

# EUROCROPS Demo Dataset

## Introduction

EUROCROPS is a dataset based on self-declared field annotations for training and evaluating methods for crop type classification and mapping, together with its process of acquisition and harmonisation. The aim of EUROCROPS is to enrich the research efforts and discussion for data-driven land cover classification via Earth observation and remote sensing. The dataset is published in different formats for researchers in remote sensing, computer vision and machine learning fields.

## Demo Dataset Description

EUROCROPS Demo Dataset contains harmonised agricultural parcel information data from 3 regions, namely Austria, Denmark and Slovenia which allows for a better representation of Europe's agricultural diversity. The dataset has been split into training and test sets as earth observation data is influenced by spatial auto-correlation, implying that using adjacent parcels for machine learning or remote sensing should be refrained from.

Country	Year	Parcels
Austria	2020	396,600
Denmark	2019	98,565
Slovenia	2020	310,236

The dataset is further presented in 3 formats to aid researchers for developing better models.

Format	Description
CSV	<b>Cols:</b> Parcel id, HCAT id, Crop name
HDF5	<b>Rows:</b> Parcel id <b>Columns:</b> Timesteps <b>Cells:</b> 13 representative reflectances (median value) for a parcel, each corresponding to one band of SENTINEL-2 satellites.*
GeoJSON	Multi polygon coordinates and its metadata

\*SENTINEL-2 satellites consist of 13 bands- 1,2,3,4,5,6,7,8,8a,9,10,11,12. For our dataset, band 8a is represented by the last reflectance value in the list i.e. the bands of our dataset are- 1,2,3,4,5,6,7,8,9,10,11,12, 8a

**HDF5:** The HDF5 file consists of the reflectance data of the 13\* bands of SENTINEL-2 satellites. Each subgroup of the dataset represents a NUTs id, with the corresponding parcel id, and time steps of the reflectance data. The metadata augments the dataset by providing the corresponding EAGLE matrix crop name and its respective crop code.

**CSV:** The demo dataset contains CSV files of each region, further divided into subregions as per the NUTs classification. Each CSV file consists of the parcel id (recno), the crop name as per the EAGLE matrix as well as its crop code.

**GeoJSON:** The GeoJSON has the multi polygon coordinates for the regions in the demo dataset along with its corresponding metadata, namely the NUTs id, crop name, crop id, country name as well as others.

\*SENTINEL-2 satellites consist of 13 bands- 1,2,3,4,5,6,7,8,8a,9,10,11,12. For our dataset, band 8a is represented by the last reflectance value in the list i.e. the bands of our dataset are- 1,2,3,4,5,6,7,8,9,10,11,12, 8a