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

1. INTRODUCTION

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:

A. the relative importance of predictor variables for corn and soybean yield prediction, and

B. the potential of ANNs for predicting corn and soybean yield.

2. STUDY AREA

The study was conducted within an experimental watershed in eastern Ontario, Canada (45.26 N, 75.18 W).

The study fields (green polygons) located within the experimental micro watershed (blue outline). The background image is a RapidEye NDVI image.

3. METHODS AND ANALYSIS

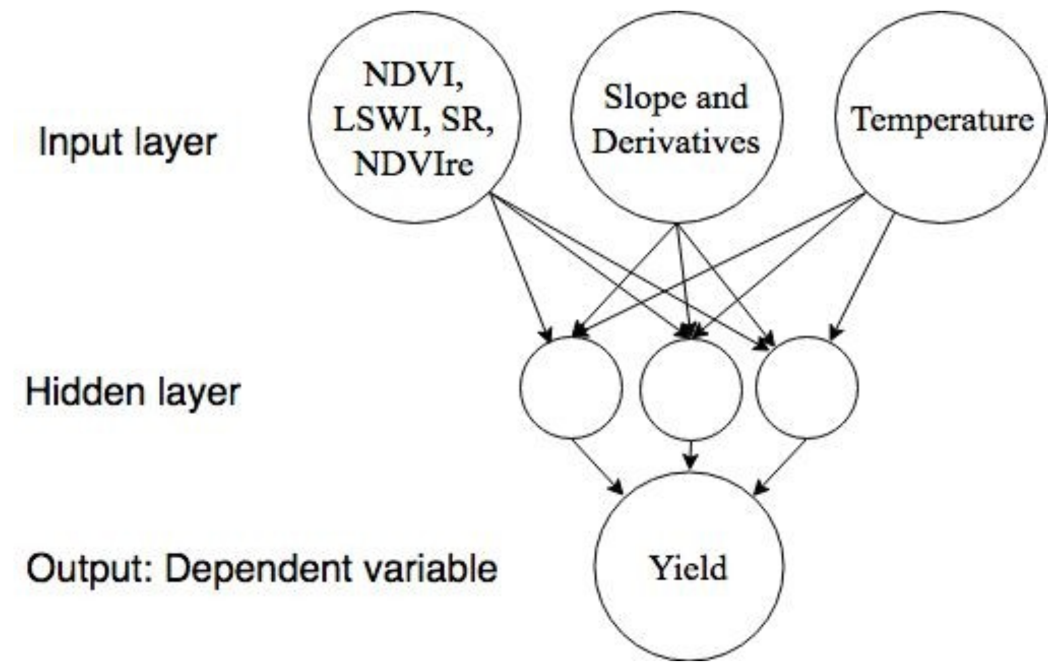
Several satellite derived vegetation indices were used as crop yield predictor variables, hypothesizing that different vegetation indices reflect different land and crop characteristics including:

- Land surface water index (LSWI), indicator of land surface water
- Normalized Difference Vegetation Index (NDVI), Simple ratio (SR), and red edge NDVI (NDVIre), indicators of plant nitrogen content and crop characteristics

Vegetation Index	Acronym	Equation*
Land Surface Water index	LSWI	$(R_{NIR}-R_{SWIR})/(R_{NIR}+R_{SWIR})$
Normalized Difference Vegetation Index	NDVI	$(R_{NIR}-R_{RED})/(R_{NIR}+R_{RED})$
Simple Ratio	SR	R_{NIR}/R_{RED}
Red Edge NDVI	NDVIre	$(R_{NIR}-R_{RED-EDGE})/(R_{NIR}+R_{RED-EDGE})$

*R = Reflectance; NIR = Near infrared

Gridded slope and temperature data were also used as predictor variables. The NDVI, SR and NDVIre were derived from RapidEye images, while the LSWI and temperature were derived from Landsat images.



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). All the variables were normalized to a scale of 0 to 1 for use in the ANN model. The results were scaled back to original data scales for error analysis.

The ANN Model was trained and tested with measured yield data from 2011 and 2012 (50% of training vs. 50% of testing sites).

A. The relative importance of the predictor variables was determined through Connection Weights and Garson's Algorithm (Garson 1991).

B. The performance of the model was evaluated by calculating the relative mean absolute error (RMAE) for crop sites using the following equations:

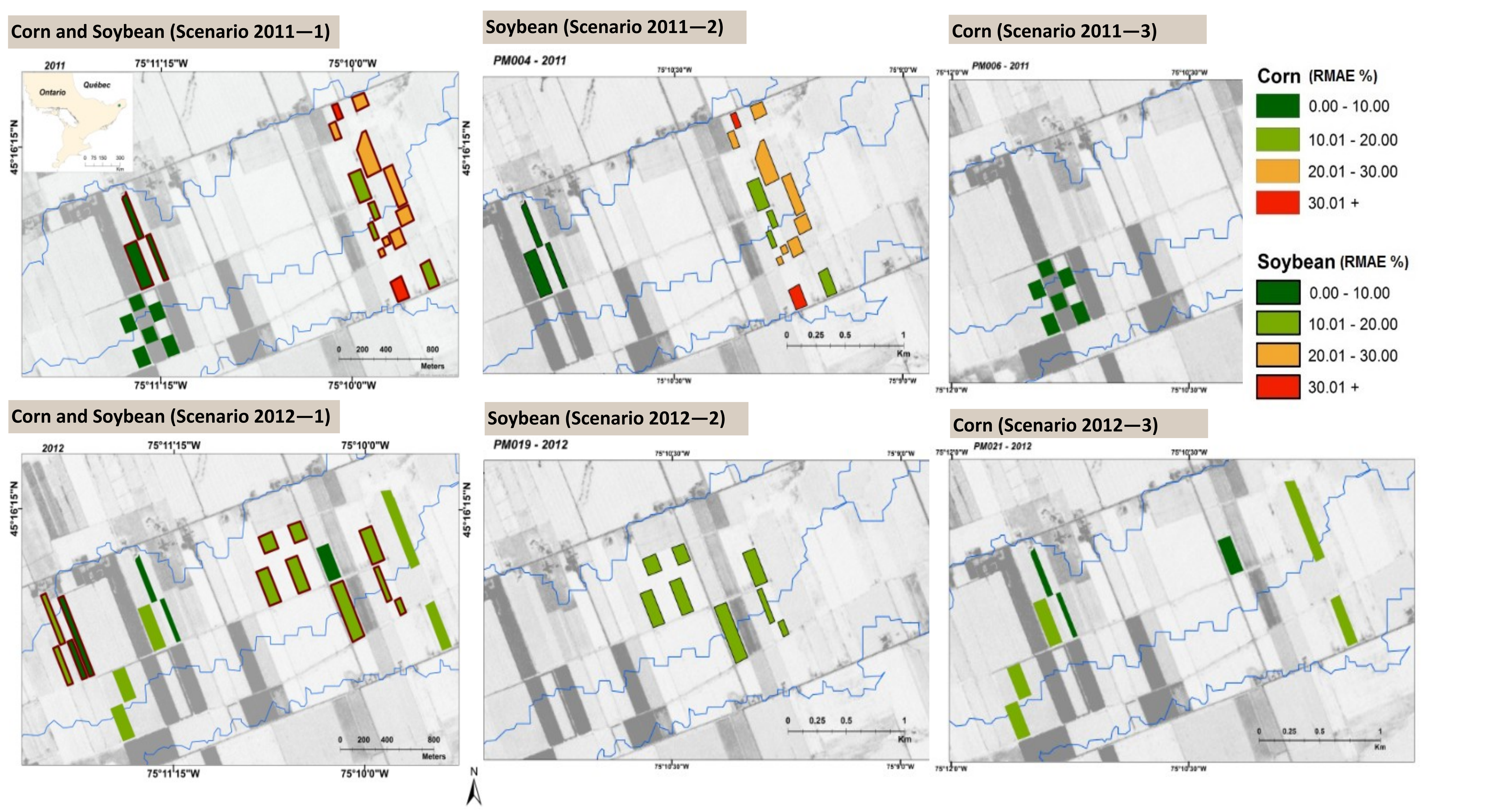
$$MAE_{Site} = \sum_i^n \frac{|Predicted\ Yield - Measured\ Yield|}{n},$$

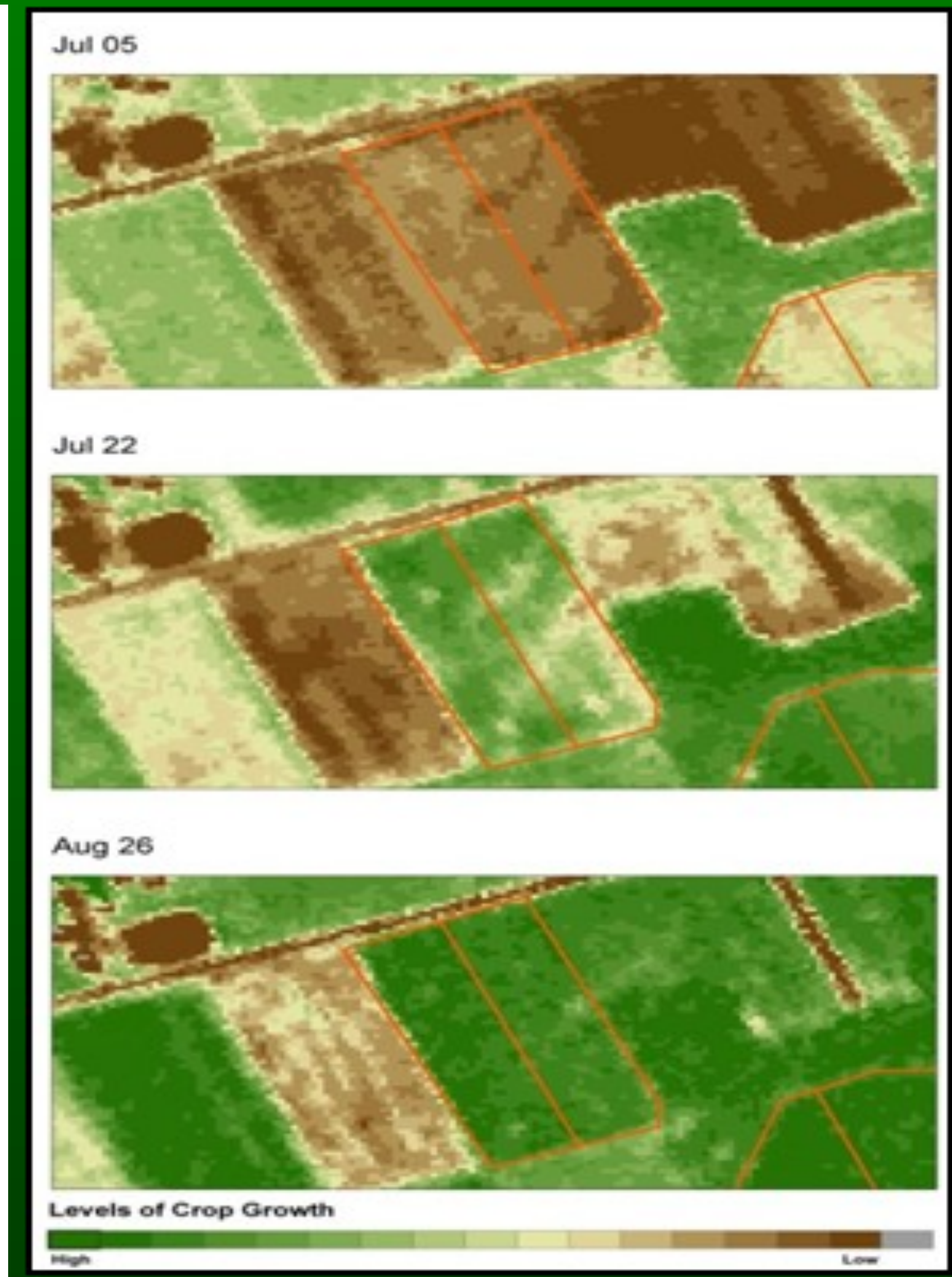
Where MAE is the Mean Absolute Error in kg/ha. Predicted Yield is the result from the ANN model and Measured Yield is yield data obtained from farmers. n is the number of pixels within a site.

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

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

4. RESULTS





A. 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** (the dates of the images, however, varied between crop types and between years). **Flow accumulation** was also important for soybean.

Scenarios	Yield Training data	Input Predictor variables	Most important predictor variables used in final optimized models
2011 - 1	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 (+)
2011 -2	Soybean (2011)		SR August 19 (+), Slope (-), SR August 12 (-), SR June 27 (-)
2011 - 3	Corn (2011)		SR June 27 (+), SR July 23 (+)
2012 - 1	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 (+)
2012 - 2	Soybean (2012)		Slope (-), SR 11 July (+), Flow accumulation (-), SR 18 August (+)
2012 - 3	Corn (2012)		Slope (-), SR 18 July (+), SR 18 August (+), SR 04 August (-)

B. 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). Most of the test sites had errors above 20% for soybean and above 10% for corn.

Three main model scenarios were explored for each year:

Scenario #1: the network was trained with yield data from corn and soybean combined for prediction of yield of the two crops;

Scenario #2: the network was trained with soybean yield data only for prediction of soybean yield, and

Scenario #3: the network was trained with corn yield data only for prediction of corn yield.

5. CONCLUSIONS

Most important predictor variables: Simple Ratio (SR) index, Slope, flow accumulation. 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 indicates that there is a need for a better spatial indicator of wetness.

Potential of ANN using satellite data: good performance but crop dependant .

- 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%.

6. ACKNOWLEDGMENTS

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

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