

# Action 2019-3-27

## Development of Downstream Applications supporting Sectoral Information System under Copernicus Climate Change Service

---

### Pilot Activity Summary Report

Dr. Pablo Marzialetti

24.03.2023

## 1. Introduction

Agriculture plays a vital role in the economy and currently various factors affect the production rate of the different crops. These factors may be rainfall, soil types, nutrients, or fertilizer, among others. And there are many serious problems that farmers are facing today as erosion, diseases, pests, weeds, drought, rainfall (Gahoi A., 2019). Abiotic stresses like shortage of water or nutrients; or by biotic stresses like insects, fungi, bacteria, or viruses are affecting the expected evolution of crops. In many cases, the exact cause of certain diseases can only be discerned through destructive sampling, where knowing the occurrence of symptoms in an environmental, spatial, and historical context helps to identify the causes (Hueni et al., 2018).

Data Mining is used in various fields of Agri-Analytics, for the Prediction of potential problems, the early detection of diseases, to optimize farmer activity results, to forecast crop water consumption, among others (Gahoi A., 2019).

Vegetation indices, based on remotely-sensed spectral reflectance in the near-infrared, red-edge and visible bands, have been widely used for monitoring vegetation cover and plant health general condition. Multi-temporal acquisition of remote sensing data facilitates the possibilities to monitor the plant health status evidencing the perturbations which affects the expected development. Regarding the electromagnetic spectrum, green vegetation exhibits strong absorption in the red range (around 670 nm); reflectance in this range,  $\rho_{Red}$ , is below 3–5 %. In the near infrared (NIR) range, green vegetation strongly reflects incident irradiation; reflectance in this region,  $\rho_{NIR}$ , reaches 40–60 %. Spectral vegetation indices were devised and used as indicators of temporal and spatial variations in vegetation structure and density (Verstrate et al. 1996). Plants experience stress if suboptimal growth conditions cause their plant physiological functions (e.g., light and dark reactions of photosynthesis) to decline from their physiological standard (Niinemets, 2010).

The main goal of the methodology, data and tools proposed and described below, is to support and facilitate the accessibility to agricultural research organizations and farmers, to the numerous capabilities offered by the Copernicus System to develop early warning procedures to detect potential unexpected anomalies in crops.

For that it was proposed the utilization of Copernicus Sentinel-2 constellation satellite data, which is publicly available via different providers, as CreoDIAS, Mundi Web Services, Onda DIAS, sobloo, Amazon, Google, Copernicus Open Access Hub, Planet, and Sentinel-Hub, among others.

The Sentinel-2 mission consists of two satellites developed to support vegetation, land cover, and environmental monitoring. The Sentinel-2A satellite was launched by ESA on June 23, 2015, and operates in a sun-synchronous orbit with a 10-day repeat cycle. A second identical satellite (Sentinel-2B) was launched March 7, 2017. Together they cover all Earth's land surfaces, large islands, and inland and coastal waters every five days (source USGS). The Sentinel-2 Multispectral (MSI) acquires 13

spectral bands ranging from Visible and Near-Infrared (VNIR) to Shortwave Infrared (SWIR) wavelengths along a 290-km orbital swath.

The number of bands offered by this constellation gives the possibility to extract several vegetation indices towards a better understanding of plant growth status. In the proposed methodology several vegetation indices have been included as:

- **SR** (Simple ratio): The band ratio parameter is defined as the difference in radiance values measured between the NIR and red bands (Rouse et al. 1973).
- **NDVI** (Normalized difference vegetative index): The NDVI is an indicator of the greenness of the biomass. It is a measure of healthy, green vegetation. This index combines information contained in two spectral bands, the red and NIR:  $NDVI = (NIR - red) / (NIR + red)$ . Is one of the most used indices that enables the assessment and monitoring of changes in canopy biophysical properties such as the vegetation fraction (VF), leaf area index (LAI), fraction of absorbed photosynthetically active radiation, and net primary production (Eitel et al. 2011).
- **GNDVI** (Green NDVI): This index combines information contained in two spectral bands, the green and NIR:  $GNDVI = (NIR - green) / (NIR + green)$  (Gitelson & Merzlyak 1998).
- **NDRE** (Red-edge NDVI): This index combines information contained in two spectral bands, the RedEdge and NIR:  $NDRE = (NIR - RedEdge) / (NIR + RedEdge)$  (Gitelson & Merzlyak 1994).
- **modNDRE** (Modified red-edge NDVI): This index combines information contained in the spectral bands blue, RedEdge and NIR:  $modNDRE = (NIR - RedEdge) / (NIR + RedEdge - (2 * blue))$  (Datt 1999).
- **EVI** (Enhanced vegetative index): EVI is calculated similarly to NDVI, it corrects for some distortions in the reflected light caused by the particles in the air as well as the ground cover below the vegetation. The EVI data product also does not become saturated as easily as the NDVI when viewing rainforests and other areas of the Earth with large amounts of chlorophyll (Huete et al. 1994; NASA Earth Observatory 2000). This index combines information contained in the spectral bands blue, red and NIR:  $EVI = (2.5 * (NIR - red)) / (NIR + (6 * red) - (7.5 * blue) + 1)$
- **EVI2** (Enhanced vegetative index 2): This index combines information contained in two spectral bands, the red and NIR:  $EVI2 = 2.5 * (NIR - Red) / [NIR + (2.4 * Red) + 1]$  (Huete et al. 1997).
- **PVR** (Photosynthetic vigour ratio): This index combines information from green and red bands.  $PVR = green / red$  (Metternicht 2003).
- **GCI** (Green chlorophyll index): This index combines information contained in two spectral bands, the green and NIR:  $GCI = (NIR / Green) - 1$  (Gitelson et al. 2003).
- **CIRE** (Red-edge chlorophyll index): This index combines information contained in two spectral bands, the RedEdge and NIR:  $CIRE = (NIR / Red-edge) - 1$  (Gitelson et al. 2003).
- **TriVI** (Triangular vegetative index): This index combines information contained in the spectral bands green, red and NIR:  $TVI = 0.5 * (120 * [NIR - Green] - 200 * [Red - Green])$  (Broge & Leblanc, 2001)
- **MTCI** (MERIS terrestrial chlorophyll): This index combines information contained in the spectral bands red, RedEdge and NIR:  $MTCI = (NIR - Red-edge) / (Red-edge - Red)$  (Dash & Curran 2004).
- **LCI** (Leaf Chlorophyll Index): This index combines information contained in the spectral bands red, RedEdge and NIR:  $LCI = (NIR - RedEdge) / (NIR + Red)$
- **TCVI** (Tasselled Cap - Green Vegetation Index MSS): This index combines information contained in the spectral bands green, red, RedEdge and NIR:  $TCVI = -0.283 * green - 0.660 * red + 0.577 * RedEdge + 0.388 * NIR$

- **GARI** (Green atmospherically resistant vegetation index): This index combines information contained in the spectral bands blue, green, RedEdge and NIR:  $GARI = \frac{NIR - (green - (blue - RedEdge))}{NIR - (green + (blue - RedEdge))}$
- **SIPI1** (Structure Intensive Pigment Index 1): This index combines information contained in the spectral bands blue, red and NIR:  $SIPI1 = \frac{NIR - blue}{NIR - red}$
- **SIPI2** (Structure Intensive Pigment Index 2): This index combines information contained in the spectral bands blue, RedEdge and NIR:  $SIPI2 = \frac{NIR - blue}{NIR - RedEdge}$
- **MCARI** (Modified Chlorophyll Absorption in Reflectance Index): The Modified Chlorophyll Absorption Ratio Index (MCARI) was designed to extract information from the chlorophyll content in plants with a resistance to the variation of the Leaf Area Index (Mouret et al., 2018). This index combines information contained in the spectral bands green, red and RedEdge:  $MCARI = ((RedEdge - red) - 0.2 * (RedEdge - green)) * (RedEdge / red)$
- **ARVI** (Atmospheric Resistant Vegetation Index): This index combines information contained in the spectral bands blue, red and NIR:  $ARVI = \frac{NIR - (2.0 * red - blue)}{NIR + (2.0 * red - blue)}$  (Kaufman & Tanre 1996)
- **OSAVI** (Optimized Soil Adjusted Vegetation Index): This index combines information contained in two spectral bands, the red and NIR:  $OSAVI = \frac{NIR - red}{NIR + red + 0.16}$  (Rondeaux et al., 1996)

Regarding the tools it was proposed the use of Python Notebook (<https://jupyter.org/>) which is a free web-based interactive interface that allows users to create, launch and share workflows in data science. And to analyze the data processed it was applied the Orange Data Mining platform (<https://orangedatamining.com/>) (Demsar et al., 2013), which is an open source machine learning and interactive data exploration and visualization tool.

And regarding data, it was selected the repository of Sentinel-2 datasets accessible by the Google Earth Engine (GEE) (<https://earthengine.google.com/>) (Gorelick et al., 2017), a planetary-scale platform for Earth Science data and analysis. GEE gives accessibility to a multi-petabyte catalogue of satellite imagery and ready to use datasets. GEE provides Sentinel-2 Level-2A orthorectified atmospherically corrected surface reflectance images available since 2017-03-28 to present. And Level-1C orthorectified top-of-atmosphere reflectance images since 2015-06-23 to present.

## 2. Consultation with Users

Regarding the use of satellite imagery in general and for the detection of crop anomalies in particular, the accessibility and manipulation of these kind of data was identified as a critical point that need to be managed (via specialized human resources not always part of the company).

A survey was conducted at the beginning of the current project where different conclusions can be extracted. The consultation underlined the willingness to:

- use new satellite based technologies even for those who are not currently using it
- to monitor different crop field sizes
- located mostly locally
- for at least in 3 months/year according the phenological status
- almost half of the them are not detecting anomalies via automated procedures, or applying early warning procedures neither.
- Regarding the Copernicus Climate Data store portfolio, more than half has limited knowledge regarding the available datasets.

During the initial survey different kind of anomalies in vegetation expected development have been identified of interests by the potential users, as those caused by pests, drought, or temperature.

In particular those related to pests presence several kinds were detailed by the users, but at the same time, it was not possible for them to collect accurate data, with precise period of time of occurrence and geolocation. The spatial distribution of some of these anomalies caused by the presence of pests, like weath aphid, is not necessarily homogenously distributed, so is required very accurate geolocation of each occurrence.

Several commercial platforms are already offering satellite based products to monitor continuously and widely agricultural regions. Despite that, the possibility to access autonomously to Copernicus portfolio of datasets, via the several DIAS partners, Amazon or Google is incrementing substantially the access and easy extraction of added value information with minimum hardware infrastructure investments.

Despite DIAS platforms, Amazon, and Sentinel HUB access platforms are known by several users, form the initial survey it was extracted that Google Earth Engine is the tools mostly used among the users who count with IT skills to manipulate satellite datasets.

*Table 1 List of potential users taking part in the consultation*

No	Institution	COUNTRY, Region	Institution profile
1	AZIENDA AGRICOLA MACCARESE S.p.A.	ITALY, Latium	AGRICULTURAL COMPANY
2	AGRICOLTURA NUOVA	ITALY, Latium	AGRICULTURAL COMPANY
3	COLDIRETTI	ITALY, Sardinia	AGRICULTURAL ASSOCIATION
4	CITTÀ METROPOLITANA ROMA CAPITALE	ITALY, Latium	GOVERNMENTAL
5	CONFEDERAZIONE ITALIANA AGRICOLTORI	ITALY, Veneto	AGRICULTURAL ASSOCIATION
6	APOFRUIT	ITALY, Latium	AGRICULTURAL COMPANY
7	UNIVERSITY OF CORODOBA	ARGENTINA, Cordoba	RESEARCH INSTITUTION
8	CONSULTANCY COMPANY	ARGENTINA, Cordoba	AGRICULTURAL SERVICES
9	AGRONOMIC CONSULTANT	ARGENTINA, Buenos Aires	AGRICULTURAL SERVICES

10	INTA	ARGENTINA, Buenos Aires	RESEARCH INSTITUTION
11	AGISENSE	POLAND, Lublin	AGRICULTURAL SERVICES
12	HODOWLA ZWIERZĄT ZARODOWYCH OSOWA SIEŃ SP. Z O.O	POLAND, Wielkopolskie voivodeship	AGRICULTURAL COMPANY
13	MŚCICE" OŚRODEK HODOWLI ZARODOWEJ SP. Z O.O	POLAND, Zachodniopomorskie	AGRICULTURAL COMPANY
14	OŚRODEK HODOWLI ZARODOWEJ DĘBOŁĘKA SP. Z O.O.	POLAND, Łódzkie	AGRICULTURAL COMPANY
15	OŚRODEK HODOWLI ZARODOWEJ GARZYN" SP. Z O.O	POLAND, Wielkopolskie	AGRICULTURAL COMPANY
16	OŚRODEK HODOWLI ZARODOWEJ W DROBNINIE	POLAND, Wielkopolska	AGRICULTURAL COMPANY
17	PRZEDSIĘBIORSTWO ROLNICZO-HODOWLANE GAŁOPOL" SP. Z O.O	POLAND, Wielkopolskie	AGRICULTURAL COMPANY
18	CESEFOR	SPAIN, Soria	FOREST SERVICES
19	KALRO	KENYA, Kisumu	RESEARCH INSTITUTION
20	UNIVERSITY OF NAIROBI	KENYA, Nairobi	RESEARCH INSTITUTION

### 3. Methodological Framework and Proposed Services

The proposed methodology is separated in two main steps:

- 1- Satellite data extraction
- 2- Data analysis

The first step regarding the Satellite Data Extraction, includes:

- a. the extraction of vegetation indices statistics of fields to train the models:
  - i. the extraction of vegetation indices statistics of fields where the occurrence of the anomaly is present.
  - ii. the extraction of vegetation indices statistics of fields where the occurrence of the anomaly is NOT present.
- b. the extraction of vegetation indices statistics of a field that want to be analyzed

While the second step regarding Data analysis, is based on the open source machine learning tool called Orange Data Mining. This tool is based on workflows which can be easily adapted for different data analysis purposes. For the current application, 3 ready-to-use workflows are uploaded in the repository:

- i. the Data Preparation Workflow,  
([https://github.com/silexclouds/AnomalyDetection/blob/main/Orange\\_Workflows/A\\_Workflow\\_dataPreparation.ows](https://github.com/silexclouds/AnomalyDetection/blob/main/Orange_Workflows/A_Workflow_dataPreparation.ows))
- ii. the Training Workflow,  
([https://github.com/silexclouds/AnomalyDetection/blob/main/Orange\\_Workflows/B\\_Workflow\\_training.ows](https://github.com/silexclouds/AnomalyDetection/blob/main/Orange_Workflows/B_Workflow_training.ows))
- iii. and the Data Prediction Workflow,  
([https://github.com/silexclouds/AnomalyDetection/blob/main/Orange\\_Workflows/C\\_Workflow\\_prediction.ows](https://github.com/silexclouds/AnomalyDetection/blob/main/Orange_Workflows/C_Workflow_prediction.ows))

The satellite data extraction module is accessible via an intuitive Python Notebook, which collects the time series of the previously detailed list of vegetation indices for a particular parcel or region of interest.

#### Extracting Sentinel-2 indices time series from Google Earth Engine

```
1 import geemap
2 import ee
3 import numpy
4 import eemont
5 import csv
6 import os
7 import io
8 import requests
9 import pandas
10 from osgeo import ogr
11
12 print('\nGoogle Earth Engine initialize..')
13 ee.Initialize()
14 print('Done.\n')
15
16 Map = geemap.Map(center=[41.8,12.5], zoom=2)
17 url = 'http://mt0.google.com/vt/lyrs=y&hl=en&x={x}&y={y}&z={z}'
18 Map.add_tile_layer(url, name='Google Map Hybrid', attribution='Google')
19 Map
```

An map interface is opened which gives the possibility to design the parcel or region of interest, required in the next step, where 4 mandatory parameters need to be defined by the user:



- **Path of fields of interest file**  
POLYGON\_PATH: ["DRAW" | shapefile path | geoson file path]  
Note: Choose "DRAW" to draw the polygon on the MAP ABOVE or write the path of a shapefile (.shp) or a .geojson file. If you chose "DRAW", draw the polygon/polygons on the map before moving forward.
- **Date range (initial and final dates of the analysis)**  
date1: ['YYYY-MM-DD'] & date2: ['YYYY-MM-DD']
- **Anomaly Presence**  
anomaly: ['true' | 'false' | 'topredict']  
*Note: choose 'true' or 'false' to create the datasets required for the model training ['true' for fields related to the presence of an anomaly (i.e. caused by pest) or 'false' in case are not related to any anomaly], Choose 'topredict' in case the fields need to be analyzed based on a already trained model.*
- **Path of output file with Vegetation Indices statistics**  
OUTPUT\_CSV: [path of output .csv files]



```

1 # INPUTS to be defined by the user
2
3 # Define Fields of Interest File path
4 # *** write "DRAW" if you want to draw the polygon on the MAP ABOVE
5 # *** or write the path of a SHP or GEOTIFF file
6 POLYGON_PATH= r'C:\AnomalyDetection\Datasets\ROIs\parcels.shp'
7
8 # SELECT THE INPUT DATES
9 date1='2018-01-01' # year-month-day FORMAT 'YYYY-MM-DD'
10 date2='2019-02-10' # year-month-day FORMAT 'YYYY-MM-DD'
11
12 # SELECT ANOMALY PRESENCE
13 anomaly='true' # 'true' or 'false' or 'topredict'
14
15 # Define CSV output directory (path with '/' at the end
16 OUTDIR CSV='C:/AnomalyDetection/Datasets/CSV/'
17
18 #-----#
19 print('SELECTED PARAMETERS\n')
20 print('  INPUT POLYGON_PATH: {}'.format(POLYGON_PATH) )
21 # if you chose "DRAW", draw the polygon/polygons on the previous map before moving forward
22 if POLYGON_PATH=="DRAW":
23     print('If you chose "DRAW", draw the polygon/polygons on the map before moving forward')
24     roiname='roi'
25 else:
26     roiname=os.path.splitext(os.path.basename(POLYGON_PATH))[0]
27 print('  ROI name : {}'.format(roiname))
28 print('  INPUT DATE RANGE: {} - {}'.format(date1,date2) )
29 print('  ANOMALY: {}'.format(anomaly) )
30 print('Step finished.\n')

```

A list of vegetation indices have been included in the Python Notebook as described in the introduction. Anyway, and depending on new requirements, new indices can be easily included in order to enrich further classification procedures. Note: In case some of them are not necessary for a particular classification, is not needed to remove them from the Python Notebook, but only in the further data mining analysis.

In the section GEE Image collection (Sentinel-2) + Preprocessing + Statistics extraction, we can see the selected collection and the statistics that will be generated for each parcel, as: mean, min, max, and stdDev.

## GEE image collection (Sentinel-2) + Preprocessing + Statistics extraction

```

1 # GEE image collection (Sentinel-2) + Preprocessing + Statistics extraction
2
3 # ROI
4 feat_collection=roi
5
6 # INDICES
7 # Supported by spectralIndices Library in GEE eeomt
8 spectralIndices_list= [ "NDVI", "SR", "GNDVI", "NDREI", "EVI", "EVI2", "CIG", "CIRE", "TrIvi", "MTCI", "GARI", "SIPI", "MCARI", "ARVI", "
9 # Unsupported by spectralIndices Library, builded by the function addIndices()
10 addIndices_list = ["modNDRE", "PVR", "GCI", "LCI", "TCVI", "SIPI1", "SIPI2"]
11 # Final order of the CSV visualization (all indexes)
12 all_index_list= [ "SR", "NDVI", "GNDVI", "NDREI", "modNDRE", "EVI", "EVI2", "PVR", "GCI", "CIRE", "TrIvi", "MTCI", "LCI", "TCVI", "GARI", "
13
14 # Google Earth Engine: get IMAGE COLLECTION (SENTINEL-2) -----
15 # A Sentinel-2 surface reflectance image.
16 print('Collecting Sentinel-2 images from GEE... ')
17 ds = 'COPERNICUS/S2_SR'
18 #ds = 'COPERNICUS/S2_SR/ANONIZED'
19 S2 = (ee.ImageCollection(ds)
20      #.closest(date1)#.first()#.scale().filter(ee.Filter.Lte('CLOUDY_PIXEL_PERCENTAGE', 40)).maskClouds(prob = 75,buffer = 30
21      .filterBounds(feat_collection)
22      .scaleAndOffset()
23      .filterDate(date1, date2)
24      .maskClouds()
25      .spectralIndices(spectralIndices_list)
26      .map(addIndices) )
27 print('Done.\n')
28 # Get statistics -----
29 print('Extracting zonal statistics time-series from GEE image-collection... ')
30 ts = S2.getTimeSeriesByRegions(reducer = [ee.Reducer.mean(), ee.Reducer.min(), ee.Reducer.max(), ee.Reducer.stdDev()],
31                                collection = feat_collection,
32                                bands = all_index_list,
33                                naValue = -9999,
34                                scale = 10)
35 print('Step finished.\n')
36

```

The last part of the code, print the following message, showing the date range of the data processed, and the location of a .csv file with the statistics of the selected parcel, for instance:

```

Number of features:1
Date list : ['2019-01-01', '2019-01-06', '2019-01-11', '2019-01-16', '2019-01-21', '2019-01-26', '2019-01-31']
Writing CSV file...
File written: C:/AnomalyDetection/Datasets/CSV/fields_anomaly_true.csv

Step finished. Output .csv file ready.

```

The values in the .csv file are distributed as: [date, Index(n)\_MIN, Index(n)\_MAX, Index(n)\_MEAN, Index(n)\_STD], Anomaly\_presence]

```

date,
SR_MIN,SR_MAX,SR_MEAN,SR_STD,
NDVI_MIN,NDVI_MAX,NDVI_MEAN,NDVI_STD,
GNDVI_MIN,GNDVI_MAX,GNDVI_MEAN,GNDVI_STD,
NDRE_MIN,NDRE_MAX,NDRE_MEAN,NDRE_STD,
modNDRE_MIN,modNDRE_MAX,modNDRE_MEAN,modNDRE_STD,
EVI_MIN,EVI_MAX,EVI_MEAN,EVI_STD,
EVI2_MIN,EVI2_MAX,EVI2_MEAN,EVI2_STD,
PVR_MIN,PVR_MAX,PVR_MEAN,PVR_STD,
GCI_MIN,GCI_MAX,GCI_MEAN,GCI_STD,
RECI_MIN,RECI_MAX,RECI_MEAN,RECI_STD,
TVI_MIN,TVI_MAX,TVI_MEAN,TVI_STD,
MTCI_MIN,MTCI_MAX,MTCI_MEAN,MTCI_STD,
LCI_MIN,LCI_MAX,LCI_MEAN,LCI_STD,
TCVI_MIN,TCVI_MAX,TCVI_MEAN,TCVI_STD,
GARVI_MIN,GARVI_MAX,GARVI_MEAN,GARVI_STD,
SIPI1_MIN,SIPI1_MAX,SIPI1_MEAN,SIPI1_STD,
SIPI2_MIN,SIPI2_MAX,SIPI2_MEAN,SIPI2_STD,
MCARI_MIN,MCARI_MAX,MCARI_MEAN,MCARI_STD,
ARVI_MIN,ARVI_MAX,ARVI_MEAN,ARVI_STD,
OSAVI_MIN,OSAVI_MAX,OSAVI_MEAN,OSAVI_STD,
anomaly

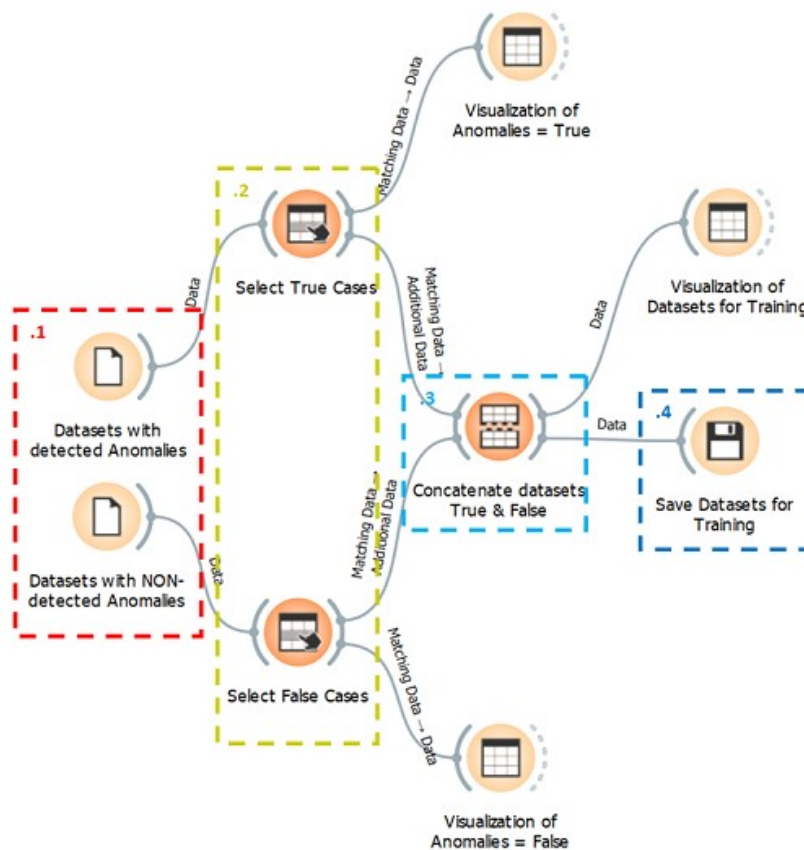
```

SR_MIN	SR_MAX	SR_MEAN	SR_STD	NDVI_MIN	NDVI_MAX	NDVI_MEAN	NDVI_STD	GNDVI_MIN	GNDVI_MAX	GNDVI
11.45631027	17.79047585	15.199478	1.488961513	0.839438796	0.893563092	0.875416577	0.01229529	0.722337008	0.765702426	0.74911
6.583038807	17.9093647	13.82984728	2.639996275	0.7362535	0.894232333	0.859906775	0.030305478	0.660797834	0.761727452	0.72688
12.28435993	19.35099411	17.17387962	1.390944948	0.849447012	0.901724696	0.889193821	0.009964646	0.732909918	0.769127488	0.75422
13.47126484	20.06920433	17.93799414	0.971313187	0.861795068	0.905074716	0.894061967	0.006483993	0.745181024	0.77170366	0.7581
13.38043499	19.63036346	18.0174886	1.36346598	0.860922158	0.903055489	0.894187247	0.009053697	0.736554384	0.784152985	0.75618
13.25340557	20.14336967	18.41993021	1.544425258	0.859682679	0.905407667	0.896188999	0.010410047	0.738447785	0.781676412	0.76252
8.544061661	17.16770172	14.28157945	2.508008476	0.790445626	0.889914513	0.864425347	0.028889651	0.703464925	0.766887009	0.74941
9.340995789	18.19682503	15.20631237	2.206005932	0.806595027	0.895816088	0.872528337	0.021110713	0.700245678	0.751011252	0.7351

As it was mentioned previously, the methodology requires to create 3 files: “fields\_anomaly\_true.csv”, “fields\_anomaly\_false.csv”, and “fields\_anomaly\_topredict.csv”, which will be the input in the next analysis with Orange Data Mining.

Three predefined workflows have been defined which are available in the repository and ready to be launched via Orange. There are different versions (Standalone, Portable, Anaconda distribution) and for different platforms (Windows, macOS, Linux) that can be used (more info at: <https://orangedatamining.com/download>).

## Data Preparation Workflow



### Procedure:

- Step 1: Open Orange Data Mining application and load the workflow, or launch it via command line:

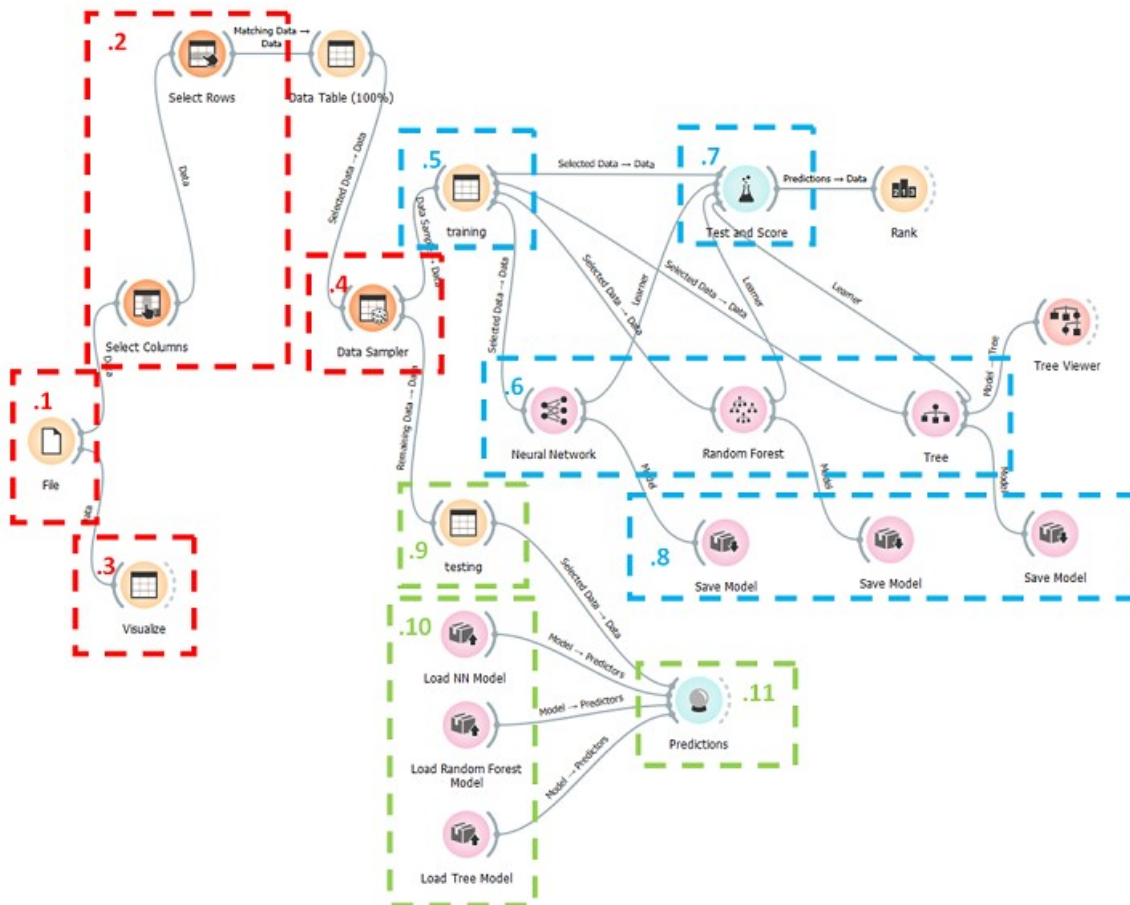
```
C:\Program Files\Orange3>orange-canvas
```

```
C:\AnomalyDetection\Orange_Workflows\A_Workflow_dataPreparation.ows
```

### Workflow description:

- 1 - Upload of datasets (.csv format) generated with VI Python Notebook (**fields\_anomaly\_true.csv** & **fields\_anomaly\_false.csv** files)
- 2 - Select particular rows (fields) if necessary
- 3 - Concatenation of True/False datasets into a single File
- 4 - Save final dataset for training **fields\_for\_training.csv** (required input for the Training Workflow)

### Training Workflow



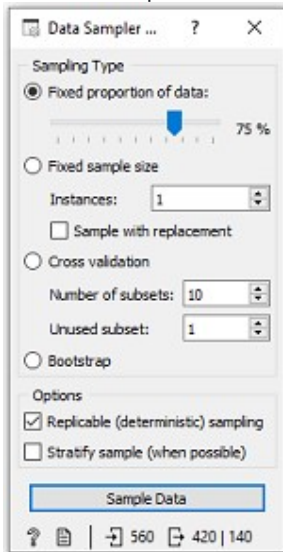
### Procedure:

- Step 1: Open Orange Data Mining application and load the workflow, or launch it via command line:  
`C:\Program Files\Orange3>orange-canvas`  
`C:\AnomalyDetection\Orange_Workflows\B_Workflow_training.ows`
- Step 2: reload the file selected in **1** (fields\_for\_training.csv)

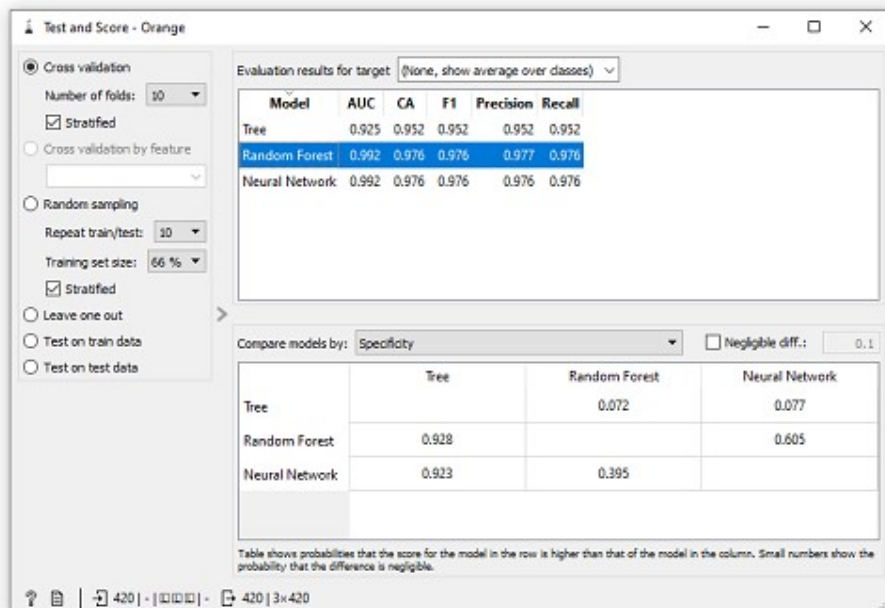
- Step 3: check the model with the best metrics from **7** (Test and Score)
- Step 4: check the model with the best metrics from **11** (Predictions)

**Workflow description:**

- 1 - Upload of the file with the information extracted from Sentinel-2 vegetation indices (fields\_for\_training.csv). *File format: Comma Separated Values (.csv)*
- 2 - Selection of Columns & Rows of interest
- 3 - Data visualization
- 4 - Data Sampler (for instance a selection of 75% of cases for training and 25% for testing)



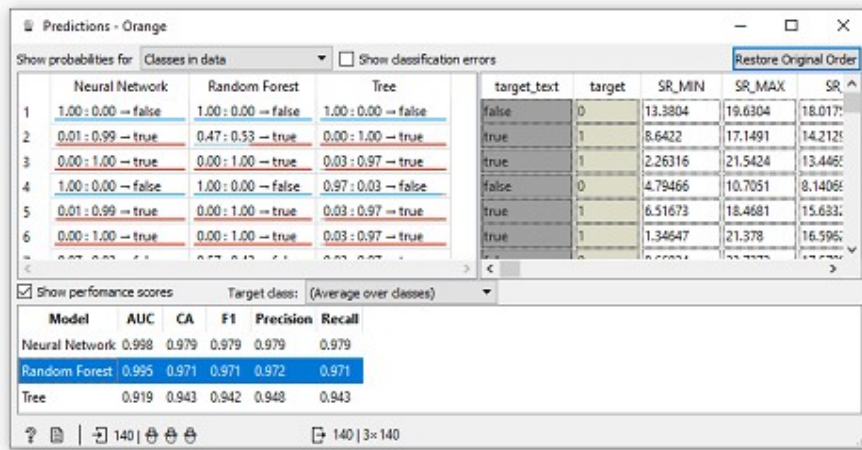
- 5 - Visualization of data for training
- 6 - Definition of model parameters
- 7 - Testing & Scoring evaluation of selected models



- Area under ROC is the area under the receiver-operating curve.
- Classification accuracy is the proportion of correctly classified examples.
- F-1 is a weighted harmonic mean of precision and recall
- Precision is the proportion of true positives among instances classified as positive, e.g. the proportion of anomaly cases correctly identified as anomalies.
- Recall is the proportion of true positives among all positive instances in the data, e.g. the number of anomaly cases among all identified as anomalies.
- Specificity is the proportion of true negatives among all negative instances, e.g. the number of non-anomaly cases among all identified as non-anomalies.

More info regarding the **Test & Score** widget at: <https://orangedatamining.com/widget-catalog/evaluate/testandscore/>

- 8 - Models trained are saved to local folder
- 9 - Visualization of data for testing
- 10 - Load of models saved on **step 8**
- 11 - Predictions based on saved trained models with a subset selected for testing



Model	AUC	CA	F1	Precision	Recall
Neural Network	0.998	0.979	0.979	0.979	0.979
Random Forest	0.995	0.971	0.971	0.972	0.971
Tree	0.919	0.943	0.942	0.948	0.943

- 12 - Selection of the model to be applied for the anomaly detection (*in the example for the demo data provided, the selected is Random Forest model*).

References of the models included for testing (source Orange Data Mining documentation):



**Tree:** is a simple algorithm that splits the data into nodes by class purity (information gain for categorical and MSE for numeric target variable). It is a precursor to Random Forest. Tree in Orange is designed in-house and can handle both categorical and numeric datasets.

The Inputs in this model are:

- Data: input dataset, and
- Preprocessor: preprocessing method(s)

While the Outputs are:

- Learner: decision tree learning algorithm, and
- Model: trained model

(more info about the parameters required at: <https://orangedatamining.com/widget-catalog/model/tree/>)



**Random Forest:** is an ensemble learning method used for classification, regression and other tasks. It was first proposed by Tin Kam Ho and further developed by Leo Breiman (Breiman, 2001) and Adele Cutler. This model builds a set of decision trees. Each tree is developed from a bootstrap sample from the training data. When developing individual trees, an arbitrary subset of attributes is drawn

(hence the term “Random”), from which the best attribute for the split is selected. The final model is based on the majority vote from individually developed trees in the forest. Random Forest works for both classification and regression tasks.

The Inputs in this model are:

- Data: input dataset, and
- Preprocessor: preprocessing method(s)

While the Outputs are:

- Learner: random forest learning algorithm, and
- Model: trained model

(more info about the parameters required at: <https://orangedatamining.com/widget-catalog/model/randomforest/>).



**Neural Network:** A multi-layer perceptron (MLP) algorithm with backpropagation. The Neural Network widget uses sklearn’s Multi-layer Perceptron algorithm that can learn non-linear models as well as linear.

The Inputs in this model are:

- Data: input dataset, and
- Preprocessor: preprocessing method(s)

While the Outputs are:

- Learner: multi-layer perceptron learning algorithm, and
- Model: trained model

(more info about the parameters required at: <https://orangedatamining.com/widget-catalog/model/neuralnetwork/> and [https://scikit-learn.org/stable/modules/neural\\_networks\\_supervised.html](https://scikit-learn.org/stable/modules/neural_networks_supervised.html))

References of the Test & Score metrics exposed in the Evaluation Results (source Orange Data Mining documentation):



**Test & Score:** The widget tests learning algorithms. Different sampling schemes are available. The widget does two things. First, it shows a table with different classifier performance measures, such as classification accuracy and area under the curve. Second, it outputs evaluation results, which can be used by other widgets for analyzing the performance of classifiers, such as ROC Analysis or Confusion Matrix.

Evaluation results for target		(None, show average over classes) v						
Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
Neural Network	0.913	0.973	0.973	0.974	0.973	0.936	0.936	0.298
Random Forest	0.942	0.973	0.973	0.974	0.973	0.936	0.936	1.006
Tree	0.883	0.919	0.920	0.922	0.919	0.813	0.913	2.982

**AUC:** Area under ROC is the area under the receiver-operating curve. An area of 1 represents a perfect test; an area of .5 represents a worthless test.

**CA:** Classification accuracy is the proportion of correctly classified examples.

**Prec:** Precision is the proportion of true positives among instances classified as positive

**Recall:** Recall is the proportion of true positives among all positive instances in the data

**F<sub>1</sub>:** is a weighted harmonic mean of previous precision and recall metrics.

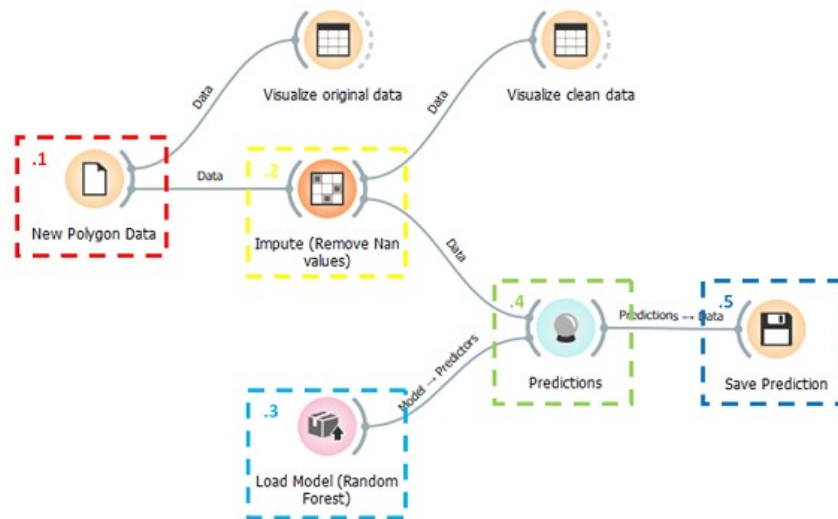
**MCC:** Matthews correlation coefficient takes into account true and false positives and negatives and .

**Spec:** Specificity is the proportion of true negatives among all negative instances.

LogLoss: takes into account the uncertainty of your prediction based on how much it varies from the actual label.

(more info at: <https://orangedatamining.com/widget-catalog/evaluate/testandscore/>)

## Prediction Workflow



### Procedure:

- Step 1: Open Orange Data Mining application and load the workflow, or launch it via command line:

```
C:\Program Files\Orange3>orange-canvas
```

```
C:\AnomalyDetection\Orange_Workflows\C_Workflow_prediction.ows
```

- Step 2: reload the file selected in **1**
- Step 3: check the selected model in **3** (default *Random Forest*)
- Step 4: check the predictions save in **5**

### Workflow description:

- 1 - Upload of new polygon/s dataset/s (in .csv format)  

```
date,SR_MIN,SR_MAX,SR_MEAN,SR_STD,NDVI_MIN,NDVI_MAX,NI
01/01/2020,2.522996058,4.033557047,3.113798426,0.36806
```
- 2 - Remove Nan values from new dataset (*cloud pixel*)
- 3 - Load of the selected trained model (example: **RandomForest\_Model.pkcls**)
- 4 - Prediction process



Predictions - Orange

Show probabilities for (None)

Random Forest	date	SR_MIN	SR_MAX	SR_MEAN	SI
1 false	2022-10-07	1.28966	2.00397	1.56897	0.1513:
2 false	2022-10-12	1.23446	1.6029	1.42338	0.0638:
3 false	2022-10-17	1.19924	1.37658	1.27601	0.0272:
4 false	2022-11-06	1.12048	1.34084	1.25576	0.0210:

- 5 - Save prediction (in .csv format)

SR_MIN	SR_MAX	SR_MEAN	SR_STD	...	SVI_MEAN	OSAVI_STD	date	Random Forest
continuous	continuous	continuous	continu	...	tinuous	continuous	discrete	false true
2.522.996.058	4.033.557.047	3.113.798.426	0.36808:		23749958	0.031689324	01/01/2020	meta false

Main description, workflows and sample testing datasets are available at: <https://github.com/silexclouds/AnomalyDetection>

#### 4. Pilot activities

- Use case: anomaly caused by pest: Psa (test case New Zealand)
- Use case: anomaly caused by pest: wheat aphid (Kenya)
- Use case: anomaly caused by drought (Poland)
- Use case: anomaly caused by temperature (Italy)
- Test Use case: crop prediction (Argentina)

#### Use case: anomaly caused by pest: Psa (test case New Zealand)

*Note: It is worth to mention that a critical point in any case of crops anomalies is the public availability of detailed and geolocated information. In particular crop illnesses occurrences, is a topic that farmers are mainly not open to publicly share. At the same time, when this information is available information as those provided by public institutions, in many cases is lacking the precise geolocation of the occurrences, which is mandatory for a robust train and validation of the proposed methodology. For this purpose, this use case which was analyzed based on public information is explained more in detail as a reference of how the methodology is applied in the different use cases (since information regarding any anomaly and the status of farmers crops, end-users mainly preferred to maintain under data privacy statements).*

In 2010, the New Zealand kiwifruit industry was hit by Psa, and the cost for this industry has been estimated to be approximately \$126 million in 2012 with an on-going cost for the next 15 years of

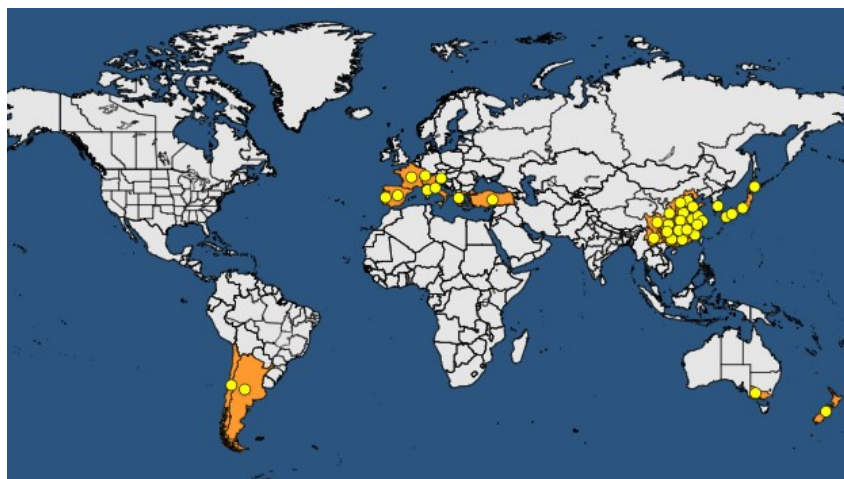
between \$740 to \$885 (Greer & Saunders 2012). Kiwifruit bacterial canker symptoms are strongly linked to environmental factors such as temperature and rainfall, which have both been shown to be associated with *Psa* bacterial population growth (Serizawa & Ichikawa 1993c, 1994; Tyson et al. 2012b). (Tyson et al. 2014c) showed in New Zealand that trap plants placed in *Psa* infected orchards were able to be infected year-round, particularly in spring, and that infection events were strongly associated with rainfall. They concluded that rain-splash and wind-blown rain were the main mechanisms of localized natural spread between and within vines in New Zealand (Froud et al., 2017).



Kiwitrees in normal conditions (left), and with *Psa* infection (right)

There are several strains of *Psa* found worldwide, some causing moderate or important damages (McCann et al. 2013). For instance, impact of high virulence has been detected in Hayward I Japan and Italy (source: Kiwifruit Vine Health, NZ). The *Hayward* kiwi is the main green variety produced for the world's markets and dominates production in most growing areas.

Regarding the presence of *Psa* (*Pseudomonas syringae* pv. *actinidiae*), according to EPPO Global Database<sup>1</sup> (<https://gd.eppo.int/>) there are more than 10 potential hosts, not only kiwi trees. To test the methodology it was selected the Kiwi trees.



Global Distribution of *Psa* (source: EPPO)

<sup>1</sup> EPPO Global Database is maintained by the Secretariat of the European and Mediterranean Plant Protection Organization (EPPO). The aim of the database is to provide all pest-specific information that has been produced or collected by EPPO. The database contents are constantly being updated by the EPPO Secretariat.

New Zealand and Italy are key exporters of kiwi fruit, representing around 29% and 20% of total exports, respectively. In Italy In particular the kiwifruit industry currently represents one of the most dynamic and profitable investments within the Italian agricultural sector. Kiwifruit production and plantation areas have been increasing since the 1970s, when the first orchards were planted. By now, the production of Italian kiwifruit is concentrated in the following districts: the province of Latina, the Romagna region, and the provinces of Cuneo and Verona (Palmieri et al., 2014).

For the case of Psa, the implementation and test of the methodology has been carried out in different parcels located in New Zealand, thanks the availability of public Psa Reports published by the Kiwifruit Vine Health of New Zealand, with detailed information regarding the presence/absence of the illness.

New Zealand kiwifruit orchards location	Orchards with Psa-V	Total orchards in region	% of hectares on an orchard with Psa-V identified
Coromandel	47	47	100%
Franklin	99	102	98%
Hawkes Bay	43	47	96%
Kaikati	458	466	99%
Kerikeri	92	110	88%
Northwest Auckland	21	23	90%
Opotiki	223	229	99%
Pouaru Bay	58	74	99%
<b>South Island</b>	<b>0</b>	<b>119</b>	<b>0%</b>
Tairāpapa	500	506	100%
<b>Te Puke</b>	<b>1098</b>	<b>1115</b>	<b>99%</b>
Waikato	40	40	100%
Waikato	82	88	91%
Wanganui/Horowhenua	13	17	93%
Whakatane	160	163	97%
Whangarei	9	58	22%

Psa statistics December 2019 (source: Kiwifruit Vine Health)



regions covering more than 1200 orchards with validated data

According to statistics from December 2019 extracted from the Kiwifruit Vine Health portal, South Island had 0 orchards with detected Psa from a total of 119 orchards. While in the Te Puke region from a total of 1115 orchards, in 99% of them it was detected the presence of Psa.

400 parcels have been digitalized and included in the analysis to test the methodology.

After the training of the different models with the input datasets, the results that can be obtained from the Test and Score widget clearly show the different performances which help to select the model which better fits the expected classification.

Evaluation results for target (None, show average over classes) ▾

Model	AUC	CA	F1	Prec	Recall	MCC
Neural Network	0.999	0.988	0.988	0.988	0.988	0.975
Random Forest	0.992	0.978	0.978	0.979	0.978	0.955
Tree	0.925	0.942	0.942	0.942	0.942	0.878

The different trained models are automatically saved as it was already described in the Training Workflow - block #8.

After the model is trained, the methodology continues applying the model in new fields not included in the testing.

The selected model, is loaded in the Prediction Workflow – block #3, with new datasets to be analyzed, and from the prediction widget (Prediction Workflow – block #4) can be seen the results of the prediction made based on the trained model.

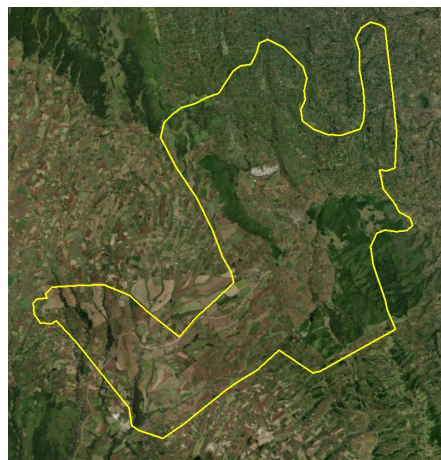
Finally the Save Prediction widget (Prediction Workflow – block #5) let the possibility to save the output as a .csv file with the probability of the field of belonging to the classes infected/not infected.

Neural Network	anomaly	SR_MIN	SR_MAX	SF
1 false	topredict	5.10086	12.1362	9.1974
2 false	topredict	11.4563	17.7905	15.199
3 false	topredict	6.58304	17.9094	13.829
4 false	topredict	12.2844	19.351	17.173
5 false	topredict	13.4713	20.0692	17.938
6 false	topredict	13.3804	19.6304	18.017

The trained model has been also applied in other unknown not monitored kiwi trees orchards that can be linked and later visualized with the support of a geographic information system together with the probabilities of belonging to a suspicious infected class.



### Use case: anomaly caused by pest: wheat aphid (Kenya)



Region identified with the presence of Wheat Aphid

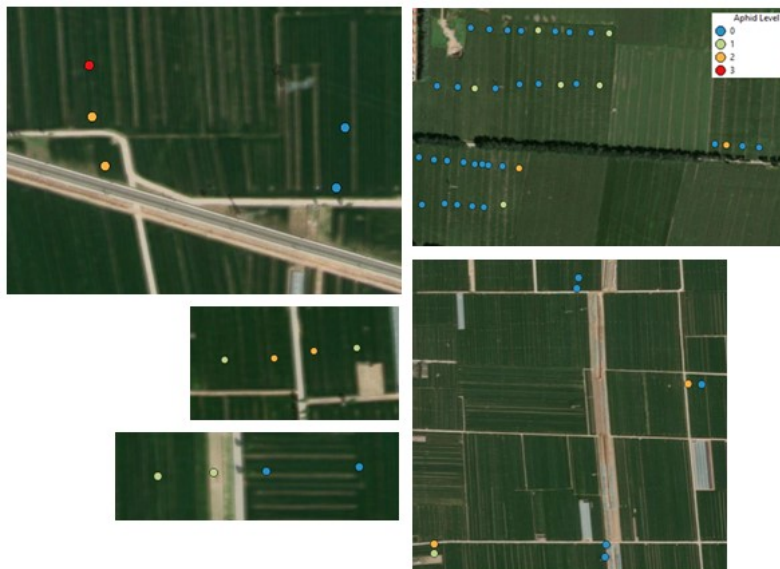
Aphids can be serious plant pests and may stunt plant growth, produce plant galls, transmit plant virus diseases, and cause the deformation of leaves, buds, and flowers. In particular when the wheat crop is ripe and the summer is approaching, the winged forms of this aphids migrate also to other plants. It is not known how the pest passes the summer and the monsoon season, but in October-November, the aphids again appear on wheat. If available, barley is preferred to wheat.



Heavily infested plants are stunted, and sometimes exhibit a flattened appearance with tillers lying almost parallel to the ground. Occasionally, particularly during cold weather, plants show a purple color. Infested leaves curl up like a soda straw and remain in a rigid upright position rather than assuming the typical drooping posture (source: University of California, Integrated Pest Management UC-IPM). The losses due to aphids have been reported up to 36%. (source Tamil Nadu Agricultural University, eagri.org), putting this threat and its analysis under a particular focus.

In particular in Kenya this illness has been first identified in several fields in 1995. It spread quickly to all the wheat growing areas of the country and it is nowadays the most important pest of wheat and barley.

For the training, testing and validation of the methodology, several parcels (with the presence of wheat aphid) have been indicated by the users and others fields where until now the presence and impact of this insect has not been reported yet. Different fields have been processed according the methodology described previously with the information provided, extracting the vegetation indices during the period of October/November.



For crop fields where the wheat aphid has been effectively detected and fields where there were no evidences of the aphid, the list of preselected vegetation indices has been extracted via the python notebook, and ingested in the explained workflows.

The selected model by the user was the Random Forest. But according to their preliminary results, up to now, they mentioned the need to obtained more detailed information about the density of this aphid in the fields, since the impact appears not to be easily detected by the vegetation indices or overestimated..

More detailed information regarding the current wheat campaign is being collected to retrain the model expecting more decisive results to apply the trained model in new not monitored fields.

### Use case: anomaly caused by drought (Poland)

Different wheat fields affected during the period 2016-2021 have been selected by the user, taking the year 2018 as a reference of those fields affected by drought, 2016 as an agricultural season without anomalies. The satellite datasets analyzed and ingested in the models have been those available during the period of time between March and June of each year. The others years have been identified by the user as season where this anomaly was present but affecting the fields not so intensively. The results reported by the user mentioned good performances of the models reaching more than 85% of good classifications.

*Note: a critical point underlined by the user that must be considered in improvements of the algorithm provided is that the statistics output file obtained by the methodology considers stats of the Vis of the parcels, which not always describe the heterogeneity inside parcels.*



### Use case: crop anomaly caused by temperature (Italy)

(Agnolucci et al. 2020) demonstrated the significance and impact of climate change with different statistical patterns in 18 crops that contribute 70% of the land and 65% of calorific value. The yield

disparity in primary crops is associated with high temperatures. High temperatures above 35°C can cause damage to rice crops. Evident injuries were observed due to high temperatures in different developmental stages.

(Zhao et al. 2017) investigated that the rise in global temperature leads to global yield loss in four significant food crops (Wheat, rice, maize, and soybean), which provide two-thirds of human caloric intake. Without CO<sub>2</sub> fertilization, effective adaptation, and genetic improvement, each degree-Celsius increase in global mean temperature would, on average, reduce global yields of wheat by 6.0%, rice by 3.2%, maize by 7.4%, and soybean by 3.1%. Results are highly heterogeneous across crops and geographical areas, with some positive impact estimates.

For this use case, have been processed several crop fields affected by temperature unexpected disturbances pointed by researchers who are monitoring the evolution of the crops in the region of interest, mainly wheat and maize crops. The results exposed by the users underlined the usefulness of satellite imagery for the detection of the stress on crops generated by anomaly temperatures, as various physiological changes during the different growth stages. As it was analyzed in (Parthasarathi et al. 2022) the reproductive stage is the most vulnerable phase of the crop's entire lifespan; this vulnerability during the reproductive stage leads to significant depletion in seed set and crop yield (Jagadish 2020, Masoomi-Aladizgeh et al. 2021).

Exposure to high temperatures reduces plant height, total biomass, and the number of tillers. Phenology, plant water relations, dry matter partitioning, and shoot growth are restricted by heat stress in bean plants (Omae et al. 2012). In some plants, it reduces total phenological duration, shorter grain filling period, and diminishes the germination period (Parthasarathi et al. 2022). These anomalies are most of them capable to be detected and monitored with the support of satellite imagery, Different patterns are being analyzed by the researchers who are collecting in-situ data for this region of interest



located in Italy with the support of the methodology proposed in the current report.

## Mention about Copernicus Contributing Missions

The potential application of datasets provided by Copernicus Contribution Missions will allow this methodology to be applied for use cases focused on monitoring anomalies in single trees, where in that case very high spatial resolution images are required (as the Pleiades, Worldview, or GeoEye constellations).

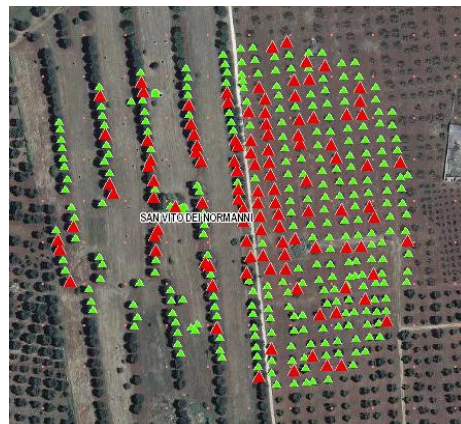
That is the case of the problem caused by the Pine Wood Nematode (PWN), or the *Xylella fastidiosa* (Xf).



PWN (left), Xf (right)

The Pine Wilt disease (PWD), is caused by the pine wood nematode (PWN), and is considered one of the most important pests in the world forest ecosystems. The pine wood nematodes are invisible to the naked eye but they are spread by sawyer beetles, and the symptom that can be detected is the rapid wilting of trees. It was listed as a quarantine pest in Europe.

Xf is a vector-transmitted bacterial plant pathogen that affects a wide range of plants (> 500 hosts). The visible symptoms are leaf scorch, foliage wilting, withering of branches. A particular use case regarding Xf, could be analyzed in the region of Puglia, where the pest is affecting many olive orchards (in Europe 93% of the olive production is coming from Spain, Italy and Greece). And the regional phytosanitary institutions provide very precise geo-located information regarding the presence/absence of the illness ([www.emergenzaxylella.it](http://www.emergenzaxylella.it)) for each single tree.



Xf presence/absence data provided by regional institutions and publicly available

## Test Use case: methodology applied to crop forecast (Argentina)

This use case was carried out to test the potentialities of the proposed methodology to forecast expected crop yields. The in-situ data was provided by agricultural harvesting machinery in different maize fields in Argentina for different campaigns. The information was provided with detailed intra field information. These datasets must be transformed from point based to parcel based during the data preparation process to match (1 to 1) the records [Crop yield forecast] & [Vegetation indices



parcel statistics]. The results of the application of the model underlined a ranking of the main vegetation indices that are enriching the model, information that can be extracted via the Rank Widget. This tool (visible in the provided Training Workflow, block 7) evidences the ranking of attributes in the classification or regression datasets, scoring the variables according to their correlation with the target variable, in this case Crop Yield Forecast.



## 5. General Outlook

A methodology for the monitoring and detection of potential crop anomalies has been presented. Satellite based imagery data provided by Copernicus systems, accessible via Google Earth Engine via its user friendly platform, is a key success point for the users to manipulate this kind of data and easily extract statistical information. At the same time, Copernicus portfolio combined with free Data Mining and Visualization applications as the one proposed in the current methodology, gives the possibility to improve final end users skills and capabilities to transform satellite data in added value information. Regarding the capabilities of this methodology to detect anomalies based on pests, a critical factor is the availability of precise geo-located information regarding illness occurrences, which is mainly not accessible or published. This issue restricts the possibility to train robustly a model and the further application to new areas which are still not registered or officially declared with the presence of a pest. At the same time, the spatial resolution of Sentinel-2 constellation, depending on the crop that want to be monitored show certain limitations. In that case, Copernicus Contribution Missions could represent an important support to cover a wide range of use cases, where single trees must be regularly monitored.

## 6. References

Gahoi A., Data Analytics for Predicting Wheat Crop Production. Artificial & Computational Intelligence / Published online: Dec 2019. [http://acors.org/ijacoi/VOL1\\_ISSUE1\\_06.pdf](http://acors.org/ijacoi/VOL1_ISSUE1_06.pdf)

Hueni A., Suarez L. , Chisholm L.A. and Held A. , "The Use of Spectral Databases for Remote Sensing of Agricultural Crops" , in Fundamentals, Sensor Systems, Spectral Libraries, and Data Mining for Vegetation ed. Prasad S. Thenkabail , John G. Lyon and Alfredo Huete (Boca Raton: CRC Press, 2018).

Verstrate M.M., Pinty B., Myneni R.B., Potential and limitations of information extraction on the terrestrial biosphere from satellite remote sensing, *Remote Sensing of Environment*, Volume 58, Issue 2, 1996, Pages 201-214, ISSN 0034-4257, [https://doi.org/10.1016/S0034-4257\(96\)00069-7](https://doi.org/10.1016/S0034-4257(96)00069-7).

Niinemets U; (2010) A review of light interception in plant stands from leaf to canopy in different plant functional types and in species with varying shade tolerance. *Ecol Res*. doi:10.1007/s11284-010-0712-4.

Rouse JW, Haas RH, Schell JA, Deering DW 1973. Monitoring vegetation systems in the great plains with ERTS. Third ERTS Symposium, NASA SP- 351 I: 309–317.

Eitel J.U.H., Vierling L.A, Long D.S., Litvak M., Eitel K.C.B. (2011 ). Simple assessment of needleleaf and broadleaf chlorophyll content using a flatbed color scanner. *Canadian journal of Forest Research*, 41, 1445-1451.

Gitelson AA, Merzlyak MN. Remote estimation of chlorophyll content in higher plant leaves. *Int J Remote Sens*. 1997;18(12):2691–2697. <https://doi.org/10.1080/014311697217558>

Gitelson AA, Merzlyak MN 1994. Spectral reflectance changes associated with autumn senescence of *Aesculus Hippocastanum* L. and *Acer Platanoides* L. leaves: spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology* 143:286–292.

Datt B. A new reflectance index for remote sensing of chlorophyll content in higher plants: tests using eucalyptus leaves. *J Plant Physiol*. 1999;154(1):30–36. [https://doi.org/10.1016/S0176-1617\(99\)80314-9](https://doi.org/10.1016/S0176-1617(99)80314-9)

Huete A, Justice C, Liu H 1994. Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of Environment* 49: 224–234.

Huete AR, Liu H, Batchily K, van Leeuwen W 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment* 59: 440–451.

Metternicht G.; (2003). Vegetation indices derived from high-resolution airborne videography for precision crop management. *Int. J. Remote Sens*. 24. 10.1080/01431160210163074.

Gitelson A.; Zur Y.; Chivkunova O.; Merzlyak M.N. Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. *Photochemistry and photobiology*, 2002, 75, pp. 272- 281.

Gitelson AA, Gritz Y, Merzlyak MN 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithm for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology* 160: 271–282.

Broge, Niels & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*. 76. 156-172. 10.1016/S0034-4257(00)00197-8.

Dash J., Curran P. (2004). The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing - INT J REMOTE SENS*. 25. 10.1080/0143116042000274015.

Mouret F., Albughdadi M., Duthoit S., Kouamé D., Rieu G., Detecting anomalous crop development with multispectral and SAR time series using unsupervised outlier detection at the parcel-level: application to wheat and rapeseed crops. 2020. fihal-02546260v2

Kaufman Y.J., Tanré D.; Atmospherically resistance vegetation index (ARVI) for EOS-MODIS *IEEE Trans. Geosci. Remote Sens.*, 30 (1992), pp. 261-270

Rondeaux G., Steven M., Frederic B.; (1996). Optimization of Soil-Adjusted Vegetation Indices. *Remote Sensing of Environment*. 55. 95-107. 10.1016/0034-4257(95)00186-7.

Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan B (2013) Orange: Data Mining Toolbox in Python, *Journal of Machine Learning Research* 14(Aug): 2349–2353.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.

Palmieri A., Pirazzoli C.; L'actinidia in Italia e nel mondo tra concorrenza e nuove opportunità. *Riv. Fruttic.* 2014, 12, 66–68.

Parthasarathi T., Firdous S., Mariya David E., Lesharadevi K., and Djanaguiraman M., 'Effects of High Temperature on Crops', *Advances in Plant Defense Mechanisms*. IntechOpen, Dec. 14, 2022. doi: 10.5772/intechopen.105945.

Jagadish SVK. Heat stress during flowering in cereals – Effects and adaptation strategies. *New Phytologist*. 2020;226(6):1567-1572. DOI: 10.1111/NPH.16429

Masoomi-Aladizgeh F, Najeeb U, Hamzelou S, Pascovici D, Amirkhani A, Tan DKY, et al. Pollen development in cotton (*Gossypium hirsutum*) is highly sensitive to heat exposure during the tetrad stage. *Plant, Cell & Environment*. 2021;44(7):2150-2166. DOI: 10.1111/PCE.13908

Omae H, Kumar A, Shono M. Adaptation to high temperature and water deficit in the common bean (*Phaseolus vulgaris* L.) during the reproductive period. *Journal of Botany*. 2012;2012:1-6. DOI: 10.1155/2012/803413