FERNANDO GBUR DOS SANTOS

Automation of Drone Data Extraction Protocol and ML Regression Algorithm

Summary of 3-month internship activities

GOALS

1. **AUTOMATE AND OPTIMIZE THE PROCESS OF EXTRACTING DATA FROM DRONE IMAGES**
2. **APPLY MACHINE LEARNING REGRESSION TO THE DRONE DATA**

- Efficient protocol, but time-consuming and with many manual processes
- Minimize manual inputs
- Simplify the process of extracting different rasters (drone images)
- Use the data acquired as inputs in the ML model to identify correlation for different measurements

A Few Important Terms

- **GIS(geographic information systems): “**GIS is a technology that is used to create, manage, analyze, and map all types of data”. Used in Data Management, Mapping and Visualization, Spatial Analysis, Communication, etc.

- **Rasters:** Image-based files. The data they contain is based on the images. The type of file that is generated with the drone images.

- **Vectors:** Spatial files: points on a map, lines, or polygons. Data contained within them is manually entered or extracted from rasters (such as plot numbers, or the mean value of a raster within the vector).

How does the protocol work?

Goal: Systematically and algorithmically identify and isolate field plots from image data and save them as data-containing vectors for analysis.

1 STEP

Import the orthomosaic to QGIS and create the vector file around the area of interest

2 STEP

Create a polarized layer to separate plots from the ground

Using A DEM(Digital Elevation Map), Vegetation Indices, etc...

3 STEP

Convert the polarized layer into new vectors.

Using the function “polygonization” of plugin GDAL

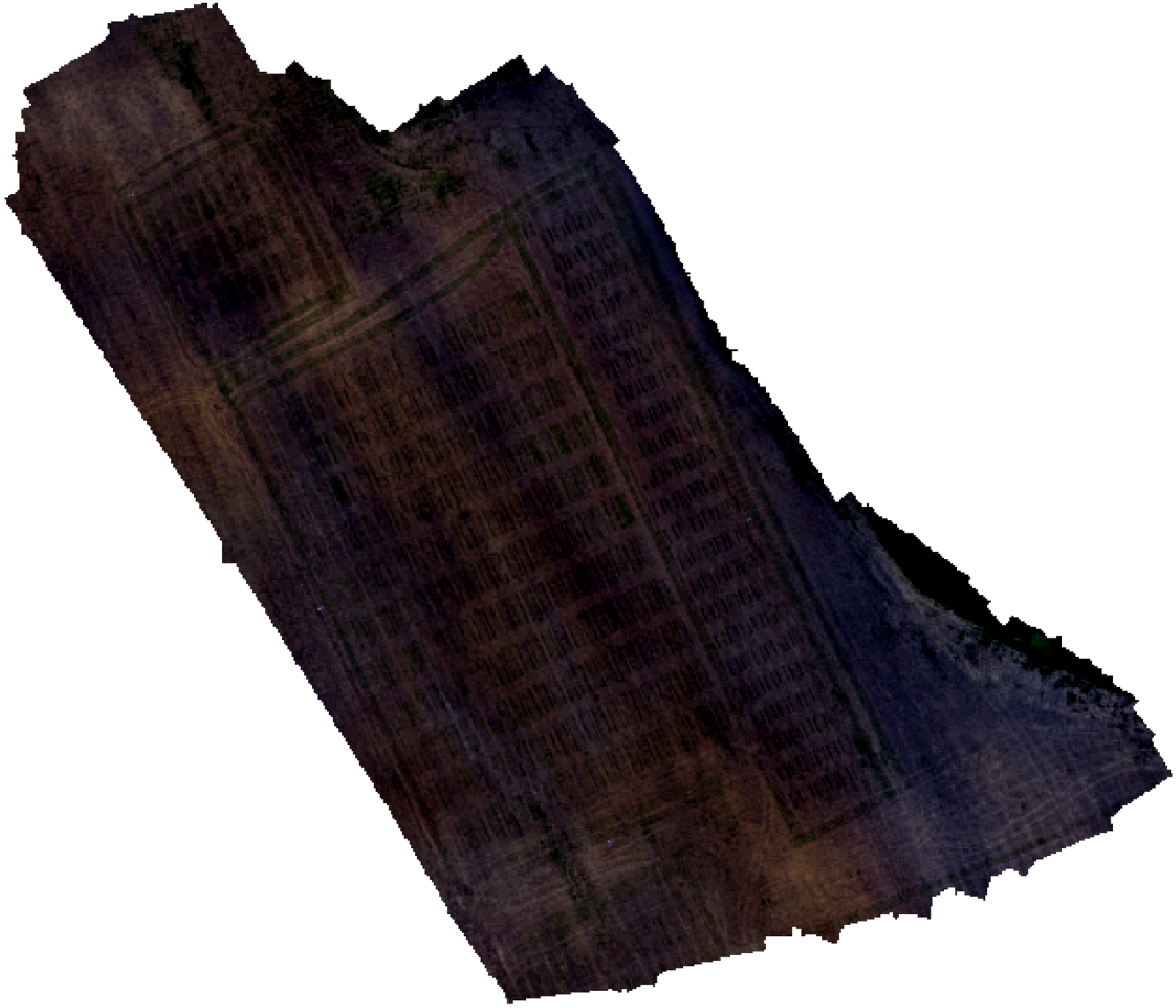
4 STEP

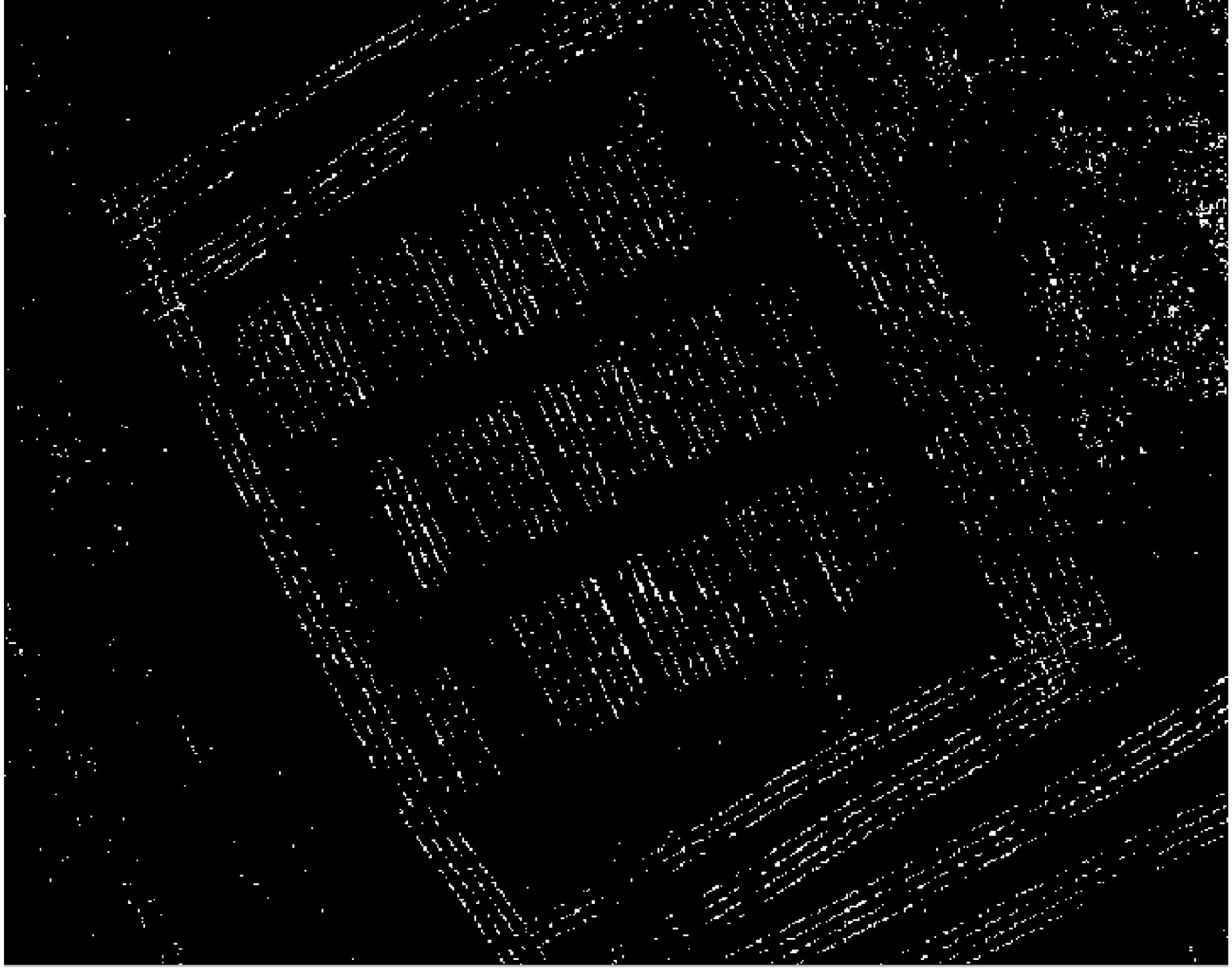
Generate a grid to analyze each individual plot and apply it to the new vectors generated

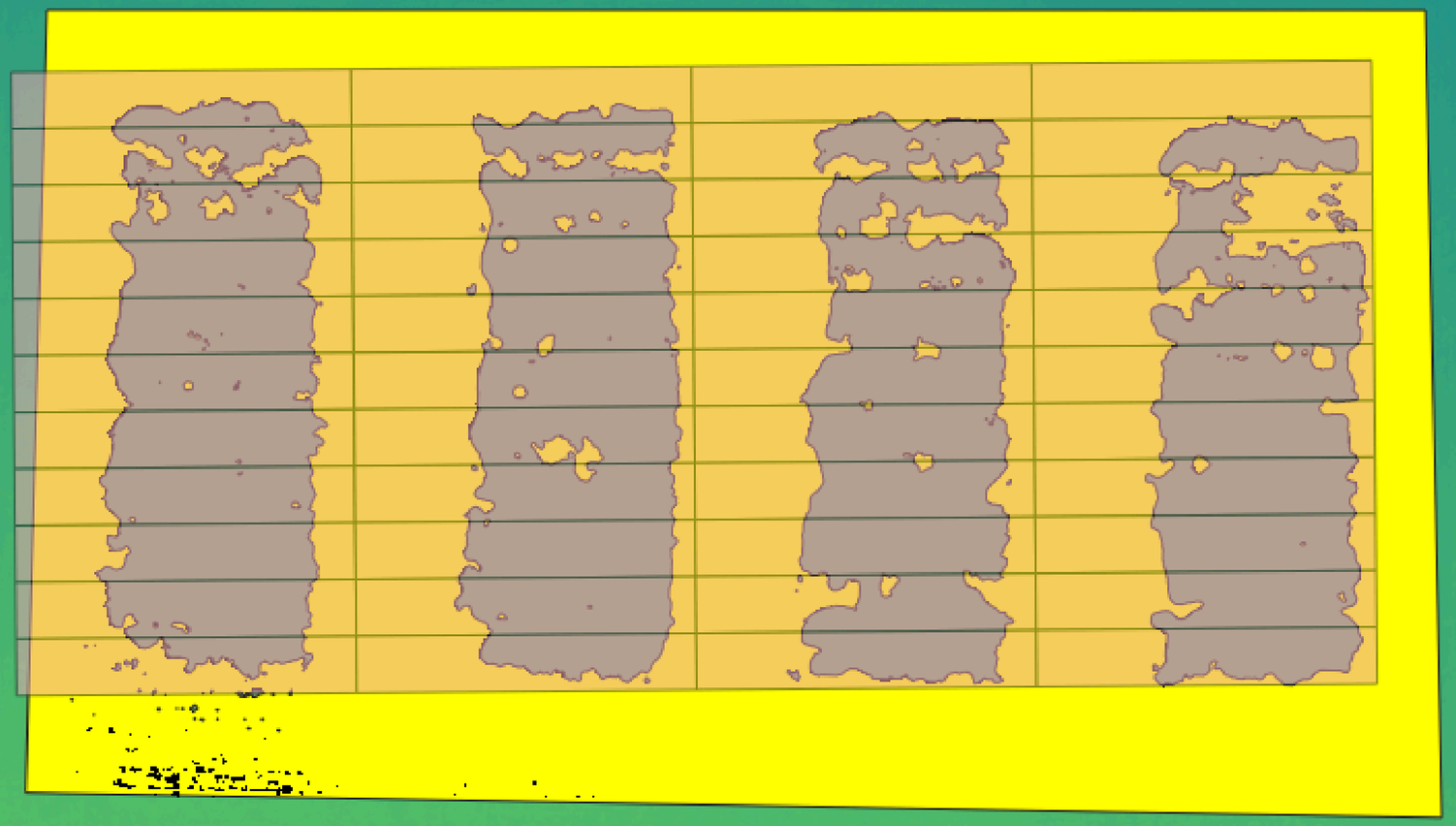
5 STEP

Import information from the raster to the new vectors and then analyze individual plot information

As mean, min, max, standart deviation, etc for each individual plot








301
302
303
304
305
306
307
308
309
201
202
203
204
205
206
207
208
209
101
102
103
104
105
106
107
108
109

How To Automate It?

- Make it scalable to multiple rasters
- What do we need as inputs? Rasters, Vectors and a polarization parameter
- Python
- GDAL plugin/PyGIS/JSON
- Polygonization Process

Getting all the raster files

- To store all the raster files, I decided to create a helper script that creates a JSON file that acts like a "list of rasters", where each entry contains the path to the raster file and the path to the corresponding vector.

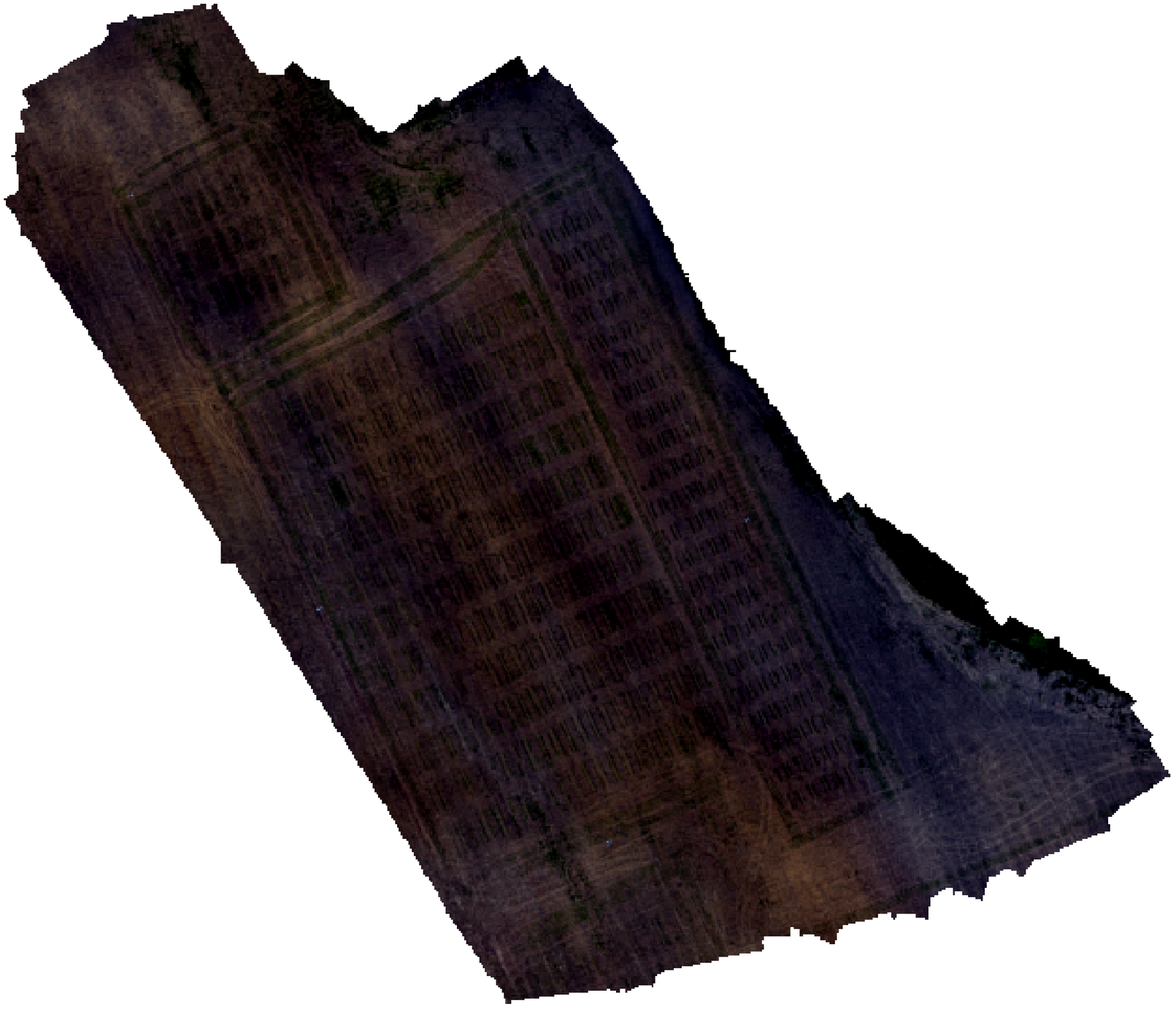
- It was necessary to create this type of configuration because each raster uses a different vector ("most of the time").

```json
[
    {
        "raster_path": "/home/gbur/Documentos/polygonization-script/raster_og/chickpeagrowout ortho 1.tif",
        "vector": "/home/gbur/Documentos/polygonization-script/vector-files/chickpea-otho-1-shape.shp"
    },
    {
        "raster_path": "/home/gbur/Documentos/polygonization-script/raster_og/6-06 upper ortho.tif",
        "vector": "/home/gbur/Documentos/polygonization-script/vector-files/chickpea-upper-new.shp"
    }
]
```

```
def main():

    with open(json_file, 'r') as file:
        data = json.load(file)

    # Opens the directory and saves its files in a list
    for item in data:
        raster_file = Path(item['raster_path'])
        vector_path = item['vector']
        if raster_file.name.endswith(".tif"):
            # Just change this line function name to use a different VI (NDVIRaster, GNDVIRaster or SAVIRaster)
            polarized_name, opt = NDVIRaster(raster_file)
            print("raster polarized")
            clipRaster(polarized_name, vector_path)
            print("raster clipped")
            polygonizeRaster(polarized_name, opt)
            print("raster polygonized")
            os.remove(polarized_name)
```



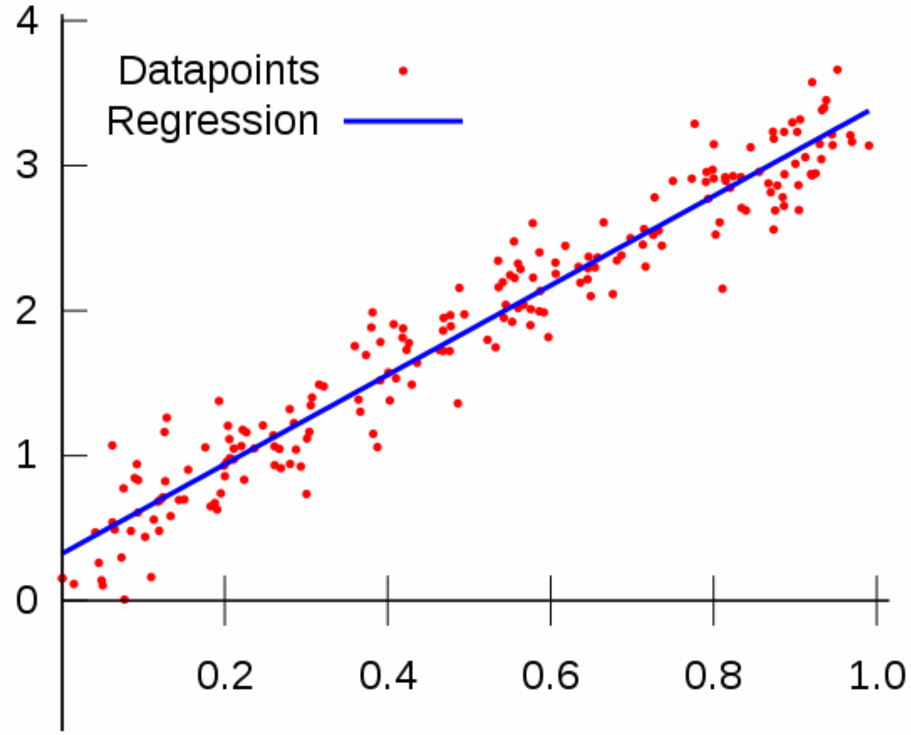
[illegible]

Script Advantages

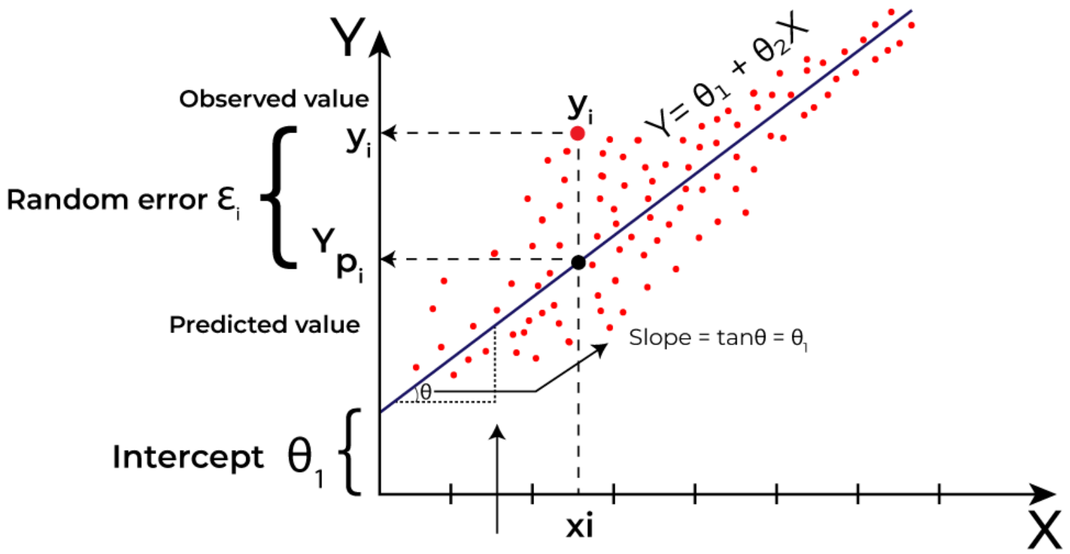
- Works with more than 1 raster at a time
- Works with any raster file of the field and vector for trials
- Work with multiple Vegetative Indexes and allows easy creation of functions that create polarized layers using a raster calculator (various vegetation indices, DEMs, etc...)
- Minimizes manual inputs during protocol execution

Machine Learning Regression

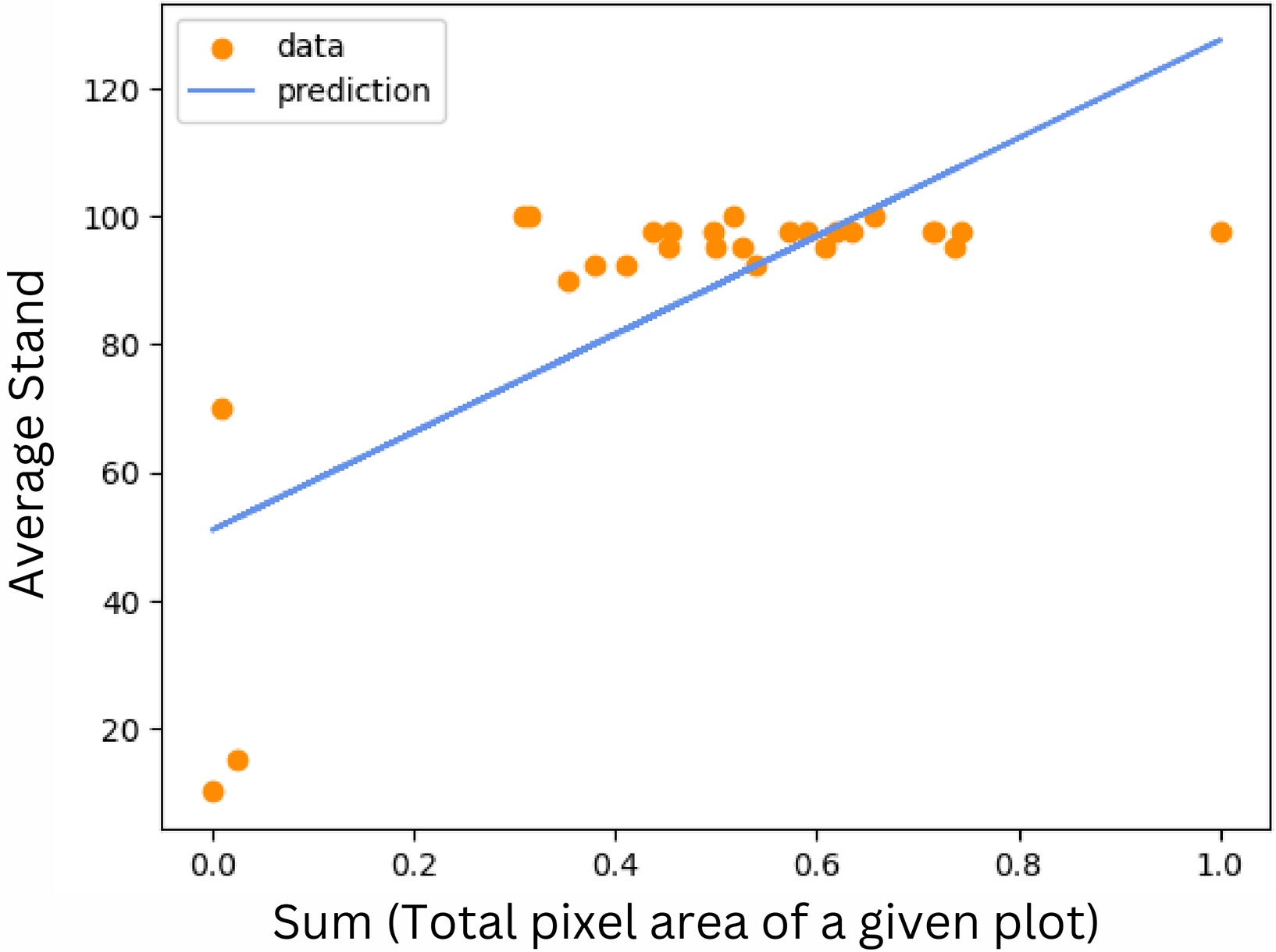
“It is a supervised machine learning technique, used to predict the value of the dependent variable for new, unseen data. It models the relationship between the input features and the target variable, allowing for the estimation or prediction of numerical values” (GeeksForGeeks)



Linear Regression



- It assumes that there is a linear relationship between the independent and dependent variables.
- $y=\beta 0+\beta 1X$
- Best-fit line (find best values for β0 and β1)
- Loss function
- Gradient descent



training score: 0.53 test score: -32.0 training score: 0.60 test score: -28.0

Other Activities

- Drone flights with Mason
- Bean evaluation with Mason
- Hand planting with Rob and Isabella
- Helped Monica in LODS
- Compute Canada

Thank You!

McGill
UNIVERSITY

Fernando Gbur dos Santos

@_gbur