# Background and motivation

Accurate, early estimation of crop yield is an important phase in the decision-making processes and it is key for managing agricultural fields during last stages of the plant growth. There are many methods available for farmers to estimate yield of various crop; from the more traditional and simpler ones based on own farmer’s experience and believes, to the more modern and complicated using state of the art remote sensing technologies. Accuracy of yield estimates depends upon an adequate approach to get the more realistic measurements in the most effective way.

This research project proposes a method to estimate yield of large cotton fields using airborne imagery. It will involve all the steps of the entire crop yield estimation problem: from 2D images processing and data extraction, to the final yield prediction.

The main goal of this project is to use normal ML methods and techniques to predict cotton yield—number of cotton bolls and total weight—from 2D images taken using UAVs. With the completion of the project we will obtain a robust and reliable ML model for fast estimation of cotton yield using aerial images.

The main topics covered by the proposed project will include:

* + Image processing and information extraction using the OpenCV library.
  + Dataset creation in Python using pandas and scikit learn libraries.
  + SVM model training, testing and validation.
  + Linear regression fitting for crop yield forecasting.

# Approach

A custom dataset will be created from several individual plot images (Figure 1. b.), and a set of manually labeled images (Figure 1. c.). Individual plots are extracted from a general orthomosaic view of the entire field (Figure 1. a.), and labeled images are obtained manually annotating and classifying each pixel as cotton and non-cotton pixels. The dataset will include a row for each image pixel, 9 features for each of these pixels (color channels extracted from the color spaces RGB, HSV, and CIE L\*a\*b\*), and a binary class (cotton: {‘yes’, ‘no’}).

|  |  |
| --- | --- |
| a. | |
| Z:\HTPProjects\Data\2020\2D analysis\plots\Individual_plots\plotid_101.tif b. | Z:\HTPProjects\Data\2020\2D analysis\labelled data\label\mask_Label_1.png c. |

Figure 1. Drone imagery for cotton yield estimation: a. Orthomosaic image of the entire field; b. individual plot extracted from the orthomosaic image; and c. labeled mask identifying cotton bolls.

A SVM model will be trained and tested using the created dataset to classify input pixels into one of the classes (cotton, or non-cotton). Hence, once each image pixel is classified, we will have the number of cotton pixels per image/plot. Additionally, the cotton of 26 plots was manually harvested, counted and weighted for ground truthing, so we will be able to fit a polynomial model that relates cotton pixel counting from 2D images with real cotton boll counting. The input data to validate the model will consist on several unseen images from the same field. These images will be analyzed by the SVM model to predict the number of cotton pixels present in each image, and then to calculate the cotton yield estimation for these specific plot.