Using Self-Organizing Map(SOM) Neural Network to Evaluate the Critical Parameters of Cotton Yield

Yawei Cheng & Zhe Lin

Department of Mathematics & Statistics, Texas Tech University

Introduction

West Texas is in the Southern High Plains area, which is the main production area of cotton (Gossypium hirsutum L.) in the U.S. Texas High Plains produces approximately 25% and 64% of the country's and Texas cotton on more than 3.5 million acres fields. The results of 2017 Replicated Agronomic Cotton Evaluations (or RACE) Trials for the Texas High Plains shows that growers on the Texas High Plains produced over 5 million bales of cotton. Most of the cotton fields are dry fields. Only 16% of the planted land area is irrigated, but the irrigated areas account for 41% of harvested yield.

Prediction the final yield based on current soil chemical and physical parameters and remote sensing data will help farmers to modify the farming arrangement in time. Since there are many parameters can be chosen to do the prediction. It will be every efficient if we can find the most critical parameters. So our purpose is to find the critical parameters that will affect the final cotton yield.

Objectives:

- To predict within field variation in cotton yield, based on different soil parameters, satellite images and supervised self-organizing maps neural network algorithm.
- To evaluate the critical crop and soil parameters affecting cotton yield.

Method and Material

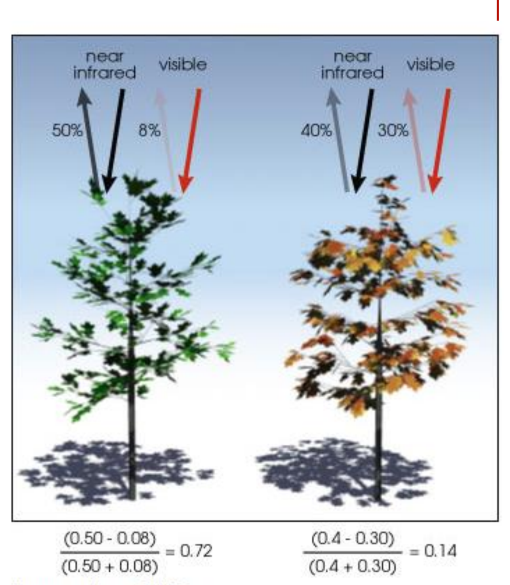
A self-organizing map (SOM) is a type of artificial neural network (ANN) that is trained using unsupervised learning to produce a low-dimensional, discretized representation of the input space of the training samples and is therefore a method to do dimensionality reduction.

Self-organizing maps learn both the distribution and topology of the input vectors they are trained on. The neurons in the layer of an SOFM are arranged originally in physical positions according to a topology function. Those neurons can be arranged in a grid, hexagonal, or random topology.

In the case of SOM, the Kohonen layer and output layers are joined together to bring up a combined layer that is updated taking into account the training regime of SOMs. Every sample (xi) and its related class vector (ci) are associated and form an input for the network.

In our project, we used some soil physical and chemical parameters such as elevation, slope, curvature, surface electric conductivity, deep electric conductivity. Four years yield data and 21 time series NDVI data were also used as the input vectors. There are 30 inputs layers in total. They are all going to working as training data. A 30 by 30 2-Dimensional topology will be presented as a output. The output nodes consisted of yield iso-frequency classes, the yield values were divided in three classes with equal number of samples.

Normalized difference vegetation index(NDVI) is a common used vegetation index in remote sensing, which quantifies vegetation by measuring the reflectance difference between red light band and near-infrared band. Vegetation strongly reflects at near-infrared band and vegetation absorbs most red light. An example shows on the right.

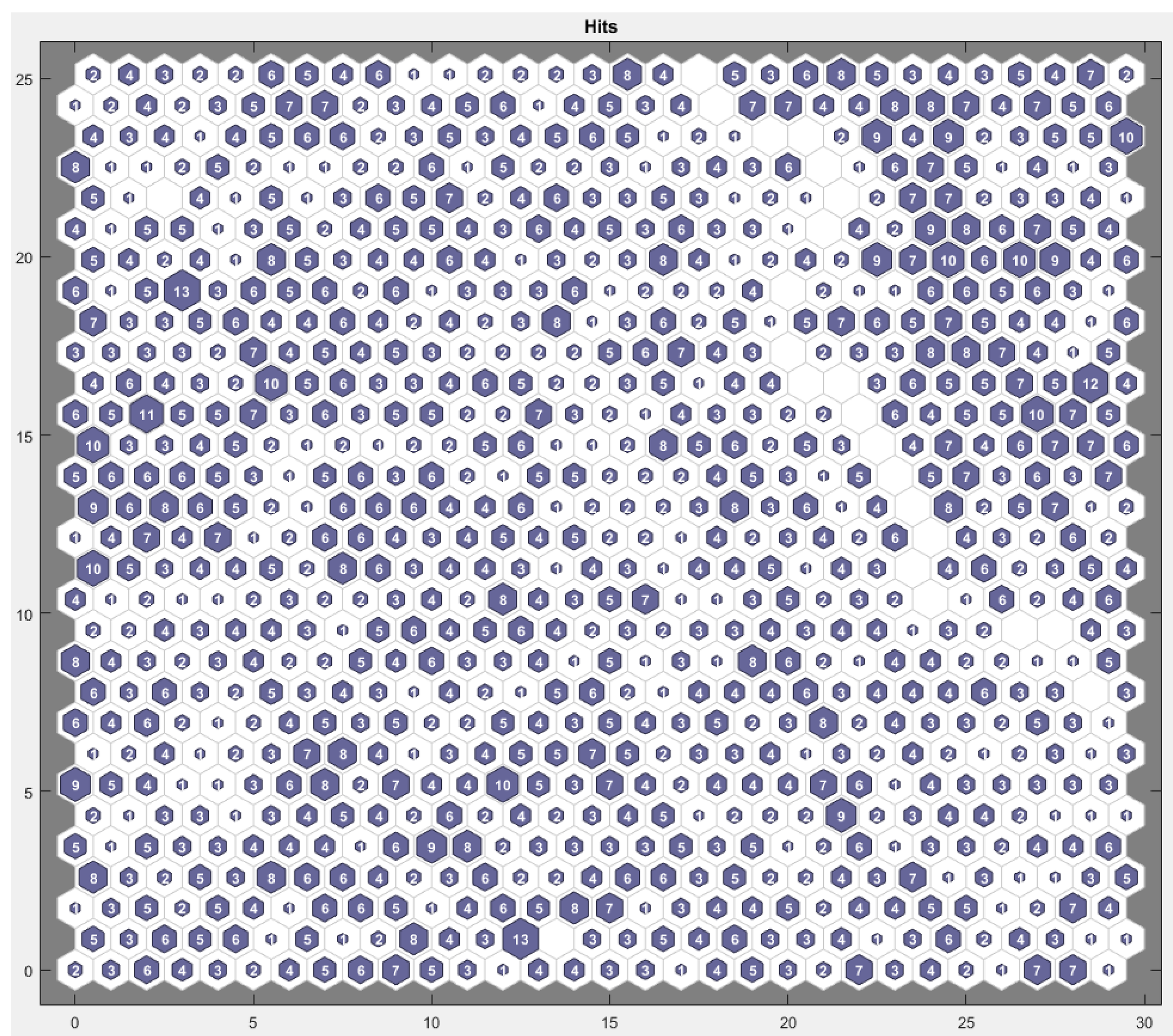


NDVI = (NIR – R)/(NIR + R)

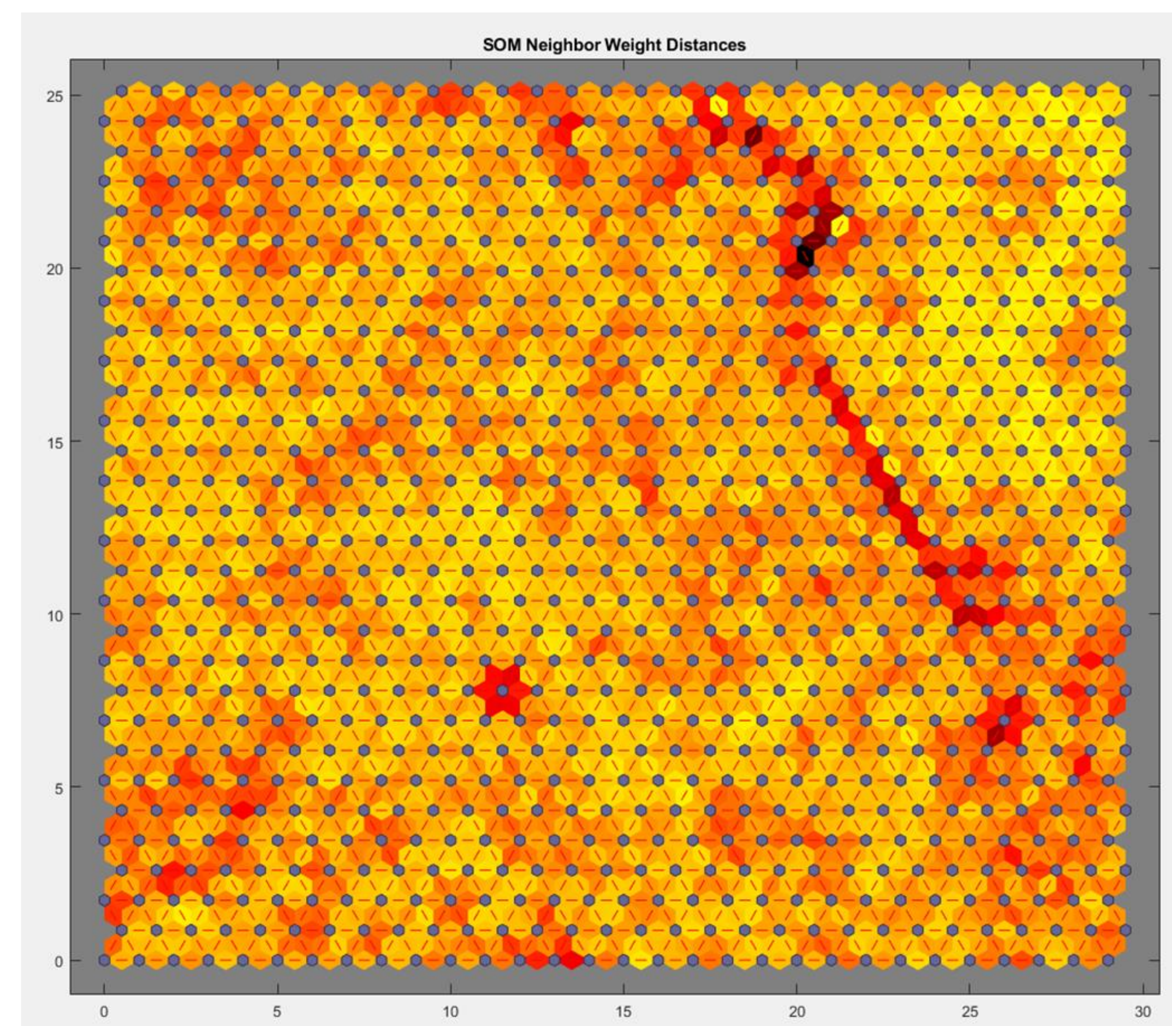
Filed location and Dataset

- North-East of Abernathy, Texas.
- Pivot irrigated and dry cotton fields.
- 2000 - 2003 yield and soil data.
- Landsat 5 & 7 satellite images with 30 meters resolution.

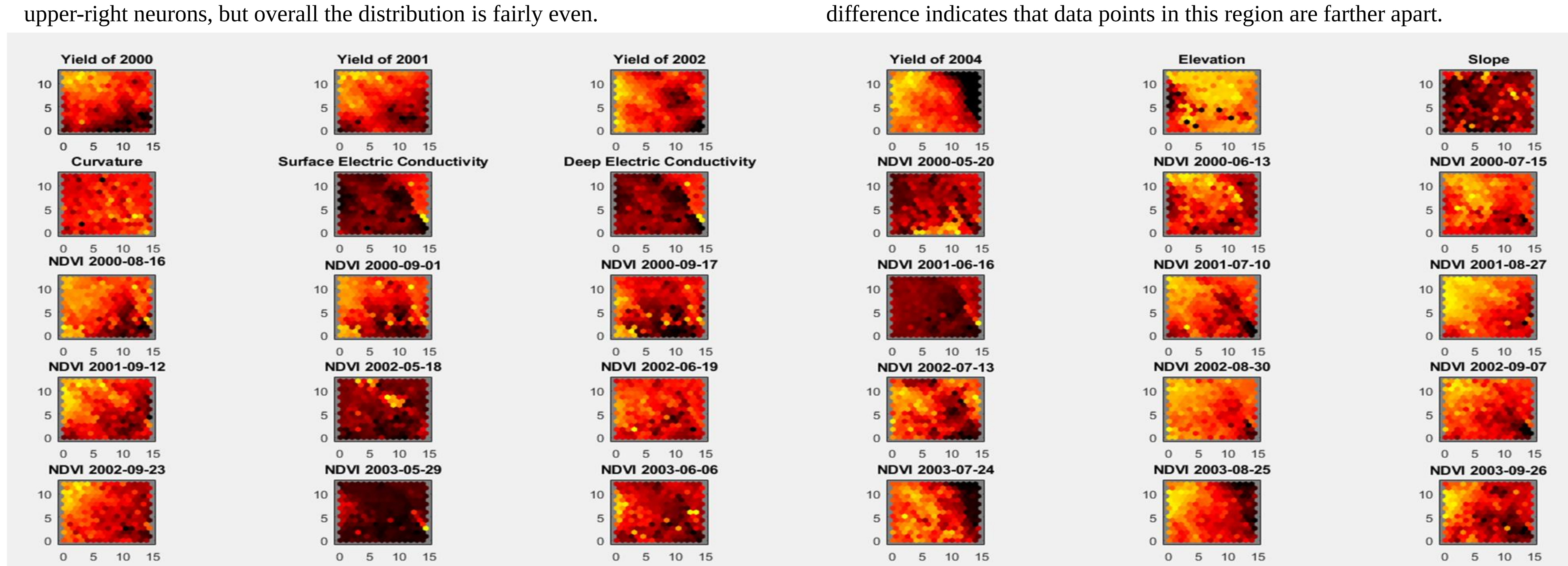
Results

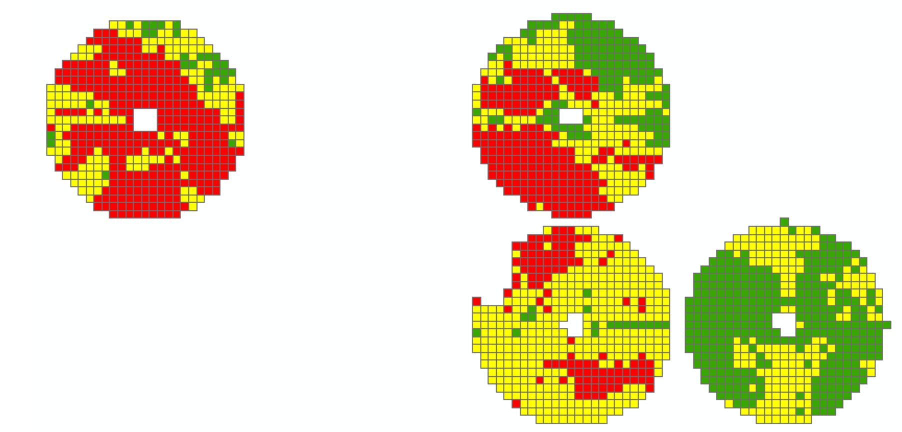


Self organizing map sample hits tell you how many data points are associated with each neuron. In this case, the data are concentrated a little more in the upper-right neurons, but overall the distribution is fairly even.

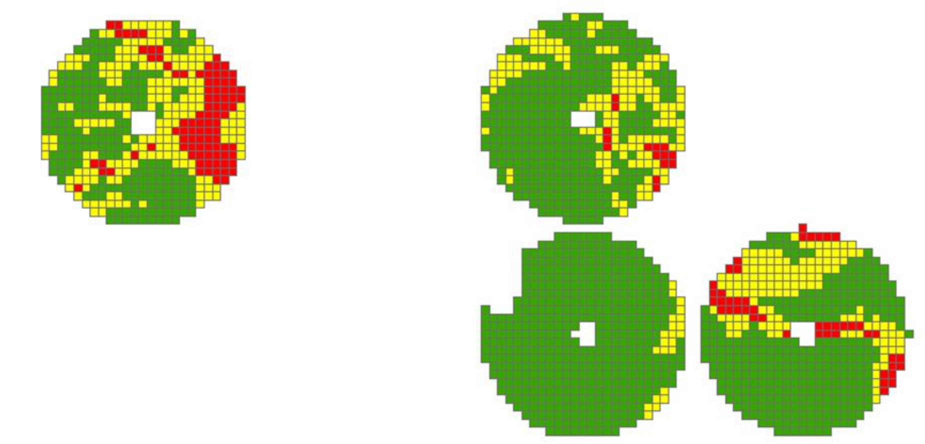


Neighbor Distances: blue hexagons represent the neurons; red lines connect neighboring neurons; dark red color represents larger distances. This color difference indicates that data points in this region are farther apart.

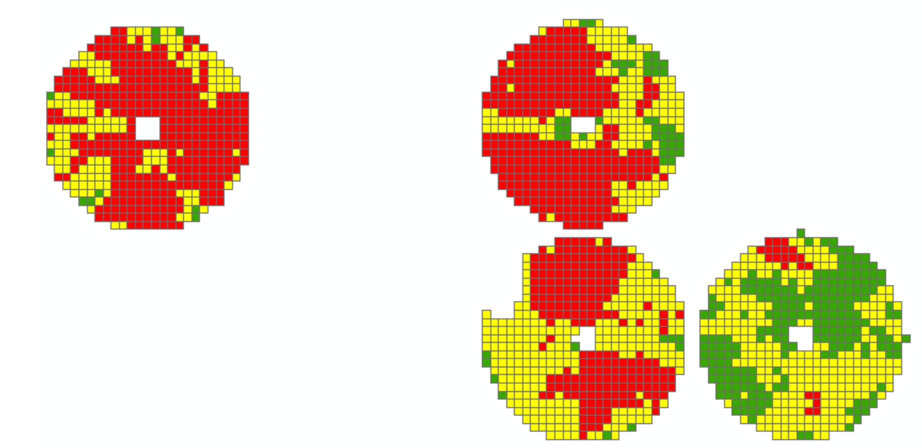




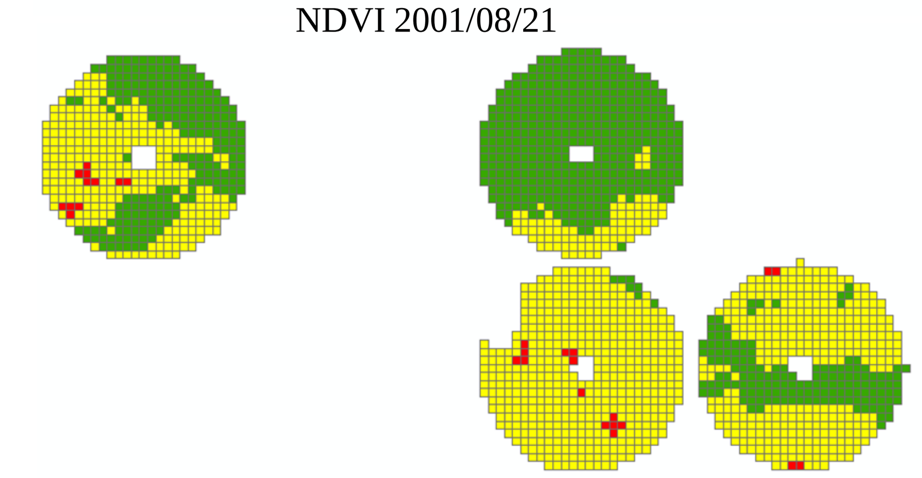
Average Yield (lb/Acre)



Slope



NDVI 2001/08/21



Surface Electric Conductivity

SOM Weight Planes are shown on the left: Each figure represents a parameter. They are visualizations of the weights that connect each input to each of the neurons. Lighter and darker colors represent larger and smaller weights, respectively. If the connection patterns of two inputs are very similar, you can assume that the inputs were highly correlated. In this case, NDVI inputs from July to September of each year are very similar with yield inputs.
The above four figures shows the real data value in each pixel for different parameters. Red color means relative high value and green color means relative low value. The overall trend of NDVI is similar with average yield, which also means NDVI has close correlation with yield prediction.
So the NDVI is the most critical parameter.