Evaluation of an Artificial Neural Network Approach for Prediction of Corn and Soybean Yield

Angela Kross[a], Evelyn Znoj[a], Daihany Callegari[a], Gurpreet Kaur[a], Mark Sunohara[b], Laura van Vliet[c], Harold Rudy[c], David Lapen[b], Heather McNairn[b]

[a]Department of Geography, Planning and environment, Concordia University, Montreal, QC; [b]Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, Canada; [c]Research and Business development, Ontario Soil and Crop Improvement Association, Guelph, ON, Canada

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

Abstract. *The ability to predict crop yield during the growing season is important for crop income, insurance projections and for evaluating food security. Yet, modeling crop yield is challenging because of the complexity of the relationships between crop growth and the interrelated predictor variables. Artificial neural networks (ANNs) are useful for such complex systems as they can capture non-linear relationships of data without explicitly knowing the underlying processes. In this study, an ANN-based method (Advangeo® Prediction Software) was used to evaluate: 1) the relative importance of predictor variables for corn and soybean yield prediction, and 2) the potential of ANNs for predicting corn and soybean yield. Several satellite derived vegetation indices (e.g. normalized difference vegetation index - NDVI, red edge NDVI, simple ratio - SR, and the land surface water index - LSWI) and slope data were used as crop yield predictor variables, hypothesizing that different vegetation indices reflect different crop and site conditions. The study identified the SR index and the slope as the most important predictor variables for both crop types during both years. The number and dates of the images however were different for the two crop types (earlier dates for corn) and for the wetter (2011) and drier (2012) years. The relative mean absolute errors (RMAEs) were overall smaller for corn compared to soybean and 100% of the corn study sites had errors below 20% in both years. The errors were more variable for soybean. The results are promising and can provide yield estimates at the farm level, unlike current county level approaches.*

Keywords. *Corn, Soybean, yield prediction, remote sensing, vegetation indices, artificial neural network*

Introduction

Crop yield can be used to determine the efficiency of a food production system. The ability to predict crop yield during the growing season is important for crop income, insurance projections and for evaluating the food security at local to global scales. Crop yield prediction requires information about nutrients and water levels, which are related to weather, soil characteristics, field hydrology, crop characteristics, crop rotation, and other management factors (Evans 1993). Modeling crop yield is challenging because of the complexity of the relationships between crop growth and the interrelated predictor variables. Artificial neural networks (ANNs) are particularly useful for studying such complex events, as they can capture non-linear relationships of data without explicitly knowing the underlying processes (Noack et al. 2014). A review study on remote sensing methods for the retrieval of terrestrial vegetation biophysical properties suggested machine learning regression algorithms (MLRAs) such as ANNs as the most promising approach for future remote sensing studies, yet very little literature was found on its application on the estimation of plant development information (Verrelst et al. 2015). ANNs were found successful for the estimation of crop yield, including corn and soybean (Jiang et al. 2004; Kaul et al. 2005; Li et al. 2007;). Yet, these studies focused on predictions at the county level which are not always suitable for use at the farm level. Within this context the objectives of this study were to use an ANN based GIS Software (Advangeo® Prediction Software v 2.0 for ArcGIS 10.0) to evaluate: 1) the relative importance of predictor variables on corn and soybean yield prediction; and 2) the potential of ANNs for predicting corn and soybean yields.

Methodology

The study was conducted within an experimental watershed in eastern Ontario, Canada (45.26 N, 75.18 W, Fig 1). Several satellite derived vegetation indices were used as crop yield predictor variables, hypothesizing that different vegetation indices reflect different land and crop characteristics including: land surface water (i.e. Land surface water index – LSWI), plant nitrogen content and crop characteristics (i.e. Normalized Difference Vegetation Index - NDVI, Simple ratio - R, and red edge NDVI - NDVIre). Gridded slope and temperature data were also used as predictor variables in 2011. The NDVI, SR and NDVIre were derived from RapidEye images, while the LSWI and temperature were derived from Landsat images.

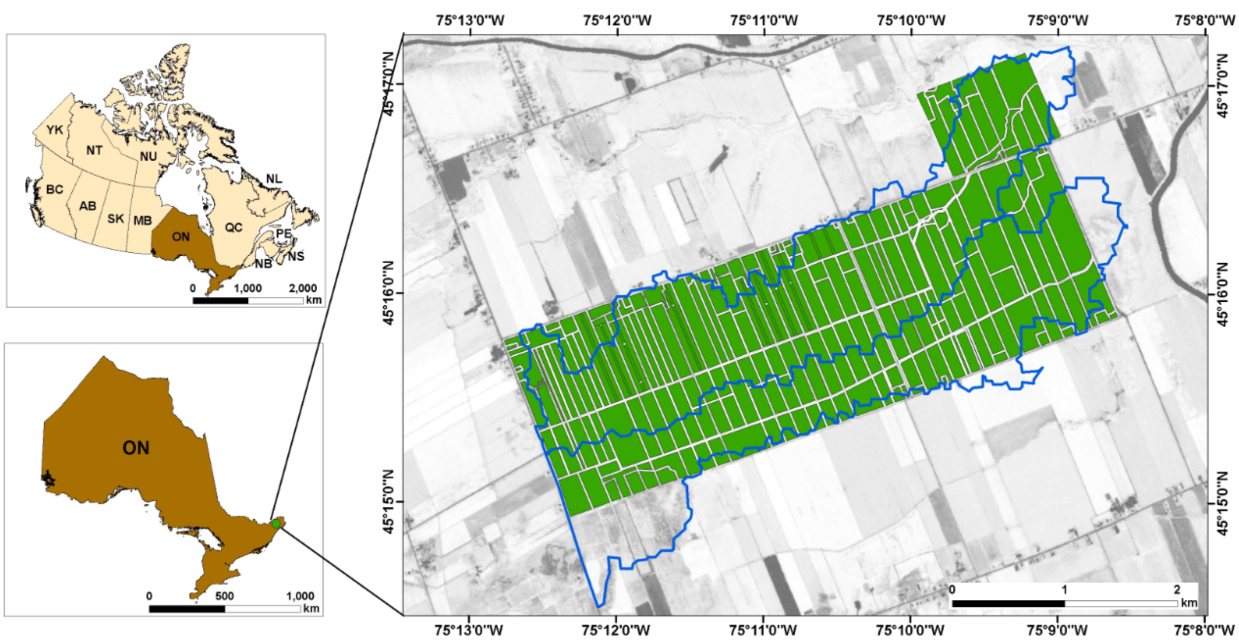

Fig 1. Overview of the study area

The study fields (green polygons) located within the experimental micro watershed (blue outline). The background image is a RapidEye NDVI image. Not all fields were used within this analysis.

The ANN Model was trained and tested with measured yield data (obtained from farmers) from 2011 and 2012 (sites were divided into 50% of training vs. 50% of testing sites). Advangeo®Prediction comes preloaded with defaults that have proven effective in many different applications (Noack et al. 2014). In this study the default parameters were used as a baseline, an overview of the model is given in Figure 2.

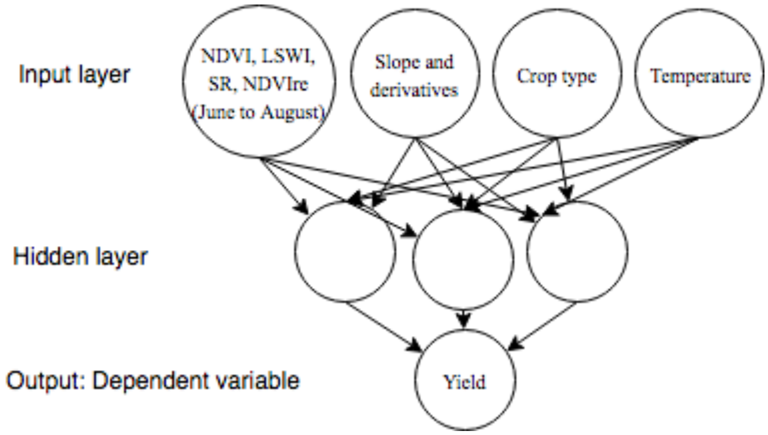

Fig 2. ANN multilayer perceptron network (MLP) for Crop yield prediction.
The input layer receives the controlling parameters, the neurons of the hidden layer(s) and the output layer process the weighted signals from the neurons of its previous layer and calculate an output value applying an activation function. Default parameters were: Network topology (fully connected input layer, and one hidden layer with three hidden neurons), Activation function (for hidden and output layer: sigmoid with a steepness of 0.5), Learning algorithm (derivative of Back propagation Algorithm), Weight initialization ('Initialize' Algorithm of Widrow and Nguyen), and Predefined stop parameters (Number of epochs = 100 and MSE border = 0.001). The input data consisted of different spatial layers of vegetation indices, slope, crop type and satellite derived temperature. All the data were normalized to scale of 0 to 1 for use in the ANN model. The results were scaled back to original data scales for error analysis.

Three main model scenarios were explored for each year: 1) the network was trained with yield data from corn and soybean combined for prediction of yield of the two crops; 2) the network was trained with soybean yield data only for prediction of soybean yield, and 3) the network was trained with corn yield data only for prediction of corn yield. The models that were trained in one year were also tested on other years to evaluate the robustness of the model. For example, an ANN model trained on yield data from 2011 was tested on data from 2011, 2012 and 2016 (these results are not shown in the preliminary results reported in this document).

The relative importance of the predictor variables was determined through Connection Weights and Garson's Algorithm (Garson 1991). The performance of the model was evaluated by calculating the relative mean absolute error (RMAE) for crop sites using the following equations:

$$MAE_{Site} = \sum_{i}^{n} \frac{|Predicted\ Yield - Measured\ Yield|}{n} \quad (1)$$

Where MAE is the Mean Absolute Error in kg/ha. Predicted Yield is the result from the ANN model and Measured Yield is yield data obtained from farmers. n is the number of pixels within a site.

$$RMAE_{Site} = \left(\frac{MAE_{Site}}{Mean\ Yield_{Site}}\right).100 \quad (2)$$

Where RMAE is in % and Mean Yield Site is the average measured yield of the site, in kg/ha.

Preliminary Results

Relative importance of predictor variables

Over 20 predictor variables were evaluated for their relevant importance to crop yield prediction. The results showed two variables with consistently high connection weights and Garson measures for both crops: the SR indices and the slope (Table 1). The dates of the images, however, varied between crop types and between years. In the wetter year (2011, cumulative May - August rainfall of 311 mm at the study area), for example, the Corn model was optimized with only two relatively early SR images: one from June 27 and one from July 23. Soybean's model used two late season images from August, one image from June 27 and the slope data. In the drier year (2012: 270mm), the images were similar for the two crops with the highest weights for July images. The slope and flow accumulation were important for soybean.

Table 1. Overview of the input predictor variables and the most important predictor variables according to Garson's algorithm and connection weights. Negative and positive signs for the most important predictor variables indicate the direction of the relationship.

Yield Training data	Input Predictor variables	Most important predictor variables used in final optimized models
Corn and Soybean (2011)	Slope, aspect, flow accumulation (2010), NDVIre, SR & NDVI (June 10, 27; July 05, 22, 23; Aug 12, 19), Temperature & LSWI (19 June)	SR August 19 (-), SR August 12 (-), Slope (-), SR June 27 (+)
Soybean (2011)		SR August 19 (+), Slope (-), SR August 12 (-), SR June 27 (-)
Corn (2011)		SR June 27 (+), SR July 23 (+)
Corn and Soybean (2012)	Slope, aspect, flow accumulation (2010), NDVIre, SR & NDVI (June 10, 27; July 05, 22, 23; Aug 12, 19), Temperature & LSWI (19 June)	SR 29 July (-), SR 18 August (-), Slope (-), SR 28 June (+)
Soybean (2012)		Slope (-), SR 11 July (+), Flow accumulation (-), SR 18 August (+)
Corn (2012)		Slope (-), SR 18 July (+), SR 18 August (+), SR 04 August (-)

The potential of ANNs for predicting corn and soybean yields

The error analysis showed low performance of the ANN models that were trained on both crop types combined (scenario 1 in Table 2). Most of the test sites had errors above 20% for soybean and above 10% for corn (Table 2).

Table 2. Overview of the performance of the ANN model measured by the RMAE of predicted yield of test sites

		RMAE (%) Soybean				RMAE (%) Corn				Total sites	
Scenario	**Yield Training data**	0-10%	10.01-20%	20.01-30%	30.01-40%	0-10%	10.01-20%	20.01-30%	30.01-40%	Soy-bean	Corn
1	**Corn and Soybean (2011)**	0	2	4	11	0	5	1	0	17	6
2	**Soybean (2011)**	3	4	8	2	N/A	N/A	N/A	N/A	17	
3	**Corn (2011)**	N/A	N/A	N/A	N/A	6	0	0	0		6
1	**Corn and Soybean (2012)**	0	0	4	8	1	2	1	4	12	8
2	**Soybean (2012)**	2	10	0	0	N/A	N/A	N/A	N/A	12	
3	**Corn (2012)**	N/A	N/A	N/A	N/A	3	5	0	0		8

The models that were trained and tested with one crop type (Scenarios 1 and 2 in Table 2) performed well for both years and for both crop types, with errors mainly below 20% (Table 2, Figure 2).

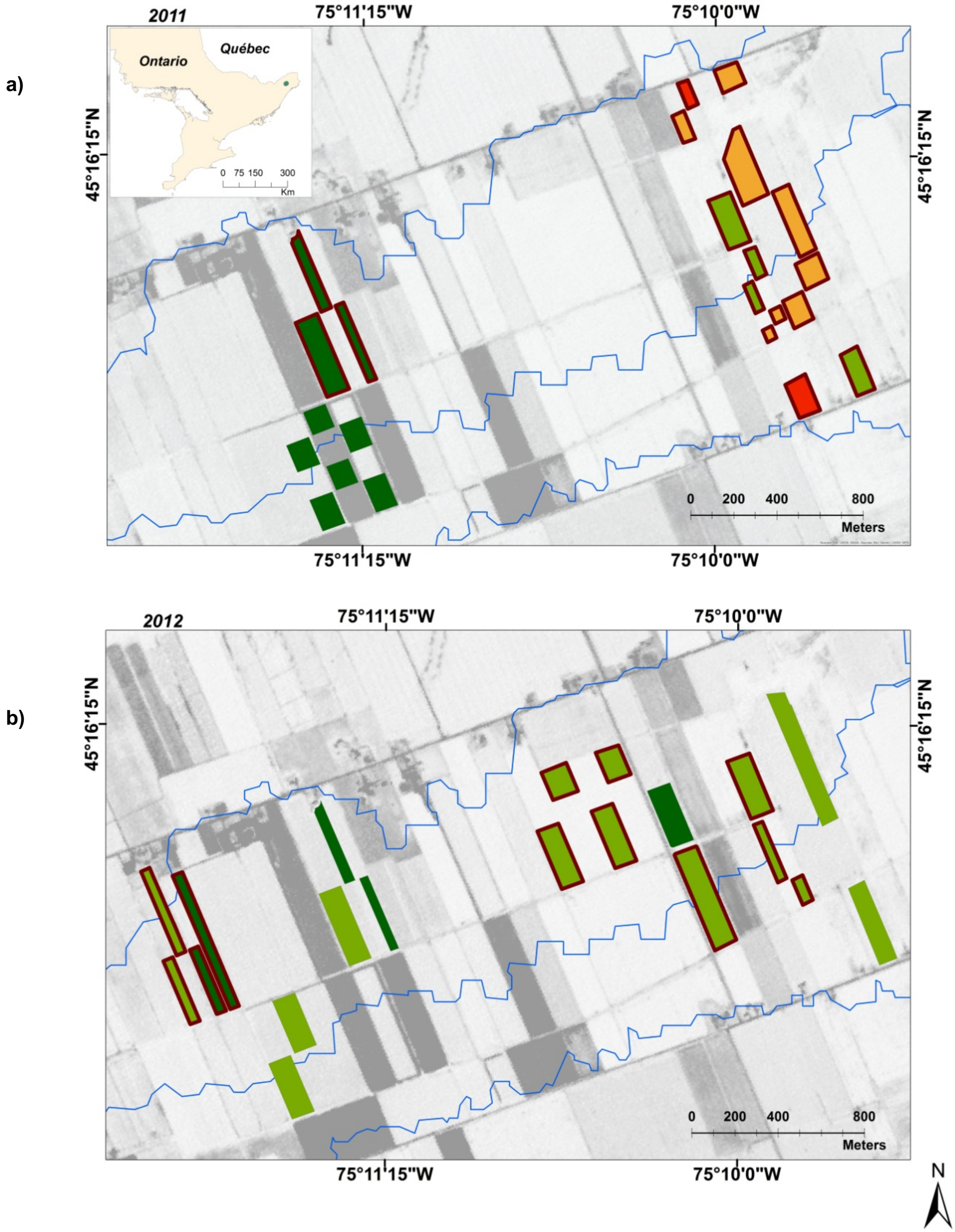

Fig 2. Overview of the results from the ANN model for 2011 (a) and 2012 (b)

Results show the errors from the models that were trained and tested with one crop. Light Green sites indicate relative mean absolute errors (RMAEs) between 0 and 10%. Dark Green sites indicate RMAEs between 10.01 and 20%. Orange sites indicate RMAEs between 20-01 and 30%. and Red sites indicate RMAEs larger than30%. The sites with the thick red outlines are soybean sites, the sites without outlines are the corn sites.

Conclusion

This study showed the potential of the Advangeo®Prediction Software ANN model for corn and soybean yield predictions in eastern Ontario. The results indicate that satellite images can be used to predict yield as long as models are created for unique crop types. The corn specific model performed better than the soybean model: all corn test sites had RMAEs lower than 10% (2012) or 20% (2011). For soybean ANN yield prediction between 42% (2011) and 100% (2012) of the soybean test sites had RMAEs lower than 20%. Slope was an important indicator overall, and future research should explore the inclusion of micro-topography metrics derived from slope data. Rainfall is uniform over the small area, but an indicator of wetness could also improve the performance of the model. The most important predictor variables were not consistent for the wetter (2011) and a drier year (2012), which may indicate that there is a need for a better spatial indicator of wetness. Future studies should also explore potential variables for the development of a crop-independent model.

Acknowledgements

Funding for this project was provided through AgriRisk Initiatives under Growing Forward 2, a federal, provincial, territorial initiative.

References

Evans, L.T. (1993). Crop Evolution. Adaptation and Yield, Cambridge University Press, Cambridge

Garson, G. D. (1991). Interpreting neural-network connection weights. AI expert, 6(4), 46-51.

Jiang, D., Yang, X., Clinton, N., & Wang, N. (2004). An artificial neural network model for estimating crop yields using remotely sensed information. *International Journal of Remote Sensing*, *25*(9), 1723-1732.

Kaul, M., Hill, R. L., & Walthall, C. (2005). Artificial neural networks for corn and soybean yield prediction. Agricultural Systems, 85(1), 1-18.

Li, A., Liang, S., Wang, A., & Qin, J. (2007). Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques. Photogrammetric Engineering & Remote Sensing, 73(10), 1149-1157.

Noack, S., Knobloch, A., Etzold, S., Barth, A., Kallmeier, E. (2014). Spatial predictive mapping using artificial neural networks. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(2):79.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G.P.W., Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. ISPRS Journal of Photogrammetry and Remote Sensing, 10(108):273-90.