# BUILDING PROACTIVE PREDICTIVE MODELS WITH BIG DATA TECHNOLOGY FOR PRECISION AGRICULTURE

**C. Belsky, C. Lai, J. Ellingson, C. Greene, S. Morgan, and B. Holub**

*Center for Optimal Control and Autonomy, School of Engineering*
*University of St. Thomas*
*St. Paul, MN, USA*

## SUMMARY OF POSTER PRESENTATION

One drawback of many traditional precision agriculture (PA) paradigms is their reactionary approach in which only the current state of the field is provided to the growers without incorporating accurate predictive forecasts. Predictive modeling can be realized by integrating modern information technologies such as Big Data Analytics, GPS (Global Positioning Systems), remote sensing technology, and GIS (Geographic Information Systems).

The research team at the University of St. Thomas (UST) strongly believe that a proactive approach will become more critical in PA applications. In the proactive approach, predictive models generate forecast reports to the growers, predicting the possible evolution of vegetation states and future risks in the field.

In order to provide highly accurate predictions to growers in real time, the predictive models must integrate tremendous amounts of information from numerous sources such as various sensor data (i.e. temperature, wind, soil pH, moisture) and aerial multi-spectral imagery data. Due to the amount of information that needs to be processed and the complexity of predictive models, the research team at UST plans to utilize a cloud computing platform such as Amazon's Cloud and SAP High performance ANAlytics (HANA) in-memory database. Without up-front infrastructure cost, this cloud computing approach is an affordable way to quickly setup a virtual computing platform to meet our processing needs. HANA provides an integrated environment for analytics solutions that allow data scientists to focus on examining data. Also, HANA can be hundreds of times faster than other big data solutions since it is a memory-based solution.

There are two major cloud-based tasks to our approach: image/data processing and prediction. The first task will process both the aerial multi-spectral image data and the ground-based sensor data. The processed data identified by the first task will then be sent to the second task where forecasts of items such as vegetation and soil health will be predicted.

Thus far, the UST team has worked on two predictive modelling techniques: 1) the "minimum-bounding-box" predictive model, and 2) the multi-variable rule-based predictive model. The proposed predictive modeling techniques take into account the *spatio-temporal* nature of the data.

The "minimum-bounding-box" predictive model was the first modeling technique developed and it focuses solely on one variable – Normalized-Difference-Vegetation-Index (NDVI), which by using multispectral analysis measures the strength of photosynthesis occurring in vegetation. The goal of this predictive model is to predict the size & location of a region of concern. The NDVI images are stitched together to produce one large NDVI map of the field of interest. The areas of concern where the NDVI values indicate stressed vegetation are clustered before a minimum-sized rectangular bounding-box (minBB) is calculated to encapsulate each area of concern. For each point in time the minBB is applied to the NDVI maps, producing a time series of minBBs. Figure 1 shows a time series of false-color NDVI data indicating a growing area of concern over time. In turn, the size, location & orientation of the minBBs can be predicted over time using classical linear regression (CLR) techniques producing a final minBB that covers the predicted area of concern at the desired point in time. This final predicted minBB identifies the area of concern that could be focused on for applying treatments to recover the vegetation (figure 2). Lastly, "velocity" vectors are also applied to the final output showing the magnitude and direction of the NDVI change amongst the time frames.

The multi-variable rule-based predictive modeling technique takes into account spatial relationships between numerous variables beyond just NDVI. This technique attempts to capture these spatial relationships over time by using a combination of a data-mining concept called decision trees (DTs) and CLR. The fundamental steps of the algorithm are:

1. For each point in time, capture the key spatial relationships by building a DT to predict the NDVI value based on the variables' spatial relationships.
2. At each point in time, predict the classified NDVI value using the DTs.
3. Calculate the delta NDVI ($\Delta$N) between points in the time.
4. Cluster the $\Delta$N to filter and focus on various areas of change.
5. Lastly, apply CLR to the points in the areas of concern to predict NDVI values at the desired future time.

The end result of this second algorithm is a prediction that attempts to take into account key spatial relationships amongst the variables. Please note that this algorithm is scalable to take into account additional variables.

These aforementioned predictive model algorithms are just the beginning of the predictive modeling development, and the UST team believes their system may begin shifting the farming paradigm to more anticipatory rather than reactionary techniques.

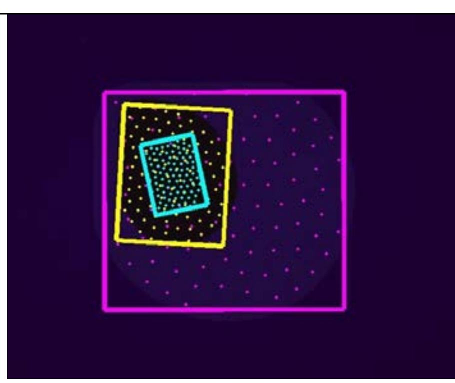

Figure 1: Clustered False-Color NDVI Data for 3-pt Time Series with minBB

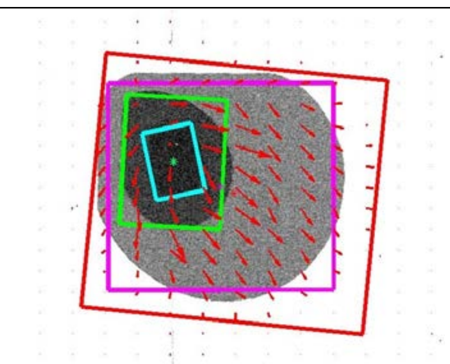

Figure 2: Predicted minBB (Shown in Red) with Velocity Vectors