Visual Data Mining of Agriculture Data

Georg Ruß, Rudolf Kruse[1], Martin Schneider, Peter Wagner[2]

[1] Otto-von-Guericke-Universität Magdeburg
[2] Martin-Luther-Universität Halle

Abstract. Precision agriculture (PA) and information technology (IT) are closely interwoven. The former usually refers to the application of nowadays' technology to agriculture. Due to the use of sensors and GPS technology, in today's agriculture many data are collected. Making use of those data via IT often leads to dramatic improvements in efficiency. For this purpose, the challenge is to change these raw data into useful information. Techniques or methods are required which use those data to their full extent – clearly being a data mining task. This paper presents experimental results on real and recent agriculture data that aid in the first part of the data mining process: understanding and visualizing the data. Self-organizing maps and multidimensional scaling techniques will be used to reduce the high-dimensional input data to two dimensions. The processed data can then be visualized appropriately on 2D maps. An analysis of correlations and interdependencies in the data set will be given, based on the visualization.

Keywords: Precision Agriculture, Data Mining, Multidimensional Scaling, Self-Organizing Maps

1 Introduction

The past decade has seen rapidly advancing information technology which has trickled down into everyday life. Not only have technological breakthroughs been made in industry and services, but also in agriculture.

Due to the adoption of modern GPS technology and the use of ever more different sensor technology on the field, major advances can be made in agriculture. Since the data resulting from the field are small-scale, precise data, this led to creation of the term *precision farming*. According to [22], precision farming is the sampling, mapping, analysis and management of production areas that recognizes the spatial variability of the cropland. It can be seen as a major step from uniform, large-scale cultivation of soil towards small-field, precise planning of, e.g., fertilizer or pesticide usage. With the ever-increasing amount of sensors and information about their soil, farmers are not only harvesting, e.g., potatoes or grain, but also harvesting large amounts of data.

In data mining terms, the area of precision farming (PF) is quite an interesting one as it involves methods and algorithms from numerous areas that the data mining community is familiar with. When analyzing the data flow that results from using PF techniques, it is clear that data mining, on top of a farmer's experience, is one of the main keys to understanding agriculture: a farmer collects data from his cropland (e.g., when fertilizing or harvesting) and would like to extract useful information from those data

and make use of this information to his (economic) advantage. A simplified data flow model can be seen in Figure 1. Therefore, it is clearly necessary to consider using data mining techniques in the light of precision farming to use those data to their full extent.

1.1 Research Target

With this contribution we aim at evaluating suitable methods to visualize agricultural data with a high degree of precision and generality. We present different data sets which shall be visualized. We present experimental results on real and recent (years 2005/2006) agricultural data. Our work helps in visualizing and understanding the available data, which are two important steps in data mining. Eventually, our work aims to proceed towards a framework for evaluating the usefulness of agricultural sensors for the purpose of yield prediction.

1.2 Article Structure

This article concentrates on the step parallel to modeling the data, as seen in Figure 1, namely visualizing the data appropriately. First, we give an overview about the data that we are dealing with in Section 2, including interesting sensor data. Thereafter, we will present selected techniques for advanced data visualization in Section 3. Section 4 is at the core of this article: the different data sets will be visualized and conclusions towards correlations and interdependencies will be drawn. The conclusions will be compared with farmers' experience. Section 5 presents a short summary and lays out our future work.

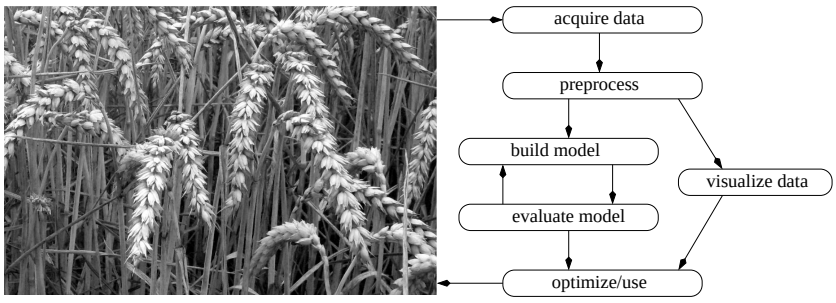


Fig. 1: Data mining for agriculture data

2 Data Description

The data available in this work have been obtained in the year 2006 on a field near Köthen, north of Halle, Germany[3]. All information available for these 72- and 32-

[3] GPS: Latitude N 51 40.430, Longitude E 11 58.110

hectare fields[4] was interpolated using kriging [19] to a grid with 10 by 10 meters grid cell sizes. Each grid cell represents a record with all available information. During the growing season of 2006, the field was subdivided into different strips, where various fertilization strategies were carried out. For an example of various managing strategies, see e.g. [18], which also shows the economic potential of PA technologies quite clearly. The field grew winter wheat, where nitrogen fertilizer was distributed over three application times during the growing season.

Overall, there are seven input attributes – accompanied by the yield in 2006 as the target attribute. Those attributes will be described in the following. In total, for the smaller field (F131) there are 2278 records, for the larger field (F330) there are 4578 records, thereof none with missing values and none with outliers.

2.1 Nitrogen Fertilizer – N1, N2, N3

The amount of fertilizer applied to each subfield can be easily measured. It is applied at three points in time into the vegetation period, which is the standard strategy for most of Northwest Europe [11].

2.2 Vegetation – REIP32, REIP49

The *red edge inflection point* (REIP) is a second derivative value calculated along the red edge region of the spectrum, which is situated from 680 to 750nm. Dedicated REIP sensors are used in-season to measure the plants' reflection in this spectral band. Since the plants' chlorophyll content is assumed to highly correlate with the nitrogen availability (see, e.g. [10]), the REIP value allows for deducing the plants' state of nutrition and thus, the previous crop growth. For further information on certain types of sensors and a more detailed introduction, see [8] or [21]. Plants that have less chlorophyll will show a lower REIP value as the red edge moves toward the blue part of the spectrum. On the other hand, plants with more chlorophyll will have higher REIP values as the red edge moves toward the higher wavelengths. For the range of REIP values encountered in the available data, see Tables 1a and 1b. The numbers in the REIP32 and REIP49 names refer to the growing stage of winter wheat, as defined in [9].

2.3 Electric Conductivity – EM38

A non-invasive method to discover and map a field's heterogeneity is to measure the soil's conductivity. Commercial sensors such as the EM-38[5] are designed for agricultural use and can measure small-scale conductivity to a depth of about 1.5 metres. There is no possibility of interpreting these sensor data directly in terms of its meaningfulness as yield-influencing factor. But in connection with other site-specific data, as explained in the rest of this section, there could be coherences. For the range of EM values encountered in the available data, see Tables 1a and 1b.

[4] We will call them *F330* and *F131*, respectively

[5] trademark of Geonics Ltd, Ontario, Canada

2.4 YIELD 2005/2006

Here, yield is measured in metric tons per hectare ($\frac{t}{ha}$), where one metric ton equals roughly 2204 pounds and one hectare roughly equals 2.47 acres. For the yield ranges for the respective years and sites, see Tables 1a and 1b. It should be noted that for both data sets the yield was reduced significantly due to bad weather conditions (lack of rain) during the growing season 2006.

2.5 Data Overview

In this work, we evaluate data sets from two different fields. A brief summary of the available data attributes for both data sets is given in Tables 1a and 1b. On each field, different fertilization strategies have been used. One of those strategies is based on a technique that uses a multi-layer perceptron (MLP) for prediction and optimization. This technique has been presented and evaluated in, e.g., [14, 16] or [21]. For each field, one data set will contain all records, thus containing all the different fertilization strategies. Another data set for each field will be a subset of the first that only contains those data records where the MLP has been used, respectively. Table 1c serves as a short overview about the resulting four different data sets.

Table 1: Overview of the F131 and F330 data sets

(a) Data overview, F131

F131	*min*	*max*	*mean*	*std*
YIELD05	1.69	10.68	5.69	0.93
EM38	51.58	84.08	62.21	8.60
N1	47.70	70	64.32	6.02
N2	14.80	100	51.71	15.67
N3	0	70	39.65	13.73
REIP32	719.6	724.4	722.6	0.69
REIP49	722.3	727.9	725.8	0.95
YIELD06	1.54	8.83	5.21	0.88

(b) Data overview, F330

F330	*min*	*max*	*mean*	*std*
YIELD05	4.64	14.12	10.62	0.97
EM38	25.08	49.48	33.69	2.94
N1	24.0	70	59.48	14.42
N2	3.0	100	56.38	13.35
N3	0.3	91.6	50.05	12.12
REIP32	719.2	724.4	721.5	1.03
REIP49	723.0	728.5	726.9	0.82
YIELD06	1.84	8.27	5.90	0.54

(c) Overview on available data sets for specific fertilization strategies for different fields

F131-all	YIELD05, EM38, N1, REIP32, N2, REIP49, N3, YIELD06, *fert. strategy*
F131-net	subset of F131-all where fertilization strategy is *neural network*
F330-all	YIELD05, EM38, N1, REIP32, N2, REIP49, N3, YIELD06, *fert. strategy*
F330-net	subset of F330-all where fertilization strategy is *neural network*

2.6 Fertilization Strategies

There were three different strategies that have been used to guide the nitrogen fertilization of the fields. F131 contains data resulting from two strategies (F, N) and F330 contains data from three strategies (F, N, S). The three strategies are as follows:

F – uniform distribution of fertilizer according to long-term experience of the farmer
N – fertilizer distribution was guided by an economic optimization with a multi-layer perceptron model; the model was trained using the above data with the current year's yield as target variable that is to be predicted (see, e.g., [16]).
S – based on a special nitrogen sensor – the sensor's measurements are used to determine the amount of nitrogen fertilizer that is to be applied.

2.7 Research Target

The overall research target is to find those indicators of a field's heterogeneity which are optimal for yield prediction. Furthermore, from the agricultural perspective, it is interesting to see how much the influencable factor "fertilization" influences the yield in the current site-year. There may be additional factors that correlate directly or indirectly with yield. These could be discovered from the acquired data using standard regression or correlation analysis techniques like regression trees or principal component analysis.

Self-organizing maps (SOMs) provide another relatively self-explanatory way to analyze those high-dimensional yield data visually. They are able to reduce the high dimension of the input data onto a two-dimensional map. They can provide insight into the underlying correlations in the data, as shown in [15]. An additional approach towards visualizing the data is to use multidimensional scaling techniques for dimensionality reduction, such as Sammon's mapping.

In this paper we will present experimental results in visualizing the available data with SOMs and multidimensional scaling techniques which helps in understanding them and will ultimately lead to new heterogeneity indicators. The following section will briefly summarize the two proposed techniques to visualize the data that we have presented before. SOMs and Sammon's mapping will be outlined briefly, with the main focus on data visualization.

3 Data Visualization

This section deals with the basic techniques that we used to visualize the agricultural yield data. The most essential feature of the visualization techniques has to be an appropriate dimensionality reduction to two or three dimensions. Furthermore, correlations between data attributes should easily be recognizable. Computational complexity is a smaller issue, but can not be neglected once the visualization goes into a production environment.

In the past, a multitude of techniques have been proposed to visualize high-dimensional data. A good overview can be found in [3]. Among others, self-organizing maps have shown to be a successful approach to this problem, as demonstrated in, e.g., [6], [12] or [15]. Multi-dimensional scaling techniques have been used to reduce high-dimensional data to two or three dimensions in, e.g., [2], [4] or [13]. SOMs and MDS will be explained shortly in the following sections and their differences will be pointed out. Since both techniques have been shown to be useful in data visualization, they will be applied to the agriculture data from the preceding section and results will be presented.

3.1 Self-Organizing Maps

Our approach of using SOMs is motivated by the need to better understand the available yield data and extract knowledge from those data. SOMs have been shown to be a practical tool for data visualization [5]. Moreover, SOMs can be used for prediction and correlation analysis, again, mostly visually [6]. As such, the main focus in explaining Self-Organizing Maps in the following will be on the visual analysis of the resulting maps.

SOM Theory Self-Organizing Maps have been invented in the 1990s by Teuvo Kohonen [7]. They are based on unsupervised competitive learning, which causes the training to be entirely data-driven and the neurons on the map to compete with each other. Supervised algorithms like MLPs or Support Vector Machines require the target attribute's values for each data vector to be known in advance whereas SOMs do not have this limitation.

Grid and Neigborhood: An important feature of SOMs that distinguishes them from Vector Quantization techniques is that the neurons are organized on a regular grid. During training, not only the Best-Matching Neuron, but also its topological neighbors are updated. With those prerequisites, SOMs can be seen as a scaling method which projects data from a high-dimensional input space onto a typically two-dimensional map, preserving similarities between input vectors in the projection.

Structure: A SOM is formed of neurons located on a usually two-dimensional grid having a rectangular or hexagonal topology. Each neuron of the map is represented by a weight vector $m_i = [m_{i1}, \cdots, m_{in}]^T$, where n is equal to the respective dimension of the input vectors. The map's neurons are connected to adjacent neurons by a neighborhood relationship, superimposing the structure of the map. The number of neurons on the map determines the granularity of the resulting mapping, which, in turn, influences the accuracy and generalization capabilities of the SOM.

Training: After an initialization phase, the training phase begins. One sample vector $\mathbf{x}$ from the input data set is chosen and the similarity between the sample and each of the neurons on the map is calculated. The Best-Matching Unit (BMU) is determined: its weight vector is most similar to $\mathbf{x}$. The weight vector of the BMU and its topological neighbors are updated, i.e. moved closer to the input vector. The training is usually carried out in two phases: the first phase has relatively large learning rate and neighborhood radius values to help the map adapt towards new data. The second phase features smaller values for the learning rate and the radius to fine-tune the map.

Visualization: The reference vectors of the SOM can be presented via a component plane visualization. The trained SOM can be seen as multi-tiered with the components of the vectors describing horizontal layers themselves and the reference vectors being orthogonal to these layers. From the component planes the distribution of the component values and possible correlations between components can easily be obtained.

Practicable Features The visualization of the component planes is the main feature of the SOMs that will be utilized in the following section. Correlations between features can easily be detected. Furthermore, in our case, the data records are labeled with different fertilization strategies which enables us to construct a labeled map. In the process of

training the self-organizing map similarities between neighboring map units can easily be computed. Those are used to construct a U-Matrix which shows those similarities and enables the distinction of homogeneous clusters and cluster boundaries.

In this work, we have used the Matlab SOM toolbox authored by [20] with the default presets and heuristics for determining map sizes and learning parameters.

3.2 Multidimensional Scaling: Sammon's Mapping

Multidimensional scaling (MDS) describes a family of methods that aim to present the underlying structure of the data in a lower number of dimensions, typically on a two-dimensional map. MDS estimates the coordinates of a set of objects $Y = \{y_1, \ldots, y_n\}$ in a feature space of low dimensionality, resulting from the data $X = \{x_1, \ldots, x_n\}$. It tries to preserve the distances between pairs of objects. Different MDS methods usually result from using different ways of computing these distances. The distances are stored in a distance matrix

$$D^x = (d_{ij}^x),\ d_{ij}^x = ||x_i - x_j||,\ i, j = 1, \ldots, n.$$

The estimation of the coordinates is carried out under the constraint that the error (called *stress*) between the distance matrix D^x of the data set and the distance matrix $D^y = (d_{ij}^y), d_{ij}^y = ||y_i - y_j||,\ i, j = 1, \ldots, n$ of the corresponding transformed data set is minimized.

A commonly used error measure is the so-called *Sammon's mapping* [17]:

$$E = \frac{1}{\sum_{i=1}^{n} \sum_{j=i+1}^{n} d_{ij}^x} \sum_{i=j}^{n} \sum_{j=i+1}^{n} \frac{(d_{ij}^y - d_{ij}^x)^2}{d_{ij}^x}$$

The above equation describes the absolute and the relative quadratic error. Error minimization can be achieved by using a gradient descent method. As the desired side effect, the transformed data set Y is computed. For further details on the iterative process of parameter optimization, we refer to [13], which also gives a good overview over advanced MDS techniques. For the purpose of this paper, we decided to generate the results using the basic Sammon's mapping. More information about different MDS methods can be found in, e.g., [1].

It should be noted that the complexity of MDS is $O(c \cdot n^2)$, where c is the (unknown) number of iterations needed for convergence of the used gradient descent. The main difference between MDS and the aforementioned SOM is that MDS does not construct an explicit mapping from the high-dimensional space to the lower-dimensional space. Instead, it tries to position the lower-dimensional feature suitably. Therefore, in contrast to SOMs, when new data points have to be visualized, they cannot be mapped directly, but the MDS procedure has to be carried out again as a whole.

For computing the mapping, the matlab script for Sammon's mapping from the University of East Anglia was used[6].

[6] `http://theoval.sys.uea.ac.uk/~gcc/matlab/#sammon`

4 Visualized Results

This section will present some of the experimental results that we have obtained using SOMs and Sammon's mapping on agricultural data. For each data set, the SOMs and the mapping results will be presented and analyzed (Sections 4.1 thru 4.4). One of the data subsets, F131-net, will be analyzed further with the mapping technique to show interesting results in Section 4.5. The data sets have been described in Section 2, an overview has been given in Table 1c.

4.1 SOM results for F330-all

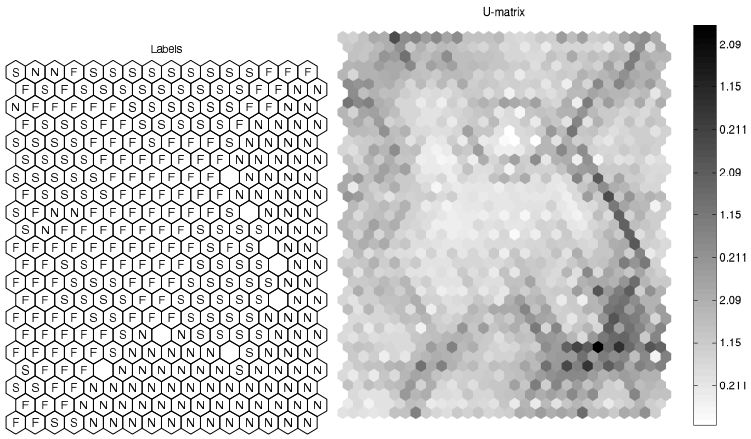
In contrast to the F131 dataset, F330 contains three different fertilization strategies. The "farm" strategy (labeled *F*), the "neural network" strategy (labeled *N*) and a third one. The third strategy (labeled *S*) is based on a special nitrogen sensor – the sensor's measurements are used to determine the amount of nitrogen fertilizer that is to be applied. In Figure 2a it can be seen that the *N* strategy is separable from the other two variants. However, the *F* and *S* strategies are not clearly separable. The U-matrix in Figure 2b also represents this behaviour. When looking at the projected values of N1, N2 and N3 in the component planes in Figures 3a to 3c, the differences between the *N* and *F* or *S* strategies are again clearly visible. There is, however, no such clear connection between the REIP49 (Figure 4a) and YIELD06 (Figure 4c) parameters. This might be due to the fact that the overall yield was significantly reduced by bad weather conditions in 2006. Nevertheless, there is a certain similarity between the relative yields that can be easily obtained by comparing YIELD05 to YIELD06 in Figures 4b and 4c.

4.2 Sammon's mapping results for F330-all

Figure 5a shows Sammon's mapping for the F330-all dataset. Visual inspection yields (at least) three clusters. However, these clusters do not represent the farming strategies (F, S, N), as had been expected. In this case, the results from the three farming strategies after the mapping of the data points have limited explanatory power. It is, however, very likely that further inspection of the mapping will yield deeper insight. In principle, the same technique as in Section 4.5 could be applied, namely splitting the data into three sub-data sets which each contain exactly one farming strategy. To demonstrate the use of this idea, we use the second presented data set in Section 4.4 which covers a field where only two farming strategies were applied. For the sake of clarity of presentation, we postpone the visual analysis of this mapping to a later stage.

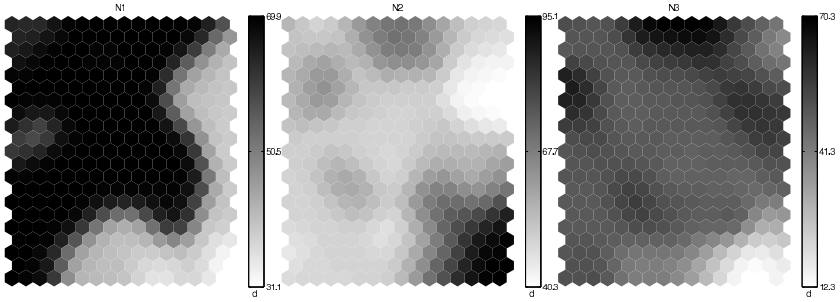
4.3 SOM results for F131-all

The complete F131-all dataset consists of two separate fertilization strategies which are known beforehand. One strategy is to fertilize the field uniformly, this is labeled as *F* (for "farm"). The other strategy is guided by neural networks which learned from past data and from current vegetation and yield indicators to predict the current year's yield – this strategy is labeled *N*. After training the SOM using the preset heuristics from the



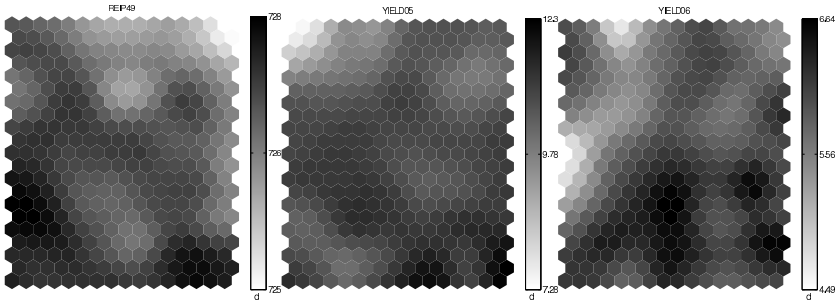
(a) Labels (b) U-Matrix

Fig. 2: F330-all, U-Matrix and Labels



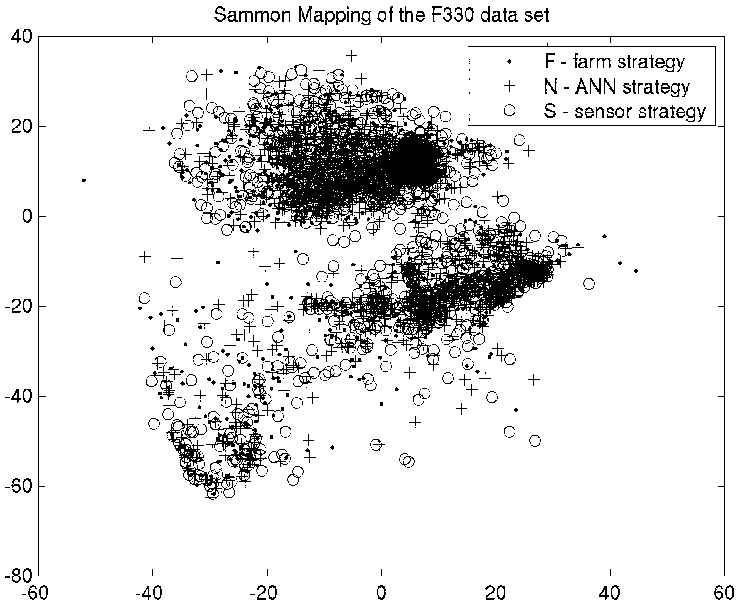
(a) N1 (b) N2 (c) N3

Fig. 3: F330-all, N1, N2, N3

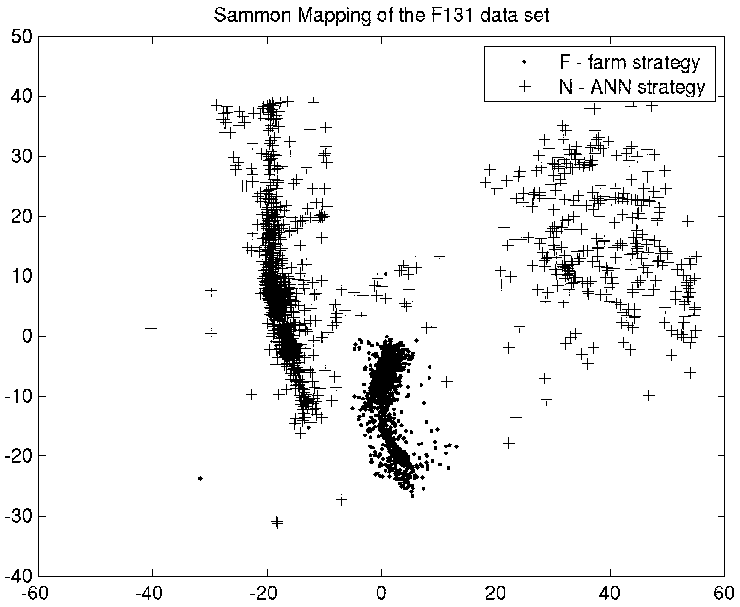


(a) REIP49 (b) YIELD05 (c) YIELD06

Fig. 4: F330-all, REIP49 vs. YIELD05 vs. YIELD06



(a) Sammon's mapping for F330-all



(b) Sammon's mapping for F131-all

Fig. 5: Sammon's mappings for F330-all and F131-all

toolbox [20], the labeled map that results is shown in Figure 6a. The corresponding U-Matrix that confirms the clear separability of those two fertilization strategies is shown in Figure 6b. In Figures 7a to 7c the amount of fertilizer for the three different fertilization times is projected onto the same SOM. On those three maps it can also be seen that the different strategies are clearly separated on the maps. Another result can be seen in Figures 8a and 8c. As should be expected, the REIP49 value (which is an indicator of current vegetation on the field) correlates with the YIELD06 attribute.

4.4 Sammon's mapping results for F131-all

Figure 5b presents Sammon's mapping of the F131-all data set with both farming strategies. Three clusters can easily be identified. Different from the F330-all data set in Section 4.2, the two farming strategies are separated in the mapping. This continues the results from the preceding section, where this particular data set was projected onto a SOM. Out of the three visible clusters, the middle one on the bottom of Figure 5b contains the data points that represent the "F" strategy. The SOM results and the mapping results lead us to the hypothesis that the "N" strategy, which employs a neural network for yield prediction and optimization, merits further research. For this purpose, the data set is split and the data that represent the "N" strategy will be analyzed in the following section.

4.5 Sammon's mapping results results for F131-net

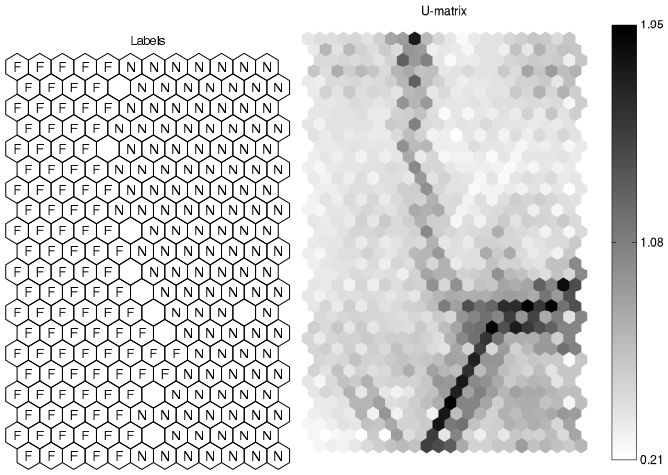
Figure 9 presents Sammon's mapping for the F131-net data set. The neural network that was trained and used for yield prediction and optimization does not divulge much information in terms of which relationships are most important. Therefore, our technique of using Sammon's mapping is applied here to reveal some of the correlations that were found by the neural network. Since the YIELD06 attribute is the variable that we are most interested in for prediction and optimization, it is considered first to separate the data points into those that are below and above a certain yield threshold. We decided to split the data points roughly in the middle of the YIELD06 range, at 5.6 $\frac{t}{ha}$. This presents us with two figures: Figures 9a and 9b convey the notion that there are three data clusters. The "high-yield" cluster is on the top right of Figure 9a. Two "low-yield" clusters are located on the bottom and in the top left of Figure 9b.

Since the three clusters are separable, the neural network seems to have learned a connection between some of the variables that separates the yield potential. The task therefore is to find possible variable splits that explain the clusters. This is presented in Figures 9c thru 9f. Four simple rules can be derived from this:

if N1>=60 then YIELD06 = low: since the range of N1 is not as wide as for N2 or N3, this rule is rather coarse.

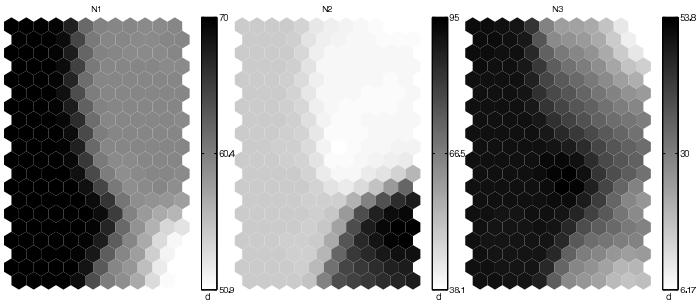
if N2>=55 then YIELD06 = high: The neural network seems to have learned that there are parts of the field where the YIELD06 potential is high so that more N will positively influence YIELD06.

if REIP49<=725.5 then YIELD06=low: Since the REIP49 value characterizes the amount of vegetation on the field at a late growing stage, it is to be expected that a low REIP49 value leads to low YIELD06.



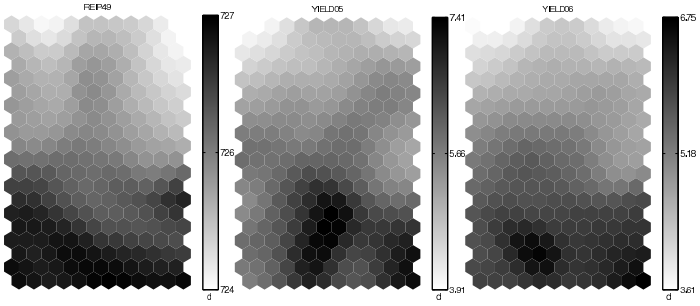
(a) Labels (b) U-Matrix

Fig. 6: F131-all, U-Matrix and Labels



(a) N1 (b) N2 (c) N3

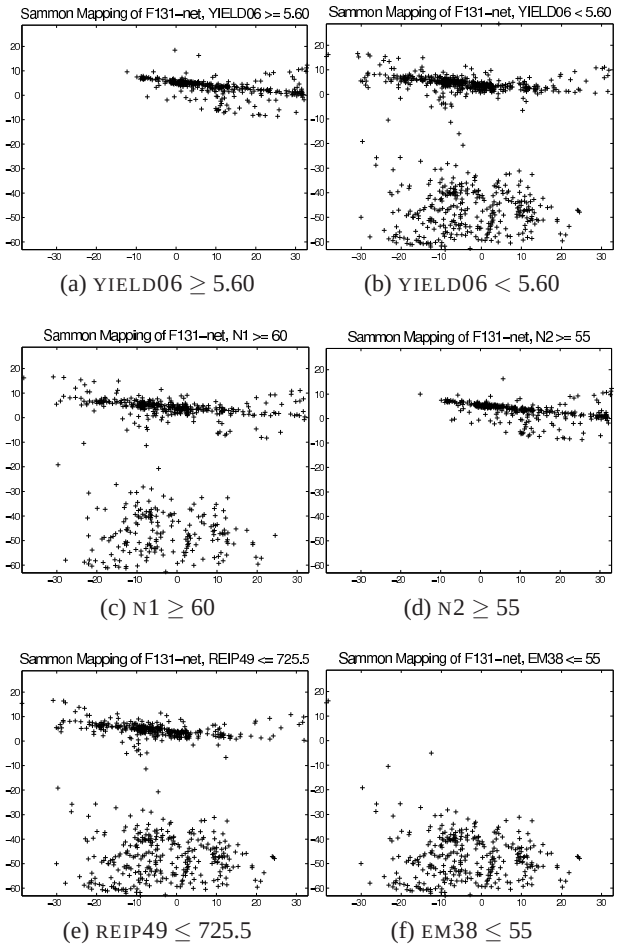
Fig. 7: F131-all, N1, N2, N3



(a) REIP49 (b) YIELD05 (c) YIELD06

Fig. 8: F131-all, REIP49 vs. YIELD05 vs. YIELD06

if EM38<=55 then YIELD06 = low: this rule would explain the bottom cluster in Figure 9b and is, like the others, plausible from an agriculture expert's point of view.



(a) YIELD06 ≥ 5.60 (b) YIELD06 < 5.60

(c) N1 ≥ 60 (d) N2 ≥ 55

(e) REIP49 ≤ 725.5 (f) EM38 ≤ 55

Fig. 9: Sammon's mapping, plot for F131-net, different variable split values. The graphs show only those points of the mapping where the given variable is above or below a certain threshold.

5 Conclusion

In this paper we have presented a novel application of self-organizing maps and a multi-dimensional scaling technique by using them on agricultural yield data. After a thor-

ough description and statistical analysis of the available data sets, we briefly outlined the advantages of self-organizing maps and Sammon's mapping in data visualization. A hypothesis on the differences between two fields could clearly be confirmed by using SOMs and the mapping. A useful correlation between different data attributes could be found using Sammon's mapping. We presented further results, which are very promising and show that correlations and interdependencies in the data sets can easily be assessed by visual inspection of the resulting component planes of the self-organizing map as well as Sammon's mapping graphs. Those results are of immediate practical usefulness and demonstrate the advantage of using data mining techniques in agriculture.

5.1 Future Work

The presented work is part of a larger data mining process. In earlier work, we have presented models and results to represent the agriculture data and use them for prediction and optimization [14, 16]. The visualization will be used on additional data sets to further substantiate its usefulness in the context of precision agriculture data. Once the data can be appropriately visualized, advanced techniques will be developed to establish the benefit of introducing additional sensors which add further data to the process. In the end, only the most effective sensors should be used for yield prediction and optimization.

5.2 Acknowledgements

The field trial data came from the experimental farm Görzig of Martin-Luther-University Halle-Wittenberg, Germany. The trial data have kindly been provided by Martin Schneider and Prof. Dr. Peter Wagner[7] of the aforementioned institution. Supplementary material, such as Matlab scripts, color plots and plot videos can be found at the author's research site `http://research.georgruss.de/?cat=19`.

References

1. Ingwer Borg and Patrick J. F. Groenen. *Modern Multidimensional Scaling: Theory and Applications (Springer Series in Statistics)*. Springer, Berlin, 2nd edition, September 2005.
2. Andreas Buja and Deborah F. Swayne. Visualization methodology for multidimensional scaling. *J. Classification*, 19:2002, 2001.
3. C. Chen, W. Härdle, and A. Unwin, editors. *Handbook of Data Visualization (Springer Handbooks of Computational Statistics) (Springer Handbooks of Computational Statistics)*. Springer, 1st edition, March 2008.
4. Trevor F. Cox and Michael A. Cox. *Multidimensional Scaling, Second Edition*. Chapman & Hall/CRC, September 2000.
5. Timo Honkela, Samuel Kaski, Krista Lagus, and Teuvo Kohonen. WEBSOM—self-organizing maps of document collections. In *Proceedings of WSOM'97, Workshop on Self-Organizing Maps, Espoo, Finland, June 4-6*, pages 310–315. Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland, 1997.

[7] {martin.schneider,peter.wagner}@landw.uni-halle.de

6. T. Kohonen, S. Kaski, K. Lagus, J. Salojarvi, J. Honkela, V. Paatero, and A. Saarela. Self organization of a massive document collection. *Neural Networks, IEEE Transactions on*, 11(3):574–585, 2000.
7. Teuvo Kohonen. *Self-Organizing Maps*. Springer, December 2000.
8. J. Liu, J. R. Miller, D. Haboudane, and E. Pattey. Exploring the relationship between red edge parameters and crop variables for precision agriculture. In *2004 IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages 1276–1279, 2004.
9. U. Meier. *Entwicklungsstadien mono- und dikotyler Pflanzen*. Biologische Bundesanstalt für Land- und Forstwirtschaft, Braunschweig, Germany, 2001.
10. E. M. Middleton, P. K. E. Campbell, J. E. Mcmurtrey, L. A. Corp, L. M. Butcher, and E. W. Chappelle. "Red edge" optical properties of corn leaves from different nitrogen regimes. In *2002 IEEE International Geoscience and Remote Sensing Symposium*, volume 4, pages 2208–2210, 2002.
11. Jacques J. Neeteson. *Nitrogen Management for Intensively Grown Arable Crops and Field Vegetables*, chapter 7, pages 295–326. CRC Press, Haren, The Netherlands, 1995.
12. B. Penn. Using self-organizing maps to visualize high-dimensional data. *Computers & Geosciences*, 31(5):531–544, June 2005.
13. Frank Rehm. *Visual Data Analysis in Air Traffic Management*. PhD thesis, Otto-von-Guericke-Universität, Magdeburg, February 2007.
14. Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. Optimizing wheat yield prediction using different topologies of neural networks. In José Luis Verdegay, Manuel Ojeda-Aciego, and Luis Magdalena, editors, *Proceedings of IPMU-08*, pages 576–582. University of Málaga, June 2008.
15. Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. Visualization of agriculture data using self-organizing maps. In Tony Allen, Richard Ellis, and Miltos Petridis, editors, *Applications and Innovations in Intelligent Systems*, volume 16 of *Proceedings of AI-2008*, pages 47–60. BCS SGAI, Springer, January 2009.
16. Georg Ruß, Rudolf Kruse, Peter Wagner, and Martin Schneider. Data mining with neural networks for wheat yield prediction. In Petra Perner, editor, *Advances in Data Mining (Proc. ICDM 2008)*, pages 47–56, Berlin, Heidelberg, July 2008. Springer Verlag.
17. John W. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, 18(5):401–409, May 1969.
18. M. Schneider and P. Wagner. Prerequisites for the adoption of new technologies - the example of precision agriculture. In *Agricultural Engineering for a Better World*, Düsseldorf, 2006. VDI Verlag GmbH.
19. Michael L. Stein. *Interpolation of Spatial Data : Some Theory for Kriging (Springer Series in Statistics)*. Springer, June 1999.
20. J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas. Self-organizing map in matlab: the SOM toolbox. In *Proceedings of the Matlab DSP Conference*, pages 35–40, Espoo, Finland, November 1999.
21. Georg Weigert. *Data Mining und Wissensentdeckung im Precision Farming - Entwicklung von ökonomisch optimierten Entscheidungsregeln zur kleinräumigen Stickstoff-Ausbringung*. PhD thesis, TU München, 2006.
22. Michael D. Weiss. Precision farming and spatial economic analysis: Research challenges and opportunities. *American Journal of Agricultural Economics*, 78(5):1275–1280, 1996.