Earth observation data and Machine Learning to estimate industrial tomato yield response to climate change in Piacenza province, Italy

Magazzino, Sara (1); Quarta, Maria Luisa (1); Fazzini, Noemi (1); Houël, Maximilien (2); Santini, Monia (3); Pellegatta, Sofia (4); Carrillo, Rob (5)
Organisation(s): 1: MEEO S.r.l., Italy; 2: SISTEMA GmbH, Austria; 3: CMCC Foundation, Italy; 4: Alma Mater Studiorum - University of Bologna, Italy; 5: Trust-IT Services, Italy

Abstract

Due to climate change, industrial tomato crops have been prone to negative impacts at global scale on yield and phenological timing, creating a snowball effect that could affect food and economic security of countries and regions with significant agricultural production.

The case study "Food Security" implemented within the Horizon Europe project EO4EU and coordinated by MEEO with the support of CMCC, aims at demonstrating a methodology to identify the most impactful climate and spatial factors using Earth Observation data and Machine Learning (ML), to obtain useful information for the implementation of mitigation actions.

The area of interest is represented by four of the most productive Italian regions, with a focus on the Piacenza province. Climate indicators were identified to define optimal and stress conditions of tomato and they were correlated with agERA5 reanalysis climate data. The adverse climate conditions to the crop yield were identified also by studying and comparing the curve of NDVI values of tomato fields over three years (2021-2023) with a reference curve derived from aggregated satellite data. Initial results show that industrial tomato so far has not been affected by climate change on all fronts and it may expand its geographical range in the future. Further analyses will confirm or refute these preliminary results.

The described methodology is important for highlighting critical climate conditions of other crops. Better and more sustainable land management and climate change mitigation practices may also result through these new ML algorithms for yield projections with the integration of climate simulations at very high (km scale) resolutions.

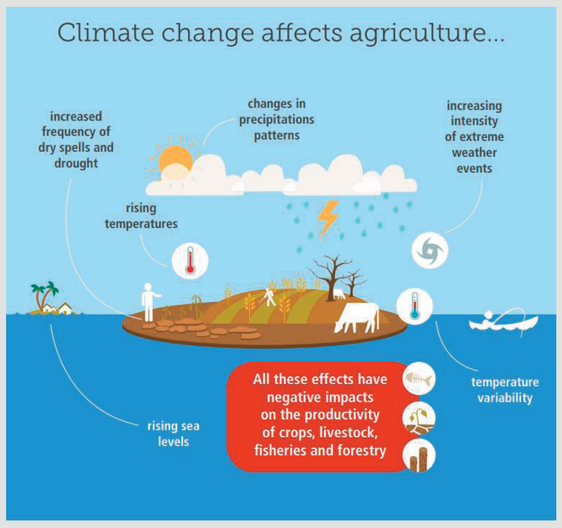
Climate change impact on crop yield

IPCC noted that climate change, with attached extreme weather events, has (and will have) significant impacts on agriculture.

Agricultural productivity is projected to decline between 3 and 16 percent by 2080, which may severely impact economies and societies, although there are large differences between territories and crops.

Therefore, ways of adapting to the impacts of climate change are needed including improved agricultural methods, landscape management and, of our interest, to the choice of crops suitable for the land in which they are grown.

Goal 2 of Agenda2030
sdgs.un.org/goals/goal2



Graphical representation of how climate change affect agriculture (From FAO 2016)
www.fao.org/3/i7084e/i7084e.pdf

EO4EU for food security

EO4EU is a European Commission funded innovation project bringing forward the EO4EU Platform which will make access and use of EO data easier for environmental, government, and even business forecasts and operations.

Use Case 3, which concerns various aspects of food security, is lead by SISTEMA, with the contribution of CMCC and MEEO.

The goal of this pilot is to analyze the impacts of climate change on agricultural crops and estimate the risk of yield loss over time through the use of satellite data and artificial intelligence. The chosen target crop is industrial tomato, because of its importance on the Italian territory (target area)

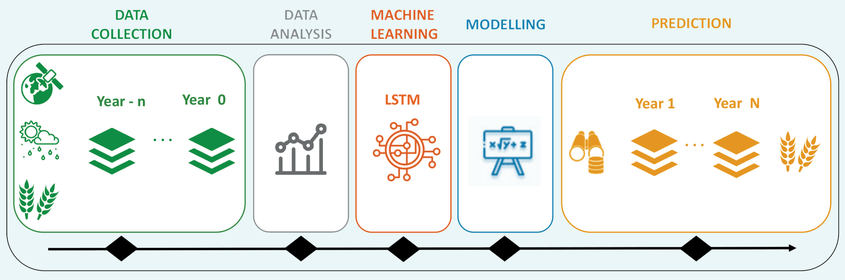
Data used:

- Copernicus ERA5
- Sentinel 2
- EURO-CORDEX - for climate simulation
- ISTAT data - for crop production data
- Parcels' polygon database AGREA

Industrial tomato, the target crop

Four italian regions, the target area

EO4EU UC3 workflow visualization



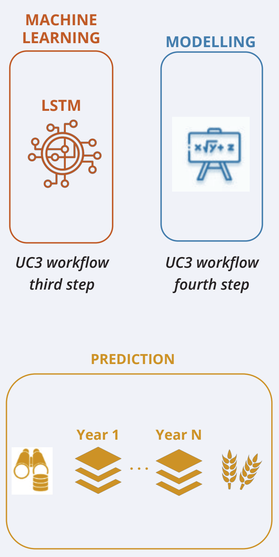
Early results and perspectives

Preliminary results:

- The most significant and most yield-correlated climatic indices on precipitation and temperature were identified.
- It was noted that the **amount of precipitation is not correlated in the same way among regions** due to different agronomic methods. Further information needs to be found to identify water availability.
- The descriptive **parameters of the provincial NDVI curves** most correlated with yield were identified. In this way, these parameters can be correlated with climatic conditions with more precise timing, increasing and improving the data at disposal.
- From the preliminary outcomes, it is clear that there is a **need to increase data availability** to obtain more accurate results.

Perspectives:

- The information obtained can be used to train **Machine Learning algorithms** in yield prediction through **regression models** and **neural networks**.
- With the use of **climate predictions** implemented by CMCC, yield can be predicted at different locations and in the future.



UC3 workflow third step

UC3 workflow fourth step

UC3 workflow fifth step

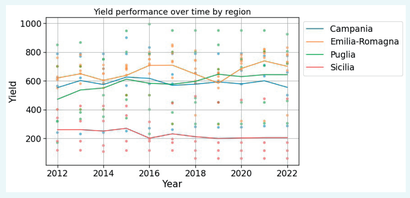
Climate indicators and crop yields correlation

Extraction of **yearly climate indicators** and comparison with the production statistical data including cultivated surfaces and crop yields

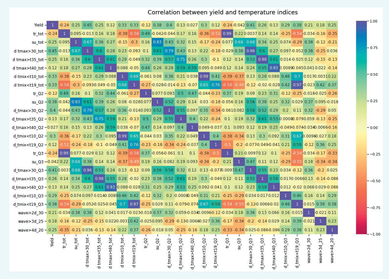
- **34 indices** were constructed **from hourly and daily precipitation and temperature** data during the months between **April and September** (beginning and end of the industrial tomato growing period). Seasonal indices (spring - Q2, and sumer - Q3) were constructed to better understand the relationship between yield and climate
- **Climate indices were correlated with yield** to understand which of these were the most significant and to detect further spatial differences.
- **Two classes (minimum yield and maximum yield)** were also defined to highlight differences in climatic conditions and to identify thresholds of index values that would ensure good yield. In this way, the most significant indices were defined.

Temperature indices:	Precipitation indices :
tr_tot,	Prec_cumulate_Q2
su_tot,	d_prec_Q2
d tmax>30_tot	prec_intensity_Q2
d tmax>t35_tot	Prec_cumulate_Q3,
d tmax>t40_tot	d_prec_Q3
d tmix<t10_tot	prec_intensity_Q3,
d tmix<t19_tot	Prec_cumulate_tot
tr_Q2	d_prec_tot
su_Q2	prec_intensity_tot
d_tmax>30_Q2	
d_tmax>t35_Q2	
d_tmax>t40_Q2	
d_tmix<t10_Q2	
d_tmix<t19_Q2	
tr_Q3	
su_Q3	
d_tmax>30_Q3	
d_tmax>t35_Q3	
d_tmax>t40_Q3	
d_tmix<t10_Q3	
d_tmix<t19_Q3	
wave>2d_30	
wave>3d_25	
wave>4d_20	
GGD	

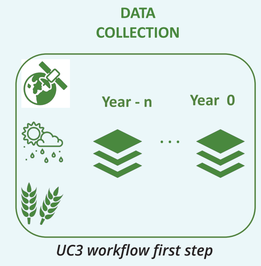
Table of indices name. 26 indices for temperature and 9 for precipitation, both based on a period between april and september



Tomato yield trend over time for the four target regions

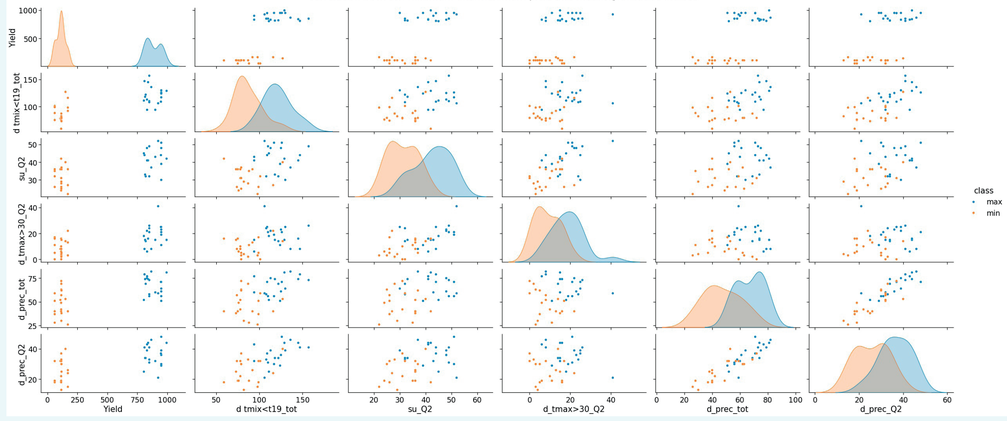


Heatmap correlation between temperature indices and yield

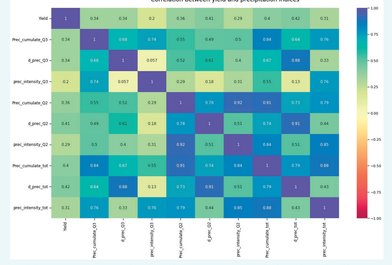


UC3 workflow first step

UC3 workflow second step



Pairplot correlation between selected indices and minimum and maximum yield



Heatmap correlation between precipitation indices and yield

Vegetation indexes and crop yield correlation

To evaluate the influence of climatic conditions on tomato plant physiology and yield we examined the **NDVI** (Normalized Difference Vegetation Index) trends of tomato fields. This study aimed to **gather production and impact insights in the absence of precise data on a territory basis**.

Data Collection:

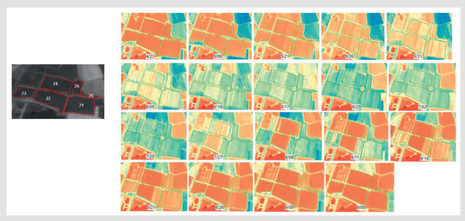
Initially, the fields cultivated with **tomato fields** from 2019 to 2023 were identified utilizing the AGREA database of the Emilia Romagna region. Then the **50 largest ones** have been selected. Subsequently, NDVI images from **Sentinel 2** data were obtained via the WEkEO Platform, along with **quality flags** indicating cloud cover.

NDVI Data Processing:

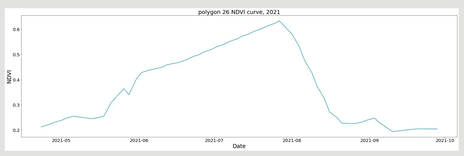
Each image was cropped according to the selected fields, and the **mean and median NDVI values** were computed for each field. Days with cloud cover exceeding 80% over a field were disregarded to prevent outliers. This resulted in a t**able of NDVI values for each date**.

NDVI Analysis:

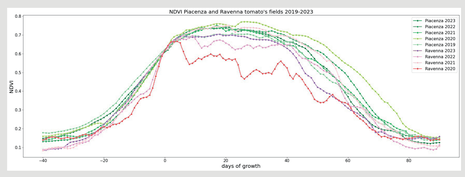
The data was interpolated to **estimate daily NDVI values**, and **trend curves were generated for each field**. Given that crops share the same growing season but exhibit **variations in start and end dates due to agricultural practices**, **provincial NDVI curves**, achieved through data normalization, were established to **correlate with the respective yield**. Additionally, an **NDVI reference curve** was formulated to facilitate comparisons across different NDVI curves, facilitating the detection of anomalies.



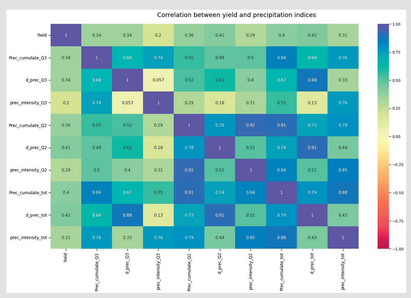
Example of NDVI satellite images in false colors during the tomato growing season in Piacenza province in 2021



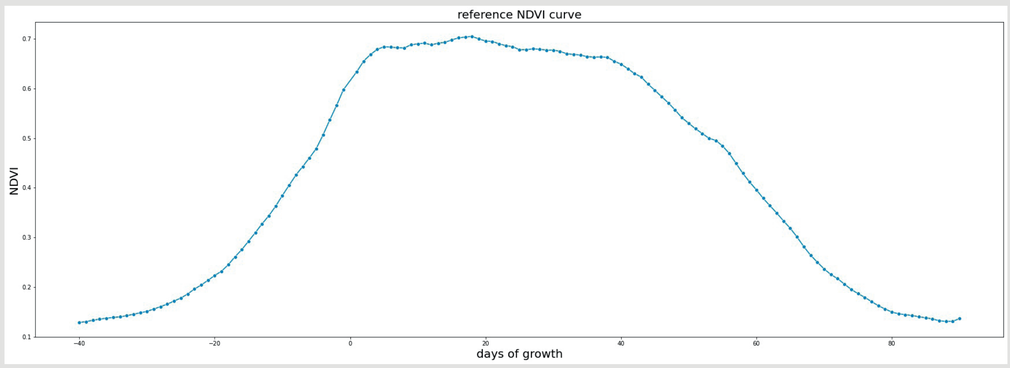
Example of NDVI curve of polygon 26, in the province of Piacenza in the year 2021



Comparison of various provincial NDVI reference curves



Correlation between parameters of provincial NDVI curves and the corresponding yield



NDVI reference curve describing the general physiological performance of the tomato plant during the growing season

Funded by the European Union