A New Approach for Zoning Irregularly-Spaced, Within-Field Data

Article *in* Computers and Electronics in Agriculture · August 2017

DOI: 10.1016/j.compag.2017.07.025

4 authors, including:

Corentin Leroux
Montpellier SupAgro

Hazaël Jones
Montpellier SupAgro

Bruno Tisseyre
Montpellier SupAgro

A New Approach for Zoning Irregularly-Spaced, Within-Field Data

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

(1) SMAG, Montpellier, France
(2) UMR ITAP, Montpellier SupAgro, Irstea, France
cleroux@smag-group.com

Abstract

Management zones can be defined as homogeneous regions for which specific management decisions are to be considered. The delineation of these management units is important because it enables or at least facilitate growers and practitioners performing site specific management. The delineation of management zones has essentially been performed by (i) clustering techniques or (ii) segmentation algorithms arising from the image processing domain. However, the first approach does not take into account the spatial relationships in the data, and is prone to generate a large number of fragmented zones while he second methodology has only been dedicated to regularly-spaced, within-field data. This work proposes a new approach to generate contiguous management zones from irregularly-spaced within-field observations, e.g. within-field yield, soil conductivity, soil samples, which are a very important source of data in precision agriculture studies. A seeded region growing and merging algorithm has been specifically designed for these irregularly-spaced observations. More specifically, a Voronoi tessellation was implemented to define spatial relationships between neighbouring observations. Seeds were automatically placed at specific locations across the fields and management zones were first expanded from these seeds. The merging procedure aimed at generating more manageable and interpretable zones. The merging algorithm was defined in a way that made it possible to incorporate machinery and technical management constraints. Experiments demonstrated that the proposed methodology was able to generate relatively compact and contiguous management zones. Furthermore, machinery and technical constraints were shown to significantly influence the results of the delineation which proved the importance of accounting for these considerations.

Keywords: irregularly distributed spatial data, management zones, seeded region growing and merging algorithm, segmentation, variable-rate fertilization

1. Introduction

Site-specific management has become a subject of growing interest due to its ability to deal with the technical, economic and environmental issues of the agricultural domain (Oliver, 2010). A popular tool to implement site-specific management is through the delineation of management zones. These management zones can significantly help growers and agronomy specialists to perform site-specific management across the fields in question. Their delineation will also greatly facilitate all kinds of machinery intervention within these fields.

There must be no confusion between the concept of management zones and the concept of management classes (McBratney et al., 2005), however. Management zones are spatially contiguous entities, i.e. closed sets from a topological point of view, over which a specific application can be performed. On the contrary, a management class is an open set which combines all the management zones over which the same treatment will be applied. It must be noted that in the literature, many authors actually delineate management classes rather than management zones. Indeed, most authors mainly use classification-based methods such as the well-known *k*-means algorithm and its fuzzy variant, the fuzzy *c*-means algorithm (Li et al. 2007; Moral et al. 2010, Peralta et al. 2015). These approaches are generally well-accepted because they systematically find patterns in the data, whether these patterns are actually interesting or not. The authors assume that the variable of interest is spatially organised and that the resulting classes will consequently be organised in zones. However, depending on the level of noise and autocorrelation of the variable under consideration, the resulting management zones may suffer from being highly fragmented within the field. Indeed, no spatial information is taken into account and the variable of interest is simply considered as aspatial information. Multiple improvements have been proposed to overcome this problem. For instance, spatial coordinates have been included in the classification-based (Oliver and Webster, 1989). Although the approach is interesting and enhances the contiguity of the management zones, there is a need to carefully consider the weighing of the spatial coordinates compared to the values of the agronomic variable under study. It must be noted that some management zones are still fragmented to a lesser or greater extent. Other approaches intended to apply spatial filters, either before or after classifying the observation methods (Ping et al., 2003). These techniques help decrease the zones fragmentation but result in the loss of meaningful information. Indeed, spatial filters smooth the information and might mask the existing heterogeneity within the fields, especially if large filters are used.

More recently, segmentation methods arising from the signal processing domain have been used to delineate management zones (Pedroso et al., 2010; Roudier et al., 2008; Zane et al., 2013). One significant advantage of segmentation approaches is that the purpose of these methods is the detection of contiguous objects (Pal and Pal, 1993). Segmentation approaches are very effective when it comes to extracting an object from a significantly different background. For example, segmentation methods are widely used in the medical domain to identify tumours or to delineate organs (Pham et al., 2000). However, in agriculture, management zones are not well-defined objects that can be clearly distinguished from other surrounding objects because the variations in agronomic information between two management zones are not crisp but graduated. There is no prototype of what the segmentation should look for, e.g. in terms of shape or colour. Segmentation approaches were usually implemented so to process regularly gridded data, yet agronomic data can be either recorded on a regular or irregular grid (Taylor et al., 2007). From a general perspective, irregularly-spaced observations can be relocated on a regular grid by interpolation. However, interpolation is likely to affect the distribution of the dataset, can be computationally intensive, and might require skilled operators to perform the process. To avoid this drawback, Pedroso et al. (2010) introduced an approach that is able to process irregularly-spaced datasets (Pedroso et al., 2010).

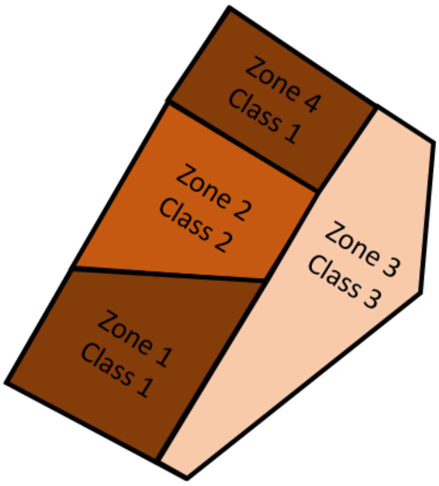

Figure 1. Management classes and management zones. *Four spatially contiguous management zones and three management classes are present within this field. One management class corresponds to one specific application.*

To be fully operational, a variable-rate application map must take into account the machine that will be used to perform the site-specific treatment (Tisseyre and McBratney, 2008), so to avoid generating management zones whose shape cannot be handled by the machinery. A variable rate application map should take into account the following main aspects: (i) the rate changing response time, i.e. the time required to change the applicate rate from a current value to a desired value (Fulton et al., 2001; Fulton et al., 2005), (ii) the accuracy of the application rates, i.e. the ability of the machine to apply a specific rate (Roudier et al. 2011) and (iii) the spatial footprint of the machine, i.e. the minimum area the machinery can deal with (Tisseyre and McBratney, 2008). Recently, some authors have proposed the delineation of rectangular management zones to fulfil the operational requirements of variable rate applications (Cid-Garcia et al., 2013). Their approach intends to find the optimal zoning of the field by minimizing the variance between rectangular management zones of different sizes. This method was proven very effective on a sparse spatial dataset consisting of soil samples. However, the optimality might be much more difficult to reach on highly-dense spatial datasets. Improvements based on semi-variogram analyses have been proposed to address the problems of (i) an insufficient number of sample sites and (ii) the optimal size of rectangular management zones (Zhang et al. 2016). However, this new approach still requires manual supervision for the estimation of variogram parameters and requires an interpolation of the variable to be segmented.

The major contribution of this work is to propose a method to generate management zones for irregularly spaced data while accounting for the technical constraints surrounding the agronomic operation to be performed. First, the proposed segmentation algorithm is presented from a theoretical standpoint. Then, an implementation of the overall approach is proposed. Next, the methodology is tested on real within-field soil phosphorus requirements with the objective of performing a variable rate fertilization. The robustness and sensitivity of the methodology are carefully evaluated.

2. A seeded region growing and merging approach

In the segmentation literature, the region growing and merging approach is one of the most common methods to detect contiguous objects (Pal and Pal, 1993). From a practical standpoint, the growing procedure aims at initializing the detection of objects by expanding small regions into larger ones. This step often leads to the identification of still relatively small zones that need to be merged to define the objects of interest more clearly. The region growing algorithm is generally driven by a set of initial regions, the seeds, from which the regions are expanded (Adams and Bischof, 1994; Mehnert and Jackway, 1997). The growing step is a very good way to account for the trade-off between over and under segmentation. Over-segmentation would occur if each observation available in the dataset was being considered as an initial zone or if the methodology was identifying a very large number of zones. In that case, the computational time of the merging algorithm would be dramatically increased because the number of fusions to be evaluated would be very large. It should be understood that the growing procedure is not compulsory. Management zones might be created by solely relying on a merging algorithm (Pedroso et al., 2010). On the other hand, the number of seeds must not be too low so as to avoid an under-segmentation effect. This latter effect would lead to a loss of information because relevant structures would be missed within the dataset. The proposed methodology involves the use of a seeded region growing and merging algorithm to perform the management zone delineation.

2.1 The Seeded Region Growing Procedure

2.1.1 Concept of seeds in region growing algorithm

From a more theoretical point of view, assume a dataset made of n observations. Let S be the set of k seeds $S_1, S_2, \ldots, S_k$ to initiate the region growing algorithm. Note that the seeds are considered as very small regions that will be expanded. The zones arising from the growing procedure are very sensitive to the choice of these seeds. To choose reliable seeds, three main rules must be followed. First, a seed must be very consistent to the observations inside its neighbourhood to ensure that the regions will be able to grow. Second, if a specific zone has to be delineated within the field, there must be at least one seed inside this zone. Last but not least, seeds for different regions must be disconnected. The selection of seeds is specified in section 3.1.2. Let T be the set of n-k observations inside the dataset that are not a seed $T_1, T_2, \ldots, T_{n-k.}$ The objective of the growing procedure is to recursively associate each observation inside T to an existing region, i.e. one of the k initial regions. To assimilate these observations, there is a need to define neighbouring relationships between observations. In the image processing domain, neighbouring observations are defined as those that share at least a vertex. As images are made of regularly-spaced pixels, these relationships are easy to set up. However, as it was previously stated, agronomic datasets generally gather irregularly-spaced observations. This requires an additional step to define the neighbouring relationship (Pedroso et al., 2010) which will be specified in section 3.1.1.

2.1.2 *Definition of a neighbourhood for each observation*

To be able to define neighbourhood relationships, each irregularly-spaced observation was converted into a small region via a Voronoi tesselation (Fig. 2). The objective was to create small contiguous zones on which the region growing and merging algorithm could be applied. The first-order neighbours and second-order neighbours of x_i will be referred to as $N_1(x_i)$ and $N_2(x_i)$ respectively (Fig. 2). The neighbourhood of each observation x_i was set to involve all the first and second-order neighbours and will be referred to as $N_{1\cup2}(x_i)$. It must be understood that $N_{1\cup 2}(x_i)$ is composed of $N_1(x_i)$ and $N_2(x_i)$. This type of neighbourhood was selected to make sure that each observation had a minimum number of neighbours, especially for the observations near the field boundaries.

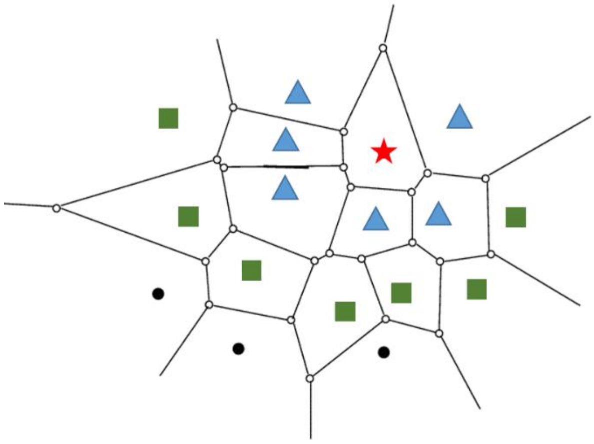

Figure 2. Voronoi tessellation and neighbourhood relationships. *Each square, triangle and circle are the centroids of the corresponding Voronoi polygons. According to the red star on the top of the plot, Voronoi polygons with a blue triangle are first-order neighbours N_1, and those with a green square are second-order neighbours N_2 (first-order neighbours of the first-order neighbours).*

2.1.3 Creation of a variance map

The objective of the zoning approach is to optimize the delimitation of management units within a given field, which requires, according to the proposed methodology, to place at least one seed inside each of these regions. To reach that objective, seeds were selected by using a variance map instead of the raw dataset for two major reasons. Firstly, it was considered that the variance within a region should be relatively homogeneous, the attribute values being consistent among others within that region. Secondly, the variance near the boundaries of two regions should significantly increase. By placing a seed into a homogeneous region, i.e. low variance, and letting the region grow until the boundaries of that region are reached, i.e. a strong increase in the variance, this region should be well delineated. This approach is relatively consistent to that of the watershed algorithm (Roudier et al., 2008) but it is here applied to irregularly-spaced observations.

Hence, for each observation x_i, a variance metric V_i was computed relatively to its first and second-order neighbourhood $N_{1\cup 2}(x_i)$ defined previously. To be more robust to possible outliers inside $N_{1\cup 2}(x_i)$, the variance metric was calculated as the median absolute deviation (Eq. 1).

$$V_i = median\left(f_A(N_{1\cup 2}(x_i)) - g_A(x_i)\right) \quad \text{Eq. 1}$$

Where f_A is an attribute function so that $f_A(N_{1\cup 2}(x_i))$ is the set containing the values of the attribute A of all the observations belonging to $N_{1\cup 2}(x_i)$ and g_A is an attribute function so that $g_A(x_i)$ is the median of the attribute A of the observations belonging to $N_{1\cup 2}(x_i)$.

2.1.4 *Seed Selection Process*

Seeds were considered to be the observations with the lowest variance with their local neighbourhood because the purpose was to spot the most homogeneous initial regions within the fields. However, by following this statement only, there would be a strong probability of selecting spatially-close seeds if a large region containing multiple observations was very homogeneous. Indeed, inside a homogeneous region, neighbouring observations are consistent between each other and there would be as many seeds as observations inside this region. This is not desirable because the region growing algorithm would expand all the neighbouring seeds and create very small management zones, i.e. with only a few points. Once a seed is placed inside a homogeneous region, the region should be able to grow until a strong step in variance is observed which would correspond to the boundaries of this region (Fig. 3). This step in variance needs to be carefully determined because it will be the threshold allowing or not the regions to expand. This threshold must not be set too low because data might be subjected to noise. This noise is likely to increase the variance locally and prevent the regions from growing if the threshold is not high enough. On the contrary, this step in variance must not be too high because two different regions should not be merged together. To define this step in variance, the amount of noise θ_i around each observation was first calculated as in Eq. 2:

$$\theta_i = sd\ (V_{N_{1\cup 2}(x_i)}) \qquad \text{Eq. 2}$$

Where sd stands for the standard deviation, $V_{N_{1\cup 2}(x_i)}$ is the set containing all the variances V of the observations belonging to $N_{1\cup 2}(x_i)$

θ_i can be understood as a criterion of variance homogeneity. The step in variance, *Thresh*, is then defined as the mean of the θ_i distribution. The seed selection process consists in the following steps:

a. Define G_1 as the group containing all the seeds and G_2 as the group containing all the non-seeds. At first, all observations belong to G_1
b. Calculate the step in variance, *Thresh*.
c. Order the observations from the lowest to highest V_i
d. Select the first seed, S_1 as the observation with the lowest V_i
e. For each observation x_i inside $N_1(S_1)$, if the step in variance is lower than *Thresh* between V_{S1} and V_{x_i}, then x_i is put in G_2 because it is considered that x_i is consistent with S_1
f. Repeat step e. for each observation x_j inside $N_1(x_i)$ 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_{x_i} and V_{x_j}.
g. Repeat step d. to f. with the next seed (that with the lowest V_i inside the new set G_1 resulting from the previous iteration) until no future seed can be selected.

In the end, the group G_1 only gathers the final seeds within the field.

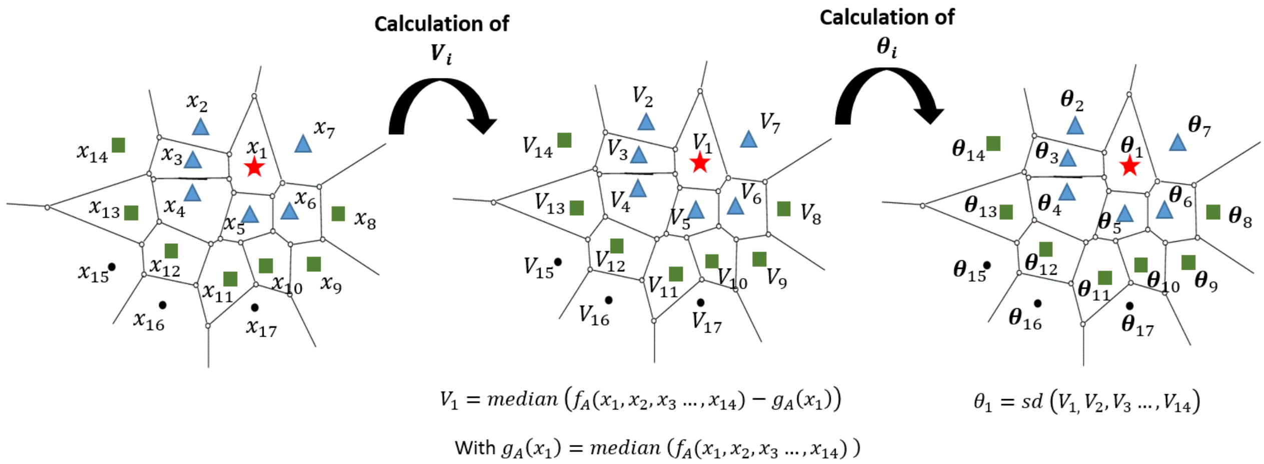

Figure 3. Calculation of a criterion of variance homogeneity for each observation.

2.1.5 *Growing the Initial Regions*

The set of k seeds S_1, S_2, …, S_k are considered as the initial regions within the field. At each iteration of the region growing algorithm, all the first-order neighbours of a given region are considered. Let $N(x_i)$ be the set of observations belonging to the neighbourhood of observation x_i and let f_A be an attribute function so that $f_A(x_i)$ is the value of the attribute A of x_i. Let g be a function so that $g(x_1, x_{2..}, x_n)$ returns the median of the observations x_1, $x_{2..}$, x_n. For a specific observation x_i, if $N(x_i)$ intersects a zone Z_j, then a similarity measure $\delta(x_i,Z_j)$ is computed between x_i and Z_j (Mehnert and Jackway, 1997). This metric is calculated as follows:

$$\delta(x_i, Z_j) = |\ f_A(x_i) - g_{x_j \in Z_j}\left(f_A(x_j)\right)| \qquad \text{Eq. 3}$$

In the case that $N(x_i)$ intersects more than one zone, x_i is associated to the region with which the similarity measure $\delta(x_i, Z_j)$ is the smallest. At each iteration of the growing procedure, the observation x_i with the smallest $\delta(x_i, Z_j)$ is merged with the zone Z_j. Note that the regions are grown one at a time, i.e. at each iteration, and not all together. This is effectively the smallest difference, concerning all the possible seeded regions, that is accounted for. As a consequence, Z_j is grown and the resulting attribute value inside Z_j is calculated as the median attribute value over all the observations belonging to Z_j. The algorithm stops when all the observations inside T have been merged with an existing zone. Note that, at the end of the procedure, there will be as many zones as the number of initial seeds.

2.2 The Region Merging Algorithm

2.2.1 *Merging the Resulting Regions from the Growing Algorithm*

The merging procedure aims at merging the resulting k regions into a set of p ($p < k$) regions. The objective being to reduce as much as possible the over-segmentation phenomenon so that the final objects are well-identified. The merging procedure is controlled by a technical opportunity index that measures how spatially manageable the variable rate application map is (See next section). As the final objective of the management zone delineation is to obtain the highest technical opportunity index possible, the approach aimed at merging iteratively the two neighbouring zones that would lead to the maximum technical opportunity index value. Be aware that the technical opportunity index is calculated over the whole field. This technical opportunity index also helps to choose the optimal number of management zones to be considered within the field.

2.2.2 *Evaluating the technical opportunity of the zoning*

Management zones have to be thought about from an operational point of view, i.e. whether these zones can be treated as they should be by the machine that will perform the application. Opportunity indices have been proposed in the literature to evaluate how fields are spatially structured (Pringle et al., 2003; Oliveira et al., 2007). More recently, new opportunity indices have intended to account for the machinery characteristics to provide a better vision of the operational possibilities for the application (Tisseyre et al., 2008; Roudier et al., 2011). More specifically, Roudier et al. (2011) have proposed a zoning index, ZOI, which evaluates the risks of making an error when performing a variable rate application over a uniform application given a proposed zoning (Eq. 4)

$$ZOI = 1 - \frac{VR_m}{U_m} \quad \text{Eq. 4}$$

Where U_m and VR_m refers to uniform and variable rate management respectively. The closer the ZOI to 1, the better the technical opportunity of the application.

The term U_m is calculated as the sum of squared differences between what should be applied over each observation and the average value of the observations within the field. The term VR_m is also a sum of squared differences but, instead of considering the average value over the field, it uses the average value of each management zones to which an observation belongs. Moreover, the term VR_m is specific to the application under consideration. Roudier et al. (2011) have defined two different risks for a variable rate application, (i) that related to the spatial footprint of the machinery, $Risk_1$ and (ii) that related to the ability for the machine to respect the given prescription, $Risk_2$. These authors have proposed to account for the machinery's spatial footprint by dilating the boundaries of the delineated zones. Here, it is proposed to consider $Risk_1$ by making use of a grid of machine's spatial footprint, so that it is possible to account for the size of the spatial footprint and for the working direction of the machine (Fig. 4). Inside each spatial footprint of the machine, only one treatment level can be applied. If the spatial footprint embraces only one management zone, it is considered that the machine will apply the treatment level associated to that one zone. If a machine footprint straddles multiple management zones, e.g. near the boundaries of these zones, it is considered that the machine will apply the treatment level of the predominant zone. The application error, $Risk_1$ is calculated as the difference between what is applied and what should have been applied. The spatial footprint can be calculated as explained in Tisseyre and McBratney. (2008):

$$Spatial\ footprint = (\beta + \delta)\times(vt + \delta) \qquad \text{Eq. 5}$$

Where β is the width of the machine, v is the speed of the machine, t is the time for the machine to alter the application rate and δ is the positioning inaccuracy.

Risk$_2$ will be evaluated in the same way as in Roudier et al. (2011) by considering the difference between what should be applied within a given region and what the machine is actually able to apply.

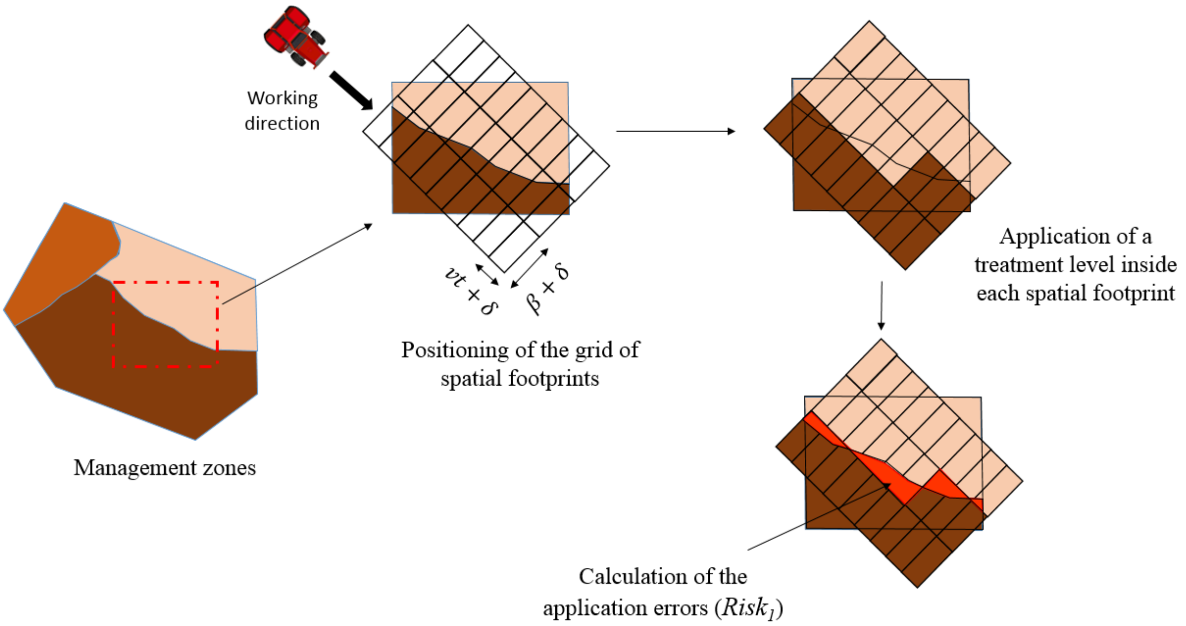

Figure 4. Use of a grid of spatial footprints on the management zones.

To prove the usefulness of the ZOI in delineating management zones, the proposed approach was compared to the methodology of Pedroso et al. (2010). These authors have come up with a zoning method that can be applied to irregularly-spaced datasets. In their approach, each observation is considered as an initial seed, i.e as an initial region. In other words, the authors did not intend to select any particular seeds within the fields. They solely relied on a region merging procedure. In their approach, the best fusion between two neighbouring regions is selected as the one that minimizes the change in the range of the attribute values between the two initial neighbouring regions and the region resulting from the merging of these two regions. In their approach, the technical opportunity of a variable rate application is not taken into account. The objective of this comparison was to evaluate the advantages of considering an opportunity index for the delineation of management zones.

2.2.3 *Finding the Optimal Number of Management Zones*

The behaviour of the ZOI indicator as management zones are merged can be divided into three major sections (Fig. 5). The right side of the plot relates to the constraints of the machinery. It corresponds to the merging of strongly constrained zones, e.g. small or very narrow zones. The ZOI is then expected to reach a maximum value, i.e. the optimal technical opportunity for the given application, because it is considered that all the remaining zones can be managed by the machine. The simplest stopping criterion would be to select the number of management zones for which the ZOI is at a maximum value, *High_MZ*, so that to ensure an optimum variable rate application. From a machinery perspective, this reasoning seems relevant. However, the resulting map might be relatively difficult to read and interpret for a farmer or an advisor because many management zones are likely to remain depending on the size of the machine footprint. As a consequence, there is a need to propose two levels of detail. One for the machine, that has been defined previously, and another more appropriate for analysis and interpretation purposes.

Fusions should be continued to overcome this issue. After reaching the maximum ZOI value, as other management zones are merged, the ZOI indicator should exhibit relatively small fluctuations and then start decreasing abruptly. This strong decline is due to the merging of management zones with strong differences in their attribute values and can be indicative of a future loss of performance. Indeed, these fusions would lead to very non-homogeneous zones and the variable rate application over these zones would be inappropriate. This second threshold, *Low_MZ*, is here defined as the point at which the change in slope is the strongest after the maximum ZOI value has been reached. *Low_MZ* can be identified by calculating angles between consecutive ZOI values and then, selecting the point for which the change in angle value is maximal. *Low_MZ* is associated with a number of management zones necessarily lower than *High_MZ.* Between these two thresholds, the ZOI remains stable which indicates that any number of management zones could be considered in the final management map. However, according to the grower's management strategies, some fusions might be much more appropriate than others. Those strategies, here referred to as management constraints, should also be used to drive the merging process of the algorithm. For instance, a farmer might want to prevent the merging of two management zones whose difference in attribute value is superior to a particular threshold because this latter is meaningful to him. It must be understood that there is a clear difference between the machinery constraints and the management constraints. The former refers to spatial and technical constraints specific to the characteristics of the machinery while the latter is related to the grower's sensitivity and management strategies. Those constraints will be illustrated in the next section.

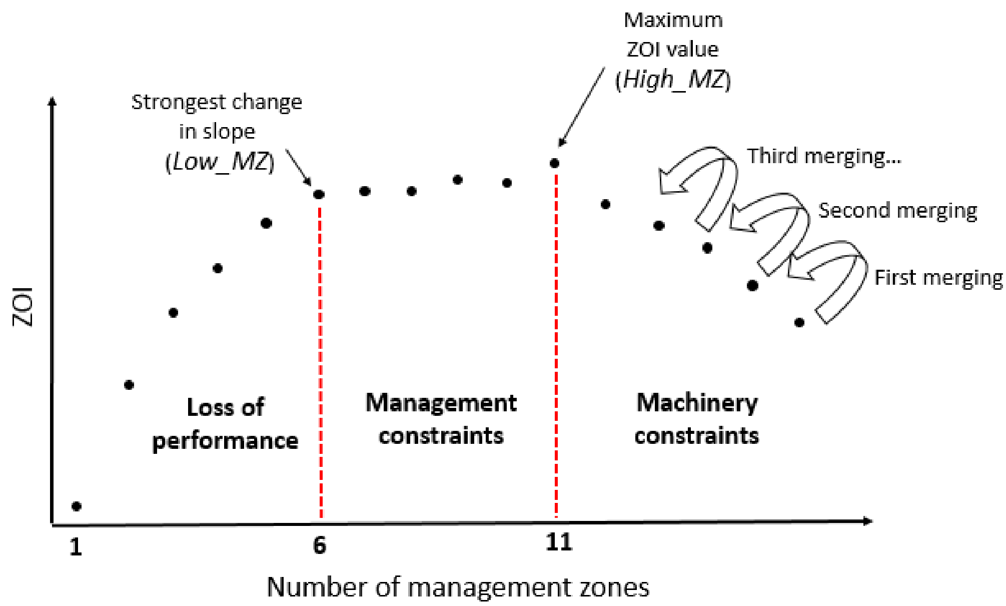

Figure 5. Evolution of the ZOI index with a decreasing number of management zones

It has been stated that the final number of management zones should be set between the two thresholds previously defined as *High_MZ* and *Low_MZ*, i.e. after the machinery constraints and before the loss of performance (Fig. 5). After the maximum ZOI value has been reached, *High_MZ*, fusions are performed until either (i) the minimum number of management zones, *Low_MZ*, has been reached or (ii) there is no more merging that fulfils the management constraints. Be aware that if no management constraints are defined, the merging process solely relies on the ZOI indicator. It should be noted that, if the number of management zones is known in advance or corresponds to a specific request from the farmer, the merging algorithm might be stopped when this number of management zones has been reached.

3. Material and Methods

3.1 Case Study: A Variable Rate of Fertilization

The methodology was tested on two within-field soil phosphorus requirement datasets (Tab. 1). The objective being to create operational and relevant phosphorus management zones. The first dataset arises from a field located close to Peterborough, in England. The second field is located near Evreux, in the north-western part of France. Both fields are cropped with a wheat and canola rotation. These datasets were obtained by using a methodology developed by fertilizer experts combining within-field yield datasets collected with an embedded yield monitor, fertility analyses, soil descriptions and crop history data amongst other criteria (Comifer, 2007 – French denomination).

Table 1. Descriptive statistics and geostatistical parameters of the two fields under consideration.

Field	Min	1st quartile	Mean	3rd quartile	Max	CV (%)	Nugget to sill ratio (%)	Range (m)	Number of points	Size (ha)
1	18.8	158.8	191.8	232.4	306.8	27.6	3.7	120	6415	10.2
2	88.7	239.9	256.0	283.4	334.7	15.4	4.5	74	1480	5.6

Table 2 reports all the standard parameters used for the case study under consideration. For the application under study, common fertilizer spreaders work with a fertilization width between 28 and 36 meters. Considering an average working speed of 10 km/h and an average lag time of 3-4 sec, the length of the spatial footprint was estimated around 10 to 15m (Fulton et al., 2005; Molin et al., 2002). Considering a positioning inaccuracy, δ, of 1m, the width and length of the spatial footprint, $\beta + \delta$ and $vt + \delta$, were set respectively to 30 and 12 m. Given the machinery available, it was considered that the application rates of the machine could be controlled every ten phosphorus units (approximately 20 kg/ha given commonly used fertilizers such as Phosphore 45). The working direction of the fertilizer spreader was considered similar to that of the major orientation of the within-field yield observations, i.e. the working direction of the combine harvester (Fig. 6).

Table 2. Input parameters in the sensitivity analysis for the creation of fertilizer management zones. *Values in bold are the standard parameters used for the zoning delineation*

Type	Criterion	Definition	Associated values
Data quality	*Small scale variations*	Gaussian noise added to each observation (mean value of 0 and an amplitude of +/- X% with regard to the attribute value of each observation)	X = **0**, 10%, 20%, 30%, 40%
Machinery constraints	*Discretization of the machine*	Graduation of application rates that the machine can consider	10, **20**,40, 60 and 80 kg/ha
	Size of the spatial footprint	Size of the spatial footprint	15, 90, **360** and 1440 m^2
	Grid orientation	Orientation relative to the major orientation of the data	**0**, 45°, 90°, 135°
	Grid transposition	Transposition of the grid relatively to the initially defined grid	**0**, 10, 15 m
Management constraints	*Attribute difference and homogeneity*	Constraints that will prevent neighbouring zones from being merged	• **No constraints**, • *M_1*: Mean attribute difference < 40 kg/ha • *M_1*: Mean attribute difference < 60 kg/ha AND *M_2*: Variance difference < 20%

3.1 Sensitivity Analysis of the Proposed Methodology

The proposed zoning algorithm was further evaluated by performing a sensitivity analysis on datasets with specific properties regarding: (i) the initial amount of noise in the dataset, (ii) the discretization rates which the machinery is able to achieve, (iii) the size of the machinery spatial footprint, (iv) the orientation of the machinery spatial footprint grid and (v) the transposition of the machinery spatial footprint grid and (vi) the type of management constraints for the operation under consideration (Table. 2). These criteria were tested one at a time while the others were given a standard value. The influence of the parameters was studied according to the resulting zone delineation and to the evolution of the ZOI indicator during the region merging process. In addition to the definition provided in Tab. 2, the following examples should make these criteria more understandable for users:

- Machinery Constraints
    - Discretization of the machine: A step of 40 kg/ha means that the machinery can only apply treatment rates every 40 kg/ha, i.e. 0, 40, 80, 120… kg/ha
    - The size of the spatial footprint is illustrated in Fig. 4.
    - Grid orientation and transposition are illustrated in Fig. 6
- Management Constraints:
    - A mean attribute difference (M_1) inferior to 40kg/ha means that neighbouring zones for which the difference in attribute value is superior to 40kg/ha cannot be merged. This management constraint, which should be defined by an expert, prevents management zones with different attribute values from being merged.
    - A variance difference (M_2) inferior to 20% means that neighbouring zones for which the variance difference is superior to 20% cannot be merged. This management constraint, which should be defined by an expert, prevents homogeneous zones from being merged with heterogeneous zones.

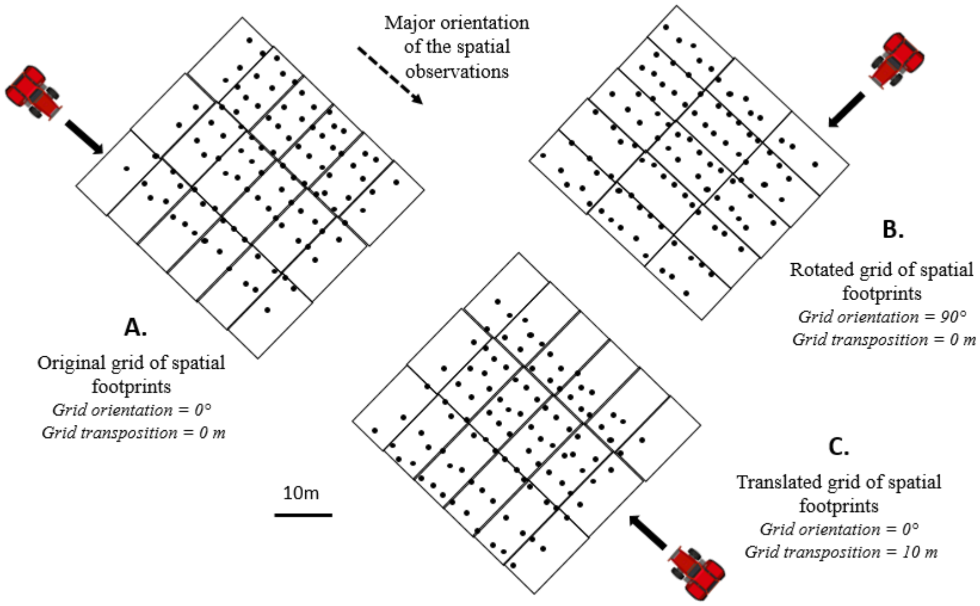

Figure 6. Rotation and Transposition of the Grid of Spatial Footprints.

The whole methodology was implemented using the R statistical environment (R Core Team, 2017).

4 Results and Discussion

4.1 Evaluation of the Seeded Region Growing and Merging Algorithm

4.1.1 The Delimitation of Management Zones

Figure 7 a-c illustrates the process of the management zones delimitation from the raw dataset to the zoning resulting from the region merging algorithm. The region growing and merging algorithm has been performed with a standard parameter setting and no management constraints were applied. The merging procedure solely relied on the ZOI indicator. Figure 7.d shows the zoning obtained following the method of Pedroso et al. (2010) with the same number of management zones as in the proposed approach. In both fields, the resulting delineation appears to be consistent with the one which would result from intuitive delineation (Fig. 7). Both fields exhibit quite a clear spatial structure with a large magnitude of variation and distance of spatial autocorrelation (Tab. 1). Visual inspection of the results shows that the proposed methodology generates more compact zones than that of Pedroso et al. (2010). The latter approach is effectively more sensitive to noise which makes the zones account for every discontinuity within the datasets. For instance, in Field 2, the low-fertilizer requirement zone in the south-east is not identified by the algorithm of Pedroso et al. (2010) to the expense of a much smaller region in the northern-western part of the field. This sensitivity to noise is due to the fact that their approach exclusively relies on a region merging algorithm in which each spatial observation is considered as an initial zone. These very small regions cannot account for the spatial structures within the datasets. Here, this sensitivity to noise is not extremely impacting but could be much more problematic in noisy datasets.

The proposed methodology generates relatively large zones in respect to the field size which means that the seeds could capture non-redundant information (Tab. 3). The region growing algorithm alone is not able to assess whether phosphorus differences between neighbouring zones are significant or not. It is efficient for this purpose in that it merges neighbouring zones that are relatively similar in terms of soil phosphorus requirements while still accounting for the machinery application. Note how large the management zones become after the fusion process (Tab. 3). It can also be seen that the merging step generates fewer zones for Field 2 because the variance of the fertilizer requirements inside this field is much lower than that in Field 1 (Tab. 1).

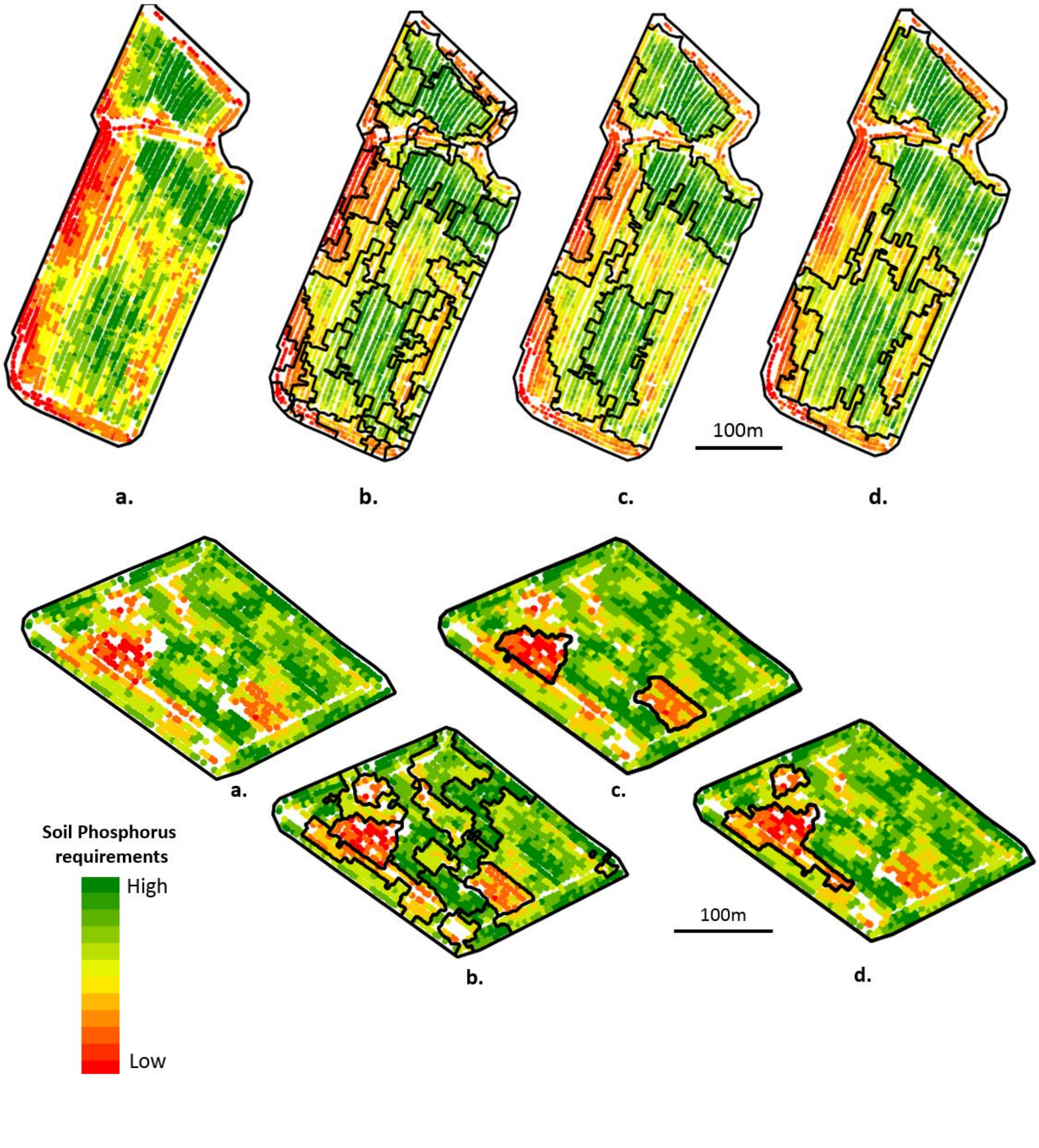

Figure 7. Zoning Soil Phosphorus Requirements for Field 1 (top) and Field 2 (bottom). *a. Raw phosphorus requirements dataset and Phosphorus management zones after the growing procedure (b), after the merging procedure (c) and with the method of Pedroso et al. (2010) (d). No management constraints were considered in that case.*

Table 3. Management Zones Resulting from the Proposed Seeded Region Growing and Merging Algorithm.

Algorithm	Field 1		Field 2	
	After region growing	After region merging	After region growing	After region merging
Number of zones	57	7	20	3
Average area per zone (ha)	0.18	1.46	0.28	1.86
Average number of points per zone	112	914	74	493

The region growing and merging algorithm generates management zones with different soil phosphorus requirements, i.e. soil phosphorus requirements in neighbouring management zones barely overlap (Fig. 8). The seven management zones of Field 1 could be associated with four management classes based on pragmatic grower decisions on fertilizer applications. Field 2 could be categorized as three management classes. The variable application maps appear to be very operational because the characteristics of the application and the machine have been accounted for. All the zones that have been delineated are effectively manageable by the machine.

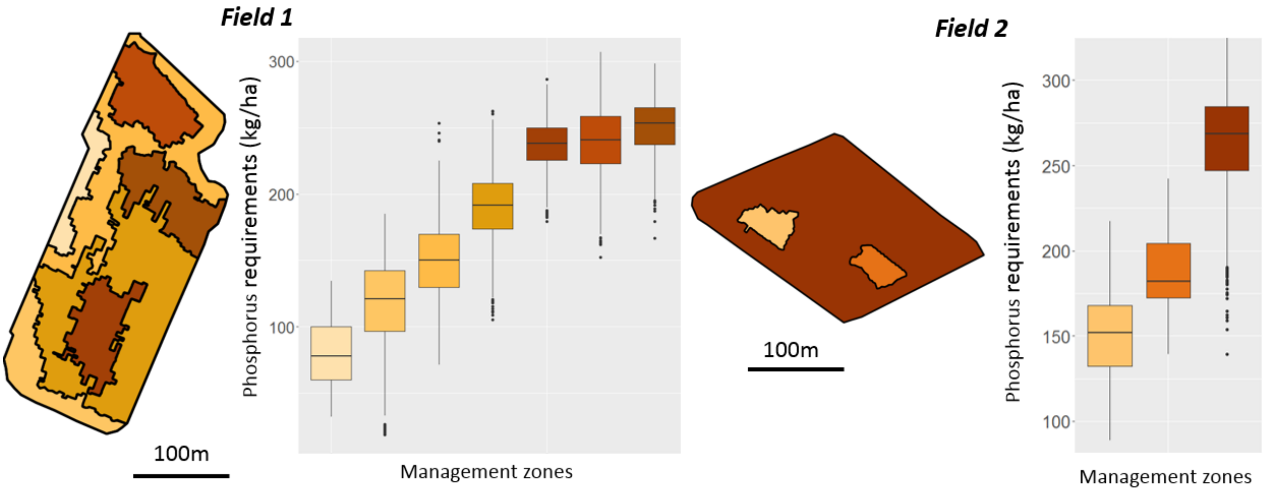

Figure 8. Management Zone Delineation and Corresponding Soil Phosphorus Requirements.

4.1.2 Usefulness of the Zoning Opportunity Index in the Delineation Process

The ZOI value is an indicator of how spatially manageable the field is and therefore also relates to the amount of errors that will be reached after performing a given application. In other words, the lower the ZOI, the higher the number of treatment errors. Field 1 exhibits higher ZOI values than Field 2 which means that Field 1 is more prone to variable rate application (Tab. 4). This could be expected because the gradient observed in Field 1 seems much clearer and because the variance of Field 1 is much higher than that of Field 2, i.e. almost twice as big (Tab. 1). For both fields, the region merging algorithm has been stopped twice, i.e. when the maximum ZOI value had been reached [*High_MZ]* and when the strongest change in slope in ZOI values has been spotted, *Low_MZ* (Fig. 5). It can be seen that, for interpretation purposes, the number of management zones is significantly reduced leading to a decrease in the operation opportunity. Note that the ZOI still remains relatively high.

Table 4. Comparison of the proposed method and that of Pedroso et al., (2010).

Algorithm	Dataset 1				Dataset 2			
	Proposed approach			Pedroso et al. (2010)	Proposed approach			Pedroso et al. (2010)
	After region growing	After region merging			After region growing	After region merging		
		High_MZ	*Low_MZ*			*High_MZ*	*Low_MZ*	
Number of management zones	57	28	7	7	20	13	3	3
ZOI	0.74	0.76	0.67	0.60	0.47	0.48	0.35	0.32

Table 5 also compares the output of the proposed approach and that of Pedroso et al. (2010). The number of management zones has been set similarly for comparison purposes. Optimizing the fusions in terms of opportunity resulted in a higher ZOI value for the proposed approach than for that of Pedroso et al. (2010). The difference in ZOI values between the two methods is not that large, between 0.03 and 0.07. Remember that the ZOI value cannot be superior to 1. This difference represents a gap of 3-7% regarding the errors in application for the same field and the same number of management zones. The methodology to automatically choose the optimal number of management zones for the grower proved relatively efficient because the change in slope in ZOI values is clear enough (data not shown). This approach provides an objective way to identify the optimal number of zones. By looking at the ZOI curve, one could have chosen a slightly different number of management zones. Several evolutions could be taken into consideration to improve the identification of the optimal number of management zones. In the next section, incorporation of management constraints in the merging process is considered.

4.2 Sensitivity analysis of the proposed approach

4.2.1 Influence of noisy datasets and machinery constraints on the delineation of management zones

Figure 9 reports the influence of noisy datasets and machinery characteristics on the ZOI indicator regarding Field 2. All the x and y-axes have been set with the same scale to facilitate the comparison between the different criteria under consideration. As expected, the spatial footprint of the machinery has a strong influence on the ZOI values. Indeed, the calculation of the ZOI indicator is based on the spatial footprint grid. Any change in the mesh of the grid will significantly influence the resulting ZOI value. As the spatial footprint decreases, the opportunity increases because the machine is able to manage smaller zones. According to the other machinery characteristics, changes in the discretization of the treatment levels also generates strong ZOI variations but there seems to be a threshold until which the machine is still considered accurate and ZOI values are consistent, e.g. 40 kg/ha in this case. The results are specific to the field under study. Contrary to the previous parameters, the transposition of the grid does not critically affect ZOI values. It can be seen that the technical opportunity is slightly higher with the standard position of the grid. On the contrary, the grid orientation has more impact, which was also highlighted by Roudier et al. (2011). If the fertilization was to be performed in a direction perpendicular to the major orientation of the observations, the opportunity of the application would be higher. This is interesting because the ZOI might help find the direction of strongest variations within the field. Noisy datasets negatively impact the ZOI indicator. Beyond 20%, the Gaussian noise applied to each observation produces a strong decrease in the ZOI values for two principal reasons: firstly, resulting management zones are heterogeneous which lowers the corresponding ZOI value and secondly, some zones are likely to be missed by the region growing and merging approach. It can also be seen that as the amount of noise increases, the initial number of management zones after the region growing process increases too.

All the ZOI curves follow the theoretical behaviour of the ZOI index that has been previously introduced (Fig. 5) relatively well. However, it must be noted that these curves are not always monotonic. For instance, with the $15m$ setting for the grid transposition parameter, the ZOI is higher for two management zones than for three. This is essentially due to the fact that the merging is performed iteratively by maximizing the ZOI resulting from the fusion of two neighbouring zones. Once a fusion has been made, it cannot be rolled back to check whether another order of merging would have ultimately led to a higher ZOI value. It must be understood that the order of fusion is critical. Here, the merging procedure is not submitted to a global optimization approach but only to a maximization approach at each iteration.

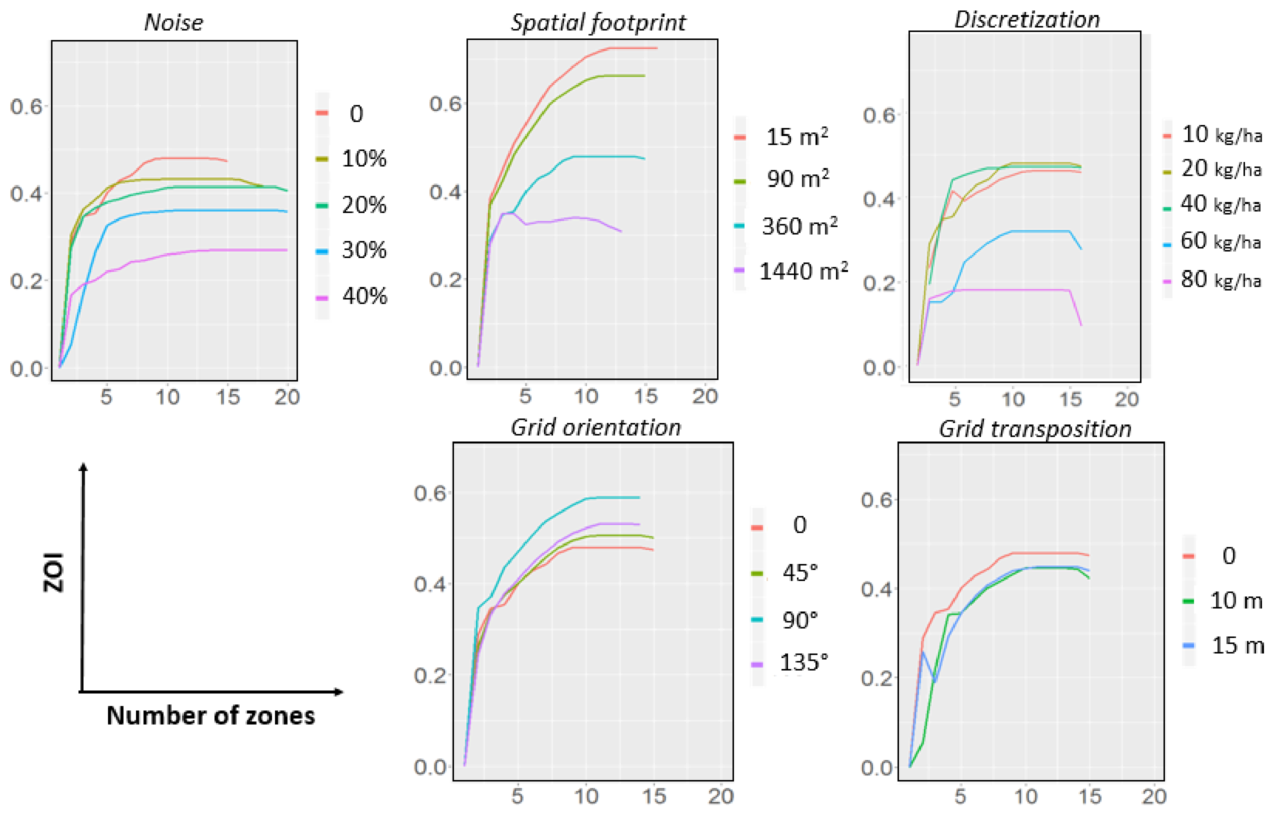

Figure 9. Impact of Noisy Datasets and Machinery Characteristics on the ZOI Indicator for Field 2.

By changing the shapes of the ZOI curves, the parameter setting of the spatial footprint grid, i.e. spatial footprint, grid orientation and grid transposition, will have an influence on the selection of the optimal number of management zones for the grower or the technical advisor. This is interesting because it demonstrates that the resulting management map is sensitive to the operation to be performed within the field. It must be clear that the order of fusion of neighbouring zones is also likely to be affected. Therefore, the resulting zoning might be different but more closely related to the future variable rate application. The influence of management constraints is discussed in the next section.

4.2.2 *Influence of the management constraints on the delineation of management zones.*

Adding management constraints to the region merging algorithm clearly influences the delineation of management zones (Fig. 10). It can be seen that multiple regions delineated after the region growing procedure (Fig. 7b) are not merged and appear in the final management map (Fig. 10). Neighbouring zones are still merged in favour of a maximization of the ZOI indicator but some fusions are not allowed given the management constraints that have been set (Tab. 2). This enables agronomical expertise to be accounted for while still considering the fertilization application to be performed. Note that the management constraints are implemented during the merging process of the algorithm which means that the initial zones, i.e. those that arise from the growing procedure before merging, are the same.

The ZOI indicator alone is a relevant parameter for the region merging algorithm. However, it has some limitations regarding the order of fusions between neighbouring zones, which makes management constraints important to consider. For instance, the ZOI resulting from the fusions of two given regions is not weighed by the surface of these regions. This implies that the merging of two small regions will lead to a lower application error than that arising from the fusion of large regions. Indeed, the ZOI is calculated as a sum of squared differences between the phosphorus requirements at a given spatial position and those actually provided by the machine. Fewer points are available in smaller zones which is likely to make the sum of squared differences lower. Therefore, small regions are likely to be merged first. This might be desirable for the grower but it still depends on management decisions. The incorporation of management constraints makes the proposed approach more sensitive to agronomic expertise and more related to the objectives of the grower. Here, simple management constraints have been proposed and especially in relation to fertilization applications.

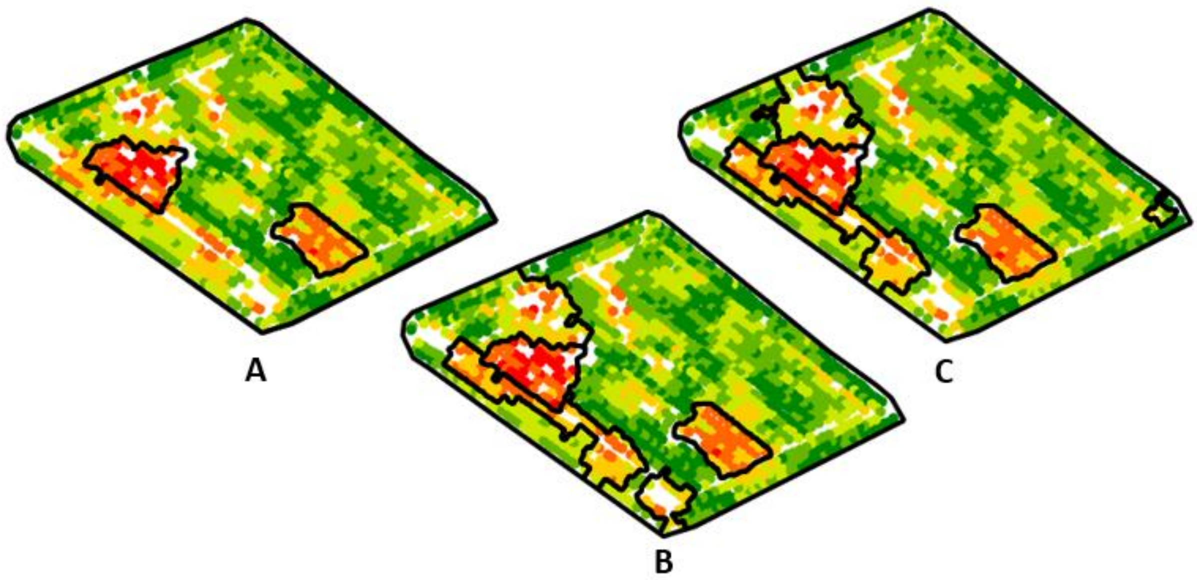

Figure 10. Impact of Management Constraints on the Resulting Variable Rate Application Map. *A*. Without management constraints. *B*. $M_1 < 40$kg/ha. *C*. $M_1 < 60$kg/ha and $M_2 < 20\%$

5 Conclusion

This work presents a region growing and merging procedure adapted to the delineation of management zones from irregularly-spaced within-field data. Overall, the proposed algorithm generates contiguous and compact management zones which are consistent with what could arise from intuitive delineation. These management zones are delineated from an operational point of view thanks to the use of machinery and technical management constraints in the zoning algorithm. Both constraints were shown to have a significant influence on the results of the delineation through a sensitivity analysis. Moreover, the methodology proposes to delineate management zones by integrating a range of possible zonings for the same field. The first zoning, very detailed, intends to maximize the technical opportunity of the application while the second, coarser, aims to simplify the representation of the field while making sure that the differences across the field are still well highlighted. This approach raises questions regarding the incorporation of other constraints, i.e. economic and environmental, which would enable to account for the specificity of the situation under consideration and to propose an optimal number of management zones. To further investigate the methodology, it could be interesting to improve the merging step of the zoning algorithm by implementing a global optimization approach. By doing so, it would be possible to roll back in the merging process and look for fusions that would ultimately lead to a better delineation of the field. Finally, the proposed approach should be tested on multiple datasets of different natures to ensure its robustness.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments which significantly improved the manuscript.

References

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

Arun, P.V. (2013). A comparative analysis of different DEM interpolation methods. *The Egyptian Journal of Remote Sensing and Space Science*, 16(2), 133-139.

Cid-Garcia, N.M., Albornoz, V., Rios-Solis, Y.A., Ortega, R. (2013). Rectangular shape management zone delineation using integer linear programming. *Computer and Electronics in Agriculture,* 93, 1–9.

Comifer (2007). Teneur en P, K et Mg des organes végétaux récoltés pour les cultures de plein champ et les principaux fourrages, Comifer, Paris, 6 pages.

de Oliveira, R. P., Whelan, B. M., McBratney, A. B., Taylor, J. A. (2007). Yield variability as an index supporting management decisions: YIELDex. *Proceedings of the 6th European Conference on Precision Agriculture,* 281–288.

Fulton, J.P., Shearer, S.A., Chabra, G., Higgins, S.F. (2001). Performance assessment and model development of a variable-rate, spinner-disc fertilizer applicator. *Transactions of the ASAE*, 44, 1071-1081.

Fulton, J. P., Shearer, S.A., Higgins, S.F., Darr, M.J., Stombaugh, T.S. (2005). Rate response assessment from various granular VRT applicators. Transactions of the ASAE, 48(6), 2095-2103.

Li, Y., Shi, Z., Li, F., Li, H.Y. (2007). Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. *Computer and Electronics in Agriculture*, 56, 174–186.

Maini, R., Aggarwal, H. (2009). Study and Comparison of Various Image Edge Detection Techniques. *International Journal of Image Processing*, 3(1), 1-11.

McBratney, A., Whelan, B., Ancev, T., Bouma, J., (2005). Future directions of precision agriculture. *Precision Agriculture,* 6, 7-23.

Mehnert, A., Jackway, V. (1997). Improved seeded region growing algorithm. *Pattern Recognition*, Letters 18(10), 1065–1071.

Molin, J. P., Menegatti, L.A.A, Pereira, L.L., Cremonini, L.C., Evangelista, M. (2002). Testing a fertilizer spreader with VRT. In Proc. World Congress of Computers in Agriculture and Natural Resources, 232-237

Moral F.J., Terron J.M., Marques da Silva, J.R. (2010). Delineation of management zones using mobile measurements of soil electrical conductivity and multivariate geostatistical techniques. *Soil & Tillage Research*, 106, 335–343

Oliver, M.A., Webster, M. (1989). A geostatistical basis for spatial weighing in multivariate classification, *Mathematical Geology*, 21, 15-35.

Oliver, M. A. (2010). *Geostatistical Applications for Precision Agriculture*, Springer, London, UK, 295 pp.

Pal, N.R., Pal, S.K. (1993). A review on image segmentation techniques. *Pattern Recognition*, 26, 1277–1294.

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.

Peralta, N.R., Costa, J.L., Balzarini, M., Franco, M.C., Córdoba, M., Bullock, D. (2015). Delineation of management zones to improve nitrogen management of wheat. , *Computer and Electronics in Agriculture*. 110, 103–113.

Pham, D.L., Xu, C.Y., Prince, J.L. (2000) A survey of current methods in medical image segmentation. *Annual review of biomedical engineering*, 315–337.

Ping, J.L., Dobermann, A. (2003). Creating spatially contiguous yield classes for site-specific management. *Agronomy Journal*, 95, 1121–1131

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

R Development Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Roudier, P., Tisseyre, B., Poilvé, H., Roger, J. (2008). Management zone delineation using a modified watershed algorithm. *Precision Agriculture*, 9, 233–250.

Roudier, P., Tisseyre, B., Poilvé, H., Roger, J. (2011). A technical opportunity index adapted to zone specific management. *Precision Agriculture,* 12, 130–145.

Taylor J., McBratney A.B., Whelan, B. (2007). Establishing management classes for broadacre agricultural production. *Agronomy Journal* 99 1366–137.

Tisseyre, B., McBratney, A. (2008). A technical opportunity index based on mathematical morphology for site-specific management: an application to viticulture. *Precision Agriculture*, 9(1-2), 101–113.

Vincent, L., Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6), 583–598.

Zane, L., Tisseyre, B., Guillaume, S., Charnomordic, B. (2013). Within-field zoning using a region growing algorithm guided by geostatistical analysis. *In Proceedings of Precision Agriculture*, 313-319

Zhang, X., Jiang, J., Qiu, X., Wang, J, Zhu, Y. (2016). An improved method of delineating rectangular management zones using a semivariogram-based technique. *Computers and electronics in Agriculture*, 121, 74-83