March 2018
DOI: 10.1016/j.compag.2018.03.029



**5 authors**, including:

Corentin Leroux
Montpellier SupAgro
**8** PUBLICATIONS **9** CITATIONS
SEE PROFILE

Hazaël Jones
Montpellier SupAgro
**45** PUBLICATIONS **154** CITATIONS
SEE PROFILE

James Arnold Taylor
National Research Institute of Science and Technology for Environ...
**75** PUBLICATIONS **666** CITATIONS
SEE PROFILE

Bruno Tisseyre
Montpellier SupAgro
**105** PUBLICATIONS **806** CITATIONS
SEE PROFILE

# A zone-based approach for processing and interpreting variability in multi-temporal yield data sets

Leroux, Corentin (1-2), Jones, Hazaël (2), Taylor, James (3), Clenet, Anthony (1), Tisseyre, Bruno (2)

(1) SMAG, Montpellier, France
(2) ITAP, Montpellier SupAgro, Irstea, Univ Montpellier, Montpellier, France
(3) School of Natural and Environmental Sciences, Newcastle University, Cockle Park Farm, Ulgham, Morpeth, UK, NE65 9EG

cleroux@smag-group.com

**Abstract**

The availability of combine yield monitors since the early 1990's means that long time-series (10+ years) of yield data are now available in many arable production systems. Despite this, yield data and maps are still under-exploited and under-valued by professionals in the agricultural sector. These historical data need to be better considered and analyzed because they are the only audited means by which growers and practitioners can assess the spatio-temporal yield response within a field. When done, time-series of yield maps are mostly processed by classification-based algorithms to generate spatial and temporal yield stability maps or to provide yield or management classes. This work details an alternate segmentation-based methodology to first generate and then characterize contiguous within-field yield zones from historical yield data. It operates on the yield data rather than interpolated yield maps. A seeded region growing algorithm is proposed that enables both the specification of seeds and zone segmentation in a multivariate (multi-temporal yield) attribute space. Novel metrics to assess the yield zoning are proposed that are derived from textural image analysis. The zoning algorithm and metrics were applied to two fields with long time-series (6+ years) of yield data in combinable crops. The two case studies showed that the proposed zone-based approach was effective in delimitating relevant within-field yield zones. The generated zones had differing temporal yield responses between neighbouring zones that were of agronomic significant and interest to the production systems. As this is a first attempt to apply a segmentation algorithm to yield data, areas for future development applications are also proposed.

**Keywords:** co-occurrence matrices**,** historical yield data, temporal stability, segmentation, within-field yield zones

## 1. Introduction

Yield monitors mounted on combine harvesters have been available since the early 1990's. However, yield data still have difficulties in being a decisive component of the decision-making process in precision agriculture studies. In terms of the utility of yield data, multiple issues have been reported by the scientific community. First of all, it is acknowledged that the yield temporal variability is often stronger than the yield spatial variability, which can hinder analyses over short and long-time periods (Blackmore et al., 2003; Bramley and Hamilton, 2004; Eghball and Power, 1995; Lamb et al., 1997). This temporal variability is essentially due to non-stable factors, such as climate patterns or the type of crops being grown each year (Basso et al., 2012). Multiple authors have stated that the number of years of yield data available to conduct yield temporal analyses was critical (Bakhsh et al., 2000; Kitchen et al., 2005) and some have even tried to propose a minimum number of years necessary to obtain reliable results (Ping and Dobermann, 2005). Secondly, it is clear that the spatial yield pattern results from an interaction of management, climate and soil conditions within a cropping season, which means that it is not possible to derive variable-rate application maps directly for a year *n* by solely relying on yield data in year *n-1*. On top of that, yield data often come with a large number of defective observations resulting from the pass of the combine harvester inside the fields. Some of these errors are widely reported in the literature, e.g. flow delay, filling and emptying times, abrupt speed changes or unknown cutting width when entering the crops (Arslan and Colvin, 2002; Sudduth and Drummond, 2007). These errors, if not accounted for, can influence agronomical decisions over the fields (Griffin et al., 2008).

However, from a precision agriculture standpoint, these high-resolution yield data are a very valuable source of information that would be aberrant not to consider (Florin et al., 2009). Yield spatial patterns are a

valuable piece of information to better characterize the sources of spatial variability across the fields. Growers are interested to know about the mean yield spatial and temporal patterns over their fields so they can make informed and reliable management decisions. It has been shown that, despite a strong temporal variability, it was often possible to detect consistent yield spatial patterns across years (Kitchen et al., 2005; Taylor et al., 2007). Be aware that some patterns were found consistent even under different crops and varying climate conditions. Furthermore, yield spatial patterns can deliver relevant information with respect to soil characteristics within the field or can help depict the influence of other external factors, such as management practices and weather conditions (Diker et al., 2004). For instance, Taylor et al. (2007) showed that, in specific portions of their field study, crop rotation management in previous years originated variations in yield spatial patterns. Other authors have found that high-yielding areas in dry years could, at the same time, be low-yielding areas in wet years which could give critical information with respect to within-field soil characteristics (Colvin et al., 1997; Sudduth et al., 1997; Taylor et al., 2007). Another strong advantage of these yield datasets is their accessibility. Indeed, in most cases, harvest has to be made which means that these data can be collected yearly once growers have invested in yield monitors.

The delineation of management zones or management units has long been a subject of interest in precision agriculture because it provides growers with a simple representation of their field. Such zones are defined as spatially contiguous areas over which specific management decisions are to be considered. More than often, management zones are found fragmented in space. This originates from a confusion between the concepts of management classes and management zones (Pedroso et al., 2010). In fact, management classes gather all the management zones over which a specific management decision is to be considered. Authors mostly use classification-based techniques, mostly *k*-means clustering and its fuzzy variant, the fuzzy c-means algorithm (Li et al., 2007; Moral et al., 2010) to delineate these management units. Some others have also proposed some post-processing methods to overcome the fragmentation issue (Ping and Dobermann, 2003). However, as non-spatial algorithms, classification-based methods do not seem to be the most relevant approaches to delineate spatially contiguous areas. One solution could be to make use of object-oriented methodologies from the image processing domain, which aim at delineating objects inside an image (Leroux et al., 2017; Pedroso et al., 2010; Roudier et al., 2008).

Despite the availability of yield data, spatio-temporal yield pattern analysis is not widely done, and when done, is typically applied in an *ad-hoc* or qualitative manner. The industry is missing effective and easily implemented approaches for spatio-temporal yield pattern analysis. The major contribution of this work is to propose a new methodology to analyze historical yield data so that growers and agronomic advisors can better understand the spatio-temporal yield variability in their fields. It must be clear that that the objective of this study is only to look information contained within yield data. It is not, as is typically done with management units, an approach to integrate and simplify crop and environmental variables. In the first instance, the method utilizes a novel multi-dimensional segmentation algorithm that can be applied directly to yield data to define within-field yield zones. The method is therefore not dependent on map production or co-location of information from disparate years. To assess the magnitude and the temporal stability of the yield response within the yield zones, novel metrics adapted from co-occurrence matrix and image textural analyses are then introduced. The algorithm and metrics are derived and then applied to two case studies from arable production systems in France and the UK. The applicability of this novel approach is then discussed including the ability to deliver the processing within a simplified framework that is applicable to non-scientific end-users. Finally, the questions and concerns requiring further work are discussed in the last section.

## 2. Material and methods

### *2.1 Study sites*

The study was conducted on a 20-ha field in England near Amble, Northumberland (WGS84 datum: E: -1.62, N: 55.37) and on a 31-ha field in the north of France near Evreux (WGS84 datum: E: 0.78, N: 48.95). Both fields are cropped in a wheat (*Triticum aestivum*) and canola (*Brassica napus*) rotation and exhibit a relatively strong yield spatial structure. For the English field, wheat yield data were acquired for six years between 2004 and 2015 with a Case combine harvester operating a 10-m cutting front. For the French field, eight years of yield mapping were available spanning the 2003-2011 period. Over the years, the field was mostly harvested with a Claas combine using a 6-m front.

## 2.2 Pre-processing multi-year yield data

Yield data were first cleaned to remove technical errors commonly reported in the literature, e.g. speed changes, unknown cutting width when entering the crop, filling and emptying times and abnormal yield values among others (Arslan and Colvin, 2002; Sudduth and Drummond, 2007). To compare yield data from multiple years with possible significant temporal variations, yield observations were standardized for each year *m* with a mean of zero and a variance of one (Eq. 1):

$$\tilde{Y}_m(i) = \frac{Y_m(i) - \bar{Y}_m}{\sigma_m} \quad \text{Eq. 1}$$

Where $\tilde{Y}_m(i)$ is the i$^{th}$ scaled and centered yield observation in year *m*, $Y_m(i)$ is the i$^{th}$ yield observation in year *m*, $\overline{Y_m}$ is the mean yield in year *m* and $\sigma_m$ is the yield standard deviation in year *m*.

Following a methodology proposed by Blackmore et al. (2003) and Marques da Silva (2006), a grid composed of 20x20m pixels, and whose orientation followed that of the harvested rows, was superimposed on the yield data. For each pixel of the grid, yield values within the pixel were first averaged by year so as to obtain one yield value for each pixel and each year. The objective was to make sure that each year had the same influence in each pixel even if the number of observations falling into each pixel was different from year to year. Empty pixels in specific years due to missing yield observations were given the mean yield value over the years in the same pixel.

## 2.3 Delineating within-field yield zones

### 2.3.1 General description of the algorithm

The objective is to delineate within-field yield zones using a time series of yield data. Within-field yield zones were derived from a seeded region growing algorithm (Adams and Bischof, 1994; Mehnert and Jackway, 1997). This procedure, arising from the image processing domain, starts by selecting a set *S* of *k* observations [$S_1$, $S_2$, …, $S_k$], referred to as the seeds, from which zones are grown. Once the seeds have been chosen, the remaining observations inside the field, i.e. the non-seeds, are recursively associated to an existing seed, given similarity measures between observations. As a consequence, this process expands and grows the zones from the selected seeds. The growing algorithm stops when all the observations have been associated to a zone. Such a procedure has already been applied in the precision agriculture domain but solely with regard to one single agronomic variable (Leroux et al., 2017; Pedroso et al., 2010; Zane et al., 2013). Here, the objective is to extend the procedure to a multi-dimensional case for which there is a need to account for several yield data at the same time. Note that the proposed methodology presents some similarities with that of Leroux et al. (2017).

### 2.3.2 The Multivariate Distance between Pixel Vectors

Several algorithms have been proposed in the literature to segment multiple layers of information, and especially multi-band images, into reliable and informative spatial objects prior to the spectral classification of these delineated objects (Fauvel et al., 2011; Fauvel et al., 2012; Noyel et al., 2007). Most of these methods make use of morphological elements or watershed algorithms, which are extended to multivariate data. Among the different approaches to pass from a one-band image to a multi-band image, Tarabalka et al. (2010) have proposed to calculate a spectral distance between pixel vectors instead of single-valued pixels, where the vector consists of all the variables of interest within a given pixel. In their study, the authors refer to this spectral distance as a vectorial gradient. It is proposed here to make use of the same approach regarding the distance computation between neighbouring pixels. Let *p* be the number of layers considered and let $x_i^p$ be the $i^{th}$ pixel vector of *p* values in the dataset. The multivariate distance between two pixel' vectors $x_i^p$ and $x_j^p$ is set as a multivariate euclidean distance (Eq. 2).

$$\mathrm{d}(x_i^p, x_j^p) = \sqrt{w_1\left(x_i^1 - x_j^1\right)^2 + w_2\left(x_i^2 - x_j^2\right)^2 + \cdots + w_p\left(x_i^p - x_j^p\right)^2} \quad \text{Eq. 2}$$

Where $\mathrm{d}(x_i^p, x_j^p)$ is the multivariate euclidean distance between the pixel vectors $x_i^p$ and $x_j^p$, $w_p$ is the weight associated to the layer *p*, $x_i^1$ is the value of the first layer of the $i^{th}$ pixel vector.

In this study, it is considered that all the available years are given the same importance, i.e all $w_p$ are set to 1.

*2.3.3 The creation of a variance map*

Within-field yield zones are defined here as contiguous spatial entities over which the yield is supposed to be homogeneous. The mean zone yield should however be relatively different to that of a neighbouring zone. As such, by considering the variance between neighbouring pixels, the variance should be relatively low within a zone and exhibit a quite strong peak between pixels belonging to different zones. Be aware that here, neighbouring pixels are pixel vectors, which means that the variance is calculated between vectors of pixels and not between single-valued pixels (Eq. 3). If a seed was to be placed inside a homogeneous zone, i.e. with low variance, and the zone was grown until the boundaries of that zone are reached, i.e. a strong increase in the variance (which will be defined in next section), this region should be well delineated.

The neighbourhood of each pixel vector $x_i^p$ was defined as follows: Let $N_4(x_i^p)$ and $N_8(x_i^p)$ be the 4-order and 8-order neighbourhood of the $i^{th}$ pixel vector respectively (Fig. 1). $H_4(x_i^p)$ is the group of observations such that $x_i^p \cup N_4(x_i^p)$. As such, $H_4(x_i^p)$ contains five observations (j=5), i.e. one central observations and four neighbouring observations. $H_4(x_i^p)_j$ is therefore the $j^{th}$ observation in the neighbourhood. Same goes for $H_8(x_i^p)$ which contains nine observations (j=9). For each pixel vector $x_i^p$, a variance metric $V_i$ was computed as follows:

$$V_i = median\left(\left|H_8(x_i^p)_j - median\left(H_8(x_i^p)\right)\right|\right) \quad \text{Eq. 3}$$

Where $\left(H_8(x_i^p)_j - median\, H_8(x_i^p)\right)$ is a set of distances, i.e. as defined in Eq. 2, between each pixel vector belonging to $H_8(x_i^p)$ and the median of the values of the pixel vectors inside $H_8(x_i^p)$

The formula beyond $V_i$ is in fact the median absolute deviation, a more robust estimate of the variance.

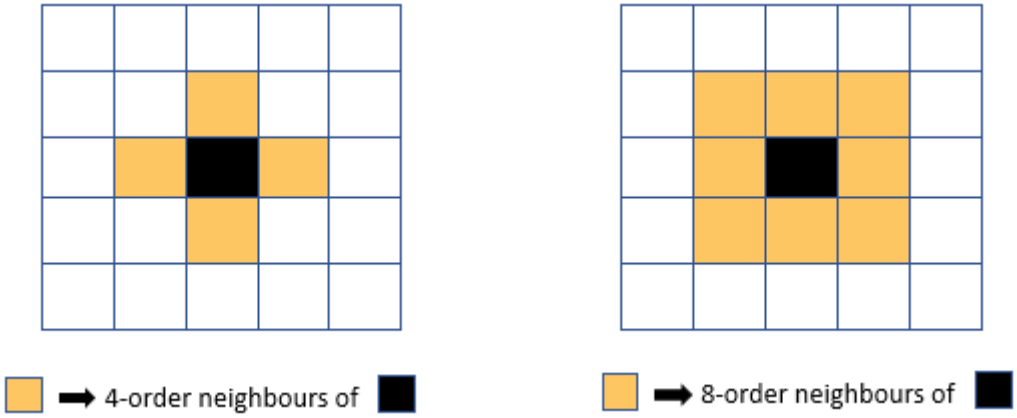



**Figure 1**. The four- and eight-order neighbourhoods of an observation.

2.3.5 *The Seed Selection Process*

At least one seed has to be placed inside each within-field zone to be delineated. As the zones are expanded from the initial *k* seeds, seeds must be carefully located inside the field. Seeds have to share relatively strong characteristics with the observations inside their neighbourhood to make sure that the resulting regions will be homogeneous. As such, seeds were selected as the observations with the lowest variance with respect to their neighbourhood, i.e. the lowest $V_i$. To prevent multiple seeds from characterizing the same within-field zone and to prevent noisy observations from strongly affecting the delineation process, a variance homogeneity criterion was put into place. This criterion is a threshold below which it is considered that there is no need to place another seed because observations are still consistent with the seed previously selected. To define this threshold, the amount of noise $\theta_i$ around each pixel $x_i^p$ was first calculated as in Eq. 4:

$$\theta_i = \sigma\left(H_8(V_i)\right) \quad \text{Eq. 4}$$

Where $\sigma$ stands for standard deviation and $H_8(V_i)$ is the set of variances belonging to the 8-order neighbourhood of the $i^{th}$ pixel.

The step in variance, *Thresh*, is then defined as the mean of the $\theta_i$ distribution. The seed selection process consists in the following steps:

a. Define *G1* as the group containing all the unassigned pixels, *G2* as the group containing all the seeds and *G3* as the group containing all the assigned pixels. At first, all observations belong to *G1*
b. Calculate the step in variance, *Thresh* as defined above
c. Order the observations from the lowest to highest $V_i$
d. Select the first seed, $S_1^p$ as the observation with the lowest $V_i$ and put it in *G2*
e. For each pixel $x_i^p$ inside $N_4(S_1^p)$, if the step in variance is lower than *Thresh* between $V_{S_1}$ and $V_{x_i}$, then $x_i^p$ is put in *G3* because it is considered that $x_i^p$ is consistent with $S_1^p$
f. Repeat step e. for each observation $x_j^p$ inside $N_4(x_i^p)$, and so on until there are no neighbours for which the step in variance is lower than *Thresh*. Be aware that here, the step in variance takes into account the spatial proximity as it is evaluated between $V_i$ and $V_j$.
g. Repeat step d. to f. with the next seed (that with the lowest $V_i$ inside the new set *G1* resulting from the previous iteration) until no future seed can be selected.

The 4-order neighbourhood $N_4(x_i^p)$ was chosen to obtain more compact zones by preventing the zones from expanding diagonally.

#### *2.3.6 The growing of the initial regions*

The set *S* of *k* seeds, i.e. the group *G2* as defined in the previous section, constitutes the starting points of the zones within the fields (see Section 2.3.1). At the end of the growing procedure, there will be as many zones as the number of initial seeds. It must be clear that the methodology detailed in section 2.3.5 was only done to select locations for seeds. The growing of the zones is detailed hereafter. Let Z be the set of *k* zones inside the fields. It must be clear that the $z^{th}$ zone $Z_z^p$ is related to the seed $S_z^p$. At each iteration of the region growing algorithm, the pixel vector $x_i^p$ with the smallest multivariate distance to a neighbouring zone $Z_z^p$, i.e. the smallest d($x_i^p$, $Z_z^p$), is associated to $Z_z^p$. Be aware that as $Z_z^p$ can contain new pixels at each iteration, each value of the p-vector associated to $Z_z^p$ is calculated as the mean of the values of all the pixels belonging to $Z_z^p$. Note that the zones are grown pixel by pixel, i.e. one pixel is attributed to an existing zone at each iteration. The process stops when all the pixels have been associated to an existing zone.

### *2.4 Evaluation of relevance of the zoning*

The objective is to evaluate whether the delineated zones encompass the yield spatial patterns for each year under consideration. If so, in each year *m*, each yield observation inside a zone $Z_z$ should be relatively similar to the mean yield in $Z_z$. On top of that, if the yield variability is spatially structured, the mean yield in $Z_z$ should be quite different to the mean yield in neighbouring zones. Here, it is proposed to make use of a variance reduction-based approach inspired from Fraisse et al. (2001). This method was extended to the multivariate case to cope with the analysis of a time-series of historical yield datasets. The variance reduction index, referred to as *RV*, depicts to what extent the zoning accounts for the spatial variability within the field or, in other words, to what extent the zoning delimitates homogeneous zones. The RV index will be first described in the univariate case, i.e. $RV_m$ for a given year *m*, to ease the understanding and will then be extended to the multivariate case. In this study, only the index RV will be computed. For a specific year *m*, the index $RV_m$ is calculated as follows:

$$RV_m = 1 - \frac{\sigma_m^2(Z)}{\sigma_m^2} \quad \text{Eq. 5}$$

Where $\sigma_m$ is the yield standard deviation in year *m* and $\sigma_m^2(Z)$ is the area-weighted yield variance in year *m* given a zoning Z. The calculation of this latter term is defined in Eq. 6:

$$\sigma_m^2(Z) = \sum_{z=1}^{k} \left(\omega_{Z_z} \times \sigma_m^2(Z_z)\right) \quad \text{Eq. 6}$$

Where $\omega_{Z_z}$ is the weighted area of the zone $Z_z$, *k* is the number of seeds and consequently of zones in the field, and $\sigma_m^2(Z_z)$ is the yield variance within the zone $Z_z$ in year *m*

To extend the $RV_m$ index to the multivariate case and, as such, evaluate the performance of the zoning algorithm over multiple years of yield data, there is a need to refine the variance terms presented in Eq. 5. and Eq. 6. In the unidimensional case, the yield variance within the zone $Z_z$ in year *m* can be simply written as a sum of squared differences between the yield of each observation $x_i$ belonging to $Z_z$ and the mean yield value inside $Z_z$:

$$\sigma_m^2(Z_z) = \frac{1}{n_Z} \sum_{x_i \epsilon\, Z_z} \left(Y_m(i) - \overline{Y_m}(Z_z)\right)^2 \quad \text{Eq. 7}$$

Where $n_Z$ is the number of observations inside the zone $Z_z$, $Y_m(i)$ is the yield of the i[th] observation in year *m* and $\overline{Y_m}(Z_z)$ is the mean yield value inside $Z_z$ in year *m*. Be aware that the calculation is done with the standardized yield values. This notation has not been added for ease of reading.

By using the multivariate euclidean distance defined in Eq. 2, it becomes possible to calculate the yield variance inside $Z_z$ over all the *p* years of study:

$$\sigma^2(Z_z) = \frac{1}{n_Z} \sum_{x_i \epsilon\, Z_z} \mathrm{d}(Y_i^p, Z_z^p)^2 \quad \text{Eq. 8}$$

Where $\mathrm{d}(Y_i^p, Z_z^p)$ is the multivariate euclidean distance between a pixel vector $Y_i^p$ containing the yield values of the *i*[th] pixel for each of the *p* years, and a vector $Z_z^p$ containing the mean yield values in the zone $Z_z$ for each of the *p* years.

The multivariate RV index can then be computed as:

$$RV = 1 - \frac{\sigma^2(Z)}{\sigma^2} \quad \text{Eq. 9}$$

Where $\sigma^2$ is the yield variance over the *p* years, $\sigma^2(Z)$ is the area-weighted yield variance of the proposed zoning over the *p* years.

Note that $\sigma^2$ is calculated similarly as $\sigma^2(Z)$, i.e. in the multivariate space, except that no zoning is considered. The RV index ranges from 0, i.e. very poor delineation to 1, i.e. perfect delineation.

### *2.5 Characterization of the within-field yield zones*

Growers are interested to know about the mean yield spatial and temporal patterns over their fields so they can make informed and reliable management decisions. In most published studies, spatial and temporal stability maps are generated by computing mean and variance yield data over the years (Blackmore et al., 2003; Ping and Dobermann, 2005). Thresholds are generally defined empirically to separate (i) high from low yielding areas and (ii) temporally stable from variable zones. Here, the spatial and temporal stability maps are proposed to be computed at the within-zone level, given that a zoning has been performed previously, and following a methodology inspired from the image processing domain, i.e. using co-occurence matrices and Haralick textural indices (Haralick et al., 1973).

### 2.5.1 *Co-occurrence matrices and yield multi-temporal analysis*

Co-occurence matrices have been originally dedicated to the analysis of texture information inside images. Mostly referred to as $P(i,j,d,\theta)$, these matrices contain the relative frequencies $p(i,j)$ with which two neighbouring pixels of an image separated by a distance $d$ with the orientation $\theta$, occur on the image, one with the information $i$ and the other with the information $j$ (Haralick et al. 1973). To perform a multi-temporal yield analysis at the within-zone level, this same approach can be used to evaluate the relative frequencies $p(i,j)$ with which a zone $Z_z$ has a yield level $i$ in year $m$ and a yield level $j$ in a consecutive year, i.e. $d$ is the temporal distance between consecutive available years. Be aware that the term 'consecutive available years' is used because yield data is missing for some years and with crop rotations, this will be the norm for any temporal yield data analysis. Regarding the proposed methodology, $\theta$ would be 0° as the analysis would be made zone by zone with a temporal sequence of yield levels (Fig. 2). In the case of a multi-temporal yield analysis, co-occurences matrices $P(i,j,d,\theta)$ can then be written $P(i,j,1,0°)$.

### 2.5.2 *Generation of a temporal sequence of yield level for each within-field zone*

For a given year $m$, each zone within the field was characterized by its mean yield level. In order to ease the computation of the co-occurrence matrix, the mean yield value of each zone in year $m$ was given a label according to a classification in $c$ classes. Given that yield values are standardized at the beginning of the method with a mean of zero and a variance of one (Eq. 1), 5 classes of equal intervals were computed between -1 and 1 and labelled 'Very Low' [-1 : -0.6], 'Low' [-0.6 : -0.2], 'Medium' [-0.2 : 0.2], 'High' [0.2 : 0.6], and 'Very High' [0.6 : 1]. For a given zone $Z_z$ and a year $m$, a mean yield level falling into one of these intervals was given the corresponding label of the interval. If the mean yield value of a zone was < -1, it was labelled as 'Very Low' and likewise if it was > 1 then it was labelled 'Very High'.

### 2.5.3 *Computation of specific Haralick indexes*

Once the temporal sequence of yield level has been created for each zone, the co-occurrence matrix $P(i,j,1,0°)$ can be generated. These matrices were normalized to lessen the influence of the number of years available for the analysis (Fig. 2). To evaluate the spatial and temporal stability of the yield patterns within the field, two textural-based indexes defined in Haralick et al. (1973) were computed, i.e. respectively the *Sum Average* and *Sum of Squares* (Fig. 2). Those metrics, referred to as the sixth and fourth Haralick indices are defined as follows:

$$Sum\ of\ Squares = \sum_{i=1}^{c}\sum_{j=1}^{c}(i-\mu)^2 \times p(i,j) \qquad \text{Eq. 10}$$

Where $\mu$ is the mean of the yield classes within the temporal yield sequence and $p(i,j)$ is the probability of having a yield class $j$ consecutively to a class $i$ in the yield temporal sequence.

$$Sum\ Average = \sum_{t=2}^{2c} t \times p_{x+y}(t) \qquad \text{Eq. 11}$$

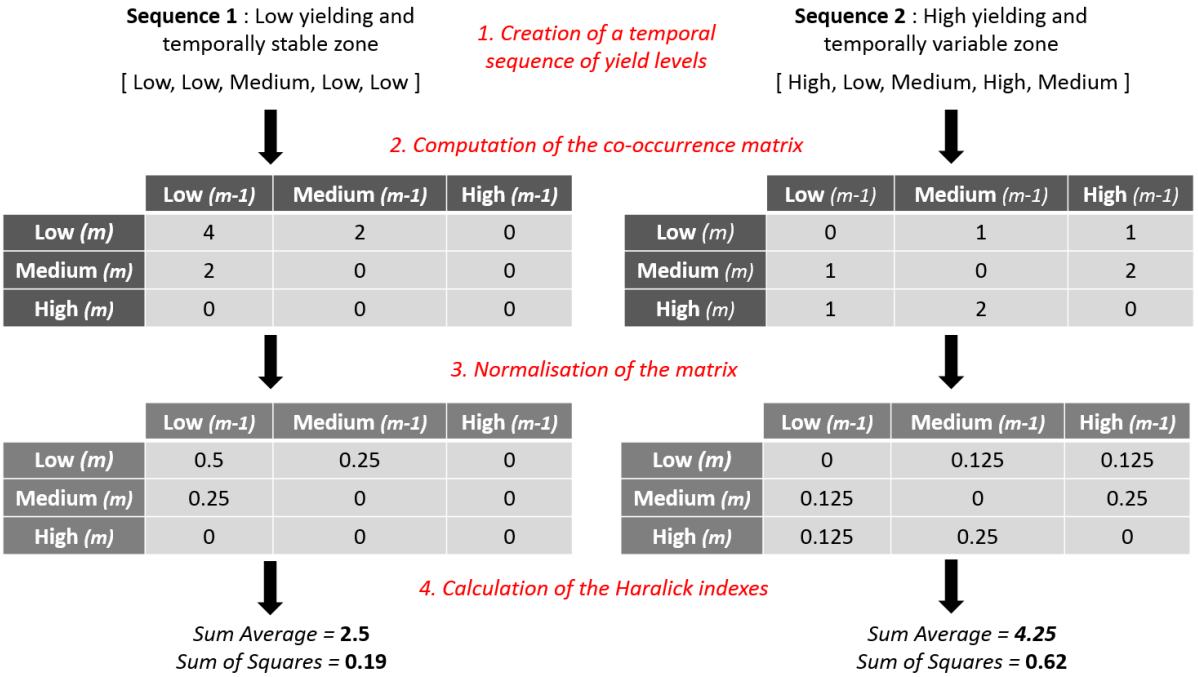
Where $p_{x+y}(t)$ is computed as stated below:

$$p_{x+y}(t) = \sum_{i=1}^{c}\sum_{j=1}^{c} p(i,j)\ , \qquad i+j=t \qquad \text{Eq. 12}$$

The higher the *Sum Average* index, the higher the production level over the years. The *Sum of Squares* ranges between 0 and 1. The closer to 0, the stronger the temporal stability of the yield patterns. The use of these two metrics will allow to obtain a range of spatial and temporal stability levels to help characterize the yield behaviour at the within-zone level across years.

The process involved in the co-occurrence matrix analysis and derivation of the Sum of Squares and Sum Average metrics is then illustrated in Fig. 2. Note that in Fig 2, the number of classes has been reduced ($c$ = 3, 'Low',

297 'Medium' and 'High') for simplicity. The process of 1. *Labelling*, 2. *Co-occurrence matrix derivation*, 3. *Matrix*
298 *normalization* and 4. *The metric calculations* are shown for two contrasting scenarios representing a yield zone
299 with low temporal variance and low yield level (Scenario 1) and yield zone with a high temporal variance and
300 medium to high yield level (Scenario 2).




| | Low (m-1) | Medium (m-1) | High (m-1) |
|---|---|---|---|
| Low (m) | 4 | 2 | 0 |
| Medium (m) | 2 | 0 | 0 |
| High (m) | 0 | 0 | 0 |

| | Low (m-1) | Medium (m-1) | High (m-1) |
|---|---|---|---|
| Low (m) | 0 | 1 | 1 |
| Medium (m) | 1 | 0 | 2 |
| High (m) | 1 | 2 | 0 |

| | Low (m-1) | Medium (m-1) | High (m-1) |
|---|---|---|---|
| Low (m) | 0.5 | 0.25 | 0 |
| Medium (m) | 0.25 | 0 | 0 |
| High (m) | 0 | 0 | 0 |

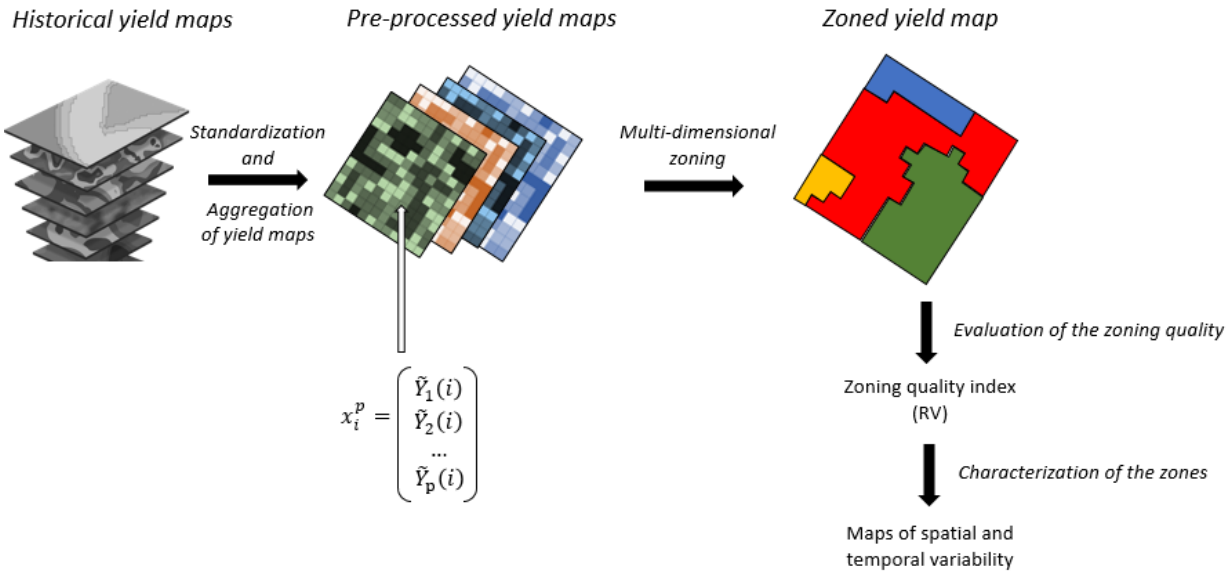
| | Low (m-1) | Medium (m-1) | High (m-1) |
|---|---|---|---|
| Low (m) | 0 | 0.125 | 0.125 |
| Medium (m) | 0.125 | 0 | 0.25 |
| High (m) | 0.125 | 0.25 | 0 |

301

302 **Figure 2**. Characterization of the within-field yield zones in terms of spatial and temporal stability. *Low(m) means*
303 *that the zone has a low yield level in year m. In the top-left hand corner of the co-occurrence matrix for the*
304 *sequence 1, the number 4 means that there were four occurrences of the zone being a low-yielding area in year*
305 *m-1 and in year m. Note that the temporal sequence is read from left to right (two occurrences) but also from right*
306 *to left (two occurrences).*

307 A simple flowchart of the proposed yield multi-temporal analysis is proposed in Figure 3. The whole methodology
308 was developed using the R statistical environment (R Core Team, 2013).

309




310

311 **Figure 3**. Workflow of the proposed yield multi-temporal analysis.

312

## 3 Results and discussion

### *3.1 Summary of yield information from the case studies.*

Figures 4 and 5 show the spatial patterns in the wheat and canola yield data for the years with available data in the two fields under investigation after the yield abnormal values were removed. Associated yield descriptive statistics can be found in Table 1. From a first visual inspection and considering each crop separately, it appears that the spatial yield patterns are consistent within both fields over time. For Field 1, in 2004, 2007 and 2015, the western-part of the field is less productive than the eastern-side. This pattern is reversed in 2012 when it clearly appears that the normally higher-yielding areas in the eastern-part of the field become the lower-yielding areas. This year, 2012, was characterized by a wet growing season that resulted in the lighter soils in the western-part of the field being less water-logged and more productive in this year (Tab. 1). In Field 1, the canola spatial yield patterns do not seem to exhibit any temporal stability, nor do they align with the wheat yield patterns. Moreover, canola observations are noisier and spatial patterns are not as visually distinguishable as those of wheat. From a general perspective, the annual yield variability is relatively low, the coefficient of variation being less than 16% (Tab. 1). Wheat production in Field 1 has significantly increased from 2004 to 2015, starting with a mean at 7.8 tons $ha^{-1}$ (2004) and rising to reaching 12.3 (2015) tons $ha^{-1}$. Note that the minimum yield values remain quite low while the maximum values increase significantly. The rainfall conditions in 2012 did not alter this increasing trend.

**Table 1**. Yield descriptive statistics for the study field over six years. *Yield values are reported in tons per hectare. Rainfall is reported from September to August.*.

| Field | Year | Rainfall (mm) | Crop | Min | 1st quartile | Mean | 3rd quartile | Max | CV (%) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2004 | 802 | Wheat | 3.3 | 7.0 | 7.8 | 8.7 | 11.0 | 15.8 |
| | 2007 | 652 | Wheat | 5.5 | 9.0 | 9.6 | 10.1 | 12.0 | 8.9 |
| | 2011 | 582 | Canola | 2.8 | 4.7 | 5.0 | 5.3 | 6.2 | 9.0 |
| | 2012 | 1160 | Winter Wheat | 5.0 | 9.2 | 9.9 | 10.7 | 13.4 | 12.1 |
| | 2014 | 631 | Canola | 2.0 | 3.3 | 3.7 | 4.1 | 5.4 | 15.6 |
| | 2015 | 401 | Winter Wheat | 5.4 | 11.3 | 12.3 | 13.4 | 16.8 | 13.1 |
| | | | | | | | | | |
| 2 | 2003 | 708 | Wheat | 4.0 | 7.7 | 9.0 | 10.7 | 15.2 | 20.4 |
| | 2004 | 614 | Canola | 0.2 | 1.7 | 3.1 | 4.3 | 8.4 | 54.0 |
| | 2005 | 576 | Wheat | 7.1 | 9.5 | 9.9 | 10.3 | 12.0 | 6.1 |
| | 2006 | 668 | Canola | 0.1 | 1.7 | 2.4 | 3.0 | 6.3 | 42.8 |
| | 2007 | 775 | Wheat | 5.6 | 8.9 | 9.5 | 10.1 | 12.0 | 9.1 |
| | 2009 | 550 | Wheat | 7.9 | 11.2 | 12.0 | 12.8 | 15.0 | 9.1 |
| | 2010 | 612 | Canola | 2.1 | 4.4 | 5.1 | 5.7 | 8.1 | 20.1 |
| | 2011 | 717 | Wheat | 4.7 | 8.5 | 9.6 | 10.7 | 13.2 | 15.4 |

Interestingly, regarding Field 2, the crop type did not affect the spatial yield patterns observed in the field (Fig. 5). In this field, the northern side exhibited generally low yield values, associated with relatively light soils, while the southern section was found to be a high-yielding area. This pattern appears to be reversed for the wheat rotation in 2004 and 2007. Note that the spatial pattern in 2007 is again due to increased in-season precipitation and further interaction with soil characteristics. In 2003, the wheat yield data appears noisier than in other years, which makes the overall yield pattern more difficult to visually detect. Yield observations are also much more variable, high coefficient of variation, when canola is cropped (Tab. 1). Contrary to Field 1, there does not seem to be a clear trend towards increasing yields over time.

When developing this methodology, the authors were aware that it may be difficult to compare the spatial patterns of different crops, such as wheat and canola that belong to different genera that have different yield levels, water requirements and root systems among others. All the spatial patterns were nonetheless plotted to see whether it was conceivable to aggregate the information arising from these two crops under the specific conditions of the field study (Fig. 4 and Fig. 5). Because the spatial canola yield patterns showed little structure and little resemblance to the wheat yield patterns in Field 1, it was decided not to include them in the historical yield data analysis. The main reason for this was to ensure growers were not provided with abnormal or irrelevant information at the end of the analysis. Unfortunately, only two years canola yield data were available for this study. More years would have certainly have enabled a clearer understanding of the spatial yield patterns for this crop in this field.

In contrast, for Field 2 the yield patterns were found to be very consistent within and between crop types so all the years were included in the analysis. Note that this study might still have been conducted for both crops separately. While the intent is to develop an automated approach to yield pattern analysis, ultimately the quality of the analysis will be linked to the choice of data used. Growers and agronomists do need to be conscious of the quality and utility of any data included into the multi-temporal yield pattern analysis.

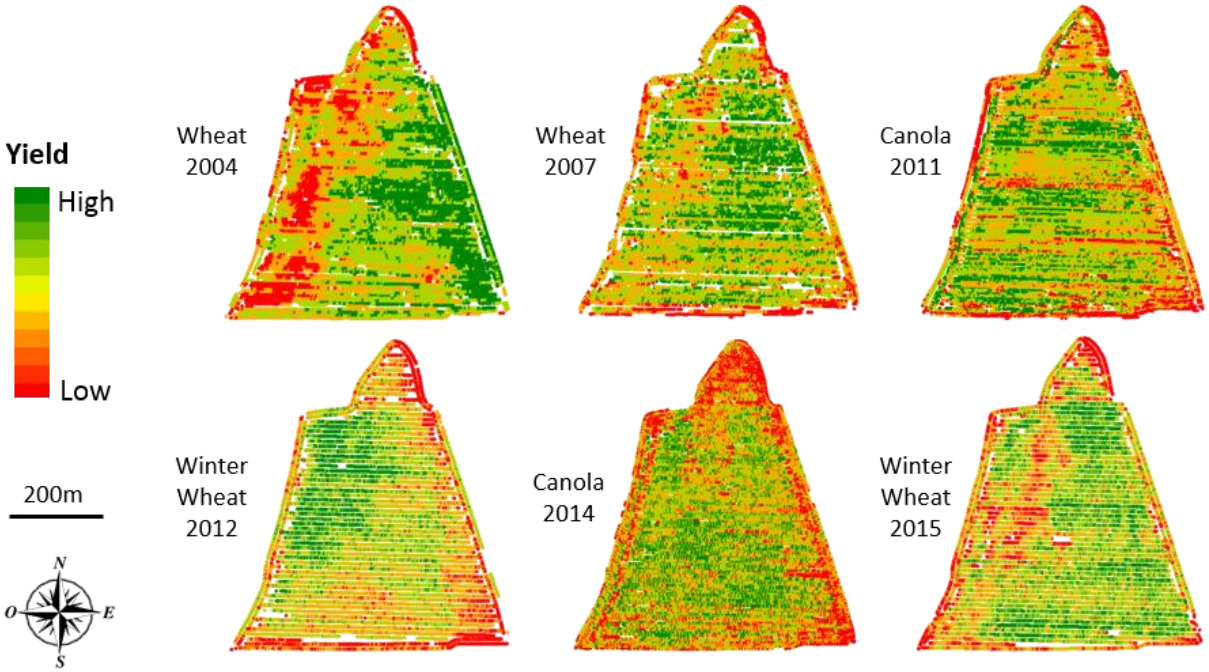



**Figure 4.** Yield spatial patterns in Field 1 for the six years over the 2004-2015 period.

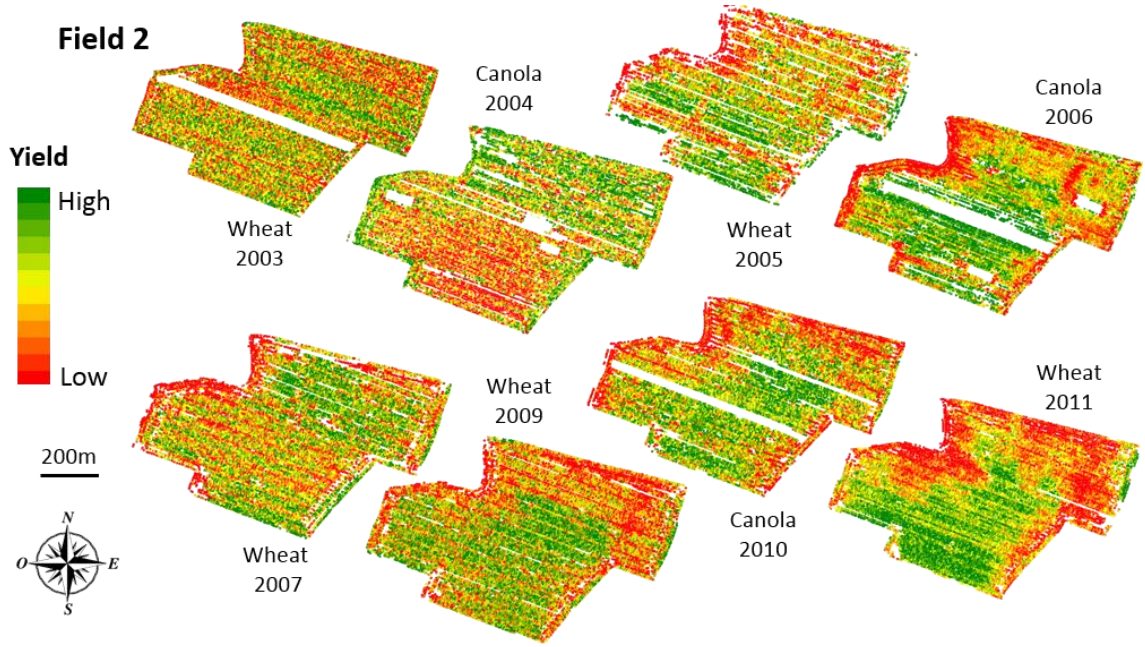



**Figure 5.** Yield spatial patterns in Field 2 for the eight years over the 2004-2011 period

### 3.2 *Evaluation of the resulting within-field yield zones*

The delineation of within-field yield zones appears to be consistent with what could stem from intuitive delineation (Fig. 6 and 7). For both fields the zoning exhibited relatively high RV values, respectively 0.65 and 0.64 for Field 1 and Field 2. Interestingly, both RV values are very similar despite the fact that the number of delineated zones and the number of years of yield mapping available are different for Field 1 and Field 2. In fact, the zoning could have been considered more reliable for Field 1 given that less zones were generated (Z = 13) but it must be understood that the zoning of Field 2 (Z = 17) involved a longer temporal yield sequence. These RV values can be considered high because the zones have been generated from simultaneous analysis of multiple yield, which means that the major yield spatial patterns across the years have been spotted. Note that while the zoning can be considered as being effective, there is still some noise and yield variance within the zones.

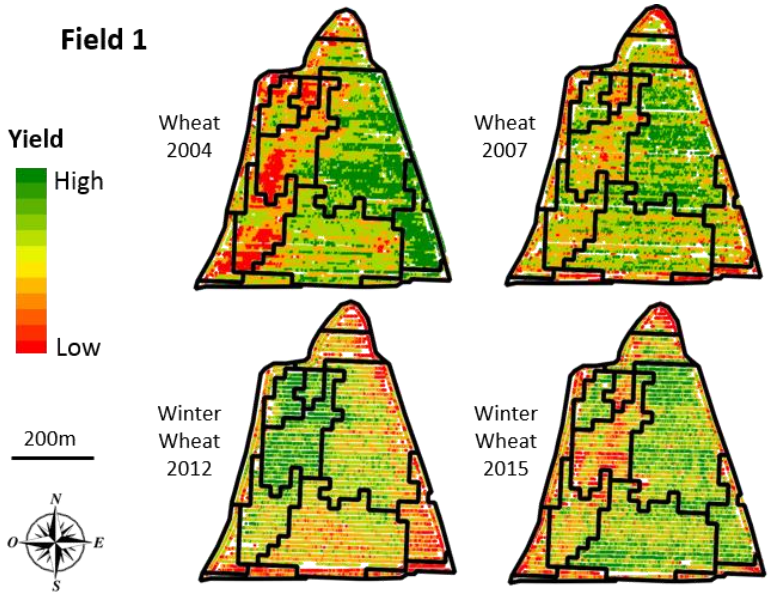



**Figure 6.** Correspondence between yield spatial patterns and within-field yield zones in Field 1.

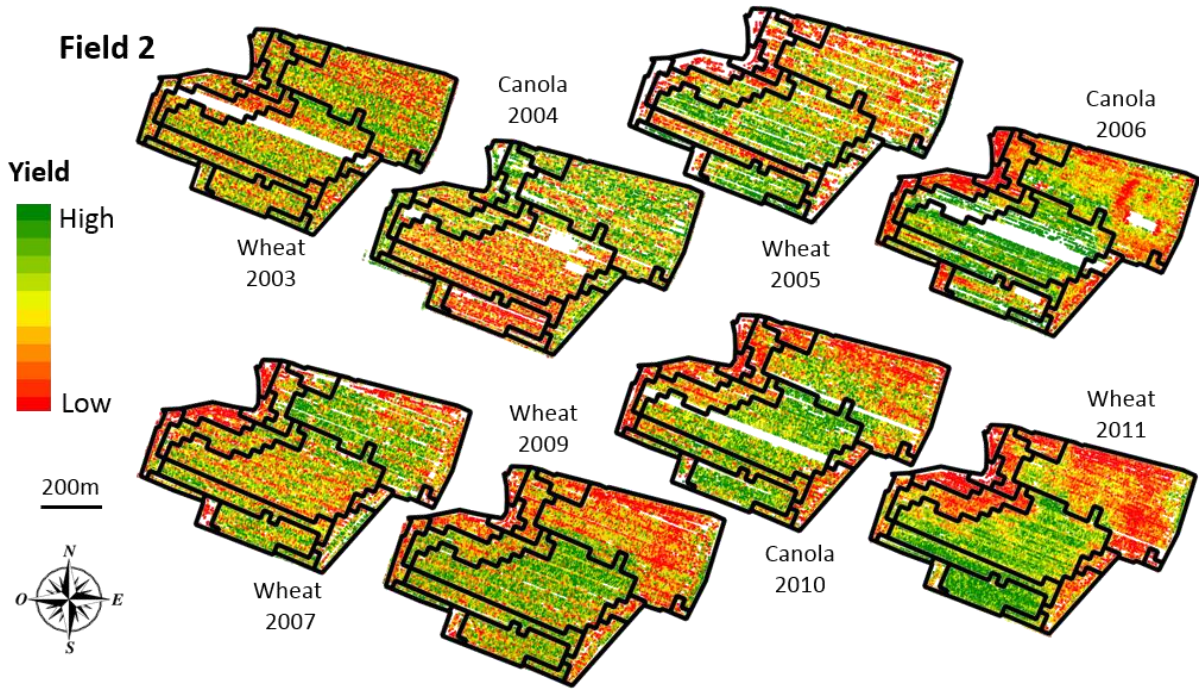



**Figure 7.** Correspondence between yield spatial patterns and within-field yield zones in Field 2.

When looking at the delineated zones more precisely, it appears that in some years, specific zones might not be considered optimal. This is the case for instance for the zone in the north-eastern part of Field 2 for years such as 2003 or 2010. However, it can also be seen that for the remaining years, this zone gathers relatively homogeneous observations. As the delineation accounts for all the years in a single run, it is sometimes difficult to spot year-specific behaviours, especially if the deviations from the general patterns are not strong. These behaviours might be identified by lowering the threshold *Thresh* that was used in the methodology (Eq. 4). From a general perspective, the proposed algorithm delineated quite large and compact zones in both fields, which is considered agronomically desirable even if it is not statistically optimal. The smallest zones are mainly located near the boundaries of the fields, which is generally the place of lower and noisier yield observations. In both fields the transition between high and low-yielding areas is quite clear and it comes with a good level of spatial autocorrelation. These specificities obviously helped the zoning algorithm to delineate relevant within-field yield zones.

Once these zones have been delineated, it is interesting to focus on the differences that these zones exhibit with their direct neighbours (Fig. 8). The major objective of the proposed approach was to delineate relevant within-field yield zones, i.e. zones whose yield behaviour should diverge with that of their neighbours. In the case of similar yield trends across the years, neighbouring zones might benefit from being merged as no clear differences exist between them. In the comparison of Zone 1 with its direct neighbours (Fig 8), it is interesting to see how the rainfall conditions in 2012 affected the yield trends. The yield in Zone 4 substantially increased in 2012, which is likely due to increased precipitation on a lighter soil type overcoming limitations in soil available water on yield in other years. However, all the other neighbouring zones exhibit a decreased yield in 2012, indicating that higher rainfall rates have a negative impact on growing conditions in these areas of the field. Given the strong accordance between the yield trends in Zones 4 and 5, it might be desirable to merge these later to facilitate the interpretation of the maps. Note however that it does not prevent the yield-affecting factors in these two zones being different but the further analysis of these zones is beyond the scope of this work. Be aware that, in this methodology, unlike for example the segmentation algorithms proposed by Leroux et al. (2017), Pedroso et al. (2011), and Roudier et al. (2008), there is no option for region (zone) merging. The number of zones is preset from the seed selection process. The only way to alter the number of zones is to influence the seed selection process. From a general perspective, neighbouring zones in both fields displayed distinct yield trends, validating the proposed zoning delineation.

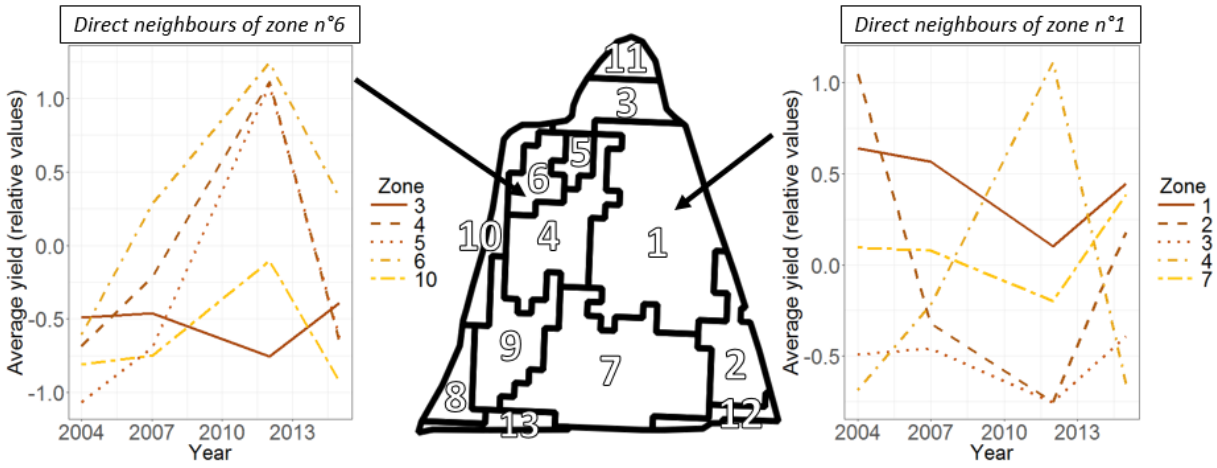



**Figure 8.** Within-field yield zones and corresponding boxplots regarding the mean yield inside neighbouring zones for Field 1.

### *3.3 Analyzing the within-field zones in terms of yield level and temporal stability*

Figure 9 displays the derived yield zones along with the temporal stability of each delineated yield zone. When considering all the study years together, these maps seem to show that Field 1 is composed of (i) a large high-yielding area in the northern-eastern part, (ii) relatively large zones with a medium yield level in the center of the field, and (iii) low-yielding areas near the boundaries. The temporal stability analysis suggests that the western side of the field has an unstable yield pattern over the years which might be of concern for future differential management. However, any differential management plan will need to consider managerial and environmental

conditions within the yield zones. The zones and conclusions here arise solely from the analysis of yield data with no other considerations. There is a need to account for the external factors that impacted the yield patterns, and more particularly the rainfall conditions. It has been discussed previously that the high amount of precipitation which occurred in 2012 in Field 1 (Tab. 1) completely reversed the expected spatial yield pattern (Fig. 4). By considering the four years of yield mapping simultaneously, Zone 4, which was most affected by the variability in rainfall conditions, was given a medium mean yield level over time and was considered temporally unstable. The analysis would have led to different conclusions if the wet growing season, i.e. 2012, had been processed separately. More specifically, Zone 4 would have been labelled a low yielding, temporally stable zone under normal growing conditions and a high yielding area in wet growing conditions. For Field 2, results showed that the spatial yield patterns were more stable over time than for Field 1. This is essentially due to the fact that only four years of yield mapping are available for Field 1 and one out of the four yield data exhibited a complete reverse yield pattern compared to the other years. Reverse yield patterns are also visible for Field 2 but the substantially higher number of years lessened their influence. It can also be seen that when more years are used, it is more difficult to reach a very low value of stability. Yield spatial stability patterns are also well represented in Field 2, with Zones 3 and 1 exhibiting the highest yield level. Contrary to the observed difference in yield temporal stability, both fields show a relatively similar range of yield variation.

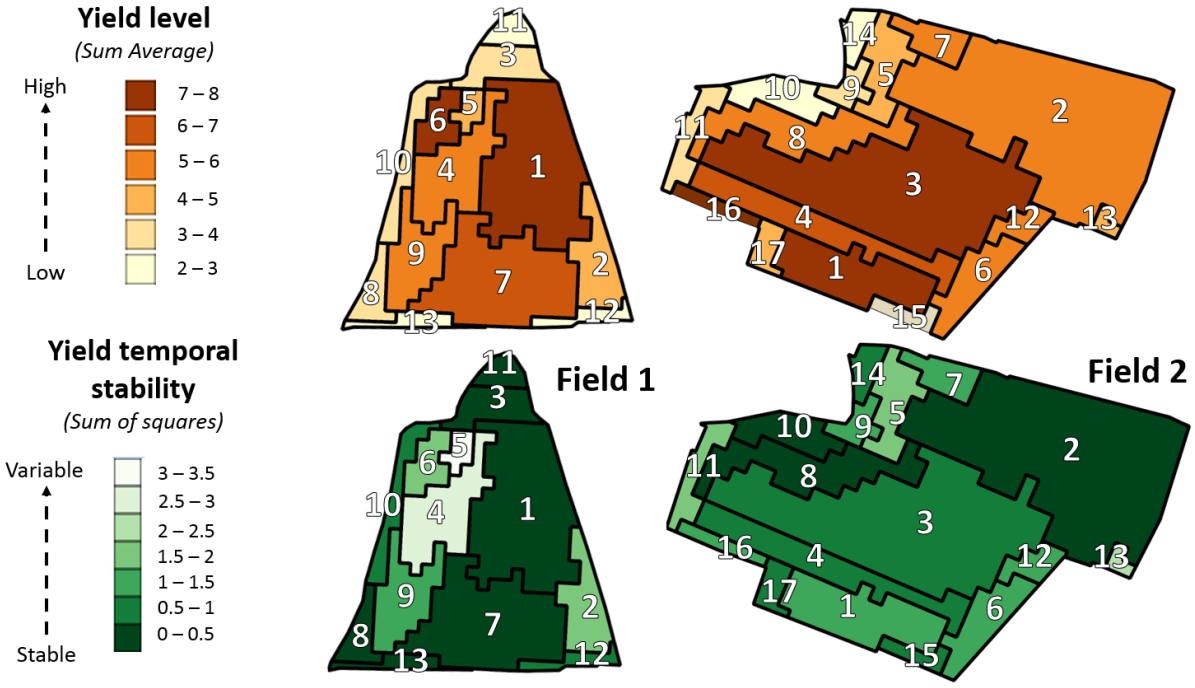



**Figure 9.** Yield level and temporal stability of the within-field zones in Field 1 and Field 2.

The spatial and temporal stability metrics, i.e. *Sum Average* and *Sum of Squares*, are of interest as they enable a quantitative analysis across a relatively wide gradient of variation. The zones are not considered either temporally stable or variable across years but rather they are given a degree of variability over the years of study. This enables growers and operators to obtain more interpretable zones and offers the potential to use personal thresholds given their knowledge of the fields. In this study, the co-occurrence matrices have been computed by considering a temporal distance of one year ($d$ = 1) for the temporal sequence of yield level in each zone. This means that the matrices are solely generated considering a specific year and the direct following or previous year available. In other words, the order of the years in the temporal sequence is taken into account. This might be questionable for annual crops such as wheat and canola. However, here, the use of the *Sum Average* and *Sum of Squares* indicators are very similar to mean and variance indicators which means that, in the end, the order of the years does not matter for these metrics. Nonetheless, this consideration of order might be much more appropriate for perennial crops, such as grape vines, for which consecutive years are much more related. For fields where fixed crop rotations are used, e.g. the alternate Wheat – Canola rotation in Field 2, and when long temporal sequences of yield mapping are available, it might be interesting to adjust the temporal distance accordingly. As such, by making use of a temporal distance of two years (d =2) for Field 2, the same analysis could be conducted on just the wheat or canola crop. This might be of particular importance in longer rotations where first and second wheat crops may want to

be considered separately. The Haralick-based temporal analysis proposed here is a first step towards reliable metrics to describe the spatial and temporal stability of zones, in this case yield zones. This approach could be enhanced further as only two Haralick indicators have been adapted here. Other Haralick indicators may be more suitable for other cropping systems, such as perennial crops, where there is a stronger inter-annual link between years.

### *3.4 Practical considerations for the delineation of within-field yield zones*

Even though the proposed algorithm has been shown to be efficient in delineating yield zones, a couple of considerations still need to be discussed. First of all, it must be said that the concatenation of multiple years of yield data makes it difficult to have a clear understanding of the absolute yield level in each management zone. Indeed, there was a need to work with standardized yield values to lessen the influence of temporal variability in the delineation of within-field spatial units. Without this pre-processing, the yield multi-temporal analysis would not have been as meaningful. Absolute mean zone yields need to be back transformed or calculated when the yield zones are being assessed otherwise growers and advisors would be forced to make decisions using relative and not absolute information. Secondly, it is clear that the history and technical management of the field are crucial when considering which yield data to include in an analysis. Crops with a similar agronomic behaviour might be considered in the same analysis. Saying that, it would be interesting to know whether there would be a possibility to consider some groups of cultivated crops for which yield data could be mixed. In this study, wheat and canola yield data were intentionally analyzed simultaneously for Field 2 given the high consistency in the yield spatial pattern for both crops.

The analysis of historical yield data cannot be considered reliable unless it is based on a significant number of years encompassing a wide range of growing conditions and affecting external factors. However, it is relatively difficult to come up with a minimum threshold of years required due to the diversity in crop production systems in the same region, let alone world-wide. However, given the analysis that was conducted here, it appears risky to obtain reliable conclusions with less than four years of yield data. Knowing the spatial response of the yield to different external factors, particularly climatic variations, will help predict and refine the expected yield spatial pattern at the end of the upcoming growing season. For instance, in Field 1 it is likely that very wet growing conditions will reproduce the spatial yield pattern that occurred in 2012. Furthermore, the proposed methodology makes it possible to vary the weight associated to each yield map, although that has not been done here. This aspect is interesting if the intent is to simultaneously analyze multiple years of yield data while lowering the weight attributed to some of the years, i.e. because of very bad growing conditions or pest/disease effects for instance.

The choice of the grid originating the change of spatial support for the yield data has been little discussed. In this work, a grid size consistent with that used in published studies has been chosen to simplify the processing chain. It must be clear however that changing the grid size will very likely generate unique within-field yield zones outcomes. For example, as the grid becomes coarser, small scale variations will be missed which will prevent small zones from being identified. This effect may be minimal however if the yield data exhibit quite a large spatial structure. One strong advantage of large grids is that they will be able to provide a simple, though less precise, zoned yield map which might help to make decisions. There is an agronomic advantage in interpretation and application to having a grid size that matches the width of field operations. The threshold *Thresh* that is proposed in the study to select the seeds from which the zones are grown is related to the size of the grid that is chosen. Even though this threshold has been selected to be relatively robust relative to the grid size, coarser grids might require the threshold to be decreased to make sure relevant information is not lost. Note also that this study solely considered one grid size for all the yield data. Nonetheless, all the yield data sets come with a different spatial resolution and the location of punctual observations do not match from year to year. The optimal grid size for different years might not be necessary similar. This raises questions regarding the choice of reliable grid sizes to aggregate yield data so as the way to combine those grid sizes if they are different from one year to another.

### *3.5. Perspectives for the analysis of the within field yield zones.*

So far, the analysis has been solely aimed at differentiating zones with specific yield behaviour across years. This work did not intend to propose any particular management of these zones nor to provide growers with variable rate application maps. However, these delineated yield zones might be useful for a further differentiate management within the field. The concept of management zones is fuzzy because it definitely depends on the grower's goal in sub-dividing the field (Kitchen et al., 2005). When using yield datasets to delimitate these zones, three dominant applications will be of interest for growers.

First of all, yield-based regions could help identify yield-limiting or at least yield-affecting factors. As the yield is the result of the combination of multiple factors that can vary over space, the division of a field into spatially homogeneous yield units would facilitate the characterization of these within-field external factors. Some of these drivers might or might not be manageable but the understanding of the underlying factors affecting the yield is decisive for the decision-making process.

Secondly, these yield-based zones can help separate the fields into areas of different potential or productivity (Bochi et al., 2007; Robertson et al., 2008; Taylor et al., 2001). Such analyses are also referred to as yield-gap analyses because there is a difference, to a greater or lesser extent, between what the field actually produces and the productivity that it could achieve (Oliver and Robertson, 2013). A large yield gap means that there is probably considerable space for improvements in management practices and agronomical decisions. There should be more focus on high-potential areas because this is where yield outcomes can be greatly increased.

Finally, the yield zones can help define economically interesting areas for the growers (Massey et al., 2008). Zones that consistently deliver insufficient returns on investments are not worth it, especially if the zones do not a great potential or if the underlying yield-limiting factors cannot be corrected. In addition to production potential, growers also need an indication of the risk associated with achieving production potential for a zone (Marques da Silva, 2006).. The metrics proposed here will assist in economic modelling and determining whether specific yield (or management) zones are worth an investment or whether some management decisions are risky.

## 4 Conclusion

This works presents a methodology to extract and characterize within-field yield zones from a temporal series of yield data. The proposed approach generates contiguous and relatively large yield zones that encompass the general spatial patterns over the years. The efficacy of the zoning algorithm was assessed for spatial and temporal stability using image-based metrics of mean and variance. The methodology was applied to two fields to good effect, with the derived zones and associated metrics raising questions regarding yield performance in space and time and spatio-temporal yield-limiting factors, particularly climatic factors. Rainfall patterns significantly influenced the spatial and temporal stability maps in the fields investigated. Yield zones could be further investigated by evaluating the risk of managing them. This risk analysis could be conducted for each zone through the characterization of multiple components such as the yield-affecting factors, the yield potential, or the return on investments among others.

## Acknowledgments

Authors would like to thank the farmers for given up their data and time on this project. James Taylor's contribution to the work was supported by a USDA-NIFA Specialty Crops Research Initiative Grant.

## References

Adams, R., Bischof, L. (1994). Seeded Region Growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16, 641-647

Arslan, S., Colvin, T. (2002). Grain yield mapping: yield sensing, yield reconstruction, and errors. *CEUR 676 Workshop Proceedings*, 1225, 41–42.

Bakhsh, A., Jaynes, D.B., Colvin, T.S., Kanxar, R.S. (2000). Spatio-temporal analysis of yield variability for a corn-soybean field in Iowa. *Agricultural and Biosystems Engineering*, 43, 31-38.

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.

Blackmore, 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.

Bramley, R.G.V., Hamilton, R.P. (2004). Understanding variability in winegrape production systems. *Australian Journal of Grape and Wine Research*, 10, 32–45

552 Colvin, T.S., Jaynes, D.B., Karlen, D.L., Laird, D.A., Ambuel, J.R. (1997). Yield variability within a central Iowa
553 field. *Transactions of the ASAE*, 40, 883–889.
554 Diker, K., Heerman, D.F., Brodahl, M.K. (2004). Frequency analysis of yield for delineating yield response zones.
555 *Precision Agriculture*, 5, 435–444.
556 Eghball, B., Power, J.F. (1995). Fractal description of temporal yield variability of 10 crops in the United States.
557 *Agronomy Journal*, 87, 152-156.
558 Fauvel, M., Chanussot, J., Benediktsson, J.A. (2011). A spatial–spectral kernel-based approach for the
559 classification of remote-sensing images. *Pattern Recognition*, 45, 381-392.
560 Fauvel, M., Tarabalka, Y., Benediktsson, J.A., Chanussot, J. Tilton, J. (2012). Advances in Spectral-Spatial
561 Classication of Hyperspectral Images. *Proceedings of the IEEE, Institute of Electrical and Electronics*
562 *Engineers*, 101, 652-675.
563 Florin, M.J., McBratney, A.B., Whelan, B.M. (2009). Quantification and comparison of wheat yield variation
564 across space and time. *European Journal of Agronomy*, 30, 212-219.
565 Fraisse, C.W., Sudduth, K.A., Kitchen, N.R. (2001). Delineation of site-specific management zones by
566 unsupervised classification of topographic attributes and soil electrical conductivity. *Transactions of the*
567 *ASAE*, 44, 155-166.
568 Griffin, T., Dobbins, C., Vyn, T., Florax, R., Lowenberg-DeBoer, J. (2008). Spatial analysis of yield monitor data:
569 case studies of on-farm trials and farm management decision making. *Precision Agriculture*, 9, 269–283
570 Haralick, R.M., Shanmugam, K., Dinstein, I. (1973). Texture features for image classification, *IEEE Transactions*
571 *on Systems, Man and Cybernetics*, 3, 610-621.
572 Kitchen, N.R., Sudduth, K.A., Myers, D.B., Drummond, S.T., Hong, S.Y. (2005). Delineating productivity zones
573 on claypan soil fields using apparent soil electrical conductivity. *Computers and Electronics in Agriculture*,
574 46, 285-308.
575 Lamb, J.A., Dowdy, R.H., Anderson, J.L., Rehm, G.W. (1997). Spatial and temporal stability of corn grain yields.
576 *Journal of Production Agriculture*, 10, 410-414.
577 Leroux, C., Jones, H., Clenet, A., Tisseyre, B. (2017). A new approach for zoning irregularly-spaced, within-field
578 data. *Computers and Electronics in Agriculture*, 141, 196-206.
579 Li, Y., Shi, Z., Li, F., Li, H.Y. (2007). Delineation of site-specific management zones using fuzzy clustering
580 analysis in a coastal saline land. *Computer and Electronics in Agriculture*, 56, 174–186.
581 Marques da Silva, J.R. (2006). Analysis of the Spatial and Temporal Variability of Irrigated Maize Yield.
582 *Biosystems Engineering*, 94, 337–349
583 Massey, R.E., Myers, D.B., Kitchen, N.R., Sudduth, K.A. (2008). Profitability maps as an input for site-specific
584 management decision making. *Agronomy Journal*, 100, 52-59.
585 Mehnert, A., Jackway, V. (1997). Improved seeded region growing algorithm. *Pattern Recognition*, Letters, 18,
586 1065–1071.
587 Moral F.J., Terron J.M., Marques da Silva, J.R. (2010). Delineation of management zones using mobile
588 measurements of soil electrical conductivity and multivariate geostatistical techniques. *Soil & Tillage*
589 *Research*, 106, 335–343
590 Noyel, G., Angulo, J., Jeulin, D. (2007). Morphological segmentation of hyperspectral images. *Image Analysis*
591 *and Stereology*, 26, 101-109.
592 Oliver, Y.M., Robertson, M.J. (2013). Quantifying the spatial pattern of the yield gap within a farm in a low rainfall
593 Mediterranean climate. *Field Crops Research*, 150, 29-41.
594 Pedroso, M., Taylor, J., Tisseyre, B., Charnomordic, B., Guillaume, S. (2010), A segmentation algorithm for the
595 delineation of management zones, *Computer and Electronics in Agriculture*, 70, 199–208.
596 Ping, J.L., Dobermann, A. (2003). Creating spatially contiguous yield classes for site-specific management.
597 *Agronomy Journal*, 95, 1121–1131
598 Ping, J.L, Dobermann, A. (2005). Processign of yield map data. *Precision Agriculture*, 6, 193-212.
599 R Core Team (2013). R: A language and environment for statistical computing. *R Foundation for Statistical*
600 *Computing*, Vienna, Austria.
601 Robertson, M.J., Lyle, G., Bowden, J.W. (2008). Within-field variability of wheat yield and economic implications
602 for spatially variable nutrient management. *Field Crops Research*, 105, 211-220.
603 Roudier, P., Tisseyre, B., Poilvé, H., Roger, J. (2008). Management zone delineation using a modified watershed
604 algorithm. *Precision Agriculture*, 9, 233–250.
605 Sudduth, K.A., Drummond, S.T., Birrell, S.J., Kitchen, N.R. (1997). Spatial modeling of crop yield using soil and
606 topographic data. *In Proceedings of the First European Conference on Precision Agriculture*, 439–447.
607 Sudduth, K., Drummond, S. T. (2007). Yield Editor: Software for Removing Errors from Crop Yield Maps.

*Agronomy Journal*, 99, 1471-1482.
Tarabalka, Y., Chanussot, J., Benediktsson, J.A. (2010). Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*, 43, 2367-2379.
Taylor, R.K., Kluitenberg, G.J., Schrock, M.D., Zhang, N., Schmidt, J.P., Havlin, J.L. (2001). Using yield monitor data to determine spatial crop production potential. *American Society of Agricultural Engineers*, 44, 1409-1414.
Taylor, J.A., McBratney, A.B., Whelan, B.M. (2007). Establishing management classes for broadacre agricultural production. *Agronomy Journal*, 99, 1366-1376.
Zane, L., Tisseyre, B., Guillaume, S., Charnomordic, B. (2013). Within-field zoning using a region growing algorithm guided by geostatistical analysis. *Precision Agriculture '13. Proceedings of the 9th ECPA, Lleida, Spain, July 7-11, 2013. J.V. Stafford (ed.). Wageningen Academic Publishers*, 313-319