



Using Deep Learning in Yield and Protein Prediction of Winter Wheat Based on Fertilization Prescriptions in Precision Agriculture

Amy Peerlinck¹, John Sheppard¹, Bruce Maxwell²

¹Gianforte School of Computing, Montana State University, Bozeman, MT.

²Land Resources & Environmental Science, Montana State University, Bozeman, MT.

**A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada**

Abstract. Precision Agriculture has been gaining interest due to the significant growth in the fields of engineering and computer science, hence leading to more sophisticated methods and tools to improve agricultural techniques. One approach to Precision Agriculture involves the application of mathematical models and machine learning to fertilization optimization and yield prediction, which is what this research focuses on. Specifically, in this work we report the results of predicting yield and protein content of winter wheat over four farms based on the levels of nitrogen fertilizer applied to the fields. The intent is to use these predictions as a basis for prescribing fertilizer application to optimize net returns on the subsequent harvest. More specifically, we compare methods based on multiple regression (linear and non-linear) and neural networks (shallow and deep). Our results indicate that a deep neural network based on the stacked autoencoder that includes spatial sampling yields the best results.

Keywords.

Precision agriculture, neural network, stacked autoencoder, yield prediction, protein prediction, fertilization optimization, spatial sampling.

1. Introduction

In recent years the amount of petroleum-based fertilizer being wasted has increased significantly. Ecologists believe this increase could be one of the leading causes for the surge in algal blooms, which in turn lead to increased soil toxicity (Michalak *et al.*, 2013). In addition, climate change is believed to influence the rate of carbon sequestration in agricultural soil (Lal, 2004). More pertinent to the current study, farmers are seeking new ways to maximize their profits, especially in the presence of climate change and limited growing seasons. This latter point is one of the primary motivations of our work. These problems are addressed in the field of Precision Agriculture, which combines recent technologies, mostly from the fields of computer science and engineering, to provide informed decision making in an agricultural setting.

There are two aspects of Precision Agriculture that could help to address these issues. The first is to calculate and predict optimized fertilizer rates in order to lessen fertilizer waste and maximize the farmer's profit. However, to accomplish this, yield and protein points have to be predicted based on climate and both current and historical properties of the field. This is the second aspect and the focus of this paper.

Our approach, which we describe in this paper, is to apply machine learning techniques to predict the yield and protein content for specific regions in target fields. In machine learning, the application of Artificial Neural Networks (ANN), especially within the context of deep learning (Goodfellow *et al.*, 2017), has been of growing interest for innumerable domains, including natural language processing, stock market prediction, and medical diagnosis. One of the most interesting advantages of ANNs is their ability to learn and recognize patterns from different input signals, where the network's "neurons" perform the analysis of the input signals simultaneously yet independently.

Applying ANNs in the field of Precision Agriculture is a relatively recent development, as will be discussed in more detail in Section 2 of this paper. Here, we present the results of our exploring two different approaches to using ANNs, each within the context of modeling localized properties of the field as well as expanding the inputs to consider spatial context. Specifically, we compare training simple feedforward neural networks to using a deep learning model known as a stacked autoencoder. We then compare the results of these two neural networks to multiple linear regression and multiple nonlinear regression. Our results indicate that the neural network models (both shallow and deep) provide an improvement over the traditional regression methods. Furthermore, we find that adding the spatial context yields a statistically significant boost in performance.

The rest of the paper is structured as follows. Related work is discussed in Section 2. In Section 3, we discuss each of the models in greater detail. In addition, this section provides background information on Precision Agriculture. Section 4 details the specifics of our experimental approach, the results of the experiments, and a discussion of these results. Finally, possible future research is presented in Section 5, and a conclusion is offered in Section 6.

2. Related Work

Precision Agriculture (PA) is a field that has been receiving a great deal of attention in recent years. Earlier research similar to the work reported here focused on linear and nonlinear regression models, but there has been an increase in the use of ANNs to predict crop yield and optimize fertilization (although, the latter has seen significantly less research). The following sections will give an overview of the most significant research using the aforementioned models in the field of PA.

2.1 Linear and Non-Linear Models

Stepwise multiple linear regression (SMLR) is one of the most commonly used methods to develop empirical models from large data sets. Osborne *et al.* (2002) used SMLR to estimate grain yield in nitrogen- and water-stressed corn. More specifically they used the REG procedure in SAS on collected data to predict plant nitrogen, biomass, grain yield, grain nitrogen concentration, and chlorophyll meter readings. Uno *et al.* (2005) explore whether ANNs perform better than SMLR. The authors note that there are three noteworthy drawbacks to using SMLR. First, SMLR is based on the assumption that a linear relationship exists between input and response variables. Second, noise in the samples is assumed to follow a normal distribution, and third, in some cases, the model tends to overfit, thus reducing its applicability to unseen data. However, their results show that ANNs and SMLR perform similarly. Because of their inconclusive results, the authors emphasize that more research needs to be conducted using both models.

2.2 Machine Learning

Machine Learning (ML) has enabled more in-depth research in a plethora of fields. Training artificial neural networks (ANNs) is one of the most popular techniques in ML and has been applied to a variety of biological and agricultural problems. An interesting and useful aspect of ANNs is their ability to find complicated associations between input and response variables without pre-defining any constraints or assumptions about the sample distribution of the data. This enables the possibility of describing complex non-linear relationships that are often present in domains such as PA resulting from a wide range of crop conditions and other influencing factors (Uno *et al.*, 2005).

This section focuses on agricultural uses of ANNs, more specifically on research concerning yield prediction and fertilization optimization. For a more general overview of ANN applications in biology and agriculture, Samborska *et al.* (2014) summarize research where ANNs have been applied successfully.

Drummond *et al.* (1998) use different methods of training ANNs on the same dataset to predict yield. They focus on training accuracy, test set generalization, and outlier rejection properties to determine the preferred method. The authors looked at backpropagation, with and without weight decay, as well as quickprop and rprop for their training techniques. Contrary to what Uno *et al.* (2005) found, Drummond *et al.* showed that any form of learning and training outperformed the linear method, and that rprop as a learning technique performed slightly better than backpropagation.

Yu *et al.* (2010) introduced an ensemble of feedforward ANNs in fertilization models. They claim there are two common issues when using single feedforward networks in this context. First, they found that assigning a maximum target yield before computation leads to a large error in determining the fertilization rates. Second, they believe that using a single ANN as their model leads to limited forecasting accuracy and generalization capacity. The authors proposed using an ANN with nutrient concentration and fertilization rate as input and yield as output. They trained several neural networks via backpropagation using a bagging procedure and used *k*-means clustering to cluster the networks. Next they selected a representative network from each cluster to form an ensemble of ANNs. They determined the combining weights for the ensemble by applying the Lagrange multiplier method where the associated constraints enforce that the weights sum to one. Then they applied the ensemble to define a nonlinear objective function based on the ensemble output and optimized the fertilizer application rates through nonlinear programming.

Since the aforementioned research, applying ANNs to PA has seen vast improvements. Dahikar *et al.* (2015) applied a feed-forward neural network trained via backpropagation to predict which crop would be best for a certain field based on different soil and atmosphere parameters. Their input parameters included pH-levels, nitrogen, potassium levels, temperature, and rainfall. Furthermore, they predicted the best fertilization rate for the crop. Their predictions were intended for farmers who are not able to perform expensive soil tests. The results showed relatively

accurate predictions and could hence assist these farmers.

A deep neural network (DNN) has been applied by Kuwata and Shibasaki in 2015. They use NDVI, Absorbed Photosynthetically Active Radiation, canopy surface temperature and water stress index as input to their DNN and a Support Vector Regressor (SVR). Using the DNN, RMSE values of approximately 8 are achieved for corn yield prediction, however, the results using the SVR are not shown (Kuwata and Shibasaki, 2015). You et al. (2017) applied a convolutional neural net (CNN) and a Long-Short Term Memory (LSTM) network to the classification of histograms created from remotely sensed images. The CNN achieves the best RMSE values overall, averaging around 7 (You et al., 2017).

Data mining is a separate field of study that frequently uses techniques from machine learning. Pantazi *et al.* (2014) used a method based on Supervised Self Organizing Maps to analyze sensor information, assess measurement data, and update initial knowledge. Their proposed model used input nodes corresponding to the main factors of wheat crop production. The model classified the data to predict wheat yield and productivity. Ru *et al.* (2008) used neural networks to predict wheat yield from freely available and usually low-quality in-season data. Furthermore, the authors applied data mining with neural networks to optimize fertilizer usage. Different networks were built and evaluated, and the results showed that the prediction accuracy of the networks rose once more data became available.

3. Background

Before going into detail on the approach proposed in this paper, some background information on Precision Agriculture in general will be given. Furthermore, Section 3.2 explains the linear and non-linear models we used and how these contribute to the yield and fertilization prediction. Both implemented neural networks—feed forward trained with backpropagation and stacked autoencoders—are clarified in Section 3.3.

3.1 Precision Agriculture

Precision Agriculture (PA) is a field where various technologies are applied to improve agricultural outcomes. The field started with the development of the Global Positioning System (GPS) technology in the 1970s, which made it possible to determine coordinates anywhere on earth, enabling machines to apply treatments with localized requirements for each field (Stafford, 2000). In recent years the fields of study that have gained a lot of traction in PA are computer science and different types of engineering. Zhang *et al.* (2002) give an overview of how different technologies have been applied to agriculture throughout the years and around the globe. Because the conducted research focuses on yield mapping and fertilization optimization, these are the subfields of PA that will be explained in the following sections.

3.1.1 Yield Mapping

Yield mapping concerns the creation of a spatially referenced, graphical representation of crop yield for a certain area. These maps can be based on physical measurements or estimates of the yield using computational models to predict the values. Yield maps are generally split into four categories: inference, prediction, interpolation, and aggregation (Pierce *et al.*, 1997). Inference maps associate yield estimates with existing map coordinates without changing the delineations on a base map, for example when specifying a yield goal with a soil map. Prediction maps fill in values on the map where a yield component is predicted rather than measured. Interpolation maps are coarser than prediction maps and are created by making yield measurements at specific locations within a defined area. Then the yield values between data points are estimated using some form of local estimation. Lastly, aggregation maps render aggregated statistics from the original data, either through measurement or prediction. Of these four types of yield mapping, the aggregation and the prediction of data are the most used in PA.

Essentially, yield mapping captures the mass or volume of grain or harvested crop by location within a specified field and requires three measurements: the yield measurement itself, the area over which the measurement was taken, and the location of the measurement within a field in that area¹ (Pierce *et al.*, 1997). Yield mapping is closely related to our objective in fertilization optimization, as it is often used as a basis to calculate the prime fertilization rate for fields.

3.1.2 Fertilization Optimization

Due to the increase in agricultural production over the past 50 years, the amount of fertilizer being applied has increased significantly. This rise in fertilization results in a boost in production, but it also increases agricultural emissions, more specifically nitrogen emission, in groundwater and surface water. This issue, combined with the goal to apply nitrogen more effectively to increase yield and thus profit, motivated research into the field of fertilization optimization (Van Alphen and Stoorvogel, 2000). The most frequently used method to improve fertilization application is to apply variable rate technology.

Bongiovanni and Lowenberg-DeBoer (2004) give an overview of different fertilization studies performed in PA. Their first literature review concerns nutrition management with a focus on research that studies how fertilization application improves yield. The authors found that most research concurs with the observation that variable rate application for site specific management performs better than using a uniform application rate. Variable rate technology uses different types of optimization models to determine what rate of nitrogen should be applied to which patches in a field. It calculates optimized nitrogen rates for each plot in a field instead of applying the same rate, a uniform rate, across the whole field and uses that as the basis for predicting return (i.e., yield, protein, or net profit). However, sprayers and spreaders that are able to apply these different rates are more expensive, and switching rates takes a higher toll on their mechanics, especially if the rate is changed frequently or if the variations involve particularly large changes. Therefore, it is no surprise that the authors report two studies where uniform rate application had higher profit (Larson *et al.*, 1997; Watkins *et al.*, 1998), especially since both of these studies were conducted when the equipment was less well developed (Bongiovanni and Lowenberg-DeBoer, 2004).

3.2 Linear and Non-linear Regression

Multiple regression is a tool for statistical analysis that uses mathematical methods to define possible relationships between variables. In linear regression, a straight line (or hyperplane) is fitted through the data in order to best match predict the response values. This method is often used to predict certain values for a variable given another variable, but can also assist in explaining certain values or to test a scientific hypothesis (Seber and Lee, 2012). Nonlinear regression works similar to linear regression but instead of constructing a linear surface, other surface shapes are implemented. This shape is determined by the overall spread of the data and can therefore vary greatly. The implemented non-linear model in this research uses a hyperbolic function to estimate the yield and protein values.

¹ In our context, we also track and predict protein content in the harvested crop. While protein is not the same as yield, the associated map is essentially just another kind of yield map.

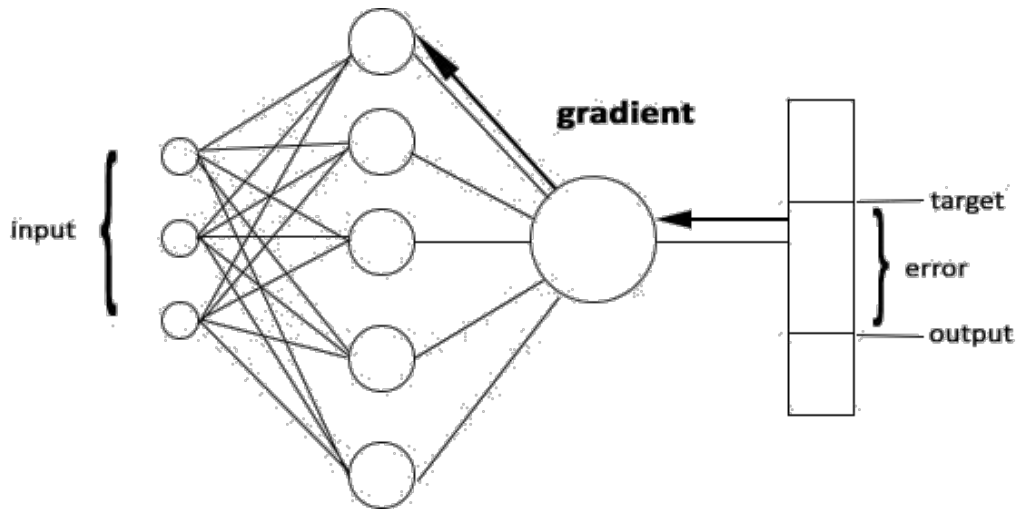


Figure 1: A feed-forward neural network being trained via backpropagation.

3.3 Neural Networks

An artificial neural network is defined as a parallel, distributed information processing structure, modeled loosely on the structure of the brain. This structure contains processing elements connected to each other with unidirectional or bidirectional signal channels (connections). Each of these processing elements possesses local memory and can therefore process local information. All of the processing within an element depends solely on the current value of the input signals and on values stored within the local memory. The elements each have a single output (with an output signal of any mathematical type) branching into as many collateral connections as necessary (Hecht-Nielsen, 1989). This section discusses the feed-forward model in more detail and introduces the concept of a stacked autoencoder as a deep method that, to our knowledge, has not been applied to PA.

3.3.1 Feed-forward Neural Network

In a feed-forward ANN, all nodes of the network pass information forward through a sequence of layers until reaching the output layer. The output of any given node in the network is computed as follows:

$$y = f\left(w_0 + \sum_{i=1}^n w_i x_i\right)$$

where $f(\cdot)$ is often referred to as an “activation function.” Letting $z = w_0 + \sum_{i=1}^n w_i x_i$, these activation functions take on many forms, including linear ($f(z) = z$), rectified linear ($f(z) = \max(0, z)$), logistic ($f(z) = 1/(1 + \exp(-z))$), hyperbolic tangent ($f(z) = \tanh z$), and radial ($f(x) = \exp(-\|x - c\|/\sigma)$).

The updates to the weights in the interior of the network are defined by the backpropagation (BP) algorithm. An example of a feed-forward ANN illustrating the backpropagation process is shown in Figure 1. BP is one of the most popular techniques for training a neural network. It was introduced to a large audience by the Parallel Distributed Processing (PDP) group at Stanford (Rumelhart *et al.*, 1988). The PDP lab also developed the first usable technique to implement backpropagation, and because of this they are usually credited for the current approach of BP, even though there were other researchers who discovered BP earlier (Bryson and Ho, 1969; Werbos, 1974). In BP, the mean squared error of the network is used as the loss function to be minimized, and the updates correspond to the following:

$$\Delta w_{ji} = \frac{\partial \text{Err}(x)}{\partial w_{ji}} = \eta \delta_i x_{ji}$$

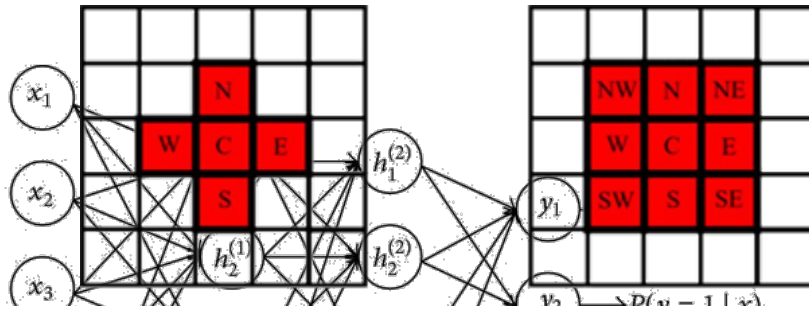


Figure 3: Neighborhood configurations for sampling. The center cell is denoted “C”. The von Neuman neighborhood is shown on the left, and the Moore neighborhood on the right.

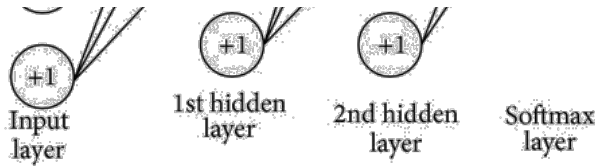


Figure 2: A stacked autoencoder with softmax classifier (Jirayucharoensak *et al.*, 2014).

where at the output layer,

$$\delta_j = (r_j - y_j)y_j(1 - y_j)$$

with r_j being the target response value and x_{ji} being the input from the i^{th} node to the j^{th} node. In the interior layers,

$$\delta_j = o_j(1 - o_j) \sum_{k \in \text{downstream}(j)} \delta_k w_{kj}$$

with o_j being the output generated by the j^{th} node. Note that this update rule assumes a logistic activation function. Other activation functions will modify the update rule slightly because of the different derivative being calculated.

3.3.2 Stacked Autoencoder

The research in this paper uses a different type of ANN that utilizes deep learning to predict yield, namely the stacked autoencoder (SAE). The concept of deep learning arose from the notion that several levels of abstraction in feature space could be the key to improving the modeling of highly nonlinear relationships among the variables and thus perform better generalizations on complex recognition tasks.

The autoencoder was introduced by Bengio *et al.* (2006) and Ranzato *et al.* (2007) as an ANN to try to copy the input it receives to its output with a hidden layer encoding the input. An autoencoder has two functions: 1) encoding the input using the hidden layer, and 2) decoding the encoded values from the hidden layer to reconstruct the input. By design, the autoencoder model is forced to prioritize certain aspects of the input to be copied, thus often learning useful properties of the data (Goodfellow *et al.*, 2017).

A stacked autoencoder is initially built as follows. In the first phase, a single autoencoder is trained using backpropagation with the inputs being treated as target values as well. In the second phase, a separate hidden-layer network is stacked onto the existing autoencoder where the input layer maps to the values generated by the hidden layer of the first autoencoder, and this new network attempts to reconstruct those hidden values at the output. By repeating this process and connecting the learned autoencoders, a “stacked” autoencoder results (Figure 2).

Since the hidden layers of each autoencoder correspond to a set of feature detectors targeted at the signals from the previous layer, this process of stacking is believed to create an abstraction hierarchy of features for the data being analyzed. Once the autoencoders have been trained and

stacked, the final step is to place another network on the top, focused on the goal of learning (e.g., classification, or in our case, yield or protein prediction). Once this is done, all the weights are trained together to learn the final mapping from the original set of inputs to the target yield/protein values (Tan and Eswaran, 2008).

3.4 Spatial Sampling

Spatial sampling is a process where observations are collected in a two-dimensional space. The sampling method is usually designed to capture as many spatial variations of the variables that are to be studied. After initial data has been sampled and documented, additional measurements may be made depending on the variations within the data. These additional measurements are generally made based on criteria to optimize the data (Delmelle, 2014).

The final collected data can be irregular due to a higher interest in certain variables of particular areas or due to terrain conditions. This irregularly spaced data is significantly more complicated to analyze than regularly spaced data. Therefore, managing these irregularities becomes important in order to analyze the datasets more efficiently (Ganesan *et al.*, 2004).

In this research, a field is laid out in a grid pattern; thus, a form of spatial sampling naturally arises from the grid-based arrangement imposed by the fields studied and the prescription maps defined for those fields. Data points reside in each grid cell; therefore, it would be natural to sample, not only from the target cell, but from the neighboring cells. Two common neighborhood functions for such sample are the von Neuman neighborhood (Figure 3 left) and the Moore neighborhood (Figure 3 right). The Moore neighborhood provides a more rounded view of the surrounding samples and was, therefore, implemented as our spatial sampling technique. Specifically, the average values of all points in a single cell were calculated. For each data point the averaged values of the 8 neighboring cells provides the spatial data for that data point. For more consistent spatial information, two alterations to the data were used to decrease noise and smooth the data. The first was the aforementioned averaging technique and second was to increase the grid cell size to contain more points in each cell.

4. Experiments

The following section (section 4.1) outlines our approach to evaluating the various modeling and machine learning methods. The results for the four implemented methods (linear regression, non-linear regression, ANN, and SAE) are presented and compared in section 4.2. Next a comparison is made between non-spatially sampled data and data that was sampled spatially using a k-nearest neighbor sampling method. Finally, these results are explained and discussed in section 4.3.

4.1 Approach

We implemented a simple linear regression through the protein and yield points based on the

Table 1: Amount of data for each field

	sec35mid	sre1314	davidsonMW	carlinW
yield points	17873	24646	11802	15621
protein points	1019	2998	560	655

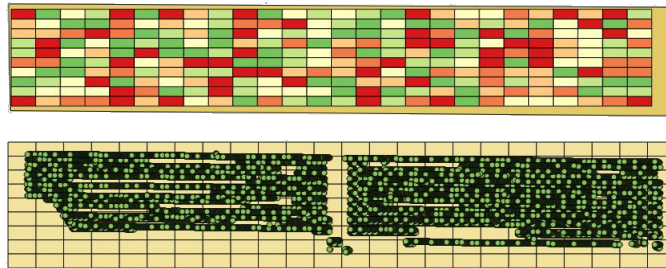
Field	Measure	Linear	NonLin	ANN	ANN Sp	SAE	SAE Sp
sec35mid	RMSE	10.16	10.01	10.64	8.97	10.09	8.95
	R ²	30.03	52.66	24.07	45.15	53.95	65.78
sre1314	RMSE	11.23	11.25	11.51	11.05	10.89	10.15
	R ²	1.37	14.78	9.16	20.69	17.69	32.69

2:	david	RMSE	15.25	15.92	16.36	12.49	16.03	12.06	Table Yield
		R ²	38.94	36.93	30.46	50.64	35.16	59.35	
	carlin	RMSE	12.07	10.20	10.23	8.06	9.93	9.15	
		R ²	1.45	6.79	-4.30	16.17	11.49	9.68	

prediction for all fields. Statistically significant differences are shown in bold. "Sp" stands for spatial data

fertilizer rate. The nonlinear model fit a hyperbolic curve through the data. Using a hyperbolic fit stems from the fact that fertilizer rate hits a saturation point after which yield and protein values no longer rise. We also implemented and evaluated non-spatial and spatial versions of the ANN and the SAE (see section 3.4), resulting in a total of six different models.

During the experimentation process, different parameters were tested to discover which architecture produced the best results. For the ANN and SAE, these parameters included the



number of hidden layers, the number of hidden nodes within those layers, and the number of epochs for the SAE. These parameters were adjusted for each dataset; however, the ANN achieved optimal performance using a single hidden layer with 15 to 100 hidden nodes, depending on the number of features available for the field. The architecture of the SAE varies between two and three hidden layers for non-spatial and spatial data respectively, where the hidden nodes for these layers diminish in each layer. For the spatial data, 500, 250 and 125 hidden nodes were used from the top to bottom layer respectively. The number of hidden nodes for non-spatial data was less consistent, depending on the number of data points available.

All of the models were evaluated using two different measures, namely the coefficient of determination (R^2) and the Root Mean Squared Error (RMSE). Each of these were calculated using 10-fold cross validation, after which the average of the results for each fold was calculated. A paired Student t-test, adjusted using the Bonferroni correction, was performed to evaluate the differences of the means (Bland and Altman, 1995).

The data we analyzed were taken from four different fields, each owned by a different farmer. The data consist of yield and protein values for a specific point, information about their location on the field, the slope at their respective location, the amount of nitrogen applied, the precipitation measured, and the Normalized Difference Vegetation Index (NDVI) for previous years. The yield and protein values are what the implemented models are trying to predict based on the other information provided. The yield and protein values were obtained directly from the farmer and come from fields with previous harvest, where the prescription map was generated based on previous yield bins. An example of such a prescription map is shown on the top in Figure 4; where red, orange, yellow, light green and dark green stand for the nitrogen rates 0, 18, 43, 69, 100 and 150 respectively. Furthermore, the number of cells in a prescription map vary from 200 to 500 depending on the field size and grid cell size.

The amount of data for protein and yield points in each field varies greatly, as can be seen in Table 1. The number of data points is reduced during the spatial sampling process because only points that have 8 neighboring cells containing data are taken into account. An example of

Table 3: Protein prediction for all fields. Statistically significant different are shown in bold. "Sp" stands for spatial data.

Field	Measure	Linear	NonLin	ANN	ANN Sp	SAE	SAE Sp
-------	---------	--------	--------	-----	--------	-----	--------

Figure 4: An example of a prescription map (top) and distribution of yield points (bottom).

sec35mid	RMSE	1.40	N/A	1.44	1.22	1.97	1.96
	R ²	24.60	N/A	23.13	49.72	0.78	17.41
sre1314	RMSE	1.22	N/A	1.16	0.90	1.15	1.13
	R ²	-56.96	N/A	-41.37	11.64	0.41	5.70
david	RMSE	1.68	N/A	1.60	1.49	1.67	1.61
	R ²	-32.20	N/A	-15.68	-9.60	0.05	0.05
carlin	RMSE	1.17	N/A	1.18	1.13	1.37	1.29
	R ²	11.15	N/A	11.12	16.45	4.82	0.48

a grid applied for spatial sampling and the distribution of yield points can be seen on the bottom in Figure 4.

4.2 Results

Tables 2 and 3 show the RMSE and R2 results for yield and protein predictions respectively. For easier readability the R2 values have been multiplied by 100 to show percentages. The non-linear model did not provide results for protein predictions, which is denoted as “N/A” in Table 3. Statistically significant differences between the RMSE values are highlighted in bold.

4.3 Discussion

The results for yield prediction suggest that spatial data generally improves a model’s accuracy (Table 2). With the exception of one field (sre1314), the spatial ANN and spatial SAE perform significantly better than the other models. The lack of improvement on this field may be due to the way the yield points are distributed across the field; the points are closer together along the length of the field but are spaced much wider apart across the width than those in other fields. This effectively increases the number of points in one cell of the grid while decreasing the number of total cells to be taken into account for the spatial data. Overall, applying machine learning techniques to spatial data outperforms the more traditional approaches to yield prediction. More research into other methods of both regression techniques and spatial sampling techniques could provide useful insights into yield prediction and further increase accuracy.

The results for protein prediction tell a different story (Table 3). Although improvement is always achieved within the results for the same model using spatial data, the results are not significantly different from each other. However, the ANNs performed significantly better than the other models on two of the fields. Interestingly, the two fields are the ones with the highest and lowest numbers of protein points. More in-depth analysis of the data may provide additional insights into the cause of this phenomenon. Furthermore, the SAE performs relatively poorly, especially when compared to its performance on yield prediction. The most logical explanation is the much lower number of data points available for training the protein models, preventing the models to learn effectively, as SAEs tend to perform better on a large number of data. However, looking at the RMSE values, the predictions themselves are fairly accurate across all models. When analyzing the data, it becomes apparent that protein values do not differ as much as yield values do, possibly making it easier to predict the correct protein values and leading to smaller RMSE values overall.

5. Future Work

There are several methods of sampling spatial context not described in this work that are worth exploring, such as Equal Distance and Random Selection. We would like to evaluate the relative performance of including these different types of spatial context. The similar performance of the ANN and SAE warrant more investigation into the workings of both models. The currently implemented stacked autoencoder was originally designed for different research, so it may be useful to see whether the autoencoder can be modified to better suit this specific problem.

The structures of the ANN and SAE are adapted depending on the data being analyzed. In other words, unique models are defined for each unique field. An area of research receiving attention is “transfer learning” where knowledge learned in one domain (i.e., a particular field) is reused in

the context of a different problem (i.e., another field). It may be interesting to explore how this could decrease training complexity or even improve overall accuracy.

Another aspect to consider for future work is provided data used to predict the yield and protein values. Investigating which values have the highest impact on accurate prediction and which make little to no difference could be a fruitful research direction. More specifically, looking at how certain values impact the different model types and comparing the differences in influence of these values for each type could provide more insight into the models and how they predict the yield and protein values.

Lastly, this paper is only concerned with the prediction of yield and protein, especially in relation to fertilizer application. As mentioned previously, predicting the optimal fertilization rate is another important aspect of our work, and we are investigating the extent to which we can use these models to facilitate creating optimized fertilization prescriptions.

6. Conclusion

Precision Agriculture has the ability to improve the yield of different crops vastly through applying various mathematical models and experimental processes. This area of study has seen an increase in computationally-based research, especially involving machine learning. By applying machine learning to train predictive models, fertilization prescription maps and yield/protein prediction could become increasingly accurate, resulting in higher productivity. Despite the amount of previous work using machine learning, there has been minimal research into applying deep learning techniques to precision agriculture. The research presented in this paper compared multiple regression, shallow feed-forward networks and stacked autoencoders (a deep method), the latter two both non-spatially and spatially, to generate models for yield and protein prediction. Our results indicate that spatial stacked autoencoders and feed-forward networks seem to provide significantly more accurate results than the competing methods on the fields studied.

Acknowledgments

We would like to thank the State of Montana for supporting this project through the Montana Research and Economic Development Initiative (MREDI). We also thank Richard McAllister for providing the stacked auto-encoder code and assistance with its utilization.

References

- Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2006). Greedy layer-wise training of deep networks. In Proceedings of the 19th International Conference on Neural Information Processing Systems, NIPS'06 (pp. 153–160). Cambridge, MA, USA: MIT Press.
- Bland, J.M., & Altman, D.G. (1995). Multiple significance tests: the bonferroni method. *British Medical Journal*, 310(6973).
- Bongiovanni R., & Lowenberg-DeBoer, J. (2004) Precision agriculture and sustainability. *Precision Agriculture*, 5, 359–387.
- Bryson, A. E., & Ho, Y.-C. (1996). Applied Optimal Control. Blaisdell.
- Dahikar, S., Rode, S., & Deshmukh, P. (2015). Artificial neural network approach for agricultural crop yield prediction based on various parameters. *International Journal of Advanced Research in Electronics and Communication Engineering*, 4(1), 94–98.
- Delmelle, E.M. (2014). Spatial sampling. In Fischer, M.M., & Nijkamp, P. (Ed.), *Handbook of Regional Science* (pp. 1385–1399). Berlin Heidelberg: Springer-Verlag.
- Drummond, S., Joshi, A., & Sudduth, K. (1998). Application of neural networks: Precision farming. In IEEE World Congress on Computational Intelligence (pp. 211–215).
- Ganesan, D., Ratnasamy, S., Wang, H., & Estrin, D. (2004). Coping with irregular spatiotemporal sampling in sensor networks. *Computer Communications Review*, 34, 125–130.
- Goodfellow, I., Bengio, Y., & Courville, A. (2017). Deep Learning. MIT Press.
- Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. In IEEE International Joint Conference on Neural Networks (pp. 65–93).
- Jirayucharoensak, S., Pan-ngum, S., & Israsena, P. (2014). EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation. *The Scientific World Journal* 2014.
- Kuwata, K., & Shibasaki, R. (2015). Estimating crop yields with deep learning and remotely sensed data. In International Geoscience and Remote Sensing Symposium (pp. 858–861).
- Lal, R. (2004). Soil carbon sequestration impacts on global climate change and food security. *Science*, 304(5677), 1623–1627.
- Larson, W.E., Lamb, J.A., Khakural, B.R., Ferguson, R.B., & Rehm, G.W. (1997). Potential of site-specific management for nonpoint environmental protection. In *The State of Site-Specific Management for Agriculture* (pp. 337–367).
- Michalak, A.M., Anderson, E.J., Beletsky, D., Boland, S., Bosch, N.S., et al. (2013). Record-setting algal bloom in lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proceedings of the National Academy of Sciences*, 110(16), 6448–6452.
- Osborne, S.L., Schepers, J.S., Francis, D.D., & Schlemmer, M.R. (2002). Use of spectral radiance to estimate in-season biomass and grain yield in nitrogen- and waterstressed corn. *Crop Science*, 42, 165–171.
- Xanthoula Eirini Pantazi, Dimitrios Moshou, Abdul Mounem Mouazen, Boyan Kuang, & Thomas Alexandridis. (2014). Application of supervised self-organizing models for wheat yield prediction. In International Federation for Information Processing (pp. 556–565).
- Pierce, F.J., Anderson, N.W., Colvin, T.S., Schueller, J.K., Humburg, D.S., & McLaughlin, N.B. (1997). Yield mapping. In *The Site Specific Management for Agricultural Systems* (pp. 211–243). Madison, USA.
- Ranzato, M.A., Boureau, Y., & LeCun, Y. (2007) Sparse feature learning for deep belief networks. In Proceedings of the 20th International Conference on Neural Information Processing Systems (pp. 1185–1192).
- Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1988). Learning representations by back-propagating errors. In James A. Anderson & Edward Rosenfeld (Ed.), *Neurocomputing: Foundations of Research* (pp. 696–699). Cambridge, MA: MIT Press.
- Ru, G., Kruse, R., Schneider, M., & Wagner, P. (2008). Data mining with neural networks for wheat yield prediction. In *Lecture Notes in Computer Science*, 5077, 47–56.
- Samborska, A.I., Alexandrov, V., Sieczko, L., Kornatowska, B., Goltsev, V., Cetner, M.D., & Kalaji, H.M. (2014). Artificial neural networks and their application in biological and agricultural research. *Signpost Open Access Journal of NanoPhotoBioSciences*, 2, 14–30.
- Seber, G.A., & Lee, A.J. (2012). Linear Regression Analysis. Wiley.
- Stafford, J. (2000) Implementing precision agriculture in the 21st century. *Journal of Agricultural Engineering Research*, 76, 267–275.
- Tan, C.C., & Eswaran, C. (2008). Performance comparison of three types of autoencoder neural networks. In Second Asia International Conference on Modeling & Simulation (pp. 213–218).
- Uno, Y., Prashera, S.O., Lacroix, R., Goela, P.K., Karimia, Y., Viauc, A., & Patel R.M. (2005). Artificial neural networks to predict corn yield from compact airborne spectrographic imager data. *Computers and Electronics in Agriculture*, 47, 149–161.
- Van Alphen, B.J., & Stoorvogel, J.J. (2000). A methodology for precision nitrogen fertilization in high-input farming systems. *Precision Agriculture*, 2, 319–332.

- Watkins, K.B., Lu, Y.C., & Huang, W.Y. (1998). Economic and environmental feasibility of variable rate nitrogen fertilizer application with carry-over effects. *Journal of Agricultural and Resource Economics*, 23(2), 401–426.
- Werbos, P. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University.
- You, J., Li, X., Low, M., Lobell, D.B., & Ermon, S. (2017). Deep gaussian process for crop yield prediction based on remote sensing data. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* (pp. 4559–4565).
- Yu, H., Liu, D., Chen, G., Wan, B., Wang, S., & Yang, B. (2010). A neural network ensemble method for precision fertilization modeling. *Mathematical and Computer Modelling*, 51, 1375–1382.
- Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture: a worldwide overview. *Computers and Electronics in Agriculture*, 36, 113–132.