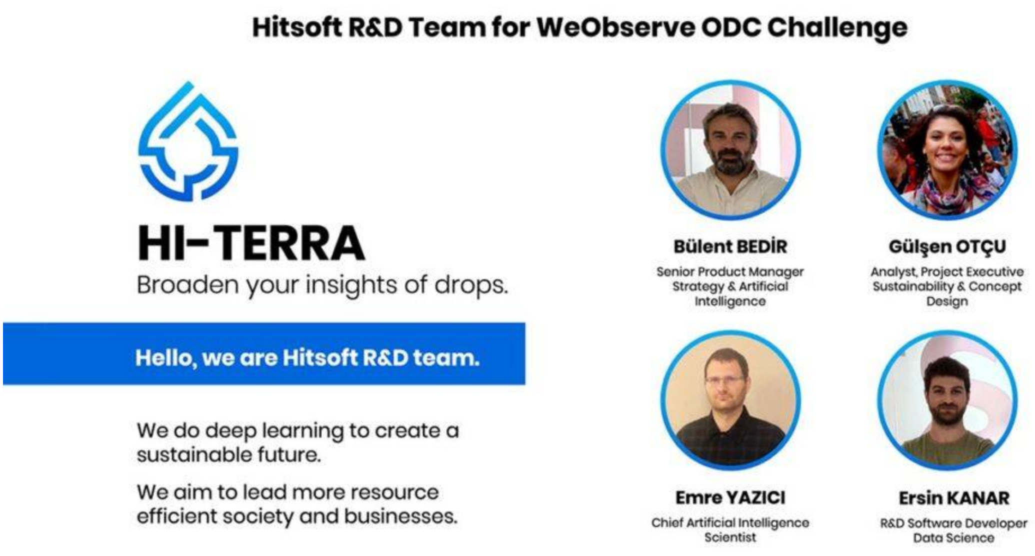
HI-TERRA

WeObserve Open Data Challenge – Final Report

May 15 – July 31

Team members

Gulsen Otcu, Bulent Bedir, Umut Şen, Ersin Kanar (Hitsoft R&D team)



Application functionalities

Hi-Terra is designed as a platform which will perform data processing to generate forecasts of soil moisture and watering. Since it has a dynamically learning capacity, the model is able to improve the forecast performance and to iteratively advance itself by using more data sets. It constitutes a sensitive, intelligent and reliable platform to produce forecasts for users both to get insights about soil moisture, watering time, amount of watering and to be notified about severe weather conditions, irrigation needs or water level anomalies. Hi-Terra provides resource efficient, cost effective and easy-to-use solution while taking its unique characteristic from the deep learning algorithms in its core.

The aim of the project is to model and forecast soil moisture based on weather conditions, soil characteristics and other parameters. The ultimate goal of Hi-Terra is to forecast soil moisture by the location in order to optimize watering, to boost smart resource use while reducing excess costs.

Hi-Terra Roadmap: ODC Phase

	TRL 1 BASIC PRINCIPLES	TRL 2 CONCEPT	TRL 3 PoC	TRL 4 LAB VALIDATION	TRL 5 VALIDATION
ACTIVITIES	Dataset selection New datasets Feature selection Idea creation Research about soil moisture and related parameters for water use Concept draft	GROW Data normalization Pre-processing Feature extraction Weather data integration Elaboration of use cases and application span User group selection Market analysis	Apply deep learning using LSTM model Train and Test Model soil moisture Improving accuracy Consolidate use-case and user groups Application scenarios Evaluation of potential	Analysis of soil datasets: ESDAC, ISRIC, Copernicus Improve model by using new features of soil type/ characteristics More train and test Improve accuracy Consolidate cases Stakeholder action plans	Apply pilot platform to make a show case Observe real-world condition and get results Develop UI /integrate Improve and consolidate the concept
METHODS	Top-down Bottom-up approach	Deep Learning method	Deep learning LSTM	*On-Going*	Improvement

During WeObserve Open Data Challenge (ODC) , our aim is to demonstrate feasibility in terms of technology, implementation scenarios, user and market uptake potential and scalability opportunity. We also analyzed the benefits and impacts as well as challenges or risks with the mitigation measures. The preliminary work enabled Hi-Terra to TRL 3 as an MVP and indicate promising outcomes which motivates us projecting the further steps. Hi-Terra, as an infrastructure, has a capability to be used within a wide spectrum of application areas from personal landscapes (gardens, yards), greenhouses, fields to golf courses, greens.

Hi-Terra is developed, trained and tested with Grow Soil Condition dataset and DarkSky Weather Dataset. It is able to generate forecast daily and weekly basis thanks to LSTM model of deep learning with the MAPE score 7.5%.

The open source repository is accessible via:
https://github.com/WeObserve/OpenDataChallenge-HiTerra/tree/master/final

Post-Challenge Roadmap

In the post-challenge phase, the impact of soil characteristics to soil moisture and system predictions is evaluated. According to the analysis, the topsoil conditions are found considerably effective to soil moisture. ESDAC[1], ISRIC[2], ESA[3] and Copernicus[4] databases are searched, analysed and dataset analysis are conducted. Especially ESDAC database has provided sets of relevant data as the primary parameter is determined as hydrogeological properties of soil. Then the soil types, characteristics and nutrient or contents has been assessed to implement in further steps.

	TRL 1 BASIC PRINCIPLES	TRL 2 CONCEPT	TRL 3 PoC	TRL 4 LAB VALIDATION	TRL 5 VALIDATION
ACTIVITIES	Dataset selection New datasets Feature selection Idea creation Research about soil moisture and related parameters for water use Concept draft	GROW Data normalization Pre-processing Feature extraction Weather data integration Elaboration of use cases and application span User group selection Market analysis	Apply deep learning using LSTM model Train and Test Model soil moisture Improving accuracy Consolidate use-case and user groups Application scenarios Evaluation of potential	Analysis of soil datasets: ESDAC, ISRIC, Copernicus Improve model by using new features of soil type/ characteristics More train and test Improve accuracy Consolidate cases Stakeholder action plans	Apply pilot platform to make a show case Observe real-world condition and get results Develop UI /integrate Improve and consolidate the concept
METHODS	Top-down Bottom-up approach	Deep Learning method	Deep learning LSTM	*On-Going*	Improvement

Topsoil Easily Available Water Capacity (EAWC) dataset which contains geolocation and eawc value, has been cleaned and processed by matching the GROW sensor locations. The data has been implemented to the Hi-Terra to train and test phases. The LSTM model has given the MAPE score of 6.9% which indicates a better forecast result and improving performance by the soil data.

Implementation considerations

The Concept: How it works

Hi-Terra forecasts the soil moisture and anomalies by using deep learning model called LSTM (Long Short-term Memory). The forecasts are based on measured data via GROW sensors and hourly weather data for the locations of the sensors. In addition, soil types and hydrogeological properties which are obtained from ESDAC (European Soil Database Center) are also considered to be used as features.

Hi-Terra is designed to be an intelligent platform to become an infrastructure to predict soil moisture and generate forecasts to lead users for better use of water and ensure the health of greens.

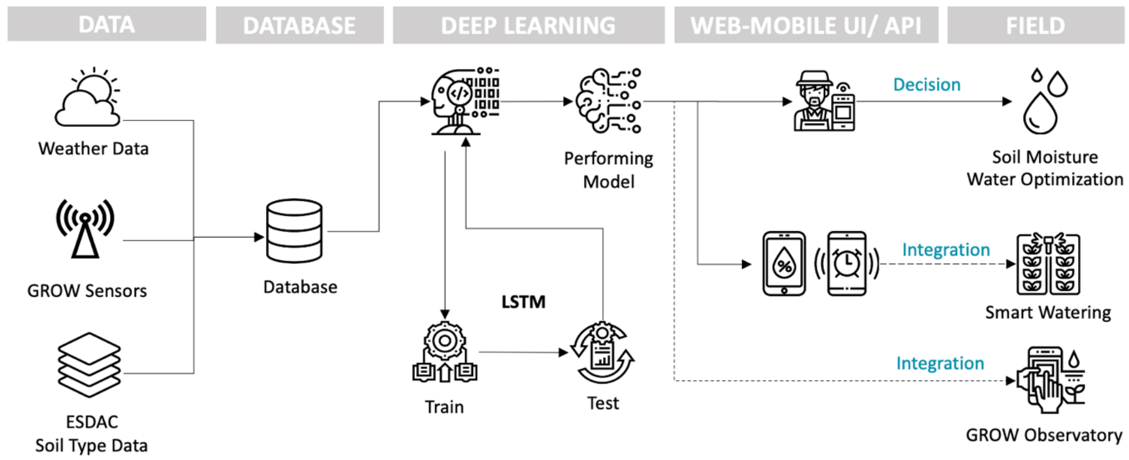
The concept has been built with the approach to understand, interpret and model soil moisture.

The core part of the data is the soil moisture data came from around 6.500 GROW sensors across Europe between 2017-2019. The second effective variable is weather which highly affects soil moisture is weather condition which shapes the model behavior. Third data set is soil type which enables to indicate correlations and effect on hydrographical profile of soil on

[1] European Soil Data Centre - https://esdac.jrc.ec.europa.eu/resource-type/datasets
[2] International Soil Reference and Information Centre - https://www.isric.org
[3] European Space Agency - https://climate.esa.int/en/odp/#/dashboard
[4] Copernicus - https://climate.copernicus.eu

moisture. When the data is conveyed to model, it processes the data on the basis of learning and improving continuously both by the data and the its forecast.

As an infrastructure, the performing model outcomes can be transferred via a User Interface (UI) or can be integrated through an API for the existing systems such as irrigation systems (smart watering), GROW Observatory App. Hi-Terra is planned to generate forecast of soil moisture and notify the user to make a decision to water the field, crops or greens. When it is integrated or embedded to existing systems, it can learn by time and generate forecasts for irrigation. If the water level is dropping, the alert will be effective to define when to water which will be optimized. Secondly, it will be able to warn the user in case of severe weather conditions or anomalies that occurs.



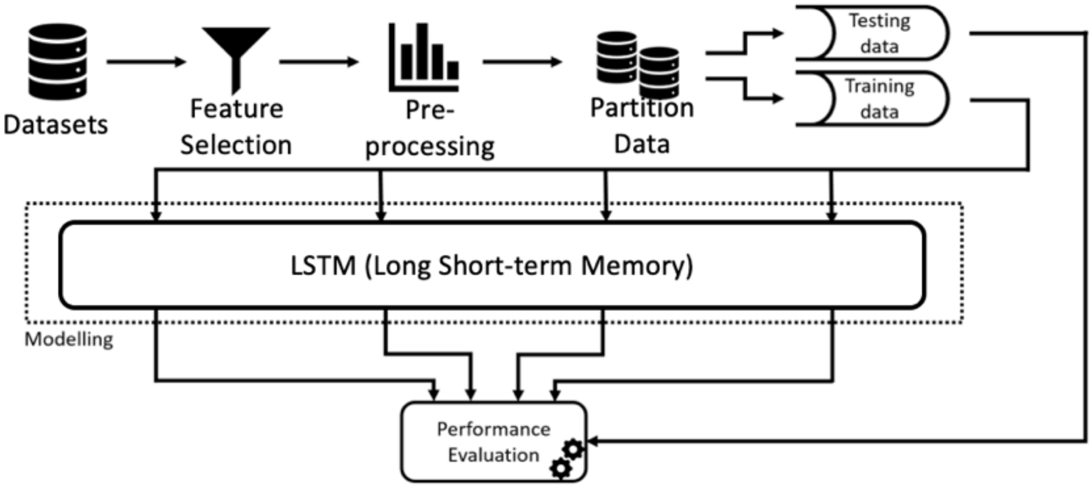
Methodology

Hi-Terra takes its unique characteristic from the deep learning algorithms in its core. Deep Learning, on the other hand, is just a type of Machine Learning, inspired by the structure of a human brain. Deep learning algorithms attempt to draw similar conclusions as humans would by continually analyzing data with a given logical structure. To achieve this, deep learning uses a multi-layered structure of model called neural networks.

Deep Learning models had been used in a vast amount of applications in recent years because of their high performance on large datasets. It is a key technology behind driverless cars, consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is achieving results that were not possible before which bring us the word, accuracy while identifying patterns, classifying different types of information and learning. We preferred deep learning to get high-accuracy forecasts and interpret the data in the best possible way. Besides, the characteristic of continuously learning mechanism and our approach helped us to generate the model.

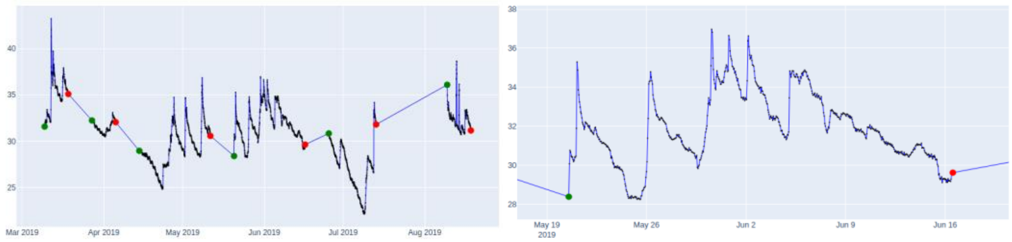
We used Long short-term memory (LSTM) architecture which is a special kind of artificial recurrent neural networks (RNN) and capable of learning long-term dependencies. LSTMs are explicitly designed to avoid the long-term dependency problem as they are remembering information for long periods of time is practically their default behavior rather than struggling to learn. Therefore, LSTM's characteristic behavior led us to choose it for GROW soil moisture data for modelling and achieving very well results of forecasts to go beyond interpretations.

Our process is standard deep learning method which is shown below:



WeObserve ODC Phase:

1. Data: GROW dataset
2. Features
   Used darksky to get **historical hourly** weather features:
   a. Humidity
   b. precip_probability
   c. precip_intensity
   d. Temperature
   e. cloud_cover
   f. apparent_temperature
   g. dew_point
   h. wind_speed
   i. wind_gust
   j. visibility
3. Data Cleaning and Normalization
   Sensors that collected moisture data had been closed for some time intervals. Our model works by taking hourly data as input sequentially, which prevents us to process missing data. We can handle this problem by using model's prediction as the true data for small intervals, but this method will result in problems with larger missing-intervals.
   So, we split data into time-series chunks that do not have missing measurements and worked with chunked data for both train and test phases. Below image depicts an example of moisture data and chunks are shown within green (representing beginning of a chunk) and red (representing end of a chunk) points. The image on the right side depicts a single chunk of this sensor.



4. Pre-processing
   a. **Measuring time:**
   Each sensor observation has a measuring date and time. Usually there is 15 minutes between subsequent observations. LSTM models expect sequential

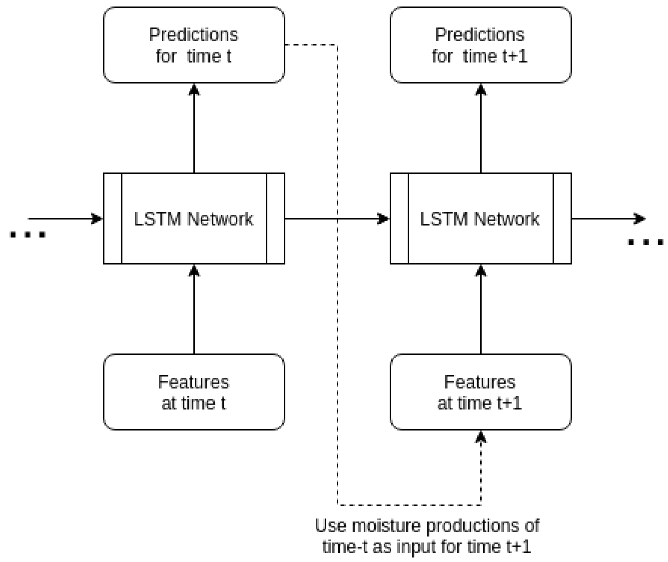
data but assumes the delay between observations are fixed. Therefore, we aggregated the moisture data for each hour, that is, we obtained the average of measures that are inside the hour as the measurement of that hour. Since moisture values contain some noise, this aggregation did not result in an information loss.

b. **Locations:**
Each sensor has a serial number and location information (latitude and longitude) in the dataset. Some sensors have the same location information, which probably indicates multiple sensors usage for a field. We averaged moisture values over sensors that are at the same location.

c. **Small chunks:**
We dropped chunks that have less than 100 hours

5. Modelling: LSTM
LSTM has an internal state that is updated at each time-step with the newly given features. We used the following features:
    - Previous-n weather measurements
    - Previous-n moisture measurements
    - Day, month, year of current time

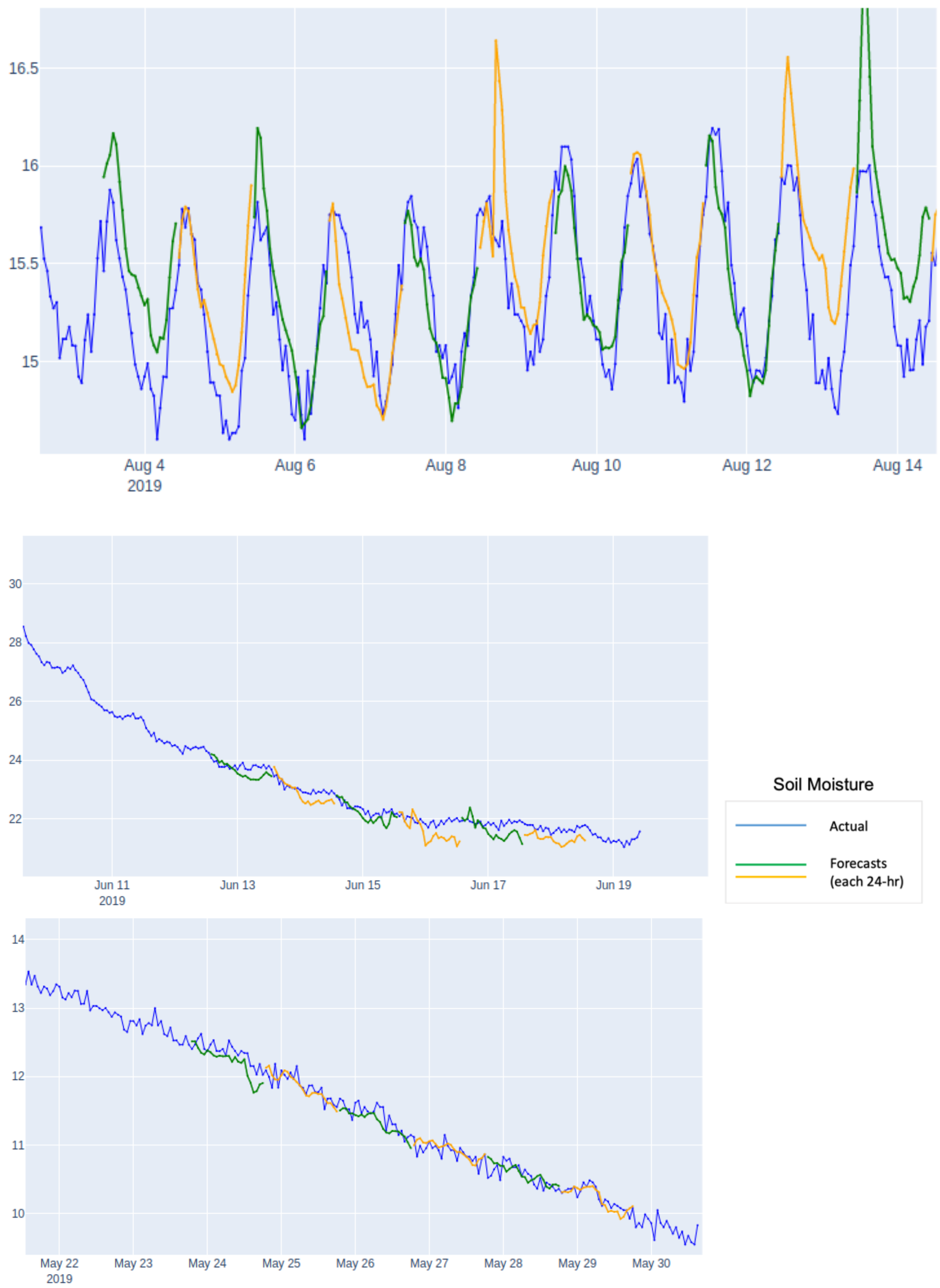
At test time, moisture predictions for previous n times before time (t+1) is used as the feature for time (t+1)



6. Experiment setup: Train & Test
We split the data randomly as train and test using location information. Some statistics of the dataset are given below:

	Train	Test	All
# of Locations	1000	274	**1274**
# of Chunks	3086	881	**3967**
# of Hours	1392707	388448	**1781155**

7. Pre-liminary Results: Obtained a MAPE error of 7.5%



Post-Challenge Period

After ODC, ESDAC, ISRIC and Copernicus Databases are assesed as one of the most effective data set Topsoil Easily Available Water Capacitiy of sensor locations as a starting parameter for soil characteristics.

1. Data: GROW dataset and DarkSky weather data
2. Features

   Used ESDAC Topsoil Easily Available Water Capacity (EAWC) to get **geolocational water retention** capacity feature:

   a. Latitude

b. Longitude
c. EAWC

3. Data Cleaning and Normalization
EAWC dataset that would be used, was obtained in Google Earth zipped file as .kmz format and converted to .kml (Keyhole Markup Language) format. In order to process the data, the file has been converted into .csv format. The data set contains X, Y data, also the points for creating a polygon with the data of "distance to next" column. Since the polygonal data has the point locations, we used only latitude and longitude columns to match with the closest sensor point. We conducted a cleaning process for the blanks.

4. Pre-processing
   a. **Locations:**
   On the main "GROW_moisture_weather – chunked" file, each sensor has a serial number named LocID. So, we had singularized each sensor based on LocID and we have taken the location information (latitude and longitude) in the dataset. In order to use EAWC dataset, we needed to combine both dataset by using location and EAWC information.
   b. **Distance Calculation:**
   We had generated a script for determining the closest distance with Grow sensor lat, lon data and EAWC measurement locations. The script had run to find the closest point. Based on the closest distance value, EAWC information is assigned and printed to a new file.
   c. **Merging:**
   Two datasets of GROW and EAWC_distance have been merged based on the location information. Therefore, final dataset of GROW_moisture_weather_EAWC chunked has been obtained.

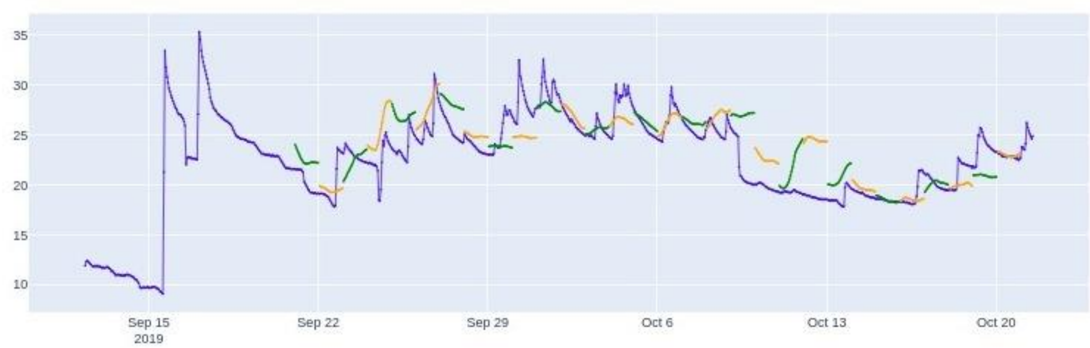
5. Modelling: LSTM
The existing model has been taken.

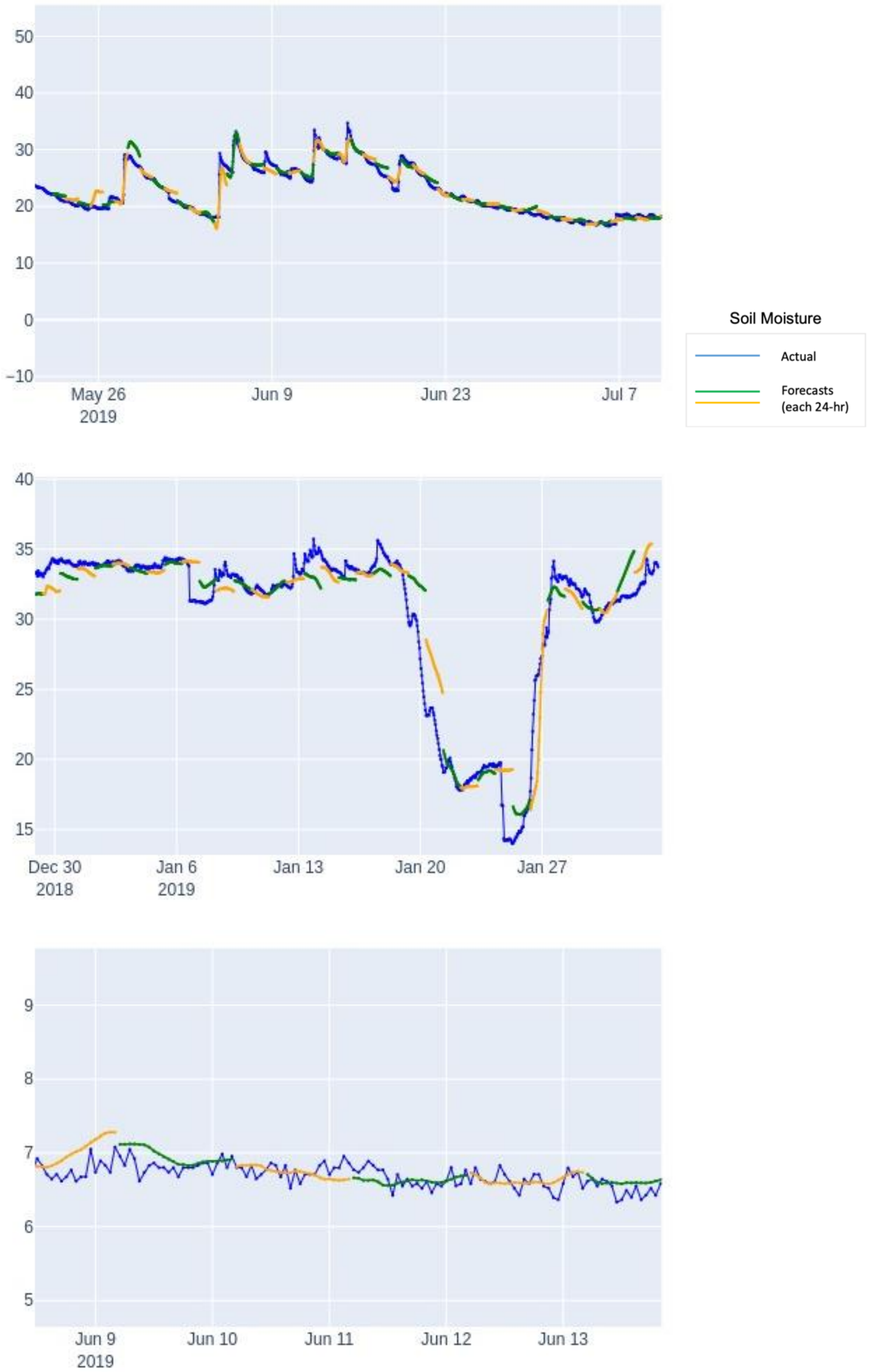
6. Experiment setup: Train & Test
We split the data randomly as train and test.

7. Results: Obtained a MAPE 6.9%

```
MAPE IS: 0.06927112961874897
```

```
tester.visualise_rand()
```




50
40
30
20
10
0
−10
May 26
2019
Jun 9
Jun 23
Jul 7
Soil Moisture
Actual
Forecasts
(each 24-hr)
40
35
30
25
20
15
Dec 30
2018
Jan 6
2019
Jan 13
Jan 20
Jan 27
9
8
7
6
5
Jun 9
2019
Jun 10
Jun 11
Jun 12
Jun 13

Expected impact

Hi-Terra, as an infrastructure, has a capability to be used within various application areas from personal landscapes (gardens, yards), greenhouses, fields to golf courses, greens. From small to large application areas, the key point is the use of artificial watering that has been done through automatic or semi-automatic irrigation systems, in some cases hand/wild watering. Most of the irrigation systems either requires human intervention to start/stop or has periodic watering approach. In all cases, the water optimization is key for both plant and soil health and saving water. Therefore, the specific market segment and users/user groups are defined in terms of Hi-Terra's forecasts to be effective.

Beneficiaries	All stakeholders and "Earth"
Contributors	Academicians and Researchers Scientific Organizations Society / Citizens
User Groups	Farmers Land-Owners/ Greenhouse Owners Golf course owners Municipalities / Utility Companies Policy Makers

Users/User Groups

Farmers/Land Owners/Greenhouse Owners: Their needs;

- Technologies for resource optimization
- When to irrigate and how much water needed
- Better knowledge and activities about soil, fields
- Low cost of investment and minimizing their costs.
- Increase yield by time

Our aim is to help water use reduction, moist and fertile soil, optimized resource use, decisions on reliable predictions and real-time data. Therefore, our reason to choose them due to their high interest and more impact potential on both agricultural practices, resources and costs.

The benefit of Hi-Terra will be a technology solution tells when to irrigate and how much water to use with its soil condition prediction. They will be able to get alerts about expected weather conditions and water levels. In order to help to ensure crop development and protect soil properties with a small investment costs and a relatively minimum data requirement.

Golf course owners: Golf is a specific niche that we aim to address through Hi-Terra. Golf has a rising popularity as the number of golf courses are increasing recently. At the same time, the maintenance of these green courses is critical as one hectare of a golf course requires 10.000 – 15.000 tons of water yearly which is equal to 12.000 people's annual water consumption. In order to optimize water use and increase the efficient resource consumption, Hi-Terra will generate benefit for them to forecast based water management platform. Secondly, the over watering increases the soil moisture dramatically and damages the soil texture as well as the grass which multiplies the loss for golf course owners. By using the intelligence of Hi-Terra, they will be able to know when to water and further how much to water. So, they will be able to cut excess costs, reduce their water use and save time/effort to sustain the golf courses.

Municipalities & Utility Companies: The municipalities, governments and utility companies need to regulate the water management since the water resource optimization, water reserve protection and reducing costs are very basic and critical touch points. Considering environmental aspect, they also need to reduce environmental impact by protecting natural resources, specifically water. They also need to take actions to increase the security of food chains and sustainability of activities.

Hi-Terra aims to provide them a continuously improving, flexible and reliable infrastructure to reduce costs, reduce water use, monitor the impacts of environmental phenomenon. They will be able to understand, interpret data in short term and get forecasts to make better action plans in mid-term to create more positive impact on natural systems and society.

Therefore, we choose them because of their coverage can escalade the impact of Hi-Terra dramatically and the need is more critical which indicates a high interest.

Policy Makers: Regarding to the issue and the need for new policies and regulations for climate change, Hi-Terra aims to provide a contributing perspective for governments and public departments who set enabling policy and regulatory frameworks to create the adapting, incentivizing and action-taking plans. Considering the smart resource use, land-use and agrifood sector requirements, we would like to contribute pro-active processes for policy makers to act in an agile way.

Contributors

Academicians, Researchers & Scientific Organizations: They will be beneficiaries but also a user group of Hi-Terra naturally. Because this infrastructure can be an intelligent platform for them to enrich their researches and scientific. In longer terms, they will be able to gather more data about the trends across Europe or more locations to detect the reasons, impacts and changes to monitor and analyze holistically. Therefore, both in scientific and humanity level, Hi-Terra will generate benefits and contribution.

Society: As people are more sensitive to nature, meaning and benefits of the solutions, HiTerra will be a sustainable platform to boost its benefit from three main dimensions such as water, greens and relatedly agri-food. In these terms, they are our beneficiary group which will be affected in a positive way from local to wider locations and from short to long term.

Besides, the solution has a capability to increase the technological uptake towards data-driven society. The new jobs and employment potential are also plus for job creation. The nature of Hi-Terra encourages a multi-disciplinary work since the new positions will generate or boost new talents to find an environment to flourish.

Beneficiaries

All stakeholders: We expect to take all internal and external stakeholders have the benefit of the system both directly and indirectly on the dimensions of water use, greens and relatedly agri-food. In these terms, they are our beneficiary group which will be affected in a positive way from local to wider locations and from short to long term on the way of achieving the sustainable development. Considering the UN's SDGs, we aim to address Climate action, Responsible consumption and production, Life on land and Industry, innovation & infrastructure goals with the provided platform. While tackling climate change and societal challenges on climate, food, land and resources, related to SDGs, the platform has the potential to provide sustainable, collaborative and contributing system for our common future.

Future outlook

During WeObserve Challenge, we had a remarkable experience about Citizen Science and climate action to create impact in a collaboration while exploring deeper perspectives about soil, resources, sustainable development to generate benefit. In a data-driven world, it is closely related to the attributes, moreover our attitudes which causes alterations in the system and impact for overcoming the climate change consequences while building our common future. So, we believe that citizen science has a great value and power as the small efforts count to tackle the challenges, to gather data to achieve meaningful insights and to spread the impact from a person to the world where everything is connected.

This era of change reveals the new opportunities and the challenges hand in hand to rethink how to grow, share and consume. The key is the emerging and cutting-edge technologies that are leverages to adapt, to participate to the solution and create positive impact. Therefore, we aim to contribute sustainable development in an understanding of big picture as well as bottom-up approach to increase comprehensiveness. We also believe the importance of collective action to innovate and find new ways to adapt and reduce the impacts of climate change phenomena. In order to create a sustainable system, our project includes various aspects. During ODC and after, it has been a great opportunity to experience to observe and analyze. After starting with one specific goal of "forecasting soil moisture", the project and our approach evolved considerably. We discovered how it is a strong parameter to dive deeper and even how one parameter affects the multiple systems. In these terms, the dynamism of technology and society-driven landscape indicate opportunities to redesign, discover and innovate. Now, we become able to put a vision to create better results and positive impact.

In this regard, we would like to advance our project Hi-Terra to the upper levels as the specific objectives are defined for further stages, as follows:

Technical Objectives	• To improve LSTM model by longer train and test periods • To include more ESDAC Soil Property datasets • To assess the usability and implementation of Copernicus data • To include more datasets which have relevantly extracted features and parameters of citizen-science data, crop/biome, soil texture, images, etc. • To increase accuracy and reliability of forecasts • To try applying methods to extend the analysis scope • To answer how much water to use for watering/irrigation • To extend the application spectrum flexibly • To make integration and embedded-use cases in real world conditions.
Business & Market Objectives	• To develop an infrastructure to be used to monitor and predict soil moisture • To expand use cases • To extend user and customer range • To improve the market proposition • To determine cost efficiency levels and income models
Environmental & Social Approach	• To boost sustainability characteristic • To be able to concretize measurable societal and economic impact • To show how to address SDGs with measurable preliminary results.

- END OF DOCUMENT -

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no 776740.