# Estimating crop yield from spectral vegetation indices

In this exercise you will build simple (linear) relationships between within-season spectral vegetation indices and crop yield in 2020. You will acquire the spectral vegetation indices from Level-2 (atmospherically-corrected) Sentinel-2 multispectral broadband and narrowband data (**Table 1**). The yield data consists of crop yield survey frames with dimensions of 60×60m$^2$. The frames were delineated using a handheld GPS. You will analyse winter wheat yields for 40 frames (**Figure 1**). **We have provided the frames for rice as well. The rice data are for those of you who have additional time and would like to be challenged further.** Crop cuttings were randomly selected in each frame. The grain was separated, dried, and weighed. The weights were averaged to obtain average dry-weight yield per frame in kg m$^{-2}$. The dimensions of the frames were chosen to account for any geographic mismatch between the frames and spatial resolution of Sentinel-2 (20×20m$^2$).

**Table 1.** Summary of Sentinel-2 sensor characteristics.

| Sentinel-2 | Band # | Designation | Spectral Range (nm) | Spatial Resolution (m) |
|---|---|---|---|---|
| | 1 | Coastal aerosol | 433-453 | 60 |
| | 2 | Blue | 458-523 | 10 |
| | 3 | Green | 543-578 | 10 |
| | 4 | Red | 650-680 | 10 |
| | 5 | Red-edge 1 | 698-713 | 20 |
| | 6 | Red-edge 2 | 733-748 | 20 |
| | 7 | Red-edge 3 | 773-793 | 20 |
| | 8a | NIR | 785-900 | 10 |
| | 8b | Narrow NIR | 855-875 | 20 |
| | 9 | Water vapor | 935-955 | 60 |
| | 10 | SWIR-Cirrus | 1360-1390 | 60 |
| | 11 | SWIR1 | 1565-1655 | 20 |
| | 12 | SWIR2 | 2100-2280 | 20 |

We conducted the field campaign over ~3850 ha of the Bonifiche Ferraresi farm, which is on the eastern edge of the Po River plain in Emilia-Romagna, Italy. The plain is a major center for agricultural production in Europe. The soil consists primarily of silty-clay or silty-clay loam and is highly fertile due to alluvial deposits from the Po River. The climate is humid subtropical (Köppen classification: Cfa) because of its hot and humid summers (average daily temperature = 23.1°C; average monthly precipitation = 54.6mm) and mild winters (average daily temperature = 3.9°C; average monthly precipitation = 41.1mm). The farm, like the surrounding area, grows important food and feed crops: alfalfa, barley, corn, rice, soybean, and wheat.

This exercise consists of three phases: (i) download Sentinel-2 reflectance data over Bonifiche Ferraresi farm for the 2020 growing season; (ii) calculate spectral vegetation indices from the Sentinel-2 reflectance data; and (iii) identify the best index and single date/multidate image set for estimating crop yield.

Learning outcomes:

At the end of the exercise, you will have

- calculated spectral vegetation indices from Sentinel-2 surface reflectance
- examined the relationship between spectral vegetation indices and crop yield through time

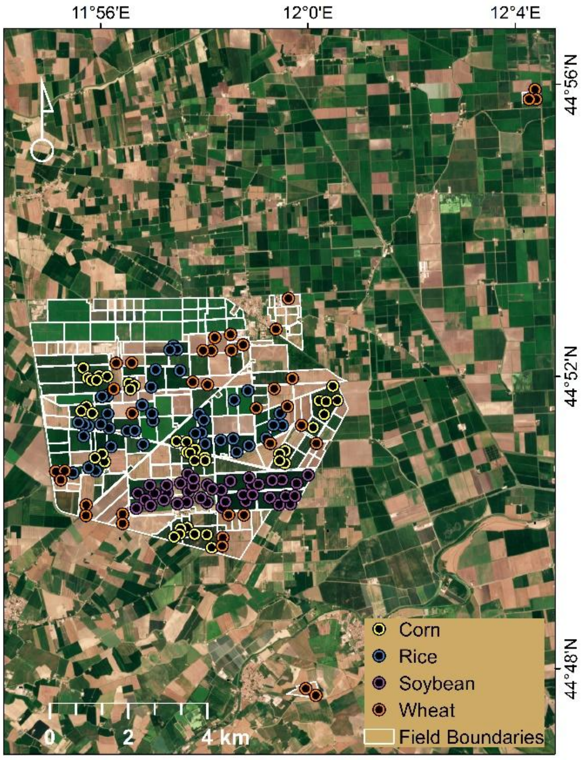


**Figure 1.** Bonifiche Ferraresi farm study area. The symbols show the location of crop yield survey frames that were collected in the 2020 growing season for four important staples.

## Download Sentinel-2 imagery over Bonifiche Ferraresi farm

Go to the Copernicus Space Ecosystem (https://dataspace.copernicus.eu/). **Please create a user account if you have not done so already** (https://identity.dataspace.copernicus.eu/auth/realms/CDSE/login-actions/registration?client_id=cdse-public&tab_id=Ksu5Mndwzk0). You will need to enter the location of the farm and the ingestion (start-end) period before proceeding further.

The farm can be located by typing "Jolanda di Savoi, Italy" into Google maps. It is just to the southwest of the town (**Figure 2**).

The winter wheat was planted and harvested on **22 October 2019 (DOY=295)** and **10 July 2020 (DOY=192)** respectively. These are your start and end dates for image acquisition. The reproductive and ripening/maturity phases began on 21 May 2020 (DOY=142) and 22 June 2020 (DOY=174) respectively. You will need these dates later in the exercise.

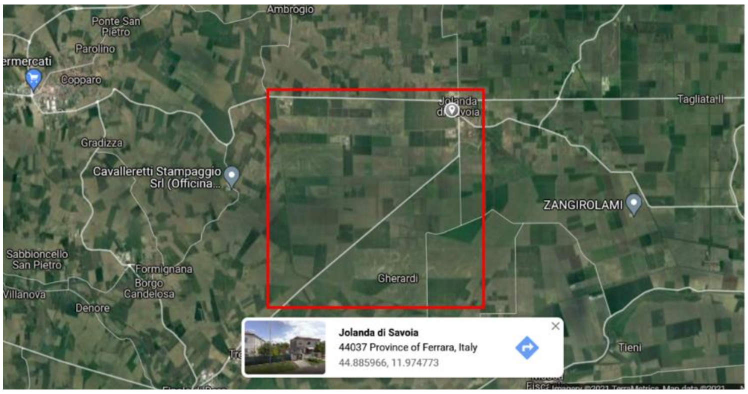


**Figure 2.** Bonifiche Ferraresi (highlighted in red) is located near Jolanda di Savoi, Italy.

On the search criteria tab in the upper left-hand corner of the Copernicus browser, enter the start and end dates for image acquisitions corresponding to the start and end of growing season. Check the L2A box. The level L2A product is surface reflectance. Level 2 data are essentially corrected and therefore more appropriate for time series analysis. You may come across the 2Ap product, which is preliminary and will not be used for this exercise. Select 40% on the cloud cover slide bar. This means the search algorithm will only select images with 0-40% cloud cover. If you open the filters tab, you will see additional criteria. Prior to 2017, the Sentinel-2 constellation consisted of only one platform with a 10-day return frequency (S2A_). Now it consists of two platforms with a 5-day return frequency (S2B_). Soon additional platforms (S2C_ and S2D_) will be added to the constellation. These handy features can simplify searches tremendously.

Now use the tools on the right side of the Copernicus browser to define the boundaries of your search (**Figure 3**). The search algorithm will use these boundaries to select relevant Sentinel-2 scenes.

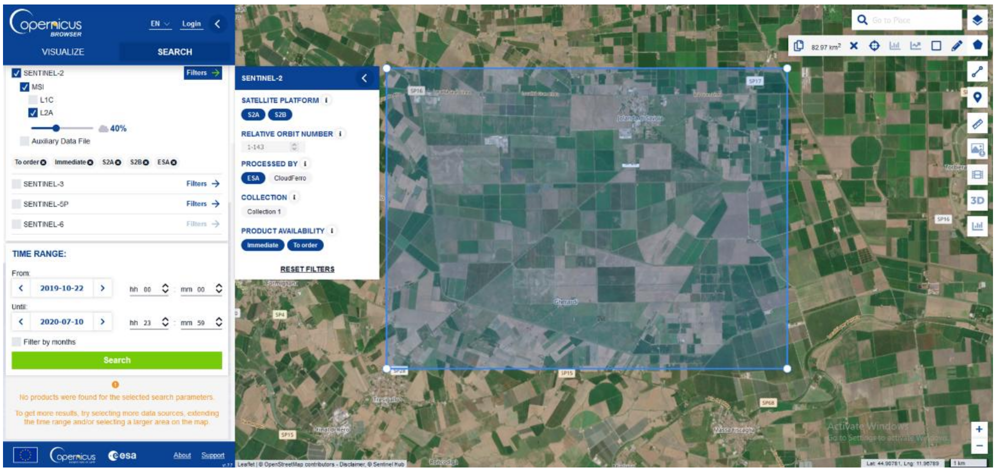


**Figure 3.** Defining your boundaries and other inputs to narrow down the mage search.

Use the visualize tab to look at true colour and other composites. Only select images that are relatively free of clouds over the study area—there are approximately seven. Omit the first three images, because these occur in the early vegetative stage before wheat is fully emergent. In the end, you will need to download four images for the analysis:

S2B_MSIL2A_20200403T100549…early vegetative phase
S2B_MSIL2A_20200423T100549…vegetative phase
S2B_MSIL2A_20200523T100559…reproductive phase
S2B_MSIL2A_20200622T100559…ripening/maturity phase

**Be sure to download the imagery well ahead of your analysis. Offline images must be ordered and will become visible in your shopping cart after some time. It can take up to 24 hours. Each image package is a zip file of ~1GB in size. They will take time to download.**

**It should be noted that Sentinel products can also be acquired from USGS Earth Explorer (https://earthexplorer.usgs.gov/). This data portal is more user friendly, but it does not offer the same range of European products as the Copernicus Space Ecosystem.**

## Calculate Sentinel-2 spectral vegetation indices

The image folders are in archive (tar) zip form, so you will need the free 7zip software (https://www.7-zip.org/) to extract the data. The most important information is located in the "GRANULE" subfolder. This folder contains the image bands and information on data quality. The image folder contains images for three spatial resolutions: 10m, 20m, and 60m. Bands 2, 3, 4, and 8 are 10m native, but have been resampled to 20m resolution to match the other 20m bands (5, 6, 7, 11, and 12). Bands 1, 9, and 10 are 60m native. They are intended for coastal and atmospheric studies, so we will not consider them in our analysis.

You are welcome to open the band images and combine them into RGB composites. Band combination 4,3,2 is a true colour composite, but is not very useful for analysis. Band combination 8,4,3 is a false colour composite that is good for analysing vegetation. Band combination 11, 8, 2 is often called the agriculture false colour composite. The band combination is good for monitoring crop health.

1. Explore the images with RGB composites. Compare and contrast the colours according to land cover type (buildings, bare soil, crops, etc.) and crop growth stage. What do the colours mean in terms of the spectral response?

Use the accompanying shapefile of survey frames (wheatframes.shp) to extract the mean spectral value for each 20m band and image date. Export these data and save them in a spreadsheet for your analysis. You can also use the point shapefile (wheatpoints.shp) to extract spectral values directly, but the frames are better, because they account for the spatial mismatch between ground and Sentinel-2 data. We have also included a shapefile containing the boundaries of the farm for your interest (bferraresi.shp).

Use the spectral values (they should range 0-1) to calculate spectral indices for each image date in a spreadsheet. **Table 2** contains a list of some important spectral vegetation indices that can be derived from Sentinel-2 bands listed in Table 1. A comprehensive list of spectral indices can be found here. For your analysis of crop yield, select at least three indices (NDVI, one red-edge index, and one SWIR index).

## Estimate crop yield with Sentinel-2 spectral vegetation indices

Use linear regression, scatterplots, and residual plots to assess the performance of the spectral vegetation indices in estimating crop yield for each image date. The crop yield data for the 40 survey frames is in wheatyield.xls. The frames are geotagged with the same ID found in the shapefile.

**Table 2.** Spectral vegetation indices and their formula in relation to Sentinel-2 bands.

| Description | Equation |
|---|---|
| Normalized Difference Vegetation Index (NDVI) | $NDVI = \frac{NIR\ B8 - Red\ B4}{NIR\ B8 + Red\ B4}$ |
| Green NDVI | $GNDVI = \frac{Green\ B3 - Red\ B4}{Green\ B3 + Red\ B4}$ |
| NDVI red-edge 1 | $NDVI\_B8B5 = \frac{NIR\ B8 - Red\ Edge\ B5}{NIR\ B8 + Red\ Edge\ B5}$ |
| NDVI red-edge 2 | $NDVI\_B8B6 = \frac{NIR\ B8 - Red\ Edge\ B6}{NIR\ B8 + Red\ Edge\ B6}$ |
| NDVI red-edge 3 | $NDVI\_B8B7 = \frac{NIR\ B8 - Red\ Edge\ B7}{NIR\ B8 + Red\ Edge\ B7}$ |
| NDVI red-edge 1 narrow | $NDVI\_B8AB5 = \frac{NIR\ B8A - Red\ Edge\ B5}{NIR\ B8A + Red\ Edge\ B5}$ |
| NDVI red-edge 2 narrow | $NDVI\_B8AB6 = \frac{NIR\ B8A - Red\ Edge\ B6}{NIR\ B8A + Red\ Edge\ B6}$ |
| Normalized Difference Vegetation Index red-edge 3 narrow | $NDVI\_B8AB7 = \frac{NIR\ B8A - Red\ Edge\ B7}{NIR\ B8A + Red\ Edge\ B7}$ |
| Plant Senescence Reflectance Index | $PSRI = \frac{NIR\ B8 - Green\ B3}{Red\ Edge\ B5}$ |
| Chlorophyll Index red edge | $CLre = \frac{Red\ Edge\ B7}{Red\ Edge\ B5} - 1$ |
| Normalized Difference red-edge 1 | $NDre1 = \frac{Red\ Edge\ B6 - Red\ Edge\ B5}{Red\ Edge\ B6 + Red\ Edge\ B5}$ |
| Normalized Difference red-edge 2 | $NDre2 = \frac{Red\ Edge\ B7 - Red\ Edge\ B5}{Red\ Edge\ B7 + Red\ Edge\ B5}$ |
| Normalized Difference Water Index | $NDWI = \frac{Green\ B3 - NIR\ B8}{Green\ B3 + NIR\ B8}$ |
| Modified Normalized Difference Water Index | $MNDWI1 = \frac{Green\ B3 - SWIR\ B11}{Green\ B3 + SWIR\ B11}$ |
| Modified Normalized Difference Water Index | $MNDWI2 = \frac{Green\ B3 - SWIR\ B12}{Green\ B3 + SWIR\ B12}$ |

Below are some questions to consider in your analysis:

2. Which spectral vegetation index produces the highest correlation with crop yield? What image date and crop stage (early vegetative, vegetative, reproductive, ripening/maturity) produces the highest correlation with crop yield? Is the relationship linear or non-linear? Why do you think the given index and crop stage are the best performing?
3. Examine the scatter and residual plots. Do you see any potential outliers? Would you remove these values? Why or why not?
4. Sometimes the data is non-linear. Transform the yield data using the natural log. Does the transformation improve the regression? Why or why not?
5. Single date images usually produce poorer estimates than multidate image composites. Now sum the indices (e.g., ∑NDVI) and take the average (e.g., meanNDVI) over the growing season. Repeat your regression analysis. Do these seasonal composites improve estimation of crop yield?

## Rice yield estimation

For the challenge, you can use the accompanying shapefiles (riceframes or ricepts) and crop yield data (riceyield.xls) to assess performance of the same spectral indices for rice. The growing season for rice is different from winter wheat, so you will have to download three additional Sentinel-2 images. Rice was planted on 27 April 2020 (DOY=118) and harvested on 8 September 2020 (DOY=252). The reproductive stage began on 4 August 2020 (DOY=217). The ripening/maturity stage began on 1 September 2020(DOY=245). Do the same spectral indices and crop stages perform the best? Why or why not? How do the sum and average seasonal composites perform?