

# Exploratory Hierarchical Clustering for Management Zone Delineation in Precision Agriculture

Georg Ruß and Rudolf Kruse

Otto-von-Guericke-Universität Magdeburg, Germany  
{russ,kruse}@iws.cs.uni-magdeburg.de

**Abstract.** Precision Agriculture has become an emerging topic over the last ten years. It is concerned with the integration of information technology into agricultural processes. This is especially true for the ongoing and growing data collection in agriculture. Novel ground-based sensors, aerial and satellite imagery as well as soil sampling provide large georeferenced data sets with high spatial resolution. However, these data lead to the data mining problem of finding novel and useful information in these data sets.

One of the key tasks in the area of precision agriculture is *management zone delineation*: given a data set of georeferenced data records with high spatial resolution, we would like to discover spatially mostly contiguous zones on the field which exhibit similar characteristics within the zones and different characteristics between zones. From a data mining point of view, this task comes down to a variant of spatial clustering with a constraint of keeping the resulting clusters spatially mostly contiguous.

This article presents a novel approach tailored to the specifics of the available data, which do not allow for using an existing algorithm. A variant of hierarchical agglomerative clustering will be presented, in conjunction with a spatial constraint. Results on available multi-variate data sets and subsets will be presented.

## 1 Introduction

In recent years, the agriculture domain has seen a vast amount of information technology being introduced. On the one hand, this is due to technological advances, such as cheaper GPS technology, novel remote sensing equipment and improved satellite and aerial imaging technology. On the other hand, there is also an economical advantage becoming more and more emergent. Based on the above technology, using the acquired data, farmers can optimize their fertilizer and pesticide applications (among other controls) to achieve an optimized outcome in terms of yield and/or economic profits. However, with the large-scale collection of georeferenced and high-resolution data sets, agriculture has turned into a data-driven discipline. Therefore, the aforementioned optimization task requires sophisticated data mining techniques tailored to the specifics of the data sets.

The available data sets which are nowadays collected are usually spatially dense, up to one data record per  $10 \times 10$ -metres square. Some of the imagery sensors yield even higher resolutions. Given the fact that measurements of different soil and vegetation properties as well as controllable inputs such as fertilizer can be acquired at more than just one point in time into the growing season, the data sets grow quickly, both on the temporal as well as the spatial scale. Naturally, each data record in these sets is georeferenced, i.e. has a specific spatial position on the field and has fixed neighbors. This, in turn, leads to neighboring data records being *not* independent of each other, a phenomenon known as *spatial autocorrelation*. Furthermore, from the physical and biological perspective there are certain assumptions as to which factors influence plant growth and yield, for example. These may now be validated using such data sets.

A task commonly occurring in agriculture is the so-called *management zone delineation*. Based on the biologically valid assumption that certain soil minerals are necessary for healthy plant growth, these minerals must be made available to the plants. Often, these minerals exist in sufficient quantities in the soil, but are not in a chemical state which allows the plants to easily tap into the mineral reservoirs. Furthermore, they may not be available at all. Therefore, basic fertilization is applied, which aims to make the minerals available. However, since the fields are usually heterogeneous, different parts of the field may require different amounts of basic fertilization. Determining these so-called management zones is therefore an important task.

In terms of data mining and knowledge discovery, the above task may be recognized as a variant of spatial clustering. The data sets consist of geo-referenced data records which have a number of attributes attached to them. Given these data sets, we aim to find, in an exploratory way, spatial clusters which exhibit the cluster property: the data records within a cluster are similar, while the similarity between clusters is low. As of now, there are only few approaches towards this problem given the type of data sets occurring in precision agriculture. We will present those approaches, outline their main issues and will develop a rather simple and straightforward approach to solving the problem of management zone delineation by adapting a constraint-based clustering algorithm.

## 1.1 Article Structure

First, this article will give an overview about the existing literature on spatial clustering. Since the existing algorithms are usually closely coupled with the available data sets, the type of data sets this article is concerned with is presented at the very beginning of the following Section 2. We present our two-stage, divide-and-conquer hierarchical spatial clustering approach in Section 3. We present the results on the data set as well as the limitations and parameter settings of our approach in Section 4. This article finishes with a conclusion and further discussion of the results.

## 2 Data Set and Existing Literature on Spatial Clustering

Clustering algorithms are usually closely interwoven with the data they are being applied to. Therefore, a data description which outlines the key characteristics of the precision agriculture data encountered here will be presented first. Based on these data, the shortcomings of the existing algorithms will be pointed out. There are also a few agricultural approaches to solving the problem which will be briefly outlined.

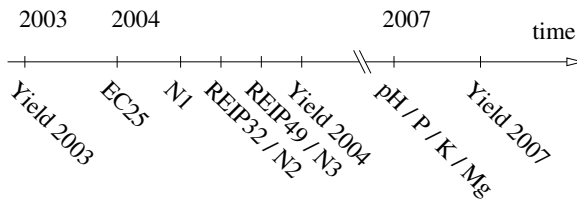
### 2.1 Precision Agriculture Data Description

The data available in this work were obtained on a precision agriculture experimental site in Northern Germany in 2004 with additional attributes from 2003 and 2007. The data are spatially distributed in regular hexagonal grid cells 25 metres in diameter, such that 16 of these grid cells represent one hectare. Overall, the data set consists of 1,080 data records, which (at the above resolution) represent a field 67.5 hectares in size.

Each data record consists of a number of attributes, which are recorded as mentioned above, using special sensors. Further details on the data are provided in [24]. The attributes are sorted chronologically and shown in the timeline in Figure 1. The temporal aspects are not considered.

Due to the origins of the data, spatial autocorrelation exists in the natural attributes of the data [21], but is less pronounced in the human-controllable fertilizer attributes. It can be easily seen from the plots in Figures 2(a) to 2(d) that spatially adjacent values for the depicted attributes are likely to be much more similar the closer they are. In a later stage of this work (Section 3.1), spatial autocorrelation will be exploited, it is therefore necessary to gain a basic understanding of this concept here. It can be shown that modeling techniques which neglect the spatial information in the data sets produce misleading results [20], which is clear from a statistical point of view [3].

It is furthermore application-dependent which of the data attributes are actually required. For elementary applications such as yield prediction, research in this area is ongoing [23], mostly in the direction of feature selection approaches, though observing the spatial nature of the data requires some additional effort [22]. Therefore, not all of the attributes are to be used in a specific task at once, but rather a smaller selection based on user experience and expert guidance.



**Fig. 1.** Timeline for data set (not to scale)

## 2.2 Review of Existing Spatial Clustering Algorithms

Given the data set presented in the preceding section, the task is to establish an algorithm which is able to delineate the field into spatially (mostly) contiguous clusters, so-called management zones. Personal experience shows that management zone delineation in practice usually relies on one attribute only because of the lack of appropriate algorithms which are able to deal with multiple attributes.

From a data mining point of view, the task is the following: given a set of data records consisting of a spatial location and a certain number of attached attributes, find a spatial tessellation of these data records such that the resulting zones can readily be used for basic fertilization. Since it is as of now unclear which of the available attributes contribute to the physical and biological underpinnings of management zones [12], the above broad task should be narrowed to the following: develop an algorithm for the above type of data sets which returns a spatially (mostly) contiguous tessellation and which can be easily parameterized by a human expert.

In precision agriculture, there are a number of approaches using standard clustering algorithms such as fuzzy c-means clustering [13,14,16]. However, these rely solely on the data records' attributes and totally neglect the spatial structure of the data records. This results in zones which are non-contiguous and spread over the whole field, as well as small islands of outliers and insignificant records which must be smoothed out manually after the clustering. A similar approach is undertaken by fuzzy classification of the data records, which exhibits the same problems [15]. In addition, there is no clear guidance available as to which input attributes enable a successful management zone delineation [4,18]. It seems, however, clear that management zones must rely on more than just yield data [12]. Based on our experience with using non-spatial models on spatial data sets, it is clear that the spatial component must not be neglected. Furthermore, as [7] points out, the farmers' long-time experience produces good results – therefore, this experience should be captured in an exploratory data mining approach.

In the area of computer science, there are, to the best of the authors' knowledge, no standard clustering algorithms which would allow tackling the above task on the given type of data sets. Density-based algorithms like DBSCAN [5], CLIQUE [1] or STING [28] usually rely on a non-uniform distribution of the data records (density differences) to find clusters. With our data sets, the records are spatially uniformly distributed, which renders the aforementioned algorithms useless. Algorithms like SKATER [2] and REDCAP [9] are different in that they explicitly incorporate spatial contiguity constraints into the clustering process. However, these algorithms may fail to report adjacent clusters correctly (SKATER) or are too strict in terms of management zone contiguity (REDCAP). In addition, they both rely on the fact that data records are spatially non-uniformly distributed, which is not the case here. This last assumption is also used by ICEAGE [10], which is therefore not applicable either. CLARANS [17] is a further algorithm designed for clustering spatial data sets but is based on the assumption that the structure to be discovered is hidden exclusively in the

spatial part of the data, which is not the case here. Finally, AMOEBA [6] works on two-dimensional spatial data by building a hierarchy of spatial clusters using a Delaunay triangulation, but lacks the extension to non-spatial attributes and also assumes that the 2D points are non-uniformly distributed in space.

One of the more common approaches to spatial clustering is a hierarchical agglomerative one: start with each point in a single cluster and subsequently merge clusters according to some criterion or constraint. Further research into constraints-based clustering [26] reveals that it may in principle be applied here. The author of [26] explicitly describes the “spatial contiguity” constraint for spatial data as a type of global clustering constraint using neighborhood information, albeit for image segmentation. The constraints are presented as “hard” or “soft”, meaning that the final clustering outcome “must” or “can” consider these constraints. The task encountered in this article, namely generating *mostly contiguous* clusters, could therefore be tackled by using a soft spatial contiguity constraint. An additional feature of constrained clustering algorithms is the existence of “must-link” and “cannot-link” pairwise constraints for data records. Although an algorithm can usually be constructed this way or the other, it seems more appropriate to model the spatial contiguity requirement as a “cannot-link” (soft) constraint for spatially non-adjacent data records or clusters. In addition, the work of [27] encounters a similar agricultural problem to the one in this article, but the focus is slightly shifted to yield prediction on a county scale with low-resolution data, rather than using high-resolution data for management zone delineation. Since the focus in this work is more on exploratory data mining in an unsupervised setup we postpone the performance question.

Additionally, hierarchical agglomerative clustering seems like a rather natural approach since the solution ultimately has to be presented to domain experts who typically prefer easy-to-understand solutions over black-box models. Therefore, our focus will be on developing a hierarchical agglomerative algorithm for zone delineation which takes the special properties of the data sets into account. Our data sets are different from the ones in existing work since the data records are located on a uniformly spaced hexagonal grid and exhibit spatial autocorrelation. This autocorrelation will be used explicitly in our approach.

### 3 Hierarchical Clustering with Spatial Constraints

This section will present an extended and refined version of the hierarchical, divide-and-conquer approach to delineating spatially mostly contiguous management zones based on precision agriculture data presented in [24,25]. Our approach can best be described as *hierarchical agglomerative clustering with a spatial contiguity constraint* and an additional (optional) initialization step which exploits the spatial autocorrelation in the data. It consists of two phases, in a divide-and-conquer manner. First, the field is tessellated into a fixed number of (spatial) clusters. Second, these clusters are merged iteratively, using a similarity measure and adhering to a spatial contiguity constraint, which shifts from being a hard constraint to a soft constraint throughout the algorithm.

### 3.1 Phase 1: Spatial Field Tessellation via k-Means (Optional)

A hierarchical agglomerative clustering starts at small clusters or single objects and consecutively merges those according to some criteria. The question whether a naïve tessellation of the field into  $N$  clusters is sufficient, where  $N$  is the number of data records, i.e. each data records occupies its own cluster. Certainly, this assumption would hold true, but due to spatial autocorrelation, spatially neighboring data records are likely to be very similar in their attributes. Therefore, by tessellating the field into a fixed number of spatial clusters  $n \leq N$ , the clusters are still very likely to contain similar (adjacent) data records while some of the ensuing computational effort of the merging step can be saved. Furthermore, the merging step requires a list of spatial neighbors for each cluster – if this can be easily computed in the tessellation step, it saves further computation time. With the above prerequisites, the simplest tessellation approach fulfilling the requirements is to perform a  $k$ -means clustering on the data records' spatial coordinates. This creates a basic tessellation, while explicitly assuming that, due to spatial autocorrelation, the resulting spatial clusters contain similar data records. Furthermore, the  $k$ -means tessellation returns a voronoi diagram of the data records' coordinates, of which the dual representation is the Delaunay triangulation. This allows for easy computation of the list of neighbors for each cluster [8]. This phase may be omitted, such that the second phase starts with each point in a single cluster.

### 3.2 Phase 2: Merging Clusters

Once the small contiguous clusters have been created in phase 1, the task is to merge these clusters consecutively into larger clusters, similar to classical agglomerative hierarchical clustering. However, in addition to the standard similarity or distance measure, a spatial constraint must be taken into account. Since the final result of the clustering is assumed to be a set of spatially mostly contiguous clusters, only those clusters should be merged which are a) similar (with regard to their attributes' values) and b) spatial neighbors (adjacent).

In classical hierarchical clustering, the standard measures for cluster similarity are single linkage, complete linkage and average linkage [11]. However, when considering the spatial data encountered here, these three criteria merit some explanation. *Single linkage* determines cluster similarity based on the smallest distance between objects from adjacent clusters. Due to spatial autocorrelation, it is likely that there are always some points at the borders of the clusters which are very similar, for each neighbor. Therefore, single linkage will not provide us with a good measure for which neighbor to choose. *Complete linkage* determines the similarity of neighboring clusters based on the distance of those objects which are farthest away from each other. Since we are considering spatially adjacent clusters, this would lead to very dissimilar clusters being merged. Due to spatial autocorrelation, these objects would also be spatially rather far away from each other, which leads to a chaining effect and less meaningful clusters. *Average linkage* determines the similarity of adjacent clusters based on the average of the

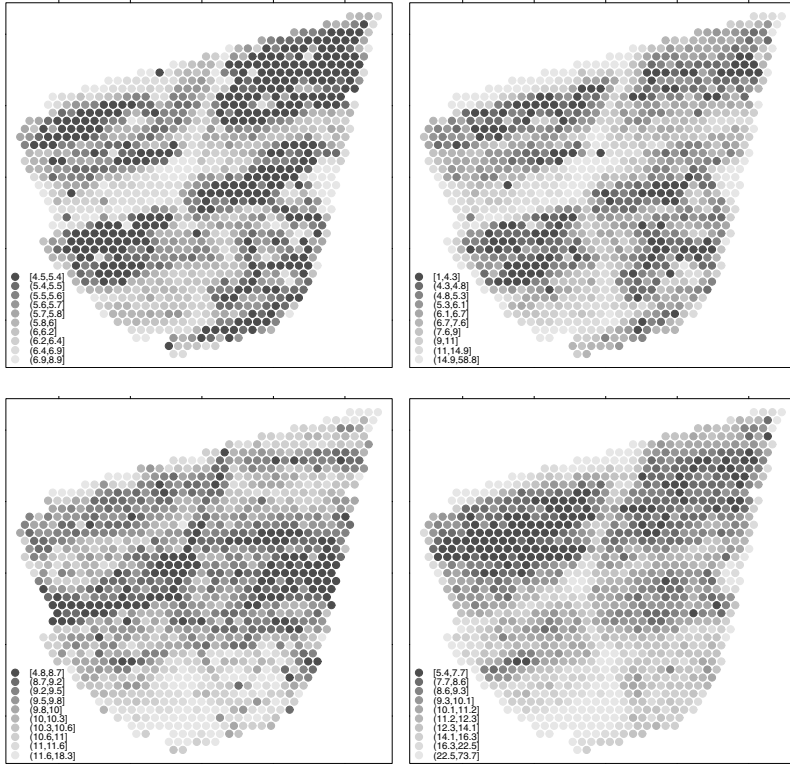
(Euclidean or other) distances between all objects in the clusters. A combination of the aforementioned arguments for single and complete linkage may be applied here: points in adjacent clusters which are spatially close/far apart are likely to also be very similar/dissimilar. Therefore, an appropriate distance for adjacent clusters may be determined by *average group linkage*: we compute an average vector for each cluster and determine the distance between these vectors.

It is not required that one zone is strictly contiguous, i.e. consists of just one spatially contiguous area on the field. It is a valid result if one zone comprises those data records which are similar but is made up of two or more larger areas on the field. This would still be considered immensely useful in practice. Since the focus of this clustering approach is on exploratory data mining rather than providing a fixed clustering, this “*mostly contiguous*” description should be seen as a soft constraint in the final merging steps. To prevent the algorithm from producing too many scattered zones, we propose to set it as a hard constraint during the beginning of the merging phase. As long as adjacent clusters are similar enough, these are merged. If this is not the case, clusters which are not direct neighbors of each other may be merged if they are similar enough. This also provides us with a user-influencable condition for when to switch from a hard to a soft constraint. Hence, we introduce a *contiguity factor*  $cf$ : we may begin merging non-adjacent clusters once the minimum average-linkage distance for adjacent clusters is  $cf$  times the minimum distance for non-adjacent clusters. In the results for Figure 3, the algorithm performs well with the hard constraint in the beginning and would switch to a soft constraint only after the bottom plot, which has 28 clusters left, with  $cf$  set to 2.

## 4 Experimental Setup and Results

We now aim to demonstrate the algorithm on multi-variate data. In order to show some of the parameter settings and the inner workings, we decide to start with a subset of the original data set: we choose the four soil sampling attributes (pH-value, P, K, Mg content). From the four plots in Figure 2 it can be seen that a certain spatial structure is emergent, with four to six visible areas, separated by another cross-shaped area in the middle. This structure is the one we would like our algorithm to discover.

The data set has 1080 spatial data records. As mentioned in the algorithm description, a hierarchical agglomerative clustering procedure may start with each of the data records forming one cluster. However, due to spatial autocorrelation, spatially neighboring data records are likely to be similar and are therefore grouped by using a  $k$ -means clustering on the spatial part only. This is depicted in the top left figure of Figure 3: we choose  $k$  to be 350, such that on average three neighboring data records are in one cluster initially. The algorithm then proceeds to consecutively merge adjacent, similar clusters. This is depicted in Figure 3, top right and bottom left plot, with 250 and 150 clusters left, respectively. The final plot in Figure 3 shows the outcome with 28 clusters left. We can roughly see six zones. Those at the borders are, of course, not (yet) zones in



**Fig. 2.** Four chosen attributes for which the management zone delineation is applied: pH value, P, Mg, K concentration (top to bottom)

the sense of the algorithm, but they are easily visually distinguishable. For an exploratory data mining task, this result is what the algorithm is supposed to deliver.

Upon further examination of the resulting six zones, it turns out that these are actually just three zones. This requires but a quick look at the four figures in Figure 2: the largest zone which covers roughly 80% of the field could be described with *low pH, low P, medium/low Mg, low K*. The border zones on the top left, the left and the bottom left of the field can be described with *high pH, high P, high Mg, high K*. The small zone at the right field border and the one extending from the left border mostly horizontally into the middle would be *high pH, high P, low Mg, high K*. For practical purposes of basic fertilization this simple characterization of a field’s principal zones is very convenient.

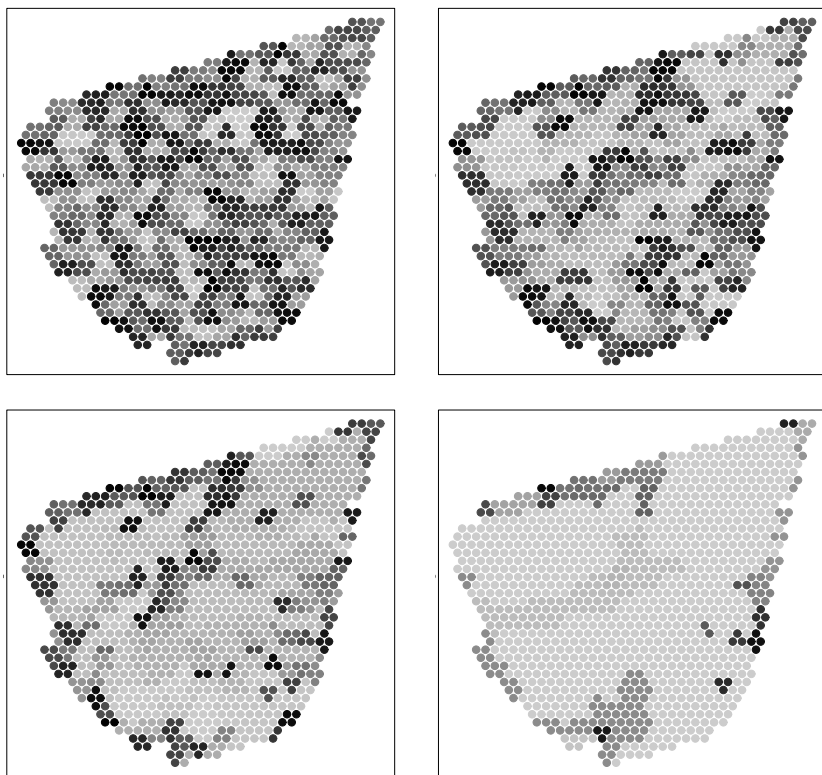
#### 4.1 Limitations and Parameter Guidelines

One of the limitations (and, at the same time, a strength) of our algorithm is the assumption that the data records are spatially autocorrelated. Since this

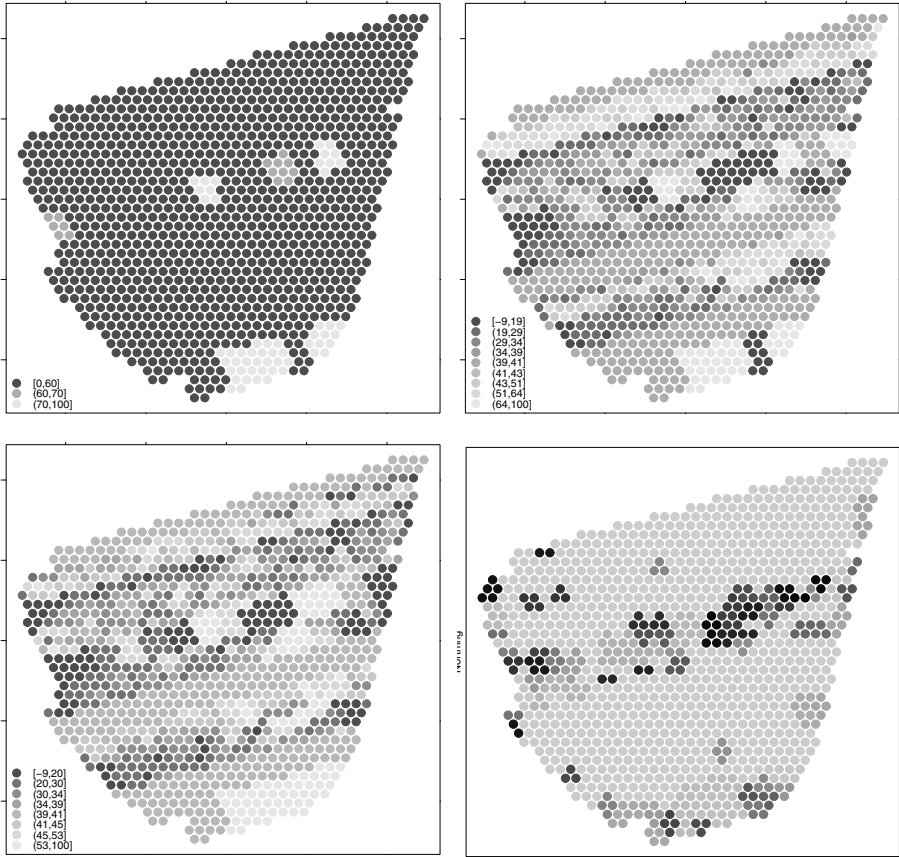


assumption has been built explicitly into the  $k$ -means tessellation step of the algorithm, a violation would lead to invalid results. This is depicted in Figure 4. Among the data set's attributes we have a few variables which can be human-controlled, namely the fertilizer applications N1, N2, N3. Since the field on which the data set has been collected also serves as a test site for fertilization strategies, the fertilizer data are not spatially autocorrelated – there are strips where different strategies were carried out and N1 was more or less uniformly applied. What happens when these data are used in our clustering algorithm can be seen in the bottom figure of Figure 4: the resulting zones are not meaningful. This is due to the first step of our algorithm, which assumes that the data are spatially autocorrelated. Therefore, data which violate this assumption must not be used with the algorithm or the first phase of the algorithm should be skipped.

Setting the parameter  $k$  for the  $k$ -means tessellation depends on the data set. For rather homogeneous fields, this can be set to a lower value such as  $\frac{N}{10}$ , where  $N$  is the number of available data records. For rather heterogeneous data sets such as the one encountered here, we may set it to as low as  $\frac{N}{3}$ , thereby combining roughly three adjacent data records into one initial cluster. If the



**Fig. 3.** Clustering on the attributes shown in Figure 2, beginning of clustering (350 clusters), after 100/200 merging steps, with 28 clusters left (left to right, top to bottom)



**Fig. 4.** The three farmer-manageable variables of nitrogen fertilizer (N1, N2, N3), along with a failed clustering approach of our algorithm (top to bottom). The data violate the algorithm’s *spatial autocorrelation* assumption.  $cf$  is set to 2 here, but has close to no influence.  $k$  is set to 350 initial clusters.

number  $k$  of initial clusters is set to  $N$ , we obtain a setting which may be used for data where no spatial autocorrelation exists.

Setting the contiguity factor  $cf$  is rather straightforward: a value much higher than 1 leads to a later switch from a hard to a soft constraint – therefore, the spatial contiguity is higher. A value larger than, but closer to 1 further weakens this hard constraint. A value smaller than 1 favors the merging of non-adjacent clusters early in the algorithm, probably resulting in rather scattered zones. The *average-linkage* similarity computation using Euclidean distance may be replaced by a different distance measure. For higher numbers of attributes, the Cosine distance measure may be employed.

## 5 Conclusion and Discussion

This article presented a hierarchical agglomerative clustering approach with a spatial constraint for the task of management zone delineation in precision agriculture. Based on the specifics of the data sets from precision agriculture, namely the uniform spatial distribution of the data records on a hexagonal grid and the existence of spatial autocorrelation, we established and recognized the shortcomings (or the lack) of existing approaches. Henceforth, we specified the requirements of a novel approach: the spatial contiguity of the resulting zones and the explicit assumption of spatial autocorrelation.

This research led to a two-phase divide-and-conquer approach. In the first phase we tessellated the field using  $k$ -means on the data records' 2D coordinates. In the second phase, we iteratively merged those spatially adjacent clusters that are similar. This was done in two sub-phases: in the first sub-phase, the spatial contiguity was a hard constraint, meaning that only adjacent clusters may be merged. In the second sub-phase, this was relaxed to a soft constraint. Switching from the hard to the soft constraint can be user-influenced by a contiguity factor  $cf$ . Proceeding like this provided us with a hierarchical structure which can then be examined by a human expert for guidance on the management zone delineation. Our focus was on providing an exploratory and easy-to-understand approach rather than a fixed, black-box solution. Our approach worked successfully for spatially autocorrelated precision agriculture data sets. The parameter setting for  $k$  (initial tessellation) was explained. An additional parameter  $cf$  was suggested for further analysis on the spatial contiguity of the resulting clusters. The algorithm was shown to return erroneous results when the assumption of spatial autocorrelation is violated.

### 5.1 Future Work

Once the clustering algorithm finishes, a certain clustering should usually be examined further. This is likely to be towards the end of the merging stage, when a human-manageable number of around ten clusters is left. These clusters may easily be examined using frequent itemset mining. Numerical attributes can be converted to a three- or five-value categorical scale and the resulting frequent sets could be generated as we did manually for the bottom plot of Figure 3. Although the *average linkage* similarity calculation turns out to work rather well in practice, it may be further researched whether different linkage criteria in combination with other similarity measures could be more appropriate. A drawback of our work is the lack of reference data sets from precision agriculture and similar domains in conjunction with a similar task. We are currently investigating the possibility of making our data sets publicly available for this purpose.

**Acknowledgements.** The implementation is carried out in R [19]. The R scripts are available on request from the first author of this article. The data in this work have been acquired on the experimental farm Görzig in the federal state Sachsen-Anhalt, in Germany. The data were obtained from Martin

Schneider and Peter Wagner from Martin-Luther-Universität Halle-Wittenberg, Germany, Lehrstuhl für landwirtschaftliche Betriebslehre.

## References

1. Agrawal, R., Gehrke, J., Gunopulos, D., Raghavan, P.: Automatic subspace clustering of high dimensional data for data mining applications. In: SIGMOD 1998: Proc. of the 1998 ACM SIGMOD Int. Conf. on Management of Data, pp. 94–105. ACM, New York (1998)
2. Assuncao, R.M., Neves, M.C., Camara, G., Da Costa Freitas, C.: Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science* 20(7), 797–811 (2006)
3. Cressie, N.A.C.: *Statistics for Spatial Data*. Wiley, New York (1993)
4. Domsch, H., Heisig, M., Witzke, K.: Estimation of yield zones using aerial images and yield data from a few tracks of a combine harvester. *Precision Agriculture* 9, 321–337 (2008)
5. Ester, M., Kriegel, H.-P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Simoudis, E., Han, J., Fayyad, U.M. (eds.) 2nd Int. Conf. on Knowledge Discovery and Data Mining, pp. 226–231. AAAI Press, Menlo Park (1996)
6. Estivill-Castro, V., Lee, I.: Multi-level clustering and its visualization for exploratory spatial analysis. *GeoInformatica* 6(2), 123–152 (2002)
7. Fleming, K.L., Westfall, D.G., Wiens, D.W., Brodahl, M.C.: Evaluating farmer defined management zone maps for variable rate fertilizer application. *Precision Agriculture* 2, 201–215 (2000)
8. Gold, C.M., Remmele, P.R.: Voronoi methods in GIS. In: van Kreveld, M., Roos, T., Nievergelt, J., Widmayer, P. (eds.) CISM School 1996. LNCS, vol. 1340, pp. 21–35. Springer, Heidelberg (1997)
9. Guo, D.: Regionalization with dynamically constrained agglomerative clustering and partitioning (redcap). *International Journal of Geographical Information Science* 22(7), 801–823 (2008)
10. Guo, D., Peuquet, D.J., Gahegan, M.: ICEAGE: Interactive clustering and exploration of large and high-dimensional geodata. *Geoinformatica* 7(3), 229–253 (2003)
11. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. *ACM Computing Survey* 31(3), 264–323 (1999)
12. Khosla, R., Inman, D., Westfall, D.G., Reich, R.M., Frasier, M., Mzuku, M., Koch, B., Hornung, A.: A synthesis of multi-disciplinary research in precision agriculture: site-specific management zones in the semi-arid western Great Plains of the USA. *Precision Agriculture* 9, 85–100 (2008)
13. King, J.A., Dampney, P.M.R., Lark, R.M., Wheeler, H.C., Bradley, R.I., Mayr, T.R.: Mapping potential crop management zones within fields: Use of yield-map series and patterns of soil physical properties identified by electromagnetic induction sensing. *Precision Agriculture* 6, 167–181 (2005)
14. Kitchen, N.R., Sudduth, K.A., Myers, D.B., Drummond, S.T., Hong, S.Y.: Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. *Computers and Electronics in Agriculture* 46(1-3), 285–308 (2005), *Applications of Apparent Soil Electrical Conductivity in Precision Agriculture*

15. Lark, R.M.: Forming spatially coherent regions by classification of multi-variate data: an example from the analysis of maps of crop yield. *International Journal of Geographical Information Science* 12(1), 83–98 (1998)
16. Li, Y., Shi, Z., Li, F., Li, H.-Y.: Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. *Comput. Electron. Agric.* 56(2), 174–186 (2007)
17. Ng, R.T., Han, J.: Clarans: A method for clustering objects for spatial data mining. *IEEE Transactions on Knowledge and Data Engineering* 14(5), 1003–1016 (2002)
18. Ortega, R.A., Santibáñez, O.A.: Determination of management zones in corn (*zea mays* l.) based on soil fertility. *Computers and Electronics in Agriculture* 58(1), 49–59 (2007), *Precision Agriculture in Latin America*
19. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2009) ISBN 3-900051-07-0
20. Ruß, G., Brenning, A.: Data mining in precision agriculture: Management of spatial information. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) *IPMU 2010. LNCS*, vol. 6178, pp. 350–359. Springer, Heidelberg (2010)
21. Ruß, G., Brenning, A.: Spatial variable importance assessment for yield prediction in precision agriculture. In: Cohen, P.R., Adams, N.M., Berthold, M.R. (eds.) *IDA 2010. LNCS*, vol. 6065, pp. 184–195. Springer, Heidelberg (2010)
22. Ruß, G., Kruse, R.: Feature selection for wheat yield prediction. In: Allen, T., Ellis, R., Petridis, M. (eds.) *Research and Development in Intelligent Systems*, vol. 26, pp. 465–478. BCS SGAI, Springer (January 2010)
23. Ruß, G., Kruse, R.: Regression models for spatial data: An example from precision agriculture. In: Perner, P. (ed.) *ICDM 2010. LNCS (LNAI)*, vol. 6171, pp. 450–463. Springer, Heidelberg (2010)
24. Ruß, G., Kruse, R., Schneider, M.: A clustering approach for management zone delineation in precision agriculture. In: Khosla, R. (ed.) *Proceedings of the Int. Conf. on Precision Agriculture 2010* (July 2010)
25. Ruß, G., Schneider, M., Kruse, R.: Hierarchical spatial clustering for management zone delineation in precision agriculture. In: Bichindaritz, I., Perner, P., Ruß, G. (eds.) *Advances in Data Mining*, Leipzig, Germany, pp. 95–104. IBAI Publishing (July 2010)
26. Wagstaff, K.L.: *Intelligent Clustering with Instance-Level Constraints*. PhD thesis, Cornell University (2002)
27. Wagstaff, K.L., Mazzoni, D., Sain, S.: HARVIST: A system for agricultural and weather studies using advanced statistical models. In: *Proceedings of the Earth-Sun Systems Technology Conference* (2005)
28. Wang, W., Yang, J., Muntz, R.: Sting: A statistical information grid approach to spatial data mining. In: *Proceedings of the 23rd VLBD Conference*, Athens, Greece, pp. 186–195. Morgan Kaufmann Publishers Inc., San Francisco (1997)