ECONOMICALLY OPTIMIZED SITE-SPECIFIC NITROGEN APPLICATION USING DATA MINING TOOLS

B. Burges and P. Wagner

Agribusiness and Farm Management Group
Martin-Luther-University Halle-Wittenberg
Halle (Saale), Germany

ABSTRACT

Future economic and environmental demands on agricultural production require a more efficient use of resources. Excessive use of nutrients may cause leaching, whereas deficits could lead to impediments in tapping full yield potential. As part of an ongoing research project, we investigated the ability to increase nitrogen efficiency for winter wheat fertilization using Artificial Neural Networks (ANN) and Support-Vector-Machines (SVM). Based on a high-resolution yield prediction, a site-specific economic optimal nitrogen amount was determined according to the maximum Nitrogen Cost-free Revenue. Results showed an increase in nitrogen efficiency of about 30 % (ANN) and 10 % (SVM) compared to uniform treatment (UT). However, a decrease in yield level of about 0.6 t ha^{-1} occurred using the ANN-based strategy.

Keywords: Precision agriculture, data mining, site-specific nitrogen fertilization, artificial neural network, support-vector-machine

INTRODUCTION

Over the last decade, economic and environmental demands on agricultural production have been consistently increasing. The realization of the European Union Water Framework Directive claims less leaching of nutrients (esp. nitrogen) into ground and surface waters. At the same time, farmers are required to maximize their economic outcomes with regard to increasing prices for fuel, pesticides or fertilizer to succeed within competitive markets. Following this, it is crucial to make use of the natural yield potential of a field. Due to heterogeneity within a field, this potential is spatially highly variable and, hence, requires site-specific management approaches. However, nitrogen fertilization is widely carried out in a uniform way based on the experience of the farmer.

Prior work has been done by Weigert (2006) who used an ANN approach to determine site-specific fertilization (Wagner et al., 2006). During the years

2005 to 2011, his approach resulted in an average nitrogen efficiency increase of 20 % concurrently with an economical benefit of about 25 EUR ha^{-1} compared to conventional uniform treatment (Wagner, 2012). However, there has been some evidence to conclude that ANN predictions might be widely imprecise at specific tasks (yield prediction) whereas SVM approaches often outperform ANN (Russ, 2009; Russ 2012).

This paper focuses on two data mining approaches to realize site-specific nitrogen fertilization for winter wheat with regard to an economically optimal amount of nitrogen. It should emphasize the advantages of a site-specific fertilization with decision rules from ANN and SVM methods especially compared to those of conventional fertilization. A major focus is put on the nitrogen efficiency of each of the approaches.

MATERIALS AND METHODS

The methodological concept for data mining approaches is based on three major steps: [1] the generation of datasets consisting of empirical training examples to present to the self-learning algorithms, [2] the training of the self-learning algorithms (ANN and SVM) based on that training data to derive prognosis models and the validation of their predictions, and [3] the application of the prognosis model in a field trial and the comparison to the outcomes of uniform treatment (UT).

Creation of training examples

Prior to in-field application, it was necessary to train both data mining models (ANN and SVM) to predict the yield for any specific position in the field at the point in time of each of the three split applications. To best illustrate the characteristic combinations of a field's properties and their resulting yield, a winter wheat field trial was set up. Parameters accounting for heterogeneous conditions of the field were measured and used to create high-resolution maps. High spatial resolution at low-cost measuring and short-term availability was preferred for this study. Thus, we used the apparent soil electrical conductivity EC_a (Geonics EM38, horizontal mode), historical yield maps (from available past combine data logs), and spectral measurements of the Red Edge Inflection Point (REIP) measured by the Yara-N-Sensor at EC development stage 32 and 49 (equivalent to Zadoks scale stage 32 and 49) for describing a field's heterogeneity. Additionally, we set up a randomized nitrogen application design for the same field trial consisting of 36 different nitrogen amount combinations for each of the three split nitrogen applications (SA1, SA2, SA3). These nitrogen combinations ranged from 0 to 270 kg ha^{-1} cumulatively (Fig.).

At the time of harvesting the field trial, a combine with yield logger was used to determine yield at any position within the field. All spatial data that reflect heterogeneity was further assigned to a certain spatially corresponding yield using ESRI ArcGIS 10.2 (nearest neighbor). Based on the distance of each assigned parameter to the closest yield value, those points that were below a certain proximity threshold were flagged valid (otherwise: non-valid). Only in the case that a yield point got all corresponding parameters within the defined threshold distance did it qualify to be used as training points for the self-learning algorithms.

Training of the self-learning algorithms

The nitrogen fertilization levels for each of the split applications along with the spatially corresponding parameters (EC_a, historical yield, REIP) were taken as input (predictors), whereas the harvested yield at the end of the field trial's growing season was taken as the output (target) for the data mining algorithms.

The training examples were used subsequent within the data mining algorithms. These were carried out separately for each split application. Thus, the included input parameters used for training were chosen according to their availability at the time of each split application. For the first yield prediction (at SA1), only historical yield maps, maps of EC_a, and the amount of applied nitrogen at SA1 were used as input. Additionally, for the second yield prediction, a canopy spectral measurement of the REIP at SA2 (EC growth stage 32) and the amount of applied nitrogen at SA2 were used (additional parameters were included in the third split application). An overview of the available parameters at each split application is given in Table 1. For the training of the three ANN and SVM models, the software package IBM Modeler 15 was used.

Table 1. Parameters used for training of the ANN and SVM at the time of each split application based on the specific parameter availability.

Prediction	Time of SA1	Time of SA1	Time of SA1
target	yield	yield	yield
predictors	historical yield EC_a N1[1]	historical yield EC_a N1[1] N2[2] REIP32[4]	historical yield EC_a N1[1] N2[2] N3[3] REIP32[4] REIP49[5]

[1] nitrogen amount at SA1, [2] nitrogen amount at SA2, [3] nitrogen amount at SA3, [4] spectral canopy measurement at EC stage 32 (Zodaks stage 32), [5] spectral canopy measurement at EC stage 49 (Zodaks stage 49)

Applying the predictive models

To apply the trained algorithms to the nitrogen fertilization in field, another field trial was set up in a completely randomized block design, consisting of four treatments with four repetitions.

For the application, spatial units were defined along the tracks as 36 x 6 m plots according to the working width of the fertilizer spreader (Rauch AGT). Within those plots, the same input parameters as used in the training were averaged and passed on to the trained algorithms. Parameter selection within each set corresponded to its availability, and hence, differed between SA1, SA2, and SA3. For each split application, the optimal amount of nitrogen needed to be determined beforehand. Therefore, the algorithms, iteratively, estimated a set of various crop yields for every possible applicable nitrogen amount for each plot. Based on this, the Nitrogen Cost-free Revenue (NCfR, Eq. 1) was calculated according to the price expectation for winter wheat of about 234 EUR t^{-1} (LFL, 2013) and a nitrogen price of 1.08 EUR kg^{-1} N (LWK, 2013). Those combinations (predicted yield and nitrogen amount) were found optimal which resulted in the highest NCfR value (considering one split application at a time).

(Eq. 1)

$$NCfR = \left(Y_{pred} \cdot P_W\right) - (N \cdot P_N)$$

with:
NCfR = Nitrogen Cost-free Revenue
Y_{pred}= predicted yield
P_W= expected price for winter wheat
N= nitrogen amount
P_N= price for nitrogen fertilizer

Results obtained by either approach were further compared to outcomes of UT via an analysis of variances (ANOVA). Considering that ANOVA does not account for spatial dependencies, results obtained from this test may be biased regarding spatial influences. Thus, we additionally attempted to mitigate such drawbacks by applying a two-step procedure based on linear models provided by the SAS proc-mixed routine (SAS 9.3, SAS Institute Inc.). This procedure ensured that small-scale autocorrelation and large-scale trends (e.g., soil quality) were considered.

RESULTS

In regards to outcomes, ANOVA shows no statistically significant yield reduction for ANN and SVM modeling strategies when compared to UT. Average yield for each of the strategies is around 10 t ha^{-1}. SVM and ANN show a monetary benefit of 28 EUR ha^{-1} and 65 EUR ha^{-1}, compared to UT (Table 2). Thus, nitrogen efficiency is improved considerably by about 10 % for the SVM approach and 34 % for the ANN approach.

Table 2. ANOVA results for the different treatments of the field trial.

Treatment	Yield [t ha^{-1}][1]	Applied nitrogen [kg ha^{1-}]	NCfR [EUR ha^{-1}]	Nitrogen efficiency [kg t^{-1}]
UT	10.2	199	1475[2]	19.5
ANN	10.1	130	1540[2]	12.8
SVM	10.2	179	1503[2]	17.5

[1] Least Square Means (Estimates), [2] based on wheat price at the time of harvest (166 EUR t^{-1})

However, when spatial dependencies are factored in, statistically significant crop yield differences are observed between the ANN and UT strategy, with a yield reduction of about 0.6 t ha^{-1} for the ANN-based approach. In contrast to these findings, no statistically significant yield differences occur between SVM and UT (Table 3). The average cumulated amount of nitrogen applied for each strategy ranges considerably between 130 kg N ha^{-1} (ANN) and 199 kg N ha^{-1} (UT). Considering the ANN-based strategy, fewer amounts of nitrogen, however, result in a lower crop yield. In contrast, using the SVM approach, less nitrogen applied (179 kg N ha^{-1}) still resulted in the same level of crop yield (as compared to UT). Differences in NCfR for each of the strategies show that only the SVM appears to have a monetary advantage of about 13 EUR ha^{-1} compared to UT. Further comparison with the UT strategy shows improved nitrogen efficiency by about 10 % with SVM and 30 % for the ANN based approach (Figure 1).

Table 3. Results for the different treatments based on the geo-statistical approach from SAS.

Treatment	Yield [t ha^{-1}][1]	NCfR [EUR ha^{-1}]	Nitrogen efficiency [kg t^{-1}]
UT	10.2	1478[3,4]	19.5[3]
ANN	9.5[2]	1436[3,4]	12.8[3]
SVM	10.2	1491[3,4]	17.5[3]

[1] Least Square Means (Estimates) based on SAS yield estimation, [2] ANN estimate significantly different compared to UT and SVM, [3] results based on SAS yield estimates and actual applied nitrogen, [4] based on wheat price at the time of harvest (166 EUR t^{-1})

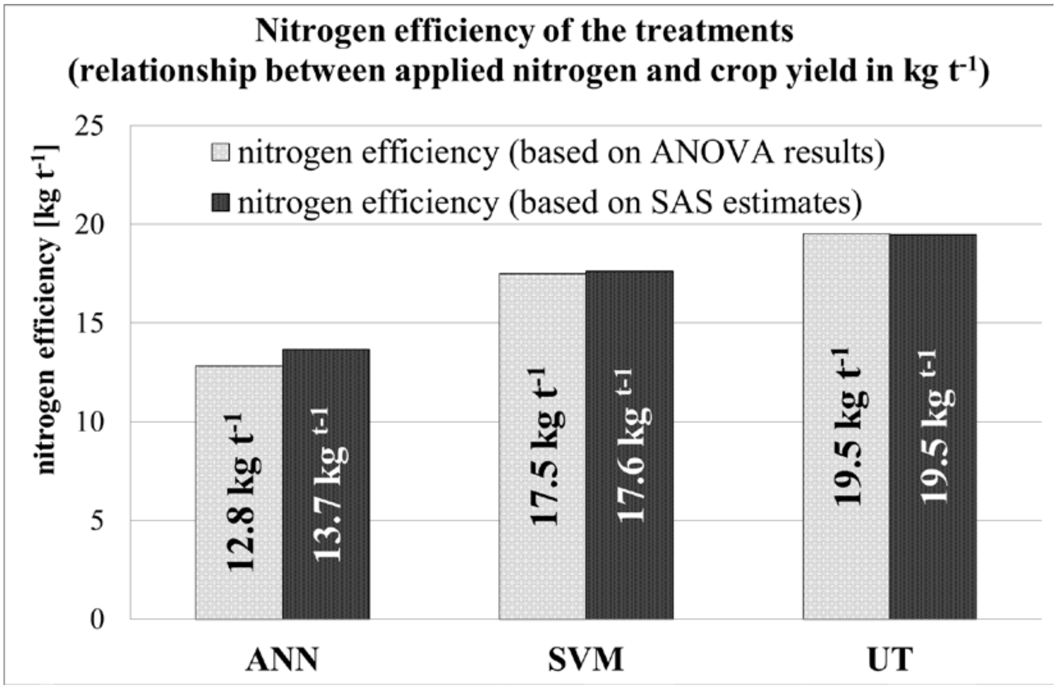

Figure 1. Nitrogen efficiency based on the average cumulated applied nitrogen for each treatment.

DISCUSSION

Considering the results of the ANOVA, the data mining approaches outperform the UT strategy significantly. However, spatial dependencies have to be considered, as it is impossible to have consistent laboratory conditions on the field trial. This leads to unbalanced preconditions for each treatment. In the case of the ANN-based strategy, average EC_a was significantly higher than those of the other treatments. EC_a was determined to be significantly important within the estimated geo-statistical crop yield model determined by SAS. Due to the higher average EC_a, the ANN-based strategy seemed to be spatially advantaged, which the SAS model tried to compensate for. Consequently, the crop yield estimate from SAS for the ANN treatment was reduced significantly. In contrast, the average historical yield for the ANN strategy was not different from those of the UT strategy but had less influence within the SAS model. The positive correlation between EC_a and crop yield, which the SAS model reflected, was also found by Wagner et al. (2006). In contrast, Huang et al. (2005) generally found the opposite relation between both parameters. Considering this, the evaluation of the results from the SAS estimation seemed to be not completely clear.

As of the current stage of this field trial, it has become apparent that economic evaluations are most reliable when using actual applied nitrogen and actual harvested crop yield. However, geo-statistical adjustments of yield were used to take spatial effects into account. Since this estimated variable was set in relation to actual applied amounts of fertilizer – instead of also using actual obtained yield – inaccuracies in estimating monetary benefits may have occurred but have not been further investigated at this point.

CONCLUSION

Based on the results we have reason to conclude that data mining tools are suitable for further optimizing the application of nitrogen. In our study, nitrogen inputs needed for growing winter wheat could be reduced to varying degrees. As nitrogen efficiency has been increased to a considerable extent, the use of data mining tools, and especially the use of SVM for fertilization shall be emphasized. Nonetheless, there are inaccuracies in the evaluation of the results that need to be further investigated.

ACKNOWLEDGEMENTS

The project is supported by funds of the Federal Ministry of Food, Agriculture and Consumer Protection (BMELV) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program.

REFERENCES

LFL (Bayerische Landesanstalt für Landwirtschaft). 2013. Winter wheat price prognosis for Aug. 2013. https://stelmf.bayern.de/idb/winterweizen.html, 11.03.2013.

LWK (Landwirtschaftskammer Niedersachsen). 2013. Average nitrogen prices Mar. 2013. *In* Zeitschrift Land und Forst, 2013, Issue No. 10, 11, 12.

HUANG, X., A. KRAVCHENKO, and K. THELEN. 2005. Yield Variability in Undulating Fields: Classification and Prediction. *In*: Proceedings of the 7th International Conference on Precision Agriculture, 2004, Minneapolis, Conference Abstracts, p.419-431, [CD-ROM] s.p..

RUSS, G. 2009. Data Mining of Agricultural Yield Data: A Comparison of Regression Models. *In*: Perner, P. (ed.) ICDM 2009. LNCS (LNAI), vol. 5633, pp. 24–37. Springer, Heidelberg.

RUSS, G. 2012. Spatial Data Mining in Precision Agriculture. Ph.D. diss., Otto-von-Guericke-Universität Magdeburg.

WAGNER, P., M. SCHNEIDER, and G. WEIGERT. 2006. The Use of Artificial Neuronal Networks to Generate Decision Rules for Site-Specific Nitrogen Fertilization. *In*: 8th International Conference on Precision Agriculture, July 23-26, 2006, Minneapolis, Conference Abstracts, p.82, [CD-ROM] s.p..

WAGNER, P. 2012. The Use of Artificial Neuronal Networks to Generate Decision Rules for Site-Specific Nitrogen Fertilization. *In*: ISPA (International Society of Precision Agriculture), Proceedings of the 11th International Conference on Precision Agriculture, July 18-21, 2012, Indianapolis, Indiana, USA, [CD-ROM] s.p..

WEIGERT, G. 2006. Data Mining und Wissensentdeckung im Precision Farming – Entwicklung von ökonomisch optimierten Entscheidungsregeln zur kleinräumigen Stickstoffausbringung. Ph.D. diss., TU München.