



Deliverable 2: xSeedScore EO module (IP held by Computomics)

*Partner in charge: Computomics, SilberGeo*

*Date: 02.03.2021*

*research and innovation programme under grant agreement No 824478.*

*This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824478.*



|  |  |
| --- | --- |
| **Project name** | **Crop Predictions Take Flight** |
| **Acronym** | **CROPTF** |
| **Start date** | 01.09.2020 |
| **End date** | 31.07.2021 |
| **Project Coordinator** | Computomics |
|  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Document History** | | | | |
| **Version** | **Issue date** | **Stage** | **Changes** | **Contributor** |
| **V2** | **03.03.2021** | **XXX** | **Split in per-deliverable reports** | **Computomics** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

|  |
| --- |
| ***Disclaimer*** |
| Any dissemination of results reflects only the author’s view and the European Commission is not responsible for any use that may be made of the information it contains |

|  |
| --- |
| ***Copyright message*** |
| **© PARSEC consortium, 2020**  This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgment of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged. |

.

Table of Contents

1. xSeedScore EO module (IP held by Computomics) 4
   1. Interface 4
   2. Data sources 4
   3. Adaptation of xSeedScore 5

# xSeedScore EO module (IP held by Computomics)

We have developed program functions that allow us to automatically access some of the EO data sets outlined in deliverables reports 2 to 4. Input to these functions are the planting locations of a crop and the time range covering the planting seasons. Outputs are the aggregated EO descriptors in a form that allows us to include them in xSeedScore model training functions.

This functionality is flexible and can be extended to cover additional EO data types in the future.

## Interface

The EO modules defines a programming interface to access EO data from for xSeedScore. It defines input parameters and expected output values and their format. For each of the data sources outlined in section 1, a function implementing this interface is present in the EO module. The interface is designed to be generalisable to all kinds of EO data, which should make it straight forward to integrate additional data sources.

#### Input parameters:

* **GPS coordinates of location**
* **Start date**: For example the planting date. Optional/ignored for static data like HWSD data
* **End date**: For example the harvest date. Optional/ignored for static data like HWSD data
* **Aggregation method**: For example how values should be binned for averages, see Deliverable report 2
* **Variable name(s):** E.g. daily precipitation, daily maximal temperature, … from NOAA

#### Output value:

* A numerical vector of aggregated environmental descriptors. Categorial variables as for example “topsoil texture type” from HWSD should be encoded as numbers.

## Data sources

We included automatic access to the following data sources. Refer to deliverable reports 2-4 for details.

#### National Oceanic and Atmospheric Administration (NOAA) Weather Data

NOAA weather data is accessed with the NOAA API, and daily values for precipitation, minimal/maximal/average temperature can be downloaded. The data is aggregated according to the requested method.

#### Harmonized World Soil Database (HWSD) – Global Stable Soil Properties

We downloaded the HWSD database file that contains the soil profile description variables, together with the grid file that assigns soil profile to world surface grid cells. Since the HWSD is static and is not updated frequently, and file sizes are moderate, a local copy makes sense.

The EO module function for HWSD data ignores start and end dates, as well as aggregation method.

#### Copernicus Data Store (CDS) - Global Dynamic Soil Properties

CDS data is accessed through the CDS python library, and daily soil moisture measurement data is downloaded. The data is aggregated according to the requested method.

## Adaptation of xSeedScore

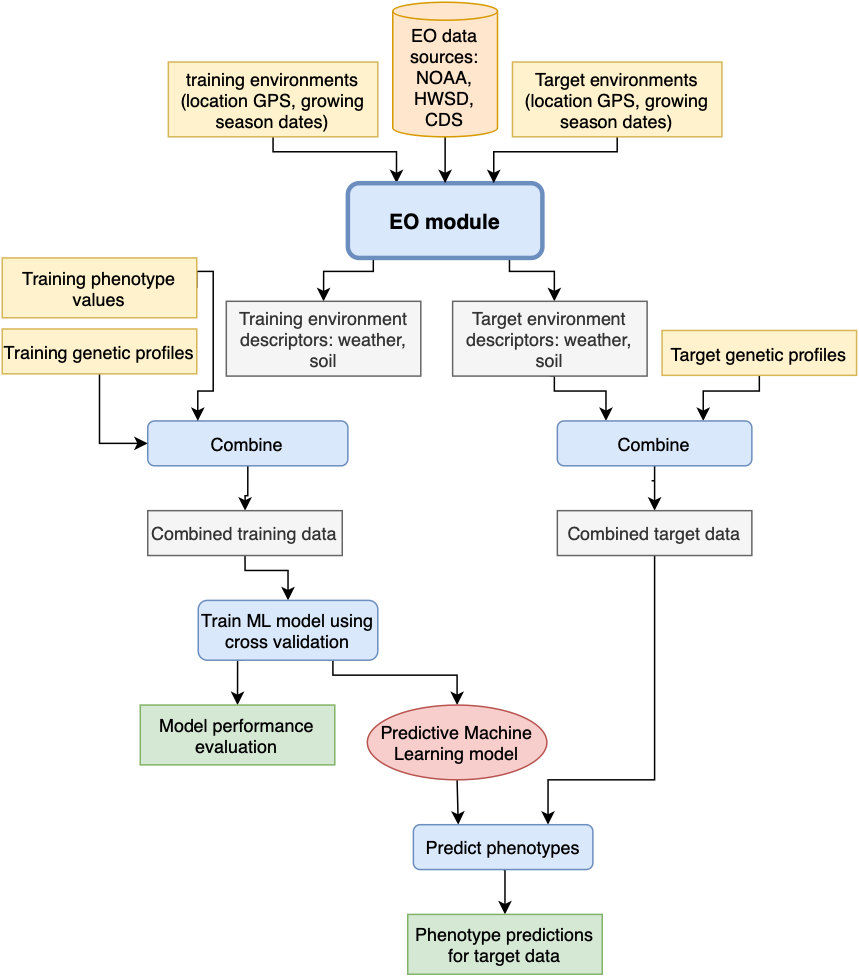


Figure 1: xSeedScore workflow with integrated EO module. Processing steps are shown in blue and with rounded edges. Input data is yellow; intermediate data is grey, results are green. EO data sources and the ML model are coloured orange and red, respectively.

We adapted our existing xSeedScore technology to be able to accept the output vectors of the EO module and combine the with the genetic profiles of the plant varieties. Each data point then consist of the genetic profile of a variety together with the description of a growing environment of the plant. xSeedScore can use these combined vectors to train machine learning models, and can predict the target values such as yield for a genetic profile in an environment. Also see our report on Deliverable 4.

In general, a customer project with xSeedScore follows the outline shown in Figure 1 and described below:

#### Customer input data

The customer provides the following data:

1. Training data:
   1. Genetic profiles of plant varieties from the breeding population
   2. GPS coordinates of the crop trial locations
   3. Dates of planting/harvesting times
   4. Phenotype measurements of the varieties from the trial, e.g. yield obtained in one location in one growing season
2. Target data:
   1. Genetic profiles of plant varieties for which phenotypes should be predicted
   2. GPS coordinates of the locations for which phenotypes should be predicted
   3. Dates of planting/harvesting times of seasons for which phenotypes should be predicted

#### xSeedScore workflow

Our xSeedScore technology follows these steps when the xSeedScore EO module is used:

1. Use **xSeedScore EO module** to download environmental descriptors for the *training* locations and growing seasons
2. Combine environmental descriptors with genetic profiles of training varieties and phenotype values
3. Train machine learning model
4. Validate model (also see Deliverable report 4)
5. Use **xSeedScore EO module** to download environmental descriptors for the *target* locations and growing seasons
6. Combine environmental descriptors with genetic profiles of *training* varieties
7. Predict phenotype values for *target* data using xSeedScore model

#### Customer deliverables

The customer will receive the following deliverables:

* A report describing model validation and EO descriptor visualisations (also see Deliverable report 3)
* Tables with phenotype predictions for target data