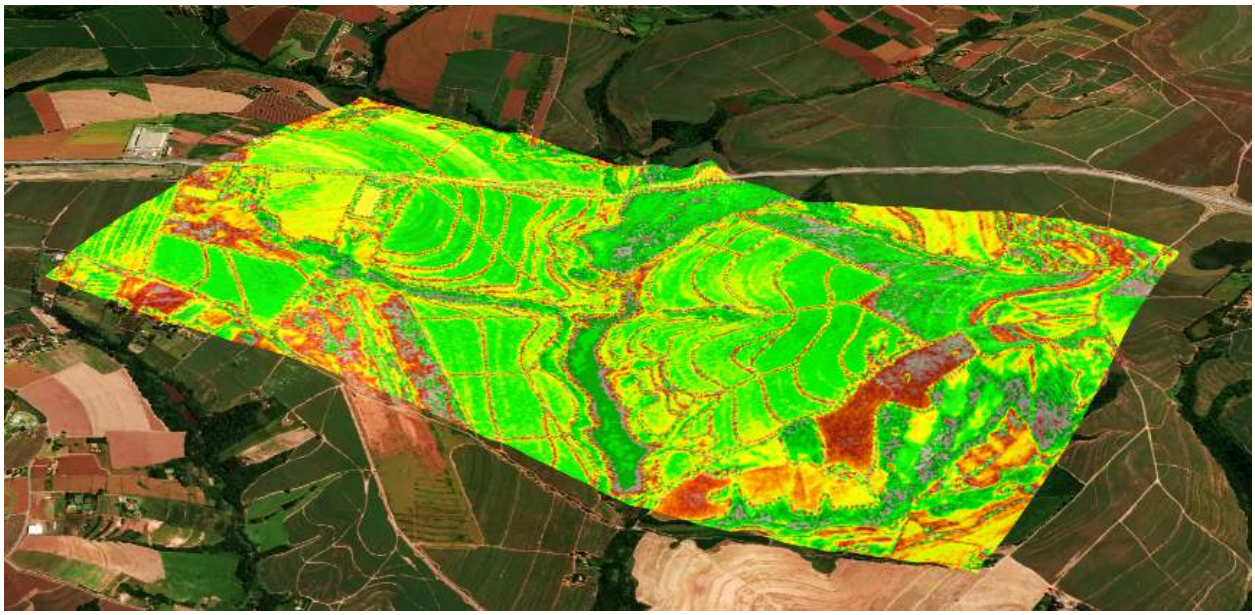
SC SOLUTIONS

Client Report

Vegetation analysis and yield prediction
HUNNICUTT FARMS



Project Deliverable Statement

SC Solutions aims to develop a comprehensive yield prediction model that intricately analyzes influential factors affecting crop production. Leveraging Earth Observation (EO) data combined with the agricultural insights of HUNNICUTT FARMS, the primary objective is to provide an in-depth analysis of expected crop yields, thereby equipping HUNNICUTT FARMS with the predictive capabilities essential for optimal resource allocation and effective harvest planning. Integral to this process is the continuous monitoring of vegetation health, ensuring any signs of stress, pest infestations, or suboptimal growth conditions are promptly addressed, thereby maximizing crop health and output.

In this first phase, we have used location and yield data from HUNNICUTT FARMS to construct a primary model to analyze the relationship between vegetation volumes and prediction. With a follow-up discussion, we are expecting to go into the second phase towards analysis with higher precision.

Objectives of the project:

- **Crop Yield Analysis:** Utilizing both EO data and HUNNICUTT FARMS' agricultural insights, our primary goal is to provide a comprehensive analysis of expected crop yields. This will empower HUNNICUTT FARMS with predictive capabilities, aiding in better resource allocation and harvest planning.
- **Vegetation Health Monitoring:** Continuous monitoring of vegetation health is paramount. Through our tools and analytics, we aim to detect early signs of stress, pest infestations, or sub-optimal growth conditions. This proactive approach ensures timely interventions, maximizing crop health and output.
- **Visualization:** Beyond mere data, our goal is to present insights in an easily digestible manner. Through dynamic charts, maps, and other visual tools, HUNNICUTT FARMS can gain a clear, real-time understanding of their farm's condition, enabling informed decisions at every stage of the crop cycle.

Initial Result Discussion

As discussed in the first meeting, we have identified the following area of interest (AOI) as the initial input from HUNNICUTT, with data provided from 2012-2022.

Area of interest as identified (Google Map)

The current available data sources have allowed us to analyze the yield from 2015 to 2022, with more optical data available from 2017-2022.

Although data is provided, more knowledge is to be shared by HUNNICUTT in the next meeting regarding the consistency of datasets, e.g. sowing and harvesting time, as from the EO data each parcels of land may have inconsistent timing.

To prepare for a highly applicable yield prediction model for HUNNICUTT farms, we have been looking into the following parameters:

Parameter	Vegetation	Vegetation	Moisture	Chlorophyll	Leaf Area

	Volumetric index	index			
Description	Assessing and monitoring vegetation and biomass, by its volume	Assessing and monitoring vegetation and biomass, by its colour and light reflection	Monitoring drought stress, irrigation management, and overall plant water content	Assessing vegetation condition, crop disease detection, and nutrient status	Assessing general leaf areas.

Detailed analysis will be discussed in the following subsections.

Yield analysis model (Initial version)

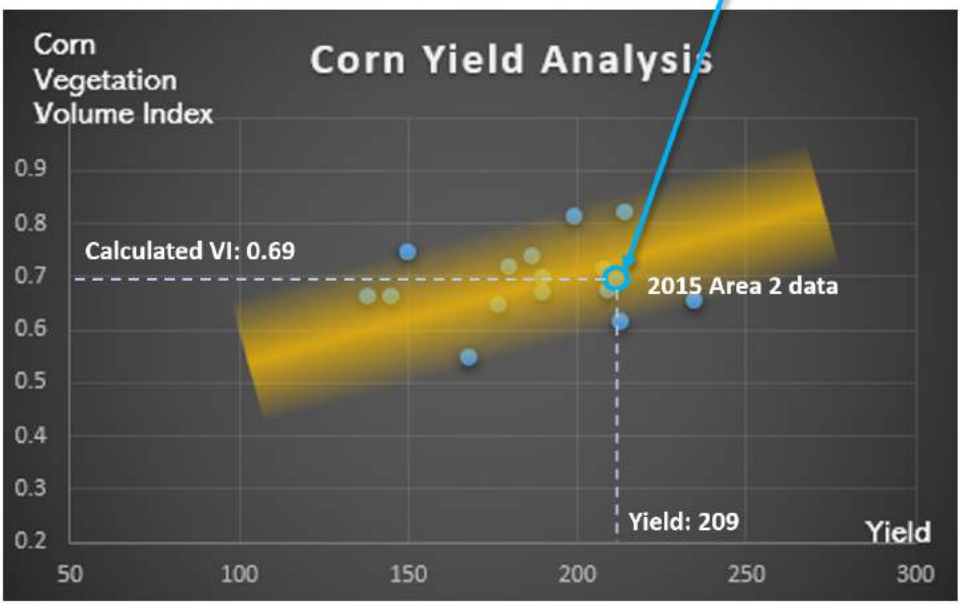
Previous research suggests that yield has positive correlation with **crop vegetation volume**, which has been leveraged as the main source of modeling.

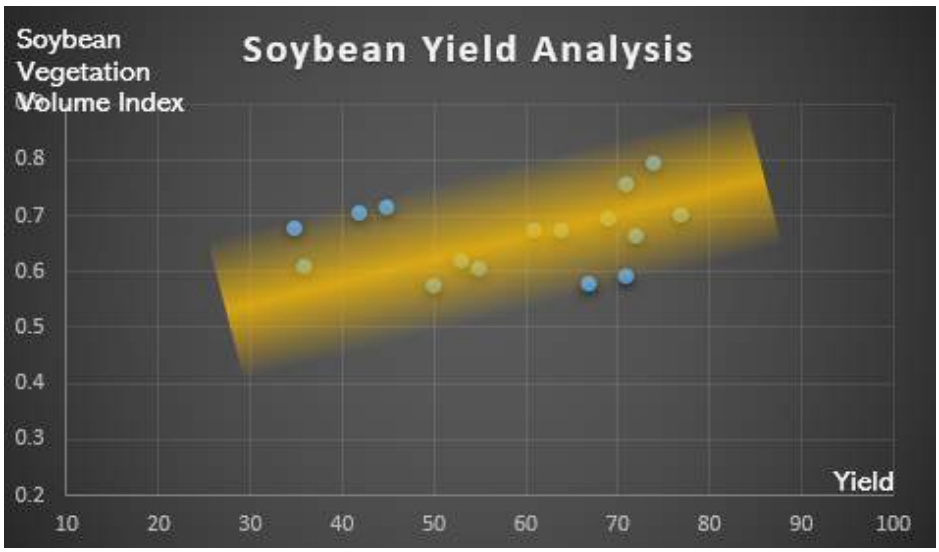
Initial evaluation has been carried out based on the **vegetation volume index**. *Plants with smaller size and thinner leaves **(evaluated from satellite imagery)** have small index value, and bigger plants have bigger index value, as is shown in the following figure.*

Small vs large volumetric index

We have put the yield data, with the analyzed vegetation index into a chart as follows.

	October	October	October	October	October
2022	190 corn	15 yellow peas	55 soybeans	150 corn	138 corn
2021	50 soybeans	63 popcorn	145 corn	64 soybeans	180 corn
2020	36 soybeans	46 yellow peas	177 corn	35 soybeans	53 soybeans
2019	67 soybeans	124 corn organic	201 corn 63 sb	61 soybeans	190 corn
2018	45 soybeans	6 sb organic	187 corn		71 soybeans
2017	79 popcorn	42 soybeans	235 corn	72 soybeans	69 soybeans
2016	77 soybeans	74 soybeans	212 corn	214 corn	199 corn
2015	168 corn	213 corn	71 soybeans	209 corn	208 corn
2014	74 soybeans	73 soybeans	225 corn	225 corn	76 soybeans
2013	242 corn	227 corn	221 corn	220 corn	81 popcorn





Soybean and corn yield analysis based on data provided

The horizontal axis of the graph depicts the yield data provided by HUNNICUTT Farms from the years 2015 to 2022, while the vertical axis showcases the Soybean Vegetation Volume Index.

Yellow line: Ideal fitting curve

Depicted in the graph is our preliminary predictive model. Here we adopted Linear Regression for constructing the fitting model. Ideally, the data points should align along the center of this yellow line, as that would represent the most accurate yield predictions. *However, it is evident that there are variations in the data, with points shifting towards the edges of the line.* This deviation indicates the influence of other factors such as moisture, leaf growth.

The fitting range, commonly referred to as the confidence interval, is typically set at **95%**. This means that when predicting new data, we can draw a line across the volume index, and we can say with 95% confidence that the predicted yields will fall within the yellow area of the horizontal line.

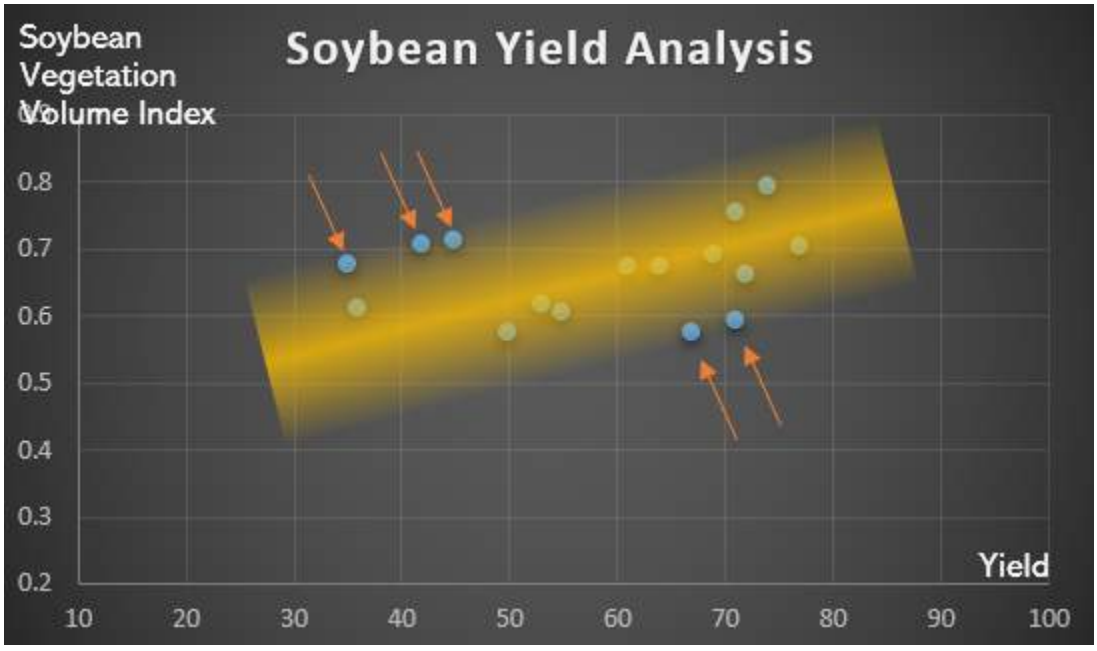
Currently, our interval is somewhat wide. This is due to our reliance on a simple model and a single index, in addition to the significant variation in our data. Moving forward, our next step will be to analyze the outliers to enhance our understanding. In this regard, we will need Hunnicutt's domain knowledge to determine if additional features are required to better train our model and reduce prediction errors.

Objectives of the next phase: Have a thorough understanding of the farm's corn field, and identify these factors that influence the shift for a better model.

Individual Case Study

Here we will focus on a thorough analysis of the outliers identified in the predictive model. Our primary objective is to scrutinize the reasons behind these data points falling outside the expected fitting range, the so-called **'outliers'** in the next figure.

By meticulously examining each data outlier, we aim to uncover the underlying factors or anomalies that may have contributed to these deviations, thereby providing a clearer understanding of the disparities observed in the model's predictions.



Data Outliers

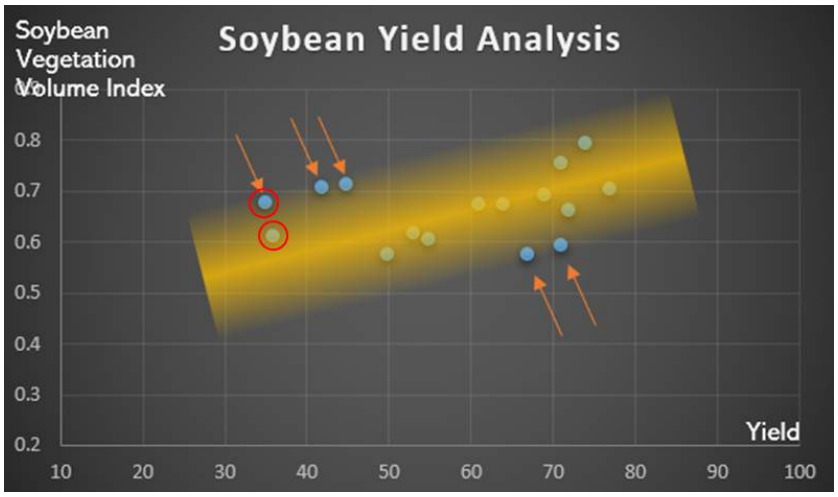
While most of the cases are within an acceptable error range of ML, the outliers are something we could pay attention to to improve the model.

We currently study three outliers.

Outlier 1:

Year: **2020**

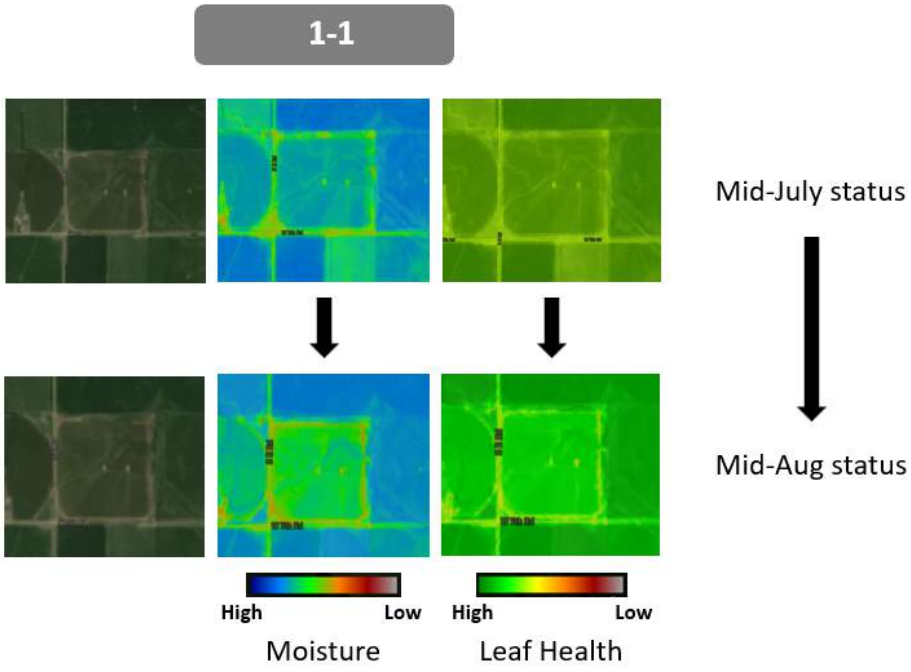
2020	Area 1-1 36 soybeans	Area 1-2 46 yellow peas	Area 1-3 177 corn	Area 2 35 soybeans	Area 3 53 soybeans



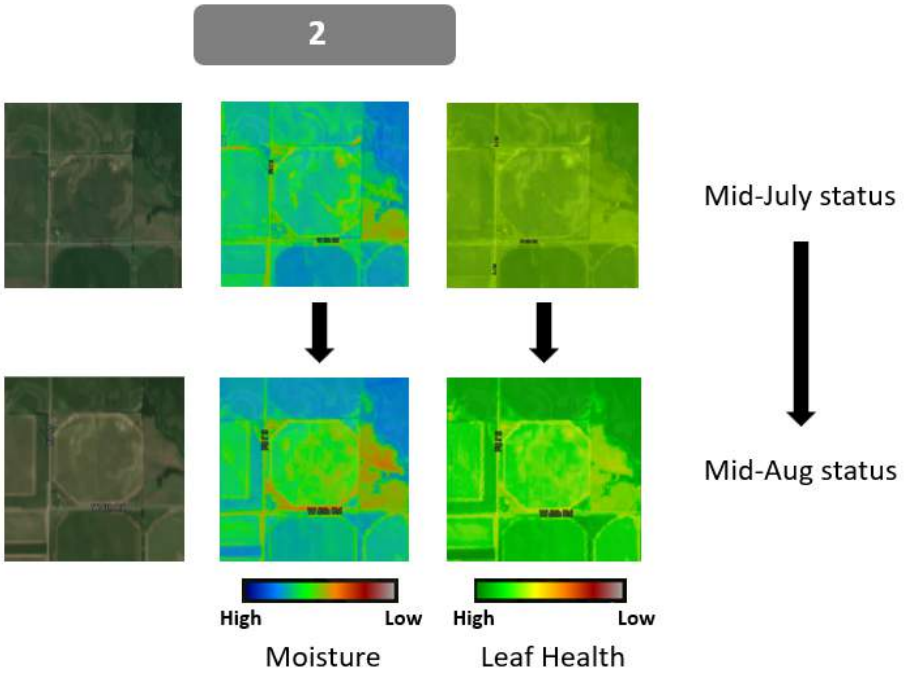
The above data indicates that the soybean plants in Area 2 are in general bigger and 'perceptually healthier' than Area 1-1. However confirmation is required from the HUNNICUTT part.

It's worth studying Area 1-1 and Area 2 in 2020, as they have potentially low yields of soybean compared with other fields provided.

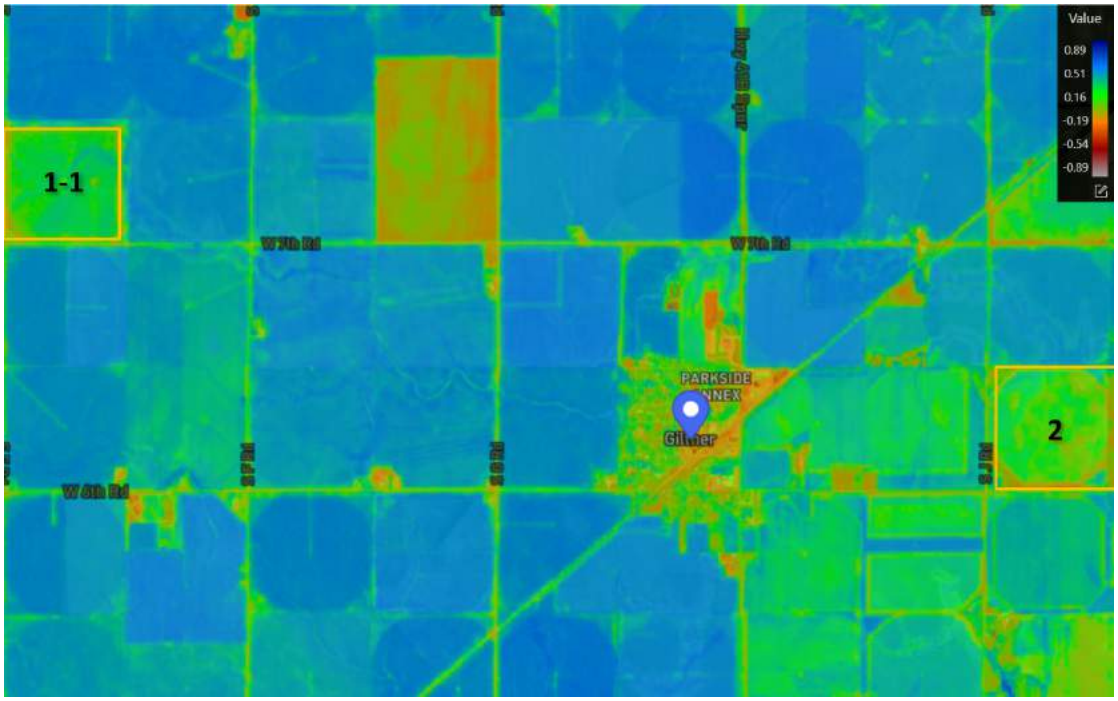
We analyzed the period potentially at V10 - R1 stage, (July - Aug, needs confirmation from HUNNICUTT). The initial results was that the moisture content was very low compared with surrounding land parcels for both 1-1 and 2. Area 2 experienced more severe drought.



Area 1-1 in July - Aug 2020



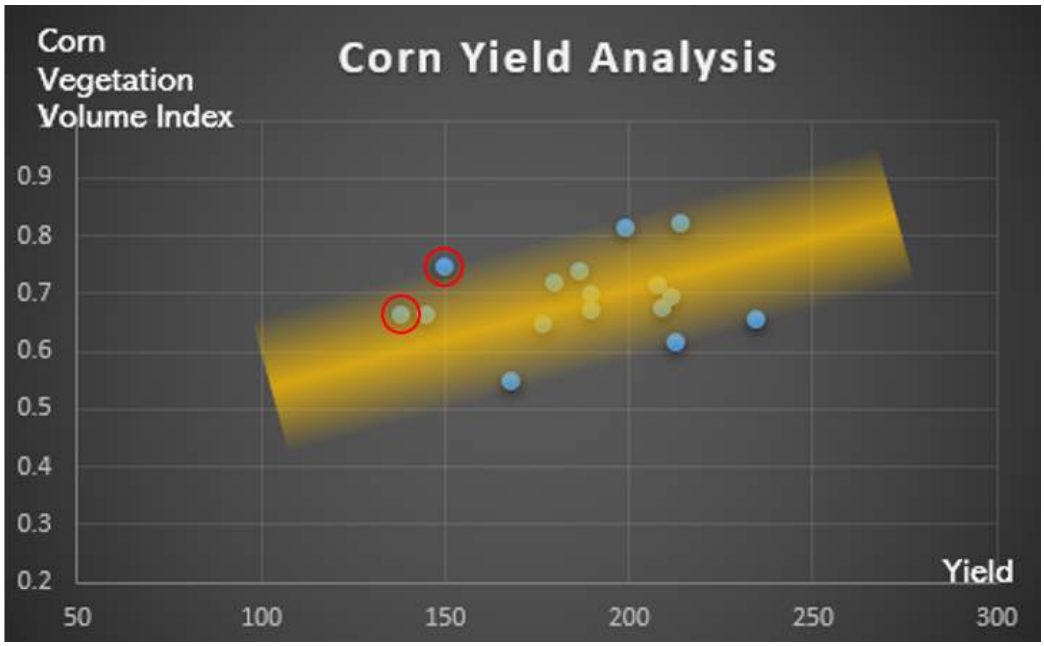
Area 2 in July - Aug 2020



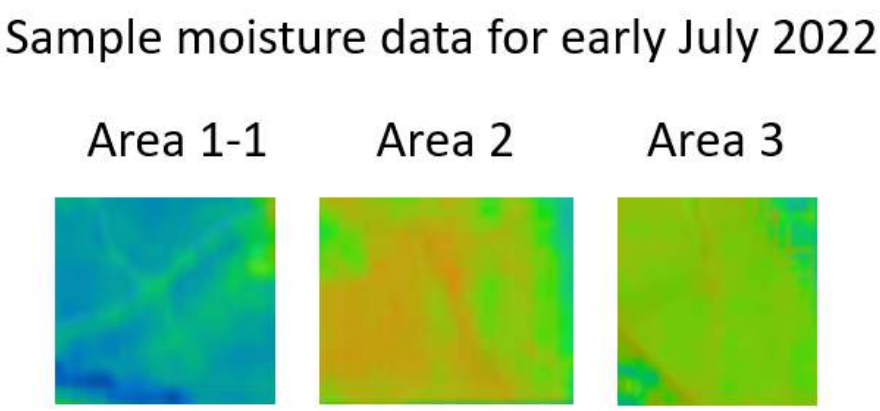
Average moisture data from July - August in area 1-1 and 2, compared with the region

Outlier 2:

2022	Area 1-1 190 corn	Area 1-2 15 yellow peas	Area 1-3 55 soybeans	Area 2 150 corn	Area 3 138 corn



Likewise, we started the moisture data analysis from June to August. However we have only identified that the moisture content from early July of Area 2 and 3 went significantly low for a week.



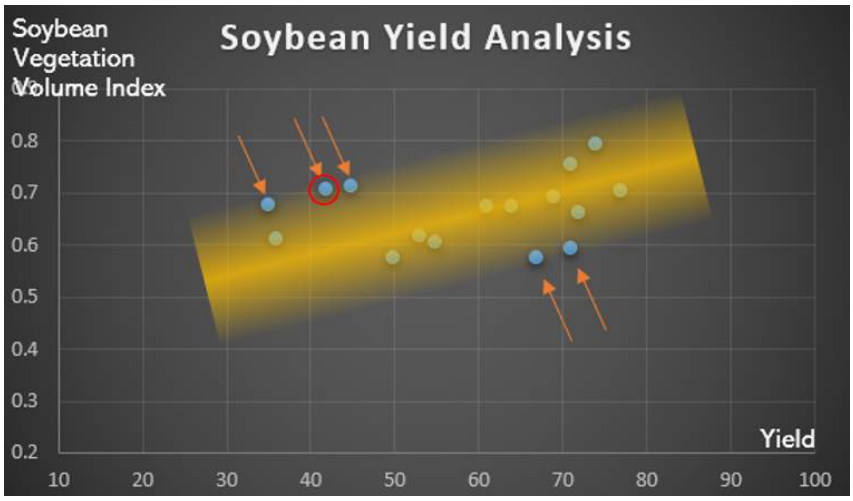
This might have resulted in a lack of chlorophyll in September before harvesting period.

Chlorophyll data for corn in September 2022

Further discussion and confirmation from HUNNICUTT regarding the growth in 2022 for better modeling.

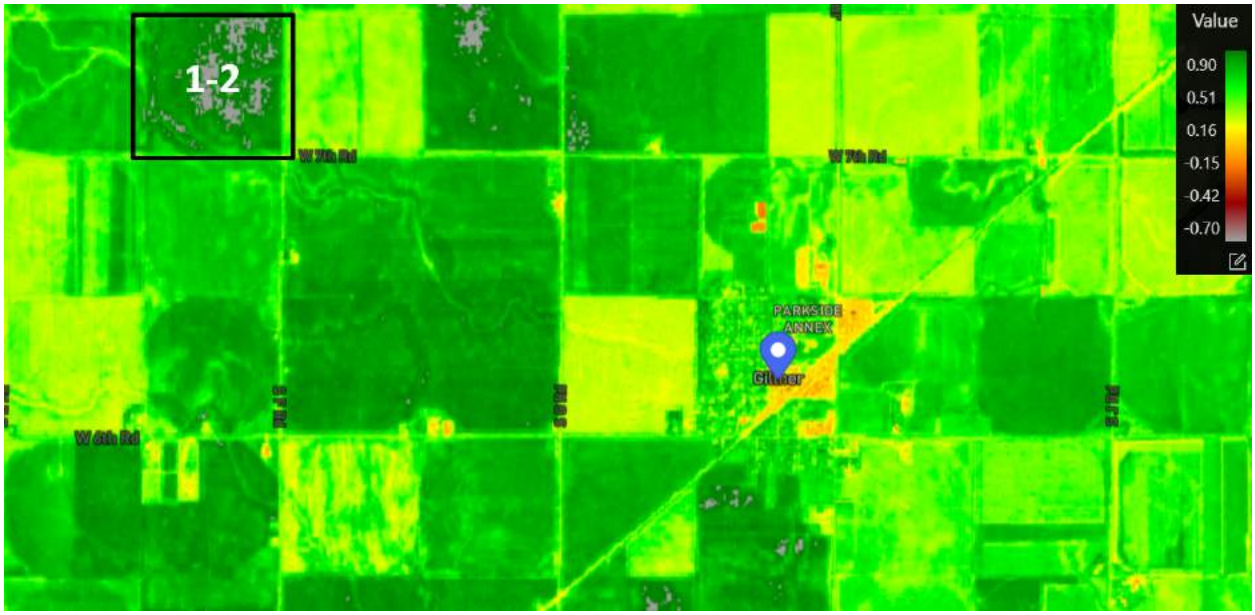
Outlier 3:

2017	Area 1-1 79 popcorn	Area 1-2 42 soybeans	Area 1-3 235 corn	Area 2 72 soybeans	Area 3 69 soybeans



Through a thorough check on moisture and other factors, we have not yet identified any reason related to drought or nutrient deficiency for the low yield.

However as is indicated in the following figure which indicates the leaf greenness and leaf area, Area 1-2 might experienced leaf overgrowth. Further discussion is required.



Visualization of parameters on platform

Visualization of the results are to be developed in the next step.

Plans for the next step

Meeting and discussion

In the next phase of our project, we plan to engage in a comprehensive discussion with HUNNICUTT to confirm our findings related to crop growth and ensure that our observations align with their on-ground experience and data. This step is crucial in validating the accuracy and relevance of our analysis.

Further data requirement and enhancement

Additionally, we aim to define and identify a few more Areas of Interest (AOIs) to gather more data and enhance our predictive model's accuracy. By broadening our data scope and incorporating more relevant variables, we hope to create a more robust and reliable prediction model that can better serve HUNNICUTT's needs and contribute to the optimization of their crop yield and overall agricultural productivity.