An iterative region growing algorithm to generate fuzzy management zones within fields

Leroux, C.[1,2], Jones, H. [2], Tisseyre, B. [3]
[1]SMAG, Montpellier, France
[2]ITAP, Montpellier SupAgro, Irstea, Univ Montpellier, Montpellier, France
cleroux@smag.tech

Abstract

Management zones are within-field homogeneous spatial units that have specific limiting factors. While most algorithms tend to delineate fixed management zones' borders, it is acknowledged that within-field variations between management zones are more graduate than crisp, mainly because of smooth variations in soil and landscape characteristics. In this work, an approach is proposed to account for the transient observations between management zones by delineating within-field fuzzy management zones. It is effectively considered that observations within transition zones could belong to more than one zone, and that this should be accounted for during the delineation process. The proposed approach requires to dispose of a pre-zoning of the data of interest, which is referred to as the initial zones. An iterative region growing algorithm is applied on within-field data to compute the frequency at which observations belong to each of these initial zones (membership frequency). Observations with a high membership frequency for a specific zone are considered to be part of the core (or kernel) of the zone, while the others are considered to be part of transitions areas (or support of the zone). The application of this iterative algorithm enables to generate a membership frequency map that helps to better visualize the size of the core and transient regions of each initial zone. These membership frequencies are then used to refine the initial zoning while accounting for user's expertise and management decisions. The proposed approach was tested and evaluated on a real within-field yield dataset in France. Theoretical fertilization-based case studies, i.e., high and weak environmental constraints, were considered to show the interest of membership frequencies to refine yield-based management zones when user's expertise and management decisions come into play.

Keywords: Fuzzy delineation, management zones, membership frequency, transition areas

Introduction

The delineation of within-field zones is widely reported as a means to improve the management of cropping systems within the scope of Precision Agriculture. Multiple works were effectively aimed at generating homogenous within-field units, mostly using classification- (Li et al., 2007; Peralta et al., 2015) or segmentation-based (Leroux et al., 2017; Roudier et al., 2008) algorithms. However, in agriculture, management zones are not well-defined spatial units that can be clearly distinguished from other surrounding spatial entities. Indeed, the variations in agronomic information between two management zones are not crisp but rather graduate, mainly because of smooth variations in soil and landscape characteristics across fields. While the borders between management zones are generally considered fixed in space, the presence of transient information between zones would be better taken into account if these borders were defined from a fuzzy perspective (Zadeh, 1964). Indeed, the membership degree of observations inside these transition areas with regard to a specific zone is not clear. These observations could effectively be associated to more than one zone.

As suggested by Paoli et al. (2007), a fuzzy delineation of management zones would be of interest to agricultural professionals because those zones could be adapted to their will and expertise, and to specific within-field applications. For instance, a wine producer might be interested in harvesting only the grapes within the top-quality zones, leaving behind those that could belong to transition zones to make sure that the vine-making process is optimized. Jones et al. (2016) have already intended to tackle this problem using simulated datasets by allowing the borders between management zones to either expand or shrink depending on a quantile separation of the data. It is also acknowledged that multiple studies have proposed to use the fuzzy variant of the classical k-means algorithm to generate management zones (Li et al., 2007; Peralta et al., 2015; Tagarakis et al., 2013). However, the main problem is that, in most cases, the observations' membership degree to each zone is seldom considered, meaning that the fuzzy approach has not much advantage with respect to a more classical k-means algorithm. Note however that Urretavizcaya et al., (2013) proposed an approach to account for membership degrees obtained with a fuzzy c-means algorithm to generate compact areas with significant oenological differences between them.

In this work, it is considered that within-field data have already been zoned, i.e., an initial zoning is available. The objective of this study is two-fold. First, an iterative region growing algorithm is applied on the within-field data to compute the frequency at which these observations belong to each initial zone. Then, the observation membership to each initial zone is used to adapt the initial zoning to the user's expertise and management decisions. The approach was tested and evaluated on a real within-field yield dataset (wheat) coming from a French farm.

Material and methods

Dataset used

The approach was tested on a real within-field yield dataset arising from a farm located near Verdun, in the north-eastern part of France (WGS84: E:5.568, N:49.378). This 10-ha field was cropped in wheat and harvested with a Claas (Harsewinkel, Germany) combine. Yield data were first filtered and the zoning algorithm of Leroux et al. (2017) was then applied on these within-field yield data, which led to the generation of three within-field zones (Fig. 1). Note that the high-yielding area in the bottom left-hand corner of zone 1 was considered too small to be delineated as a specific zone by the algorithm of Leroux et al. (2017). For visualisation purposes, yield data were separated into five quantiles.

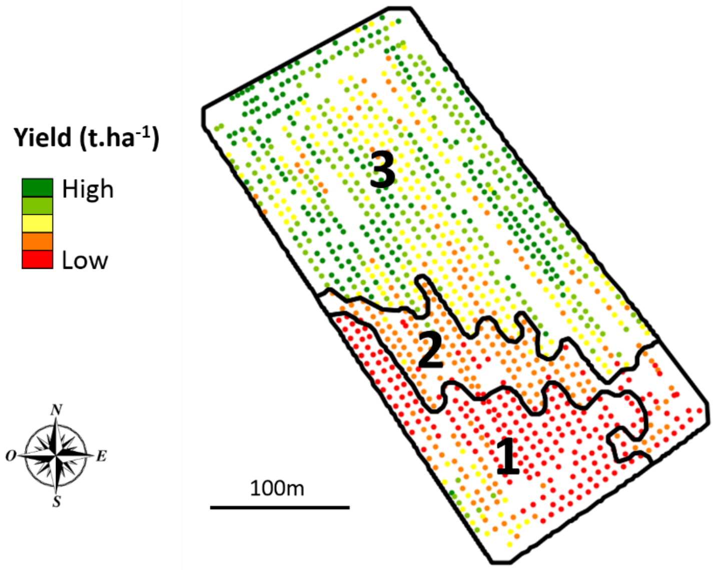

Figure 1. Within-field yield dataset under study with the associated initial zoning. *Yield data are separated into five quantiles.*

An iterative region growing algorithm

On Figure 1, within-field initial zones were generated by the region growing and merging algorithm proposed in Leroux et al., (2017). In simple terms, a set of observations, referred to as the seeds, was selected in homogeneous regions within the field as the core of within-field zones. Observations were aggregated to these seeds following a similarity measure to form the zones (region growing). As the number of resulting regions is generally too high for management purposes, these zones were merged by maximizing an opportunity criterion (region merging). It must be clear that each observation is necessarily associated to one and only one zone. Interested readers are referred to Leroux et al. (2017) for more information.

To generate fuzzy maps, it is proposed here to run the region growing algorithm of Leroux et al. (2017) multiple times with a fixed number of seeds, i.e., the number of zones in the initial zoning (three seeds in this case study). At each iteration, the region growing algorithm is run with a different set of seeds (each one being selected in one of the initial zones), so that in the end, all the zonings are somehow different but the number of zones are necessary the same given that the number of seeds is fixed. At the end of all the iterations, each observation can be given a membership frequency to each one of the initial zones. It can be considered that if observation x_i always belongs to zone Z_j across iterations, then x_i effectively belongs to the core (or kernel) of Z_j. On the contrary, if x_i can belong to multiple zones, then x_i might be part of a transition area (or support of a zone). A step by step description of this approach is presented below:

Algorithm 1: Membership map generation

1. Select randomly one observation within each one of the N initial zones. This set of observations is referred to as the set of seeds S $(S_1, S_2, \ldots, S_N)$
2. Run the region growing algorithm of Leroux et al. (2017) with the previously selected set of seeds.
3. Repeat n times steps 1 to 2
4. For each observation x_i, compute the membership frequency $\mu_{i,j}$ to each zone Z_j as the ratio of the number of times x_i belongs to Z_j to the number of iterations n. It must

be noted that $\mu_{i,j}$ lies between 0 (observation x_i never belongs to Z_j) and 1 (observation x_i always belongs to Z_j)

5. Generate the membership frequency map

In this study, the number of iterations n was set to 30. Be aware that, by construction, as each observation is necessarily associated to one and only one zone at each iteration, $\sum_{j=1}^{N} \mu_{i,j} = 1$.

Adapting the initial zoning to users' expertise and management decisions

The membership map that can be generated with the previously defined algorithm is interesting in itself as it can be used to visualise to what extent an observation belongs to a zone rather than another. It is stressed here that the membership frequency of each observation to the initial zones could be used to refine the initial zoning for specific management decisions. Two theoretical fertilization-based use-cases (UC) are considered:

- UC1: The field is subjected to strong environmental constraints and the farmer cannot accept any risk of overdosing. Zones should therefore be limited to their core or kernel
- UC2: The farmer is using a low-cost fertilizer for which the risks on environment or the impact on yield are limited. Zones could therefore be extended to their support to a certain level.

To provide a zoning that copes with these two different management strategies and users' expertise, the following approach is proposed.

Algorithm 2: Zoning adaptation

1. Choose the zone $Z_{j'}$ for which the border should be refined
2. Choose a membership threshold α above which it is considered that the observations definitely belong to $Z_{j'}$.
3. Update the membership frequency of each observation as follows:

$$if\ \mu_{i,j'} > \alpha\ then \begin{cases} \mu_{i,j'} = 1 \\ \mu_{i,j(j\neq j')} = 0 \\ \mu_{i',j'(i\neq i')} = 0 \\ Rescale\ \mu_{i',j}\ so\ that \sum_{j=1}^{N} \mu_{i',j}\ remains\ equal\ to\ 1 \end{cases}$$

4. For each observation x_i, compute a weighted yield value $Y_i^w = \sum_{j=1}^{N} \mu_{i,j} \times \overline{Y_{Z_j}}$ where $\overline{Y_{Z_j}}$ is the average yield of the initial zone Z_j
5. Select one seed per zone as the one with the maximum membership value
6. Run the region growing algorithm of Leroux et al. (2017) using the weighted yield attribute and the previously selected seeds.

Results and discussion

Membership maps and fuzzy zoning

Figure 2 shows the membership frequency of each observation for the three zones of interest. It appears clearly that the spatial extent of membership frequencies for zone n°2 is much larger than for the two other zones, meaning that zone n°2 could be seen as a transition zone between zones n°1 and 3. The core of zone n°3 seems to be located on the western, northern, and eastern parts of the field as the southern region of zone n°3 (centre of the field) has a colour ranging from purple to blue. It can be seen that this is the region where observations have a lower yield value compared to the rest of zone n°3 (Fig. 1). Figure 2 highlights that observations near the border of each zone have a significative lower membership frequency than the remaining observations. Interestingly, observations within the region on the bottom right hand corner of the field, belonging to the initial zone n°2, have a similar membership frequency for zones n°1 and 2 (Fig. 2). It can be seen that the membership frequency map is also interesting to evaluate the surface of the transition zones.

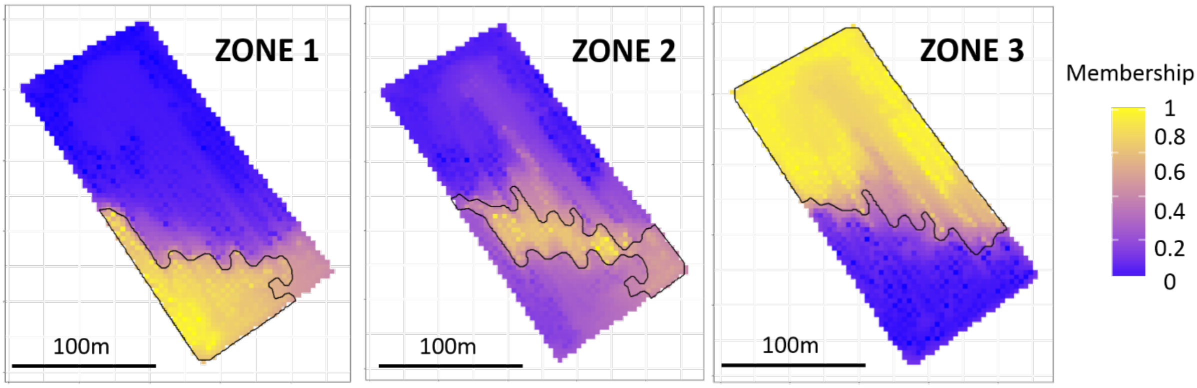

Figure 2. Membership frequency map to each zone of interest. *For visualisation purposes, membership values were interpolated via inverse distance weighing with a power distance of 2 on a 2x2m grid. Interpolated values are solely used for mapping. All the algorithms make use of raw (non-interpolated) data.*

Figure 3 shows how the zoning could be refined for zones n°1 and 3 depending on the management strategies defined by users. As the threshold α increases, zones n°1 (top) and 3 (bottom) shrink as transient observations are removed from the zones. These new zonings would be more adapted to the first use-case (UC1) previously defined where environmental constraints are strong and the farmer must not take any risk with respect to the fertilization of regions out of zones n°1 or 3 (Fig. 3, right) On the contrary, as the threshold α decreases, zones n°1 (top) and 3 (bottom) expand as transient observations are aggregated to the zones, i.e., the zone's support is considered to be part of the zone (Fig. 3, left). In this case, the zonings would be more suitable to UC2 as the fertilization-based risk is much lower. Figure 3 demonstrates that the expansion and shrinking of the zones is not simply a buffer applied on the borders as the new zones follow the yield values across the field, e.g., when α is set to 0.7 for zone n°3.

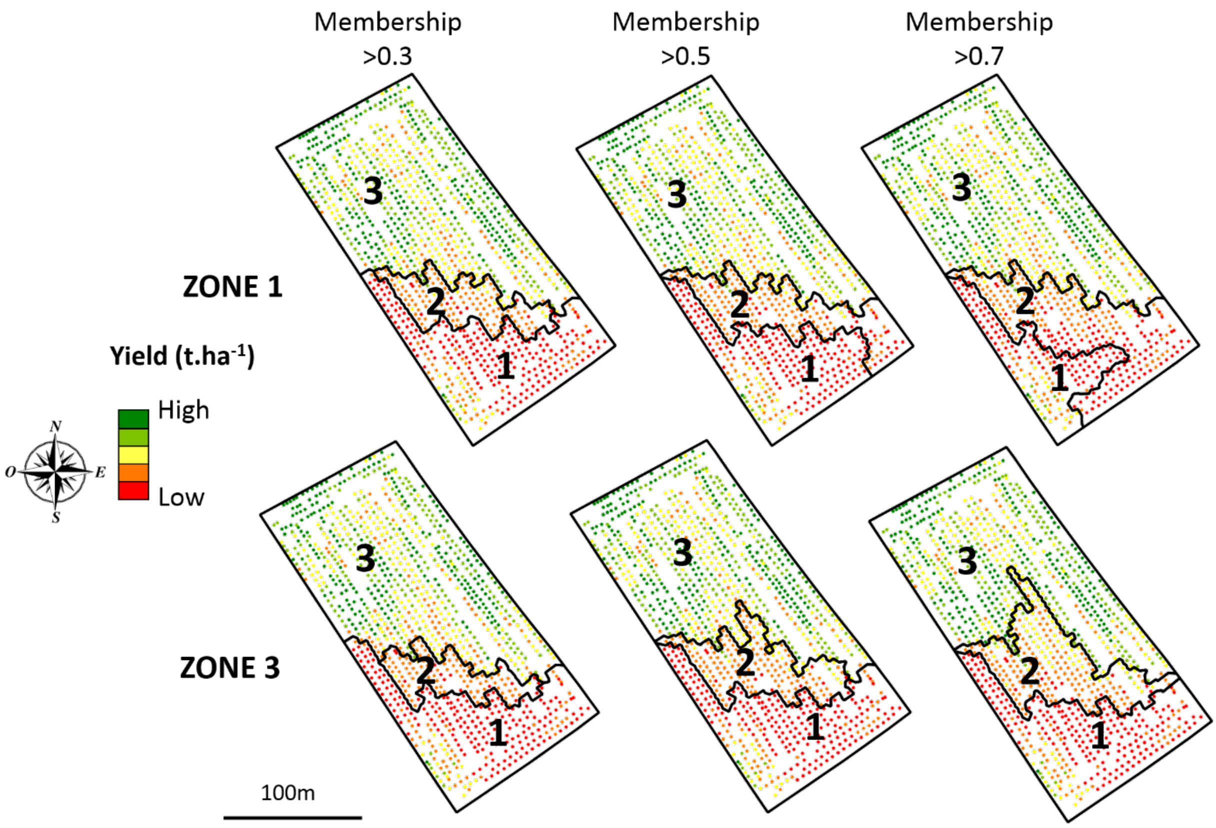

Figure 3. Adaptation of the zoning to specific management decisions. *Zones are either expanded or shrunk depending on the membership threshold that is chosen. Plots on the top represent zoning with respect to zone 1. Plots on the bottom represent zoning with respect to zone 3.*

Further considerations

Readers must be aware that the initial zoning has a great importance in the proposed methodology. In fact, this initial zoning conditions the number of zones and the possible locations where the seeds can be selected in each zone. Different initial zonings would necessarily lead to different fuzzy zonings. It must however be clear that the initial zoning, derived from the approach of Leroux et al. (2017), was motivated by operational constraints and was considered reliable here. The high-yielding region in the south-west portion of the field was effectively considered too small to be delineated.

It must be said that the concept of fuzzy zones has been tackled in this study with a frequentist or probabilistic approach. Indeed, the membership of each observation x_i to a zone Z_j was defined as the number of times x_i belonged to Z_j. This frequency membership is different from the more common degree membership that can be obtained with specific fuzzy algorithm such as the fuzzy k-means. Further work should be carried out to evaluate the similarities or differences between those two information.

The membership frequency maps presented in this study were obtained after the application of an iterative region growing algorithm, which used seeds randomly selected within the initial zones Z_j. It is clear that the choice of the algorithm to generate the zones and the way seeds are selected within the zones will have an impact on the membership frequency of each observation to the zones. Other seed sampling processes might be considered. For instance, the seed selection could be constrained to regions near the zones' border. This would certainly allow zones to expand further (zones would have a larger support). Another possibility could be to select seeds with a yield value near the centre of the yield distribution of each zone to ensure that noisy observations are not selected.

For now, no rules have been set up for the selection of the membership threshold α. A set of maps with different thresholds was shown to evaluate how the zoning could be adapted to specific management decisions. The threshold should be chosen in accordance with the user's expertise on the field and with respect to the management decisions to be made. There might be cases where some zones have multiple cores, i.e., transition areas are not solely at the border between zones but also within the zones. This might be problematical to adapt the zoning as there would be a need to account for these multiple cores. These multiple cores would however be seen relatively easily on the membership frequency map. If such phenomenon happens, there might be a need to work on the initial zoning to make sure that the zone with multiple cores is separated into multiples zones.

For a given membership threshold α, the proposed approach generates one map per zone of interest. As the number of zone increases, the number of maps will consequently increases, which will make the reading and interpretation of the zoning more difficult by end users. Further work will have to investigate on how to combine these maps to facilitate the decision-making process.

Finally, fuzzy zonings would also benefit from quality criteria that would help assess how well the zones were treated. Even zones are considered homogeneous, it is clear that the natural intrinsic variability of plants will lead to some treatment errors that would be interesting to know about. It must be noted that this criterion will help to evaluate the application error in the small region in the bottom left hand corner of zone n°1, which was considered too small to be delineated in the initial zoning.

Conclusion

This work has made use of the fuzzy set theory to improve the delineation of management zones. Maps of membership frequency were built with an iterative region growing algorithm to evaluate to what extent observations belonged to specific within-field zones. These membership frequencies were then used to refine the borders of management zones so as to cope with theoretical fertilizer-based case studies where environmental constraints were either strong or weak. So far, the membership threshold above which observations were considered to be part of the core of the zones was not defined automatically nor by expertise. Maps with different threshold's settings were simply displayed to assess how the zoning could be possibly adapted. Future work will intend to combine the multiple fuzzy zonings that were generated to facilitate the decision-making process.

References

Jones, H., Guillaume, S., Loisel, P., Charnomordic, B., & Tisseyre, B. (2016). Generation of Plateau-Approximated Fuzzy Zones. *In Proceedings of the Conference on Spatial Accuracy*, Montpellier, France.

Leroux, C., Jones, H., Clenet, A., & Tisseyre, B. (2017). A new approach for zoning irregularly-spaced, within-field data. *Computers and Electronics in Agriculture*, 141, 196-206.

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

Paoli, J.N., Strauss, O., Tisseyre, B., Roger, J.M., & Guillaume, S. (2007). Spatial data fusion for qualitative estimation of fuzzy request zones: Application on precision viticulture. *Fuzzy Sets and Systems*, 158, 535-554.

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.

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

Tagarakis, A., Liakos, V., Fountas, S., Koundouras, S., & Gemtos, T.A. (2013). Management zones delineation using fuzzy clustering techniques in vines. *Precision Agriculture*, 14, 18-39

Urretavizcaya, I., Santesteban, L. G., Tisseyre, B., Guillaume, S., Miranda, C., and Royo, & J. B. (2013). Oenological significance of vineyard management zones delineated using early grape sampling, Precision Agriculture, 9, 101–103

Zadeh, L. (1964) Fuzzy sets. *Information and Control*, 8, 338–353.